SCHOLAR COMMONS
USF UNIVERSITY OF SOUTH FLORIDA

University of South Florida
Scholar Commons

Graduate Theses and Dissertations

Graduate School

6-24-2014

From CO2 to Cell: Energetic Expense of Creating Biomass Using the Calvin-Benson-Bassham and Reductive Citric Acid Cycles Based on Genomic Data

Mary Ann Mangiapia
University of South Florida, mangiapia@mail.usf.edu

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the Biology Commons, and the Microbiology Commons

From CO$_2$ to Cell: Energetic Expense of Creating Biomass Using the Calvin-Benson-Bassham

and Reductive Citric Acid Cycles Based on Genomic Data

by

Mary Mangiapia

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Biology
with a concentration in Environmental and Ecological Microbiology
Department of Integrative Biology
College of Arts and Sciences
University of South Florida

Major Professor: Kathleen Scott, Ph.D.
Valerie J. Harwood, Ph.D.
James Riordan, Ph.D.

Date of Approval:
June 24, 2014

Keywords: carbon fixation, energy, hydrothermal vents, biosynthesis

Table of Contents

List of Tables

List of Figures

Abstract

The ubiquity of the Calvin-Benson-Bassham cycle (CBB) amongst autotrophic organisms suggests that it provides an advantage over a wide range of environmental conditions. However, in some habitats, such as hydrothermal vents, the reductive citric acid cycle (rCAC) is an equally predominant carbon fixation pathway. It has been suggested that the CBB cycle poses a disadvantage under certain circumstances due to being more energetically demanding compared to other carbon fixation pathways. The purpose of this study was to compare the relative metabolic cost of cell biosynthesis by an autotrophic cell using either the CBB cycle or the rCAC. For both pathways, the energy, in ATP, required to synthesize the macromolecules (DNA, RNA, protein, and cell envelope) for one gram of biomass was calculated, beginning with CO_2. Two sulfur-oxidizing chemolithoautotrophic proteobacteria, *Thiomicrospira crunogena* XCL-2, and *Sulfurimonas autotrophica* were used to model the CBB cycle and rCAC, respectively while *Escherichia coli* was used to model both pathways because it has had its cell composition extremely well-characterized. Since these organisms have had their genomes sequenced, it was possible to reconstruct the biochemical pathways necessary for intermediate and macromolecule synthesis. Prior estimates, based solely on the ATP cost of pyruvate biosynthesis, suggested that the cellular energetic expense for biosynthesis from the CBB cycle was more than that from the rCAC. The results of this study support this conclusion; however the difference in expense between the two pathways may not be as extreme as suggested by pyruvate synthesis. Other factors, such as oxygen sensitivity, may act in concert with energetic expense in contributing to the selective advantages between different autotrophic carbon fixation pathways.

Introduction

Autotrophs use carbon fixation to synthesize cell biomass from inorganic carbon ($= CO_2$ + HCO₃- + CO_3^{-2}). Organic compounds synthesized by autotrophs are cycled into the biosphere and serve as the building blocks for all life on Earth. Six carbon fixation pathways have been elucidated (Berg 2011, Hugler and Sievert 2011) and the presence of organisms that fix carbon without using any of the known pathways suggests that there are more yet to be discovered (Berg 2011). It is puzzling that so many pathways exist in order to perform the same task and it is intriguing that the six known carbon fixation pathways are not equally distributed amongst autotrophs. Because biomass production is an energetically expensive endeavor, it has been suggested that the uneven distribution of carbon fixation pathways may be due to the efficiency with which these pathways operate under various environmental conditions (Hugler and Sievert 2011).

The most well studied and widely distributed carbon fixation pathway, the Calvin-Benson-Bassham cycle (CBB cycle), is a conglomeration of gluconeogenesis and the pentose phosphate pathway (Figure 1) but is identified by the unique enzymes ribulose 1, 5 bisphosphate carboxylase/oxygenase (Rubisco) and phosphoribulokinase. Rubisco incorporates a molecule of carbon dioxide into a molecule of ribulose-1, 5-bisphosphate (RuBP). This carboxylation reaction results in two molecules of 3-phosphoglycerate which are reduced to glyceraldehyde-3-phosphate via the addition of six NADPH and nine ATP molecules. The glyceraldehyde-3-phosphate molecule can be used for biomass production or for the regeneration of RuBP

1

(Bassham 1954). The CBB cycle is primarily used for autotrophic growth; however, it can also be used to remove excess electrons from the cell (Wang 1993). The enzymes of the CBB cycle are oxygen tolerant but oxygen is a competitive substrate with CO_2 for Rubisco and results in a wasteful oxygenase reaction. Aerobic cells must minimize the chances of the oxygenase reaction, which results in the loss of a carbon dioxide molecule as well as the consumption of ATP to regenerate RuBP (Tabita et al. 2008).

The reductive citric acid cycle (rCAC) is a reversal of the oxidative citric acid cycle and fixes carbon dioxide into acetyl-CoA (Figure 1). This cycle requires two NADPH, two to three ATP, one quinol (or other unknown donor for fumarate reductase) and four reduced ferredoxin in order to produce one molecule of pyruvate (Fuchs 2011). The reductive version of the cycle is characterized by the presence of three key enzymes: ATP citrate lyase, α-ketoglutarate synthase, and fumarate reductase. These replace the irreversible enzymes of the oxidative citric acid cycle: citrate synthase, α-ketoglutarate dehydrogenase and succinate dehydrogenase, respectively (Buchanan and Arnon 1990, Evans et al. 1966). Pyruvate synthase is also required for the rCAC and replaces pyruvate dehydrogenase in the oxidative citric acid cycle. Pyruvate synthase and α-ketoglutarate synthase are typically inactivated in the presence of oxygen which generally restricts this cycle to anaerobes and microaerobes (Bar-Even et al. 2012). However, some species such as *Hydrogenbacter thermophilus* and *Aquifex pyrophilus* live in microaerophilic environments and may have oxygen tolerant versions of these enzymes (Shiba et al. 1985, Beh et al. 1993).
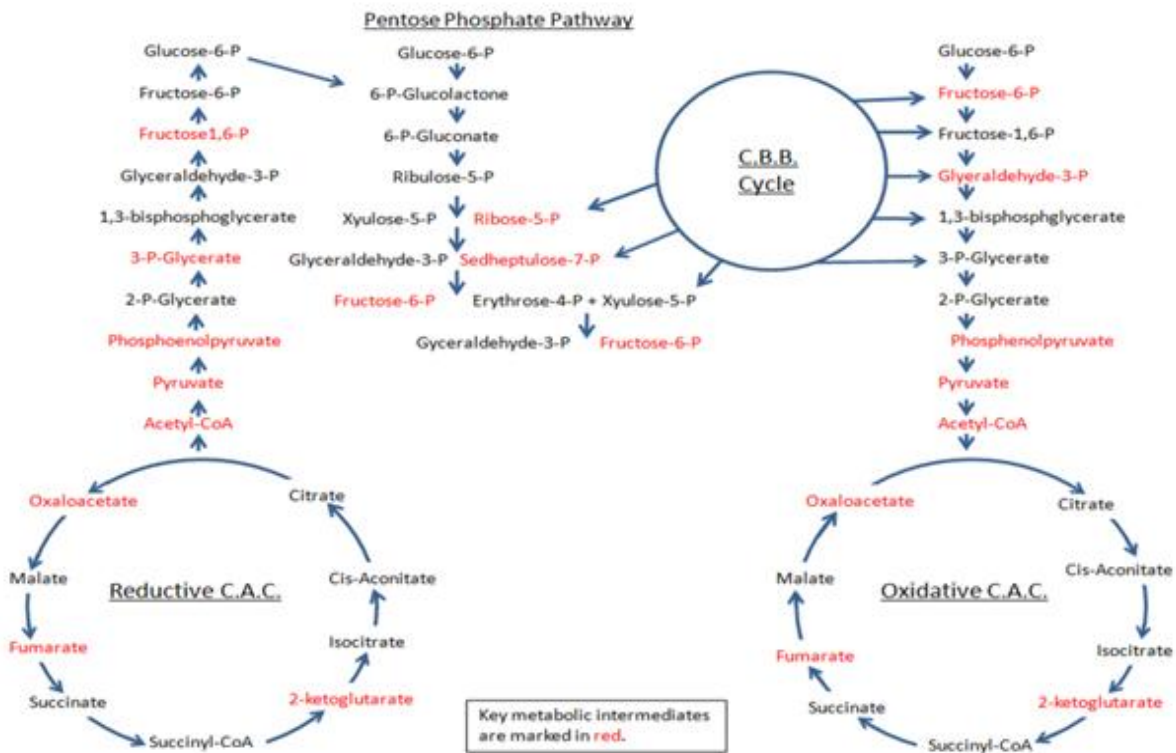
Figure 1| Key metabolic intermediates of the Calvin-Benson-Bassham cycle and the reductive citric acid cycle. These intermediates provide carbon skeletons for the biosynthesis of macromolecules.

The four other known carbon fixation pathways are the reductive acetyl-CoA pathway, or Wood–Ljungdahl pathway (Wood et al. 1986, Ragsdale and Pierce 2008), the 3-hydroxypropionate/ malyl-CoA cycle (Herter et al. 2002) and two variations of the 3-hydroxypropionate cycle: the 3-hydroxypropionate/4- hydroxybutyrate cycle (Berg et al. 2007) and the dicarboxylate/4- hydroxybutyrate cycle (Huber et al. 2008). Like the CBB cycle, the 3-hydroxypropionate/4- hydroxybutyrate cycle operates under aerobic conditions while all of the other known extant pathways function in anaerobic conditions.

The divergence and distribution of these pathways are driven by strong selective forces such as oxygen availability and toxicity as well as the availability of electron donors and the efficiency with which autotrophs can use them. This study investigated the metabolic efficiency of the two most widely studied carbon fixation pathways, the CBB cycle and the rCAC. Hydrothermal vents provide a unique environment to compare the efficiency of these pathways since chemolithoautotrophic primary productivity at these sites is provided predominantly by microorganisms using either the CBB cycle or the rCAC (Campbell and Cary 2004, Hugler et al. 2005, Wirsen et al. 1993, Nakagawa and Takai 2008).

Deep sea hydrothermal vents occur when cold, oxygen-rich ocean water seeps deep into the oceanic crust at crustal spreading centers and subduction zones and interacts with magma or rock heated by the Earth's mantle. This water is heated to over 350º C and undergoes a variety of chemical transformations. The sulfide and metal enriched water then flows out of the crust and mixes with ocean bottom water, both before and after emission from the crust as dilute hydrothermal fluid (Van Dover 2000). As dilute hydrothermal fluid is emitted from the crust and meets cold bottom water, it creates turbulent eddies. Accordingly, temperatures and the correlating chemistry at the vents show variation on timescales ranging from seconds to hours (Johnson et al. 1988, Goffredi et al. 1997).

Ocean bottom water is cold (2 ºC), oxic and alkaline (pH 7.9 to 8.1), while the dilute hydrothermal fluid is warm (2-45ºC), sulfidic, anoxic and acidic (pH 5-8). These two bodies of water differ markedly with respect to dissolved inorganic carbon (DIC; $CO_2 + HCO_3 + CO_3^{-2}$) concentration and composition. Bottom water DIC concentrations are typically around 2 mM and due to the alkaline pH, the dominant forms of DIC are HCO_3^- and CO_3^{-2}. Dilute hydrothermal

fluid DIC concentrations can be quite elevated (2-5 mM DIC) and when coupled with the more acidic pH, result in substantially elevated CO_2 concentrations (Goffredi et al. 1997). Due to the eddies, microorganisms growing attached to surfaces at vents have their habitat alternate between one in which reductant is abundant and the CO_2 concentration is high as 1.8 mM to one where oxidant is abundant and the CO_2 concentration is much lower ($\sim$ 20 µM). Hydrogen levels can vary from near zero up to 2.5 mM (Takai et al. 2004) while hydrogen sulfide levels can vary from 0.33- 1.90 mM/ kg^{-1} (Johnson et al. 1988, Shank et al. 1998).

Two major groups of autotrophic microbes are found at the vents: gammaproteobacteria that use the CBB cycle (Nakagawa and Takai 2008) and epsilonproteobacteria that use the rCAC (Campbell and Cary 2004, Hugler et al. 2005). The gammaproteobacteria are generally found where the vent fluid has mixed thoroughly with ocean bottom water (2-40°C) (Fisher et al. 2007). The epsilonproteobacteria can tolerate a higher range of temperatures, from 20° C to 60°, (Takai et al. 2003) and are often found in warmer, less oxygenated water around the vents (Campbell and Carey 2004). This study utilized representative members from each of these two groups, whose genomes had been sequenced, which facilitated elucidation of biochemical pathways likely to be utilized by the organisms based on the presence of genes encoding for enzymes within the pathway.

The gammaproteobacterium *Thiomicrospira crunogena* XCL-2 is an obligate chemolithoautotroph. This species was originally isolated from vents near the Galapagos Islands (Jannasch et al. 1985) and subsequently has been detected in hydrothermal vents in both the Pacific and the Atlantic oceans. *T. crunogena* is a free-living mesophile (15-35 ℃) that can oxidize a wide variety of sulfur compounds present at the vents such as thiosulfate, hydrogen

sulfide and sulfide minerals but can only reduce oxygen as a final electron acceptor (Scott et al. 2006). It uses the CBB cycle to fix CO_2 (Scott et al. 2006) as it grows rapidly with a minimum doubling time of approximately one hour (Jannasch et al. 1985).

Sulfurimonas autotrophica OK 10 is an epsilonproteobacterium found in deep-sea hydrothermal sediments that grows at an optimal temperature of 25° C but can handle temperatures ranging from 10-40 °C (Inagaki et al. 2003, Sikorski et al. 2010). *S. autotrophica* oxidizes sulfur, sulfide and thiosulfate in order to power carbon fixation via the rCAC. Despite using a carbon fixation pathway known for having oxygen sensitive enzymes, *S. autotrophica* can only use oxygen as its final electron acceptor (optimal growth 1-15% v/v oxygen) (Inagki et al. 2003).

The presence of two predominant autotrophic pathways at the vents suggests that they are equally successful in allowing autotrophic bacteria to efficiently fix carbon into biomass; however, there is an ongoing controversy suggesting that the rCAC is less bioenergetically expensive than the CBB. The energetic efficiency of carbon fixation pathways is generally compared in terms of the ATP and NADPH required for the synthesis of individual central carbon metabolic intermediates from carbon dioxide. For example, the CBB cycle requires 3 CO_2, 5 NADPH and 7 ATP to produce pyruvate while the rCAC requires 3 CO_2, 2 NADPH, 1 quinol, 2 ATP, and 4 molecules of reduced ferredoxin. (Fuchs 2011) Based on this stoichiometry, it has been suggested that the rCAC is a less energetically expensive pathway based solely on the ATP requirements to synthesize pyruvate (Hugler and Sievert 2011, Markert et al. 2007).

Comparing the efficiencies of carbon fixation pathways based on the production of a single intermediate may not be an accurate reflection of how these pathways are used in nature. Because autotrophic cells use their carbon fixation pathways primarily for the production of biomass, several variables must be taken into account when determining the actual cost incurred by use of a specific pathway. Carbon fixation pathways are used for the production of all central carbon intermediates which are in turn used to produce the entire organic biomass of the cell. The production of different metabolic intermediates from carbon dioxide requires varying amounts of ATP and NADPH. These intermediates are also required in different amounts for the synthesis of macromolecules such as proteins, lipids, and nucleic acids. The composition of macromolecules varies between species which results in even greater variation in metabolic intermediate composition.

The objective of this study was to clarify the degree to which biomass synthesis expense varies between the CBB cycle and the rCAC. The availability of genome data in Integrated Microbial Genomes (IMG- http://img.jgi.doe.gov/) and the Kyoto Encyclopedia of Genes and Genomes (KEGG- http://www.genome.jp/kegg/) allowed for metabolic pathways to be reconstructed for individual species. For comparison, three cell types were synthesized *in silico* from CO_2 using both the CBB cycle and the rCAC. The three cells all had genome sequence data available and together represented a CBB cycle autotroph (*Thiomicrospira crunogena XCL-2*), an rCAC autotroph (*Sulfurimonas autotrophica OK 10*) and a heterotroph (*Escherichia coli*) which was used as it had readily available biomass composition values. Pathway maps were used to trace the biosynthesis of cell biomass, from carbon dioxide and cellular electron carriers, to central carbon intermediates, to building blocks (amino acids, lipid components etc.) to macromolecules (proteins, lipids, nucleic acids etc).

Chapter 2: Materials and Methods

CO₂

CBB Cycle · rCAC

ATP to build each central carbon intermediate from cellular reductant

fructose 6-P	sedoheptulose 7-P
pyruvate	oxaloacetate
2-ketoglutarate	3-phosphoglycerate
acetyl Coa	fumarate
ribose-5-P	erythrose-4-P
	phosphoenolpyruvate

ATP to synthesize building blocks from central carbon intermediates, ATP and cellular reductant

Amino Acids · Nucleotides · Lipid Components · Cell Wall Components

ATP to synthesize macromolecules from building blocks, ATP and cellular reductant

Proteins · Nucleic Acids · Lipids · Peptidoglycan

ATP to synthesize one gram of cell biomass from macromolecules

T. cru · S. auto · E. coli

Figure 2 | Methods overview: Steps used to determine the amount of ATP required to build each cell type using the CBB cycle and the rCAC from CO_2.

Central Carbon Intermediates from CO$_2$

Metabolic pathways maps from the Kyoto Encyclopedia of Genes and Genomes which are based on complete genome sequences were used to determine the pathway for the synthesis of central carbon intermediates from CO$_2$, ATP and cellular reductant using both the CBB cycle and the rCAC (Figure 1). To facilitate pathway comparison, cellular electron carriers were converted to ATP equivalents. Based on genome data, both *T. crunogena* and *S. autotrophica* use the same respiratory electron transport complexes: NDH-1-type NADH dehydrogenase, bc1 complex, and cbb3-type cytochrome oxidase (http://img.jgi.doe.gov). Based on a NDH-1$\rightarrow$ bc1 $\rightarrow$ cbb3 electron transport chain, the number of protons translocated from a cell per two electrons introduced via NAD(P)H oxidation is 12 (4 from NDH-1, 4 from bc1 and 4 from cbb3) (Rauham et al. 2012) The number of protons per electron pair introduced from quinol is 8 (4 from bc-1 and 4 from cbb-3). Assuming 1 ADP phosphorylation per 3 H+ re-entering the cell via ATP synthase, this translates to 4 ATP/ NAD(P)H and 2.67 ATP/ quinol. Since ferredoxin has redox potential suitable to pass electrons to NAD(P)H via transhydrogenase and since 2 ferredoxin are needed to reduce 1 NAD(P)H, 2 ferredoxin were assumed to be equivalent to 1 NAD(P)H (= 4ATP).

Building Blocks from Central Carbon Intermediates

The pathways necessary to synthesize cellular building blocks from central carbon intermediates were identified from KEGG. These building blocks were the components necessary to synthesize macromolecules. For proteins, these building blocks included the twenty canonical amino acids, for nucleic acids they were the five nucleotides and their appropriate deoxy cognates. For the cell membrane, the building blocks were fatty acids, polar lipids and lipopolysaccharides. For the cell wall, the building blocks were N-acetylglucosamine (NAG), N-

9

acetylmuramic acid (NAM), and tetrapeptide linker amino acids. The ATP necessary to

synthesize one mole of each building block from central carbon intermediates was calculated.

Macromolecules from Building Blocks

Proteins

General calculation

Mole % amino acids

E.coli   R. japonicum   S. aureus

×

gram molecular weight

grams amino acid in 1 mol amino acids

÷

[Repeat for all amino acids]

$$\frac{grams\ individual\ amino\ acid}{grams\ mixed\ amino\ acids}$$
(= grams protein)

×

$$\frac{grams\ protein}{gram\ of\ cell\ biomass}$$

$$\frac{grams\ individual\ amino\ acid}{gram\ of\ cell\ biomass}$$

÷

$$\frac{grams\ individual\ amino\ acid}{mole\ amino\ acid}$$

$$\boxed{\frac{moles\ individual\ amino\ acid}{gram\ of\ cell\ biomass}}$$

Example calculation for alanine

$$\frac{13\ moles\ alanine}{100\ moles\ total\ amino\ acids}$$

×

$$\frac{89.1\ grams}{mol\ alanine}$$

11.5 grams of alanine in 1 mol of amino acids

÷

$$\frac{124.5\ grams\ mixed\ amino\ acids}{mol\ amino\ acids}$$

$$\frac{0.092\ grams\ alanine}{grams\ mixed\ amino\ acids}$$

×

$$\frac{0.55\ grams\ protein}{1\ gram\ cell}$$

$$\frac{0.05\ grams\ alanine}{gram\ of\ cell\ biomass}$$

÷

$$\frac{89.1\ grams\ alanine}{mole\ alanine}$$

$$\boxed{\frac{0.000569\ moles\ alanine}{gram\ of\ cell\ biomass}}$$

Figure 3 | Overview of amino acid calculations. On the left, is a general formula for turning the molar percent composition values from literature sources into moles of individual amino acid per gram of cell biomass. A specific example is given for the amino acid alanine.

To estimate the amino acid composition of total cellular protein for all three species,

values were obtained from the limited data available for hydrolyzed protein extracts from

bacteria. Since amino acid composition values are not available for *T. crunogena* and *S. autotrophica*, the moles of each amino acid per mole of total amino acids was averaged from *E. coli (*Roberts et al. 1957)*, Rhizobium japonicum* (Lillich and Elkan 1973) and *Staphylococcus aureus* (Okayasu et al. 1997) (Table 1). Protein hydrolysate literature values do not separate aspartate from asparagine or glutamate from glutamine. Therefore, the individual proportions of these amino acids were determined by obtaining the amino acid sequences of all protein encoding genes from Integrated Microbial Genomes (IMG) and calculating their relative frequencies for the three species used for this study. These frequencies provided a ratio which was multiplied by the asp+asn and glu+gln sums and resulted in the individual proportion of each amino acid.

Table 1 | Moles percent amino acid composition form literature sources.

Amino Acid	*E. coli* (Roberts)	*E. coli* (Okayasu)	*R. japonicum* (Lillich)	*S. aureus* (Okayasu)	Average
Alanine	12.7	11.97	12.3	14.95	13.20
Arginine	5.3	4.4	4.3	2.8	3.98
Cysteine	1.7	0.19	n/a	0.18	0.56
Glycine	7.8	9.07	9.4	16.43	11.42
Histidine	0.97	1.96	1.6	1.53	1.53
Isoleucine	4.6	7	4.6	5.94	5.45
Leucine	7.9	9.46	7.5	6.22	7.47
Lysine	7.0	5.52	9.1	7.82	7.73
Methionine	3.4	2.92	2.7	1.88	2.58
Phenylalanine	3.3	4.04	3.7	3.32	3.56
Proline	4.6	4.37	3.8	3.42	3.90
Serine	6.1	4.87	6.4	4.47	5.45
Threonine	4.7	5.58	5.4	4.23	4.92
Tryptophan	1.04	n/a	n/a	n/a	1.04
Tyrosine	2.1	2.67	2.7	2.22	2.44
Valine	5.5	8.23	6.9	5.88	6.55
Aspartate+asparagine	9.9	11	9	9.9	9.78
Glutamate+glutamine	10.5	11.3	10.6	11.37	10.96

To calculate the moles of ATP necessary to synthesize the protein component of 1 gram of dry cellular biomass, it was necessary to convert the relative abundances of the amino acids into moles of each amino acid.   The above calculations provided molar ratios of each amino acid (e.g., moles of alanine per mole of total amino acids in proteins). The total moles of amino acids necessary to make 1 gram of dry cellular biomass has not been reported. Instead, the amount of protein in 1 gram of dry cellular biomass has been determined (0.55 g protein/ grams dry weight (gdw); Neidhart et al. 1990). Because of this, the initial molar ratios had to be converted to mass ratios, and multiplied by 0.55 g protein/gdw, to obtain the grams of a particular amino acid per gram dry cellular biomass, which could then be converted back to moles of a particular amino acid. The total amino acid value (124 g/mol total amino acids) was calculated by summing the gram weight of all individual amino acids in one mole of protein.

To convert the relative abundances of amino acids into moles of each amino acid, each molar ratio was multiplied by the gmw of the appropriate amino acid (e.g., 0.13 mol alanine/mol amino acids $\times$ 89.1 g alanine/mol alanine = 11.5 g alanine per mol total amino acids). This value was then divided by the combined mass of 1 mol of the amino acid mixture with the composition calculated above (e.g., for alanine, 11.5 g alanine per mol total amino acids $\div$ 124 g total amino acids per mol total amino acids = 0.092 g alanine per g protein). This value was then multiplied by the amount of protein per gram dry cellular biomass to obtain the grams of a particular amino acid per gram dry weight (0.092 g alanine per g protein $\times$ 0.55 g protein per gram dry weight = 0.051 g alanine per gram dry weight). This mass was converted to moles by dividing by the cognate amino acid gram molecular weight (0.051 g alanine per gram dry weight $\div$ 89.1 g alanine/mol alanine = 0.00057 mol alanine per gram dry weight).

The ATP necessary to synthesize enough of each amino acid to synthesize 0.55 g protein was calculated by multiplying the number of moles of each amino acid by the number of moles of ATP necessary to synthesize 1 mol of each amino acid from CO_2 (as calculated above). The sum of these values from each amino acid was then taken to provide an estimate of the total amount of ATP necessary to synthesize the amino acids which constitute 0.55 g protein. To account for the cost of amino acid activation and peptide bond formation, per amino acid, an additional 2 ATP were added to represent the cost of activating the amino acids and forming aminoacyl- tRNAs; an additional ATP was added per amino acid to account for GTP hydrolysis by elongation factor G during ribosome translocation along the mRNA (Novelli, 1967).

Nucleic Acids

The A, T, G and C content of DNA was determined directly from the chromosome sequences of *T. crunogena*, *S. autotrophica* and *E. coli* listed in IMG. To calculate the amount of ATP necessary to synthesize sufficient nucleotide triphosphates to polymerize into DNA, it was necessary to convert the relative abundances of nucleotides into moles of nucleotides. As for proteins, it was necessary to convert molar ratios into mass ratios to make use of the reported amount of DNA per gram dry weight cellular material (0.03 g DNA/ gram dry weight of cell, from *E. coli* (Neidhart et al. 1990). These mass ratios were then converted to moles as described for proteins above.

For example, the moles of deoxyguanine triphosphate (dGTP) per gram dry weight of cellular biomass were calculated as follows. The grams of dGTP per mole of nucleotides in genomic DNA (gDNA) was calculated by multiplying the fraction of gDNA that is dGMP (0.21 in *T. crunogena*) and multiplying by the gram molecular weight of dGMP. This value was then

divided by the sum of the grams of dNTPs per mole of nucleotides in gDNA, to obtain the grams of dGMP per gram gDNA. This value was then multiplied by the amount of DNA per gram cellular biomass (0.03 g/ gram dry cellular biomass; Neidhart et al. 1990) to obtain the grams of dGMP per gram dry cellular biomass. This value was then divided by the gram molecular weight of dGMP to obtain the number of moles of dGMP per gram dry cellular biomass. This process was repeated for dATP, dCTP, and dTTP. For each dNTP, the amount of ATP necessary to synthesize enough for one gram of dry biomass was calculated by multiplying the amount of each dNTP per gram dry biomass by the moles of ATP necessary to synthesize 1 mol of dNTP.

Ribonucleic acids were separated into rRNA, tRNA, and mRNA. The A, U, G and C content of the RNA was determined from the 23S, 16S and 5S gene sequences of *T. crunogena*, *S. autotrophica* and *E. coli* obtained from IMG. These values were converted into mass ratios as described above for dNTPs so that the published amounts of these RNAs per cell could be used to estimate the moles of each NTP per gram dry weight of cell (Neidhart et al. 1990).

Fatty Acids

The composition of fatty acids in cell membrane constituents was based on values published for *E. coli* (Ratledge, C. 1989) and *T. crunogena* (Conway, N. 91). Values for *S. autotrophica* were averaged from other epsilonproteobacteria *Sulfurimonas paralvinellae*, *Sulfurimonas autotrophica*, and *Nitratifractor salsuginis* (Takai et al. 2006). These molar ratios of fatty acids were converted into mass ratios as described above for amino acids and nucleotides, as a step in calculating the moles of each fatty acid per gram dry weight of cellular material by using the published value for grams of fatty acids per gram dry weight of cellular material (Table 2).

Polar Lipids

Polar Lipids were divided into cardiolipin (CL), phosphatidylglcerol (PG), phosphatidylserine (PS) and phosphatidylethanolamine (PE). The relative amount of these components in *E. coli*'s cell membrane has been published (Epand 2010 and Ratledge, C. 1989). For *T. crunogena,* these values were based on those determined for chemolithoautotrophic betaproteobacterium *Thiobacillus neapolitanus* (Ratledge, C. 1989); the pathways necessary to synthesize these polar lipids were verified to be possible in *T. crunogena* by examining KEGG maps. Values for *S. autotrophica* were based on an average of seven species of *Helicobacter* and *Campylobacter jejuni* (Mendz, GL et al.. 200) which are members of the epsilonproteobacteria. These molar compositional data were treated identically to similar data for amino acids and nucleotides above, to obtain the ATP needed to synthesize sufficient polar lipids per gram dry weight of cellular material.

Lipopolysaccharides

Lipopolysaccharides were separated into lipid A plus the core and further broken into the building blocks UDP-n-acetylglucosamine 6-P, ADP-glyceromannoheptose, ethanolamine, UDP-galactose, UDP-glucose, CMP-3-deoxy-D-manno-octulosonate (CMP-KDO), 3-OH-14:0 and 12:0 (Caroff 2003). Cellular lipopolysaccharide composition was based on published values for *E. coli* (Neidhart et al.. 1990) (Table 1) and verified through KEGG. The molar composition was calculated in a similar fashion to data for amino acids and nucleotides in order to obtain the ATP necessary to synthesize the lipopolysaccharide for one gram of cell.

Peptidoglycan

As the contribution of peptidoglycan to total cell mass has not been determined for many organisms, values from *E. coli* were used for all three organisms (Table 6). The genetic capability for peptidoglycan biosynthesis was verified for *S .autotrophica* and *T. crunogena* by examining KEGG maps. The monomer from which peptidoglycan is polymerized is a NAG-NAM dimer with a tetrapeptide linker covalently bonded to NAM (Izaki, 1968) The number of moles of this monomer per gram dry weight of cell was calculated by dividing the grams of peptidoglycan per gram dry weight of cellular material (Table 1) by the molecular weight of the monomer (~1030 g/mol). The number of moles of the monomer is equal to the number of moles of NAG, NAM, and tetrapeptide linker. The amount of ATP necessary to synthesize sufficient monomer per gram dry weight of cell was calculated by taking the sum of the ATP necessary to synthesize the NAG, NAM, and tetrapeptide linker.

Table 2 | Macromolecular composition in grams per dry weight of cells

Cellular composition [a]	Grams/ gram dry cell weight
Protein	0.55
DNA	0.03
r-RNA	0.16
m-RNA	0.01
t-RNA	0.03
fatty acid	0.06
polar lipids	0.09
Lipopolysaccharides	0.03
Peptidoglycan	0.03
Sum	1.00

[a] From *E. coli* hydrolysate values (Neidhart et al.. 1990)

Chapter 3: Results and Discussion

ATP to build central carbon intermediates from CO_2

The moles of ATP required to build one mole of central carbon intermediates from CO_2 were determined from KEGG pathway maps (Table 3, Figure 4). The ATP requirements were calculated from metabolic pathways that are possible in each species based upon the presence of the genes encoding all of the enzymes present in the pathways. The CBB cycle requires more ATP than the rCAC to produce all intermediates (Table 3, Figure 4). For example, the CBB cycle requires 27.3 more ATP to produce fumarate than the rCAC. Organisms growing via the CBB cycle generate fumarate through glycolysis and the oxidative citric acid cycle. In contrast, fumarate is a direct product of the rCAC.

The ATP values in Table 2 appear higher than previously published ATP costs (Fuchs 2011, Bar-Even et al. 2012) because they include the ATP equivalent cost of cellular electron carriers. For example, in (Fuchs 2011) it was calculated that 5 NADPH and 7 ATP are required to produce pyruvate via the CBB cycle. When the NADPH is converted to its ATP equivalent (4 ATP/NADPH) the total ATP value comes out to 27 ATP to make pyruvate via the CBB cycle (Table 3, Figure 4). Previously published values (Fuchs 2011, Bar-Even et al. 2012) state the rCAC requires 2 ATP, 2 NADPH, 1 quinol and 2 Fd_{red} pairs to produce a molecule of pyruvate. When the values are converted to their ATP equivalents, this equates to (8 ATP + 2 NADPH (8ATP) + 1 quinol (2.66 ATP) + 4 Fd_{red} (8ATP) for a total of 20.66 ATP.

The fractional values for the rCAC (Table 3) result from the fractional ATP equivalent of quinol (2.66 ATP/quinol) and the need to balance the number of carbons during synthesis

calculations. For example, four molecules of the three-carbon compound glyceraldehyde-3-phosphate are used to synthesize three molecules of the four-carbon compound erythrose-4-phosphate. In order to tally the amount of ATP to synthesize a single molecule of erythrose-4-phosphate, the amount of ATP needed to synthesize four molecules of glyceraldehyde-3-phosphate were divided by three. The same approach was used for ribose-5-phosphate and sedoheptulose-7-phosphate.

Table 3 | Moles ATP required to make one mole of central carbon intermediates from CO_2.

Intermediates	CBB ATP	rCAC ATP	(CBB-rCAC)
Fructose 6-P	66	55.32	10.68
Sedoheptulose 7-P	77	64.54	12.46
Pyruvate	27	20.66	6.34
Oxaloacetate	27	22.66	4.34
2-ketoglutarate	51	34.32	16.68
3-phosphoglycerate	28	22.66	5.34
Acetyl-Coa	24	16.66	7.34
Fumarate	54	26.66	27.34
PEP	28	22.66	5.34
Erythrose-4-P	44	36.88	7.12
Ribose-5-P	55	46.1	8.9
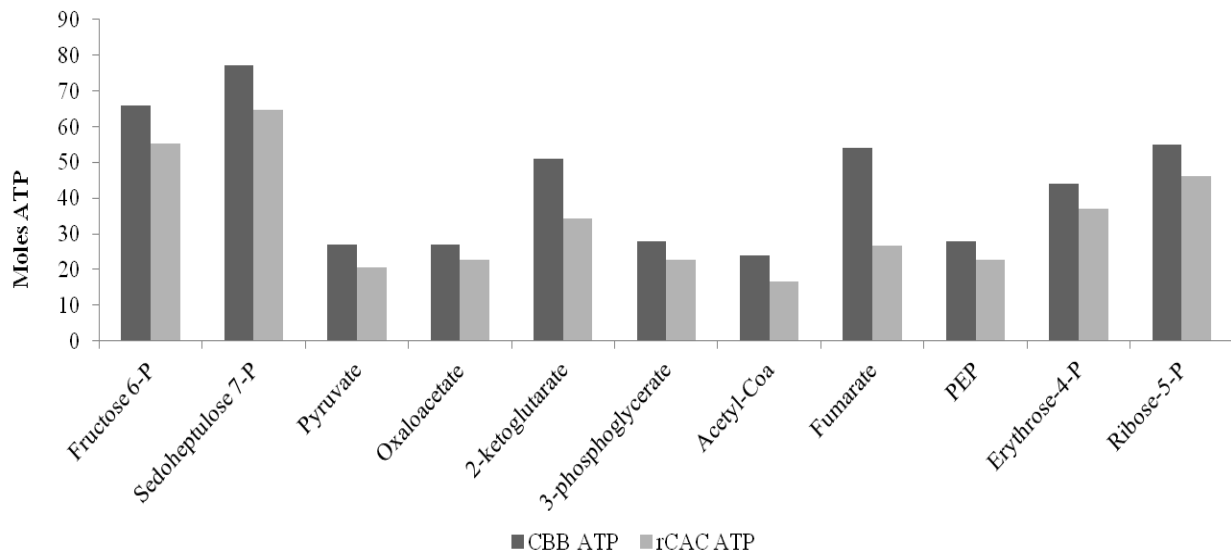
**Figure 4** | ATP required to make one mole of central carbon intermediates from CO_2

Building blocks from central carbon intermediates

The moles of intermediates to synthesize building blocks for macromolecules (i.e amino acids to build proteins, nucleotides to build nucleic acids etc…) were tallied for all three species. The production of amino acids required large amounts of pyruvate, oxaloacetate and cellular energy carriers in the form of ATP, and NAD(P)H (Table 4, Figure 5). Pyruvate provides a basic carbon skeleton for the formation of the amino acids alanine, valine and leucine and is also required for the production of lysine and isoleucine. Oxaloacetate provides a carbon skeleton for the amino acids glutamine, aspartate, asparagine, arginine, threonine, methionine and isoleucine. Although oxaloacetate is used for the production of a wider variety of amino acids, the amount of pyruvate required for the amount of amino acids in a gram of cell biomass is higher. This can be attributed to the high percentage of alanine in the hydrolyzed cell extracts used for the biomass composition values for this study (Table 1) (Roberts et al. 1957, Okayasu et al. 1997, Lillich and Elkan 1973). As expected, large amounts of cellular reductant were required to synthesize amino acids from central carbon intermediates (Table 4, Figure 5). The values only reflect the amount of ATP to build each amino acid and do not take into account the amount of energy required to generate peptide bonds. The negative values for CO_2, NADH and fumarate indicate that they are produced during the synthesis of certain amino acids. Carbon dioxide is a byproduct in the synthesis of several amino acids that have carbon skeletons provided by pyruvate, phosphoenolpyruvate or erythrose-4-phosphate. NADH is produced during the synthesis of amino acids based on 3-phosphoglycerate such as cysteine and glycine. Fumarate is a byproduct of arginine synthesis and is generated when citruline and aspartate form L-arginosuccinate which is broken into fumarate and arginine by the enzyme arginosuccinatelyase (KEGG).

19

Table 4 | Moles of intermediates to build amino acids for proteins in one gram of cell biomass

Intermediates[a]	*T. crunogena*	*S. autotrophica*	*E. coli*
Pyruvate	2.30E-03	2.30E-03	2.31E-03
Oxaloacetate	1.48E-03	1.48E-03	1.51E-03
2-ketoglutarate	8.12E-04	8.12E-04	8.59E-04
3-phosphoglycerate	7.06E-04	7.06E-04	5.87E-04
Acetyl-Coa	6.29E-04	6.29E-04	6.15E-04
Fumarate	-1.72E-04	-1.72E-04	-2.06E-04
PEP	6.07E-04	6.07E-04	6.03E-04
Erythrose-4-P	3.03E-04	3.03E-04	3.01E-04
Ribose-5-P	6.60E-05	6.60E-05	6.22E-05
CO_2	-1.80E-03	-1.80E-03	-1.47E-03
ATP	8.24E-03	8.02E-03	8.27E-03
NADH	-4.94E-04	-1.10E-03	-5.15E-04
NADPH	8.27E-03	8.76E-03	8.48E-03

[a]Intermediate values to synthesize amino acids based on genomic data from KEGG and cell composition literature.
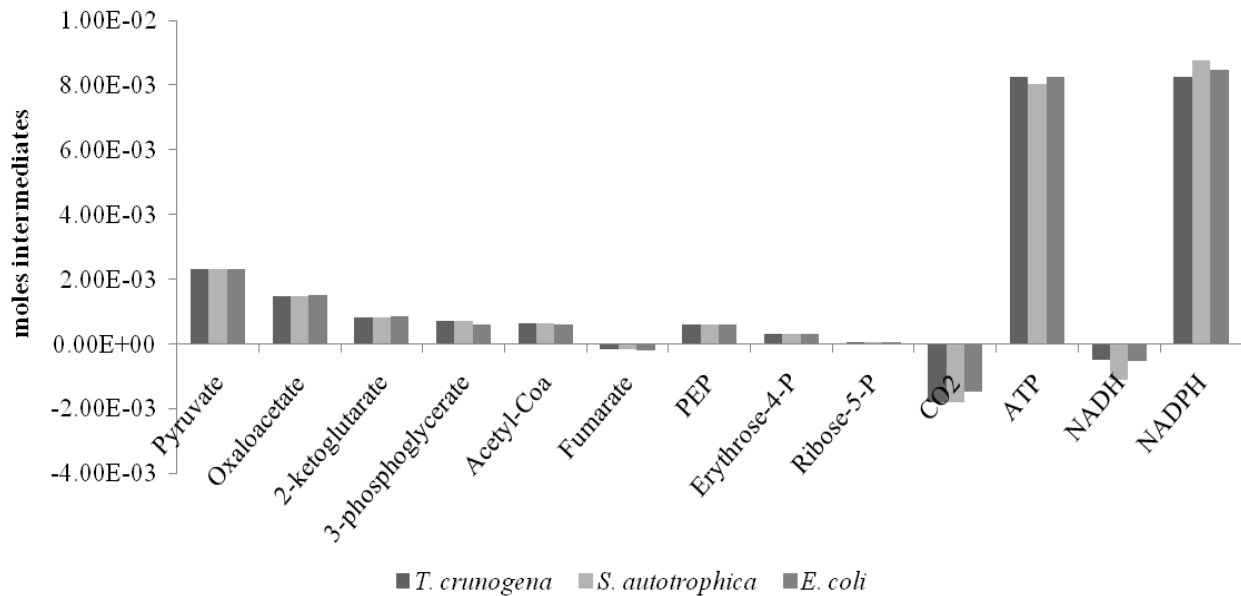
Figure 5 | Moles of intermediates to build amino acids for proteins in one gram of cell biomass.
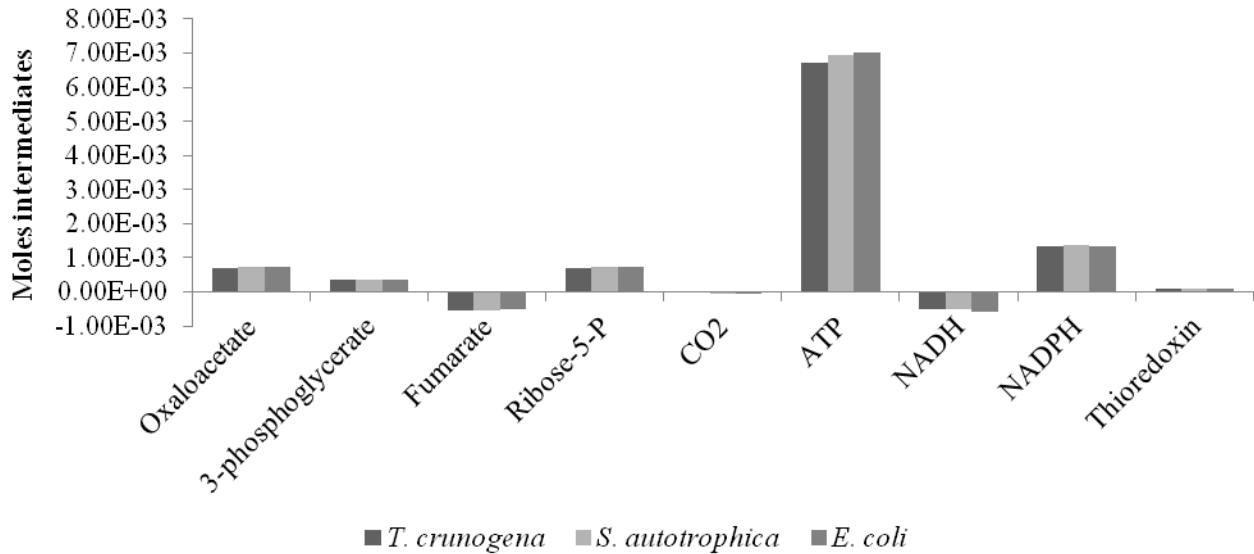
The moles of intermediates to build nucleotides for DNA and the three forms of RNA (mRNA, rRNA and tRNA) in a gram of cell biomass were tallied for all three cell types (Table 5, Figure 6). Oxaloacetate, 3-phosphoglycerate and ribose-5-P were most utilized intermediates for

nucleotide synthesis. Slight variations in these values for the three cell types are mainly due to

differences in nucleotide composition and not variations in biosynthetic pathways. Oxaloacetate

provides the carbon skeleton for aspartate which is a key component for purine and pyrimidine

rings. For the synthesis of purines, aspartate provides a nitrogen atom for the larger of the two

rings, while the carbon skeleton is removed as fumarate. For pyrimdine synthesis, aspartate

provides three of the four carbons in the single ring as well one of the two nitrogen atoms (Yates,

1956). Three-phosphoglycerate is incorporated into purine biosynthesis by providing the carbon

skeleton for glycine. Ribose-5-p is incorporated into all purines and pyrimidines by providing the

carbon skeleton for 5-phosphoribosyl-pyrophosphate (PRPP). Large amounts of ATP and

NAD(P)H are required for nucleotide synthesis while NADH is produced as a byproduct of the

orotate synthesis for pyrimidines. Reduced thioredoxin provides the hydride ion needed to

generate deoxynucleoside diphosphates from nucleosides.

Table 5 | Moles of intermediates to build nucleic acids for one gram of cell biomass

Intermediates[a]	T. crunogena	S. autotrophica	E. coli
Oxaloacetate	6.88E-04	7.25E-04	7.28E-04
3-phosphoglycerate	3.49E-04	3.44E-04	3.47E-04
Fumarate	-5.35E-04	-5.45E-04	-5.06E-04
Ribose-5-P	6.88E-04	7.25E-04	7.28E-04
CO_2	9.65E-06	-3.74E-05	-3.35E-05
ATP	6.72E-03	6.93E-03	7.02E-03
NADH	-5.02E-04	-5.23E-04	-5.70E-04
NADPH	1.32E-03	1.36E-03	1.33E-03
Thioredoxin	9.63E-05	9.18E-05	9.53E-05

[a] Nucleotide content based on chromosome data for each organism from IMG.

Figure 6 | Moles of intermediates to build nucleic acids for one gram of cell biomass

The moles of intermediates required to build lipids for a gram of cell biomass was tallied for all three cell types (Table 6, Figure 7). Fatty acid synthesis required the largest amounts of acetyl-CoA, ATP and NADPH. Fatty acids are synthesized via the sequential condensation and reduction of acetyl-CoA molecules. In order to continue this process, ATP is needed to activate the acetyl-CoA. Acetyl-CoA and malyonyl CoA combine to form 3-ketoacyl-SACP which reduced by NADPH. The requirement of 3-phosphoglycerate is tied to the production of serine and dihydroxyacetone phosphate (DHAP), which is used as the carbon skeleton for the construction of other polar lipids.

Lipopolysaccharides (Table 7, Figure 8), from which the outer leaflet of gram negative bacteria is comprised, are synthesized using the same pathways for all three cell types. The large requirement for acetyl-CoA, ATP and NADPH is due to the synthesis of tetradecanoic acid (3-OH-14:0) and dodecanoic acid (12:0), fatty acids found within the lipid A molecule of

the lipoloysaccharide. Fructose-6-phosphate inputs are directed into the synthesis of UDP-n-acetylglucosamine 6-phosphate, and UDP glucose and galactose. Sedoheptulose-7-phosphate is used in the synthesis of ADP-glyceromannoheptose while ribose-5-phosphate and phosphoenolpyruvate are used for the synthesis of CPM-3-deoxy-*manno*-octulosonate (CMP-KDO).

Table 6| Moles of intermediates to build phospholipids and other polar lipids for one gram of cell biomass**.**

Intermediates	*T. crunogena*	*S. autotrophica*	*E. coli*
Fructose 6-P	4.54E-06	5.05E-06	5.24E-05
3-phosphoglycerate	2.26E-04	2.35E-04	2.43E-04
Acetyl-Coa	1.19E-03	1.18E-03	1.14E-03
ATP	1.31E-03	1.28E-03	1.30E-03
NADPH	2.15E-03	2.14E-03	2.08E-03
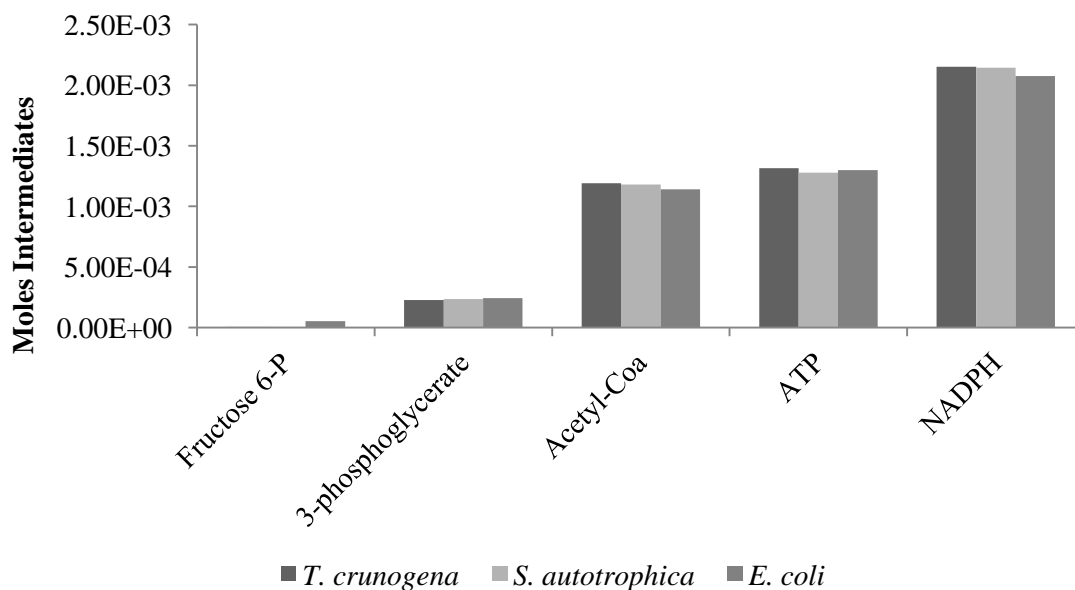
Figure 7| Moles of intermediates to build phospholipids and other polar lipids for one gram of cell biomass**.**

Table 7 | Moles of intermediates to build lipopolysaccharides for one gram of cell biomass.

Intermediates	*T. crunogena*	*S. autotrophica*	*E. coli*
Fructose 6-P	5.24E-05	5.24E-05	5.24E-05
Sedoheptulose 7-P	2.24E-05	2.24E-05	2.24E-05
3-phosphoglycerate	7.48E-06	7.48E-06	7.48E-06
Acetyl-Coa	3.22E-04	3.22E-04	3.22E-04
PEP	2.24E-05	2.24E-05	2.24E-05
Ribose-5-P	2.24E-05	2.24E-05	2.24E-05
ATP	4.94E-04	4.94E-04	4.94E-04
NADPH	4.79E-04	4.79E-04	4.79E-04
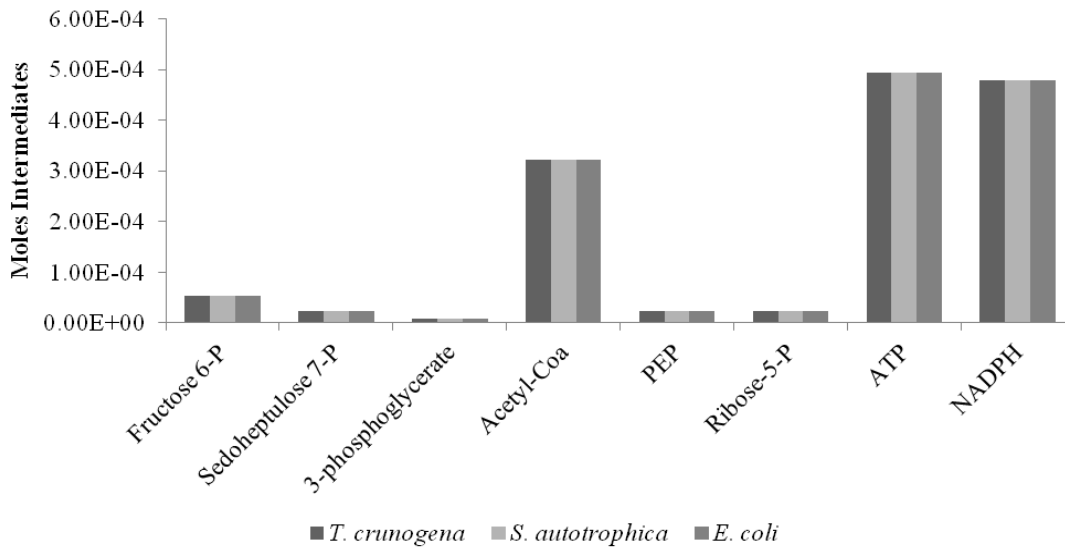
Figure 8 | Moles of intermediates to build lipopolysaccharides for one gram of cell biomass.

The moles of intermediates to build peptidoglycan for all three cell types were based on data available for *E. coli* (Table 8, Figure 9). The peptidoglycan layer is constructed of alternating molecules of N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) which are connected by tetrapeptide cross links. Fructose-6-phosphate provides the six carbon sugar for UDP-NAG and UDP-NAM while acetyl-CoA provides an acetyl group. The synthesis of NAM from NAG requires phosphoenolpyruvate which donates a lactyl unit to form a NAG-3-

enolpyruvalether intermediate which is further reduced by NAD(P)H . Pyruvate requirements

stem from the production of D-alanine which is used in creating peptide cross links in the

bacterial cell wall. High ATP costs result from the synthesis of NAG and NAM and for the

formation of tetrapeptide linkers.

Table 8 | Moles of intermediates to build peptidoglycan for one gram of cell biomass**.**

Intermediates	*T. crunogena, S. autotrophica, E.coli*
Fructose 6-P	4.86E-05
Pyruvate	7.28E-05
Oxaloacetate	2.43E-05
2-ketoglutarate	2.43E-05
Acetyl-Coa	4.86E-05
PEP	2.43E-05
ATP	3.16E-04
NADPH	1.70E-04
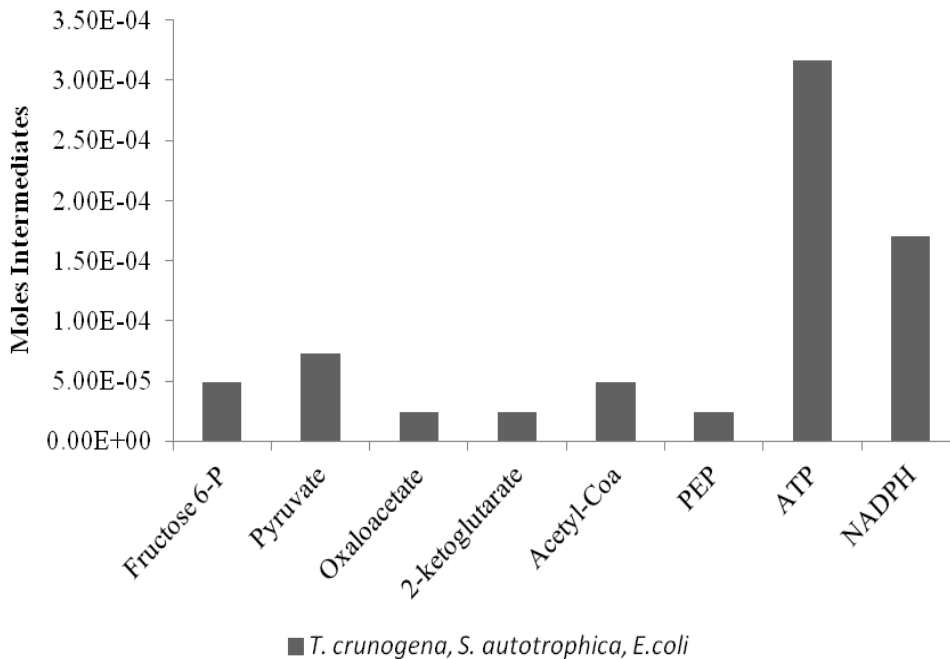
Figure 9| Moles of intermediates to build peptidoglycan for one gram of cell biomass**.**

The amounts of metabolic intermediates needed to make all macromolecules were summed (Tables 4-8) and multiplied by the moles of ATP needed to synthesize each intermediate from CO_2 (Table 3) in order to estimate the amount of ATP necessary to make enough of each metabolic intermediate for one gram of cell biomass (Figure 10). The CBB cycle requires more ATP to build all the intermediates required for a gram of cell biomass. However, the amount of ATP and NADPH required to turn these intermediates into macromolecules slightly mitigates their overall difference in producing biomass. While both cycles require more ATP to build larger intermediates such as sedoheptulose-7-phosphate, these large intermediates are not required in high amounts for biomass synthesis. Therefore large intermediates do not have high ATP costs per gram of biomass of cell even though they are more expensive to make than smaller intermediates such as pyruvate.

As noted in Table 3, fumarate production from CO_2 requires twice as much ATP using the CBB cycle instead of the rCAC. The negative ATP values for fumarate are indicative of its production during arginine biosynthesis (Cunin et al. 1986) and therefore net ATP is not consumed for fumarate synthesis when building cell biomass. In other words, the negative value for the amount of ATP required to build enough fumarate for a gram of cell biomass results from multiplying the negative intermediate input value by the positive ATP requirement to synthesize fumarate from CO_2. While the actual amount of fumarate that is produced is quite small, the ATP values are amplified due to the expense of building fumarate from CO_2.

Table 9 | Moles ATP to build intermediates for one gram of cell biomass for three species.

Intermediates	*T. cru* CBB	*S. auto* CBB	*E. coli* CBB	*T. cru* rCAC	*S. auto* rCAC	*E. coli* rCAC
Acetyl CoA	5.26E-02	6.06E-02	5.11E-02	3.65E-02	4.21E-02	3.55E-02
Erythrose-4-p	1.33E-02	1.33E-02	1.33E-02	1.12E-02	1.12E-02	1.11E-02
Fructose-6-p	6.96E-03	7.00E-03	6.66E-03	5.83E-03	5.87E-03	5.58E-03
Fumarate	-3.82E-02	-3.87E-02	-3.84E-02	-1.88E-02	-1.91E-02	-1.90E-02
2-ketoglutarate	4.26E-02	4.26E-02	4.51E-02	2.87E-02	2.87E-02	3.03E-02
Oxaloacetate	5.93E-02	6.03E-02	6.12E-02	4.98E-02	5.06E-02	5.14E-02
Phosphenolpyruvate	1.77E-02	1.83E-02	1.82E-02	1.43E-02	1.48E-02	1.47E-02
3-phosphoglycerate	3.61E-02	2.67E-02	3.32E-02	2.92E-02	2.16E-02	2.68E-02
Pyruvate	6.40E-02	6.40E-02	6.44E-02	4.90E-02	4.90E-02	4.93E-02
Ribose-5-p	4.27E-02	4.47E-02	4.47E-02	3.58E-02	3.75E-02	3.75E-02
Sedoheptulose-7-p	1.73E-03	1.73E-03	1.73E-03	1.45E-03	1.45E-03	1.45E-03
ATP	1.72E-02	1.68E-02	1.74E-02	1.72E-02	1.68E-02	1.74E-02
NADH	-2.99E-03	-4.85E-03	-3.25E-03	-2.99E-03	-4.85E-03	-3.25E-03
NADPH	3.72E-02	3.81E-02	3.76E-02	3.72E-02	3.81E-02	3.76E-02
Thioredoxin	2.89E-04	2.67E-04	2.86E-04	2.89E-04	2.67E-04	2.86E-04
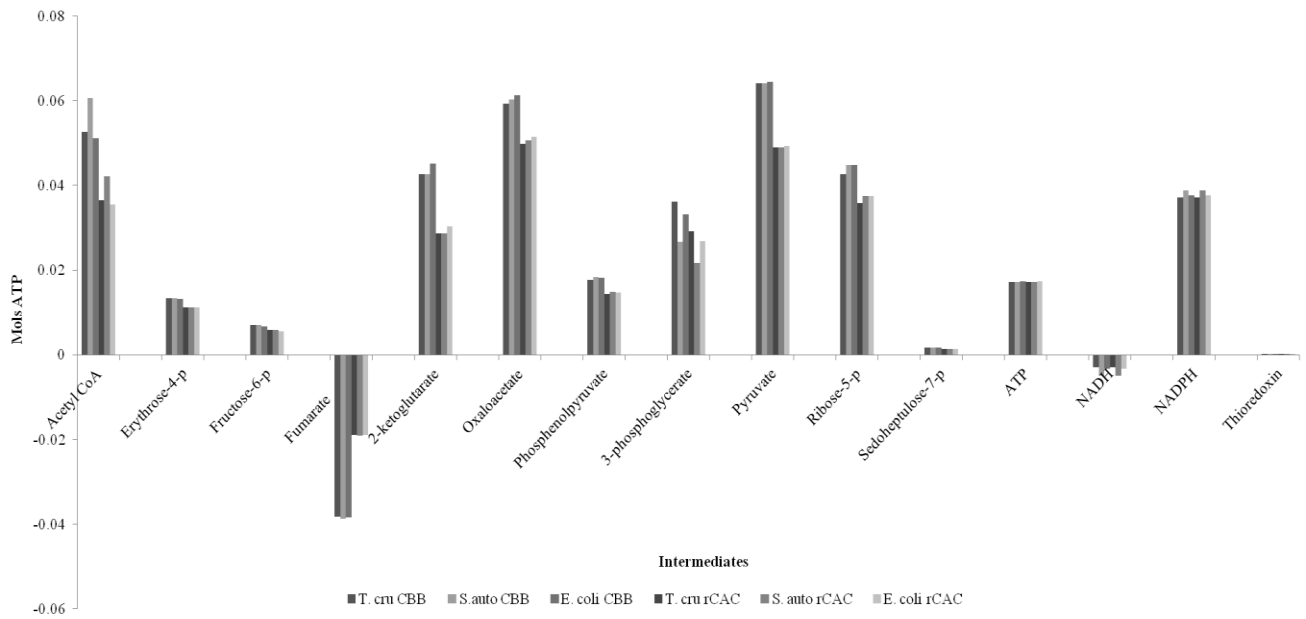
Figure 10| Moles ATP to build intermediates for one gram of cell biomass for three species. The total moles ATP to build metabolic intermediates in one gram of cell biomass were tallied for each cell type using both the CBB cycle and the rCAC. Intermediates built using the CBB cycle are clustered in columns to the left of each label and those built using the rCAC are clustered to the right.

Table 10 | The total molar amount of ATP to build one gram of cell biomass. Values include the cost to polymerize amino acids into proteins. The native carbon fixation pathways for *T. crunogena* and *S. autotrophica* are marked in bold.

Organism	CBB Cycle	rCAC
T. crunogena	**0.36**	0.31
S. autotrophica	0.36	**0.31**
E. coli	0.37	0.31

The total molar amount of ATP to build one gram cell biomass was tallied for each organism using both carbon fixation pathways (Table 10). The CBB cycle requires about 0.05 mols of ATP more than the rCAC to synthesize one gram of cell biomass from CO_2 when cells are compared using their native carbon fixation pathways (i.e. *T.crunogena* → CBB and *S. autotrophica* → rCAC). This difference is retained even with varying biomass composition. This study supports previously published literature which suggests that the CBB cycle is more expensive than the rCAC. However, the differences between the two cycles are mitigated by process downstream from intermediate synthesis such as the cost to turn the intermediates into building blocks and macromolecules. For the production of pyruvate using the rCAC, the ATP costs are approximately 77% of the CBB (ATP needed to synthesize 1 mol pyruvate from rCAC ÷ ATP needed to synthesize 1 mol pyruvate from CBB = 20.66÷27=0.77). For the production of biomass using the rCAC, the ATP costs are approximately 84% of the CBB (ATP needed to synthesize 1g biomass from rCAC ÷ ATP needed to synthesize 1g biomass from CBB = 31 ÷ 37 = 0.84). Although the CBB cycle requires more ATP to synthesize biomass than the rCAC, the predominance of both pathways in hydrothermal vent communities suggests that CBB cycle must confer some selective advantage to the microbes utilizing it. Oxygen tolerance, enzyme

kinetics, structural costs, phylogenetic affiliation or other factors may be the driving forces behind the distribution and use of these pathways.

In some cases, a cell may adapt a less than ideal carbon fixation pathway for its environment simply because it has inherited the pathway from ancestors that were adapted to different environmental conditions. For example, species that have been found with variations of the rCAC with relatively oxygen tolerant enzymes may have transitioned to a more oxygen rich environment and through selection their descendants evolved more oxygen tolerant pathways. These species would have to maintain a high level of metabolic efficiency in order to remain competitive in the new environment. Alternatively, if the rCAC is not prevalent in a lineage, it may be that an aerobic organism acquired rCAC genes from an rCAC autotroph while incorporating oxygen tolerant enzymes of its own.

Enzyme kinetics also contributes to the expensive of a particular pathway. Enzymes with a low maximal velocity (V_{max}) must be produced in higher quantities by the cell in order to process large amounts of substrate. A cell using a pathway with low V_{max} enzymes may have to synthesize more enzymes in order to metabolize the same amount of substrate as a pathway with higher V_{max} enzymes. Indeed, Rubisco enzymes have very low V_{max} values compared to other enzymes (Tabita, F.R. 1999), which may boost the expense of this pathway.

Increased enzyme production is not the only structural cost that must be taken into consideration when comparing the efficiency of a carbon fixation pathway. In the case of the CBB cycle, autotrophs must take precautions to avoid a costly oxygenase reaction by Rubisco (Tabita, F.R. et al. 2008). Under low DIC conditions, *T. crunogena* will synthesize protein microcompartments called carboxysomes which isolate Rubisco in order to maximize its ability

to function as a caroboxylase (Dobrinski et al. 2005). When grown under low CO_2 conditions, *T. crunogena* and many other autotrophic bacteria that use the CBB cycle form multiple carboxysomes per cell; the added cost of forming these subcellular structures is large, and therefore is carefully regulated (Yeates et al. 2008). Given this added cost, the CBB cycle does not seem like an advantage; however, it appears to be the only one that functions well at low CO_2 concentrations (Fuchs 2011).

The CBB cycle and the rCAC are both "patchwork" pathways in that they are composed primarily of other central carbon pathways. They become carbon fixation pathways by the presence of two or three enzymes that catalyze irreversible steps to drive reactions in the reductive direction. For example, the CBB cycle is a conglomeration of the pentose phosphate pathway and the glycolytic pathway and is functional as carbon fixation pathway only with the addition of Rubisco and the irreversible phosphatase, phosphoribulokinase. It can be surmised that the addition of a few key enzymes via horizontal gene transfer may allow a heterotrophic species to gain autotrophic abilities. Indeed, genes encoding key enzymes for autotrophic pathways have been found to be widely transferred by horizontal gene transfer. Most cyanoobacteria have form 1B Rubisco while some marine cyanobacteria have form 1A Rubisco which is believed to have been transferred to the cyanobacteria (Rae et al. 2011). Some dinoflagellates also have form II Rubisco which appears to have been horizontally transferred from the proteobacteria. (Bachvaroff et al. 2004) Thus, the acquisition of autotrophic metabolism via insertion of a key carboxylase or other key enzyme is quite feasible.

Despite its apparent expense with respect to metabolite synthesis, the CBB cycle may have an advantage based on its reliance on NADPH instead of ferredoxin as its main electron

carrier. Some enzymes of the rCAC require ferredoxin (-400 mV) as a substrate (pyruvate synthase, alphaketoglutarate synthase). Cognate enzymes from CBB cells require NADPH as a substrate (-320 mV; e.g., glyceraldehyde 3-phospate dehydrogenase). Many organisms using the rCAC utilize hydrogen as their electron donor (~-400 mV standard reduction potential at pH 7); transfer of electrons from hydrogen to ferredoxin (~-400 mV standard reduction potential) is facile. In contrast, cells using other electron donors, such as hydrogen sulfide (~-200 mV), would find reduction of ferredoxin to be an energetically expensive proposition; reduction of NADP (~320 mV) is still costly, but less so. (Buckel and Thauer 2013) Use of an autotrophic pathway (CBB) which consumes electrons from less 'costly' cellular electron carriers (e.g., NADPH) may be an advantage when environmental electron donors with less negative reduction potentials (e.g., H_2S) predominate.

References

Bachvaroff, TR , Concepcion, GT , Rogers, CR, Herman, EM, Delwiche, CF. 2004. Dinoflagellate Expressed Sequence Tag Data Indicate Massive Transfer To The Nuclear Genome Sequence Tag Data Of Chloroplast Genes. Protist **155**: 65-78

Bar-Even A, Noor E, Milo R. 2012. A Survey of Carbon Fixation Pathways Through A Quantitative Lens. Journal of Experimental Botany **63:**2325-2342.

Bassham JA, Benson AA, Kay LD, Harris AZ, Wilson AT, Calvin M. 1954. The Path of Carbon in Photosynthesis .21. The Cyclic Regeneration of Carbon Dioxide Acceptor. Journal of the American Chemical Society **76:**1760-1770.

Beh M, Strauss G, Huber R, Stetter KO, Fuchs G. 1993. Enzymes of the Reductive Citric-Acid Cycle In The Autotrophic Eubacterium *Aquifex-Pyrophilus* And In The Archaebacterium *Thermoproteus-Neutrophilus*. Archives of Microbiology **160:**306-311.

Berg IA. 2011. Ecological Aspects of the Distribution of Different Autotrophic CO2 Fixation Pathways. Applied and Environmental Microbiology **77:**1925-1936.

Berg IA, Kockelkorn D, Buckel W, Fuchs G. 2007. A 3-hydroxypropionate/4-hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. Science **318:**1782-1786.

Biebl H, Pfennig N. 1978. Growth Yields of Green Sulfur Bacteria In Mixed Cultures With Sulfur And Sulfate Reducing Bacteria. Archives of Microbiology **117:**9-16.

Braakman R, Smith E. 2012. The Emergence and Early Evolution of Biological Carbon-Fixation. Plos Computational Biology **8:**16.

Buchanan BB, Arnon DI. 1990. A Reverse Krebs Cycle in Photosynthesis - Consensus at Last. Photosynthesis Research **24:**47-53.

Buckel W, Thauer RK. 2013. Energy Conservation via Electron Bifurcating Ferredoxin Reduction And Proton/Na+ Translocating Ferredoxin Oxidation. Biochimica Et Biophysica Acta-Bioenergetics **1827:**94-113.

Caldwell RB, Lattemann CT. 2004. Simple and Reliable Method to Precipitate Proteins From Bacterial Culture Supernatant. Applied and Environmental Microbiology **70:**610-612.

Campbell BJ, Cary SC. 2004. Abundance of Reverse Tricarboxylic Acid Cycle Genes in Free-Living Microorganisms At Deep-Sea Hydrothermal Vents. Applied and Environmental Microbiology **70:**6282-6289.

Cannon GC, Shively JM. 1983. Characterization of a Homogenous Preparation of Carboxysomes from Thiobacillus-Neapolitanus. Archives of Microbiology **134:**52-59.

Caroff M, Karibian D. 2003. Structure of Bacterial Lipopolysaccharides. Carbohydrate Research **338:**2431-2447.

Conway N, Capuzzo JM. 1991. Incorporation and Utilization of Bacterial Lipids in the Solemya-Velum Symbiosis. Marine Biology **108:**277-291.

Cunin R, Glansdorff N, Pierard A, Stalon V. 1986. Biosynthesis and Metabolism of Arginine in Bacteria. Microbiological Reviews **50:**314-352.

Dennis PP, Bremer H. 1974. Macromolecular-Composition During Steady-State Growth of *Escherichia-Coli* B-R. Journal of Bacteriology **119:**270-281.

Dobrinski KP, Boller AJ, Scott KM. 2010. Expression and Function of Four Carbonic Anhydrase Homologs in the Deep-Sea Chemolithoautotroph *Thiomicrospira crunogena*. Applied and Environmental Microbiology **76:**3561-3567.

Dobrinski KP, Enkemann SA, Yoder SJ, Haller E, Scott KM. 2012. Transcriptional Response of the Sulfur Chemolithoautotroph *Thiomicrospira crunogena* to Dissolved Inorganic Carbon Limitation. Journal of Bacteriology **194:**2074-2081.

Dobrinski KP, Longo DL, Scott KM. 2005. The Carbon-Concentrating Mechanism Of The Hydrothermal Vent Chemolithoautotroph *Thiomicrospira crunogena*. Journal of Bacteriology **187:**5761-5766.

Dobrinski KP, Scott KM. 2006. Genes Encoding the Carbon Concentrating Mechanism of the Hydrothermal Vent Chemolithoautotroph *Thiomicrospira crunogena*. Abstracts of the General Meeting of the American Society for Microbiology **106:**402.

Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA, Fraser CM. 2002. The Complete Genome Sequence of *Chlorobium Tepidum* TLS, a Photosynthetic, Anaerobic, Green-Sulfur Bacterium. Proceedings of the National Academy of Sciences of the United States of America **99:**9509-9514.

Evans MCW, Buchanan BB, Arnon DI. 1966. A New Ferredoxin-Dependent Carbon Reduction Cycle in a Photosynthetic Bacterium. Proceedings of the National Academy of Sciences of the United States of America **55:**928-&.

Fuchs G. 2011. Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life?, p. 631-+. *In* Gottesman S, Harwood CS (ed.), Annual Review of Microbiology, Vol 65, vol. 65. Annual Reviews, Palo Alto.

Fisher, CR, Takai, K, Le Bris, N. 2007. Hydrothermal Vent Ecosystems. Oceanography, **20**:14-23

Ghosh W, Dam B. 2009. Biochemistry and Molecular Biology Of Lithotrophic Sulfur Oxidation By Taxonomically And Ecologically Diverse Bacteria And Archaea. Fems Microbiology Reviews **33:**999-1043.

Glaubitz S, Lueders T, Abraham WR, Jost G, Jurgens K, Labrenz M. 2009. C-13-Isotope Analyses Reveal That Chemolithoautotrophic Gamma- And Epsilonproteobacteria Feed a Microbial Food Web in a Pelagic Redoxcline of the Central Baltic Sea. Environmental Microbiology **11:**326-337.

Goffredi SK, Childress JJ, Desaulniers NT, Lee RW, Lallier FH, Hammond D. 1997. Inorganic Carbon Acquisition by the Hydrothermal Vent Tubeworm *Riftia Pachyptila* Depends Upon High External P-CO2 And Upon Proton-Equivalent Ion Transport by the Worm. Journal of Experimental Biology **200:**883-896.

He PQ, Liu Y, Yue WJ, Huang XH. 2013. Key Genes Expression of Reductive Tricarboxylic Acid Cycle from Deep-Sea Hydrothermal Chemolithoautotrophic *Caminibacter Profundus* In Response To Salinity, Ph And O-2. Acta Oceanologica Sinica **32:**35-41.

Herter S, Fuchs G, Bacher A, Eisenreich W. 2002. A Bicyclic Autotrophic CO_2 Fixation Pathway in *Chloroflexus Aurantiacus*. Journal of Biological Chemistry **277:**20277-20283.

Huber H, Gallenberger M, Jahn U, Eylert E, Berg IA, Kockelkorn D, Eisenreich W, Fuchs G. 2008. A Dicarboxylate/4-Hydroxybutyrate Autotrophic Carbon Assimilation Cycle in the Hyperthermophilic Archaeum *Ignicoccus Hospitalis*. Proceedings of the National Academy of Sciences of the United States of America **105:**7851-7856.

Hugler M, Sievert SM. 2011. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean, p. 261-289. *In* Carlson CA, Giovannoni SJ (ed.), Annual Review of Marine Science, Vol 3, vol. 3. Annual Reviews, Palo Alto.

Hugler M, Wirsen CO, Fuchs G, Taylor CD, Sievert SM. 2005. Evidence for Autotrophic CO_2 Fixation via the Reductive Tricarboxylic Acid Cycle by Members of the Epsilon Subdivision of Proteobacteria. Journal of Bacteriology **187:**3020-3027.

Inagaki F, Takai K, Hideki KI, Nealson KH, Horikishi K. 2003. *Sulfurimonas Autotrophica* Gen. Nov., Sp Nov., A Novel Sulfur-Oxidizing Epsilon-Proteobacterium Isolated From Hydrothermal Sediments in the Mid-Okinawa Trough. International Journal of Systematic and Evolutionary Microbiology **53:**1801-1805.

Izaki K, Matsuhas.M, Stroming.Jl. 1968. Biosynthesis of Peptidoglycan of Bacterial Cell Walls .13. Peptidoglycan Transpeptidase and D-Alanine Carboxypeptidase - Penicillin-Sensitive Enzymatic Reaction in Strains of *Escherichia Coli*. Journal of Biological Chemistry **243:**3180-&.

Jannasch HW, Wirsen CO, Nelson DC, Robertson LA. 1985. *Thiomicrospira-Crunogena* Sp-Nov, A Colorless, Sulfur-Oxidizing Bacterium From A Deep-Sea Hydrothermal Vent. International Journal of Systematic Bacteriology **35:**422-424.

Javor BJ, Wilmot DB, Vetter RD. 1990. Ph-Dependent Metabolism of Thiosulfate and Sulfur Globules in the Chemolithotrophic Marine Bacterium *Thiomicrospira-Crunogena*. Archives of Microbiology **154:**231-238.

Johnson KS, Childress JJ, Beehler CL. 1988. Short-Term Temperature Variability in the Rose Garden Hydrothermal Vent Field - An Unstable Deep-Sea Environment. Deep-Sea Research Part a-Oceanographic Research Papers **35:**1711-&.

Larsen RA, Wilson MM, Guss AM, Metcalf WW. 2002. Genetic Analysis Of Pigment Biosynthesis in *Xanthobacter Autotrophicus* Py2 Using A New, Highly Efficient Transposon Mutagenesis System That Is Functional In A Wide Variety Of Bacteria. Archives of Microbiology **178:**193-201.

Lillich TT, Elkan GH. 1973. Analysis Of Intracellular Amino-Acid Pool And Proteins From Whole Cells of *Rhizobium-Japonicum*. Journal of Applied Bacteriology **36:**315-319.

Lovering AL, de Castro LH, Lim D, Strynadka NCJ. 2007. Structural Insight Into The Transglycosylation Step Of Bacterial Cell-Wall Biosynthesis. Science **315:**1402-1405.

Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hugler M, Albrecht D, Robidart J, Bench S, Feldman RA, Hecker M, Schweder T. 2007. Physiological Proteomics of the Uncultured Endosymbiont of *Riftia Pachyptila*. Science **315:**247-250

McCollom TM, Amend JP. 2005. A Thermodynamic Assessment of Energy Requirements for Biomass Synthesis by Chemolithoautotrophic Micro-Organisms in Oxic and Anoxic Environments. Geobiology **3:**135-144.

Mendz GL, Smith MA, Finel M, Korolik V. 2000. Characteristics of the Aerobic Respiratory Chains Of the Microaerophiles *Campylobacter Jejuni* And *Helicobacter Pylori*. Archives of Microbiology **174:**1-10.

Menning KJ, Dobrinski KP, Scott KM. 2009. Identifying the Components of the Carbon Concentrating Mechanism of Deep-sea Hydrothermal Vent Chemolithoautotrophic Bacterium *Thiomicrospira crunogena*. Abstracts of the General Meeting of the American Society for Microbiology **109**.

Metcalf WW, Jiang WH, Daniels LL, Kim SK, Haldimann A, Wanner BL. 1996. Conditionally Replicative and Conjugative Plasmids Carrying Lacz Alpha for Cloning, Mutagenesis, and Allele Replacement in Bacteria. Plasmid **35:**1-13.

Nakagawa S, Takai K. 2008. Deep-Sea Vent Chemoautotrophs: Diversity, Biochemistry and Ecological Significance. Fems Microbiology Ecology **65:**1-14.

Neidhart, F.C., Ingraham, J.L., Schaechter, M. 1990. Physiology of the Bacterial Cell. Sinauer Associates, Inc., Sunderland, Massachusetts, 01375, U.S.A.

Novelli GD. 1967. Amino Acid Activation for Protein Synthesis. Annual Review of Biochemistry **36:**449-&.

Okayasu T, Ikeda M, Akimoto K, Sorimachi K. 1997. The Amino Acid Composition of Mammalian and Bacterial Cells. Amino Acids **13:**379-391.

Olins HC, Rogers DR, Frank KL, Vidoudez C, Girguis PR. 2013. Assessing the Influence of Physical, Geochemical And Biological Factors on Anaerobic Microbial Primary Productivity Within Hydrothermal Vent Chimneys. Geobiology **11:**279-293.

Perner M, Gonnella G, Hourdez S, Bohnke S, Kurtz S, Girguis P. 2013. In Situ Chemistry and Microbial Community Compositions in Five Deep-Sea Hydrothermal Fluid Samples from Irina II In The Logatchev Field. Environmental Microbiology **15:**1551-1560.

Ragsdale SW, Pierce E. 2008. Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. Biochimica Et Biophysica Acta-Proteins and Proteomics **1784:**1873-1898.

Ratledge, C. 1989. Microbial Lipids, Volume 2. Academic Press. Univ. of Wisconsin-Madison, Wisonsin, U.S.A

Rae, BD. Forster, B, Badger, M.R. Price, G.D. 2011. The CO_2-concentrating mechanism of *Synechococcus* WH5701 is composed of native and horizontally-acquired components. Photosynthesis Research **109**: 59-72

Rauham V, Bloch DA, Wikstrom M. 2012. Mechanistic stoichiometry of proton translocation by cytochrome cbb_3. Proceedings of the National Academy of Sciences **109:** 7286-7291

Roberts RB, Cowie DB, Abelson PH, Bolton ET, Britten RJ. 1957. Studies of biosynthesis of *Escherichia coli.* 2nd printing, with addenda . Washington : Carnegie Institution

Schauder R, Widdel F, Fuchs G. 1987. Carbon Assimilation Pathways in Sulfate-Reducing Bacteria .2. Enzymes of A Reductive Citric-Acid Cycle In The Autotrophic *Desulfobacter-Hydrogenophilus*. Archives of Microbiology **148:**218-225.

Scott KM, Fisher CR, Vodenichar JS, Nix ER, Minnich E. 1994. Inorganic Carbon and Temperature Requirements for Autotrophic Carbon Fixation By The Chemoautotrophic Symbionts Of The Giant Hydrothermal Vent Tube Worm, *Riftia-Pachyptila.* Physiological Zoology **67:**617-638.

Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA, Boller AJ, Chain PSG, Clark JA, Davis CR, Detter C, Do KF, Dobrinski KP, Faza BI, Fitzpatrick KA, Freyermuth SK, Harmer TL, Hauser LJ, Hugler M, Kerfeld CA, Klotz MG, Kong WW, Land M, Lapidus A, Larimer FW, Longo DL, Lucas S, Malfatti SA, Massey SE, Martin DD, McCuddin Z, Meyer F, Moore JL, Ocampo LH, Paul JH, Paulsen IT, Reep DK, Ren QH, Ross RL, Sato PY, Thomas P, Tinkham LE, Zeruth GT. 2006. The Genome of Deep-Sea Vent Chemolithoautotroph *Thiomicrospira crunogena* XCL-2. Plos Biology **4:**2196-2212.

Shank, TM, Fornari, DJ, Von Damm, KL, Lilley, MD, Haymon, RM, Lutz, RA. 1998. Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9 '50N, East Pacific Rise). Deep-Sea Research II **45:** 465-515

Shiba H, Kawasumi T, Igarashi Y, Kodama T, Minoda Y. 1985. The Co2 Assimilation via The Reductive Tricarboxylic-Acid Cycle In An Obligately Autotrophic, Aerobic Hydrogen-Oxidizing Bacterium, *Hydrogenobacter-Thermophilus*. Archives of Microbiology **141:**198-203.

Sievert SM, Scott KA, Klotz MG, Chain PSG, Hauser LJ, Hemp J, Hugler M, Land M, Lapidus A, Larimer FW, Lucas S, Malfatti SA, Meyer F, Paulsen IT, Ren Q, Simon J, Class USFG. 2008. Genome of the Epsilonproteobacterial Chemolithoautotroph *Sulfurimonas Denitrifican*s. Applied and Environmental Microbiology **74:**1145-1156.

Sikorski J, Munk C, Lapidus A, Djao ODN, Lucas S, Del Rio TG, Nolan M, Tice H, Han C, Cheng JF, Tapia R, Goodwin L, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Mikhailova N, Pati A, Sims D, Meincke L, Brettin T, Detter JC, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Rohde M, Lang E, Spring S, Goker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP. 2010. Complete Genome Sequence of *Sulfurimonas autotrophica* type strain (OK10(T)). Standards in Genomic Sciences **3:**194-202.

Tabita FR. 1999. Microbial Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase: A Different Perspective. Photosynthesis Research **60:**1-28.

Tabita FR, Hanson TE, Satagopan S, Witte BH, Kreel NE. 2008. Phylogenetic and Evolutionary Relationships of Rubisco and the Rubisco-Like Proteins And the Functional Lessons Provided by Diverse Molecular Forms. Philosophical Transactions of the Royal Society B-Biological Sciences **363:**2629-2640.

Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, Nealson KH, Horikoshi K. 2003. Isolation and Phylogenetic Diversity of Members of Previously Uncultivated Epsilon-Proteobacteria in Deep-Sea Hydrothermal Fields. Fems Microbiology Letters **218:**167-174

Takai, K., T. Gamo, U. Tsunogai, N. Nakayama, H. Hirayama, K.H. Nealson, and K. Horikoshi. 2004. Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. *Extremophiles* **8**:269‑282

Takai K, Suzuki M, Nakagawa S, Miyazaki M, Suzuki Y, Inagaki F, Horikoshi K. 2006. Sulfurimonas paralvinellae sp nov., a novel mesophilic, hydrogen- and sulfur-oxidizing

chemolithoautotroph within the Epsilonproteo-bacteria isolated from a deep-sea hydrothermal vent polychaete nest, reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas denitrificans* comb. nov and emended description of the genus Sulfurimonas. International Journal of Systematic and Evolutionary Microbiology **56:**1725-1733.

Van Dover, C.L. 2000.The Ecology of Deep-Sea hydrothermal Vents. Princeton University Press, Princeton,New Jersey

Wang X, Falcone DL, Tabita FR. 1993. Reductive Pentose Phosphate-Independent Co2 Fixation In *Rhodobacter-Sphaeroides* And Evidence That Ribulose-Bisphosphate Carboxylase Oxygenase Activity Serves To Maintain The Redox Balance Of The Cell. Journal of Bacteriology **175:**3372-3379.

Weissbach A, Horecker BL, Hurwitz J. 1956. Enzymatic Formation Of Phosphoglyceric Acid From Ribulose Diphosphate and Carbon Dioxide. Journal of Biological Chemistry **218:**795-810.

Wirsen CO, Jannasch HW, Molyneaux SJ. 1993. Chemosynthetic Microbial Activity at Mid-Atlantic Ridge Hydrothermal Vent Sites. Journal of Geophysical Research-Solid Earth **98:**9693-9703.

Wood HG, Ragsdale SW, Pezacka E. 1986. The Acetyl-Coa Pathway of Autotrophic Growth. Fems Microbiology Letters **39:**345-362.

Yates RA, Pardee AB. 1956. Pyrimidine Biosynthesis in *Escherichia Coli*. Journal of Biological Chemistry **221:**743-756.

Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM. 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments Nature Reviews Microbiology **6:** 681-691.

Appendices

Appendix A: Random and site-directed mutagenesis of *T. crunogena* to identify genes responsible for its CO2-concentrating mechanism

Introduction

Deep in the world's oceans, unique communities of organisms thrive on reductant-enriched water flowing from hydrothermal vents. Ubiquitous at these hydrothermal vents are *Thiomicrospiras*, which are obligate chemolithoautotrophs. The ecophysiology of carbon fixation by these organisms, or any other chemolithoautotroph, has not been well-studied. The genome of *Thiomicrospira crunogena* has been sequenced, and provided some clues (Scott et al. 2006). Based on gene complement, it has adaptations for CO2 scarcity (carboxysomes), and based on physiological studies, it can generate intracellular CO2 concentrations considerably higher than extracellular. The purpose of this study was to elucidate the molecular mechanism for dissolved inorganic carbon (DIC) uptake by *T. crunogena.* To find the genes encoding system(s) for (DIC) uptake, random knockout mutants were generated and screened for inhibited growth under low CO_2 conditions.

Methods

Growth Medium

Thiomicrospira crunogena was grown in chemostats (New Brunswick Scientific BioFlo 110) in TASW medium, an artificial saltwater medium supplemented with 40 mM thiosulfate and 100 mM Na HEPES to maintain pH 8 (Dobrinski et al. 2005). Cells were grown under DIC limitation (0.1 mM DIC) and had the NaCl content dropped to 20% (65 mM) in order to facilitate mating with *E. coli* for random and site-directed mutagenesis. *Escherichia coli* was grown in TYE medium (Larsen et al.., 2002) supplemented with kanamycin (25 mg/L). TCTYE medium used for the mating of *T. crunogena* and *E. coli* consisted of tryptone (10 g/L), yeast extract (5g/L), NaCl (3.8g/L), $(NH_4)_2SO_4$ (0.8g/L), $MgSO_4 \cdot 7H_2O$ (1.4g/L), $CaCl_2 \cdot 2H_2O$ (0.2g/L), K_2HPO_4 (0.6g/L), $NaHCO_3$ (4.2g/L), $Na_2S_2O_3$ (anhydrous, 6.3 g/L), NaHEPES (26g/L), agar (15g/L) and SL-8 (Biebl and Pfennig, 1978).

Random Mutagenesis

Thiomicrospira crunogena cells were harvested and resuspended at an optical density of 20 at 600nm (OD_{600}) in TASW medium. *Escherichia coli* strain BW20760, carrying plasmid pRL-27 (Larsen et al.., 2002), were grown to an OD_{600}, washed twice in TASW and resuspended as described for *T. crunogena* above. Plasmid pRL-27 contains a transposon that has a kanamycin resistance cassette as an identifiable marker (Figure A11). The transposase gene does not move with the transposon into the chromosome, resulting in a stable insertion. To mate the cells, 50 μl of a 1:1 *T. crunogena/E. coli* suspension was added to solid TCTYE. The suspension was allowed to dry and form a biofilm. Plates were incubated overnight at 32ºC under a 5% CO_2 headspace. Biofilms were resuspended, washed three times with TASW, spread

onto recovery plates (TASW medium + 25 mg/ml kanamycin) and incubated under a 5% CO_2 headspace.
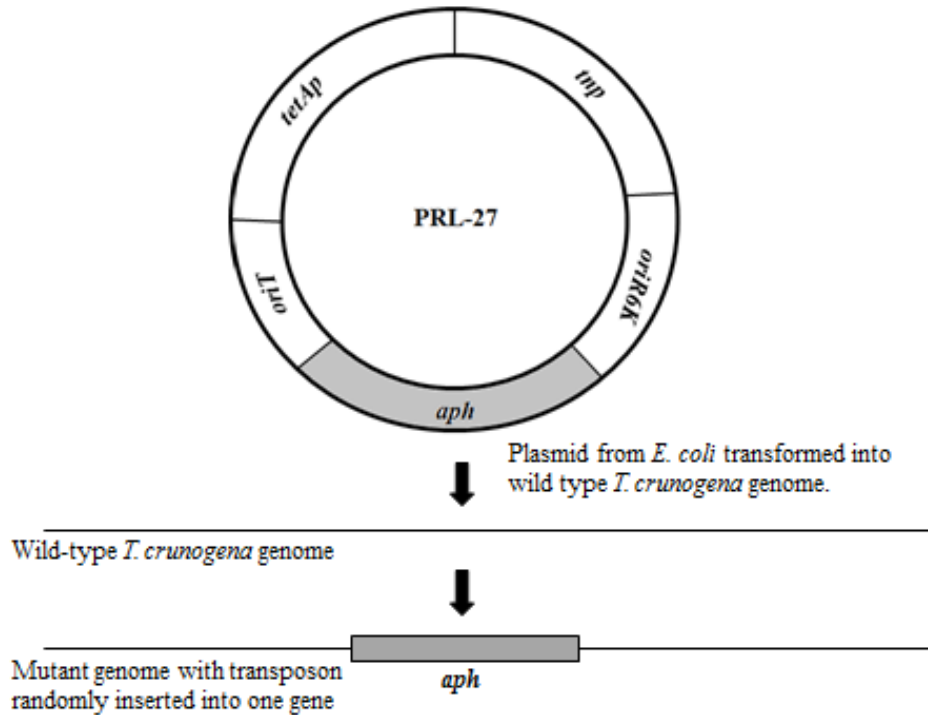
Figure A11| Generation of random knockout mutants. Plasposon PRL27 was propagated in *Escherichia coli* and transformed into *T. crunogena* via conjugation. The transposon on pRL27, which carries a kanameycin resistance gene as a selectable marker, inserted randomly in the wild type *T. crunogena* genome generating a random mutant library.

Site-directed Mutagenesis

Site-directed mutagenesis was used to generate strains of *T. crunogena* to use as controls for screening strains generated by random mutagenesis for CO_2 sensitivity. Two strains with mutations in genes encoding carboxysomal proteins were created (*Tcr_0838:* Rubisco large subunit *cbbL; Tcr_0841:* carboxysomal carbonic anhydrase *csoSCA*), anticipating that these strains would be sensitive to the CO_2 concentration available during growth (Cannon and

Shively, 1983), and could therefore function as positive controls when screening for this phenotype. A strain that could function as a negative control (insensitive to CO_2 concentration during growth) was generated by targeting a gene predicted to encode a conserved hypothetical protein located within a prophage (*Tcr_0668*). Transcript abundance from this gene is insensitive to the concentration of CO_2 during growth (Dobrinski et al. 2012), suggesting the protein product would be less likely to be part of a CCM. Further, its location within the prophage suggests that its protein product would not be necessary for *T. crunogena* growth in general.

Site-directed mutagenesis of *Tcr_0838, Tcr_0841,* and *Tcr_0668* was conducted as described in (Metcalf, 1996). Target genes were PCR-amplified from *T. crunogena* genomic DNA using Taq polymerase (Qiagen *Taq* PCR Master Mix Kit) and primers directed to each gene (Table 1; Integrated DNA Technologies, Inc.) which were modified at the 5' termini to carry a BamHI restriction sequence (AGGATCC). PCR products as well as a carrier plasmid (pcDNA3.1; Invitrogen, Inc.) were cleaved with BamHI. Cleaved plasmids were treated with Antarctic alkaline phosphatase (New England Biolabs, Inc.); linearized dephosphorylated plasmids and PCR products were then purified via phenol/chloroform and ethanol precipitation. Amplified target genes were then ligated into pcDNA3.1 (T4 Ligase; New England Biolabs, Inc.), and introduced into chemically competent *E. coli*Top10 cells (Invitrogen, Inc.). Transformants were selected by growth on solid Luria agar supplemented with 100 mg/L ampicillin. *E. coli* colonies were screened via PCR for the presence of one of the target *T. crunogena* genes, and plasmids were purified from PCR positives. Purified pcDNA3.1 plasmids carrying the target *T. crunogena* genes were subjected to Tn5-transposon mutagenesis using the EZ-Tn5<kan-2> kit (Epicentre, Inc.), transformed into chemically competent *E. coli*Top10 cells,

and selected by cultivation on solid Luria agar supplemented with ampicillin (100 mg/L) and

kanamycin (25 mg/L). Transformants with plasmids in which Tn5 inserted itself into the *T.*

crunogena gene carried by the plasmid were selected via colony-pick PCR, using primers for the

target *T. crunogena* genes; insertion of Tn5 into the *T. crunogena* gene increased the size of its

PCR product by 1221 nucleotides. Plasmids containing Tn5-interrupted *T. crunogena* genes

were purified, and mutated genes were sequenced to elucidate the position of Tn5-insertion

(Macrogen, USA). Plasmids carrying *T. crunogena* genes in which Tn5 was bracketed at 5' and

3' ends by at least 200 nucleotides of the target *T. crunogena* gene were carried forward, to

facilitate eventual RecA-mediated recombination into the *T. crunogena* genome (Metcalf, 1996).

Mutated genes were amplified from these plasmids using primers (Table 8) modified at

their 5' ends to include SpeI restriction sites (AACTAGT). PCR products, as well as *E. coli*

mating plasmid pLD55 (Metcalf, 1996), were digested with SpeI (New England Biolabs, Inc.),

purified and ligated as described above. This DNA was introduced into chemically competent *E.*

*coli*BW20767 (Metcalf, 1996), and transconjugants were selected by plating onto solid Luria

agar supplemented with 100 mg/L ampicillin and 25 mg/L kanamycin. To introduce mutated

target genes into *T. crunogena,* these *E. coli* BW20767 cells were then mated with *T. crunogena*

as described above for random mutagenesis.

Kanamycin-resistant *T. crunogena* were screened for the presence of mutated target

genes via PCR. A single RecA-mediated crossover event between the plasmid carrying the

mutated *T. crunogena* gene and the *T. crunogena* chromosome results in integration of the

plasmid into the *T. crunogena* chromosome (Figure A12); as a result, these cells carry two copies

of the target gene: one wild-type and one interrupted by Tn5. Cells that successfully mated with

E. coli and integrated the plasmid into the chromosome will therefore have two PCR products when screened using primers directed to the target gene. *T. crunogena* strains with two bands from PCR using primers to the target genes were carried forward.

These strains were next cultivated under conditions favoring cells in which a second RecA-mediated crossover event had occurred, removing the plasmid and the wild-type copy of the gene (Metcalf, 1996). Cells were cultivated under high-CO_2 conditions in liquid TASW supplemented with 25 mg/L chlortetracycline (added before autoclaving; degradation products induce the *tet*^{R} gene without killing tet^S cells) and 25 mg/L kanamycin. A set of 5 ml cultures was harvested via centrifugation and resuspended in TASW. A dilution series of the resuspended cells was plated on solid TASW medium supplemented with 25 mg/L chlortetracycline (degraded by autoclaving), 25 mg/L kanamycin, and 6 mg/L fusaric acid, which preferentially destabilizes cytoplasmic membranes of cells expressing TetR (Metcalf, 1996). The most rapidly growing colonies from these plates were screened for loss of the wild-type target gene via colony-pick PCR.
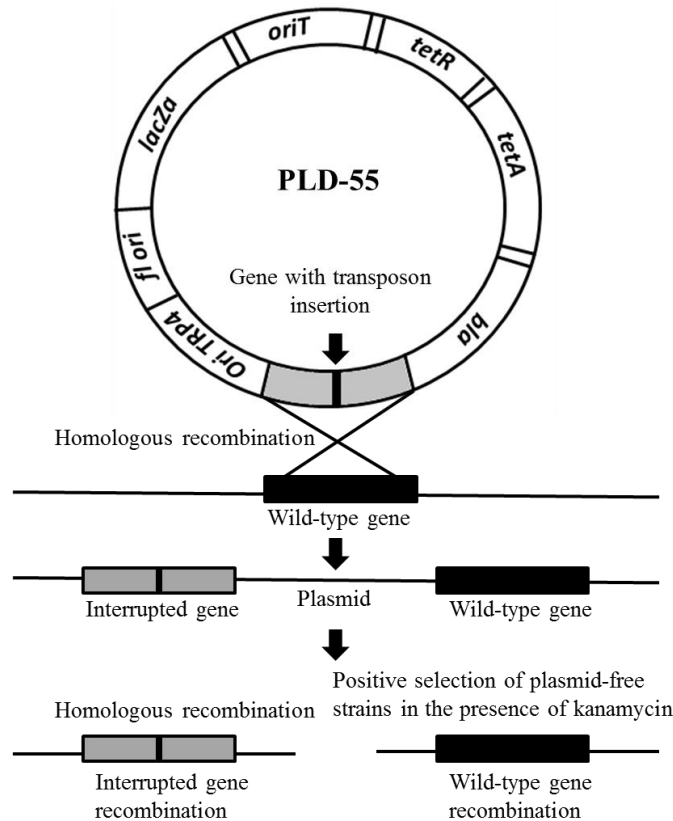
Figure A12 |Generation of site-directed mutants**.** pLD55 plasmid carrying mutant gene interrupted with a transposon containing Kanamycin resistance (indicated by black line through interrupted gene). Homologous recombination with the wild-type gene results in a genome containing both mutant and wild-type copies of the gene. A second round of homologous recombination results in the loss of the plasmid and the generation of strains carrying either a wild-type or mutant copy of the gene.

Assay to detect CO_2 sensitivity in random knockout mutants

A high throughput assay for growth was based on sulfuric acid production from thiosulfate by *T. crunogena*. Phenol red was added to TASW (P-TASW); growth resulted in medium color changing from red (pH 8) to yellow (pH < 6). Random knockout colonies were inoculated into 96 well plates (200 µl P-TASW/ well) and placed in 5% CO_2 headspace growth chambers at room temperature for approximately two weeks. Strains were then transferred into

two new 96 well plates containing either high or low DIC P-TASW. High DIC strains were grown under a 5% CO_2 headspace while low were kept at ambient (~.03% CO_2). Strains that grew well under high DIC conditions but were inhibited under low were selected for further investigation. Site-directed mutants served as controls for the assay. The positive control strain was the carboxysomal carbonic anhydrase mutant (*csoSCA:Tn5*) that resulted in a carbon sensitive phenotype while the negative control strain was the embedded prophage mutant that was predicted to be unlikely to have a CO_2 sensitive phenotype (*Tcr_0668:Tn5*).

Growth and protein assays to verify CO_2 sensitivity

CO_2 sensitivity was confirmed by monitoring growth of larger volume (40 ml) cultures of candidate strains under low and high DIC conditions via OD_{600}. Growth was measured until cells reached stationary phase and the culture was centrifuged to collect the cells. Proteins were extracted by vortexing cells in a solution of extraction buffer (1% triton and 20 mM Tris) and 3.5% beta-mercaptoethanol. Proteins were quantified using a RC DC protein assay kit microfuge tube protocol (BioRad Inc.).

PCR to eliminate carboxysome mutants and verify presence of transposon

It was anticipated that mutants whose carboxysomal genes were interrupted would exhibit a CO_2 sensitive phenotype (Cannon and Shively 1983). To remove these mutants and focus on those which might have mutated transport-related genes, each strain was screened with four PCR reactions, with four sets of primers (Table A11) which together covered the whole carboxysome operon (Tcr_0840-0846). Increased band size or missing bands indicated an interruption within the operon. Strains that had wild-type carboxysome genes but were CO_2 sensitive were selected for further scrutiny. To verify that kanamycin sensitivity was due to the

presence of the Tn5 transposon, strains were screened via PCR for the presence of the

transposon-encoded kanamycin gene (Table A11).

Table A11 | Primer sets, forward (f) and reverse (r), for kanamycin resistance (aph) which was used a selectable marker for transposon insertion. Tcr_0668 is a hypothetical protein found in a prophage embedded in the *T. crunogena* genome. Tcr_0841 encodes the large subunit of Rubisco (*cbbL*) while Tcru_0838 encodes carboxysomal carbonic anhydrase (*csoSCA*). Four primer sets were used to cover the entire length of the *T. crunogena* operon Tcr_0840-Tcr_0846.

Primers for Kanamycin resistance	
APH F	5'-CGA GGC CGC GAT TAA ATT CCA ACA-3'
APH R	5'-AGG CAG TTC CAT AGG ATG GCA AG -3'
Primers for site-directed mutagenesis	
Tcr_0668 F	5'-ACGTTCATGGCGACAACTGTTTTGATA-3'
Tcr_0668 R	5'-TTGGCCACTTCTCTCCATTTTCCTG-3'
Tcr_0841 F	5'-ATGAATCGTTTGAAAAAAAGTCATC-3'
Tcr_0841 R	5'-CTATGCGGTTCTTTGCTT-3'
Tcr_0838 F	5'-CAGCGGTAGCAGCAGAAAGTTCAA-3'
Tcr_0838 R	5'-ACCACCACCGAACTGAAGAACAGA-3'
Primer set for *T.crunogena* carboxysome operon	
Carboxy F 1	5'-TTG CGC GCT CCC GAT ATC TG-3'
Carboxy R 1	5'GAG TAT AGT TGT CAT ACC CAA CCA AAC GGA TC-3'
Carboxy F 2	5'-GGT ACA CGC GAT CCA AAT GAA GTT TTG G-3'
Carboxy R 2	5'-GTA ATT ACG CTC GTG CAC CAC CTG-3'
Carboxy F 3	5'-TAA CAA TGA AGT TCC GAT GAG CCC GAT T-3'
Carboxy R 3	5'-GAT GCC GCC ACC CGT TAA ATC TGT TAA T-3'
Carboxy F 4	5'-AAC AAA AGA TGG TGA CTG GGT TTT TAC CAT CG-3'
Carboxy R 4	5'-CGT TTG AGC GTT TTC GAC AGA AGC TTT ACT AG-3'

Results

Growth and protein assays to verify CO₂ sensitivity

Forty random knockout strains exhibited a CO_2 sensitive phenotype 'in the high-throughput 96-well assay and were grown in 40ml cultures to verify the results of the CO_2 sensitivity assay. Thirty-eight of these strains exhibited the same CO_2 sensitive phenotype when grown in flask cultures (Figure A13). Because elemental sulfur produced by the cells during growth cause false-high absorbance readings, protein measurements were taken for each strain. The protein assays confirmed that the strains did not grow as well under CO_2 limited conditions (Table A12).
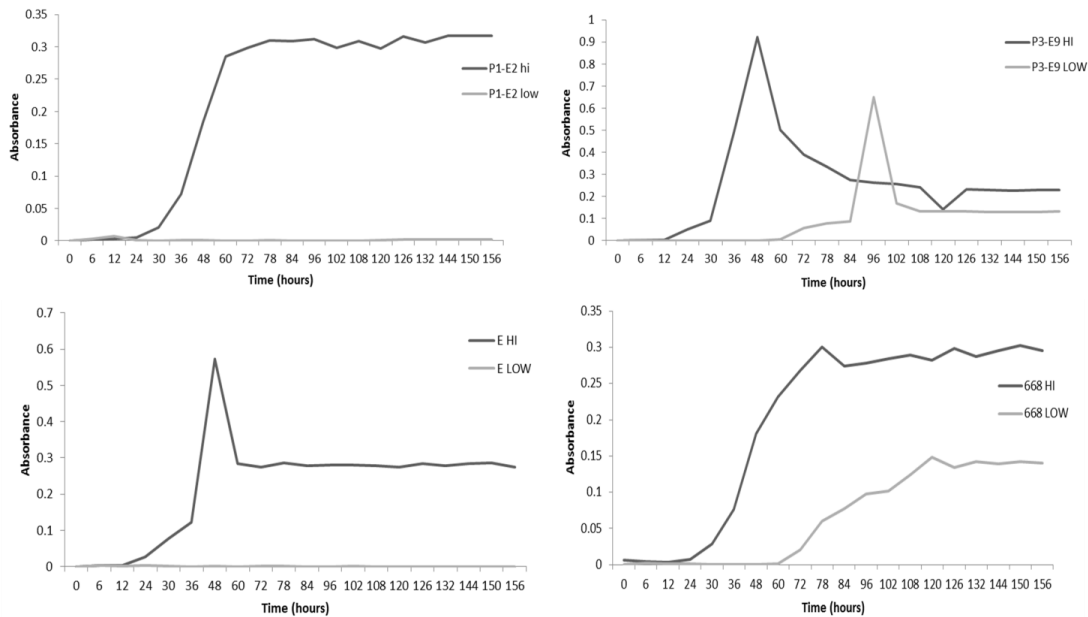

Figure A13 |Growth curves for random knockout strains. (Clockwise from top left). Strain P1-E2 shows strong growth at high CO_2 and no growth at low indicating carbon sensitivity. Strain P3-E9 showing strong growth at high CO_2 and delayed growth at low CO_2. The spike in absorbance is due to sulfur globules in the media. *T. crunogena* eventually reabsorbed the sulfur causing the drop in absorbance. Control strain *CsoSCA:Tn5* shows strong growth at high CO_2 and no growth at low CO_2. Control strain *Tcr_0668:Tn5* shows a slight delay in growth at low CO_2.

Table A12| Protein assay absorbance readings for a selection of random-knockout mutant strains.

Strain ID	High DIC	Low DIC
P1-E2	.520	.008
P1-H8	.721	.330
P2-A9	.520	.314
P2-G8	.679	.306
P3-E9	.520	.285
P3-F9	.552	.246

PCR to eliminate carboxysome mutants and verify presence of transposon

Thirty-eight random knockout strains expressed CO_2 sensitivity in the growth and protein assays (Figure A14). These strains were scanned via PCR in order to eliminate strains that contained mutations within the *T. crunogena* carboxysome. Strains *Tcr_0668:Tn5* and wild-type served as positive controls since it was known that they have wild-type carboxysomes while strain *cbbL:Tn5* served as a negative control due to its interruption within the caroboxysome operon. Out of the thirty-eight strains tested, sixteen exhibited wild-type carboxysome operons while the remaining strains had mutations within their carboxysome operons.
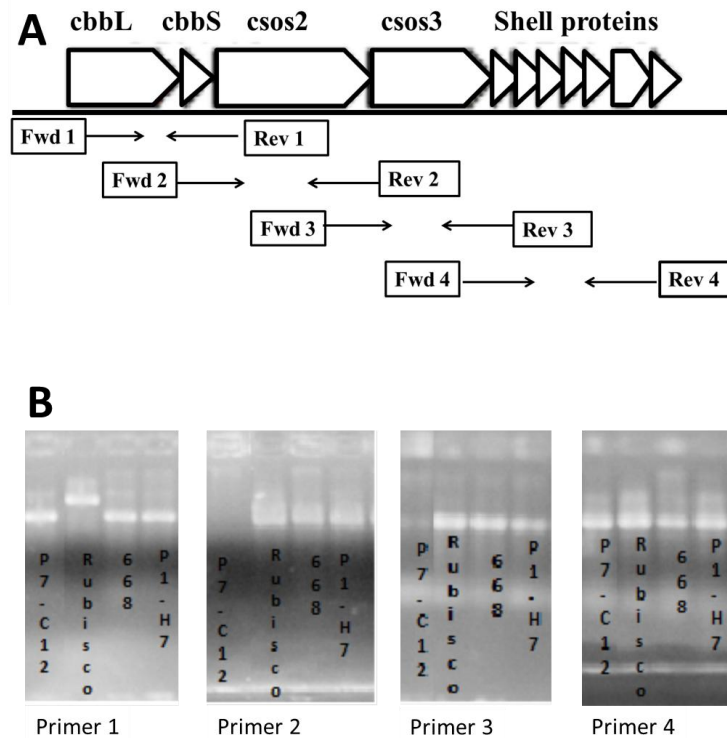
Figure A14 | PCR to detect carboxysome mutants. **A:** cbbL and cbbS represent the large and small subunits of Rubisco. $CSOS_2$ is a carboxysome shell peptide and CSOS 3 is the carboxysome shell carbonic anhydrase. **B:** Sample of PCR results from strains treated with four primers directed towards the carboxysome operon. For strain P7-C12, the missing band in primer 2 indicates that the cbbS and/or the csos2 genes were interrupted by the transposon. The larger primer 1 band for Rubisco results from site-directed interruption of the cbbL gene by a transposon. Strains 668 and P1-H7 have normal bands across all four primer sets indicating that they have uninterrupted wild-type carboxysomes.

PCR to detect presence of transposon within the mutant genome

The sixteen random knockout strains that expressed CO_2 sensitivity while having wild-type carboxysomes underwent PCR to scan for the Kanamycin (aph) gene (Figure A15). This was to ensure that the kanamycin resistance expressed by these strains was due to the presence of the transposon and not to a random mutation that conferred kanamycin resistance. Strains *Tcr_0668:Tn5*, *csoSCA:Tn5* and pRL27 served as controls as they are known to contain the transposon encoded APH gene. Only eight of the sixteen strains retained the aph gene. These

eight candidate strains were further investigated in order to find the site of the transposon

insertion within the genome. Preliminary attempts to use inverse PCR to characterize the location

of the transposon in the eight CO_2 sensitive strains with intact carboxysome operons were

unsuccessful.
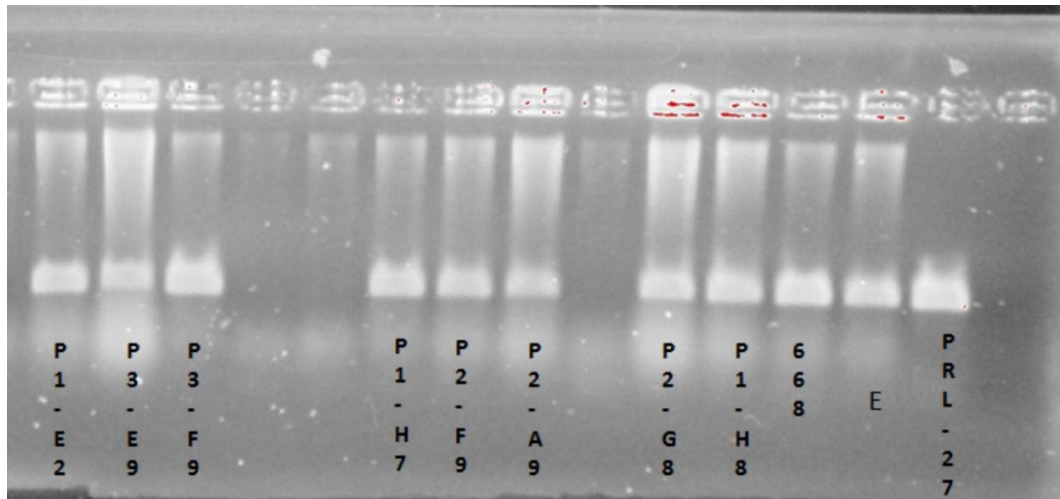
Figure A15 | Eight candidate strains containing the kanamycin (aph) gene. To ensure that the kanamycin resistance exhibited by the sixteen strains was due to the presence of the transposon, a PCR was run with primers directed towards the kanamycin resistance gene (aph).