Thiomicrospira crunogena:
A Chemoautotroph
With a Carbon Concentrating Mechanism

by

Kimberly P. Dobrinski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Biology
College of Arts and Sciences
University of South Florida

Major Professor: Kathleen M. Scott, Ph.D.
James Garey, Ph.D.
Valerie Harwood, Ph.D.
John Paul, Ph.D.

Date of Approval:
July 13, 2009

Keywords: Thiomicrospira crunogena, carbon concentrating mechanism, chemoautotroph, carbon fixation, carbonic anhydrase

©Copyright 2009, Kimberly P. Dobrinski
Dedication

Thank you Mom (the first scientist in the family) for fun discussions about Biology and all your encouragement. A warm thank you to Dad, Mike, Aunt Sallie and Uncle Jim for unending support. Also thank you Cathy, Joe, Donna and all my family and friends for believing in me. Thank you Sondra for being the friend of a scientist.

I give a very special thank you to my beautiful children: Lenna, Jason, Michael (and you too Matt) for all the sacrifice you endured so that I could follow my dream of becoming a scientist. Thank you Emma for always making me smile!

My most special gratitude goes to Joe who has always believed in me and worked two jobs (sometimes more) to support my science “habit”. Without his love and support a teenage mom could have never achieved a Ph.D. This is all for you!
Acknowledgements

I owe interminable gratitude to Dr. Kathleen Scott who provided me with guidance, patience, and unending support. She believed in me enough to let me take on RNA, microarrays and qRT PCR on my own. Because of her belief in me, I worked hard to “push back the frontiers of science”; and it is because of her I believe I can now call myself a scientist. (She also has great engineering skills – she makes a mean flowered arch!) I thank her for all the hard work and all the good times.

I would like to thank Dr. Valerie Harwood for all the knowledge she shared with me inside and outside of the classroom and the confidence she showed in me. I gained from Dr. Harwood a great love and understanding of Microbiology and also great confidence in myself. Thank you to Dr. John Paul for all that you taught me not only about viruses and bacteria but also a love of Microbial Ecology. Thank you also to Dr. James Garey for stimulating conversations and kind advice about my work. I give a warm thank you to Amanda Boller, Phaedra Thomas, Kristy Menning, Ishtiaque Quasem, Vedad Delic, Rachel Ross, Chantel Barrett and Kelly Fitzpatrick for good times and hard work on Scott Island, exciting ASM meetings and also for interesting Journal Club discussions. Thank you Jessica Moore for encouraging my eukaryotic interests and love of cancer research. I will miss our talks. A special thank you to Sean Yoder and Dr. Steve Enkemann who provided me with a bench to work on
and constructive feedback to my questions. Without their guidance *Thiomicrospira crunogen*a microarrays would have never been a reality. Thank you Dr. Richard Pollenz for always encouraging me. Thank you Dr. Jeffrey Yoder for the first year and a half of my Ph.D. I learned a tremendous amount in that time. Thank you Jennifer Montgomery, Matt Buzzeo, and Patoula Panagos for insight and assistance at Yoder lab. Thank you to Dr. Lee Adair for giving me my first real glimpse of research and instilling in me a love of science that will carry me through the rest of my life.

My work was supported by the National Science Foundation, DOE/JGI and the USDA.
Table of Contents

List of Tables iii
List of Figures iv
Abstract vi

Chapter 1 Background 1
  Hydrothermal vents 2
  Carbon concentrating mechanisms 2
  Dissolved inorganic uptake 3
  RubisCO 3
  Carboxysomes 5
  Carbonic anhydrase 6
  Thiomicrospira crunogena 6
  References 9

Chapter 2 The Carbon Concentrating Mechanism of the
  Hydrothermal Vent Chemolithoautotroph Thiomicrospira crunogena 14
  Abstract 15
  Introduction 16
  Materials and Methods 18
  Results 24
  Discussion 29
  Acknowledgements 32
  References 33

Chapter 3 Expression and Function of Three Carbonic Anhydrase
  Enzymes in Thiomicrospira crunogena 37
  Introduction 37
  Materials and Methods 40
  Results 48
  Discussion 53
  References 58

Chapter 4 Transcriptone Response in Thiomicrospira crunogena
  to the Dissolved Inorganic Carbon Concentration 62
  Introduction 62
  Materials and Methods 64
  Results and Discussion 70
List of Tables

Table 1. Primers and probes used to target *T. crunogena* carbonic anhydrase genes 46

Table 2. Response of *T. crunogena* carbonic anhydrase gene transcription to the DIC concentration available during growth 48

Table 3. Activity of heterologously expressed carbonic anhydrase genes from *T.crunogena* 49

Table 4. Primers and probes used to target *T. crunogena* conserved hypothetical genes via qRT-PCR 68

Table 5. Genes transcribed more frequently in cells grown under high-DIC conditions 72

Table 6. Genes transcribed more frequently in cells grown under low-DIC conditions 72

Table 7. Response of conserved hypothetical gene transcription to changes in growth conditions 76
List of Figures

Figure 1 Transmission electron micrograph of *T. crunogena* cells 7

Figure. 2. Steady-state growth rate constants and DIC concentrations for *T. crunogena* 24

Figure. 3. Carbon fixation rates for *T. crunogena* under DIC limitation 25

Figure. 4. Carbon fixation rates for *T. crunogena* harvested from DIC-limited chemostats supplied with CO$_2$ and HCO$_3^-$ 26

Figure. 5. Intracellular DIC accumulation by *T. crunogena* cultivated under DIC limitation 28

Figure 6. Energy-dependence of intracellular DIC accumulation by *T. crunogena* 29

Figure 7. (A.) SDS PAGE and (B.) Western blot depicting heterologous expression of *T.crunogena* carbonic anhydrase in *E.coli*. 49

Figure 8. Effect of acetazolamide and ethoxyzolamide on carbonic anhydrase activity of crude extracts prepared from *E. coli* 50

Figure 9. Carbonic anhydrase activity and inhibition in *T. crunogena* crude extract. 51

Figure 10. Carbonic anhydrase activity and inhibition in *T. crunogena* whole cells. 51
Figure 11. Effect of acetazolamide and ethoxyzolamide on DIC uptake and fixation by *T. crunogena* 52

Figure 12. Effect of AZA and EZA on the intracellular DIC pool and carbon fixation in intact low- and high-DIC *T. crunogena* cells. 53

Figure 13. Model of *T. crunogena* cell with carbonic anhydrase locations and inhibitor sensitivities 57

Figure 14. A sampling of housekeeping genes that do not have a measureable change in transcription under low- versus high-DIC conditions 70

Figure 15. Genes that have increased transcription under high-DIC conditions 74

Figure 16. Genes encoding conserved hypothetical proteins that have increased transcription under high-DIC conditions 75

Figure 17. Carboxysome genes that are transcribed more frequently under low-DIC conditions 75

Figure 18. Transcription of carbonic anhydrase-encoding genes under low- and high-DIC conditions 77

Figure 19. Genes encoding conserved hypothetical proteins that have increased transcription under low-DIC conditions 78
Thiomicrospira crunogena: A Chemoautotroph With a Carbon Concentrating Mechanism

Kimberly P. Dobrinski

ABSTRACT

Gammaproteobacterium *Thiomicrospira crunogena* thrives at deep-sea vents despite extreme oscillations in the environmental supply of dissolved inorganic carbon (DIC; \(=\text{CO}_2 + \text{HCO}_3^- + \text{CO}_3^{2-}\)). Survival in these habitats is likely aided by the presence of a carbon concentrating mechanism (CCM). Though CCMs are well-documented in cyanobacteria, based on this study *T. crunogena* is the first chemolithoautotroph to have a physiologically characterized CCM. *T. crunogena* is capable of rapid growth in the presence of 20 \(\mu\text{M}\) DIC, has the ability to use both extracellular \(\text{HCO}_3^-\) and \(\text{CO}_2\), and generates intracellular DIC concentrations 100-fold greater than extracellular, all of which are consistent with a CCM analogous to those present in cyanobacteria. Interestingly, however, the *T. crunogena* genome lacks apparent orthologs of many of the components of the cyanobacterial CCM (e.g., \(\text{HCO}_3^-\) transporters). However, despite this lack, several candidate genes were identified during genome annotation as likely to play a role in DIC uptake and fixation (three carbonic anhydrase genes: \(\alpha\text{-CA}\), \(\beta\text{-CA}\) and \(\text{csoSCA}\), as well as genes encoding three RubisCO enzymes: *cbbLS*,
CScbbLS, and cbbM, which encode a cytoplasmic form I RubisCO, a carboxysomal form I RubisCO, and a form II RubisCO, respectively).

In order to clarify their possible roles in DIC uptake and fixation, \(\alpha\)-CA, \(\beta\)-CA and csoSCA transcription by low-DIC and high-DIC \textit{T. crunogena} were assayed by qRT PCR, heterologous expression in \textit{E. coli}, and potentiometric assays of low-DIC and high-DIC \textit{T. crunogena}. Transcription of \(\alpha\)-CA and \(\beta\)-CA were not sensitive to the DIC concentration available during growth. When overexpressed in \textit{E. coli}, carbonic anhydrase activity was detectable, and it was possible to measure the effects of the classical carbonic anhydrase inhibitors ethoxyzolamide and acetazolamide, as well as dithiothreitol (DTT; recently determined to be a carboxysomal CA inhibitor). The \(\alpha\)-CA was sensitive to both of the classical inhibitors, but not DTT, \(\beta\)-CA was insensitive to all inhibitors tested, and the carboxysomal carbonic anhydrase was sensitive to both ethoxyzolamide and DTT. The observation that the CA activity measurable potentiometrically with intact \textit{T. crunogena} cells is sensitive to classical inhibitors, but not DTT, strongly suggests the \(\alpha\)-CA is extracellular. The presence of carbonic anhydrase activity in crude extracts of high-DIC cells that was resistant to classical inhibitors suggests that \(\beta\)-CA may be more active in high-DIC cells. Incubating cells with ethoxyzolamide (which permeates cells rapidly) resulted in inhibition of carbon fixation, but not DIC uptake, while incubation with acetazolamide (which does not permeate cells rapidly) had no apparent effect on either carbon fixation or DIC uptake. The observations that inhibition of \(\alpha\)-CA has no effect on DIC uptake and fixation, and that the \(\beta\)-CA is not transcribed
more frequently under low-DIC conditions, make it unlikely that either play a role in DIC uptake and fixation in low-DIC cells. Further studies are underway to determine the roles of α-CA and β-CA in *T. crunogenae*.

To assay the entire genome for genes transcribed more frequently under low-DIC conditions, and therefore likely to play a role in the *T. crunogenae* CCM, oligonucleotide arrays were fabricated using the *T. crunogenae* genome sequence. RNA was isolated from cultures grown in the presence of both high (50 mM) and low (0.05 mM) concentrations of DIC, directly labeled with cy5 fluorophore, and hybridized to microarrays. Genes encoding the three RubisCO enzymes present in this organism demonstrated differential patterns of transcription consistent with what had been observed previously in *Hydrogenovibrio marinus*. Genes encoding two conserved hypothetical proteins were also found to be transcribed more frequently under low-DIC conditions, and this transcription pattern was verified by qRT-PCR. Knockout mutants are currently being generated to determine whether either gene is necessary for growth under low-DIC conditions. Identifying CCM genes and function in autotrophs beyond cyanobacteria will serve as a window into the physiology required to flourish in microbially-dominated ecosystems where noncyanobacterial primary producers dominate.
Chapter 1

Background

Before 1.8 billion years ago (Gyr) atmospheric CO$_2$ concentrations were more than 100 times greater than they are today (33). It was in this high CO$_2$ environment that the three domains of life diverged (5), and autotrophic members of all three domains provided the input of organic carbon into microbial food webs. Between 2.45–1.85 Gyr ago, cyanobacterial photosynthesis caused atmospheric oxygen levels to rise allowing the shallow oceans to become mildly oxygenated while the deep oceans remained anoxic. Oxygen levels remained at the same levels until 0.54 Gyr ago, when the shallow oceans were oxygenated and the deep oceans fluctuated between oxic and anoxic conditions (18).

As oxygen levels increased, a corresponding decrease in CO$_2$ occurred due to cyanobacterial carbon fixation (14). With the drop in atmospheric CO$_2$, organisms that had once thrived in its abundance now had to adapt to its relative scarcity. Autotrophic microorganisms have successfully adapted, and continue to fulfill the role of primary producer in diverse habitats from hydrothermal vents to acid mine drainage and from terrestrial habitats to the open ocean (14). Substantial variation in the chemistry of these habitats presents these organisms with pH values ranging from 1-14, and DIC (dissolved inorganic carbon, $= CO_2 +$
HCO$_3^-$ + CO$_3^{2-})$ concentrations ranging from extraordinarily abundant (>20 mM) to extremely low (<20 µM; 37, 10, 12).

**Hydrothermal Vents**

One environment that is particularly challenging due to extreme temporal heterogeneity is the hydrothermal vent environment. In this environment, warm hydrothermal fluid is emitted from cracks in the crust (17). This hydrothermal fluid is anoxic, contains hydrogen gas, hydrogen sulfide, and other reduced compounds (22). It is also acidic (pH ~5; 25) and has an elevated DIC concentration (5-6 mM; 10). When this warm (~30-40ºC) fluid meets alkaline (pH 8), oxic seawater at the bottom of the ocean (2ºC), it creates turbulent eddies due to the differences in temperature (13, 22). These eddies cause large changes in temperature, and corresponding chemistry (e.g., pH, DIC, sulfide, oxygen), with oscillations occurring over timescales ranging from seconds to hours (22). Under these conditions the predominant form of DIC is HCO$_3^-$, with concentrations of 5 to 7 mM (13, 22). The pH fluctuations cause dramatic ebbs and spikes in CO$_2$ concentrations, which vary from 20 to 2000 µM (13). One would expect that this habitat would result in a selective advantage for autotrophic organisms having adaptations to maintain a steady supply of CO$_2$.

**Carbon Concentrating Mechanisms**

One key adaptation that enhances microbial carbon fixation in chronically or episodically low CO$_2$ and/or HCO$_3^-$ concentrations is a carbon concentrating mechanism (CCM). CCMs have been extensively studied in cyanobacteria, but
are likely to be common beyond this phylum. Cyanobacteria with CCMs typically demonstrate a higher whole-cell affinity for DIC when grown under low DIC conditions (3). Cyanobacteria that are cultivated under low DIC conditions capture HCO$_3^-$ by using high affinity transporters to deliver it to the cytosol (3). Within the cytosol, this intracellular HCO$_3^-$ pool is consumed by carboxysomes. Carboxysomes have a protein shell and contain a trace of carbonic anhydrase (CA) which facilitates the conversion of HCO$_3^-$ to CO$_2$ within the shell (3, 24, 39). Carboxysomes are also packed with form I ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), which fixes the majority of CO$_2$ before it can diffuse back into the cytosol (3).

**Dissolved Inorganic Carbon Uptake**

Among cyanobacteria, there are three known phylogenetically independent HCO$_3^-$ transporters. BCT1 is an ABC-transporter (34), while SbtA is a Na$^+$-dependent HCO$_3^-$ transporter (44). The third transporter, SulP, is evolutionarily related to sulfate transporters (40).

Some cyanobacteria have an additional adaptation, CO$_2$ traps, to prevent the loss of DIC transported to the cytosol by these HCO$_3^-$ transporters. In these cells, any CO$_2$ that escapes from the carboxysomes is reconverted to HCO$_3^-$ in the cytoplasm, where a thylakoid-associated complex couples CO$_2$ hydration to electron transfer from NAD(P)H to plastoquinone (44, 29, 3).

**RubisCO**
A carbon concentrating mechanism raises intracellular inorganic carbon concentrations, which facilitate carbon fixation by RubisCO, which is a poor catalyst (54). RubisCO is the carboxylase of the Calvin-Benson-Bassham (CBB) cycle. This enzyme has rather low $k_{cat}$ values (55) and can use both $CO_2$ and $O_2$ as a substrate (55, 6). The oxygenase reaction, in which ribulose 1,5-bisphosphate (RuBP) is oxygenated rather than carboxylated, is wasteful for the cell and requires energetic expenditure to regenerate the RuBP necessary to keep the CBB cycle functioning (9).

Imperfect as it is, this enzyme is found in a wide variety of microorganisms and can be subdivided into three forms (I – III; 60). Form I RubisCO consists of two subunits (large and small), encoded by $cbbL$ and $cbbS$, and is further subdivided into four groups (IA – ID; 59, 60). Form IA RubisCO is found in many autotrophic proteobacteria and some marine cyanobacteria (46) while form IB is found in other cyanobacteria and in green plastids (60). Form IC is found in some facultatively autotrophic proteobacteria (60), while form ID is present in many marine eukaryotic algae, including diatoms, coccolithophores, and many dinoflagellates (60). Form II RubisCO consists of only one type of subunit (CbbM), which is evolutionarily related to the large subunit of form I RubisCOs (35, 15, 7, 52, 53, 55).

Many microorganisms have multiple RubisCO genes present in their genomes (9, 1, 59). Unlike cyanobacteria, which only have a single RubisCO enzyme encoded in their genome (11, 23), proteobacteria can have as many as three (59). For example, the obligate hydrogen-oxidizing chemolithotroph
*Hydrogenovibrio marinus* has three RubisCO enzymes encoded in its genome (59). Two are form IA enzymes, one of which is carboxysomal (59), while one encodes form II RubisCO. The form II gene (*cbbM*) is constitutively expressed, though *cbbM* transcription as well as enzyme concentration decrease when culture DIC concentrations are low (59). The two form IA RubisCO enzymes in *H. marinus* are also differentially expressed: carboxysomal form IA RubisCO is transcribed under low-DIC conditions, while the non-carboxysomal enzyme is preferentially transcribed under high-DIC conditions (59).

**Carboxysomes**

Carboxysomes are composed of RubisCO, carbonic anhydrase and at least seven other polypeptides which are responsible for building the shell surrounding the RubisCO and carbonic anhydrase contained by these inclusions (6). It has been proposed that the role of the carboxysomal shell is to serve as a selective barrier, allowing the influx of HCO$_3^-$ into the carboxysome, while limiting O$_2$ diffusion into the carboxysome, where it prevents O$_2$ inhibition of CO$_2$ fixation by RubisCO (13).

Though all share the same function, carboxysomes can be divided into two distinct categories: alpha carboxysomes, which contain form IA RubisCO, and have shell peptides distinct from beta carboxysomes, which contain form IB RubisCO (39, 6). Both alpha and beta carboxysomes have shell proteins with shared evolutionary history (CsoS1, CcmK, CcmO), but some proteins are unique: in alpha carboxysomes, CsoS2 and CsoSCA proteins, which are a shell
protein and a carboxysomal CA, do not appear to have orthologs in beta carboxysomes, which are distinct from alpha carboxysomes due to the presence of CcmM and CmmN proteins (6). However, the CcmM protein has a domain which is homologous to $\gamma$-CA and form Ib small subunit RubisCO (39).

**Carbonic Anhydrase**

Carbonic anhydrase, which can play a major role in DIC uptake and fixation in autotrophic organisms (2, 51, 31), has arisen independently multiple times. Four forms have been described thus far: $\alpha$, $\beta$, $\gamma$, and $\delta$. $\alpha$-CA enzymes are widespread among bacteria and are also present in animals (51). $\beta$-CA is found in plants, bacteria, and archaea (51); this group includes the CsoSCA carbonic anhydrase present in alpha carboxysomes (6, 39). The only $\gamma$-CA that has been biochemically characterized is found in a methanoarchaeon, however, putative $\gamma$-CA from sequence analysis have been found in bacteria and archaea (60). $\delta$–CA is found in marine algae; in dinoflagellate *Lingulodinium polyedrum*, it plays a role in dissolved inorganic carbon uptake (27).

**Thiomicrospira crunogen**

*T. crunogen* is an obligate chemoautotrophic gamma proteobacterium that was originally isolated from scrapings taken from *Riftia pachytila* tubeworms at the East Pacific hydrothermal vents (21). This mesophilic organism grows best at circumneutral p[H] (57) and can only use a limited subset of redox substrates: it is an obligate aerobe that does not denitrify (21), and can only use reduced sulfur
compounds (thiosulfate, sulfide, or elemental sulfur) as electron donors (21, 43, 44, 57).

As described above, growth in the hydrothermal vent habitat exposes this organism to fluctuations in temperature, pH and DIC (13, 22). The fluctuations in CO₂ availability are particularly problematic for this organism, since *T. crunogena* uses the Calvin-Benson-Bassham cycle and RubisCO for carbon fixation (appendix). It was reasonable to hypothesize that this organism has a CCM based on the observation that it is capable of rapid growth at concentrations of DIC as low as 20 μM, as well as the presence of carboxysomes (FIG. 1; 47).

Figure 1 Transmission electron micrograph of *T. crunogena* cells, with arrows indicating carboxysomes. Micrograph courtesy of M. Bright and K. Scott.

The genome of *T. crunogena* does encode three carbonic anhydrase (CA) enzymes (one α-CA, one β-CA, and one β-like carboxysomal CA) which were
hypothesized to play a role in DIC uptake and fixation (appendix). There are also three RubisCO enzymes, two form IAs and one form II, as well as a complete carboxysome operon.

The focus of this study was to examine the carbon concentrating mechanism in *T. crunogena*. The *T. crunogena* CCM was physiologically characterized (Chapter 2; 8) by determining the $K_s$ and $V_{max}$ for HCO$_3^-$ and CO$_2$ uptake by cells grown under DIC-limiting conditions and DIC-replete conditions. The ability of cells to generate elevated intracellular concentrations of DIC, as well as the necessity of the presence of an electron donor for this process, were also measured.

The sequence of its genome was then examined to find genes that might play a role in enabling *T. crunogena* to grow under low DIC conditions (appendix). Three carbonic anhydrase genes were apparent (one $\alpha$-CA, and two $\beta$-CA enzymes, one of which is carboxysomal). To examine their function (Chapter 3), these genes were heterologously expressed in *E. coli*, assayed for activity, and tested for sensitivity to classical carbonic anhydrase inhibitors. *T. crunogena* cells were grown under low-DIC and high-DIC conditions, and assayed for CA gene transcription via qRT-PCR, and CA enzyme activity and inhibition patterns. The impact of these CA enzymes on DIC uptake and fixation was inferred by measuring the impact of CA inhibitors on these parameters.

To uncover other genes that could play a role in the CCM, microarrays were used to compare genome-wide gene expression in low- and high-DIC cells (Chapter 4). It was hypothesized that genes playing a role in the CCM would be
transcribed more frequently under low-DIC conditions. Genes that appeared to be transcribed more frequently when assayed with the microarray were examined further via qRT-PCR.

*T. crunogena* and relatives are able to adapt to low concentrations of DIC during growth. Understanding this adaptation provides insights into how the ancestors of these organisms coped with the historic drop in atmospheric CO2, and also provides insights into the ecophysiology of carbon fixation in the diverse habitats where this process is catalyzed by many phylogenetically and physiologically distinct microorganisms.

References


47. Scott K., Bright M., Fisher C. (1998) The burden of independence: Inorganic carbon utilization strategies of the sulphur chemoautotrophic hydrothermal vent isolate *Thiomicrpsira crunogena* and the symbionts of


60. Zimmerman, S., Ferry, J., Supuran, C. (2007) Inhibition of the Archaeal β-class (Cab) and γ-class (Cam) carbonic anhydrases. *Current Topics in Medicinal Chemistry*, 7: 901-908.
Chapter 2

The Carbon Concentrating Mechanism

of the Hydrothermal Vent Chemolithoautotroph

*Thiomicrospira crunogena*

Kimberly P. Dobrinski, Dana L. Longo, and Kathleen M. Scott*

*Biology Department, University of South Florida, Tampa, FL USA*

*Corresponding author. Mailing address: 4202 East Fowler Avenue, SCA 110; Tampa, FL 33620; USA. Phone: (813)974-5173. Fax: (813)974-3263. E-mail: kscott@cas.usf.edu.*

Running title: Chemolithoautotrophic carbon concentrating mechanism

Abstract

Chemolithoautotrophic bacteria grow in habitats with a variety of dissolved inorganic carbon (DIC) concentrations, and are likely to have transport-related adaptations to DIC scarcity. Carbon concentrating mechanisms (CCMs) are present in many species of cyanobacteria, enabling them to grow in the presence of low concentrations of CO₂ by utilizing bicarbonate transporters and CO₂ traps to generate high intracellular concentrations of DIC. Similar CCMs may also be present in many other autotrophic bacteria. The sulfur-oxidizing γ-proteobacterial chemolithoautotroph *Thiomicrospira crunogena* experiences broad fluctuations in DIC availability at its hydrothermal vent habitat, and may use a CCM to facilitate growth during periods of CO₂ scarcity. *T. crunogena* was cultivated in chemostats under DIC limitation to determine whether it has a CCM. Its $K_{DIC}$ for growth was 0.22 mM, with a $\mu_{\text{max}}$ of 0.44 hr⁻¹. In short-term incubations with DI¹⁴C, DIC-limited cells had higher affinities for DIC (0.026 mM) than DIC-sufficient cells (0.66 mM). DIC-limited cells demonstrated an ability to use both extracellular CO₂ and HCO₃⁻, as assayed by isotopic disequilibrium incubations. These cells also accumulated intracellular DIC to concentrations 100X higher than extracellular, as determined using the silicone oil centrifugation technique. Cells that were not provided with an electron donor did not have elevated intracellular DIC concentrations. The inducible changes in whole-cell affinity for DIC, the ability to use both extracellular CO₂ and HCO₃⁻, and the energy-dependent generation of elevated intracellular concentrations of DIC are all consistent with the presence of a CCM in *T. crunogena*.
2.7 billion years ago, this planet was geochemically, ecologically, and biologically distinct from what it is today. Atmospheric carbon dioxide concentrations were 1-3 orders of magnitude higher (16, 30). Representatives from all three domains of life (Archaea, Bacteria, and Eukarya) were already present and had begun to diversify (3). Confronted with the precipitous fall of atmospheric and oceanic CO₂ concentrations in the late Proterozoic, many autotrophic lineages likely responded with adaptations to maintain an adequate supply of CO₂ for growth (27, 35).

Carbon concentrating mechanisms (CCMs) can facilitate rapid autotrophic growth in environments where the CO₂ and/or HCO₃⁻ concentrations are chronically or episodically low. CCMs are present in many species of cyanobacteria, and generate an elevated intracellular concentration of HCO₃⁻ by using active HCO₃⁻ transport (22, 32) and CO₂ traps (33). Carboxysomal carbonic anhydrase (EC 4.2.1.1) converts intracellular HCO₃⁻ to CO₂, which is fixed by ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco, EC 4.1.1.39; 2, 15, 26). The elevated intracellular concentrations of dissolved inorganic carbon resulting from active transport expedite carbon fixation by Rubisco by enhancing substrate availability and mitigating the Rubisco oxygenase reaction (15).

CCMs have not been rigorously studied for any other prokaryotic autotrophs (e.g., autotrophic Proteobacteria, Planktomycetes, Green Sulfur Bacteria, Aquificales, Archaea). This is surprising, as CCMs are likely to be quite relevant to primary productivity in the diverse habitats where autotrophic
microorganisms are found. CCMs with similarities to those present in cyanobacteria may be widespread. For example, the recently-sequenced genomes of the photosynthetic α-proteobacterium *Rhodopseudomonas palustris* and the ammonia-oxidizing β-proteobacterium *Nitrosomonas europaea* contain genes for carbonic anhydrase and potential HCO$_3^-$ transporters (4, 18). Furthermore, it has recently been demonstrated that carboxysomes from the chemolithoautotroph *Halothiobacillus neapolitanus* contain carbonic anhydrase, and are believed to function similarly to those present in cyanobacteria (34).

A CCM could facilitate the growth of chemolithoautotrophs at hydrothermal vents, where there is an enormous degree of spatial and temporal variability in the concentration of CO$_2$ (9). The hydrothermal vent γ-proteobacterium *Thiomicrospira crunogena* is an obligate sulfur-oxidizing chemoautotroph that was originally isolated from a deep-sea hydrothermal vent habitat where the CO$_2$ concentration oscillates between 20 µM and 1 mM, and HCO$_3^-$ is always the most abundant form of dissolved inorganic carbon (DIC, equal to the sum of CO$_2$, HCO$_3^-$, and CO$_3^{2-}$; (9, 11). This is one of the fastest-growing chemoautotrophs, with a doubling time as low as one hour (11). It continues to grow rapidly in batch culture even after drawing the concentration of DIC down to less than 20 µM, and it has carboxysomes (31). Both characteristics are consistent with the presence of a CCM.

Detailed physiological experiments were undertaken to determine whether this organism has a carbon concentrating mechanism. Understanding how *T. crunogena* and other autotrophic microorganisms adapt to low concentrations of
DIC during growth is relevant to understanding the physiology of these unique organisms, and will provide insights into the response of autotrophic carbon fixation to changes in global geochemistry over Earth history.

Materials and Methods

**Analytical methods and reagents.** Dissolved inorganic carbon (DIC; \( \text{CO}_2 + \text{HCO}_3^- + \text{CO}_3^{2-} \)) was quantified with an Agilent 6890N gas chromatograph equipped with an extractor to permit stripping dissolved gases from aqueous samples (5). Ammonia (\( \text{NH}_3 + \text{NH}_4^+ \)) was assayed using a commercial colorimetric kit (Sigma Inc.). After sonicating the cells for 30 sec in the presence of glass beads, total protein was measured with a Lowry-type assay (BioRad Inc).

The DI\(^{14}\)C used to measure carbon uptake and fixation rates was purchased as a sterile pH 9.5 solution (2 mCi ml\(^{-1}\), 50 mM DIC, MP Biomedicals Inc.). Upon receipt, 0.5 ml portions were sealed into glass vials with gas-tight gas chromatograph septa, and stored at 4ºC until use. These stock DI\(^{14}\)C solutions had stable counts over the course of this study (unpubl. data).

**Bacterial strains and growth conditions.** *Thi microspira crunogena* XCL-2 (1) was cultivated at 25ºC on liquid and solid TASW media modified from (11). TASW medium consists of artificial seawater supplemented with 40 mM thiosulfate, which *T. crunogena* utilizes as an electron donor, and NaHEPES to maintain the pH at 8 (100 mM in batch culture, and 10 mM for continuous culture). The strain was maintained long-term in 15% glycerol/TASW medium (v/v) at -80ºC.
**Cultivation under nutrient limitation.** *T. crunogena* was cultivated in chemostats (New Brunswick Scientific BioFlo 110) to grow the cells under DIC or ammonia limitation. dO2/pH controllers monitored the pH and O2 concentrations in the growth chambers with electrodes, maintaining optimal growth conditions by adding 10N KOH to keep the pH between 7.8 and 8, and by periodically pulsing the growth chamber with O2 gas to maintain its concentration between 3 and 25 µM. The growth chamber was supplied with TASW medium (2.5 mM DIC, 6.6 mM (NH4)2SO4) from a 10L reservoir at a range of dilution rates (0.03 – 0.44 hr⁻¹).

**Measurement of inducible changes in the half-saturation constant for DIC.** To determine whether *T. crunogena* has inducible adaptations to cope with lower concentrations of DIC during growth, DIC-limited and DIC-sufficient (but ammonia-limited) cells were cultivated in chemostats at a dilution rate of 0.1 vessel volume hr⁻¹ and their whole-cell affinities for DIC were measured. For DIC-sufficient cells, the reservoir [DIC] was raised to 10 mM, and the ammonia concentration was dropped from 13.2 mM to 0.5 mM. When the dilution rate was 0.1 vessel volume hr⁻¹, the steady-state [DIC] in the growth chamber was 0.08 mM for DIC-limited cells, and 5.5 mM for DIC-sufficient cells. The growth chamber ammonia concentration for DIC-sufficient cells was below the limit of detection for the assay used (< 10 µM). DIC or ammonia was confirmed to be limiting growth by observing higher biomass densities in the growth chamber, assayed as protein concentrations, when either DIC or ammonia concentrations (as appropriate) were raised.
To harvest the cells, 150 ml portions were removed from the growth chamber and centrifuged (5000 g, 4°C, 10 min). Pellets were resuspended in 3 ml TASW medium (for DIC-limited cells: trace DIC, 13.2 mM ammonia; for ammonia-limited cells: 5.5 mM DIC and 0.5 mM ammonia), and kept on ice until the experiment was completed (less than 30 mins). 10 µl aliquots of the suspended cells were added to 7 glass reaction vials with stir bars, filled with 1.98 ml TASW medium (pH 8, 0.02 to 10 mM DIC, supplemented with 14C-DIC to a specific activity of 2-40 Ci/mol). Once per minute, over a timecourse of 4 min, a 400 µl aliquot was removed from each reaction vial and injected into a scintillation vial containing 200 µl 65°C glacial acetic acid. These acidified samples were gently sparged with air until dry to remove the 14C-DIC. Scintillation cocktail was added to quantify the organic 14C. Initial activities were measured by injecting 10 µl portions of the incubations into scintillation vials containing 3 ml scintillation cocktail plus 50 µl β-phenethylamine.

**Bicarbonate and carbon dioxide uptake and fixation.** To determine whether DIC-limited cells can use both extracellular bicarbonate and carbon dioxide, DIC-limited cells were cultivated and harvested as described above, and resuspended in DIC-free TASW medium and bubbled with soda lime-treated (carbon dioxide-free) air until [DIC] = 0. Cell suspensions were then placed on ice and gently sparged with carbon dioxide-free air until use. Incubations with 14C were conducted as described above, with the following modifications. Instead of using DI14C equilibrated at pH 8, an isotopic disequilibrium technique was used, in which either H14CO3− or 14CO2 was added (6). The 14C stock solution was
pH 9.5 in distilled water, and therefore < 0.1% CO₂, ~30% HCO₃⁻, and 70% CO₃²⁻. Upon injection into a well-buffered pH 8 solution, the HCO₃⁻ concentration instantaneously jumps to 93% due to protonation of most of the CO₃²⁻. To prepare dissolved ¹⁴CO₂, ¹⁴C stock solution was added to DIC-free 1 mM H₃PO₄ in a sealed glass vial and allowed to equilibrate ~30 mins before use to quantitatively convert the DI¹⁴C to ¹⁴CO₂.

10 µl aliquots of suspended cells were added to reaction vials filled with 2 ml DIC-free TASW medium. H¹⁴CO₃⁻ or ¹⁴CO₂ were added to the vials to begin the reaction, and timepoints were taken at 10 sec intervals. Upon addition, ¹⁴CO₂ generates H¹⁴CO₃⁻ with a half-dehydration time of ~17 sec, while ¹⁴CO₂ forms from H¹⁴CO₃⁻ with a half-hydration time of 26 mins (37). Accordingly, incubations in which ¹⁴CO₂ or H¹⁴CO₃⁻ was initially added were limited to 20 sec or 40 sec, respectively.

**Intracellular DIC accumulation and pH.** The silicone oil centrifugation method was used to measure the size of the intracellular DIC pool and the intracellular pH (modified from 14). DIC-limited cells were grown and harvested as described above. 0.6 ml eppendorf tubes were prepared containing two immiscible layers of fluid: a dense bottom layer, consisting 20 µl of a killing solution (2:1 (v:v) 1M glycine, pH 10:triton) overlain with 65 µl of silicone oil (Dow Chemicals SF 1156). Cell suspensions (200 µl) were pipetted on top of the silicone oil layer, mixed with radiolabelled solute, and at timed intervals, were centrifuged for 20 sec. The cells passed through the silicone layer and pelleted in the killing solution, carrying intracellular radiolabelled solute. Immediately after
centrifugation, the tubes were frozen in liquid nitrogen and the bottom layer with
the cells was clipped into scintillation vials containing 200 µl glacial acetic acid
(for measuring intracellular fixed carbon) or directly into 3 ml scintillation
cocktail plus 50 µl β-phenethylamine (for quantifying intracellular fixed and
inorganic carbon, cell volumes, or pH).

For measuring intracellular DIC, TASW medium containing 3 to 240 µM
DI^{14}C (40 Ci mol⁻¹) was allowed to equilibrate at pH 8 before the experiment. 200
µl portions of this solution were pipetted on tope of the silicone layers in
eppendorf tubes, and 10 µl of suspended cells in DIC-free TASW medium were
added. After 30 sec, the tubes were spun at 14,000 g for 20 sec before freezing
and processing, as preliminary timecourse experiments indicated that the
intracellular DIC pool was constant after 20 sec. For each concentration of DIC,
samples were run in parallel to measure intracellular DIC plus fixed carbon
(clipped into scintillation cocktail alkalinized with β-phenethylamine) and fixed
carbon (clipped into glacial acetic acid). Values from the samples clipped into
glacial acetic acid were subtracted from those clipped into alkalinized scintillation
cocktail to calculate the intracellular DIC pool. For estimating intracellular
volume, which is necessary for calculating intracellular solute concentrations,
incubations were also conducted with 9 µCi ml⁻¹ D-sorbitol ([U-^{14}C], MP
Biomedicals) and 3 µCi ml⁻¹ tritiated water (Amersham Biosciences; 14, 29). The
intracellular concentrations of both of these substances reached equilibrium
before 2 min (unpubl. data). Accordingly, incubations with sorbitol or tritium
were terminated at 2 min by centrifugation.
To assess whether DIC accumulation was energy-dependent or simply driven by intracellular alkalinization, the intracellular pH of DIC-limited cells was measured using the silicone oil centrifugation technique described above, with methylamine (ME; methylamine hydrochloride, [¹⁴C]; MP Biomedicals Inc). ME has a pKa of 10.7 and accumulates inside cells as its positively charged, conjugate acid when the cytoplasmic pH is lower than the extracellular pH (29, 38). ME was added to cell suspensions (pH 8) to an activity of 0.45 μCi ml⁻¹ and allowed to equilibrate for 3 min, which was sufficient time to reach equilibrium concentrations inside the cells (K. Scott, unpubl. data). Parallel incubations with ¹⁴C-sorbitol and tritiated water were conducted with the cell suspensions to make it possible to estimate the intracellular concentrations of the ME.

Due to concern that ME might be accumulating in the cells via uptake by ammonia permeases (20), intracellular pH was also measured with DMO (dimethyloxazolidine-2,4 dione 5,5 [²⁻¹⁴C]; American Radiolabeled Chemicals; 29, 38). DMO has a pKa of 6.2 (29, 38) and accumulates inside cells as its negatively charged, conjugate base. Incubating cells with DMO under conditions where the extracellular pH is much more alkaline than the intracellular pH prevents DMO from accumulating within the cells. To raise intracellular DMO concentrations, the pH used for incubations with this compound was 7.3 instead of 8.0. A second series of incubations with ME were also conducted at pH 7.3.

**Energy-dependence of intracellular DIC accumulation.** Cells were cultivated under DIC-limiting conditions as described above. To prepare them for these experiments, it was necessary to wash them four times with thiosulfate-free
artificial seawater medium as this electron donor is present in their growth medium at a high concentration (40 mM). After the final wash, cells were resuspended in DIC-free, thiosulfate-free medium and put on ice while being gently bubbled with CO₂-free air. Silicone oil centrifugation experiments with radiolabelled DIC were conducted as described above, at a range of thiosulfate concentrations (0-1 mM).

Results

Cultivation under DIC limitation. Steady-state exponential growth, confirmed by protein and DIC assays, occurred after ~6 liters of TASW medium had passed through the 1 liter growth chamber. The steady-state DIC concentrations in the growth chamber were substantially lower than in the reservoir due to consumption by *T. crunogena* (FIG. 2). A rectangular hyperbola was fitted to the data via the direct linear plot method (8) to estimate the $K_{DIC}$ (0.22 mM) and the $\mu_{max}$ (0.44 hr⁻¹).

![Graph showing steady-state growth rate constants (μ) and dissolved inorganic carbon (DIC) concentrations for T. crunogena cells grown in chemostats under DIC limitation. Using the direct linear plot method, a rectangular hyperbola was fitted to the data to estimate the $K_{DIC}$ (0.22 mM) and the $\mu_{max}$ (0.44 hr⁻¹).]
linear plot method (8), a rectangular hyperbola was generated to model the data (shown as a curve in the figure; \( K_{DIC} = 0.22 \text{ mM}; \mu_{\text{max}} = 0.44 \text{ hr}^{-1} \)). The error bars indicate the standard error of the measurements (n=3).

**Measurement of inducible changes in the half-saturation constant for DIC.** DIC-limited cells had much higher whole-cell affinities for DIC \( (K_{DIC} = 0.026 \text{ mM}) \) than DIC-sufficient cells did \( (K_{DIC} = 0.66 \text{ mM}; \text{FIG. 3}) \), which is consistent with inducible changes in transport and/or fixation occurring when cells were growing under DIC-limiting conditions.

**FIG. 3.** Carbon fixation rates for *T. crunogena* harvested from DIC-limited (solid circles) and DIC-sufficient (ammonia-limited; open circles) chemostats with \( \mu = 0.1 \text{ hr}^{-1} \). Data are presented with rectangular hyperbolae derived as in (8). For DIC-limited cells, the \( K_{DIC} (0.026 \text{ mM}) \) was lower than for DIC-sufficient cells \( (K_{DIC} = 0.66 \text{ mM}) \). The \( V_{\text{max}} \) for both was similar \( (133 \text{ and } 120 \text{ nmol min}^{-1} \text{ mg protein}^{-1}, \text{respectively}) \). The error bars indicate the standard deviations of the slopes.

**Bicarbonate and carbon dioxide uptake and fixation.** *T. crunogena* demonstrated an ability to use both extracellular carbon dioxide and bicarbonate (FIG. 4). Some interconversion of carbon dioxide and bicarbonate did occur over
the timecourse of these experiments and needs to be considered when interpreting the results. When $^{14}$CO$_2$ was initially added, the H$^{14}$CO$_3^-$ formed was unlikely to contribute substantially to the observed carbon fixation rates, as the initial concentration of CO$_2$ was low (1 to 6 µM) relative to the K$_{HCO_3^-}$ (53.6 µM, FIG. 4). Carbon fixation rates for *T. crunogena* harvested from DIC-limited chemostats with $\mu = 0.1$ hr$^{-1}$, when supplied with CO$_2$ (A) K$_{CO_2} = 1.03$ µM; $V_{max} = 97.2$ nmol min$^{-1}$ mg protein$^{-1}$) or HCO$_3^-$ (B) K$_{HCO_3^-} = 53.6$ µM; $V_{max} = 149$ nmol min$^{-1}$ mg protein$^{-1}$). Data are presented with rectangular hyperbolae derived as in (8). In (B), error bars indicate the standard deviations of the slopes. Error bars are not presented in (A) since the carbon fixation rates were based on 2 timepoints.

FIG. 4. Carbon fixation rates for *T. crunogena* harvested from DIC-limited chemostats with $\mu = 0.1$ hr$^{-1}$, when supplied with CO$_2$ (A) K$_{CO_2} = 1.03$ µM; $V_{max} = 97.2$ nmol min$^{-1}$ mg protein$^{-1}$) or
HCO$_3^-$ (B) $K_{HCO_3^-} = 53.6$ µM; $V_{max} = 149$ nmol min$^{-1}$ mg protein$^{-1}$). Data are presented with rectangular hyperbolae derived as in (8). In (B), error bars indicate the standard deviations of the slopes. Error bars are not presented in (A) since the carbon fixation rates were based on 2 timepoints.

For the experiments in which H$^{14}$CO$_3^-$ was added, the $^{14}$CO$_2$ formed from H$^{14}$CO$_3^-$ over the 40 sec timecourse probably contributed to the observed carbon fixation rates since the cells demonstrate such a high affinity for CO$_2$. To account for this, a pseudo-first-order rate constant for CO$_2$ formation from HCO$_3^-$ in seawater was used to calculate the concentration of CO$_2$ present in the incubations at each timepoint (37). The contribution of this CO$_2$ to the measured rates of carbon fixation was estimated using the $K_{CO_2}$ (1.03 µM) and $V_{max}$ (97.2 nmol min$^{-1}$ mg protein$^{-1}$) for CO$_2$-dependent carbon fixation. When this estimate of CO$_2$-dependent carbon fixation was subtracted from the rates measured in these experiments, the $K_{HCO_3^-}$ was 11 µM and the $V_{max}$ was 52 nmol min$^{-1}$ mg protein$^{-1}$.

The data are consistent with bicarbonate use contributing substantially to growth under DIC limitation. At pH 8 and 80 µM DIC, the conditions under which these cells were grown, the bicarbonate concentration was 74 µM and the carbon dioxide concentration was ~0.7 µM (19). Using the $K_s$ and $V_{max}$ values (FIG 3; parameters for HCO$_3^-$ use corrected as described above) and the Michaelis-Menton equation, the carbon fixation rates due to bicarbonate and carbon dioxide use are 45 and 39 nmol min$^{-1}$ mg protein$^{-1}$, respectively. Whether DIC-sufficient cells also demonstrate an ability to use extracellular bicarbonate
remains to be determined; our manipulations of the DIC concentration in the cell suspensions to prepare them for these experiments would likely induce the expression of the same traits observed in DIC-limited cells.

FIG. 5. Intracellular DIC accumulation by *T. crunogen* cultivated under DIC limitation in a chemostat (DIC = 0.1 mM, \( \mu = 0.1 \text{ hr}^{-1} \)). A unity line (dashed; intracellular DIC = extracellular DIC) is presented with the data for comparison, and error bars are the standard errors of the intracellular concentrations (n=3).

**Intracellular DIC accumulation and pH.** Intracellular concentrations of DIC exceeded the extracellular concentration by 100X (FIG. 5), consistent with energy-dependent bicarbonate transport and accumulation in the cytoplasm. The intracellular pH of *T. crunogen* is \(~7\) at an extracellular pH of 8 (ME: 6.88 ± 0.10) and 7.3 (ME: 7.03 ± 0.06; DMO: 7.01 ± 0.32; pH ± std. err., n=3), precluding intracellular alkalinization as the driving force for intracellular DIC accumulation.

**Energy-dependence of intracellular DIC accumulation.** Cells accumulated elevated intracellular concentrations of DIC when thiosulfate was
present at concentrations greater than 1 µM (FIG. 6). This correlation of DIC accumulation and thiosulfate presence is consistent with intracellular DIC accumulation relying on either the membrane potential or ATP synthesis resulting from thiosulfate oxidation.

![FIG. 6. Energy-dependence of intracellular DIC accumulation by T. crunogena. Cells were cultivated under DIC limitation with thiosulfate (Na$_2$S$_2$O$_3$) as the electron donor. [DIC]$_i$ = intracellular DIC concentration, [DIC]$_e$ = extracellular DIC concentration (0.1 mM), and error bars indicate the standard error (n=3).](image)

**Discussion**

*T. crunogena* has a CCM that enables it to grow in the presence of low concentrations of CO$_2$ by generating an elevated concentration of intracellular DIC. The results presented here are consistent with active transport of bicarbonate or carbon dioxide playing a role. Active transport could create elevated intracellular DIC concentrations despite the intracellular pH being lower
than extracellular, and would require the presence of thiosulfate or other electron donor for energy.

A CCM may serve to supplement the supply of CO₂ available at *T. crunogena*’s hydrothermal vent habitat. Temperature differences between bottom water (2°C) and dilute hydrothermal fluid (up to 35°C) create turbulent eddies at the vents. In turn, these eddies are responsible for seconds-to-hours-long oscillations of environmental CO₂ from 1 mM down to ~20 µM CO₂ (9, 12). During the lower CO₂ periods, a CCM would enable *T. crunogena* to continue to grow rapidly.

The CCM present in *T. crunogena* has many parallels with cyanobacterial CCMs. In both cases, whole-cell affinities increase in response to the concentration of DIC available during growth. Energy-dependent transport generates an elevated concentration of intracellular DIC which, presumably in *T. crunogena* as in cyanobacteria, is “harvested” by carboxysomes (2, 26, 34). Use of extracellular HCO₃⁻ (FIG. 4) is consistent with the presence of HCO₃⁻ transporters, and a high cellular affinity for CO₂ (1 µM) compared to Rubisco (30-140 µM for form IA, 100-250 µM for form II; 10) may be indicative of an active CO₂ uptake system.

*T. crunogena*’s genome (Scott, unpubl. data) encodes some genes whose products may function similarly to the components of cyanobacterial CCMs. An α-type carboxysome operon, which includes an ε-class carbonic anhydrase gene (34), is present. Additionally, a gene for a SulP-type anion transporter has been
found, similar to the SulP-type transporters from marine cyanobacteria that have recently been demonstrated to have \(\text{HCO}_3^-\)-transporting activity (24).

It is also likely that the \textit{T. crunogena} CCM will have several features that distinguish it from cyanobacterial CCMs. In contrast to cyanobacteria, whose genomes have a single form IA or form IB Rubisco gene (13, 21, 23, 28), \textit{T. crunogena}’s genome carries three genes for this enzyme (two form IA Rubiscos, one form II Rubisco). The two form I Rubisco genes are expressed when cells are grown under low DIC conditions, while the form II Rubisco gene is preferentially expressed under high DIC conditions (K. Scott, unpubl. data), similar to what has been observed for \textit{Hydrogenovibrio marinus} (36). Also different from cyanobacteria, \textit{T. crunogena}’s genome lacks any apparent homologs for the cyanobacterial bicarbonate transporter genes \textit{cmpABCD} (22) and \textit{sbtA} (32), as well as \textit{chpX} and \textit{chpY}, which encode key components of the cyanobacterial CO\(_2\) uptake system (25, 33). It is possible that novel \(\text{HCO}_3^-\) and CO\(_2\) transporters will be found in this organism, and it will be interesting to determine whether transporter and Rubisco gene expression is coordinately regulated. The generation of random and directed knockout mutants is underway (Dobrinski and Scott, unpubl. data), with the objective of deciphering the mechanism for intracellular DIC accumulation in this organism.

\textit{T. crunogena} is not likely to be the only chemolithoautotroph with a CCM. CCMs may be present in other chemolithoautotrophs that have a high demand for DIC (e.g., due to rapid growth rates), utilize a sufficiently abundant or electronegative electron donor to offset the energetic burden of a CCM, and
inhabit an environment with periods of DIC or CO₂ scarcity. Mechanistic and genetic studies of CCMs in several lineages of Bacteria and Archaea has the potential to illuminate how their ancestors coped with the two-orders-of-magnitude drop in CO₂ availability occurring over Earth history (16, 30).

Autotrophs may have addressed this dilemma with a single solution that was spread by horizontal gene transfer (as Rubisco genes were; 7). Alternatively, each lineage may have come up with a unique solution. It is reasonable to anticipate lineage-specific innovation, based on cyanobacterial CCMs. In cyanobacteria, three nonhomologous bicarbonate transporting systems and two forms of carboxysomes are scattered among the different clades (2, 24). Other autotrophs inhabit microhabitats even more disparate than those where cyanobacteria flourish (17), and embrace an astounding degree of phylogenetic and physiological diversity (at least four divisions of Bacteria; Archaea). Perhaps this ecological and phylogenetic diversity is reflected in a genetic and mechanistic diversity of CCMs.

Acknowledgements
Thanks are due to Douglas Nelson for providing the strain of *T. crunogena* used for this study, to Gordon Cannon for illuminating discussions, to Stacy Guerin and Darinka Obradovich for expert assistance with the experiments, to Peter Girguis, Shana Goffredi, and Zoe McCudden for editing early versions of this manuscript, and to two anonymous reviewers for helpful suggestions. This work was supported in part by a New Researcher Grant (USF).
References


Chapter 3

Expression and Function of Three Carbonic Anhydrase

Enzymes in *Thiomicrospira crunogena*

*Introduction*

*Thiomicrospira crunogena*, a deep-sea hydrothermal vent sulfur oxidizing chemoautotroph (17), lives in a spatially and temporally heterogeneous environment where it experiences broad sweeps in pH (5 to 8) and dissolved inorganic carbon (DIC; = CO$_2$ + HCO$_3^-$ + CO$_3^{2-}$; 2 mM to 7 mM; 12) due to interactions between hydrothermal fluid emitted from the seafloor and the overlaying seawater that percolates through the fluid and trickles into the earth’s crust (38). This results in turbulent eddies caused by mixtures between bottom sea water and dilute hydrothermal fluid that results in temperature and chemical changes occurring on a temporal scale from seconds to days (18). These oscillations in pH and DIC result in fluctuating CO2 concentrations, from 20 to 2000 μM (12).

To cope with these oscillations in DIC composition and concentration, *T. crunogena* has a carbon concentrating mechanism (CCM; Chapter 2). *T. crunogena* can grow rapidly despite DIC concentrations of less than 20 μM in batch culture, and when grown at low concentrations of DIC, its whole-cell
affinity for DIC (0.026 mM) is markedly smaller than when cultivated at high DIC concentrations (0.66 mM; Chapter 2). *T. crunogena* can use both extracellular CO$_2$ and HCO$_3^-$, and is able to accumulate intracellular DIC to concentrations 100x higher than extracellular (Chapter 2). The inducible molecular mechanism responsible for generating high intracellular concentrations of DIC has yet to be characterized. The genome does not encode any apparent orthologs to any of the HCO$_3^-$ or CO$_2$ transporters that have been characterized in cyanobacteria (appendix, 19, 3).

The genome of *T. crunogena* does encode three carbonic anhydrase (CA) enzymes (one $\alpha$-CA, one $\beta$-CA, and one $\beta$-like carboxysomal CA) which may play a role in DIC uptake and fixation (appendix). $\alpha$-CA enzymes are the best biochemically characterized carbonic anhydrases, are present in animals and are widespread among bacteria (46). $\beta$-CA enzymes are phylogenetically unrelated to $\alpha$-CA, and are found in plants, bacteria, and archaea (46). The shared ancestry of $\beta$-CA and $\beta$-like carboxysomal CA enzymes (CsoSCA) is not apparent based on sequence comparisons; structural comparisons revealed a congruence of form indicative of extremely distant relatedness (42).

In sulfur-oxidizing beta proteobacterial autotroph *Halothiobacillus neapolitanus*, CsoSCA is present in small amounts in carboxysomes, which are protein-bound inclusions packed with RubisCO (ribulose 1,5-bisphosphate carboxylase/oxygenase; 15). CsoSCA converts HCO$_3^-$ to CO$_2$ within the carboxysome, where it is fixed by RubisCO (15). Given that the *csoSCA* gene in *T. crunogena* is present in a carboxysome operon (appendix), and carboxysomes
are present in this microorganism, it is likely that CsoSCA plays a similar role in *T. crunogena*.

Extracellular CA has been implicated in facilitating DIC uptake in a variety of organisms. For example, *Chlamydomonas reinhardtii* cells, when treated with the CA inhibitors acetazolamide (AZA) and dextran-bound sulfonamide, to which cells are relatively impermeable, are unable to replenish intracellular CO$_2$ levels from extracellular bicarbonate, prevalent at basic pH. Intracellular CO$_2$ pools were not sensitive to these inhibitors when *C. reinhardtii* were incubated at pH 5.1, when extracellular CO$_2$ was abundant (26). In *Rhodopseudomonas palustris*, growth is inhibited when its periplasmic $\alpha$-CA is inactivated either by AZA or mutation; in either case, growth rates were restored with elevated CO$_2$ concentrations or lower pH (35).

The objective of this study was to determine whether the multiple CA enzymes found in *T. crunogena* are involved in DIC uptake and fixation. The response of transcription of CA genes to the concentration of DIC available during growth was monitored. To verify that all three genes encoded functional CA enzymes, as well as to characterize patterns of inhibition, the $\alpha$-, $\beta$- and carboxysomal CA genes were overexpressed in *E.coli*. CA activity was also measured in *T. crunogena* cell extracts and in incubations with whole cells to infer the cellular location of the enzymes, and carbon fixation and DIC uptake rates were measured in the presence of CA inhibitors to determine whether these processes were affected by CA inactivation.
Materials and Methods

Analytical methods and reagents. DIC was quantified with a gas chromatograph (Chapter 2). Protein concentrations were determined by using the RC DC Protein Assay (Biorad, Hercules, CA).

Bacterial strains and cultivation. Thiomicrospira crunogena XCL-2 was cultivated in artificial seawater medium supplemented with 40 mM thiosulfate and 10mM Na HEPES, pH 8 (‘TASW’; 17, Chapter 2). Cells were grown in chemostats (Bioflo 110, New Brunswick Scientific) under DIC limitation (‘low-DIC cells’: 0.1 mM DIC, 13 mM (NH₄)₂SO₄) or ammonia limitation (‘high-DIC cells’: 50 mM DIC, 0.8 mM (NH₄)₂SO₄). The pH (≈8) and oxygen concentration (~20 -100 µM) were maintained in the growth chambers by using pH and O₂ electrodes which directed 10 N KOH addition and O₂-sparging (pure O₂ was used for the low-DIC cells, while 5% CO₂, balance O₂ was used for the high-DIC cells; Chapter 2).

One Shot Mach1™ – T1R and BL21 (DE3) One Shot® E. coli (Invitrogen, Carlsbad, CA) used for transformation and expression studies were cultivated in Luria broth supplemented with the appropriate antibiotic (see below; 41).

Transcription of CA genes in T. crunogena. Thiomicrospira crunogena cells were harvested by centrifugation (10,000×g, 5 min, 4°C), flash frozen in liquid nitrogen, and stored at -80°C. RNA was isolated from low- and high-DIC cells by the Ribopure system (Ambion, Austin, TX), and cDNA was reverse transcribed using the Improm II RT system (Promega, Madison, WI), with primers targeting the gene of interest (Table 1). Taqman primers and probe for
qRT PCR were designed using Primer Express software (Table 1; ABI, Carlsbad, CA). qRT PCR was carried out with Taqman and the Step One qPCR system (ABI, Carlsbad, CA), using the following parameters as recommended by the manufacturer: a two-step holding stage (50°C for 2 min, 95°C for 10 min), and a two-step cycling stage (95°C for 15 sec, 60°C for 1 min, 40 cycles).

To verify that amplification efficiencies were similar for primer/probe sets directed against the 16S (=calibrator) and target genes (α-CA, β-CA, and csoSCA), qPCR using these primer/probe sets was conducted on a serial dilution of template cDNA. In all cases, a plot of $\Delta C_T (=C_T$ target - $C_T$ calibrator, where $C_T$ is the qPCR cycle where fluorescence of the reaction has crossed the value considered to be baseline) versus log [template cDNA] had a slope of less than 0.1, indicating that primer/probe amplification efficiency was constant for all primer/probe sets (22). To determine the concentration of primer/probe and template that resulted in $C_T$ values falling within 10-20 cycles, a dilution series of 16S primer/probe and cDNA concentrations were run. Based on the results of these experiments, 50 ng of cDNA template, 900 nM primers, and 250 nM probe were used. To verify that the 16S gene was expressed at the same level for both low- and high-DIC cells and was therefore suitable for use as a calibrator, the $C_T$ value for qPCR directed against 16S in cDNA from low- and high-DIC cells, at a range of template concentrations, was captured. No difference in $C_T$ values, and therefore 16S gene expression, between low- and high-DIC cells was detected. The amount of 16S RNA in cDNAs from low and high-DIC cells was found to be ~40% of total cDNA used in each qPCR via Megascript Reaction (Promega, Madison, WI).
Fold differences in transcription between low- and high-DIC cells were calculated as $2^{-\Delta\Delta CT}$, where $\Delta\Delta CT = [(C_T \text{ target} - C_T \text{ calibrator})-(C_T \text{ targetRef} - C_T \text{ calibratorRef})]$, $C_T \text{ target}$ and $C_T \text{ calibrator}$ are the $C_T$ values for target and 16S amplification in high-DIC cells, and $C_T \text{ targetRef} - C_T \text{ calibratorRef}$ are the corresponding values from low-DIC cells (22).

**Expression of T. crunogena CA genes in E. coli.** CA genes and form II Rubisco (negative control) were PCR-amplified from *T. crunogena* genomic DNA and cloned into the pET SUMO plasmid (Table 1; Invitrogen, Carlsbad, CA). The construct was then introduced into competent One Shot Mach1™ – T1R E. coli. Transformants were selected from colonies growing on Luria plates with 50 µg/ml kanamycin, and validated via PCR. Plasmid DNA was isolated with spin columns (Qiagen, Germantown, MD) and transformed into BL21 (DE3) One Shot E. coli for expression. After verifying gene presence with PCR, these *E. coli* cells were cultivated in Luria Broth supplemented with 1% glucose and 50 µg/ml kanamycin. To induce gene expression, once cultures reached an OD$_{600}$ ~ 0.3, IPTG was added to a final concentration of 1 mM and the cultures were incubated overnight at room temperature while agitated at 100 rpm, harvested the next day (10,000×g, 5 min, 4°C), and flash-frozen with liquid nitrogen. To verify target gene expression, samples from each *E. coli* culture were analyzed via SDS-PAGE (41), followed by Western blotting using antisera directed against the polyhistidine tag added to the amino terminus of the proteins when they are expressed from the pET-SUMO vector (anti-His G-alkaline phosphatase-conjugated Antibody, Invitrogen).
CA assays and inhibition. *E. coli* cells expressing *T. crunogena* CA genes, as well as low- and high-DIC *T. crunogena* cells, were harvested by centrifugation (10,000×g, 5 min., 4°C). Cells were resuspended in assay buffer (5 mM HEPES, pH 8.0, 0.1 mM ZnSO₄) and sonicated on ice using acid-washed glass beads (≤106 µm; Sigma, St. Louis, MO) with 3×15 second blasts (Sonic Dismembrator, Fisher Scientific, Pittsburgh, PA). Portions of 2 ml crude extract were placed into a 4°C reaction vial with a stir bar, and a pH electrode. Once temperature was stable at 4°C, 1 ml ice-cold CO₂-saturated distilled water was injected, and the pH was monitored. For reactions where inhibitors were used, 2 ml crude extract were stirred for 1 min after inhibitor addition, and then CO₂-saturated distilled water was added and the pH was monitored as it fell. Bovine CA (1 µg/ml final concentration), used as a positive control, was dissolved in 5 mM HEPES, pH 8.0, 0.1 mM ZnSO₄. For a negative control, CA activity was measured in samples that had been autoclaved for 1 hr. Units of activity (U = (tₜₐₖ-t)/t, where t and tₜₐₖ are the time (in seconds) required for the pH to decrease from 8.0 to 7.0 in the sample and autoclaved control, respectively; 4) were calculated and normalized for the protein concentration in the assay (U mg⁻¹). For CA assays conducted on whole cells, *T. crunogena* were harvested via centrifugation (5,000 × g, 15 min., 4°C) and washed three times in assay buffer that was rendered isosmotic with growth medium by adding NaCl (5 mM HEPES, pH 8.0, 0.1 mM ZnSO₄, 65 mM NaCl).

Carbon fixation assays and inhibition. To measure the effects of CA inhibitors on DIC uptake and fixation, low- and high-DIC *T. crunogena* were
cultivated in chemostats as described above, and a 350 ml portion of culture 
\( \text{OD}_{600} \sim 0.1 \) was harvested by centrifugation (5,000 \( \times \) g, 15 min., 4\(^\circ\)C, SLC-1500 
rotor, Sorvall RC5C, Waltham, MA). Cells were washed 3× with ice-cold wash 
buffer (artificial seawater medium, buffered with 10 mM NaHepes pH 8) 
unsupplemented with thiosulfate or dissolved inorganic carbon, and resuspended 
to a final volume of 3 ml in this wash buffer. This cell suspension was sparged 
with soda lime-scrubbed (CO\(_2\)-free) air for 30 mins to minimize the DIC 
concentration. A 20 \( \mu l \) portion of cell suspension was added to 1.98 ml TASW 
supplemented with radiolabelled DIC at a range of concentrations (0.07 – 11 mM 
for low-DIC cells; 0.47 – 43 mM for high-DIC cells; specific activity = 2 – 30 
Ci/mol). For incubations with inhibitors, 20 \( \mu l \) of cell suspension were either 
added to TASW supplemented with acetazolamide or ethoxyzolamide (250 \( \mu M \)) 
to measure any immediate effects from these compounds, or incubated on ice for 
one hour in the presence of 250 \( \mu M \) inhibitor, and added to TASW brought to this 
concentration of inhibitor as well. A 4 min timecourse was taken by injecting 0.4 
ml portions of the incubation into 0.5 ml glacial acetic acid in scintillation vials at 
1 min intervals. After allowing the \(^{14}\text{C-DIC} \) to dissipate overnight, scintillation 
cocktail was added for quantification of the acid-stable \(^{14}\text{C} \) via scintillation 
counting.

To measure the effects of acetazolamide and ethoxyzolamide on DIC 
uptake, cell suspensions were prepared and sparged with CO\(_2\)-free air as above. 
Portions of 10 \( \mu l \) cell suspension were added to 200 \( \mu l \) TASW, 0.3 mM 
radiolabelled DIC (10 Ci/mol), which were layered on top of 65 \( \mu l \) silicone oil
(Dow Chemicals SF 1156) overlaying 20 µl killing solution (2:1 (v:v) 1M glycine, pH 10:triton) in a 0.6 ml eppendorf tube (Chapter 2). After a 30 sec incubation, these tubes were centrifuged for 40 sec and flash-frozen with liquid nitrogen. Cell pellets were clipped into scintillation vials primed with 50 µl β-phenethylamine and 3 ml scintillation cocktail (to measure fixed + intracellular inorganic carbon) or 0.5 ml glacial acetic acid (to measure fixed, acid-stable carbon). After clipping the pellets into glacial acetic acid, they were stirred and $^{14}$CO$_2$ was allowed to dissipate overnight before adding scintillation cocktail. To estimate the intracellular volume, which is necessary for calculating the intracellular concentration of DIC, cells were also incubated with $^{14}$C-sorbitol (to measure periplasmic volume) and $^3$H$_2$O (to measure periplasmic + cytoplasmic volume), and centrifuged through silicone oil as described above. Cytoplasmic volume was calculated by subtracting the sorbitol-permeable space from the $^3$H$_2$O-permeable space (Chapter 2).
<table>
<thead>
<tr>
<th>Target (locus tag)</th>
<th>Purpose</th>
<th>Function</th>
<th>Sequence (5'-3')</th>
<th>Location on gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>α CA (Tcr_1545)</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>AATTCGTAAACCGTGTTGGCTCGG</td>
<td>852</td>
</tr>
<tr>
<td></td>
<td>qRT PCR</td>
<td>F primer</td>
<td>GGTGATCGCTATGAATTGTTGCAAT</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>ACCAAGTGCAATTCCATCGGATAAT</td>
<td>467</td>
</tr>
<tr>
<td></td>
<td></td>
<td>probe</td>
<td>FAM ACGCCTCAGAACCACC MGBNFQ</td>
<td>436</td>
</tr>
<tr>
<td></td>
<td>Heterologous expression</td>
<td>F primer</td>
<td>ATGAAGAAACCGTGTAGCTTT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>TTAATAAATTCGTAACCGTGTTT</td>
<td>858</td>
</tr>
<tr>
<td>β CA (Tcr_0421)</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>AAGCCGTGAATGGCTAGATG</td>
<td>538</td>
</tr>
<tr>
<td></td>
<td>qRT PCR</td>
<td>F primer</td>
<td>GCCGTGGAAATTCTAAGGTCAAA</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>CTGGGATTGTGTCTCTCCATTGATG</td>
<td>338</td>
</tr>
<tr>
<td></td>
<td></td>
<td>probe</td>
<td>FAM CACAGGCCGTAATGTC MGBNFQ</td>
<td>311</td>
</tr>
<tr>
<td></td>
<td>Heterologous expression</td>
<td>F primer</td>
<td>ATGTGCGCATCAATGTGAC</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>TTAGCTATCGTGTTCCTTAAG</td>
<td>636</td>
</tr>
<tr>
<td>csoSCA (Tcr_0841)</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>GCTTGCTTGGCCTCAGATC</td>
<td>1027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F primer</td>
<td>TCTAAGGCAGACCTACACATCA</td>
<td>760</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>CGCCGCTTTATGGTCACTACT</td>
<td>801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe</td>
<td>FAM CATGAGCTGCACACCC MGBNFQ</td>
<td>784</td>
</tr>
<tr>
<td></td>
<td>Heterologous expression</td>
<td>F primer</td>
<td>ATGAATCGTGGAAAAAAGGTCATC</td>
<td>1</td>
</tr>
<tr>
<td>Gene</td>
<td>Stage</td>
<td>Primer Type/Probe Type</td>
<td>Primer Sequence</td>
<td>Length</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>16S (Tcr_R0053)</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>CTATGC&lt;sup&gt;GGTTCTTTGCTT&lt;/sup&gt;</td>
<td>1570</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>TTTATGAGATTTCGCGCAGT</td>
<td>1206</td>
</tr>
<tr>
<td></td>
<td>qRT PCR</td>
<td>F primer</td>
<td>CGAATATGCTCTACGGAGTAAGG</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>CGCGGGCTCATCCTTTAG</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td></td>
<td>probe</td>
<td>VIC CCCTCTCCTTGGGAAGGT MGBNFQ</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Megascript F primer</td>
<td>TAATACGACTCTATAGGGCGAGT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Megascript R primer</td>
<td>AAGGGCCATGATGACTTGAC</td>
<td>1129</td>
</tr>
<tr>
<td>cbbM (Tcr_0424)</td>
<td>Heterologous expression</td>
<td>F primer</td>
<td>ATGGATCAGTCGAATCGTTATG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>TTATTTGTGCAACGCAATTTTC</td>
<td>1357</td>
</tr>
</tbody>
</table>

*Numbers listed reflect the number of nucleotides 3’ of the start codon of the gene. *FAM and VIC refer to the fluorescent tags while MGB NFQ refer to the major groove binder and non-fluorescent quencher (ABI, Carlsbad, CA). *cbbm = form II RubisCO.*
Results

Transcription of CA genes in *T. crunogena*. The transcription of *csoSCA* was particularly sensitive to the DIC concentration during growth; low-DIC cells had *csoSCA* RNA levels substantially higher than high-DIC cells (Table 2). Transcription levels for *α-CA* were similar for low- and high-DIC cells (Table 2). Low-DIC cells may transcribe *β-CA* somewhat less than high-DIC cells do, but the differences were not statistically significant (Table 2).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold Increase a</th>
<th>ΔΔCT ± s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-CA</td>
<td>1.2</td>
<td>-0.3 ± 1.8</td>
</tr>
<tr>
<td>β-CA</td>
<td>0.39</td>
<td>1.4 ± 1.2</td>
</tr>
<tr>
<td>csoSCA</td>
<td>173</td>
<td>-7.4 ± 1.7</td>
</tr>
</tbody>
</table>

a qRT PCR was carried out on RNA extracted from low- and high-DIC *T. crunogena* cells using the 16S gene as the calibrator. Fold increase is the frequency of transcription in low-DIC cells divided by the frequency in high-DIC cells.

Expression of *T. crunogena* CA genes in *E. coli*. When *α-CA*, *β-CA*, and *csoSCA* genes were expressed in *E. coli*, the proteins they encode were apparent via SDS-PAGE and Western blot analysis (Fig. 7), and CA activity was measurable in *E. coli* cell extracts (Table 3). *α-CA* activity was found to be sensitive to both ethoxyzolamide and acetazolamide (Fig. 8), with activity completely inhibited at ethoxyzolamide concentrations as low as 2.5 µM (data not shown).
TABLE 3. Activity of heterologously expressed carbonic anhydrase genes from *T. crunogenae*

<table>
<thead>
<tr>
<th></th>
<th>Units mg protein$^{-1}$ ± SD$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-CA</td>
<td>79 ± 3</td>
</tr>
<tr>
<td>$\beta$-CA</td>
<td>7.1 ± 0.3</td>
</tr>
<tr>
<td>csoSCA</td>
<td>1.24 ± 0.02</td>
</tr>
<tr>
<td>cbbM</td>
<td>1.34 ± 0.05</td>
</tr>
</tbody>
</table>

$^a$ *T. crunogenae* CA genes were individually expressed in *E. coli* and activity was measured in crude extract. cbbM, which encodes a form II ribulose-1,5-bisphosphate carboxylase/oxygenase, served as a negative control. Results are from three independent trials.

$\beta$-CA activity, however, was insensitive to ethoxyzolamide and acetazolamide (FIG. 8), and neither enzyme was inhibited by dithiothreitol (data not shown).

CsoSCA activity was low, and inhibited by both ethoxyzolamide and dithiothreitol (FIG. 8), similar to carboxysomal carbonic anhydrase activity from *Halothiobacillus neapolitanus* (15).

FIG. 7. (A.) SDS PAGE and (B.) Western blot depicting heterologous expression of *T. crunogenae* carbonic anhydrase in *E. coli*. All were expressed with an amino-terminal polyhistidine tag, which added 13 kDa to the molecular weight of each protein. Lane 1—CbbM, 63 kDa; lane 2—marker;
lane 3—α−CA, 46 kDa; lane 4—β−CA, 38 kDa; lane 5—CsoSCA, 72 kDa. Cbbm was used as a negative control for all potentiometric assays in *E. coli* cells.

![Graph](image)

FIG. 8. Effect of acetazolamide (AZA) and ethoxyzolamide (EZA) on carbonic anhydrase activity of crude extracts prepared from *E. coli* in which (A.) α−CA and β−CA and (B.) csoSCA from *T. crunogena* were expressed. [inhibitors] = 250 µM for acetazolamide and ethoxyzolamide. Error bars indicate standard error.

**CA activity and inhibitors in low and high-DIC T. crunogena.** CA activity was measurable in crude extract from *T. crunogena* cells (FIG. 9). CA activity in crude extracts from low-DIC *T. crunogena* was completely inhibited by 250 µM ethoxyzolamide, but not by dithiothreitol, consistent with this activity being due primarily to α−CA. α−CA was also very active in cell extracts from high-DIC cells, based on substantial inhibition by ethoxyzolamide. For crude extract from high-DIC cells, some CA activity was resistant to both ethoxyzolamide and dithiothreitol, consistent with the presence of β−CA activity (FIG. 9). CA activity in intact cells from low-DIC and high-DIC *T. crunogena* was completely inhibited by 250 µM acetazolamide (FIG. 10).
**Effect of CA inhibitors on DIC uptake and fixation.** For intact *T. crunogena* cells, ethoxyzolamide had a pronounced and immediate inhibitory effect on carbon fixation rates in both low- and high-DIC cells, while acetazolamide did not (FIG. 11). These lower carbon fixation rates do not appear to result from inhibition of DIC uptake, as the concentration of intracellular DIC was not measurably affected by either acetazolamide or ethoxyzolamide (FIG. 12).
FIG. 11. Effect of acetazolamide (AZA) and ethoxyzolamide (EZA) on DIC uptake and fixation by *T. crunogena* cells. Inhibitors (250 µM) were added to cell suspensions and carbon fixation was measured immediately or after incubating on ice for 1 hr. (A) Cells grown under DIC-limitation (DIC during growth ~ 0.1 mM). (B) Cells grown under NH₄⁺ limitation (DIC during growth ~ 60mM). The standard deviation of the carbon fixation is so small relative to the value of the rate that it is too small to see on these figures.
FIG. 12. Effect of 250 µM acetazolamide (AZA) and ethoxyzolamide (EZA) on the size of the intracellular DIC pool and carbon fixation in intact (A.) low- and (B.) high-DIC *T. crunogena* cells. *T. crunogena* were grown under DIC-limitation (DIC ~ 0.1 mM; low-DIC) or NH₄⁺ limitation (DIC ~ 60 mM; high-DIC). Extracellular [DIC] = 0.3 mM. Error bars indicate standard error.

**Discussion**

*T. crunogena* has multiple CA enzymes (FIG. 13), and it was not clear from the genome sequence whether all of them played a role in DIC uptake and fixation in this organism. Based on the results presented here, only the CsoSCA enzyme is implicated in this process. Transcription levels for *csoSCA* are greatly enhanced under low-DIC conditions (Table 2), consistent with its role in a carbon
concentrating mechanism. It appears that CsoSCA, which is sensitive to EZA (FIG. 8), plays a role in carbon fixation, but not DIC uptake, as EZA did not inhibit DIC uptake in intact cells, though it did affect the rate of carbon fixation (FIG. 11). The greater degree of inhibition of carbon fixation by EZA in low-DIC <i>T. crunogena</i> than in high-DIC cells (40% inhibition and 23% inhibition, respectively) may be due to CsoSCA playing more of a role in low-DIC cells, which is consistent with enhanced transcription of <i>csoSCA</i> and other carboxysomal genes under low-DIC conditions (data not shown).

In contrast to <i>csoSCA</i>, <i>α−CA</i> is equally transcribed in low- and high-DIC cells, and <i>α−CA</i> enzyme is the dominant carbonic anhydrase activity measureable in crude extracts from these cells, as CA activity in crude extracts was strongly inhibited by EZA, which targets the <i>α−CA</i> and CsoSCA, but not by DTT, which inhibits CsoSCA. An analysis of the amino acid sequence of <i>α−CA</i> with TMHMM (Center for Biological Sequence Analysis, University of Denmark) predicts its location to be periplasmic, due to an amino-terminal hydrophobic sequence that is likely to be a signal peptide. Whole-cell potentiometric assays confirm the location of <i>α−CA</i> to be either extracellular or periplasmic, due to CA inhibition by AZA (FIG. 10), which is relatively membrane-impermeable (5,7). Similarly, a periplasmic <i>α−CA</i> is present in the alphaproteobacterial purple photosynthetic bacterium <i>Rhodopseudomonas palustris</i>. In <i>R. palustris</i>, <i>α−CA</i> functions to convert bicarbonate to CO₂, which facilitates CO₂ diffusion into the cytoplasm (5). However, in <i>T. crunogena</i>, <i>α−CA</i> does not appear to play a role in
DIC uptake, as incubating cells in the presence of AZA, which inhibits α−CA, does not have a measurable effect on carbon uptake or fixation (FIGS. 11 and 12).

The possibility that α−CA does not play a role in DIC uptake and fixation is also supported by the genomic context of the α−CA gene; in T. crunogena the α−CA is not located in the genome near any other genes whose products are involved in carbon fixation (appendix). A possible role in pH homeostasis, as seen for the α−CA in Helicobacter pylori (6) is unlikely in T. crunogena, as α−CA is transcribed at similar levels in cells grown at pH 6.5 and 8 (data not shown).

Another possible role might be to ‘trap’ CO₂ that is diffusing out of T. crunogena cells. Perhaps the α−CA may be converting periplasmic CO₂ into bicarbonate, which is then transported into the cell. However, such a role is not supported by the high rates of DIC uptake and fixation measured for T. crunogena incubated in the presence of AZA. Currently the role α−CA plays in T. crunogena is elusive and merits further investigation.

Similar to α−CA, it does not appear that β-CA plays a role in DIC uptake in T. crunogena cells. β-CA is transcribed similarly in low and high-DIC cells; if it were a necessary part of the T. crunogena CCM, it would be expected to be transcribed more frequently in low-DIC cells. In fact, β-CA activity may be higher in high-DIC cells, as 15% of CA activity is resistant to EZA in crude extract from high-DIC cells, while no CA activity is apparent in crude extract from low-DIC cells treated with this inhibitor (FIG. 9). A role for β-CA in high-DIC cells is suggested by its genome context, as the β-CA gene is directly downstream from the gene encoding a form II RubisCO (appendix). The form II

55
Rubisco gene is transcribed more frequently under high-DIC conditions (data not shown), consistent with what has been noted for Hydrogenovibrio marinus (49), though it does not appear this is the case for the β-CA gene.

*T. crunogena* has three distinct CAs and it is clear that CsoSCA is associated with carboxysomes and plays a major role in carbon fixation, particularly in low-DIC cells. However, carbon fixation is affected in both low- and high-DIC cells by EZA, which is consistent with CsoSCA being associated with carbon fixation in both (FIG. 11). CsoSCA has been shown to be transcribed in both low- and high-DIC cells by Northern blot (data not shown) independent of other genes contained in the carboxysome operon. Indeed a core promoter is present immediately upstream from this gene, indicating that the csoSCA may be independently regulated from the other carboxysome genes. Both α−CA and β-CA appear to be constitutively expressed; pH does not affect the level of transcription of these genes, and DIC concentration has, at most, a minor (β) or no (α) apparent effect. None of the CA enzymes appear to play a role in DIC uptake, as the intracellular DIC concentration was not affected by the presence of AZA or EZA (which inhibit α−CA and CsoSCA; FIGS.8 and 12), and neither α−CA nor β-CA were preferentially expressed in low-DIC cells (Table 2). Knock-out mutations of all three CAs are currently underway for a better understanding of the role α−CA and β-CA are fulfilling in *T. crunogena*.

*T. crunogena* and close relatives are ubiquitous isolates from hydrothermal vents worldwide (48) and experience temporal and spatial changes in DIC availability in situ (12). CsoSCA and carboxysomes facilitate growth during low-
CO₂ periods. The advantages that the α- and β-CA confer on this organism warrant further study to uncover their roles in facilitating the survival of this deep-sea microorganism.

FIG. 13. Model of *T. crunogena* cell with carbonic anhydrase locations and inhibitor sensitivities
References


Chapter 4

Transcriptome response in *Thiomicrospira crunogena*

to the dissolved inorganic carbon concentration

*Introduction*

Hydrothermal fluid emitted from cracks in the earth’s tectonic plate system provides reduced chemicals (e.g., H$_2$S, H$_2$, CH$_4$, and Fe$^{+2}$) for use by vent microbes for carbon fixation (14, 30, 8). This deep-sea hydrothermal vent environment, while being one of the most productive ecosystems on the planet (23, 16) also presents challenges to which vent organisms must adapt. In this habitat, turbulent eddies of dilute hydrothermal fluid (30°C), which has a low pH and carries H$_2$S, mix with bottom seawater (2°C), which is alkaline in pH and carries O$_2$, causing wide fluctuations in habitat chemistry (Johnson et al., 1988). The concentration of dissolved inorganic carbon (DIC = CO$_2$ + HCO$_3^-$ + CO$_3^{2-}$; 2 mM to 7 mM) and pH values (~5-8) vary considerably, presenting very divergent concentrations of CO$_2$ (20 - 2000 µM) and HCO$_3^-$ to the autotrophs growing there (16), which may necessitate adaptations to maintain a steady supply of DIC, despite environmental fluctuations.

One such adaptation is a carbon concentrating mechanism (CCM; Price et al., 2002). Ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), the
carboxylase of the Calvin-Benson-Bassham cycle, has a low affinity for CO$_2$, and can use both CO$_2$ and O$_2$ as substrates (25). CCMs compensate for the low efficiency of RubisCO and help minimize the wasteful oxygenase reaction by active HCO$_3^-$ transport, which generates a high intracellular concentration of this compound (24), and carboxysomes, protein-bound inclusions in which a trace of carbonic anhydrase catalyzes the conversion of HCO$_3^-$ to CO$_2$, which is then fixed by the massive amounts of RubisCO packed within these inclusions (3, 35, 19). CCMs have been well studied in cyanobacteria and facilitate the growth of these cells under low CO$_2$ conditions. Cyanobacterial CCMs vary among species (4), but typically include high-affinity HCO$_3^-$ transporters. Thus far, three evolutionarily distinct HCO$_3^-$ transporters have been uncovered: BCT1, an ABC-type transporter, which is induced under carbon limitation (Omata et al., 1999), a Na$^+$-dependent transporter (42), and SulP, which is evolutionarily related to sulfate transporters (38).

Hydrothermal vent $\gamma$-proteobacterial chemolithoautotroph *Thiomicrospira crunogena* has a CCM, which may enable it to grow steadily despite environmental CO$_2$ fluctuations (Chapter 2). *T. crunogena* utilizes the Calvin-Benson-Bassham cycle for CO$_2$ fixation, and grows rapidly even when the culture DIC concentration is below 20 µM (Chapter 2). *T. crunogena* cells cultivated under low DIC conditions are capable of generating intracellular DIC concentrations 100-fold higher than extracellular (Chapter 2), and can utilize CO$_2$ and HCO$_3^-$ for carbon fixation. Whole-cell affinities for DIC respond to the DIC concentration present during growth; when DIC concentrations are low, cell
affinities for this substrate are substantially higher (\(K_{\text{DIC}} = 0.026\) mM) than when cells are cultivated under elevated DIC concentrations (\(K_{\text{DIC}} = 0.66\) mM; Chapter 2).

Some components of a typical CCM are apparent in the genome of \(T.\ crunogenae\), but key components are not (appendix). A carboxysome operon is present, which encodes the shell proteins, carbonic anhydrase (\(csoSCA\)), and carboxysomal Form I RubisCO. Elsewhere in the genome, two other RubisCO enzymes are encoded (one form I and one form II), as well as an \(\alpha\)- and two \(\beta\)-CA genes (including \(csoSCA\)). However, no orthologs to the bicarbonate transporters found in cyanobacteria are apparent in the \(T.\ crunogenae\) genome (appendix).

The objective of this study was to compare the transcriptomes of \(T.\ crunogenae\) cultivated under low- and high-DIC conditions to identify the genes whose expression is stimulated by growth under low-DIC conditions, as a step in resolving all of the components of this proteobacterial CCM.

**Materials and Methods**

**Analytical methods and reagents.** Quantification of DIC was carried out via gas chromatograph (Chapter 2). Protein concentrations were determined by the RC DC Protein Assay (Biorad, Hercules, CA).

**Bacterial strains and cultivation.** *Thiomicrospira crunogenae* XCL-2 was cultivated in artificial seawater medium supplemented with 40 mM thiosulfate and 10 mM Na HEPES (‘TASW’; Chapter 2, 22). Cells were grown in chemostats (Bioflo 110, New Brunswick Scientific) under DIC limitation (‘low-DIC cells’):
0.1 mM DIC, 13 mM (NH₄)₂SO₄, 3.1 mM PO₄), ammonia limitation (‘high-DIC cells’: 50 mM DIC, 0.8 mM (NH₄)₂SO₄, 3.1 mM PO₄) or phosphate limitation (‘low-PO₄’, 50mM DIC, 13 mM (NH₄)₂SO₄, 0.025 mM PO₄). In order to maintain the pH (=8) and oxygen concentration (~20-100 µΜ), pH and O₂ electrodes directed 10 N KOH addition and O₂-sparging (pure O₂ was used for the low-DIC cells, while 5% CO₂, balance O₂ was used for the high-DIC cells; Chapter 2). After harvesting by centrifugation (10,000×g, 5 min, 4ºC), T. crunogena cells were flash frozen in liquid nitrogen, and stored at -80ºC.

**RNA isolation and transcriptional profiling.** Oligonucleotide arrays were fabricated with probes designed to represent all genes within the *T. crunogena* genome, with two or three probes per gene, and a probe length of 35 nucleotides (Combimatrix, Mukilteo, WA). RNA was isolated from cells grown in six chemostats (three low-DIC and three high-DIC) using the Ribopure system (Ambion, Austin, TX). RNA was purified further with the RNeasy Minelute cleanup kit (Qiagen, Germantown, MD), which also served to remove EDTA remaining from the Ribopure system, and was eluted in RNAse-free water. One µg total RNA was directly labeled for one hour with a *LabelIT* cy5 labeling reaction (Mirus Bio, Madison, WI). After labeling, the RNA underwent EtOH precipitation (5 min at -80ºC and 30 min centrifugation). The RNA pellet was resuspended in 16 µl of water and 4 µl fragmentation buffer (final concentration: 40 mM Tris Acetate pH 8.1, 100 mM KOAc, 30 mM MgOAc).

Prior to hybridization, Combimatrix arrays were rehydrated at 65ºC with water for 10 min and then incubated at 45ºC for two hours in prehybridization
solution (final concentrations: 6X SSPE (Lonza Accugene), 0.05% Tween-20, 20mM EDTA, 5X Denhardt’s solution, 100 ng/µl herring sperm, 0.05% SDS. Labeled RNA was then added to hybridization solution (final concentrations: 6X SSPE, 0.05% Tween-20, 20mM EDTA, 25% distilled formamide, 100 ng/µl herring sperm DNA, 0.04% SDS) and the array was incubated overnight at 45ºC. Arrays were washed and imaged per Combimatrix CustomArray™ 12K Microarray: Hybridization and Imaging Protocol (Combimatrix, Mukilteo, WA). All arrays used were stripped and rehybridized for a maximum of three times each, as subsequent stripping and hybridization resulted in substantial loss of signal (data not shown). Global normalizing was used for all six arrays (three low-DIC, and three high-DIC) to compensate for between-array variation in overall signal strength; to normalize, the average signal strength for each array was calculated using fluorescence intensity for all spots on each array. Spot intensities on all arrays were normalized so that each array would have the same average signal (Combimatrix, Mukilteo, WA). Fold changes were calculated by comparing the average values from 3 low-DIC microarrays to 3 high-DIC microarrays.

Quantification of microarray results via qRT-PCR. RNA was isolated from low- and high-DIC cells (as well as low-PO₄ cells) as above, and cDNA was synthesized via reverse transcription using the Improm II RT system (Promega, Madison, WI), with primers targeting the gene of interest (Table 4). Taqman primers and probe for qRT PCR were designed using Primer Express software (Table 4; ABI, Carlsbad, CA). qRT-PCR reactions were assembled by adding
primers, probe, and template to Taqman Universal PCR Master Mix and a Step
One qPCR system was used to amplify target genes (ABI, Carlsbad, CA), with
parameters recommended by the manufacturer (a two-step holding stage (50°C for
2 min, 95°C for 10 min), and a two-step cycling stage (95°C for 15 sec, 60°C for
1 min, 40 cycles).

Verification of amplification efficiencies was carried out with
primer/probe sets directed against the 16S (=calibrator) and target genes
(Tcr_1019, Tcr_1315, Tcr_466, and Tcr_2018), and qPCR using these
primer/probe sets was conducted on a serial dilution of template cDNA. The line
in which \( \Delta C_T = C_T \text{ target} - C_T \text{ calibrator}, \) where \( C_T \) is the qPCR cycle where
fluorescence of the reaction has crossed the value considered to be baseline was
regressed against log [template cDNA] had a slope of less than 0.1, indicating that
primer/probe amplification efficiencies were constant for all primer/probe sets
(27). Based on previously described conditions for optimization of reaction
conditions (Dobrinski et al., in prep), 50 ng of cDNA template, 900 nM primers,
and 250 nM probe were used. Transcription levels of the 16S gene were invariant
for all growth conditions explored here, verifying its use as a calibrator
(Dobrinski et al., in prep).

Fold differences in transcription between low- and high-DIC cells were
calculated as \( 2^{\Delta \Delta C_T} \), where \( \Delta \Delta C_T = [(C_T \text{ target} - C_T \text{ calibrator})-(C_T \text{ targetRef} - C_T \text{ calibratorRef})] \), \( C_T \) target and \( C_T \) calibrator are the \( C_T \) values for target and 16S
amplification in high-DIC cells, and \( C_T \) targetRef - \( C_T \) calibratorRef are the
(corresponding values from low-DIC cells (Livak and Schmittgen, 2001).
<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Purpose</th>
<th>Function</th>
<th>Sequence (5’-3’)</th>
<th>Location on gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tcr_1019</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>ACGCGTTAGATCCCATTG</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>qRT PCR</td>
<td>F primer</td>
<td>AGAAAGCCGGCCGCTAAA</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>CCGGTCTCTTTTTTCAGTTGT</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td></td>
<td>probe</td>
<td>FAM CCGGTCCAAACAG NFQ</td>
<td>150</td>
</tr>
<tr>
<td>Tcr_1315</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>GGTATACGAGCAGTCATTG</td>
<td>653</td>
</tr>
<tr>
<td></td>
<td>qRT PCR</td>
<td>F primer</td>
<td>CCGTCGGATTTTGGAATGAAACC</td>
<td>427</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>GGGTTACGCTAAACGCCATAA</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td></td>
<td>probe</td>
<td>FAM ACGAACCACCACTTT NFQ</td>
<td>452</td>
</tr>
<tr>
<td>Tcr_0466</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>ACGGCACATCTTTTGTTC</td>
<td>541</td>
</tr>
<tr>
<td></td>
<td>qRT PCR</td>
<td>F primer</td>
<td>GGGTTGACCGCATGTAAACGA</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R primer</td>
<td>GATCGTAAAGCCACCAGAACC</td>
<td>382</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe</td>
<td>FAM CTGCCGACAGATTTA NFQ</td>
<td>368</td>
</tr>
<tr>
<td>Tcr_2018</td>
<td>cDNA synthesis</td>
<td>R primer</td>
<td>ATATCGGTCTTTTGACAGCA</td>
<td>1423</td>
</tr>
</tbody>
</table>
### qRT PCR

| F primer  | GCCGATATCGCTGGTGTAGAATT | 1075 |
| R Primer  | GTTTTGCCAGGTTGAAAGTCTATT | 1117 |
| probe     | **FAM CCAAGGCCATATT** **NFQ** | 1099 |

### cDNA synthesis

| R primer  | TTTATGAGATTCCGCACCTG | 1206 |

### qRT PCR

| F primer  | CGAATATGCTCTACGGAGTAAAGGT | 110 |
| R primer  | CGCGGGCTCATCTTTTAG | 157 |
| probe     | **VIC CCCTCTCCTTGGGAAGGT NFQ** | 136 |

---

Numbers listed reflect the number of nucleotides 3' of the start codon of the gene. FAM and VIC refer to the fluorescent tags while MGB NFQ refer to the major groove binder and non-fluorescent quencher (ABI, Carlsbad, CA). Tcr_1019, Tcr_1315, Tcr_0466, Tcr_2018 are conserved hypothetical proteins. Locus tags represent gene names.
Results and Discussion

There are many genes present in the *T. crunogena* genome that appeared likely to play a role in DIC uptake and fixation, including those encoding the carboxysomal components, three RubisCO enzymes (two form I enzymes, one of which is carboxysomal, and one form II), and three carbonic anhydrase enzymes. What was unclear upon annotation was whether these genes were differentially transcribed, and the molecular mechanism for generating high intracellular concentrations of DIC (e.g., bicarbonate transporters; appendix, Chapter 2).
FIG. 14. A sampling of housekeeping genes that do not have a measureable change in transcription under low- versus high-DIC conditions, as determined with microarrays. 'High/low' is the fold change in transcription, calculated by dividing the average spot intensity for microarrays hybridized with labeled mRNA purified from high-DIC cells by the spot intensity for those hybridized with labeled mRNA from low-DIC cells. Genes whose results are shown here include those encoding: (A.) subunits of ATP synthase, (B.) subunits of the bc$_1$ complex, (C.) subunits of the cbb$_3$ complex, (D.) subunits of DNA polymerase, and (E.) subunits of RNA polymerase. Two or three probes were designed to target each gene and are illustrated here with individual bars on the graph. Multiple bars per gene represent individual probes for that gene. Error bars represent standard error.

As expected, transcription levels for housekeeping genes, including ATP synthase, components of the electron transport chain (the bc$_1$ complex and the cbb$_3$ complex), RNA and DNA polymerase were similar for low- and high-DIC cells (FIG. 14). However, many other genes were transcribed at different levels under the different growth conditions tested (Tables 5 and 6).
### TABLE 5. Fold changes (high/low-DIC) by microarray

<table>
<thead>
<tr>
<th>Gene product</th>
<th>Locus tag</th>
<th>Probe #</th>
<th>Fold Difference $^a \pm$ SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CbbM</td>
<td>Tcr_0424</td>
<td>733_767</td>
<td>4.84 $\pm$ 0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>925_959</td>
<td>3.42 $\pm$ 0.39</td>
</tr>
<tr>
<td>CbbL</td>
<td>Tcr_0427</td>
<td>760_794</td>
<td>5.57 $\pm$ 1.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>940_974</td>
<td>4.01 $\pm$ 0.68</td>
</tr>
<tr>
<td>CbbS</td>
<td>Tcr_0428</td>
<td>23_58</td>
<td>4.80 $\pm$ 0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99_136</td>
<td>4.54 $\pm$ 0.52</td>
</tr>
<tr>
<td>Cons. hyp. protein$^b$</td>
<td>Tcr_0466</td>
<td>157_191</td>
<td>6.37 $\pm$ 1.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>201_235</td>
<td>8.89 $\pm$ 1.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>501_535</td>
<td>7.35 $\pm$ 0.67</td>
</tr>
<tr>
<td>PII</td>
<td>Tcr_1499</td>
<td>165_199</td>
<td>6.83 $\pm$ 2.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>79_113</td>
<td>7.49 $\pm$ 2.19</td>
</tr>
<tr>
<td>AMT</td>
<td>Tcr_1500</td>
<td>502_536</td>
<td>8.18 $\pm$ 0.77</td>
</tr>
<tr>
<td>Cons. hyp. protein$^b$</td>
<td>Tcr_2018</td>
<td>2103_2137</td>
<td>3.62 $\pm$ 0.37</td>
</tr>
</tbody>
</table>

$^a$Fold differences are high-DIC fluorescence values/low-DIC fluorescence values for a particular probe. $^b$Conserved hypothetical protein. CbbM = form II RubisCO, CbbL, CbbS = subunits of form I RubisCO, PII = regulatory protein in the glutamine synthesis cascade, AMT = ammonia transporter

### TABLE 6. Fold changes (low/high-DIC) by microarray

<table>
<thead>
<tr>
<th>Gene product</th>
<th>Locus tag</th>
<th>Probe #</th>
<th>Fold Difference $^a \pm$ SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phage tail</td>
<td>Tcr_0690</td>
<td>91_125</td>
<td>2.23 $\pm$ 0.01</td>
</tr>
<tr>
<td>CS_cbbL</td>
<td>Tcr_0838</td>
<td>1125_1159</td>
<td>4.48 $\pm$ 0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>993_1027</td>
<td>5.21 $\pm$ 0.18</td>
</tr>
<tr>
<td>CS_cbbS</td>
<td>Tcr_0839</td>
<td>177_211</td>
<td>9.24 $\pm$ 0.87</td>
</tr>
</tbody>
</table>

$^a$Fold differences are low-DIC fluorescence values/high-DIC fluorescence values for a particular probe.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Tcr_0000</th>
<th>Start</th>
<th>Stop</th>
<th>Coverage</th>
<th>Avg CD47</th>
</tr>
</thead>
<tbody>
<tr>
<td>CsoS2</td>
<td>Tcr_0840</td>
<td>1086</td>
<td>1120</td>
<td>14.33</td>
<td>± 0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1582</td>
<td>1616</td>
<td>12.88</td>
<td>± 0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1902</td>
<td>1936</td>
<td>2.82</td>
<td>± 0.04</td>
</tr>
<tr>
<td>CsoSCA</td>
<td>Tcr_0841</td>
<td>1180</td>
<td>1214</td>
<td>3.25</td>
<td>± 0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1368</td>
<td>1402</td>
<td>3.72</td>
<td>± 0.04</td>
</tr>
<tr>
<td>CSorfA</td>
<td>Tcr_0842</td>
<td>113</td>
<td>147</td>
<td>5.06</td>
<td>± 0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>193</td>
<td>227</td>
<td>4.22</td>
<td>± 0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53</td>
<td>88</td>
<td>3.27</td>
<td>± 0.62</td>
</tr>
<tr>
<td>CSorfB</td>
<td>Tcr_0843</td>
<td>201</td>
<td>235</td>
<td>2.43</td>
<td>± 0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65</td>
<td>99</td>
<td>2.73</td>
<td>± 0.03</td>
</tr>
<tr>
<td>CsoS1-1</td>
<td>Tcr_0844</td>
<td>26</td>
<td>60</td>
<td>12.78</td>
<td>± 0.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>114</td>
<td>6.08</td>
<td>± 0.25</td>
</tr>
<tr>
<td>CsoS1-2</td>
<td>Tcr_0845</td>
<td>204</td>
<td>238</td>
<td>19.61</td>
<td>± 0.64</td>
</tr>
<tr>
<td>CsoS1-3</td>
<td>Tcr_0846</td>
<td>25</td>
<td>59</td>
<td>14.57</td>
<td>± 0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>273</td>
<td>307</td>
<td>12.57</td>
<td>± 1.40</td>
</tr>
<tr>
<td>BFR</td>
<td>Tcr_0847</td>
<td>145</td>
<td>179</td>
<td>4.47</td>
<td>± 0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>97</td>
<td>131</td>
<td>11.19</td>
<td>± 0.70</td>
</tr>
<tr>
<td>Orf2-11</td>
<td>Tcr_0848</td>
<td>129</td>
<td>165</td>
<td>5.49</td>
<td>± 0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>188</td>
<td>222</td>
<td>3.62</td>
<td>± 0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>236</td>
<td>271</td>
<td>4.94</td>
<td>± 0.03</td>
</tr>
<tr>
<td>Hyp. protein</td>
<td>Tcr_0851</td>
<td>333</td>
<td>367</td>
<td>2.74</td>
<td>± 0.05</td>
</tr>
<tr>
<td>Cons. hyp. protein</td>
<td>Tcr_1019</td>
<td>17</td>
<td>51</td>
<td>2.95</td>
<td>± 0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>89</td>
<td>126</td>
<td>3.01</td>
<td>± 0.02</td>
</tr>
<tr>
<td>Cold Shock Protein</td>
<td>Tcr_1057</td>
<td>925</td>
<td>959</td>
<td>2.60</td>
<td>± 0.14</td>
</tr>
<tr>
<td>Cons. hyp. protein</td>
<td>Tcr_1315</td>
<td>520</td>
<td>554</td>
<td>3.25</td>
<td>± 0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>832</td>
<td>866</td>
<td>2.35</td>
<td>± 0.04</td>
</tr>
</tbody>
</table>
Fold differences are low-DIC fluorescence values/ high-DIC fluorescence values for a particular probe. Conserved hypothetical protein. Hypothetical protein. CScbbL, CScbbS = subunits for carboxysomal form I RubisCO. CsoS2, CSoorfA, CSoorfB, CsoS1-1, CsoS1-2, CsoS1-3, Orf2-11 = carboxysome shell proteins. BRF = bacterial peptide chain release factor. CsoSCA = carboxysomal carbonic anhydrase.

Under low-NH₄⁺ (high-DIC) growth conditions, genes associated with a nitrogen starvation response (1) had increased transcription levels (FIG. 15). This included glutamine synthetase, ammonium transporters Amt I and Amt II, and the regulatory protein PII, which stimulates transcription of the gene encoding glutamine synthetase, as well as post-translational modification of this enzyme via adenylation/deadenylation (10, 15).

FIG. 15. Genes that have increased transcription under high-DIC conditions, as determined with microarrays. ‘High/low’ is the fold change in transcription, calculated by dividing the average spot intensity for microarrays hybridized with labeled mRNA purified from high-DIC cells by the spot intensity for those hybridized with labeled mRNA from low-DIC cells. Multiple bars per gene represent individual probes for that gene. (A.) Genes associated with nitrogen starvation. (B.) RubisCO genes that are preferentially transcribed under high-DIC conditions, where cbbL = noncarboxysomal form I RubisCO large subunit, cbbS = noncarboxysomal form I RubisCO small subunit, cbbM = form II RubisCO. Error bars represent standard error.

The genes encoding the enzyme glutamine synthetase also appeared to have slightly enhanced transcription under these conditions, though the fold changes were not statistically significantly different from 2.0. Genes encoding two
conserved hypothetical proteins, *Tcr_0466*, and *Tcr_2018*, were found to have increased transcription under low-NH$_4^+$ conditions (FIG. 16, Table 6), as confirmed by qRT PCR (Table 7). The function for these conserved hypothetical proteins could not be discerned based on their amino acid sequences.

**FIG. 16.** Genes encoding conserved hypothetical proteins that have increased transcription under high-DIC conditions, as determined with microarrays. ‘High/low’ is the fold change in transcription, calculated by dividing the average spot intensity for microarrays hybridized with labeled mRNA purified from high-DIC cells by the spot intensity for those hybridized with labeled mRNA from low-DIC cells. *Tcr_0466* and *Tcr_2018* are the locus tag designators of the two genes predicted to encode proteins whose function is not possible to infer based on sequence. Multiple bars per gene represent individual probes for that gene Error bars represent standard error.

However, both have predicted aminoterminal signal peptides and an absence of downstream transmembrane helices, suggesting a periplasmic location. The observation that transcription of these genes was insensitive to the DIC concentration when PO$_4$ was the limiting nutrient (Table 7) indicates that stimulated transcription under high-DIC, low-NH$_4^+$ conditions was a response to NH$_4^+$-limitation, not DIC abundance (Table 7).
FIG. 17. Carboxysome genes that are transcribed more frequently under low-DIC conditions, as determined with microarrays. ‘Low/high’ is the fold change in transcription, calculated by dividing the average spot intensity for microarrays hybridized with labeled mRNA purified from low-DIC cells by the spot intensity for those hybridized with labeled mRNA from high-DIC cells. Gene names: CScbbL = large subunit of form I carboxysomal RubisCO, CScbbS = small subunit of form I carboxysomal RubisCO, csoS2 = carboxysome shell protein, csoSCA = carboxysomal carbonic anhydrase, csoS1-1 = carboxysome shell protein. Multiple bars per gene represent individual probes for that gene Error bars represent standard error.

Similar to what has been observed in other microorganisms (9, 47), the genes in the carboxysome operon, including the gene encoding carboxysomal carbonic anhydrase (csoSCA), were transcribed more frequently under low DIC conditions (FIG. 17; Table 6). RubisCO genes were also differentially transcribed: the carboxysomal form I had higher transcription levels under low-DIC conditions while the other two (noncarboxysomal form I RubisCO and form II RubisCO) had increased transcription under high-DIC conditions (FIG. 15). This pattern is consistent with what has been observed in close relative H. marinus, a gammaproteobacterial hydrogen-oxidizing autotroph with an identical RubisCO operon structure (47).

TABLE 7. Response of T. crunogena conserved hypothetical gene transcription to changes in growth conditions

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>response to low DIC</th>
<th>response to low NH$_4^+$</th>
<th>response to low PO$_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fold$^{ab}$ increase</td>
<td>$\Delta C_{\text{T, s.d.}}$</td>
<td>fold$^{ac}$ increase</td>
</tr>
<tr>
<td>Tcr_1019</td>
<td>4.4</td>
<td>-2.1 ± 0.17</td>
<td></td>
</tr>
<tr>
<td>Tcr_1315</td>
<td>6.7</td>
<td>-2.8 ± 0.33</td>
<td></td>
</tr>
<tr>
<td>Tcr_0466</td>
<td>202.72</td>
<td>-7.66 ± 0.96</td>
<td>0.77</td>
</tr>
<tr>
<td>Tcr_2018</td>
<td>6.95</td>
<td>-2.79 ± 1.2</td>
<td>0.69</td>
</tr>
</tbody>
</table>

$^a$qRT PCR was carried out on RNA extracted from low- and high-DIC T. crunogena cells using the 16S gene as the calibrator. $^b$Fold increase is the frequency of transcription in low-DIC cells divided by the frequency in high-DIC cells. $^c$Fold increase is the frequency of transcription in low-NH$_4^+$ cells divided by the frequency in high-NH$_4^+$ cells. $^d$Fold increase is the frequency of transcription in low-PO$_4$ cells divided by the frequency in high-PO$_4$ cells. Locus tag represents gene name.
It is unclear, based on transcription patterns, whether \( \alpha \)- and \( \beta \)-CA play a role in the \( T. \ crunogena \) CCM. \( \alpha \)–CA is transcribed at similar levels under low- and high-DIC conditions (FIG. 18) and does not appear to play a role in DIC uptake and fixation (Dobrinski et al., in prep). The \( \beta \)–CA gene, though found in the form II RubisCO operon (appendix) does not appear to have increased transcription levels under high-DIC growth conditions as demonstrated by microarray (FIG. 18) and confirmed by qRT PCR (Dobrinski et al., in prep).

![Graph showing transcription of carbonic anhydrase-encoding genes under low- and high-DIC conditions as determined with microarrays. ‘Low/high’ is the fold change in transcription, calculated by dividing the average spot intensity for microarrays hybridized with labeled mRNA purified from low-DIC cells by the spot intensity for those hybridized with labeled mRNA from high-DIC cells. csoSCA = carboxysomal carbonic anhydrase. Multiple bars per gene represent individual probes for that gene Error bars represent standard error.]

Two genes that were transcribed more frequently under low-DIC conditions (\( Tcr_1019 \) and \( Tcr_1315 \)) may be novel components of the CCM of \( T. \ crunogena \) (Table 6, Fig. 19). While the function of the protein encoded by \( Tcr_1019 \) is not possible to infer from its sequence, this protein is predicted to have an aminoterminal signal peptide without downstream transmembrane alphahelices, and therefore is likely to be localized in the periplasm. Likewise, for
Tcr_1315, function assignment is not possible, but a likely cellular location can be inferred from the predicted amino acid sequence. It is likely that the protein encoded by this gene would fold into a beta barrel, and would therefore likely be present in the outer membrane, and has a signal peptide which would facilitate its translocation to the periplasm as a step in reaching its final destination. Knock out studies are under way to determine if either protein plays a role in the CCM.

A full understanding of how the CCM works in chemoautotrophs is instrumental in gaining insight into how these organisms are successful in environments such as the hydrothermal vents. Furthermore, identifying proteobacterial CCM components in T. crunogena makes it possible to identify them in other microorganisms. Until then, the understanding of DIC uptake and fixation is quite limited in the many noncyanobacterial autotrophs that fix carbon dioxide in their bizarre, challenging habitats.
References


*marinus* SS120, a nearly minimal oxyphototrophic genome. *PNAS*, 100:10020-10025


Chapter 5

Overall conclusions

It is clear that \textit{T. crunogena} has a CCM. It is able to grow under extremely low-DIC conditions and changes its whole-cell affinity for DIC in response to the DIC concentration available during growth (Chapter 2; 8). Extracellular HCO$_3^-$ and CO$_2$ can both be taken up and fixed by intact \textit{T. crunogena} cells (Chapter 2; 8). \textit{T. crunogena} is capable of generating an intracellular DIC concentration 100X its extracellular concentration when grown under low-DIC conditions and the ability to generate increased intracellular DIC concentration is energy dependent (Chapter 2; 8).

Genomic analysis of \textit{T. crunogena} uncovered some CCM components similar to those that been found in cyanobacteria (appendix). This includes a carboxysome operon which contains genes for shell proteins, RubisCO, and CA. However, some CCM components were not apparent from genome analysis; HCO$_3^-$ transporters orthologous to those present in cyanobacteria are absent.

Three CA genes from the genome were examined to determine whether they play a role in the CCM. The \(\alpha\)-CA is unlikely to play a role as it is equally transcribed under low- and high-DIC conditions. Functional studies indicate that inhibition with AZA does not affect DIC uptake and fixation. Therefore, it is unlikely that \(\alpha\)-CA is fulfilling a role in DIC uptake and fixation (Chapter 3). \(\beta\)-
CA is also unlikely to play a role in the CCM as it is equally transcribed under low- and high-DIC conditions. Even though the β-CA gene is found in the form II RubisCO operon (appendix), there does not appear to be significant increased transcription levels for high-DIC growth conditions as demonstrated by microarray and confirmed by qRT PCR (Chapter 3). CsoSCA does play a role in carbon fixation and is transcribed more frequently under low-DIC conditions. Inhibition by EZA results in inhibition of DIC fixation, but not uptake, which is consistent with its role in carboxysomes in assisting CO₂ fixation by RubisCO (Chapter 4).

Microarrays were used to scan the whole genome for genes transcribed more frequently under low-DIC conditions and possibly elucidate the genes responsible for active bicarbonate transport. As was expected, the genes in the carboxysome operon were transcribed more frequently under low-DIC conditions. Genes encoding two conserved hypothetical proteins (Tcr_1019 and Tcr_1315) were also found to be transcribed more frequently under low-DIC conditions. Under high-DIC, low NH₄⁺ conditions, the non-carboxysomal form I RubisCO and form II RubisCO genes were transcribed more frequently. Genes encoding two conserved hypothetical proteins (Tcr_0466 and Tcr_2018) also appeared to be transcribed more frequently under these conditions.

Autotrophic microorganisms are found in many phyla. In the Bacteria, they are found in the Cyanobacteria, Proteobacteria, Firmicutes, Planctomycetes, Green Sulfur Bacteria, Green Nonsulfur Bacteria, and others. Among Archaea, they are present in both the Crenarchaeota as well as the Euryarchaeota. Based on
extensive physiological and genetic characterization, CCM components (transporters, carboxysomes) can be found encoded in genomes of cyanobacteria. However, CCMs have not been studied in the other phyla. As a result, carboxysomes are the only CCM components that can be found in these noncyanobacterial organisms using genome data. Finding the CCM genes in *T. crunogen* will facilitate the discovery of CCM genes in noncyanobacterial autotrophs. This will substantially enhance the understanding of the ecophysiology of carbon fixation and evolution of mechanisms to cope with periodic or chronic scarcity of dissolved inorganic carbon.

References

hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology* 164: 165-172.


48.
Appendix A

The Genome of Deep Sea Vent Chemolithoautotroph

*Thiomicrospira crunogena* XCL-2
The Genome of Deep-Sea Vent Chemolithoautotroph

*Thiomicrospira crunogena* XCL-2

Running head: *T. crunogena* genome

Kathleen M. Scott*1, Stefan M. Sievert2, Fereniki N. Abril1, Lois A. Ball1, Chantell J. Barrett1, Rodrigo A. Blake1, Amanda J. Boller1, Patrick S. G. Chain3,4, Justine A. Clark1, Carisa R. Davis1, Chris Detter4, Kimberly F. Do1, Kimberly P. Dobrinski1, Brandon I. Faza1, Kelly A. Fitzpatrick1, Sharyn K. Freyermuth5, Tara L. Harmer6, Loren J. Hauser7, Michael Hügler2, Cheryl A. Kerfeld8, Martin G. Klotz9, William W. Kong1, Miriam Land7, Alla Lapidus4, Frank W. Larimer7, Dana L. Longo1, Susan Lucas4, Stephanie A. Malfatti3,4, Steven E. Massey1, Darlene D. Martin1, Zoe McCuddin10, Folker Meyer11, Jessica L. Moore1, Luis H. Ocampo Jr.1, John H. Paul12, Ian T. Paulsen13, Douglas K. Reep1, Qinghu Ren13, Rachel L. Ross1, Priscila Y. Sato1, Phaedra Thomas1, Lance E. Tinkham1, and Gary T. Zeruth1

Biology Department, University of South Florida, Tampa, Florida USA1; Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts USA2; Lawrence Livermore National Laboratory, Livermore, California USA3; Joint Genome Institute, Walnut Creek, California USA4; Department of Biochemistry, University of Missouri, Columbia, Missouri USA5; Division of Natural Sciences and Mathematics, The Richard Stockton College of New Jersey, Pomona, New Jersey USA6; Oak Ridge National Laboratory, Oak Ridge, Tennessee USA7; Molecular Biology Institute, University of California, Los Angeles, California USA8; University of Louisville, Louisville USA9; The Monsanto Company, Ankeny, Ia USA10; Center for Biotechnology, Bielefeld University, Germany11; College of Marine Science, University of South Florida, St. Petersburg, Florida USA12; The Institute for Genomic Research, Rockville, Maryland USA13

*Corresponding author. Mailing address: 4202 East Fowler Avenue; SCA 110; Tampa, FL 33620. Phone: (813)974-5173. Fax: (813)974-3263. E-mail: kscott@cas.usf.edu.
Appendix A continued

Presented here is the complete genome sequence of *Thiomicrospira crunogena* XCL-2, representative of ubiquitous chemolithoautotrophic sulfur-oxidizing bacteria isolated from deep-sea hydrothermal vents. This gammaproteobacterium has a single chromosome (2,427,734 bp), and its genome illustrates many of the adaptations that have enabled it to thrive at vents globally. It has 14 methyl-accepting chemotaxis protein genes, including four that may assist in positioning it in the redoxcline. A relative abundance of CDSs encoding regulatory proteins likely control the expression of genes encoding carboxysomes, multiple dissolved inorganic nitrogen and phosphate transporters, as well as a phosphonate operon, which provide this species with a variety of options for acquiring these substrates from the environment. *T. crunogena* XCL-2 is unusual among obligate sulfur oxidizing bacteria in relying on the Sox system for the oxidation of reduced sulfur compounds. The genome has characteristics consistent with an obligately chemolithoautotrophic lifestyle, including few transporters predicted to have organic allocrits, and Calvin-Benson-Bassham cycle CDSs scattered throughout the genome.
Appendix A continued

Introduction

Deep-sea hydrothermal vent communities are sustained by prokaryotic chemolithoautotrophic primary producers that use the oxidation of electron donors available in hydrothermal fluid (H₂, H₂S, Fe⁺²) to fuel carbon fixation [1,2,3]. The chemical and physical characteristics of their environment are dictated largely by the interaction of hydrothermal fluid and bottom water. When warm, reductant- and CO₂-rich hydrothermal fluid is emitted from fissures in the basalt crust, it creates eddies as it mixes with cold, oxic bottom water. As a consequence, at areas where dilute hydrothermal fluid and seawater mix, a microorganism’s habitat is erratic, oscillating from seconds to hours between dominance by hydrothermal fluid (warm; anoxic; abundant electron donors; 0.02 to > 1mM CO₂) and bottom water (2°C; oxic; 0.02 mM CO₂) [4,5].

Common chemolithoautotrophic isolates from these “mixing zones” from hydrothermal vents include members of the genus *Thiomicrospira*, a group which originally included all marine, spiral-shaped sulfur oxidizing bacteria. Subsequent analyses of 16S rDNA sequences have revealed the polyphyletic nature of this group; members of *Thiomicrospira* are distributed among the gamma and epsilon classes of the Proteobacteria. *T. crunogena*, a member of the cluster of *Thiomicrospiras* in the gamma class, was originally isolated from the East Pacific Rise [6]. Subsequently, *T. crunogena* strains were cultivated or detected with molecular methods from deep-sea vents in both the Pacific and Atlantic, indicating a global distribution for this phylotype [7]. Molecular methods in combination with cultivation further confirmed the ecological importance of *Tms. crunogena* and closely related species at deep-sea and shallow-water hydrothermal vents [8,9].

To provide the energy necessary for carbon fixation and cell maintenance, *T. crunogena* XCL-2 and its close relatives *Tms. spp.* L-12 and MA-3 are capable of using hydrogen sulfide, thiosulfate, elemental sulfur, and sulfide minerals (e.g., pyrite, chalcopyrite) as electron donors; the only electron acceptor they can use is oxygen [6,10,11,12].

Given its temporally variable habitat, *T. crunogena* XCL-2 is likely adapted to cope with oscillations in the availability of the inorganic nutrients necessary for chemolithoautotrophic growth. One critical adaptation in this habitat is its carbon concentrating mechanism [13,14]. This species is capable of rapid growth in the presence of low concentrations of dissolved inorganic carbon, due to an increase in cellular affinity for both HCO₃⁻ and CO₂ under low CO₂ conditions [14]. The ability to grow under low CO₂ conditions is likely an advantage when the habitat is dominated by relatively low CO₂ seawater. Further adaptations in nutrient acquisition and microhabitat sensing are likely to be present in this organism.

*T. crunogena* XCL-2 [15] is the first deep-sea autotrophic hydrothermal vent bacterium to have its genome completely sequenced and annotated. Many other autotrophic bacterial genomes have been examined previously, including several species of cyanobacteria (e.g., [16,17], nitrifiers [18], purple nonsulfur
[19] and green sulfur [20] photosynthetic bacteria, as well as an obligately
chemolithoautotrophic sulfur-oxidizer [21] and a hydrogen-oxidizer [22]. These
genomes have provided insight into the evolution of autotrophy among four of the
seven phyla of Bacteria known to have autotrophic members.

The genome of *T. crunogena* XCL-2 was sequenced to illuminate the
evolution and physiology of bacterial primary producers from hydrothermal vents
and other extreme environments. It was of interest to determine whether any
specific adaptations to thrive in an environment with extreme temporal and spatial
gradients in habitat geochemistry would be apparent from the genome. It was
predicted that comparing its genome both to the other members of the
gammaproteobacteria, many of which are pathogenic heterotrophs, and also to
autotrophs from the Proteobacteria and other phyla, would provide insights into
the evolution and physiology of autotrophs within the Gammaproteobacteria.
Further, this genome provides a reference point for uncultivated (to date)
chemoautotrophic sulfur-oxidizing gammaproteobacterial symbionts of various
invertebrates.

**Results/Discussion**

Genome structure

*Thiomicrospira crunogena* XCL-2 has a single chromosome consisting of 2.43 Mbp,
with a GC content of 43.1% and a high coding density (90.6 %; Figure 1). The
GC skew shifts near the gene encoding the DnaA protein (located at ‘noon’ on the
circular map; *Tcr0001*), and thus the origin of replication is likely located nearby.
One region with a deviation from the average %GC contains a phosphonate
operon and has several other features consistent with its acquisition via horizontal
gene transfer (see ‘Phosphorus Uptake’ below). Many genes could be assigned a
function with a high degree of confidence (Table 1), and a model for cell function
based on these genes is presented (Figure 2).

**TABLE 1. Thiomicrospira crunogena** XCL-2 genome summary

<table>
<thead>
<tr>
<th>Item</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomes</td>
<td>1</td>
</tr>
<tr>
<td>Basepairs</td>
<td>2,427,734</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>43.1</td>
</tr>
<tr>
<td>% coding</td>
<td>90.6</td>
</tr>
<tr>
<td>RNA-encoding genes</td>
<td></td>
</tr>
<tr>
<td>tRNAs</td>
<td>43</td>
</tr>
<tr>
<td>16S-Ile tRNA&lt;sub&gt;GAT&lt;/sub&gt;-Ala tRNA&lt;sub&gt;TGC&lt;/sub&gt;-23S-5S RNA operons</td>
<td>3</td>
</tr>
<tr>
<td>Genes in each COG category</td>
<td></td>
</tr>
<tr>
<td>DNA replication, recombination, and repair</td>
<td>113</td>
</tr>
<tr>
<td>Transcription</td>
<td>84</td>
</tr>
<tr>
<td>Translation, ribosomal structure and biogenesis</td>
<td>153</td>
</tr>
<tr>
<td>Posttranslational modification, protein turnover, chaperones</td>
<td>115</td>
</tr>
</tbody>
</table>
## Appendix A continued

<table>
<thead>
<tr>
<th>Category</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy production and conversion</td>
<td>117</td>
</tr>
<tr>
<td>Carbohydrate transport and metabolism</td>
<td>79</td>
</tr>
<tr>
<td>Amino acid transport and metabolism</td>
<td>167</td>
</tr>
<tr>
<td>Nucleotide transport and metabolism</td>
<td>50</td>
</tr>
<tr>
<td>Lipid transport and metabolism</td>
<td>39</td>
</tr>
<tr>
<td>Coenzyme transport and metabolism</td>
<td>102</td>
</tr>
<tr>
<td>Secondary metabolite biosynthesis, transport, catabolism</td>
<td>37</td>
</tr>
<tr>
<td>Cell wall/membrane/envelope biogenesis</td>
<td>142</td>
</tr>
<tr>
<td>Inorganic ion transport and metabolism</td>
<td>120</td>
</tr>
<tr>
<td>Cell motility</td>
<td>79</td>
</tr>
<tr>
<td>Signal transduction mechanisms</td>
<td>147</td>
</tr>
<tr>
<td>Cell cycle control, cell division, chromosome partitioning</td>
<td>18</td>
</tr>
<tr>
<td>Intracellular trafficking, secretion, and vesicular transport</td>
<td>65</td>
</tr>
<tr>
<td>General function</td>
<td>188</td>
</tr>
</tbody>
</table>
Three rRNA operons are present, and two of them, including their intergenic regions, are 100% identical. In the third rRNA operon, the 16S and 5S genes are 100% identical to the other two, but the 23S gene has a single substitution. The intergenic regions of this third operon also has several substitutions compared to the other two, with three substitutions between the tRNA-Ile-GAT and tRNA-Ala-TGC genes, six substitutions between the tRNA-Ala-TGC and 23S genes, and one substitution between the 23S and 5S genes. Having three rRNA operons may provide additional flexibility for rapid shifts in translation activity in response to a stochastic environment, and may contribute to this species’ rapid doubling times [6]. Forty-three tRNA genes were identified by tRNA-scan SE [23] and Search For RNAs. An additional region of the chromosome was identified by Search For RNAs, the 3’ end of which is 57% identical with the sequence of the tRNA-Asn-GTT gene, but has a 47 nucleotide extension of the 5’ end, and is a likely tRNA pseudogene.

Figure 1. Circular map of the *Thiomicrospira crunogena* XCL-2 genome. The outer two rings are protein-encoding genes, which are color-coded according to COG category. Rings 3 and 4 are tRNA and rRNA genes. Ring 5 indicates the location of a prophage (magenta), phosphonate/heavy metal resistance island (cyan), and four insertion sequences (red; two insertion sequences at 2028543 and 2035034 are superimposed on this figure). The black circle indicates the deviation from the average %GC, and the purple and green circle is the GC
skew (= [G-C]/[G+C]). Both the %GC and GC skew were calculated using a sliding window of 10,000 bp with a window step of 100.

Prophage

A putative prophage genome was noted in the *T. crunogea* chromosome. The prophage genome is 38,090 bp and contains 54 CDSs, 21 of which (38.9%) had significant similarity to genes in GenBank. The prophage genome begins with a tyrosine integrase (*Tcr0656*), and contains a cI-like repressor gene (*Tcr0666*), features common to lambdoid prophages (Figure 3; [24]). These genes define a probable “lysogeny module” [25] and are in the opposite orientation from the rest of the phage genes (the replicative or “lytic module”). The lytic half of the prophage genome encodes putative genes involved in DNA replication and phage assembly (Figure 3). Beginning with a putative DNA primase (*Tcr0668*) is a cluster of genes interpreted to represent an active or remnant DNA replication module (including an exonuclease of DNA polymerase, a hypothetical DNA binding protein, and a terminase large subunit; *Tcr0669, 0670, 0672*). Terminases serve to cut the phage DNA in genome sized fragments prior to packaging. Beyond this are eight CDSs of unknown function, and then two CDSs involved in capsid assembly, including the portal protein (*Tcr0679*) and a minor capsid protein (*Tcr0680*) similar to GPC of λ. Portal proteins are ring-like structures in phage capsids through which the DNA enters the capsid during packaging [26]. In λ, the GPC protein is a peptidase (S49 family) that cleaves the capsid protein from a scaffolding protein involved in the capsid assembly process [27]. Although no major capsid protein is identifiable from bioinformatics, capsid proteins are often difficult to identify from sequence information in marine phages [28]. A cluster of P2-like putative tail assembly and structural genes follows the capsid assembly genes. The general organization of these genes (tail fiber, tail shaft and sheath, and tape measure; *Tcr0691; Tcr0690; Tcr0695; Tcr0698*) is also P2-like [24]. The complexity of these genes (10 putative CDSs involved in tail assembly) and the strong identity score for a contractile tail sheath protein strongly argues that this prophage was a member of the *Myoviridae*, ie. phages possessing a contractile tail. The final gene in the prophage-like sequence was similar to a phage late control protein D, gpD (*Tcr0700*). In λ, gpD plays a role in the expansion of the capsid to accommodate the entire phage genome [29].
Figure 2. Cell model for *Thiomicrospira crunogena* XCL-2, with an emphasis on ultrastructure, transport, energy, carbon metabolism, and chemotaxis. Genes encoding virtually all of the steps for the synthesis of nucleotides and amino acids by canonical pathways are present, and are omitted here for simplicity. Electron transport components are yellow, and abbreviations are: NDH—NADH dehydrogenase; UQ—ubiquinone; bc_{1}—bc_{1} complex; Sox—Sox system; cytC—cytochrome C; cb_b—cb_b-type cytochrome C oxidase. Methyl-accepting chemotaxis proteins (MCP) are fuchsia, as are MCP’s with PAS domains or PAS folds. Influx and efflux transporter families with representatives in this genome are indicated on the figure, with the number of each type of transporter in parentheses. ATP-dependent transporters are red, secondary transporters are sky blue, ion channels are light green, and unclassified transporters are purple. Abbreviations for transporter families are as follows: ABC—ATP-binding cassette superfamily; AGCS—Alanine or glycine:cation symporter family; AMT—Ammonium transporter family; APC—amino acid-polyamine-organocation family; ATP syn—ATP synthetase; BASS—Bile acid:Na^{+} symporter family; BCCT—Betaine/carnitine/choline transporter family; CaCA—Ca^{2+}:cation antiporter family; CDF—cation diffusion facilitator family; CHR—Chromate ion transporter family; CPA—Monovalent cation:proton antiporter-1, -2, and -3 families; DAACS—Dicarboxylate/amino acid:cation symporter family; DASS—Divalent anion:Na^{+} symporter family; DMT—Drug/metabolite transporter superfamily; FeoB—Ferrous iron uptake family; IRT—Iron/lead transporter superfamily; MATE—multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily, MATE family; MscS—Small conductance mechanosensitive ion channel family; MFS—Major facilitator superfamily; MgtE—Mg^{2+} transporter-E family; MIT—CorA metal ion transporter family; NCS2—Nucleobase:cation symporter-2 family; NRAMP—Metal ion transporter family; NSS—Neurotransmitter:sodium symporter family; P-ATP—P-type ATPase superfamily; Pit—Inorganic phosphate transporter family; PNaS—Phosphate:Na^{+} symporter family; PnnC—Nicotamide mononucleotide uptake permease family; RhtB—Resistance to homoserine/threonine family; RND—Resistance-nodulation-cell division superfamily; SSS—Solute:sodium symporter family; SulP—Sulfate permease family; TRAP—Tripartite ATP-independent periplasmic transporter family; TRK—K^{+} transporter family; VIC—Voltage-gated ion channel superfamily.
The high similarity of the CDSs to lambdoid (lysogeny and replication genes) and P2-like (tail module) temperate coliphages is surprising and unprecedented in marine prophage genomes [30]. A major frustration encountered in marine phage genomics is the low similarity of CDSs to anything in GenBank, making the interpretation of the biological function extremely difficult. The lambdoid siphophages are generally members of the *Siphoviridae* whereas the P2-like phages are *Myoviridae*, which the *T. crunogena* XCL-2 prophage is predicted to be. Such a mixed heritage is often the result of the modular evolution of phages. The general genomic organization of the *T. crunogena* XCL-2 prophage-like element (integrase, repressor, DNA replicative genes, terminase, portal, capsid, tail genes) is common to several known prophages, including those of *Staphylococcus aureus* (ie. φMu50B), *Streptococcus pyogenes* (prophages 370.3 and 370.2), and *Streptococcus thermophilus* (prophage O1205; [31]).

**Figure 3.** Prophage genome within the *Thiomicrospira crunogena* XCL-2 genome. Lysogenic and lytic genes are delineated, as are predicted gene functions.

**Redox substrate metabolism and electron transport**

Genes are present in this genome that encode all of the components essential to assemble a fully functional Sox-system that performs sulfite-, thiosulfate-, sulfur-, and hydrogen-sulfide dependent cytochrome c reduction, namely, SoxXA (*Tcr0604, Tcr0601*), SoxYZ (*Tcr0603, Tcr0602*), SoxB (*Tcr1549*), and SoxCD (*Tcr0156, Tcr0157*) [32,33]. This well-characterized system for the oxidation of reduced sulfur compounds has been studied in facultatively chemolithoautotrophic, aerobic, thiosulfate-oxidizing alphaproteobacteria, including *Paracococcus versutus* GB17, *Thiobacillus versutus*, *Starkeya novella* and *Pseudoaminobacter salicylatoxidans* ([32,34] and references therein). This model involves a periplasmic multienzyme complex that is capable of oxidizing various reduced sulfur compounds completely to sulfate. Genes encoding components of this complex have been identified, and it has further been shown that these so-called “sox” genes form extensive clusters in the genomes of the aforementioned bacteria. Essential components of the Sox-system have also been identified in genomes of other bacteria known to be able to use reduced sulfur compounds as electron donors, resulting in the proposal that there might be a common mechanism for sulfur oxidation utilized by different bacteria [32,34]. Interestingly, *T. crunogena* XCL-2 appears to be the first obligate
chemolithoautotrophic sulfur-oxidizing bacterium to rely on the Sox system for oxidation of reduced sulfur compounds.

Figure 4

Figure 4. Phylogenetic relationship of *Thiomicrospira crunogena* XCL-2 SoxB to sequences of selected bacteria. Sequences were aligned using the program package MacVector. Neighbor-joining and parsimony trees based on the predicted amino acid sequences were calculated using PAUP 4.0b10. Bootstrap values (1,000 replicates) are given for the neighbor-joining (first value) and parsimony analyses (second value).

Genome analyses also reveal the presence of a putative sulfide:quinone reductase gene (*Tcr1170*; SQR). This enzyme is present in a number of phototrophic and chemotrophic bacteria and is best characterized from *Rhodobacter capsulatus* [35]. In this organism it is located on the periplasmic surface of the cytoplasmic membrane, where it catalyzes the oxidation of sulfide to elemental sulfur, leading to the deposition of sulfur outside the cells. It seems reasonable to assume that SQR in *T. crunogena* XCL-2 performs a similar function, explaining the deposition of sulfur outside the cell under certain conditions (e.g., low pH or oxygen; [36]). The Sox system, on the other hand, is expected to result in the complete oxidation of sulfide to sulfate. Switching to the production of elemental sulfur rather than sulfate has the advantage that it prevents further acidification of the medium, which ultimately would result in cell lysis. An interesting question in this regard will be to determine how *T. crunogena* XCL-2 remobilizes the sulfur globules. The dependence on the Sox system, and possibly SQR, for sulfur oxidation differs markedly from the obligately autotrophic sulfur-oxidizing betaproteobacterium *Thiobacillus denitrificans*, which has a multitude of pathways for sulfur oxidation, perhaps facilitating this organism’s ability to grow under aerobic and anaerobic conditions [21].

In contrast to the arrangement in facultatively autotrophic sulfur-oxidizers [34], the *sox* components in *T. crunogena* XCL-2 are not organized in a single cluster, but in different parts of this genome: *soxXYZA*, *soxB*, and *soxCD*. In particular, the isolated location of *soxB* relative to other *sox* genes has not been observed in any other sulfur-oxidizing organisms. The components of the Sox
system that form tight interactions \textit{in vivo} are collocated in apparent operons (SoxXYZA, SoxCD; [37]), which is consistent with the ‘molarity model’ for operon function (reviewed in [38]), in which cotranslation from a single mRNA facilitates interactions between tightly-interacting proteins, and perhaps correct folding. Perhaps for obligate chemolithotrophs like \textit{T. crunogena} XCL-2 that do not have multiple sulfur oxidation systems, in which \textit{sox} gene expression is presumably constitutive and not subject to complex regulation [39], \textit{sox} gene organization into a single operon may not be strongly evolutionarily selected. Alternatively, the \textit{T. crunogena} XCL-2 \textit{sox} genes may not be constitutively expressed, and may instead function as a regulon.

The confirmation of the presence of a \textit{soxB} gene in \textit{T. crunogena} XCL-2 is particularly interesting, as it is a departure from previous studies with close relatives. Attempts to PCR-amplify \textit{soxB} from \textit{T. crunogena} ATCC 700270\textsuperscript{T} and \textit{T. pelophila} DSM 1534\textsuperscript{T} were unsuccessful [40]. In contrast, a newly isolated \textit{Thiomicrospira} strain obtained from a hydrothermal vent in the North Fiji Basin, \textit{T. crunogena} HY-62, was positive, with phylogenetic analyses further revealing that its \textit{soxB} was most closely related those from alphaproteobacteria, such as \textit{Silicibacter pomeroyi} [40]. The \textit{soxB} gene from \textit{T. crunogena} XCL-2 falls into a cluster containing the green-sulfur bacterium \textit{Chlorobium} and the purple-sulfur gammaproteobacterium \textit{Allochromatium vinosum}, and separate from the cluster containing \textit{soxB} from \textit{S. pomeroyi} and \textit{T. crunogena} HY-62 (Figure 4). This either indicates that \textit{T. crunogena} XCL-2 has obtained its \textit{soxB} gene through lateral gene transfer from different organisms, or that the originally described \textit{soxB} gene in \textit{T. crunogena} HY-62 was derived from a contaminant. The fact that both \textit{soxA} and \textit{soxX} from \textit{T. crunogena} XCL-2 also group closely with their respective homologs from \textit{Chlorobium} \textit{spp} argues for the latter (data not shown). Also, the negative result for the two other \textit{Thiomicrospira} strains is difficult to explain in light of the observation that sulfur oxidation in \textit{T. crunogena} XCL-2 appears to be dependent on a functional Sox system. It is possible that \textit{T. crunogena} ATCC 700270\textsuperscript{T} and \textit{T. pelophila} DSM 1543\textsuperscript{T} also have \textit{soxB} genes, but that the PCR primers did not target conserved regions of this gene.

Up to this point, obligate chemolithoautotrophic sulfur oxidizers were believed to use a pathway different from the Sox system, i.e., the SI4 pathway [41] or a pathway that represents basically a reversal of dissimilatory sulfate reduction, by utilizing the enzymes dissimilatory sulfite reductase, APS reductase, and ATP sulfurylase [42]. In this context, it is interesting to note that \textit{T. crunogena} also seems to lack enzymes for the assimilation of sulfate, i.e., ATP sulfurylase, APS kinase, PAPS reductase, and a sirohaem-containing sulfite reductase, indicating that it depends on reduced sulfur compounds for both dissimilation and assimilation. \textit{T. crunogena} XCL-2 apparently also lacks a sulfite:acceptor oxidoreductase (SorAB), an enzyme evolutionarily related to SoxCD that catalyzes the direct oxidation of sulfite to sulfate and that has a wide distribution among different sulfur-oxidizing bacteria (Figure S1). The presence of the Sox system and the dependence on it in an obligate chemolithoautotroph also raises the question of the origin of the Sox system. Possibly, this system first evolved in obligate autotrophs before it was transferred into facultative
autotrophs. Alternatively, *T. crunogena* XCL-2 might have secondarily lost its capability to grow heterotrophically.

Genes for Ni/Fe hydrogenase large and small subunits are present (*Tcr2037; Tcr2038*), as well as all of the genes necessary for large subunit metal center assembly (*Tcr2035 - 6; Tcr2039 - 2043*) [43]. Their presence and organization into an apparent operon suggest that *T. crunogena* XCL-2 could use H₂ as an electron donor for growth, as its close relative *Hydrogenovibrio* does [44,45]. However, attempts to cultivate *T. crunogena* with H₂ as the sole electron donor have not been successful [46]. A requirement for reduced sulfur compounds, even when not used as the primary electron donor, is suggested by the absence of genes encoding the enzymes necessary for assimilatory sulfate reduction (APS reductase; ATP sulfurylase), which are necessary for cysteine synthesis in the absence of environmental sources of thiosulfate or sulfide. Alternatively, this hydrogenase could act as a reductant sink under periods of sulfur and oxygen scarcity, when starch degradation could be utilized to replenish ATP and other metabolite pools (see “Central Carbon Metabolism”, below).

The redox partner for the *T. crunogena* XCL-2 hydrogenase is suggested by the structure of the small subunit, which has two domains. One domain is similar to other hydrogenase small subunits, while the other is similar to pyridine nucleotide-disulphide oxidoreductases and has both an FAD and NADH binding site. The presence of a NADH binding site suggests that the small subunit itself transfers electrons between H₂ and NAD(H), unlike other soluble hydrogenases, in which this activity is mediated by separate “diaphorase” subunits [43], which *T. crunogena* XCL-2 lacks. The small subunit does not have the twin arginine leader sequence that is found in periplasmic and membrane-associated hydrogenases [47], suggesting a cytoplasmic location for this enzyme.

All 14 genes for the subunits of an electrogenic NADH:ubiquinone oxidoreductase (NDH-1) are present (*Tcr0817 - 0830*) and are organized in an apparent operon, as in other proteobacteria [48,49]. A cluster of genes encoding an RNF-type NADH dehydrogenase, which is evolutionarily distinct from NDH-1 [50], is present in the *T. crunogena* XCL-2 genome (*Tcr1031 - 1036*), and may shuttle NADH-derived electrons to specific cellular processes (as in [51]).

In this species, ubiquinone ferries electrons between NADH dehydrogenase and the bc₁ complex; all genes are present for its synthesis, but not for menaquinone. Unlike most bacteria, *T. crunogena* XCL-2 does not synthesize the isopentenyl diphosphate units that make up the lipid portion of ubiquinone via the deoxyxylulose 5-phosphate pathway. Instead, most of the genes of the mevalonate pathway (HMG-CoA synthase, *Tcr1719*; HMG-CoA reductase, *Tcr1717*; mevalonate kinase/phosphomevalonate kinase, *Tcr1732, Tcr1733*; and diphosphomevalonate decarboxylase, *Tcr1734*) are present. The single “missing” gene, for acetyl-CoA acetyltransferase, may not be necessary, as HMG-CoA reductase may also catalyze this reaction as it does in *Enterococcus faecalis* [53]. Interestingly, the mevalonate pathway is found in Archaea, eukaryotes, and is common among gram positive bacteria [52,54]. Thus far, the only other proteobacterium to have this pathway is from the alpha class, *Paracoccus zeaxanthinifaciens* [55]. Examination of unpublished genome data
from the Integrated Microbial Genomes webpage (http://img.jgi.doe.gov/v1.1/main.cgi), and queries of Genbank did not uncover evidence for a complete set of genes for the mevalonate pathway in other proteobacteria.

The three components of the bc1 complex are represented by three genes in an apparent operon, in the typical order (Rieske iron-sulfur subunit; cytochrome b subunit; cytochrome c1 subunit; Tcr0991 – 3; [49]). Consistent with its microaerophilic lifestyle and inability to use nitrate as an electron acceptor [6], the only terminal oxidase present in the T. crunogena XCL-2 genome is a cbb3-type cytochrome c oxidase (Tcr1963 - 5). To date, Helicobacter pylori is the only other sequenced organism that has solely a cbb3-type oxidase, and this has been proposed to be an adaptation to growth under microaerophilic conditions [49], since cbb3-type oxidase has a higher affinity for oxygen than aa3-type oxidase does [56].

In searching for candidate cytochrome proteins that facilitate electron transfer between the Sox system and the bc1 complex and cbb3 cytochrome c oxidase, the genome was analyzed to identify genes that encode proteins with heme-coordinating motifs (CxxCH). This search yielded 28 putative heme-binding proteins (Table S1), compared to 54 identified in the genome of T. denitrificans [21]. Thirteen of these genes encode proteins that were predicted to reside in the periplasm, two of which (Tcr0628; Tcr0628) were deemed particularly promising candidates as they met the following criteria: 1) they were not subunits of other cytochrome-containing systems, 2) they were small enough to serve as efficient electron shuttles, 3) they were characterized beyond the level of hypothetical or conserved hypothetical, and 4) they were present in Thiobacillus denitrificans, which also has both a Sox system as well as cbb3 cytochrome c oxidase, and had not been implicated in other cellular functions in this organism. Tcr0628 and Tcr0629 both belong to the COG2863 family of cytochromes c553, which are involved in major catabolic pathways in numerous proteobacteria. Interestingly, genes Tcr0628 and Tcr0629, which are separated by a 147-pb spacer that includes a Shine-Delgarno sequence, are highly likely paralogues and a nearly identical gene tandem was also identified in the genome of T. denitrificans (Tbd2026, Tbd2027). A recent comprehensive phylogenetic analysis of the cytochrome c553 proteins, including the mono-heme cytochromes from T. crunogena and T. denitrificans, revealed existence of a large protein superfamily that also includes proteins in the COG4654 cytochrome c551/c552 protein family (M.G. Klotz and A.B. Hooper, unpublished results). In ammonia-oxidizing bacteria, representatives of this protein superfamily (NE0102, Neut2204, NmulA0344 in the COG4654 protein family; Noc0751, NE0736, Neut1650 in the COG2863 protein family) are the key electron carriers that connect the bc1 complex with complex IV as well as NOx-detoxifying reductases (i.e., NirK, NirS) and oxidases (i.e., cytochrome P460, cytochrome c peroxidase) involved in nitrifier denitrification ([57] and references therein). In Epsilonproteobacteria such as Helicobacter pylori and hepaticus, cytochromes in this family (jhp1148; HH1517) interact with the terminal cytochrome cbb3 oxidase. Therefore, we propose that the expression products of genes Tcr0628 and
Appendix A continued

*Tcr0629* likely represent the electronic link between the Sox system and the *bc1* complex and *cbb3* cytochrome c oxidase in *T. crunogena*. It appears worthwhile to investigate experimentally whether the small difference in sequence between these two genes reflects an adaptation to binding to interaction partners with sites of different redox potential, namely cytochrome c₁ in the *bc1* complex and cytochrome FixP (subunit III) in *cbb3* cytochrome c oxidase.

Given the presence of these electron transport complexes and electron carriers, a model for electron transport chain function is presented here (Figure 2). When thiosulfate or sulfide are acting as the electron donor, the Sox system will introduce electrons into the electron transport chain at the level of cytochrome c [32]. Most will be oxidized by the *cbb3*-type cytochrome c oxidase to create a proton potential. Some of the cytochrome c electrons will be used for reverse electron transport to ubiquinone and NAD⁺ by the *bc1* complex and NADH:ubiquinone oxidoreductase. The NADH created by reverse electron transport must contribute to the cellular NADPH pool, for use in biosynthetic pathways. No apparent ortholog of either a membrane-associated [58] or soluble [59] transhydrogenase is present. A gene encoding a NAD⁺ kinase is present (*Tcr1633*), and it is possible that it is also capable of phosphorylating NADH, as some other bacterial NAD⁺ kinases are [60].

**Transporters and nutrient uptake**

One hundred sixty nine transporter genes from 40 families are present in the *T. crunogena* XCL-2 genome (Figure 5), comprising 7.7% of the CDSs. This low frequency of transporter genes is similar to other obligately autotrophic proteobacteria and cyanobacteria as well as intracellular pathogenic bacteria such as *Xanthomonas axonopodis, Legionella pneumophila, Haemophilus influenzae,* and *Francisella tularensis* (Figure 5; [61,62]). Most heterotrophic gammaproteobacteria have higher transporter gene frequencies, up to 14.1% (Figure 5), which likely function to assist in the uptake of multiple organic carbon and energy sources, as suggested when transporters for sugars, amino acids and other organic acids, nucleotides and cofactors were tallied (Figure 5).
Figure 5

Figure 5. Transporter gene frequencies within the genomes of *Thiomicrospira crunogena* XCL-2 (marked with an arrow) and other proteobacteria. *N. winogradskyi* is an alphaproteobacterium, *N. europaea* is a betaproteobacterium, and *N. oceani* and *M. capsulatus* are gammaproteobacteria. Bars for intracellular pathogens are lighter red than the other heterotrophic gammaproteobacteria.

**Carbon dioxide uptake and fixation**

*T. crunogena* XCL-2, like many species of cyanobacteria [63], has a carbon concentrating mechanism, in which active dissolved inorganic carbon uptake generates intracellular concentrations that are as much as 100X higher than extracellular [14]. No apparent homologs of any of the cyanobacterial bicarbonate or carbon dioxide uptake systems are present in this genome. *T. crunogena* XCL-2 likely recruited bicarbonate and perhaps carbon dioxide transporters from transporter lineages evolutionarily distinct from those utilized by cyanobacteria. Three carbonic anhydrase genes are present (one α-class, *Tcr1545*; two β-class, *Tcr0421, Tcr0841* [64,65,66], one of which (α-class) is predicted to be periplasmic and membrane-associated, and may keep the periplasmic dissolved inorganic carbon pool at chemical equilibrium despite selective uptake of carbon dioxide or bicarbonate. One β-class enzyme gene is located near the gene for a form II RubisCO (see below) and may be coexpressed with it when the cells are grown under high-CO₂ conditions. The other β-class (formerly ε-class; [66]) carbonic anhydrase is a member of a carboxysome operon and likely functions in this organism’s carbon concentrating mechanism. Unlike
many other bacteria [67], the gene encoding the sole SulP-type ion transporter (Tcr1533) does not have a carbonic anhydrase gene adjacent to it.

The genes encoding the enzymes of the Calvin-Benson-Bassham (CBB) cycle are all present. Three ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) enzymes are encoded in the genome: two form I (FI) RubisCOs (Tcr0427-8 and Tcr0838-9) and one form II (FII) RubisCO (Tcr0424). The two FI RubisCO large subunit genes are quite similar to each other, with gene products that are 80% identical at the amino acid level. The FII RubisCO shares only 30% identity in amino acid sequence with both FI enzymes. The operon structure for each of these genes is similar to Hydrogenovibrio marinus [68]: one FI operon includes RubisCO structural genes (cbbL and cbbS) followed by genes encoding proteins believed to be important in RubisCO assembly (cbbO and cbbQ; Tcr429-30) [69,70]. The other FI operon is part of an α-type carboxysome operon (Tcr0840-6) [71] that includes carboxysome shell protein genes csoS1, csoS2, and csoS3 (encoding a β-class carbonic anhydrase; [65,66]). In the FII RubisCO operon, cbbM (encoding FII RubisCO) is followed by cbbO and cbbQ genes, which in turn are followed by a gene encoding a β-class carbonic anhydrase (Tcr0421 – 3) [64]. Differing from H. marinus, the noncarboxysomal FI and FII RubisCO operons are juxtaposed and divergently transcribed, with two genes encoding LysR-type regulatory proteins between them (Tcr0425-6).

The genes encoding the other enzymes of the CBB cycle are scattered in the T. crunogena XCL-2 genome, as in H. marinus [68]. This differs from facultative autotrophic proteobacteria, in which these genes are often clustered together and coregulated [72,73,74]. Based on data from dedicated studies of CBB operons from a few model organisms, it has been suggested that obligate autotrophs like H. marinus do not have CBB cycle genes organized into an apparent operon because these genes are presumably constitutively expressed, and therefore do not need to be coordinately repressed [68].
Figure 6. Calvin-Benson-Bassham cycle gene organization in Proteobacteria.

Rubisco genes (cbbLS and cbbM) are green, phosphoribulokinase genes (cbbP) are red, other genes encoding Calvin-Benson-Bassham cycle enzymes are black, and carboxysome structural genes are grey. For species where cbbP is not near cbbLS or cbbM, the distance from the Rubisco gene to cbbP in kbp is indicated in parentheses. *Thiobacillus denitrificans* has two cbbP genes, so two distances are indicated for this species. Names of organisms that are unable to grow well as organoheterotrophs are boxed. Abbreviations and accession numbers for the 16S sequences used to construct the cladogram are as follows: *A. ehrlichei*—Alkalilimnicola ehrlichei, AF406554; *Brady. sp.—Bradyrhizobium* sp., AF338169; *B. japonicum—Bradyrhizobium japonicum*, D13430; *B. xenovorans—Burkholderia xenovorans*, U86373; *D. aromatica—Dechloromonas aromatica*, AY032610; *M. magneticum—Magnetospirillum magneticum*, D17514; *M. capsulatus—Methylloccus capsulatus BATH*, AF331869; *N. hamburgensis—Nitrobacter hamburgensis*, L11663; *N. winogradskyi—Nitrobacter winogradskyi*, L11661; *N. oceani—Nitrosococcus oceani*, AF363287; *N. europaea—Nitrosomonas europaea*, BX321856; *N. multiformis—Nitrosospira multiformis*, L35509; *P. denitrificans—Paracoccus denitrificans*, X69159; *R. sphaeroides—Rhodobacter sphaeroides*, CP000144; *R. ferrireducens—Rhodoferax ferrireducens*, AF435948; *R. palustris—Rhodopseudomonas palustris*, NC 005296; *R. rubrum—Rhodospirillum rubrum*, D30778; *R. gelatinosus—Rubrivivax gelatinosus*, M60682; *S. meliloti—Sinorhizobium meliloti*, D14509; *T. denitrificans—Thiobacillus denitrificans*, AJ43144; *T. crunogena—Thiomicrospira crunogena*, AF064545. The cladogram was based on an alignment of 162 bp of the 16S rRNA genes, and is the most parsimonious tree (length 2735) resulting from a heuristic search with 100 replicate random step-wise addition and TBR branch swapping (PAUP*4.0b10; [113]). Sequences were aligned using ClustalW [114], as implemented in
BioEdit. Percent similarities and identities for \textit{cbbL}, \textit{cbbM}, and \textit{cbbP} gene products, as well as gene locus tags, are provided as supporting information (Table S4).

Experimental evidence suggests that the CBB cycle is constitutively expressed in \textit{T. crunogena} XCL-2. This species cannot grow chemoorganoheterotrophically with acetate, glucose, or yeast extract as the carbon and energy source ([10]; Table S2). When grown in the presence of thiosulfate and dissolved inorganic carbon, Rubisco activities were high both in the presence and absence of these organic carbon sources in the growth medium (Table S3).

Many sequenced genomes from autotrophic bacteria have recently become available and provide a unique opportunity to determine whether CBB gene organization differs among autotrophs based on their lifestyle. Indeed, for all obligate autotrophs, RubisCO genes are not located near the genes encoding the other enzymes of the CBB cycle (Figure 6; Table S4). For example, the distance on the chromosome of these organisms between the genes encoding the only two enzymes unique to the CBB cycle, RubisCO (\textit{cbbLS} and/or \textit{cbbM}) and phosphoribulokinase (\textit{cbbP}), ranges from 139 – 899 kbp in Proteobacteria, and 151 – 3206 kbp in the Cyanobacteria. In contrast, for most facultative autotrophs, \textit{cbbP} and \textit{cbbLS} and/or \textit{cbbM} genes are near eachother (Figure 6); in most cases, they appear to coexist in an operon. In the facultative autotroph \textit{Rhodospirillum rubrum}, the \textit{cbbM} and \textit{cbbP} genes occupy adjacent, divergently transcribed operons (\textit{cbbRM} and \textit{cbbEFPT}). However, these genes are coordinately regulated, since binding sites for the regulatory protein \textit{cbbR} are present between the operons [75]; perhaps they are coordinately repressed by a repressor protein that binds there as well. The lack of CBB enzyme operons in obligate autotrophs from the Alpha-, Beta-, and Gammaproteobacteria, as well as the cyanobacteria, may reflect a lack of selective pressure for these genes to be juxtaposed in their chromosomes for ease of coordinate repression during heterotrophic growth.

Central carbon metabolism

\textit{3-phosphoglyceraldehyde} generated by the Calvin-Benson-Bassham cycle enters the Embden-Meyerhoff-Parnass pathway in the middle, and some carbon must be shunted in both directions to generate the carbon “backbones” for lipid, protein, nucleotide, and cell wall synthesis (Figure 7). All of the enzymes necessary to direct carbon from \textit{3-phosphoglyceraldehyde} to fructose-6-phosphate and glucose are encoded by this genome, as are all of the genes needed for starch synthesis. To convert fructose 1,6-bisphosphate to fructose 6-phosphate, either fructose bisphosphatase or phosphofructokinase could be used, as this genome encodes a reversible PP$_i$-dependent phosphofructokinase (\textit{Tcr1583}) [76,77]. This store of carbon could be sent back through glycolysis to generate metabolic intermediates to replenish levels of cellular reductant (see below). Genes encoding all of the enzymes necessary to convert \textit{3-phosphoglyceraldehyde} to phosphoenolpyruvate and pyruvate are present, and the pyruvate could enter the citric acid cycle via pyruvate dehydrogenase, as genes encoding all three subunits of this complex are represented (\textit{Tcr1001 - 3}) and activity could be measured with
cell-free extracts of cultures grown in the presence and absence of glucose (Hügler and Sievert, unpublished data).

**Figure 7.** Models for glycolysis, gluconeogenesis, and the citric acid cycle in *Thiomicrospira crunogena* XCL-2. Models for central carbon metabolism for cells under environmental conditions with A. sufficient reduced sulfur and oxygen; B. sulfide scarcity; C. oxygen scarcity; Green arrows represent the two ‘non-canonical’ citric acid cycle enzymes, 2-oxoglutarate oxidoreductase (2-OG OR) and malate: quinone oxidoreductase (MQO).

All of the genes necessary for an oxidative citric acid cycle (CAC) are potentially present, as in some other obligate autotrophs and methanotrophs [18,78]. However, some exceptions from the canonical CAC enzymes seem to be present. The *T. crunogena* XCL-2 genome encodes neither a 2-oxoglutarate dehydrogenase nor a typical malate dehydrogenase, but it does have potential substitutions: a 2-oxoacid:acceptor oxidoreductase (α and β subunit genes in an apparent operon, *Tcr*1709 - 10), and malate: quinone-oxidoreductase (*Tcr*1873), as in *Helicobacter pylori* [79,80]. 2-oxoacid:acceptor oxidoreductase is reversible, unlike 2-oxoglutarate dehydrogenase, which is solely oxidative [79,81]. An overall oxidative direction for the cycle is suggested by malate: quinone oxidoreductase. This membrane-associated enzyme donates the electrons from malate oxidation to the membrane quinone pool and is irreversible, unlike malate dehydrogenase, which donates electrons to NAD⁺ [80]. The 2-oxoacid:acceptor oxidoreductase shows high similarity to the well-characterized 2-oxoglutarate:acceptor oxidoreductase of *Thauera aromatica* [82], suggesting that it might catalyze the conversion 2-oxoglutarate rather than pyruvate as a substrate. However, cell-free extracts of cells grown autotrophically in the presence and absence of glucose have neither 2-oxoglutarate- nor pyruvate:acceptor oxidoreductase activity (Hügler and Sievert, unpubl. data); thus, the citric acid cycle does not appear to be complete under these conditions.

A wishbone-shaped reductive citric acid pathway is suggested by this apparent inability to catalyze the interconversion of succinyl-CoA and 2-
oxoglutarate. However, even though genes are present encoding most of the enzymes of the reductive arm of the reductive citric acid pathway, from oxaloacetate to succinyl CoA (phosphoenolpyruvate carboxylase, Tcr1521; fumarate hydratase, Tcr1384; succinate dehydrogenase/fumarate reductase, Tcr2029-31; succinyl-CoA synthetase; Tcr1373 - 4), the absence of malate dehydrogenase and malic enzyme genes, and the presence of a gene encoding malate:quinone-oxidoreductase (MQO) suggests a blockage of the reductive path as well.

A hypothesis for glycolysis/gluconeogenesis/citric acid cycle function is presented here to reconcile these observations (Figure 7). Under conditions where reduced sulfur compounds and oxygen are sufficiently plentiful to provide cellular reductant and ATP for the Calvin cycle and other metabolic pathways, some carbon would be directed from glyceraldehyde 3-phosphate through gluconeogenesis to starch, while some would be directed to pyruvate and an incomplete citric acid cycle to meet the cell’s requirements for 2-oxoglutarate, oxaloacetate, and other carbon skeletons. Succinyl-CoA synthesis may not be required, as in most bacteria [83], this genome encodes the enzymes of an alternative pathway for porphyrin synthesis via 5-amino levulinate (glutamyl-tRNA synthetase, Tcr1216; glutamyl tRNA reductase, Tcr0390; glutamate 1-semialdehyde 2,1 aminomutase; Tcr0888). Should environmental conditions shift to sulfide scarcity, cells could continue to generate ATP, carbon skeletons, and cellular reductant by hydrolyzing the starch and sending it through glycolysis and a full oxidative citric acid cycle. Should oxygen become scarce instead, cells could send carbon skeletons derived from starch through the incomplete citric acid cycle and oxidize excess NADH via the cytoplasmic Ni/Fe hydrogenase, which would also maintain a membrane proton potential via intracellular proton consumption. Clearly, the exact regulation of the CAC under different growth conditions promises to be an interesting topic for future research.

Genes encoding isocitrate lyase and malate synthase are missing, indicating the absence of a glyoxylate cycle, and consistent with this organism’s inability to grow with acetate as the source of carbon (Table S2).

Nitrogen uptake and assimilation are described in Protocol S1 and Table S5.

**Phosphorus uptake**

*T. crunogena* XCL-2 has all of the genes for the low affinity PiT system (Tcr0543 - 4) and an operon encoding the high affinity Pst system for phosphate uptake (Tcr0537 - 9) [84]. *T. crunogena* XCL-2 may also be able to use phosphonate as a phosphorus source, as it has an operon, phnFDCEGHIJKLMNOP (Tcr2078 – 90), encoding phosphonate transporters and the enzymes necessary to cleave phosphorus-carbon bonds (Figure 8). This phosphonate operon is flanked on either side by large (>6500bp) 100% identical direct repeat elements. These elements encode three predicted coding sequences (Tcr2074 – 6; Tcr2091 - 3): a small hypothetical, and two large (>2500 aa in length) coding sequences with limited similarity to a phage-like integrase present in *Desulfuromonas*.
acetoxidans, including a domain involved in breaking and rejoining DNA (DBR-1, DBR-2). It is interesting to note that two homologs found in the draft sequence of the high GC (~65%) gammaproteobacterium *Azotobacter vinelandii* AvOP have a similar gene organization to the large putative integrases DBR-1/DBR-2. Directly downstream of the first copy of this large repeat element (and upstream of the phosphonate operon) lies another repeat, one of the four IS911-related IS3-family insertion sequences [85] present in this genome (Figure 1). Along with the presence of the transposase/integrase genes and the flanking large repeat element (likely an IS element), the strikingly different G+C of this entire region (39.6%) and the direct repeats (35.9%) compared to the genome average (43.1%) suggest that this region may have been acquired by horizontal gene transfer.

![Figure 8](image)

**Figure 8.** *Thiomicrospira crunogena* XCL-2 phosphonate operon. The DBR-1 genes are identical to each other, as are the DBR-2 genes. Gene abbreviations are: DBR-1 and 2 — DNA breaking-rejoining enzymes; hyp — hypothetical protein; phnFDCEGHJKLMN — phosphonate operon; chp — conserved hypothetical protein. An asterisk marks the location of a region (within and upstream of tRNA-phe) with a high level of similarity to the 5' ends of the two direct repeat sequences noted in the figure. The transposase and integrase are actually a single CDS separated by a frameshift.

Interestingly, immediately downstream of this island lies another region of comparatively low G+C (39.6%) that encodes a number of products involved in metal resistance (e.g., copper transporters and oxidases, heavy metal efflux system). Directly downstream of this second island lies a phage integrase (Tcr2121) adjacent to two tRNAs, which are known to be common phage insertion sites. Strikingly, there is a high level of similarity between the 5' region of the first tRNA – and its promoter region – and the 5' regions of the large repeat elements, particularly the closest element (Figure 8). Taken together, it is proposed that this entire region has been horizontally acquired. Interestingly, it appears that the phosphonate operon from the marine cyanobacterium *Trichodesmium erythraeum* was also acquired by horizontal gene transfer [86]. Phylogenetic analyses reveal that the PhnJ protein of *T. crunogena* XCL-2 falls into a cluster that, with the exception of *Trichodesmium erythraeum*, contains sequences from gamma- and betaproteobacteria, with the sequence of *Thiobacillus denitrificans*, another sulfur-oxidizing bacterium, being the closest relative (Figure S2). The potential capability to use phosphonates, which constitute a substantial fraction of dissolved organic phosphorus [87], might
provide *T. crunogena* XCL-2 a competitive advantage in an environment that may periodically experience a scarcity of inorganic phosphorous. Any excess phosphate accumulated by *T. crunogena* XCL-2 could be stored as polyphosphate granules, as polyphosphate kinase and exopolyphosphatase genes are present (*Tcr1891 - 2*).

**Regulatory and signaling proteins**

Despite its relative metabolic simplicity as an obligate autotroph, *T. crunogena* XCL-2 allocates a substantial fraction of its protein-encoding genes (8.9%) to regulatory and signaling proteins (Table 2). In order to determine whether this was typical for a marine obligately chemolithoautotrophic gammaproteobacterium, the numbers of regulatory and signaling protein-encoding genes from this organism were compared to the only other such organism sequenced to date, *Nitrosococcus oceani* ATCC 19707 [88]. It was of interest to determine whether the differences in their habitats (*T. crunogena*: attached, and inhabiting a stochastic hydrothermal vent environment, vs. *N. oceani*: planktonic, in a comparatively stable open ocean habitat; [89]) would affect the sizes and compositions of their arsenals of regulatory and signaling proteins. Noteworthy differences between the two species include a high proportion of genes with EAL and GGDEF domains in *T. crunogena* XCL-2 compared to *N. oceani* (Table 2). These proteins catalyze the hydrolysis and synthesis of cyclic diguanylate, suggesting the importance of this compound as an intracellular signaling molecule in *T. crunogena* XCL-2 [90]. In some species the abundance of intracellular cyclic diguanylate dictates whether the cells will express genes that facilitate an attached vs. planktonic lifestyle [90]. Given that *T. crunogena* was isolated by collecting scrapings from hydrothermal vent surfaces [6,15], perhaps cyclic diguanylate has a similar function in *T. crunogena* as well.

**TABLE 2.** *Thiomicrospira crunogena* XCL-2a and *Nitrosococcus oceani* ATCC 19707 regulatory and signaling proteins

<table>
<thead>
<tr>
<th>Number:</th>
<th><em>T. crunogena</em></th>
<th><em>N. oceani</em></th>
<th>Transcription/Elongation/Termination Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>104</td>
<td></td>
<td>Transcription/Elongation/Termination Factors</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td></td>
<td>Sigma Factors</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td></td>
<td>Anti/Anti Sigma Factors</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td></td>
<td>Termination/Antitermination Factors</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td>Elongation Factors</td>
</tr>
<tr>
<td>52</td>
<td>76</td>
<td></td>
<td>Transcription factors</td>
</tr>
<tr>
<td>128</td>
<td>75</td>
<td></td>
<td>Signal Transduction proteins</td>
</tr>
</tbody>
</table>

Chemotaxis Signal Transduction

112
Appendix A continued

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1</td>
<td>Methyl-accepting chemotaxis proteins</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>CheA signal transduction histidine kinase</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>CheW protein</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Response regulator receiver modulated CheW protein</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>MCP methyltransferase, CheR-type</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Response regulator receiver, CheY</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Response regulator receiver modulated CheB methylesterase</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>CheD, stimulates methylation of MCP proteins</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>CheZ chemotaxis phosphatase</td>
</tr>
</tbody>
</table>

### Non-Chemotaxis Signal Transduction (101 total, *T. crunogena*)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>18</td>
<td>Signal Transduction Histidine Kinase</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>Diguanylate phosphodiesterase</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Response regulator receiver modulated diguanylate phosphodiesterase</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>Diguanylate cyclase</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Response regulator receiver modulated diguanylate cyclase</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>Diguanylate cyclase with PAS/PAC sensor</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>Diguanylate cyclase/phosphodiesterase</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Periplasmic sensor hybrid histidine kinase and response regulator receiver modulated diguanylate cyclase/phosphodiesterase</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>Diguanylate cyclase/phosphodiesterase with PAS/PAC sensor</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Response regulator receiver modulated diguanylate cyclase/phosphodiesterase</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Cyclic nucleotide-binding protein</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Cyclic-AMP phosphodiesterase</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Adenylate/guanylate cyclase</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>PTS NTR Regulator proteins</td>
</tr>
<tr>
<td>34</td>
<td>29</td>
<td>Miscellaneous</td>
</tr>
</tbody>
</table>

200 179 Total

*A list of locus tags for these genes is present in Table S6*

Many of these EAL and GGDEF-domain proteins, and other predicted regulatory and signaling proteins, have PAS domains (Table 2; Table S6), which often function as redox and/or oxygen sensors by binding redox or oxygen-sensitive ligands (e.g., heme, FAD; [91]). Twenty PAS-domain proteins predicted from *T. crunogena* XCL-2’s genome sequence include 4 methyl-accepting chemotaxis proteins (see below), 3 signal transduction histidine kinases, 6 diguanylate cyclases, and 7 diguanylate cyclase/phosphodiesterases. *N. oceani* has 14 predicted gene products with PAS/PAC domains; notable differences from *T. crunogena* XCL-2 are an absence of PAS/PAC domain methyl-accepting...
chemotaxis proteins, and fewer PAS/PAC domain proteins involved in cyclic
diguanylate metabolism (7 diguanylate cyclase/phosphodiesterases).

Despite its metabolic and morphological simplicity, *T. crunogena* XCL-2
has almost as many genes encoding transcription factors (52) as the cyst and
zoogloea-forming *N. oceani* does (76; Table 2; [89]). Indeed, most free-living
bacteria have a considerably lower frequency of genes encoding regulatory and
signaling proteins (5.6% in *N. oceani* [88]; 5-6% in other species [19]). Other
organisms with frequencies similar to *T. crunogena* XCL-2 (8.6%) include the
metabolically versatile *Rhodopseudomonas palustris* (9.3%; [19]). Although *T.
crunogena* XCL-2 is not metabolically versatile, it has several apparent operons
that encode aspects of its structure and metabolism that are likely to enhance
growth under certain environmental conditions (e.g., carboxysomes; phosphonate
metabolism; assimilatory nitrate reductase; hydrogenase). Perhaps the relative
abundance of regulatory and signaling protein-encoding genes in *T. crunogena*
XCL-2 is a reflection of the remarkable temporal and spatial heterogeneity of its
hydrothermal vent habitat.

**Chemotaxis**

Genes encoding the structural, regulatory, and assembly-related
components of *T. crunogena* XCL-2’s polar flagellae are organized into *flg*
(*Tcr*1464 - 77) and *fla/fli/flh* clusters, similar to *Vibrio* spp. [92]. However, the
*fla/fli/flh* cluster is split into two separate subclusters in *T. crunogena* XCL-2
(*Tcr*0739 – 47; *Tcr*1431 – 53).

Fourteen genes encoding methyl-accepting chemotaxis proteins (MCPs)
are scattered throughout the genome, which is on the low end of the range of
MCP gene numbers found in the genomes of gammaproteobacteria. The function
of MCPs is to act as nutrient and toxin-sensors that communicate with the
flagellar motor via the CheA and CheY proteins [93]. As each MCP is specific to
a particular nutrient or toxin, it is not surprising that *T. crunogena* XCL-2 has
relatively few MCPs, as its nutritional needs as an autotroph are rather simple.
Interestingly, however, the number of MCP genes is high for obligately
autotrophic proteobacteria (Table 2; Figure 9), particularly with respect to those
containing a PAS domain or fold (Figure 9). The relative abundance of MCPs in
*T. crunogena* XCL-2 may be an adaptation to the sharp chemical and redox
gradients and temporal instability of *T. crunogena* XCL-2’s hydrothermal vent
habitat [4].
Figure 9

Figure 9. Numbers of methyl-accepting chemotaxis protein genes in *Thiomicrospira crunogena* XCL-2 and other proteobacteria. *T. crunogena* is marked with an arrow.

Adhesion

A cluster of genes encoding pilin and the assembly and secretion machinery for type IV pili is present (*fip tadE cpaBCEF tadCBD; Tcr1722 - 30*). In *Actinobacillus actinomycetemcomitans* and other organisms, these fimbriae mediate tight adherence to a variety of substrates [94]. *T. crunogena* was originally isolated from a biofilm [6]. Adhesion within biofilms may be mediated by these fimbriae.

Heavy metal resistance

Despite being cultivated from a habitat that is prone to elevated concentrations of toxic heavy metals including nickel, copper, cadmium, lead, and zinc [95, 96], *T. crunogena* XCL-2’s arsenal of heavy metal efflux transporter genes does not distinguish it from *E. coli* and other gammaproteobacteria. It has eleven sets of Resistance-Nodulation-Cell Division superfamily (RND)-type transporters, five Cation Diffusion Facilitator family (CDF) transporters, and six P-type ATPases, far fewer than the metal-resistant *Ralstonia metallidurans* (20 RND, 3 CDF, 20 P-type; [97]), and lacking the arsenate, cadmium, and mercury
Appendix A continued

detoxification systems present in the genome of hydrothermal vent heterotroph *Idiomarina loihiensis* [98]. To verify this surprising result, *T. crunogena* XCL-2 was cultivated in the presence of heavy metal salts to determine its sensitivities to these compounds (Table 3). Indeed, *T. crunogena* XCL-2 is not particularly resistant to heavy metals; instead, it is more sensitive to them than *E. coli* [99]. Similar results were found for hydrothermal vent archaea [100]; for these organisms, the addition of sulfide to the growth medium was found to enhance their growth in the presence of heavy metal salts, and it was suggested that, *in situ* at the vents, sulfide might “protect” microorganisms from heavy metals by complexing with metals or forming precipitates with them [100]. Potentially, this strategy is utilized by *T. crunogena* XCL-2. Alternatively, hydrothermal fluid at its mesophilic habitat may be so dilute that heavy metal concentrations do not get high enough to necessitate extensive adaptations to detoxify them.

### TABLE 3. Growth-inhibiting concentrations (mM) of heavy metals for *Thiomicrospira crunogena* XCL-2 and *Escherichia coli*.

<table>
<thead>
<tr>
<th>Heavy metal ion</th>
<th><em>T. crunogena</em></th>
<th><em>E. coli</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hg$^{2+}$</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Cu$^{2+}$</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>Ag$^{+}$</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Cd$^{2+}$</td>
<td>0.05</td>
<td>0.5</td>
</tr>
<tr>
<td>Co$^{2+}$</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Ni$^{2+}$</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Zn$^{2+}$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cr$^{2+}$</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Mn$^{2+}$</td>
<td>2</td>
<td>20</td>
</tr>
</tbody>
</table>

*a* *T. crunogena* XCL-2 was cultivated on solid thiosulfate-supplemented artificial seawater media with metal salts added to the final concentration listed (0.01 to 20 mM). For both species, the concentration at which growth ceased is listed.

*b* Data from [99].

**Conclusions**

Many abilities are apparent from the genome of *T. crunogena* XCL-2 that are likely to enable this organism to survive the spatially and temporally complex hydrothermal vent environment despite its simple, specialized metabolism. Instead of having multiple metabolic pathways, *T. crunogena* XCL-2 appears to have multiple adaptations to obtain autotrophic substrates. Fourteen methyl-accepting chemotaxis proteins presumably guide it to microhabitats with characteristics favorable to its growth, and type IV pili may enable it to live an attached lifestyle once it finds these favorable conditions. A larger-than-expected arsenal of regulatory proteins may enable this organism to regulate multiple mechanisms for coping with variations in inorganic nutrient availability. Its three RubisCO genes, three carbonic anhydrase genes, and carbon concentrating
mechanism likely assist in coping with oscillations in environmental CO$_2$
availability, while multiple ammonium transporters, nitrate reductase, low- and
high- affinity phosphate uptake systems, and potential phosphonate use, may
enable it to cope with uncertain supplies of these macronutrients.

In contrast, systems for energy generation are more limited, with only one,
i.e., Sox, or possibly two, i.e., Sox plus SQR, systems for sulfur oxidation and a
single low-oxygen adapted terminal oxidase (cbb$_3$-type). Instead of having a
branched electron transport chain with multiple inputs and outputs, this organism
may use the four PAS-domain or –fold methyl-accepting chemotaxis proteins to
guide it to a portion of the chemocline where its simple electron transport chain
functions. It is worth noting, in this regard, that *Thiobacillus denitrificans*, which
has several systems for sulfur oxidation, has fewer MCPs than *T. crunogena*
XCL-2 (Figure 9). Differential expression of portions of the citric acid cycle may
enable it to survive periods of reduced sulfur or oxygen scarcity during its
‘transit’ to more favorable microhabitats.

Up to this point, advances in our understanding of the biochemistry,
genetics, and physiology of this bacterium have been hampered by a lack of a
genetic system. The availability of the genome has provided an unprecedented
view into the metabolic potential of this fascinating organism and an opportunity
use genomics techniques to address the hypotheses mentioned here and others as
more autotrophic genomes become available.

**Materials and Methods**

**Library construction, sequencing, and sequence quality.** Three DNA
libraries (with approximate insert sizes of 3, 7, and 35 kb) were sequenced using
the whole-genome shotgun method as previously described [18]. Paired-end
sequencing was performed at the Production Genomics Facility of the Joint
Genome Institute (JGI), generating greater than 50,000 reads and resulting in
approximately 13X depth of coverage. An additional ~400 finishing reads were
sequenced to close gaps and address base quality issues. Assemblies were
accomplished using the PHRED/PHRAP/CONSED suite [101,102,103], and gap
closure, resolution of repetitive sequences and sequence polishing were performed
as previously described [18].

**Gene identification and annotation.** Two independent annotations were
undertaken: one by the Genome Analysis and System Modeling Group of the
Life Sciences Division of Oak Ridge National Laboratory (ORNL), and the other
by the University of Bielefeld Center for Biotechnology (CeBiTec). After
completion, the two annotations were subjected to a side-by-side comparison, in
which discrepancies were examined and manually edited.

Annotation by ORNL proceeded similarly to [18] and is briefly described
here. Genes were predicted using GLIMMER [104] and CRITICA [105]. The
lists of predicted genes were merged with the start site from CRITICA being used
when stop sites were identical. The predicted coding sequences were translated
and submitted to a BLAST analysis against the KEGG database [106]. The
BLAST analysis was used to evaluate overlaps and alternative start sites. Genes with large overlaps where both had good (1e-40) BLAST hits were left for manual resolution. Remaining overlaps were resolved manually and a QA process was used to identify frameshifted, missing, and pseudogenes. The resulting list of predicted coding sequences were translated and these amino acid sequences were used to query the NCBI nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. PFam and TIGRFam were run with scores > trusted cutoff scores for the HMMs. Product assignments were made based on the hierarchy of TIGRFam, PRIAM, Pfam, Smart (part of InterPro), UniProt, KEGG, and COGs.

Annotation by CeBiTec began by calling genes using the REGANOR strategy [107], which is based on training GLIMMER [104] with a positive training set created by CRITICA [105]. Predicted coding sequences were translated and these amino acid sequences were used to query the NCBI nonredundant database, SwissProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. Results were collated and presented via GenDB [108] for manual verification. For each gene, the list of matches to databases was examined to deduce the gene product. Specific functional assignments suggested by matches with SwissProt and the NCBI nonredundant database were only accepted if they covered over 75% of the gene length, had an e-value < 0.001, and were supported by hits to curated databases (Pfam or TIGRFam, with scores > trusted cutoff scores for the HMMs), or were consistent with gene context in the genome (e.g., membership in a potential operon with other genes with convincing matches to curated databases). When it was not possible to clarify the function of a gene based on matches in SwissProt and the nonredundant database, but evolutionary relatedness was apparent (e.g., membership in a Pfam with a score > trusted cutoff score for the family HMM), genes were annotated as members of gene families.

When it was not possible to infer function or family membership, genes were annotated as encoding hypothetical or conserved hypothetical proteins. If at least three matches from three other species that covered >75% of the gene’s length were retrieved from SwissProt and the nonredundant database, the genes were annotated as encoding conserved hypothetical proteins. Otherwise, the presence of a Shine-Dalgarno sequence upstream from the predicted start codon was verified and the gene was annotated as encoding a hypothetical protein. For genes encoding either hypothetical or conserved hypothetical proteins, the cellular location of their potential gene products was inferred based on TMHMM and SignalP [109,110]. When transmembrane alpha helices were predicted by TMHMM, the gene product was annotated as a predicted membrane protein. When SignalP Sigpep probability and max cleavage site probability were both >0.75, and no other predicted transmembrane regions were present, the gene was annotated as a predicted periplasmic or secreted protein.

Comparative genomics. All CDSs for this genome were used to query the TransportDB database [111]. Matches were assigned to transporter families to facilitate comparisons with other organisms within the TransportDB database [http://www.membranetransport.org/]. To compare operon structure for genes encoding the Calvin-Benson-Bassham cycle, amino acid biosynthesis,
Appendix A continued

phosphonate metabolism, and to find all of the genes encoding methyl-accepting chemotaxis proteins, BLAST-queries of the microbial genomes included in the Integrated Microbial Genomes database were conducted [112]. Comparison of operon structure was greatly facilitated by using the “Show Neighborhoods” function available on the IMG website (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi).

Nucleotide sequence accession number. The complete sequence of the *T. crunogena* XCL-2 genome is available from the nonredundant database (GenBank accession number CP000109).
Supplemental figures and tables

Supporting protocol S1. Nitrogen uptake and assimilation

*Thiomicrospira crunogena* XCL-2 is capable of growing with nitrate or ammonia as its nitrogen source ([6]; K. Scott, unpubl. data). Accordingly, it has an apparent operon encoding the components of a NasFED-type nitrate transporter (*Tcr1153 - 5*) [115], cytoplasmic assimilatory nitrate (*nasA; Tcr1159*) and nitrite reductase (*nirBD; Tcr1157-8*) genes, as well as four Amt-family ammonia transporters (*Tcr0954; Tcr1340; Tcr1500; Tcr2151*).

Ammonia originating from environmental sources or produced from nitrate reduction is incorporated into the *T. crunogena* XCL-2 organic nitrogen pool by glutamine synthetase and NADPH-dependent glutamate synthase. *T. crunogena* XCL-2 has three different glutamine synthetase genes: one encodes a GlnA-type enzyme (*Tcr0536*) while the others are both GlnT-type (*Tcr1347, Tcr1798*) [116]. Perhaps these three glutamine synthetase genes are differentially expressed under different nitrogen conditions.

Genes encoding the majority of the enzymes necessary to synthesize all 20 L-amino acids were detected; exceptions and omissions are described here. The first enzyme of aromatic amino acid synthesis, 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthetase, is encoded by a single gene. More commonly, bacteria have 2-3 isoforms of this enzyme, which are differentially regulated by the concentrations of tyrosine, phenylalanine, and tryptophan [117]. *T. crunogena* XCL-2 also has a single copy of the genes encoding the large and small subunits of acetolactate synthase, which is the “gatekeeper” enzyme of valine, leucine, and isoleucine biosynthesis, while *E. coli* has three isoforms that are differentially sensitive to leucine, valine and isoleucine [118]. It would be of interest to determine whether the single isoforms of these enzymes present in *T. crunogena* XCL-2 are sensitive to the concentration of any of their three amino acid endproducts.

Only a single gene was identified for chorismate mutase, as part of a PheA-like bifunctional protein for the production of phenylpyruvate for phenylalanine synthesis. How the prephenate necessary for tyrosine synthesis is supplied is not apparent. The gene encoding 4-hydroxyphenylpyruvate-producing prephenate dehydrogenase (*tyrC*) does not appear to be bifunctional, as it lacks a domain that could catalyze the chorismate mutase step. Searches for *aroQ* and *aroH*-type chorismate mutase genes [119] yielded only *pheA*.

With regard to lysine synthesis, the identity of the gene encoding N-succinyl-LL-diaminopimelate aminotransferase is not clear. Currently, the gene (*dapC*) encoding this enzyme has only been unambiguously identified in *Bordetella pertussis* [120]. A BLAST search of the *T. crunogena* XCL-2 genome with the *B. pertussis* DapC amino acid sequence did not yield an apparent homolog. A gene identified as *argD* is present, which encodes N-acetyl-ornithine aminotransferase, which can also catalyze the DapC reaction [121]. However, it is unclear whether the *argD* gene product functions in lysine synthesis, as cell extracts from other species contain a heretofore unidentified enzyme specific to N-succinyl-LL-diaminopimelate whose gene locus has not been determined [121].
Alanine synthesis is perplexing in this organism, as genes encoding alanine dehydrogenase, alanine-oxo-acid transaminase, alanine transaminase and alanine-glyoxylate transaminase are all absent, nor are the enzymes present necessary to synthesize alanine from cysteine. Perhaps one of the many transaminase genes detected, the specificity of which could not be deduced based on its sequence, is capable of using pyruvate as a substrate.

Amino acid degradation capabilities are quite limited, as in the obligate chemoautotroph *Nitrosomonas europaea* [18]. The genome sequence was searched for amino acid-catabolizing enzymes common in other bacteria, and many were lacking. For example, both histidine and serine ammonia-lyase are absent. Enzymes necessary for histidine degradation via urocanate or histamine are all absent. Aromatic amino acids cannot be degraded, based on the lack of genes encoding tryptophanase, tryptophan monoxygenase, aromatic-L-amino-acid decarboxylase, aromatic-amino-acid transaminase, tyrosine aminotransferase, or any of the other enzymes necessary to recycle their carbon skeletons through central carbon metabolism. Carbon skeletons from valine, isoleucine, and leucine cannot be routed to central carbon metabolism via acetyl-CoA or succinyl-CoA, as acyl-CoA dehydrogenase and enoyl-CoA hydratase genes are not present.

With the exceptions of glutamine and asparagine, aminoacyl-tRNA synthetases for all amino acids are present. Two copies of the *gltX* gene are present, and it is not possible to deduce whether they have differential specificities for glutamine or glutamate based on their sequences. In *T. crunogena* XCL-2, as in other organisms [122], GltX is probably capable of aminoacylating both tRNA(glu) and tRNA(gln) with glutamate, as glutaminyl-tRNA synthetase is missing. Genes for the three subunits of a tRNA-dependent amidotransferase (adt) are present, and their gene products could convert the glutamyl-tRNA(glu) and aspartyl-tRNA(asp) to glutaminyl-tRNA(gln) and aspartanyl-tRNA(asn). Glycyl- and phenylalanyl-tRNA synthetases are present as heterodimeric forms. In addition to a canonical histidinyl-tRNA synthetase gene, a paralog is present (*hisZ*) and may function as a regulatory protein that interacts with ATP phosphoribosyltransferase, which catalyzes the first step of histidine synthesis [123]. The paralog may act as an intracellular histidine sensor [123]. Two copies of lysyl-tRNA synthetase genes are also present. One is truncated (as in [124]), and may have a different function.

The arginine, tryptophan, leucine, and histidine operons present in the Enterobacteria, Pasteurellales, Vibrionales, and Alteromonidales are disrupted in *T. crunogena* (Table S1), as they are in other “deeply branching” members of the γ-Proteobacteria (*Nitrosococcus oceanii, Methylococcus capsulatus* Bath, and *Pseudomonas aeruginosa*; K. Scott, unpubl. data).

**Nucleotide metabolism**

Genes encoding all of the enzymes necessary to synthesize purine and pyrimidine nucleotides are present. However, the salvage enzymes are poorly represented, as in *Nitrosomonas europaea* [18]. Adenosine and cytidine deaminases are absent, as are purine or pyrimidine phosphorylases and hydrolases. Since *T. crunogena* XCL-2 is an obligate autotroph, an extensive arsenal of salvage enzymes may not
be necessary, as it probably does not rely on environmental sources of these compounds either for nucleotide synthesis, or to use as carbon sources.

**TABLE S1.** Proteins with a heme-coordinating motif (CxxCH)

<table>
<thead>
<tr>
<th>Predicted location</th>
<th>Locus tag #</th>
<th>Predicted function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoplasm</td>
<td>Tcr0028</td>
<td>Cyclic diguanylate phosphodiesterase</td>
</tr>
<tr>
<td></td>
<td>Tcr0157</td>
<td>SoxD</td>
</tr>
<tr>
<td></td>
<td>Tcr0324</td>
<td>UvrABC system protein A</td>
</tr>
<tr>
<td></td>
<td>Tcr0387</td>
<td>ThiF family protein</td>
</tr>
<tr>
<td></td>
<td>Tcr0719</td>
<td>DNA-directed DNA polymerase</td>
</tr>
<tr>
<td></td>
<td>Tcr0871</td>
<td>Chaperone protein DnaJ</td>
</tr>
<tr>
<td></td>
<td>Tcr0994</td>
<td>DNA repair protein RadA</td>
</tr>
<tr>
<td></td>
<td>Tcr1506</td>
<td>Ribonuclease, Rne/Rng family protein</td>
</tr>
<tr>
<td></td>
<td>Tcr1803</td>
<td>Beta-lactamase family protein</td>
</tr>
<tr>
<td></td>
<td>Tcr1975</td>
<td>Conserved hypothetical protein</td>
</tr>
<tr>
<td>Membrane</td>
<td>Tcr0993 bc1</td>
<td>Complex cytochrome c₁ subunit</td>
</tr>
<tr>
<td></td>
<td>Tcr1963 cbb3</td>
<td>Cytochrome c oxidase CcoP subunit</td>
</tr>
<tr>
<td></td>
<td>Tcr1964 cbb3</td>
<td>Cytochrome c oxidase CcoO subunit</td>
</tr>
<tr>
<td></td>
<td>Tcr0776</td>
<td>Diguanylate cyclase/phosphodiesterase</td>
</tr>
<tr>
<td></td>
<td>Tcr1187</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>Periplasm</td>
<td>Tcr0601</td>
<td>SoxA</td>
</tr>
<tr>
<td></td>
<td>Tcr0604</td>
<td>SoxX</td>
</tr>
<tr>
<td></td>
<td>Tcr0111</td>
<td>Thioredoxin-related protein</td>
</tr>
<tr>
<td></td>
<td>Tcr0865</td>
<td>Di-heme hypothetical protein</td>
</tr>
<tr>
<td></td>
<td>Tcr1063</td>
<td>Penta-heme hypothetical protein</td>
</tr>
<tr>
<td></td>
<td>Tcr2105</td>
<td>COG3258 single domain family</td>
</tr>
<tr>
<td></td>
<td>Tcr1575</td>
<td>COG2010 Ccca protein family; Cu detox.?</td>
</tr>
<tr>
<td></td>
<td>Tcr2114</td>
<td>“</td>
</tr>
<tr>
<td></td>
<td>Tcr1266</td>
<td>Hypothetical protein; similar to Daro3004 and Tbd1840</td>
</tr>
<tr>
<td></td>
<td>Tcr1885</td>
<td>Hypothetical small mono-heme cytochrome protein</td>
</tr>
<tr>
<td></td>
<td>Tcr0952</td>
<td>COG2863 family of cytochromes c553</td>
</tr>
<tr>
<td></td>
<td>Tcr0628</td>
<td>“ and likely to transfer electrons to the bc₁ complex and cbb₃ cytochrome c oxidase</td>
</tr>
<tr>
<td></td>
<td>Tcr0629</td>
<td>“</td>
</tr>
</tbody>
</table>
TABLE S2. Growth of *Thiomicrospira crunogena* XCL-2 on solid artificial seawater medium supplemented with carbon and electron sources

<table>
<thead>
<tr>
<th>Supplement</th>
<th>Air (0.03% CO₂) headspace</th>
<th>CO₂-free air headspace</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glucose</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Acetate</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thiosulfate</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Thiosulfate + glucose</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Thiosulfate + acetate</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Thiosulfate + yeast extract</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^{a}\)Cells were cultivated in chemostats under dissolved inorganic carbon limitation in artificial seawater medium supplemented with thiosulfate [14], and were spread on solid artificial seawater media supplemented with thiosulfate and organic carbon sources (40 mM thiosulfate, 0.02% w/v acetate, 0.02% w/v glucose, or 0.1% w/v yeast extract). Two replicates of each plate were made: one set was incubated with an air headspace while the other set was incubated in a BBL Gaspak jar continuously purged with air that had passed through a 30 cm column of soda lime to remove atmospheric CO₂.

\(^{b}\)Symbols indicate the presence (+) and absence (-) of growth twelve days after inoculation.

TABLE S3. *Thiomicrospira crunogena* XCL-2 Rubisco activity when grown in the presence and absence of organic carbon

<table>
<thead>
<tr>
<th>Supplement</th>
<th>Rubisco activity (μmol min⁻¹ mg protein⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.41 ± 0.02</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.32 ± 0.04</td>
</tr>
<tr>
<td>Acetate</td>
<td>0.36 ± 0.05</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>0.42 ± 0.05</td>
</tr>
</tbody>
</table>

\(^{a}\)Cells were grown in chemostats in artificial seawater supplemented with thiosulfate (40 mM) under dissolved inorganic carbon limitation [14], in the presence and absence of organic carbon (0.02% w/v acetate, 0.02% w/v glucose, or 0.1% w/v yeast extract). Cells were harvested (10,000g, 10 min, 4°C),
Appendix A continued

sonicated (three 15-sec sonications in the presence of glass sonication beads), and assayed for protein [14] and ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) activity [125].

**TABLE S4.** Percent similarities and identities of proteobacterial *cbbL*, *cbbM*, and *cbbP* genes

<table>
<thead>
<tr>
<th>Supplement</th>
<th>Rubisco activity (µmol min⁻¹ mg protein⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.41 ± 0.02</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.32 ± 0.04</td>
</tr>
<tr>
<td>Acetate</td>
<td>0.36 ± 0.05</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>0.42 ± 0.05</td>
</tr>
</tbody>
</table>

*a* Cells were grown in chemostats in artificial seawater supplemented with thiosulfate (40 mM) under dissolved inorganic carbon limitation [14], in the presence and absence of organic carbon (0.02% w/v acetate, 0.02% w/v glucose, or 0.1% w/v yeast extract). Cells were harvested (10,000g, 10 min, 4°C), sonicated (three 15-sec sonications in the presence of glass sonication beads), and assayed for protein [14] and ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) activity [125].

**TABLE S5.** Amino acid biosynthesis gene organization in *Thiomicrospira crunogena* XCL-2 vs. *Escherichia coli*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th><em>E. coli</em> gene clusters</th>
<th><em>T. crunogena</em> gene clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>argECBH&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Genes present individually</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>trpE(GD)(CF)BA&lt;sup&gt;b&lt;/sup&gt;</td>
<td>trpEGDC &lt;br&gt; trpFBA</td>
</tr>
<tr>
<td>Threonine</td>
<td>thrABC</td>
<td>thrAC &lt;br&gt; thrB</td>
</tr>
<tr>
<td>Leucine</td>
<td>leuABCD</td>
<td>leuA &lt;br&gt; leuCDB</td>
</tr>
<tr>
<td>Ile, leu, val</td>
<td>ilvEDA</td>
<td>Genes present individually</td>
</tr>
<tr>
<td>Histidine</td>
<td>hisGDC(NB)HAF(IE)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>hisGD</td>
</tr>
</tbody>
</table>
Appendix A continued

<table>
<thead>
<tr>
<th>Library</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hisC</td>
<td></td>
</tr>
<tr>
<td>hisN</td>
<td></td>
</tr>
<tr>
<td>hisBHAFE</td>
<td></td>
</tr>
<tr>
<td>hiSl</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) \textit{argE} is divergently transcribed from GBH

\(^b\) \textit{trp(GD)} and \textit{trp(CF)} are gene fusions of \textit{trpG} and \textit{trpD}, and \textit{trpC} and \textit{trpF}, respectively.  

\(^c\) \textit{his(NB)} and \textit{his(IE)} are gene fusions of \textit{hisN} and \textit{hisB} and \textit{hiSl} and \textit{hisE}. 

### TABLE S6. *Thiomicrospira crunogena* XCL-2 regulatory and signaling proteins

<table>
<thead>
<tr>
<th>Number</th>
<th>Item</th>
<th>Locus Tag Number (Tcr…)*a</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>Transcription/Elongation/Termination Factors</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Sigma Factors</td>
<td>0747  1205  1771  1806  2134  2156</td>
</tr>
<tr>
<td>6</td>
<td>Anti/Anti-Sigma Factors</td>
<td>0730  0972  1038  1390  1225  1478</td>
</tr>
<tr>
<td>6</td>
<td>Termination/Antitermination Factors</td>
<td>0126  0283  0769  1122  1151  1396</td>
</tr>
<tr>
<td>2</td>
<td>Elongation Factors</td>
<td>0515  0880</td>
</tr>
<tr>
<td>52</td>
<td>Transcription factors</td>
<td>0004  0035  0054  0188  0205  0425  0426  0440  0460  0470  0540  0541  0617  0786  0788  0852  0858  0868  0904  0906  0969  1057  1203  1224  1234  1301  1338  1382  1405  1418  1444  1446  1487  1555  1588  1660  1702  1782  1789  1862  1872  1877  1884  1928  1947  2002  2078  2101  2126  2183  2185  2187</td>
</tr>
<tr>
<td>128</td>
<td>Signal Transduction proteins</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Chemotaxis Signal Transduction proteins (27 total)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CheA signal transduction histidine kinase</td>
<td>0750  1612</td>
</tr>
<tr>
<td>3</td>
<td>CheW protein</td>
<td>0728  0751  0755</td>
</tr>
<tr>
<td>2</td>
<td>response regulator receiver modulated CheW protein</td>
<td>1271  1476</td>
</tr>
<tr>
<td>1</td>
<td>MCP methyltransferase, CheR-type</td>
<td>0757</td>
</tr>
<tr>
<td>2</td>
<td>CheY, response regulator receiver</td>
<td>0748  0754</td>
</tr>
<tr>
<td>1</td>
<td>response regulator receiver modulated CheB methylesterase</td>
<td>0758</td>
</tr>
<tr>
<td>1</td>
<td>CheD, stimulates methylation of MCP proteins</td>
<td>1613</td>
</tr>
<tr>
<td>1</td>
<td>CheZ chemotaxis phosphatase</td>
<td>0749</td>
</tr>
</tbody>
</table>

*Chemotaxis Signal Transduction proteins (27 total)*
Appendix A continued

Non-Chemotaxis Signal Transduction (102 total)

17  Signal Transduction Histidine Kinase

0005  0066  0189  0226  0542  0905  0968  1040
1041  1202  1383  1445  1913  1924  2068  2180
2189
Figure S1. Phylogenetic relationships of SoxC sequences of *Thiomicrospira crunogena* XCL-2 with SoxC/SorA sequences of selected bacteria.

Sequences were aligned using the program package MacVector. Neighbor-joining and parsimony trees based on the predicted amino acid sequences were calculated using PAUP 4.0b10. At the base of the three main groups, bootstrap values (1,000 replicates) are given for the neighbor-joining (first value) and parsimony analyses (second value). In (B) bootstrap values are depicted only at the base of the three main groups. *A. thaliana* represents a plant assimilative nitrate reductase and *D. melanogaster* represents a eukaryotic sulfite oxidase.
Figure S2

Figure S2. Phylogenetic relationships of PhnJ sequences of *Thiomicrospira crunogena* with PhnJ sequences of selected bacteria. Sequences were aligned using the program package MacVector. Neighbor-joining and parsimony trees based on the predicted amino acid sequences were calculated using PAUP 4.0b10. Bootstrap values (1000 replicates; neighbor-joining/parsimony analyses) are depicted only at the base of the three main groups and for the branch grouping *Tb. denitrificans* and *T. crunogena*. 
Appendix A continued

References


Appendix A continued

Appendix A continued

the microbiological community, release 314, http://link.springer-
nycom/link/service/books/10125/. New York: Springer-Verlag.
Phylogenetic position of an obligately chemoheterotrophic, marine
hydrogen-oxidizing bacterium, Hydrogenovibrio marinus, on the basis of
16S rRNA gene sequences and two form I RubisCO gene sequences. Arch
Analysis of the molecular species of hydrogenase in the cells of an
obligately chemolithoautotrophic, marine hydrogen-oxidizing bacterium,
nov., sp. nov., a marine obligately chem lithoautotrophic hydrogen-
reactive Ni/Fe-hydrogenase of Wollinella succinogenes. Eur J Biochem
and eukarya and its module common with membrane-bound multisubunit
respiratory chains of the microaerophiles Campylobacter jejuni and
oxidoreductases: an extension to the complex-I family of primary redox
pumps. Biochim Biophys Acta 1505: 45-56.
localization, topology, and mutual stabilization of the rnfABC gene
products in Rhodobacter capsulatus and implications for a new family of
52. Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis:
The evolution of two ancient and distinct pathways across genomes. Proc
Natl Acad Sci USA 97: 13172-13177.
Enterococcus faecalis acetocetate-coenzyme A thiolase/3-hydroxy-3-
methylglutaryl-coenzyme A reductase, a dual-function protein of
Identification, Evolution, and Essentiality of the Mevalonate Pathway for
Isopentenyl Diphosphate Biosynthesis in Gram-Positive Cocci. J Bacteriol
182: 4319-4327.
isoprenoid biosynthesis in Paracoccus zeaxanthinifaciens. Gene 297: 129-
139.
high-affinity cbb3-type cytochrome oxidase terminates the symbiosis-
Appendix A continued


Appendix A continued

Appendix A continued

Appendix A continued


Appendix A continued

Acknowledgments

This work was performed under the auspices of the United States Department of Energy by Lawrence Livermore National Laboratory, University of California, under contract W-7405-ENG-48. Genome closure was funded in part by a USF Innovative Teaching Grant (KMS). SKF, CAK, and KMS gratefully acknowledge support from the United States Department of Agriculture Higher Education Challenge Grants Program (Award # 20053841115876). SMS kindly acknowledges support through a fellowship received from the Hanse Wissenschaftskolleg in Delmenhorst, Germany (http://www.h-w-k.de). MH was supported by a WHOI postdoctoral scholarship. We would like to thank Hannah Rutherford for her assistance in studies to ascertain the sensitivity of *T. crunogena* to heavy metals, Marian Arada for her help in preparing genomic DNA, Jennifer Mobberly for her assistance with inducing phage, and Shana K. Goffredi and Shirley A. Kowalewski for their thoughtful suggestions on this manuscript. Doug Nelson and three anonymous reviewers provided constructive comments that substantially improved the manuscript.
About the author

Kimberly Dobrinski did her undergraduate work at the University of South Florida and graduated December 2000 with a Bachelors of Science degree in Biology. While there she did undergraduate research under Dr. Lee Adair at the University of South Florida Medical School. She began her graduate work at the University of South Florida in 2003 under the supervision of Dr. Jeffrey Yoder studying immune receptors in zebrafish. In 2004 Dr. Yoder left the university, and Kim moved to the lab of Dr. Kathleen Scott where she used Molecular Biology to explore the adaptation of a carbon concentrating mechanism in a chemoautotrophic bacteria from the hydrothermal vents. While pursuing her graduate degree, she was the recipient of a full scholarship to Cold Spring Harbor’s Advanced Bacterial Genetics Course, the University of South Florida’s Distinguished Graduate Achievement Award and also was a winner in the 2008 Poster Symposium and Competition “Global Challenges for the 21st Century”. After graduation, Kimberly will continue her love of science in a postdoctoral fellowship under Dr. Kim Brown and Dr. Charles Lee at Harvard Medical School, Brigham Women’s Hospital. She is married to a wonderful husband, Joseph, has three beautiful children: Ilena, Jason and Michael, and one granddaughter, Emma.