Taming the Wild RubisCO: Explorations in Functional Metagenomics

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Brian Hurin Witte, M.S.

Graduate Program in Microbiology

The Ohio State University

2012

Dissertation Committee
:
F. Robert Tabita, Advisor

Joseph Krzycki

Birgit E. Alber

Paul Fuerst

Abstract

Ribulose bisphosphate carboxylase/oxygenase (E.C. 4.1.1.39) (RubisCO) is the most

abundant protein on Earth and the mechanism by which the vast majority of carbon enters

the planet's biosphere. Despite decades of study, many significant questions about this

enzyme remain unanswered. As anthropogenic CO_2 levels continue to rise,

understanding this key component of the carbon cycle is crucial to forecasting feedback

circuits, as well as to engineering food and fuel crops to produce more biomass with few

inputs of increasingly scarce resources. This study demonstrates three means of

investigating the natural diversity of RubisCO. Chapter 1 builds on existing DNA

sequence-based techniques of gene discovery and shows that RubisCO from uncultured

organisms can be used to complement growth in a RubisCO-deletion strain of autotrophic

bacteria. In a few short steps, the time-consuming work of bringing an autotrophic

organism in to pure culture can be circumvented. Chapter 2 details a means of entirely

bypassing the bias inherent in sequence-based gene discovery by using selection of

RubisCO genes from a metagenomic library. Chapter 3 provides a more in-depth study

of the RubisCO from the methanogenic archaeon *Methanococcoides burtonii*. *Mc.

burtonii* RubisCO (MBR) is unique in being intermediate between two previously-

recognized families of RubisCO, as well as having an unprecedented C-terminal loop

structure. Deletion of all or part of the loop appears to improve the oxygen tolerance of MBR, while simultaneously disrupting ability of the protein to form a decameric holoenzyme. This is the first report of a structural feature in RubisCO that can prevent the association of RubisCO dimers into higher-order structures without eliminating the catalytic activity of the enzyme.

Dedication

I would like to dedicate this document to the friends and family who have supported me through what seemed at times a Sisyphean project. In particular, I want to thank Julia for her steadfast belief that I would someday roll that rock over the crest of the hill.

Thank you.

Acknowledgments

I would like to acknowledge the help and guidance I received from members of my lab and department over the course of this research. In particular, I would like to thank Dr. Sriram Satagopan for his insight in to the mysteries of RubisCO and structural modeling. My undergraduate assistant Kim McCabe also made several valuable contributions to the study of *Methanococcoides burtonii* RubisCO. Drs. John Paul and David John of the University of South Florida and Dr. Boris Wawrik, currently of the University of Oklahoma, provided invaluable assistance in metagenomics research. My fellow graduate students, now Drs., David Longstaff and Jitesh Soares provided discussions on the nature of research and helped troubleshoot many problems large and small.

And, of course, I would like to thank my committee for their help, insight and patience. In particular, I appreciate the opportunity to be a member of Dr. Tabita's research group, and the insight, encouragement and interesting research projects he has provided.

Vita

1992..Columbia River H.S.

1998..B.S. Botany, University of Washington

2002 to present ....Graduate Teaching Associate/Graduate
Research Associate, Department of
Microbiology, The Ohio State University

2002-2003 ....University Fellowship, The Ohio State
University

2003-2006 ....National Science Foundation, Pre-doctoral
Fellowship

2008..M.S. Microbiology, The Ohio State
University

Publications

Blum JS, Han S, Lanoil B, Saltikov C, Witte B, Tabita FR, Langley S, Beveridge TJ, Jahnke L, Oremland RS. 2009. Ecophysiology of "*Halarsenatibacter silvermanii*" strain SLAS-1T, gen. nov., sp. nov., a facultative chemoautotrophic arsenate respirer from salt-saturated Searles Lake, California. Appl Environ Microbiol 75(7):1950-60.

Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, Santini JM, Oremland RS. 2007. *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int J Syst Evol Microbiol 57(Pt 3):504-12.

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 of London.Series B, Biological Sciences 363(1504):2629-40.

Witte B, John D, Wawrik B, Paul JH, Dayan D, Tabita FR. 2010. Functional prokaryotic RubisCO from an oceanic metagenomic library. Appl Environ Microbiol 76(9):2997-3003.

Fields of Study

Major Field: Microbiology

Table of Contents

List of Tables

List of Figures

Introduction

Ribulose bisphosphate carboxylase/oxygenase (RubisCO) (EC 4.1.1.39) is an enzyme of

remarkable diversity in terms of both sequence and function. In its role as the central

enzyme of the Calvin-Benson-Bassham (CBB) cycle, it catalyzes both the carboxylation

and oxygenation of ribulose 1,5-bisphosphate (RuBP) to generate 2 molecules of 3-

phosphoglycerate (3PGA) or one molecule of 3PGA and one molecule of 2-

phosphoglycolate (2PG), respectively. (Figure 1) Of the estimated 210 petagrams of CO_2

removed from the atmosphere each year, approximately 57% leaves due to biological

activity (Lal 2008) – and the vast majority of that biological activity is due to

photosynthesis in conjunction with the CBB cycle. The reactions of the CBB cycle are

outlined in Figure 2. The CBB cycle is employed by a wide array of organisms,

including bacteria oxidizing iron in virgin benthic basalt (Edwards 2011); autotrophic sea

slugs (Rumpho *et al* 2008); hyper-acidic speleo-snottites (Jones *et al* 2011); and arsenite-

respiring bacteria in desert soda lakes (Hoeft *et al* 2007) and the RubisCO enzymes they

harbor are correspondingly diverse. Making matters even more complicated, RubisCO is

a frequently trafficked enzyme - there is extensive evidence of horizontal gene transfer

(Horken and Tabita 1999a; Schopf 2011; Tabita *et al* 2008b; Turova and Spiridonova

2009) of RubisCO individually and in conjunction with large segments of the CBB cycle.

RubisCO is not limited just to the CBB cycle however. Many euryarchaea, for example,

1

possess fully functional RubisCO without a CBB cycle. Rather, the archaeal RubisCO

appears to participate in an AMP-recycling pathway in *Thermococcus kodakarensis* (Sato

et al 2007) (see Figure 3 for an illustration) or a 5-phospho-D-ribose-1-pyrophosphate

scavenging reaction (Finn and Tabita 2004, Estelmann *et al* 2011) in *Archaeoglobus spp.*

and *Methanosarcina acetivorans*. None of the cited examples, however, appear to fix

CO_2 at a rate sufficient to provide a significant portion of cell carbon. More recently, a

number of studies have shown that RubisCO is able to catalyze a reaction in the

methionine salvage pathway in addition to the canonical carboxylation/oxygenation

reactions (Singh and Tabita 2010).

 Finally, the "RubisCO Like Proteins" (RLPs) are structurally nearly identical to "bona

fide" RubisCO, and catalyze reactions involving small sulfur-containing substrates,

including in methionine salvage pathways (Nakano et al; Tabita et al 2008a; Tabita et al

2008b). The reaction mechanism of some RLPs remain as obscure as their physiological

function, but some reaction mechanisms are known with a fair degree of precision

(Ashida et al 2008; Imker et al 2008; Singh and Tabita 2010). In the case of the RLP

from Geobacillus kaustophilus, the enzyme catalyzes the enolization of the methionine

recycling pathway intermediate 2,3-diketo-5-methylthiopentane 1-phosphate (DK-MTP

1-P) to 2-hydroxy-3-keto-5-methylthiopent-1-ene 1-phosphate (HK-MTP 1-P). (See

Figure 4 for an illustration.) A perusal of the tree presented in Figure 5 shows that even

with the small sample of RLPs used to construct this tree, there are deep divisions,

indicating significant diversity. Although this study does not touch directly on the RLPs, it should be remembered that they are, functionally and structurally, part of the great family of RubisCO proteins, and a further indicator of how much remains to be learned.

The diversity of "bona fide" RubisCO (that is, capable of catalyzing the canonical oxygenation and carboxylation of RuBP) is due in part to the important role it plays in autotrophic carbon fixation. Although other biochemical pathways provide more efficient means of obtaining cellular carbon (Bar-Even *et al* 2010; Pereto *et al* 1999; Shively *et al* 1998) the CBB cycle is unique in being oxygen tolerant and wide-spread. Alternate pathways, such as the Wood-Ljungdahl (Wood 1991) and reductive citric acid cycle (Evans *et al* 1966) are both strongly inhibited by the presence of molecular oxygen; the 3-hydroxypropionate (Strauss and Fuchs 1993) bicycle is oxygen tolerant, but found in very few taxa. Since the advent of an aerobic biosphere some 2.5 billion years ago, (see Figure 6) the ability to fix carbon despite the presence of oxygen has been essential. RubisCO may well predate even that ancient era, as evidenced by the presence of cyanobacteria-like microfossils from 3.5 billion years ago (Schopf 1993) and isotopic signatures in geological formations consistent with an origin in RubisCO (Schopf 2011).

Understanding the role of RubisCO in the biosphere is more important now than ever before given the steadily increasing concentration of fossil carbon liberated by human activity. (See Figure 7) The importance includes gaining a better understanding of natural sinks of CO_2 (e.g. monitoring RubisCO transcript levels to track carboxylation activity in a variety of marine ecological niches (John *et al* 2007a,b).

3
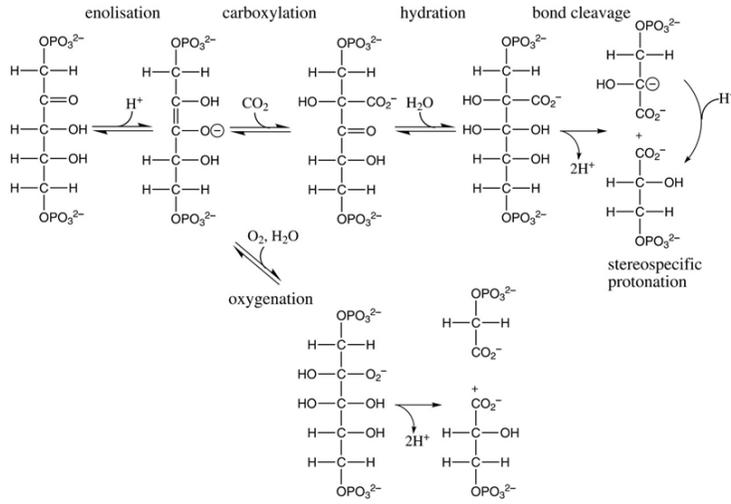
Figure 1 - RubisCO Reaction

Carboxylation and oxygenation reactions of RuBP catalyzed by RubisCO.
From Andersson (2008)
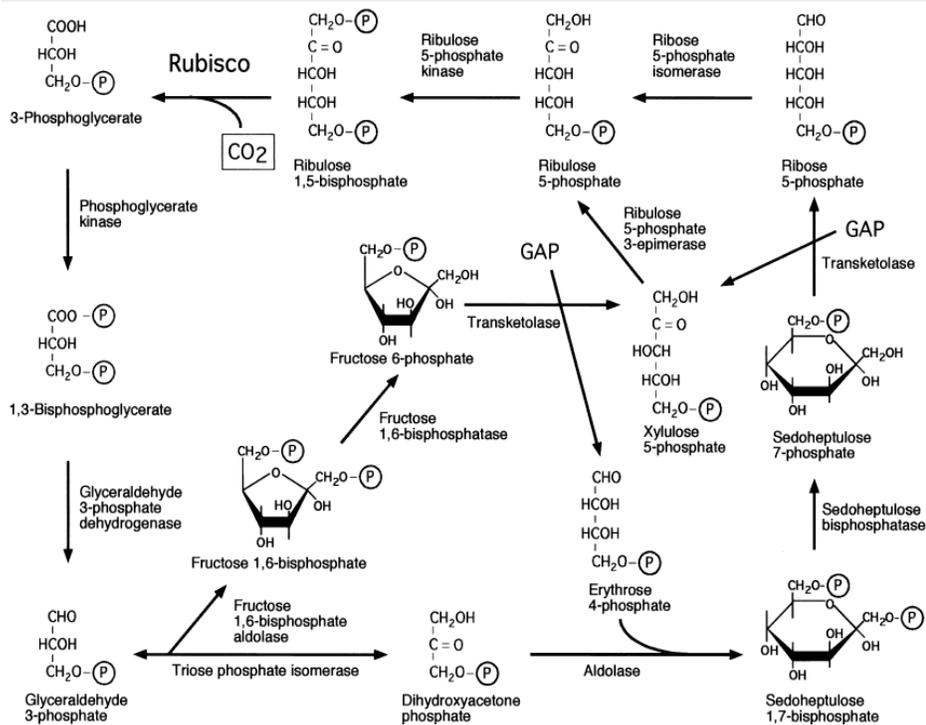
Figure 2 - Calvin-Benson-Bassham Cycle

From Atomi (2002). Reactions are only outlined in respect to carbon flow. GAP refers to Glyceraldehyde 3-phosphate.
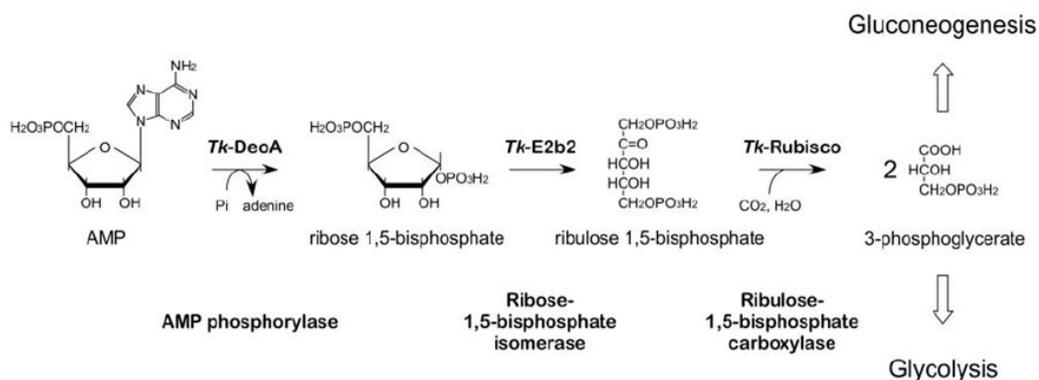
Figure 3 - AMP Recycling Pathway

The AMP recycling pathway in *Thermococcus kodakariensis* as proposed by Sato (2007).
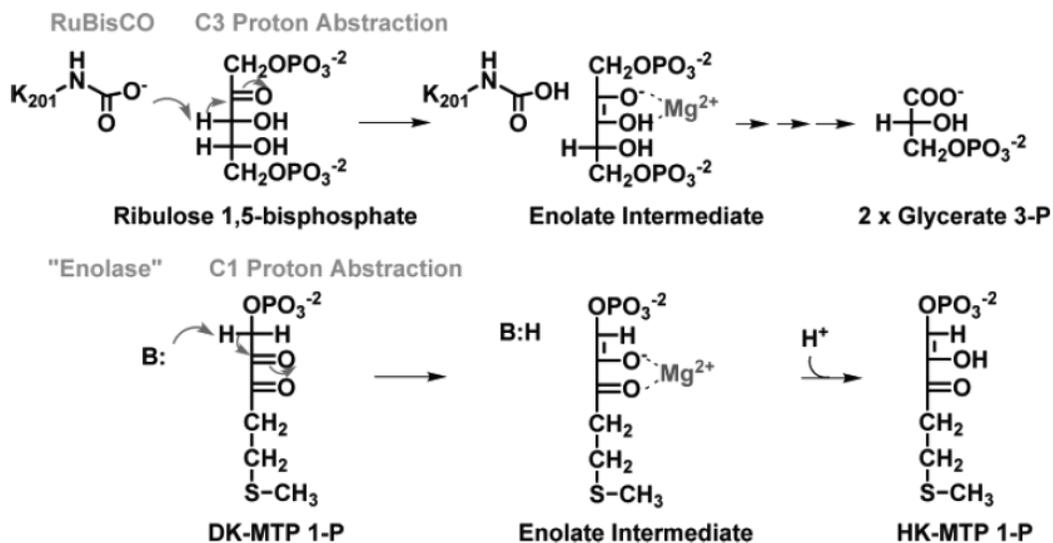
Figure 4 - RubisCO-Like Protein Enolization Reaction

From Imker *et al* (2007) DK-MTP 1-P = 2,3-diketo-5-methylthiopentane 1-phosphate; HK-MTP 1-P = 2-hydroxy-3-keto-5-methylthiopent-1-ene 1-phosphate
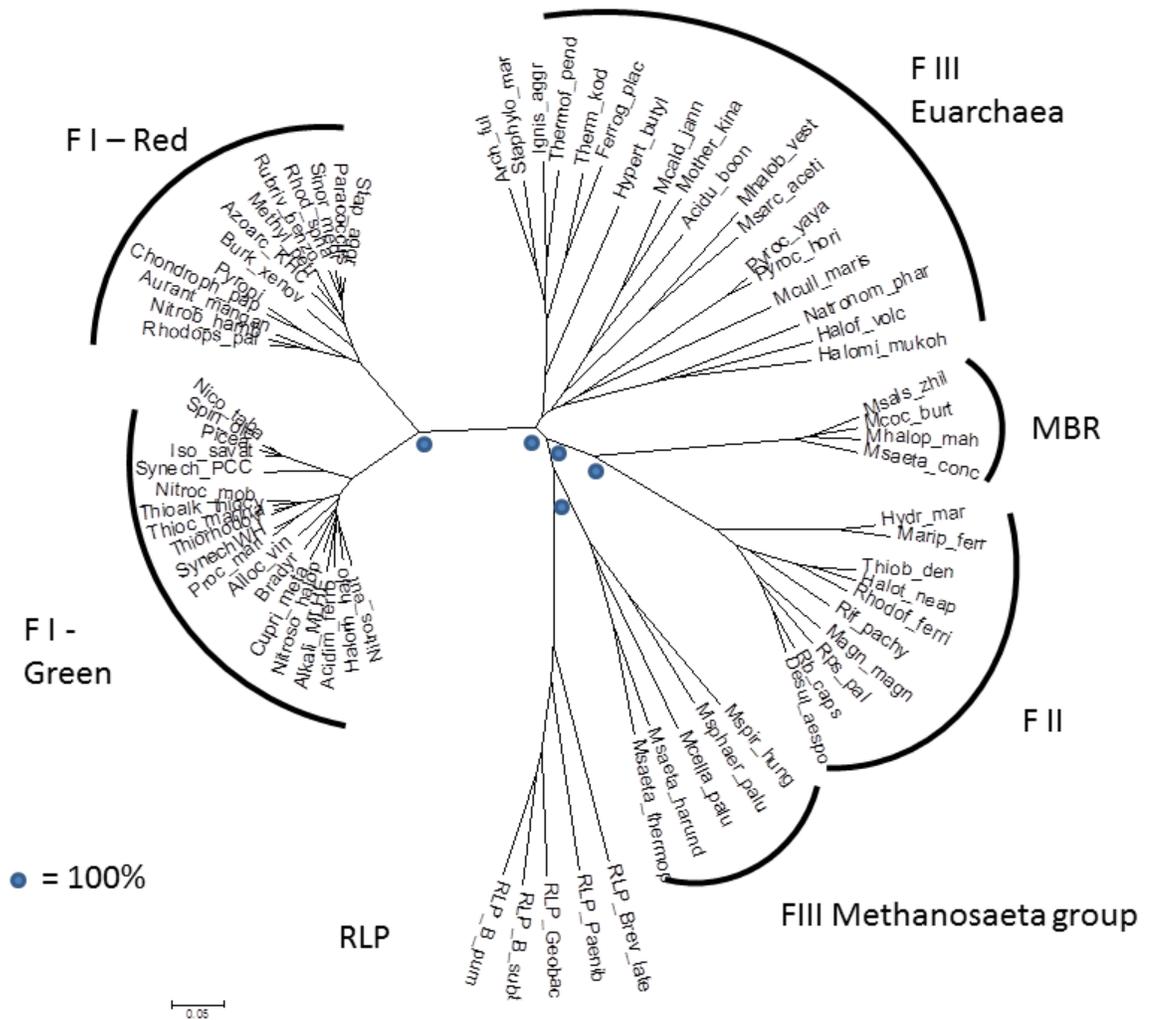
Figure 5 - Phylogenetic Tree of RubisCO Large subunit sequences

Minimum evoluition tree of RubisCO large subunit protein sequences. FI, FII and FIII refers to form I, II and III of rubisCO. Form I – Green can be further subdivided in to IA and IB, which correspond to the land plant and *Synechococcus* group and the bacterial group, respectively. Form I – Red can be further subdivided in to IC and ID groups, corresponding to RubisCOs from purple bacteria and non-green algae, respectively (Delwiche and Palmer 1996; Watson and Tabita 1996, 1997). MBR here refers to the sequences that share significant similarity to the *Methanococcoides burtonii* RubisCO. RLP refers to the so-called RubisCO-like proteins. Details of tree construction are presented in Chapter 3 – Methods. Dots at branch points refer to 100% confidence after 1000 bootstrap iterations.
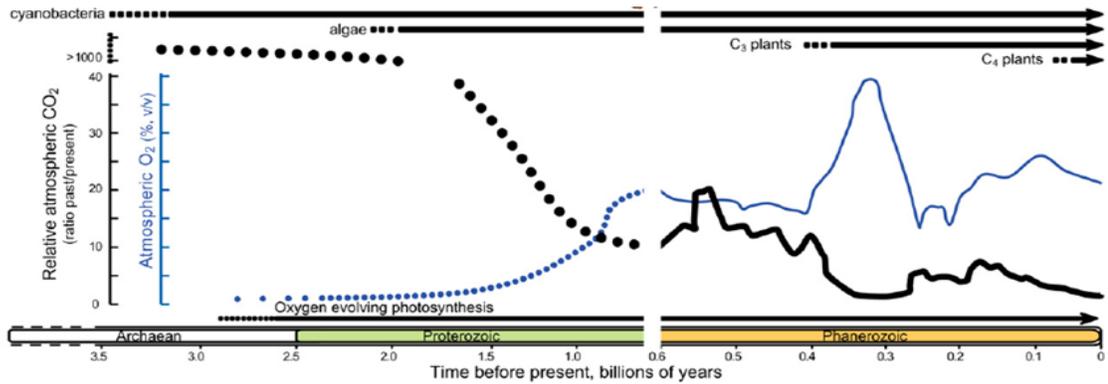
Figure 6 - Atmosphere Through Time

Change in atmospheric CO_2 and O_2 through time. Figure from Whitney *et al* (2011).
Their hypothesis is that RubisCO evolved in the anaerobic biosphere, but that the Calvin
Cycle became the predominant CO_2 fixation scheme as oxygenic photosynthesis gradually
made other autotrophic pathways impracticable in many habitats.
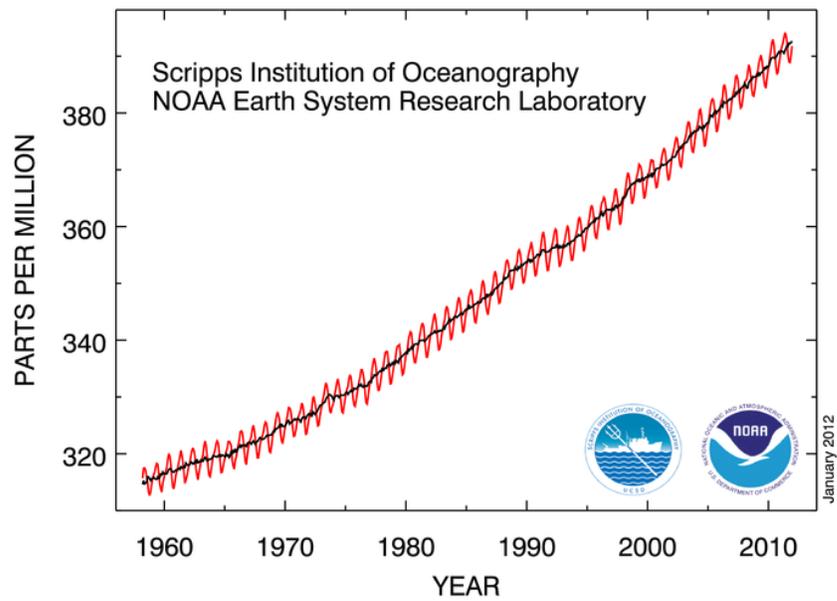
Figure 7 - Atmospheric CO_2 Concentration at Mauna Loa Observatory

Change in CO_2 in the modern era. (Tans and Keeling 2011) Red line records seasonal changes,
while the black line presents the running average.

Alternatively, better understanding of how RubisCO sequence translates to function could allow for engineering better, more efficient plants and algae, in terms of carbon captured per units of nitrogen or water consumed. More efficient crops means better water use efficiency in an era of declining availability of fresh water and increasingly scarce food and fuel.

RubisCO, of course, is only one of a great many enzymes with an ancient lineage. What differentiates RubisCO is that it is still directly affected by the molecular oxygen with which it was recruited to cope. RubisCO is the enzyme which directly mediates the withdrawal of CO_2 from the atmosphere and that consumes the energy captured by the oxygen-evolving photosystem of plants, algae and cyanobacteria. There may be no other enzyme that so intimately ties the living biosphere to its geochemical context.

Although Figure 5 presents a more complicated picture of RubisCO phylogeny, Tabita *et al* (2008b) notes four "forms" of RubisCO based on phylogenetic analysis of large subunit protein sequences. The forms (I-IV) of RubisCO, are based on phylogenetic analysis, but also possess consistent within-group structural features. Higher order structures vary, but all RubisCOs occur as $(L_2)_x$ where x indicates that multimeric holoenzymes are comprised of dimers of large subunits (LSUs). All known examples form two distinct active sites at the interface between two subunits. The residues that directly interact with the substrates (see Figure 8) are invariant. See Tabita *et al* (2008b)

```
                         C         R                                  C
I Synechococcus 6301   D E A G A A I A A E S S T G T W T T V W T |68   S V T N I L T S I V G N V F G F K A I R S L  130
I R.capsulatus         E E A A A A V A A E S S T A T W T T V W T |64   S V V N V F T S L V G N V F G F K A V R A L  126
I Prochlorococcus      E E V A A A V A A E S S T G T W S T V W S |63   S I T N V L T S L V G N V F G F K A L R H L  125
II Rs.rubrum           V A T A A H F A A E S S T G T N V E V C T  58   M I A S F L T L T M G N N Q G M G D V E Y A  121
II Ri. pachyptila      L E A A A H F A A E S S T G T N V E V S T  61   M L A S F L T L T I G N N Q G M G D I E Y A  123
II/III Mc burtonii     L Q A A A E I A A E S S T G T N I K V S T  68   N V Q N I L T Y I I G N I L G M K E I Q A L  125
III Ms.acetivorans C2A E E A S T H M A G E S S I D S W T E I S T  52   S V P Q V L S A V A G N I F S M K I V D N L  109
III A.fulgidus         E D A A G A V A A E S S T G T W T S L H P  57   N M P G L L A S I A G N V F G M K R V K G L  118
III T.kodakaeransis    E Q A A G A V A A E S S T G T W T T L Y P  60   N L P G L L A S I A G N I F G M K R V K G L  121

                       C   C C                                       R   R R
I Synechococcus 6301   C T I K P K L G L S A K N Y G R A V Y E C L R G G L D F T K D D E N I N S Q  206
I R.capsulatus         C T I K P K L G L A A K N Y G R A V Y E C L R G G L D F T K D D E N V N S Q  202
I Prochlorococcus      C T I K P K L G L S G K N Y G R V V Y E C L R G G L D L T K D D E N I N S Q  201
II Rs.rubrum           T I I K P K L G L R P K P F A E A C H A F W L G G - D F I K N D E P Q G N Q  199
II Ri. pachyptila      T I I K P K L G L R P E P F A E A A Y G F W L G G - D F I K N D E P Q G N Q  201
II/III Mc burtonii     T I V K P K M G L T S A E Y A E V C Y D F W V G G G D F V K N D E P Q A N Q  201
III Ms.acetivorans C2A T I V K P K V G L T S E M H A E V A Y N A F A G G C D L V K D D E N L T D Q  185
III A.fulgidus         T V P K P K V G Y S A E E V E K L A Y E L L S G G M D Y I K D D E N L T S P  194
III T.kodakaeransis    V V P K P K V G Y S P E E F E K L A Y D L L S N G A D Y M K D D E N L T S P| 197

                             C R                                                  R              C
I Synechococcus 6301   V L L H I H R A M H A V I D R - Q R N H G I H F R V L A K C L R L S G G D H L H S G T V - V G K L  332
I R.capsulatus         M L L H V H R A M H A V M D R - N P N H G I N F R V L A K I L R L M G G D H L H S G T V - V G K L  328
I Prochlorococcus      M L L H I H R A M H A V I D R - N P K H G I H F R V L A K C L R L S G G D L H T G T V - V G K L  327
II Rs.rubrum           N F L H Y H R A G H G A V T S P Q S K R G Y T A F V H C K M A R L Q G A S G I H T G T M G F G K M  330
II Ri. pachyptila      Q F L H F H R A G H G A V T S P Q S K R G Y T A F V H I K M T R L L G A S G M H V G T M G Y G K M  332
II/III Mc burtonii     V F L H F H R A A H G A F T R Q E N P I G F S V L V L S K F A R L A G A S G I H T G T A G I G K M  331
III Ms.acetivorans C2A L A L H A H R C M H S A F T R - N P R H G V S M L L V A K L C R L I G L D Q L H I G T V - V G K M  311
III A.fulgidus         L A I H G H R A M H A A F T R - N A K H G I S M F V L A K L Y R I I G I D Q L H I G T A G A G K L  320
III T.kodakaeransis    L A I H G H R A M H A A F T R - N P Y H G I S M F V L A K L Y R L I G I D Q L H V G T A G A G K L  323

                         R   R                                      R R
I Synechococcus 6301   A S G G I H V W H M P A L V E I F G - D D S V L Q F G G G T L G H P W G N A P G A T A N R V A L E  422
I R.capsulatus         A S G G I H V W H M P A L V S I F G - N D S V L Q F G G G T L G H P W G N A A G A C A N R V A L E  418
I Prochlorococcus      A S G G I H V W H M P A L L A I F G - D D S C L Q F G G G T H G H P W G S A A G A A A N R V A L E  417
II Rs.rubrum           I S G G M N A L R M P G F F E N L G N A N V I L T A G G G A F G H I D G P V A G A R S L R Q A W Q  415
II Ri. pachyptila      I S G G M N A L R L P G F F E N L G H V N T A G G G T Y G H I D G P V A G A R S L R Q A Y E  417
II/III Mc burtonii     V S G G L N P V K L K P F I D V M E N V D F I T T M G S G V H S H P G G T Q S G A K A L V Q A C D  446
III Ms.acetivorans C2A A S G G L A P T M I P D L Y S I F G - K E V I M Q F G G G I H A H P M G T A A G A A A C R Q A L E  400
III A.fulgidus         S S G G L H P G N L E P V I D A L G - K E I V I Q V G G G V L G H P M G A K A G A K A V R Q A L D  410
III T.kodakaeransis    S S G G L H P G N I Q P V I E A L G - T D I V L Q L G G G T L G H P D G P A A G A R A V R Q A I D  413
```

Figure 8 - Conserved Residues in RubisCO

The protein sequences aligned using ClustalW (Larkin et al 2007; Goujon et al 2010) The accession numbers for each large subunit gene sequence are as follows: *A. fulgidus* RbcL2, O28635; *T. kodakaraensis* RbcL, O93627; *Synechococcus* PCC 6301 RbcL, P00880; *Prochlorococcus marinus* str MIT 9301 RbcL YP_001090800; *Rhodospirillum rubrum* CbbM P04718; *Riftia pachyptila* CbbM AAC38280; *Methanococcoides burtonii* RbcL ABE53176. C = catalytic residues, R = ribulose bis-phosphate binding residues
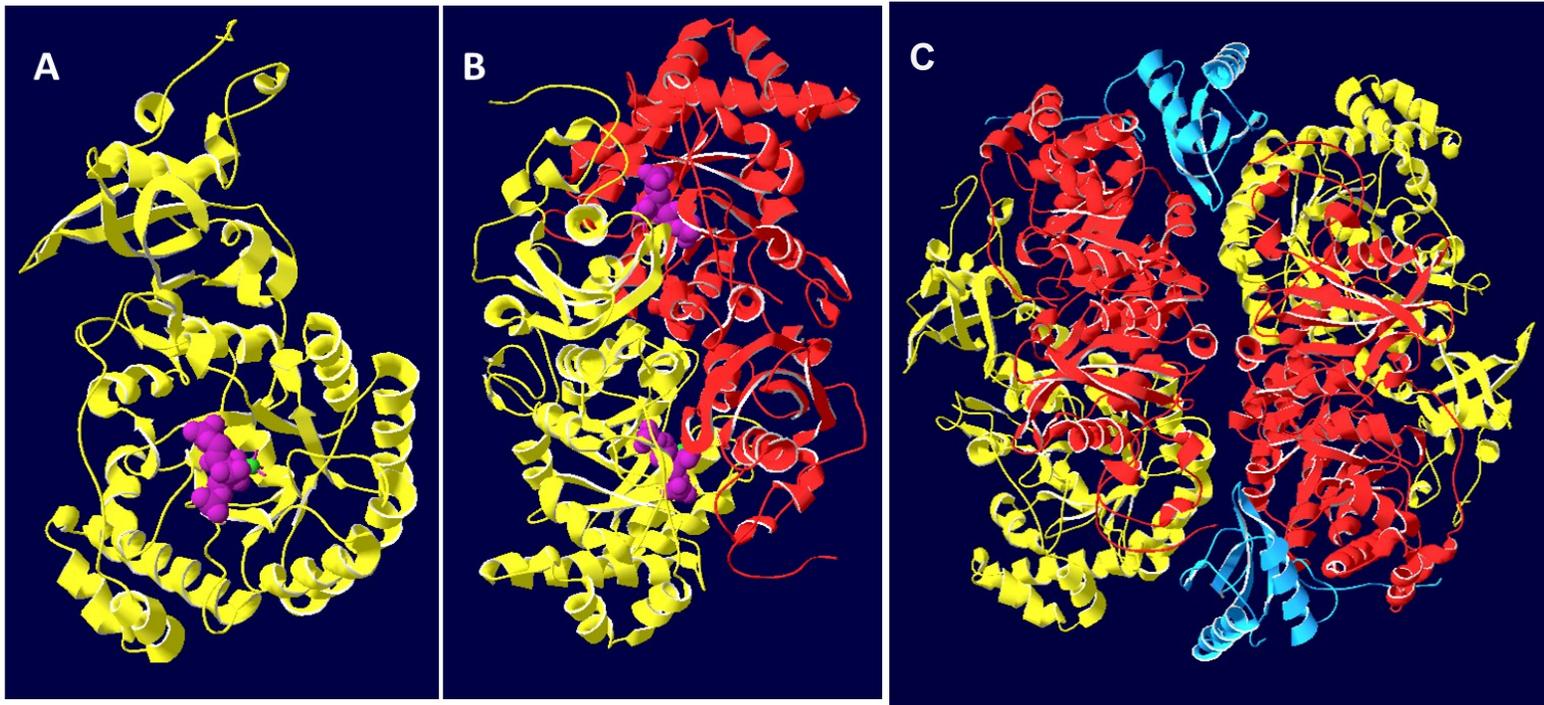
Figure 9 - Form I RubisCO Large and Small Subunits

The 1RBL structure of *Synechococcus* PCC6301 RubisCO was used as the model. A: a single large subunit with the substrate analog 2-carboxyarabinitol bisphosphate rendered in purple. Only one of two active sites is shown. B: The view from A rotated 90° and the other subunit of the dimer is pictured. The two active sites are identifiable by the positions of the 2-carboxy arabinitol bisphosphate (CABP - a non-reactive substrate analog of RuBP) molecules at the subunit interface. C: Four large subunits with two small subunits (blue) showing how the small subunits are positioned between large subunit dimers
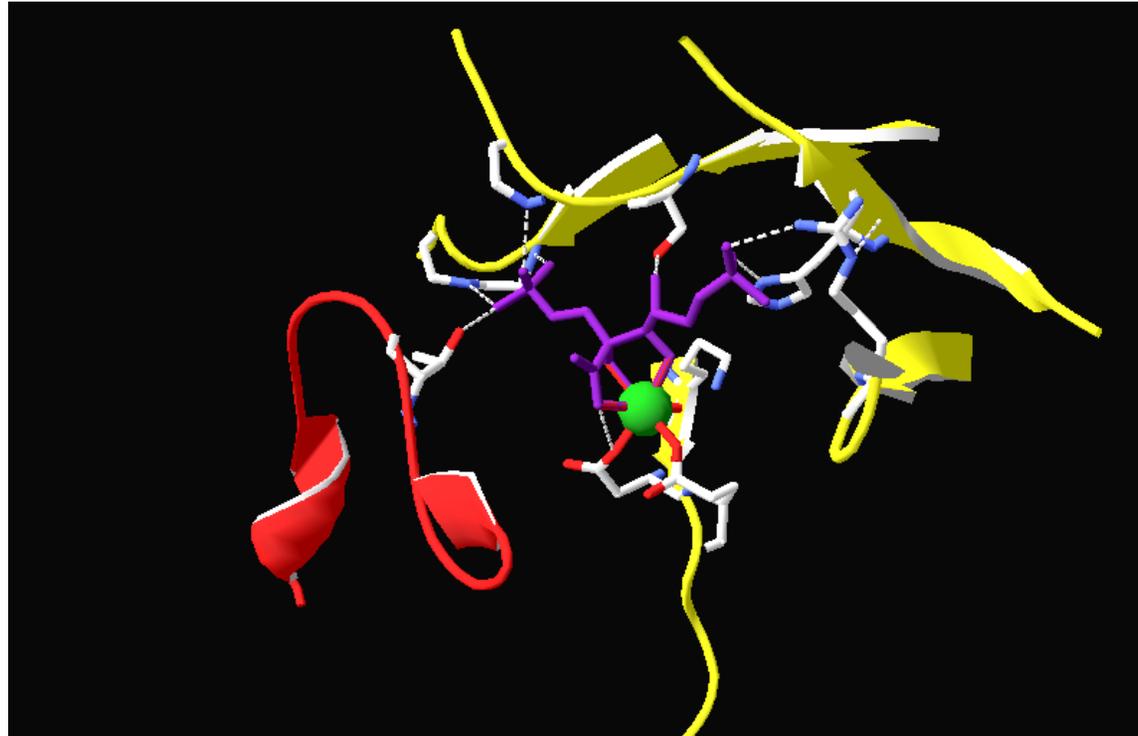
Figure 10 - RubisCO Active Site

Substrate analog 2-carboxy arabinitol bis-phosphate (CABP) is rendered in purple, the Mg^{2+} in green. The residues that bind the substrate are indicated by stick figures (denoted as R in Figure 8 - Conserved Residues in RubisCO). Nitrogen atoms in the side chains are indicated in blue, oxygen in red. Hydrogen atoms are not shown. The hydrogen bonds that connect the side chains and the substrate are shown as white dashed lines. Two large subunits participate in forming the active site (yellow and red ribbons). See Figure 9 - Form I RubisCO Large and Small Subunits, parts A and B for the position of the active site(s) in relation to the dimeric LSUs. Figure based on 1RBL – *Synechococcus* PCC6301
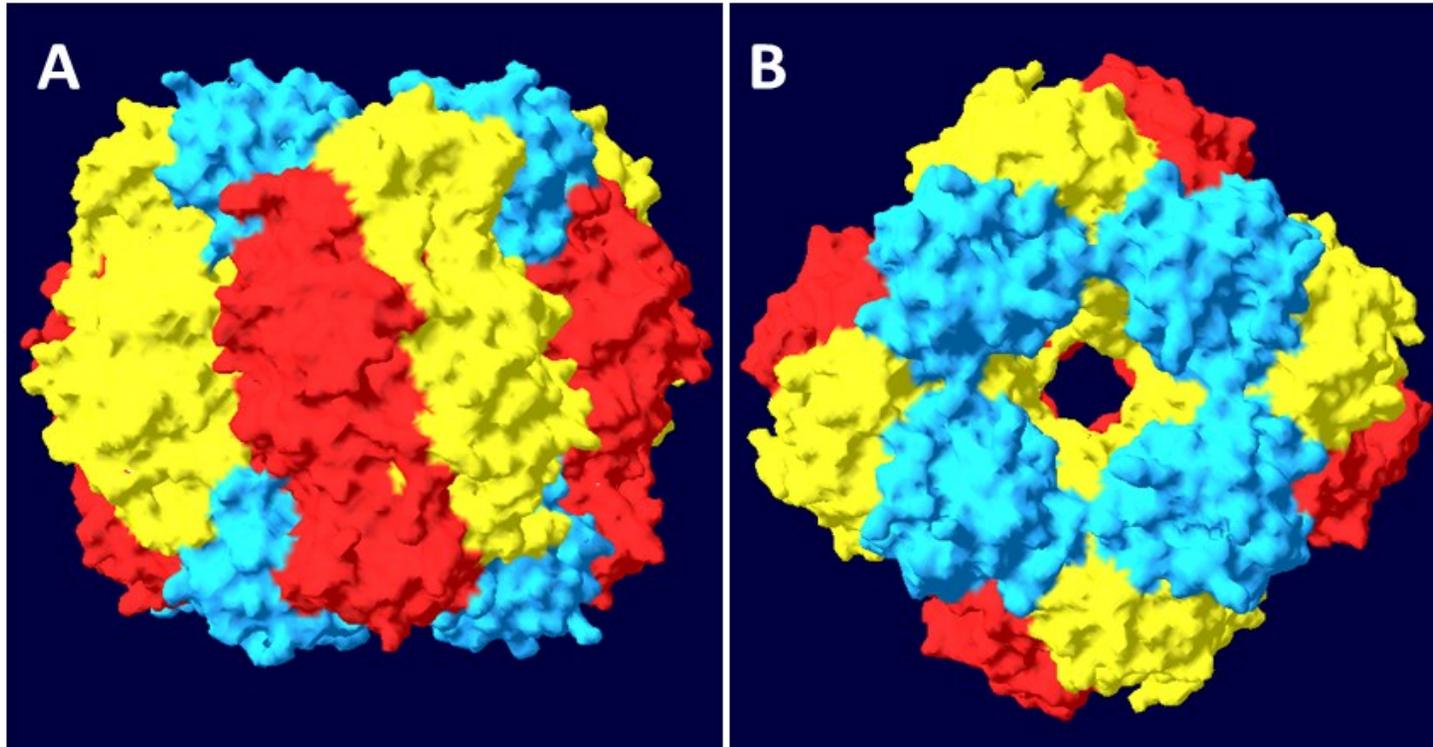
Figure 11 - Form I Holoenzyme

Surface model of 1RBL, form I RubisCO. **A**: Side view of L_8S_8 holenzyme. One red and one yellow large subunits make a catalytic dimer with two active sites. Small subunits are shown in blue. **B**: The view from the top of the barrel, along the two-fold axis of symmetry.
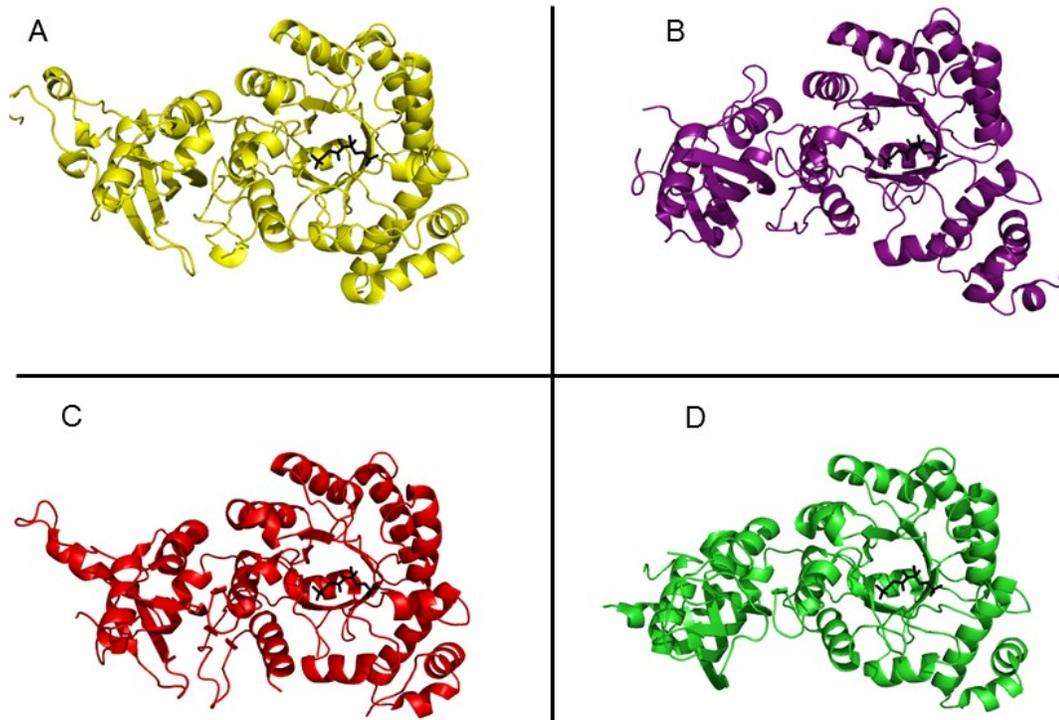
Figure 12 - Form I, II, III, IV Large Subunits

Single LSUs from the four major groups of RubisCO and RubisCO-like proteins. The approximate position of the substrate is shown as a black stick figure. A: Form I, *Spinacia oleracia* 1UPP; B: Form II *Rs. rubrum* 1RUB; C: Form III *Thermococcus kodakariensis*, 3A12; D: Form IV/RLP *Chlorobium tepidum*, 1YKW

and Hügler and Sievert (2011) for recent reviews of RubisCO distribution and Tabita *et al* (2008b) and Andersson and Backlund (2008) for recent reviews of structural diversity In Form I RuibsCO, the small subunits (SSUs) do not pair as the large subunits do, but rather associate with the crevices between the large subunit dimers and form four-unit rings at each pole of the central barrel of eight large subunits. All known form I enzymes occur as L_8S_8 (or more precisely as $(L_2)_4(S_4)_2$) where 4 pairs of large subunits form a barrel with a cap of 4 small subunits circling each end of the central pore. Form I can be

further divided in to the green group (land plants, cyanobacteria and some facultatively autotrophic bacteria) and the red group (purple bacteria and eukaryotic marine algae). These subgroups can be further divided, with the green group to forms I A and IB, and the red group in to forms IC and ID (Delwiche and Palmer 1996; Watson and Tabita 1996, 1997). Notably, RubisCO shows abundant evidence of lateral gene transfer, and the phylogeny of RubisCO proteins differs markedly from the 16S RNA-based trees used to infer overall relatedness (Delwiche and Palmer 1996).

Form II RubisCO is found exclusively in bacteria and in the plastids of dinoflagellates. In its simplest arrangement, as in *Rhodospirillum rubrum*, it occurs as just the dimer of large subunits (Tabita and McFadden 1974) although larger concatenations of dimers are found in the RubisCOs of other bacteria.

Form III RubisCO is similar to form II in that it occus as large subunits only. Some appear to be just L_2 (Finn and Tabita 2003) while others such as *Thermococcus kodakarensis* KOD1 RbcL have the form $(L_2)_5$ (Kitano *et al* 2001). All known form III RubisCOs are found in archaea. The converse, however, may not be true as some archaea appear to possess RubisCO sequences that are intermediate between forms II and III. The RubisCO of *Methanococcoides burtonii*, the first form II/III intermediate described, is discussed in some detail in Chapter 3.

Finally, the form IV RubisCO Like Proteins (RLPs) are also comprised.of large subunits only. As mentioned previously, the RLPs do not carry out the carboxylation or

oxygenation of RuBP, but in the examples studied do perform reactions with similar chemistry. Additionally, the RLPs have tertiary and quaternary structures that are very similar to the other forms of RubisCO. Only a small sample of RLP sequences were used in the construction of Figure 5. RLPs are distributed throughout the bacteria and archaea. See Figure 12 for an illustration comparing a single large subunit of each of the four families of RubisCO. As mentioned above, RubisCO, and by extension the CBB cycle, can tolerate molecular oxygen, the enzyme is still inhibited by it. From a physiological standpoint, the persistence of the oxygenation reaction despite billions of years of selective pressure presents a puzzle. The oxygenation reaction, after all, results in a net loss of fixed carbon, energy (via ATP consumption) and reducing power as 2-PG is recycled to cell material. In land plants and cyanobacteria, as much as 25% of cell carbon can be lost via photorespiration (Sharkey 1988). Some researchers speculate that the oxygenation reaction acts as a "dump" for excess reducing power that prevents damage to the metabolically costly photosynthetic apparatus when leaves are in a low-CO_2 environment or saturating light (Taiz and Zeiger 2010).

The problem with this argument is that RubisCO from chemoautotrophic organisms with no photosynthetic apparatus to protect still has robust oxygenation activity. It is, however, interesting to contrast the properties of RubisCO enzymes that operate in high-oxygen environments (e.g. C3 plants, cyanobacteria without carbon concentration mechanisms, Knallgass bacteria) with those that experience low oxygen or anoxic environments (C4 plants, non-oxygenic photosynthetic bacteria, cyanobacteria with CCMs, archaea).

Other suggestions for the persistence of the oxygenation reaction lie in its role in shunting carbon to necessary amino acid catabolism (Harley and Sharkey 1991) or in supporting nitrogen assimilation in plants (Bloom *et al* 2010; Rachmilevitch *et al* 2004). The actual mechanism connecting photorespiration to nitrogen assimilation remains obscure, but a strong correlation has been observed. These questions have not been addressed at all in prokaryotes, but presumably a diffusion-based means of nitrogen assimilation should have obviated the need for oxygenation in RubisCO.

And yet, oxygenation persists in all studied RubisCOs. The explanation remains elusive, although a precise vocabulary exists for describing the interaction between this complex enzyme and its substrates. Specificity is the most facile means of summarizing the relative preference of an enzyme for carboxylation versus oxygenation reactions. And, as might be expected of such a diverse enzyme, catalytic properties vary widely. See Table 1 - Diversity of RubisCO Kinetic Constants. The specificity factor, Ω, is defined as V_cK_o/V_oK_c where V is the velocity with respect to each partial reaction and K_c or K_o is the Michaelis-Menten constant for each substrate. There appears to have been significant selective pressure to reduce the oxygenation reaction (i.e. to increase Ω) as the highest specificities are found in RubisCOs from organisms which fix carbon in the presence of molecular oxygen without a carbon concentration mechanism (e.g. *Ralstonia eutropha*) Significantly lower specificities are seen in RubisCO from organisms with the ability to artificially concentrate CO_2 near the RubisCO (e.g *Synechococcus* WH8102), either via subcellular structures such as carboxysomes or by segregation of carbon fixation from

atmosphere by tissue specialization (C4 plants) or time (Crassulacean Acid Metabolism plants). The lowest specificities of all are seen in enzymes isolated from organisms that never experience molecular oxygen, such as methanogenic archaea (e.g. *Methanosarcina acetivorans* C2A).

Solid descriptions of the basic reaction mechanism have been proposed. We know with a fair degree of precision which portion of which residues contact which portions of the substrate and the order in which the reaction steps proceed (Andersson 2008; Cleland *et al* 1998; Gutteridge and Gatenby 1995; Schneider *et al* 1992).

The proposed mechanism, however, does not account for the differences in specificity seen among various enzymes. Much of this detailed knowledge has come from painstaking site directed mutagenesis and subsequent kinetic studies on purified protein. The limits of reside-by-residue studies, however, are reached when trying to explain the basis of oxygenation versus carboxylation in a particular enzyme. The residues that directly interact with the substrates are universally conserved throughout all the known "bona fide" RubisCOs and changing them results in a near-total loss of activity. Since the active site residues are so stringently conserved, the basis of specificity must lie in how those residues are positioned relative to each other (and the substrates) and how the substrates reach the active site. In other words, it seems that the entire structure of the enzyme affects specificity.

17

Table 1 - Diversity of RubisCO Kinetic Constants

RubisCO type *Organism*	$\Omega = \dfrac{V_{CO2}K_{O2}}{V_{O2}K_{CO2}}$	K_{CO2} (µM)
Type IA		
Rhodobacter capsulatus I	25	30
Hydrogenovibrio marinus Ia	25	ND
Chromatium vinosum	40	35
Thiobacillus denitrificans I	45	140
Vent symbiont	30	80
Type IB Cyanobacteria		
Synechococcus 6301	40	175
Anabaena 7120	35	150
Green algae		
Chlamydomonas reinhardtii	60	30
Plants – many species	80	10–30
Type IC Purple bacteria class		
Bradyrhizobium japonicum	75	65
Xanthobacter flavus	45	100
Rhodobacter sphaeroides I	60	25
Ralstonia eutropha	75	ND
Type ID Marine algae		
Cylindrotheca sp. strain N1	105	30
Olisthodiscus luteus	100	60
Porphyridium cruentum	130	20
Cylindrotheca fusiformis	110	35
Cyanidium caldarium	225	5
Galdieria partita	240	5
Type II - bacteria		
Rhodospirillum rubrum	15	100
Rhodobacter sphaeroides II	10	100
Type III - archaea		
Archaeoglobus fulgidus	4*	51*
Thermococcus kodakarensis	11	52

* 83°C; *A. ful* from (Kreel and Tabita 2007); *T. kod* from (Nishitani *et al* 2010);
All other numbers from the review in (Tabita 1999).

Once the basic mechanism of RubisCO was elucidated, the most interesting studies were those that showed an impact on specificity when mutations are made in residues distant from the active sites – or even in the non-catalytic small subunits of form I enzymes. Read and Tabita (1992b) for example, show that single residue substitutions (I87V, R88K, G91V, and F92L) in the small subunit of *Synechococcus* PCC6301 RubisCO can have a significant, negative, impact on carboxylation without affecting assembly or the holoenzyme structure.

An even more subtle effect was detected by Flachmann *et al* (1997). The P108L substitution in PCC6301 RubisCO resulted in a three-fold increase in the production of the "misfire" product xylulose 1,5-P (XuBP). XuBP is produced from RuBP when the enediol reaction intermediate fails to react with CO_2 or O_2 and is misprotonated at C-3.

These examples should serve to illustrate the profound effects that can be caused by small structural changes very far from the active site. The ultimate goal, aside from simple understanding of course, is to engineer a better enzyme. Better in this context means maintaining a relatively high rate of carboxylation (V_c) while increasing specificity. A number of recent papers (Gutteridge and Pierce 2006; Tcherkez *et al* 2006) have gone so far as to argue that a better RubisCO is not possible. The conclusions of those papers, however, are based on a very narrow sample set. Most of the data for kinetic constants was obtained from crude extracts of land plants. As Figure 5 demonstrates, however, the greatest diversity of RubisCO lies in the prokaryotic realm.

There are good data points for some prokaryotic RubisCOs, but very little of the sequence space has been explored in the laboratory. Further, the relatively recent explosion of whole genome and metagenome sequences has made clear that there is a great deal of unknown diversity in RubisCO sequences. A comprehensive understanding of this vital and perplexing enzyme would rely on a multi-component approach. Certainly, more detailed studies linking structure with kinetic properties are necessary.

In the near term, however, efforts to improve RubisCO must rely on the tools and knowledge at hand. Our understanding of the link between structure and function beyond the active site is still rudimentary, however. There is certainly more to be learned by measured investigations of RubisCO sequence using the traditional tools: site directed mutagenesis to test structure-function hypotheses and random mutagenesis to identify negative mutantions and suppressors thereof.

As previously mentioned, however, the residues that contact the substrates of RubisCO are universally conserved, even among the most disparate known examples. Site-directed and random mutagenesis studies have shown that changing any of these residues results in the abolition of catalysis (Parry *et al* 2003). If one were to attempt to improve RubisCO, then, it is clear that the focus must be on subtle alterations to the holoenzyme structure, rather than wholesale alteration of the active site. With the active site thoroughly explored, each of the remaining ~460 residues of a typical large subunit could be altered to one of 20 amino acid residues at each position, resulting in 20^{460} possible sequences – a number that is considerably larger than the 10^{80} atoms in the observable

universe (Anonymous, Wikipedia).

Sampling that sequence space is clearly impossible on human, or even graduate student, time scales. Nature, however, has been making mutations and testing the results for billions of years in trillions of organisms. While the available sequence space cannot yet (or ever) be fully tested, the range of sequences tested by natural selection is more vast by far than the range that can be explored by deliberate manipulation.

There are ways, however, to begin to sample some of the vast diversity of "wild" RubisCOs. Chapter 1 demonstrates how existing technology can be used to study RubisCO genes isolated from uncultured "wild" bacteria. Chapter 2 expands the "proof of concept" from Chapter 1 to show a new selection-based, sequence-independent means of capturing RubisCO genes from uncultured organisms. In Chapter 3 we take a more in-depth look at a particular RubisCO gene that serves as an example of the unexpected diversity that would be missed by a strictly sequence-based approach.

Chapter 1 – RubisCO from a BAC Library

Introduction

Looking to nature to catalog enzyme diversity is not a new concept. In the case of

RubisCO, however, this traditionally has been done via laborious culture-based studies

that focus on isolating new organisms or by sequence-based catalogs of diversity that rely

on PCR-derived libraries. Culture based studies can provide unparalleled information on

the context of a particular enzyme in the sense of learning whether the RubisCO in

question is the sole means of obtaining fixed organic carbon (obligate autotrophy) or

whether the RubisCO is used only when fixed carbon and a suitable electron acceptor are

absent (facultative autotrophy). Or, as with RubisCO from archaea, whether the role of

RubisCO is even to support autotrophy at all.

The main draw-back of culture-based approaches is primarily one of time – sampling

diverse environments, obtaining pure cultures of a particular organism, exploring a range

of temperatures, nutrient availabilities and a host of other culture conditions, determining

whether RubisCO is present in the genome and if so, under what conditions it is

expressed. Add to this the difficulty of growing many environmental isolates under

laboratory conditions, and it quickly becomes apparent that any survey of RubisCO

functional diversity would be impractical.

The alternative, sequence-based approach can leverage the power of PCR to amplify segments of RubisCO genes from metagenomic samples (Alfreider *et al* 2009; Nigro and King 2007; Swan *et al* 2011; Wawrik *et al* 2002). This approach relies on the observation that, like many large proteins, different regions of RubisCO are conserved to varying degrees over evolutionary time. The more highly conserved regions have been used to develop PCR primers with minimal degeneracy that can be used to amplify portions of the coding region of the large subunit of RubisCO from form I or form II enzymes (Alfreider *et al* 2003; Alfreider *et al* 2009). The advantage of this approach is the broad net it casts and the rapidity with which one can assess some of the diversity of a sample of environmentally-derived DNA. The primary drawback of this approach, however, is that even when the PCR reaction is successful, only a portion of the large subunit can be sequenced. Even for form II enzymes where no small subunit is required the 600 bases obtained in a conserved-site PCR reaction is much too little for obtaining a functional enzyme.

Previous studies (Horken and Tabita 1999a) have shown that hybrids of large subunits rarely assemble at all, let alone produce an enzyme capable of carrying out the RubisCO reaction(s). Further, a fragment from the interior of an ORF reveals very little of the context of the gene. Very closely related RubisCO sequences can come from otherwise very distantly related organisms (e.g. *Synechococcus* PCC6301 RbcL is 72% identical to *Rhodobacter capsulatus* SB1003 CbbL (Altschul *et al* 1997)).

One solution to these conundra is to combine several well-understood techniques to produce novel results. In particular, a library derived from a sample of environmentally-derived genomic DNA (a metagenomic library) can be screened using sequence-based techniques (PCR or Southern-blot hybridization) and the clones possessing full-length RubisCO genes can be studied in greater detail. Obtaining functional information about these wild RubisCOs requires an additional step, however.

This chapter details a means of examining the vast untapped functional diversity of RubisCO from uncultivated organisms by adapting an extant gene expression system to a new purpose. A RubisCO-deletion strain of *Rb. capsulatus* (designated strain SBI/II-) was developed as a means of selecting for positive or negative mutations in prokaryotic RubisCO genes via the ability or inability to restore autotrophic growth in strain SBI/II- (Paoli *et al* 1998). *Rb. capsulatus* is a metabolically diverse α-proteobacterium capable of non-oxygenic photoautotrophic growth in a reducing atmosphere, as well as aerobic chemoautotrophic growth with energy obtained via the Knallgas reaction. Strain SBI/II-, in which both form I and II RubisCO genes were deleted and replaced by antibiotic resistance cassettes, is incapable of autotrophic growth, while still able to grow photoheterotrophically and chemoheterotrophically in complex media. Strain SBI/II- can grow autotrophically only with vector-encoded RubisCO genes.

There are several advantages to using this expression system. The first is that growth of the host organism demonstrates very clearly whether a recombinant gene encodes for a

functional RubisCO. Assay-based screens of multiple clones for RubisCO activity are a non-trivial undertaking that requires considerable time, consumables and potential exposure of the researcher to radioactive $^{14}CO_2$.

Secondly, *Rhodobacter* has been been shown to support high levels of expression of functional RubisCO under modes of growth, including those that exclude or include oxygen. One of the most common problems encountered with *Escherichia coli*-based expression systems is that a protein may be produced, yet may accumulate in insoluble inclusion bodies. RubisCO, in particular, has at times proven to be difficult to refold into a functional conformation (Goloubinoff *et al* 1989; Lee and Tabita 1990; Li and Tabita 1997; Lorimer 1996). *Rhodobacter*, by contrast, expresses high levels of chaperones homologous to GroEL/ES under autotrophic growth conditions (Lee *et al* 1997; Lee *et al* 1998). In the current study, it is demonstrated that fully functional RubisCO can be produced from ORFs cloned from environmentally-derived samples from uncultured organisms, and that these genes will complement photoautotrophic growth of the RubisCO-deletion strain of *Rb. capsulatus*.

Previous studies have demonstrated the possibility of obtaining functional enzymes from metagenomic libraries (Andexer *et al* 2006; Beloqui *et al* 2006; Coque *et al* 2002; Diaz-Torres *et al* 2003; Findley *et al* 2011; Findley *et al* 2011; Gao *et al* 2011; Hu *et al* 2011; LeCleir *et al* 2004; Lim *et al* 2005; Sim *et al* 2011; Uyaguari *et al* 2011; Voget *et al* 2006; Yun *et al* 2004) using *E. coli* as the expression host. None of these studies, however, focused on enzymes essential for carbon fixation or of proteins that play such a key role

25

in determining biogeochemical cycles. Moreover, this robust system provides high levels

of the environmentally critical RubisCO enzyme such that facile column

chromatography-based purification of tagged or untagged proteins may be employed to

obtain enough purified enzyme suitable for basic enzyme kinetic measurements. These

studies open an intriguing new method for understanding the functional properties of key

catalysts of global carbon cycles obtained from organisms that have never been in

cultivation, thus dispensing with obstacles inherent in culture-based methods.

Materials and Methods

Bacterial strains, plasmids and growth conditions. All cloning steps were performed in

E. coli Top10 strain (F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG

recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$) (Invitrogen). *E.

coli* was grown in Luria-Bertani (LB) broth containing (per liter of water) 5 g yeast

extract, 10 g tryptone, 10 g NaCl, at 37 ^{o}C and near-neutral pH. LB plates used 1.5%

agar. When required for plasmid maintenance, antibiotics were used at appropriate

concentrations (12.5 μg/ml tetracycline, 100 μg/ml ampicillin, 25 μg/ml spectinomycin,

25 μg/ml streptomycin, 50 μg/ml kanamycin, 12.5 μg/ml chloramphenicol). *Rb.

capsulatus* SBI/II- was grown aerobically at 30 ^{o}C in peptone-yeast extract (PYE) broth

or on 1.5% PYE agar plates. PYE contained (per liter of water) 3 g peptone, 3 g yeast

extract and Ormerod's trace elements (Ormerod 1961) supplemented with 0.1 mg biotin

and 0.1 mg riboflavin. Antibiotics for PYE media and plates were used at the following

concentrations: 2 μg/ml tetracycline, 10 μg/ml spectinomycin, 5 μg/ml kanamycin, 100 μg/ml rifampicin.

Conjugation required helper strain of *E. coli* HB101 (Δ(gpt-proA)62 leuB6 thi-1 lacY1 hsdS$_B$20 recA rpsL20 (Str^R) ara-14 galK2 xyl-5 mtl-1 supE44 mcrB$_B$) (Boyer and Roulland-Dussoix 1969) containing the plasmid pRK2013 (Kahn *et al* 1979) which provides the transfer genes. Conjugation was performed according standard protocol (Smith and Tabita 2003). Briefly, a single colony of I/II- was used to inoculate liquid PYE and grown for 36-48 hours. Overnight cultures of the donor strain and the helper strain were grown in LB such that they were mature at the same time as the I/II- culture. A 1 ml aliquot of I/II- and 0.5 ml of both donor and helper strain were harvested by centrifugation and washed separately in PYE. Finally, the washed pellets were resuspended in a total of 50 μl of PYE and the mixture of strains was deposited on to a prewarmed PYE plate for incubation at 30 ^{o}C for 24 hours. The mating mixture was then transferred to selective media (see below).

Autotrophic *Rhodobacter* cultures were grown in Ormerod's minimal salts media (OM) (Ormerod *et al* 1961). Liquid cultures used 500 ml flat sided glass bottles. Gasses were injected through glass tubes that passed through a butyl rubber stopper in the mouth of the bottles. Photoautotrophic cultures were bubbled with premixed gas (5% or 1.5% CO_2, balance H_2) and the bottles were incubated at 30 ^{o}C with a bank of incandescent lights for illumination. Minimal media plates contained 1.5% Noble agar (USB) and were incubated in air tight jars containing a CO_2/H_2 generating system (5-6% CO_2, BBL

27

GasPak System, Becton Dickson Microbiology Systems). The jars contained a palladium catalyst to ensure complete removal of molecular oxygen. The jars were partially submerged in a 30 °C water bath in front of a bank of incandescent lights.

Exogenous genes were expressed in *Rb. capsulatus* SBI/II- using the broad host range plasmid pRPS-MCS3, a pBBR1-derived vector (Smith and Tabita 2003). The pRPS-MCS3 vector expresses the *Rs. rubrum cbbR* transcriptional regulator. The CbbR protein binds to the *cbb* operon promoter that controls expression of ORFs in the vector MCS such that the coding region is maximally transcribed under autotrophic conditions and repressed under chemoheterotrophic growth conditions. (See Figure 18, p. 52, for a diagram of pRPS-MCS3.)

Chemicals – Ribulose 1,5-bisphosphate was obtained from Sigma-Aldrich (St. Louis, MO). A 10 mM solution was stored at -80 °C until needed for assays. Tritiated RuBP ([1-^{3}H]RuBP) had been prepared by Dr. Sriram Satagopan enzymatically as described in (Horken 1998), using [2-^{3}H]glucose as the starting material. NaH[^{14}C]O$_3$ (40-60 mCi/mol) was obtained from Perkin-Elmer (Waltham, MA) or ViTrax (Placentia, CA) 40-60 mCi/mmol. All other chemicals were obtained from Sigma-Aldrich or Fisher unless otherwise noted.

BAC Library Methods - Sample collection and cloning. Sampling for the 4N23 BAC clone sequence was previously described (John *et al* 2006). Briefly, 240 l seawater was collected by rosette sampler from a depth of 40 m at an oligotrophic station in the Gulf of

Mexico. Station coordinates were Lat. 25.2630, Long. -84.2202. For the clone B15

sequence, surface seawater (100 l) was collected from the Long-Term Ecosystem

Observatory site (LEO-15), near the Rutgers University Marine Field Station, Tuckerton,

New Jersey (Lat. 39.4640, Long. -74.2600) from a depth of 1 m. For both libraries,

methods employed were as previously described (John *et al* 2006). Plankton cells were

concentrated by vortex-flow filtration (Membrex) with a 100 kDa membrane filter, and

then further concentrated by centrifugation at $18,500 \times g$ for 15 min. Cell pellets were

suspended in 1% molten SeaPlaque LMP agar. Cells in agarose were solidified into 75-

µl plugs, then subject to chemical lysis (via sarcosyl, sodium deoxycholate, and

lysozyme) and proteinase K digestion. DNA in the agarose plugs was partially digested

using HindIII and size-separated using pulsed-field gel electrophoresis. The gel areas

corresponding to the 100-125 kb size range were excised and DNA was electroeluted

with dialysis tubing at 3 V cm^{-1} for 3 h, at which time the current was reversed for 30 sec

to facilitate DNA removal from the dialysis tubing. The sample bag was then dialyzed

twice for 1 h in deionized water at 4°C. Cloning was performed by ligation of 40 ng DNA

into 50 ng HindIII-digested pIndigoBAC cloning vector (Epicentre) according to the

manufacturer's recommendations, followed by transformation via electroporation into 20

µl Transformax EPI300 electrocompetent *E. coli* cells (Epicentre) using a Bio-Rad

MicroPulser. Transformants were stored in 25% glycerol at -80°C.

BAC Library Methods - Clone screening and selection. Transformants were spread and

grown on LB plates containing chloramphenicol, and colonies were robotically picked

into 384-well plates. Cells from plates were robotically arrayed onto positively-charged

nylon membranes (Roche, Indianapolis, USA) via the BioGrid array system
(BioRobotics, Cambridge, UK) to enable screening for target genes by probe
hybridization. After arraying, the membrane was grown on LB agar with
chloramphenicol at 37 ^{o}C overnight. DNA was immobilized on membranes as follows:
cells were lysed via 20% sodium dodecyl sulfate in water, wicked from below by
chromatography paper (3 mm); membranes were microwaved approximately 2 min until
dry and cell proteins were removed by proteinase K digestion (10 µg ml^{-1}). Membranes
were then rinsed with 2X salt sodium citrate and subjected to UV crosslinking to
immobilize DNA to the membrane. Probes to select clones containing *rbcL/cbbL*
(RubisCO large subunit gene) were created as previously described (Paul *et al* 1999). In
short, RNA probes were generated from form IA and form ID *rbcL/cbbL* fragments by *in
vitro* transcription (Riboprobe Combination System SP6/T7, Promega) according to the
manufacturer's instructions. Probes were labeled with ^{35}S-UTP (Amersham Bio-
sciences/GE Healthcare). Hybridization and washing of membranes was also as
previously described (Paul *et al* 1999). Probed membranes were screened on a Bio-Rad
PMI Molecular Imager and images analyzed to determine which plates and wells
contained clones of interest. Recovered clones were further verified to contain *rbcL*
genes by PCR with primers specific to either form IA or ID *rbcL*. Primers used for
screening were form ID primers GATGATGARAAYATTAACTC forward, ATTT
GDCCACAGTGDATACCA reverse; form IA primers CTGAGIGGIAARAACTACGG
forward, GGCATRTGCC-ANACRTGRAT reverse. Selected clones were sequenced at
the U.S. Department of Energy Joint Genome Institute (Walnut Creek, CA) or the Broad
Institute at Massachusetts Institute of Technology.

Genetic Techniques. Other than the original BACs, all DNA sequencing was performed by the Plant-Microbe Genomics Facility (PMGF) at The Ohio State University. Plasmids were isolated from *E. coli* and *Rb. capsulatus* using 3 ml of culture and a standard miniprep kit (Qiagen). BACs were isolated using either alkaline lysis/ethanol
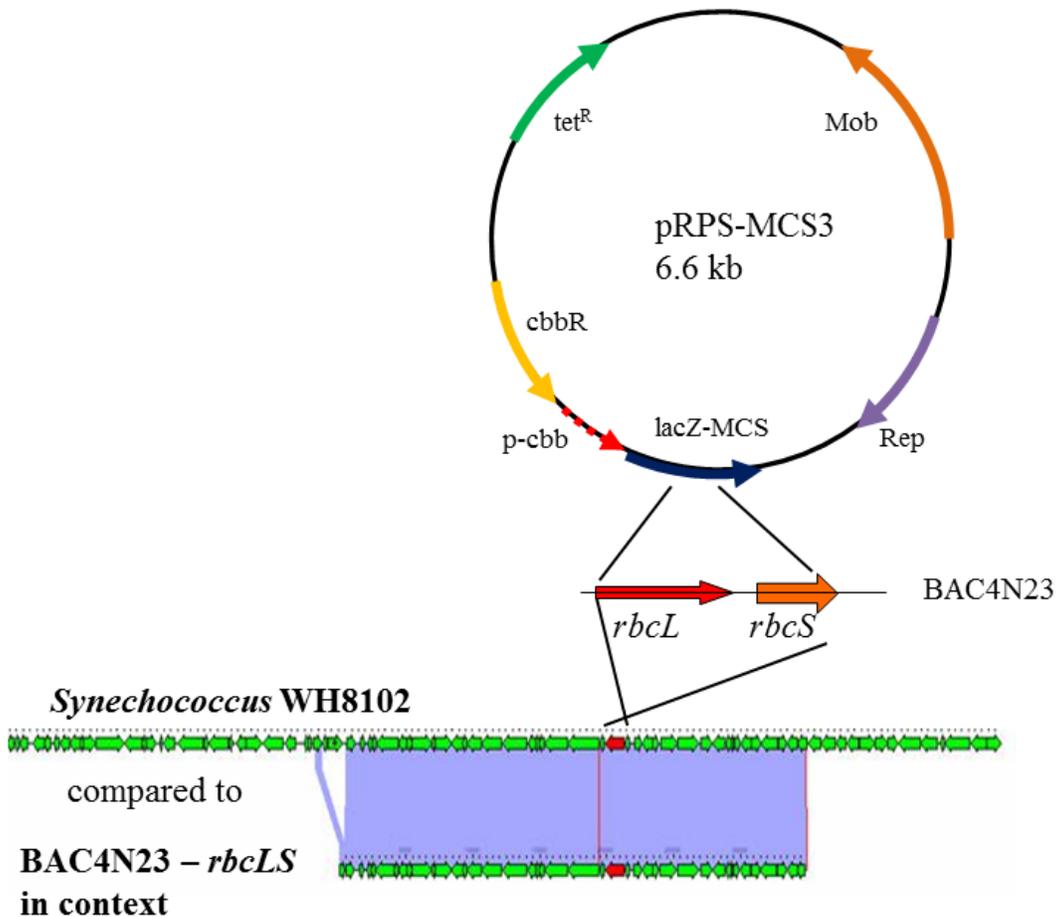
Figure 13 - Schematic of Metagenomic Clone

The high degree of synteny with *Synechococcus* WH8102 is illustrated. Only one ORF is transposed. The *rbcLS* ORFs were cloned from the BAC using specific PCR primers and the *rbcLS* product was inserted in to pRPS-MCS3 multiple cloning site.

precipitation (Sambrook and Russell 2001) or a Spin Doctor BAC Prep kit (Gerard

Biotech, Oxford, Ohio). Gel purification of DNA fragments was performed with a

standard kit (Qiagen). The *rbcLS* ORF from BAC4N23 was amplified using Pfu high

fidelity polymerase (Stratagene). The forward primer (GACTGGGCCCTTCACCGA-

CCTAACGG) added a 5' ApaI restriction site. The reverse primer (TCTAGACCAAT-

GGTTCGAAGGATCAGCGTCC) incorporated the endogenous stop codon of *rbcS* and a

3' XbaI restriction site. The 1.8 kb product was cloned using the Topo-Blunt kit

(Invitrogen) and sequenced using the M13 forward and reverse primers flanking the

cloning site. The plasmid was digested with ApaI and XbaI and the 1.8 kb fragment was

purified from an agarose gel. Separately, pRPSMCS3 was digested with ApaI and XbaI

and gel purified. Plasmid pRPS-4N23 was then constructed using aliquots of both

fragments and T4 DNA ligase (Invitrogen). The pRPS-4N23 plasmid was subsequently

repurified and sequenced using custom primers RPSF (AGTGAGCGCGCGTAATA-

CGAC) and RPSR (GGTCGACGGTATCGATAAGCTTG) that flanked the cloning site

in pRPS-MCS3. Also, an internal sequencing primer 4N23I (GGGGTGGCCGATTGC-

ATCAACCGG) was used. The *cbbLS* ORF from BAC15 was cloned using identical

techniques and the resulting plasmid was named pRPS-B15. Forward primer:

ACTAGTATAGTACCCATTGTCCCTCGACAC (contains 5' SpeI site). Reverse:

GAGCTCTTAACGACCCTCGTGTACTACGAAG (contains 3' SacI site). The complete

sequences of BAC4N23 and BAC15 were deposited in NCBI as Accession #s DQ325541

and EU795144, respectively.

Complementation Studies. The plasmids pRPS-4N23 and pRPS-B15 were transferred to

strain SBI/II- via conjugation. A starter culture of 20 ml was inoculated with a single

colony in PYE broth containing kanamycin, spectinomycin, and tetracycline. After 48 h,

the stationary-phase culture ($OD_{600} \sim 2.0$) was harvested by centrifugation in sterile

bottles and washed twice with sterile OM. The pellet was resuspended in 1 ml OM and

250 μl was used to inoculate 350 ml OM for photoautotrophic growth. Bottles were

maintained at 30 ^{o}C in an illuminated water bath and continuously bubbled with either

5% CO_2/balance H_2 or 1.5% CO_2/balance H_2. A 2 ml sample was withdrawn aseptically

once per day from each bottle using a custom-built sampling port and A_{600} was measured.

SBI/II- complemented with the *rbcLS* genes from *Synechococcus* PCC6301 served as a

positive growth control (plasmid pRPS-6301). The construction of pRPS-6301 has been

previously described (Smith and Tabita 2003). Bottles were inoculated in triplicate and

all bottles shared the same water bath, bank of lights and tank of gas.

Protein Purification. Recombinant RubisCO was purified from photoautotrophically-

grown strain SBI/II- harvested by centrifugation at late log/early stationary phase (OD_{660}

1-2). Cells were pelleted at 4 ^{o}C at 10,000 x g and washed twice in ice-cold TEMDB (50

mM Tris-HCl pH 8.0, 1 mM EDTA, 10 mM $MgCl_2$, 1 mM DTT, 10 mM $NaHCO_3$). Cell

pellets were frozen at -80 ^{o}C. Later, frozen pellets were thawed on ice, resuspended in a

small volume of TEMDB and lysed via French press (three passes at 14,000 psi). Small

aliquots of each bottle of culture were assayed for RubisCO activity. Further, plasmids

were isolated from each bottle of culture and the insert sequenced with 2X coverage.

Purification of recombinant protein followed previously-published protocols (Read and

Tabita 1992a; Smith and Tabita 2003). After lysis strain SBI/II- containing the BAC15

RubisCO genes, RubisCO activity was confirmed using standard assay procedures (Smith and Tabita 2003). The lysate was then centrifuged for 10 min at 14,000 x g at 4 $^{\circ}$C. The supernatant was transferred to a new tube and centrifuged at 45,000 x g at 4 $^{\circ}$C for 1 h (Beckman J2-21 Centrifuge). The supernatant was filtered through a 0.2 μm pore size nylon syringe filter (Fisher). Subsequent column chromatography steps were performed at 4 $^{\circ}$C using a Biorad Duo Flow Workstation.

The first chromatography step employed a HI-PREP DEAE-Sepharose FF column (Amersham/GE Healthcare) and fractions were eluted with a 10 mM to 500 mM NaCl gradient over 20 column volumes. The highest activity fractions were pooled and RubisCO was precipitated at 70% ammonium sulfate saturation on ice. The precipitate was collected via centrifugation at 14,000 x g. The pellet was gently resuspended in TEMDB and the solution was desalted using a disposable 10DG column (Bio-Rad). Activity was verified once again, and the desalted eluate was loaded onto the top of a TEMDB-sucrose step gradient (0.2, 0.4, 0.6, 0.8, 1.0 M sucrose dissolved in TEMDB) and centrifuged in a swinging bucket rotor for 22 h at 25,000 x g at 4 $^{\circ}$C. 1 ml fractions were removed and assayed for activity. The highest-activity fractions were pooled and the protein was dialyzed against a low-ionic-strength buffer (10 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl pH 8.0, 1 mM $MgCl_2$, 1 mM $NaHCO_3$, 1 mM DTT). Dialysis was performed overnight at 4 $^{\circ}$C using 25,000 MW-cutoff cellulose (SpectraPor). The final step of purification used a UnoQ column (Bio-Rad) with a 10 to 500 mM NaCl gradient. Purified protein was dialyzed into TEMDB. Glycerol was added to 20% final concentration before the protein was stored at -80 $^{\circ}$C.

Strain SBI/II- complemented with *rbcLS* sequences from BAC 4N23 was lysed similarly, but only the lysate centrifugation, DEAE-Sepharose and sucrose gradient ultracentrifugation steps were used due to the extreme lability of this enzyme (see discussion, below). Purified protein was flash frozen in liquid nitrogen and stored at -80 ^{o}C directly in the sucrose solution.

Enzyme Assays. RubisCO specific activity was determined using standard $^{14}CO_2$-based techniques (Smith and Tabita 2004) Briefly, a small (<20 μl) volume of protein or cell lysate was added to 100 μl buffer A (50 mM 2-(Bis(2-hydroxyethyl)amino)acetic acid/NaOH (Bicine) pH 8.0, 10 mM $MgCl_2$) in a 10 mM borosilicate glass tube. Specific activity was determined with at least two concentrations of protein or lysate to ensure protein-dependence of the activity detected. Assay tubes were prepared in triplicate for each protein concentration. The tube with buffer A and protein was kept on ice until the start of the assay. Immediately prior to the beginning of the assay powdered sodium bicarbonate was added to ice cold buffer B (50 mM Bicine pH 8.0, 10 mM $MgCl_2$) to a final concentration of 50 mM. The "hot B" buffer was made by aliquotting a volume of buffer B equal to (# assay tubes + 1) x 100 μl. Then 10 μl of hot sodium bicarbonate was added per ml buffer B. 10 μl of hot B buffer was added to cold B buffer and 100 μl of that mix was counted in 3 ml of scintillation fluid in order to determine the counts per μmol CO_2. The assay tubes were transferred to a heating block prewarmed to 30 ^{o}C and allowed to incubate for 2 minutes before 100 μl of hot B was added per assay tube. After 5 minutes, 20 μl of 10 mM RuBP was added to each assay at precise intervals. Two tubes

containing protein or cell lysate received hot B but not RuBP and served as negative

controls to determine background counts. 100 µl of propionic acid was used to stop each

assay 5 minutes after the addition of RuBP. The assay tubes were then briefly mixed and

placed in a bench top centrifuge to spin at 5000 x g for 30-60 minutes. After

centrifugation 200 µl of each assay was added to 3 ml of scintillation cocktail, mixed and

read in a Beckman LS-5000TD scintillation counter. Protein concentration was

determined with a Bradford assay (Bio-Rad) with a bovine serum albumin standard. The

protein concentration standard curve was determined separately each time the assay was

performed. Additionally, the curve was determined in the presence of a buffer identical

to that of the unknown sample.

CO_2/O_2 substrate specificity assays were also performed according to standard protocols

(Harpel *et al* 1993; Smith and Tabita 2003; Smith and Tabita 2004) See Figure 14 for an

illustration of the mechanism. Each assay was performed in triplicate. *Synechococcus*

PCC6301 RubisCO purified for this study served as the positive control to ensure proper

technique. Briefly, assays were conducted in 1 ml Wheaton vials capped with butyl

rubber stoppers and crimped metal seals. For each assay ten units of enzyme were added

to sufficient volume of buffer A (50 mM Bicine-NaOH, pH 8.3, 10 mM $MgCl_2$) to make

a total volume of 360 µl. Each vial was kept on ice while being flushed with UHP-grade

O_2 for 15 minutes. Just prior to transfer to a 30 ^{o}C water bath, each assay was injected

with 20 µl 100 mM $NaHCO_3$. The vials were then transferred to a prewarmed 30 ^{o}C

water bath and allowed to incubate for 5 minutes. 20 µl of 130 mM ^{3}H-RuBP was added

to each vial (for a final concentration of 6.55 mM) using a gas-tight Hamilton syringe.

Each assay was allowed to proceed for 60 minutes before being terminated with 20 µl of

40 mM NaBH$_4$. After a 15 minute incubation, 20 µl of 160 glucose was added to

consume the residual NaBH$_4$. Each vial was then uncapped and 1 ml of water was added

to each assay before the contents of the vials were separately centrifuged through a

10,000 MW Amicon filter in order to remove the protein. The filtrate was stored at -20

ºC until immediately before being read.
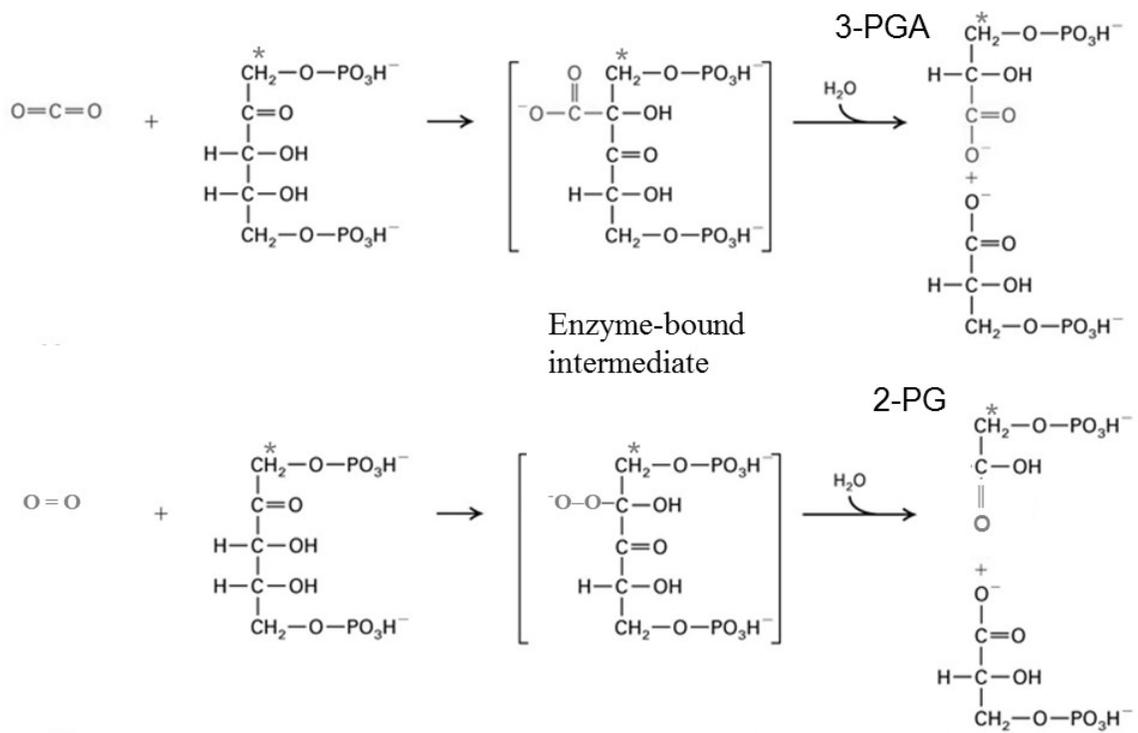
Figure 14 - Specificity Assay

* = ^{3}H label.

When RuBP is labeled with ^{3}H in the C-1 position, the carboxylation reaction of RubisCO produces a
^{3}H-labeled molecule of 3PGA (as well as one unlabeled 3PGA). The oxygenation reaction produces a
^{3}H-labeled molecule of 2PG as well as one unlabeled 3PGA. The two labeled products can be
separated chromatographically and the peaks detected by an in-line scintillation counter.

The ratio of 3-PGA to 2-PG was determined with a 1 ml BioRad Uni-Q1 anion exchange column attached to a Dionex DX500 HPLC. 1 ml of sample was loaded on to the column in 10 mM $NaBO_4$ pH 8.0 and eluted with a gradient of NH_4Cl in 10 mM $NaBO_4$ pH 8.0. An IN/US β-RAM with a 500 µl flow-through scintillation cell was used in real time to monitor the column eluate. The eluate was mixed 1:5 with Eco-Scint scintillation fluid (National Diagnostics).

Results and Discussion

BAC sequences and their genomic context. Bacterial Artificial Chromosome 4N23, and the library from whence it originated, have been previously described (John *et al* 2007) Briefly, clone 4N23 contains a 19.6 kB fragment of metagenomic DNA from oligotrophic waters of the Gulf of Mexico with 68% nucleotide identity to *α- Synechococcus* sp. WH8102 (John *et al* 2006). The ORF encoding the large subunit of RubisCO, *rbcL*, is 96% identical at the nucleotide level (1366/1418 nucleotides) to *Synechococcus* WH8102 *rbcL* and ~99% identical at the amino acid level (466/471 residues). Further, a high level of synteny was shown with strain WH8102 in regions immediately upstream and downstream of the 4N23 *rbcL*, including a putative seven-ORF operon encoding proteins of an α-carboxysome carbon concentrating mechanism (CCM) (e.g., *ccmK1* immediately upstream of *rbcL*, and genes that encode carboxysome peptides B, A, *csoS3, csoS2* and *csoS1/ccmK2* downstream of *rbcS*.). Figure 15 presents a schematic diagram of metagenomic ORFs from BAC clones 4N23 and B15.

By blastp (Altschul *et al* 1997), the RbcL peptide sequence derived from BAC clone B15 possessed a high degree of similarity to the RubisCO from *Synechococcus* sp. RS9917, *Prochlorococcus marinus* str. CCMP1375 and the chemoautotroph *Nitrococcus mobilis* (88%, 86% and 86% identity, respectively). However, in the case of clone B15, the immediate genomic context of the ORF coding for RubisCO provides more clues as to the original source of this gene. ORFs coding for carboxysome components appear immediately upstream and downstream of the BAC15 *rbcLS* genes, and the sequence of these genes indicates that they are most similar to the carboxysome proteins of *Nitrococcus*. Further up and downstream of the *rbcLS* genes, however, there is little direct similarity to *Nitrococcus*, indicating that the original organism is one that has not been previously examined. (See Appendix B, Figure 34, p. 146, for a chart of the closest blastp matches.)

Growth studies. In this study, a major objective was to determine if the *Rb. capsulatus* strain SBI/II- selection system could be utilized to detect and isolate functional RubisCO molecules from uncultured environmental samples. This was accomplished by determining whether metagenomic DNA, when cloned into compatible vectors, could complement photoautotrophic (CO_2-dependent) growth of *Rb. capsulatus* strain SBI/II-. RubisCO genes from clones B15 and 4N23, when inserted into vector pRPS, both complemented strain SBI/II- and allowed the host strain to grow as well or, in some cases, even better than with the *Synechococcus* 6301 *rbcLS* genes that are known to provide functional RubisCO (Smith and Tabita 2003) (Figure 16). Notably, *Rb. capsulatus* strain SBI/II- complemented with clone 4N23 exhibited no growth defect

relative to the other strains, despite the apparent lability of the 4N23 RubisCO in cell lysates (see Enzyme Purification discussion below). If there were a significantly higher turnover rate due to instability and/or rapid degradation of the 4N23 RubisCO, we would expect to see a metabolic cost reflected in increased generation times. It thus seems likely that intracellular factors were able to maintain structural stability of the enzyme *in vivo*. selection conditions were relatively generous in that no molecular oxygen was present to compete with CO_2 as a substrate for RubisCO.

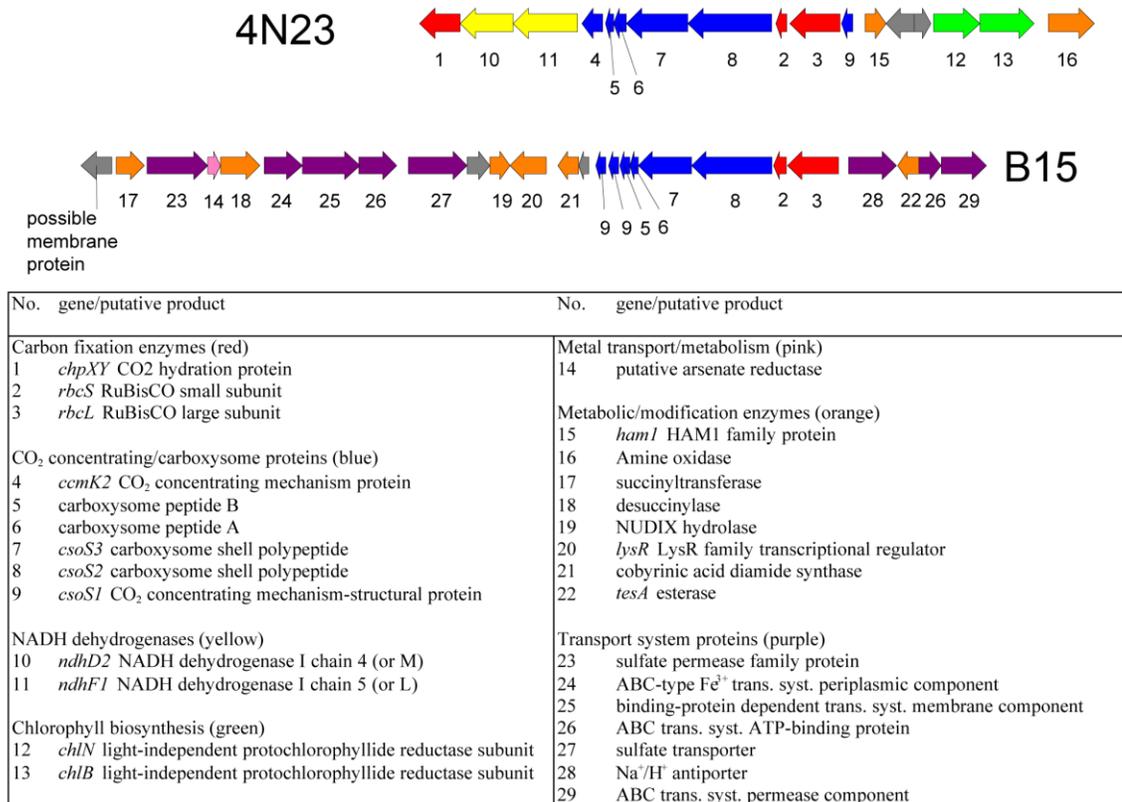| No.	gene/putative product	No.	gene/putative product
	Carbon fixation enzymes (red)		Metal transport/metabolism (pink)
1	*chpXY* CO2 hydration protein	14	putative arsenate reductase
2	*rbcS* RuBisCO small subunit		
3	*rbcL* RuBisCO large subunit		Metabolic/modification enzymes (orange)
		15	*ham1* HAM1 family protein
	CO_2 concentrating/carboxysome proteins (blue)	16	Amine oxidase
4	*ccmK2* CO_2 concentrating mechanism protein	17	succinyltransferase
5	carboxysome peptide B	18	desuccinylase
6	carboxysome peptide A	19	NUDIX hydrolase
7	*csoS3* carboxysome shell polypeptide	20	*lysR* LysR family transcriptional regulator
8	*csoS2* carboxysome shell polypeptide	21	cobyrinic acid diamide synthase
9	*csoS1* CO_2 concentrating mechanism-structural protein	22	*tesA* esterase
	NADH dehydrogenases (yellow)		Transport system proteins (purple)
10	*ndhD2* NADH dehydrogenase I chain 4 (or M)	23	sulfate permease family protein
11	*ndhF1* NADH dehydrogenase I chain 5 (or L)	24	ABC-type Fe^{3+} trans. syst. periplasmic component
		25	binding-protein dependent trans. syst. membrane component
	Chlorophyll biosynthesis (green)	26	ABC trans. syst. ATP-binding protein
12	*chlN* light-independent protochlorophyllide reductase subunit	27	sulfate transporter
13	*chlB* light-independent protochlorophyllide reductase subunit	28	Na^+/H^+ antiporter
		29	ABC trans. syst. permease component

Figure 15 - BAC15 and 4N23 Context

Context of the RubisCO ORFs from metagenomic BACs. Synteny of ORFs on two metagenome-derived BACs. ORFs are color-coded based on putative function. Gray shaded ORFs are of unknown function. See Figure 34, p. 146, for a chart of ORF similarities by blastp
.

The In future investigations it should be possible to vary the growth conditions in order to select for enzymes with desirable enzymatic properties (e.g. high substrate specificity factor and high specific activity). *Rb. capsulatus*, as mentioned in the Introduction, is capable of aerobic chemoautotrophic growth (in an atmosphere of $O_2/H_2/CO_2$).

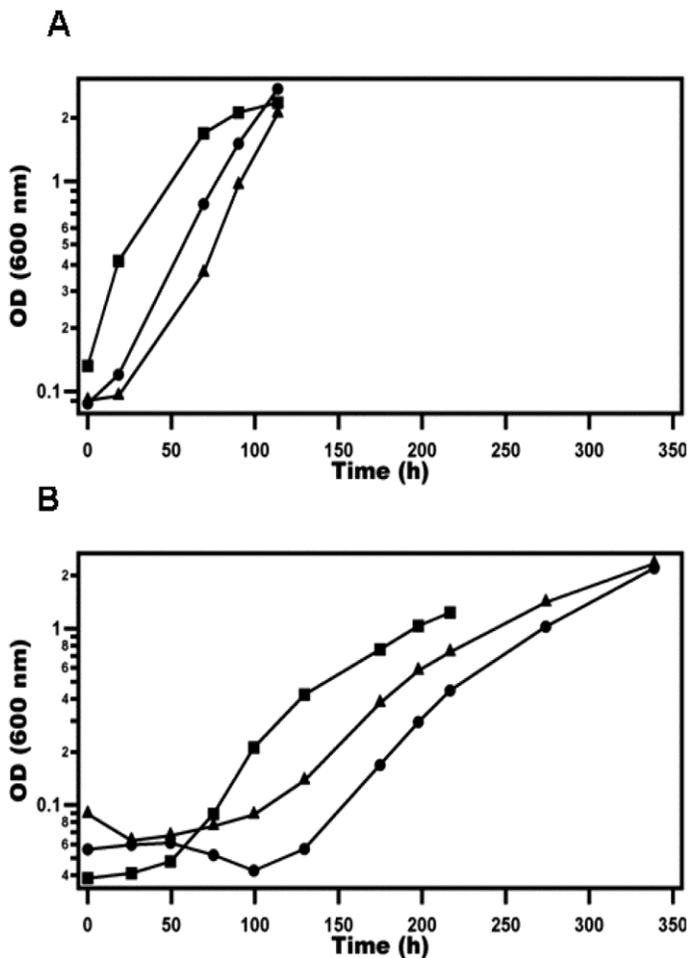**Figure 16 - Compelementation of *Rb. capsulatus* I/II- with Metagenomic RubisCO**

Photoautotrophic growth of *Rb. capsulatus* strain SBI/II- complemented with plasmids pRPS-4N23 (▲), pRPS-B15 (■) and pRPS-6301 (●) (positive control containing the *Synechococcus* sp. strain PCC 6301 *rbcLS* genes). (A) Minimal salts media bubbled with 5% CO_2/balance H_2; (B) minimal salts media bubbled with 1.5% CO_2/balance H_2.
.

41

Because *Rb. capsulatus* has no discernible CCM, growth under aerobic chemoautotrophic conditions is partly indicative of the RubisCO enzyme's ability to discriminate between CO_2 and O_2. Thus, further selection of RubisCO molecules after growth of complemented strain SBI/II- under aerobic chemoautotrophic growth conditions in strain SBI/II-, or at very low (<1%) CO_2 concentrations under anaerobic photoautotrophic conditions, could possibly be reflective of a very efficient RubisCO with high specificity or favorable K_o or K_c.

Both metagenome-derived RubisCO genes used in this study were able to support photoautotrophic growth in strain SBI/II-. The doubling time of approximately 24 h observed for strains complemented with both genes was comparable to the doubling time observed in the strain complemented with the well-characterized *Synechococcus* PCC6301 RubisCO genes.

The pRPS-MCS3 vector carries a tetracycline cassette. This is useful for maintaining the plasmid on heterotrophic media, in both *E. coli* and in strain SBI/II-. Tetracycline is degraded by light and thus cannot be used for plasmid maintenance under photoautotrophic conditions in SBI/II-. However, transconjugants containing a functional RubisCO conferred a selective advantage under autotrophic growth conditions without antibiotics present in the media. To ensure that the observed growth was due to complementation from expressed metagenomic RubisCO genes, rather than from contaminating DNA, plasmids were extracted from samples of each photoautotrophically grown SBI/II- complemented culture and subsequently analyzed. Plasmids isolated from

42

each culture showed the expected restriction pattern for the pRPS-MCS3 plasmid containing the expected insert. Further, the presence of the expected insert was verified by direct sequencing of the cloned DNA from each plasmid sample.

Enzyme purification and properties. Crude lysates demonstrated equivalent RubisCO activity levels when derived from either fresh or frozen (-80 $^{\circ}$C) photoautotrophically grown strain SBI/II- cells complemented with *rbcLS* genes. Native RubisCO was successfully purified from photoautotrophically grown SBI/II- complemented with *Synechococcus* 6301 or metagenomic RubisCO genes. The RubisCO proteins deriving from BAC15 and 6301 genes were purified to greater than 95% homogeneity without affinity tags (Figure 17A). The recombinant 4N23 enzyme, however, proved to be very unstable once cell pellets were lysed. Samples of 4N23 protein kept on ice at any stage of the purification process, from crude lysate through the sucrose density gradient centrifugation step, lost ~90% activity per 24 h time period. By contrast, the BAC15 and 6301 samples lost only ~10% of activity per 24 h time period. Reducing agents (DTT, β-mercaptoethanol), protease inhibitors and a wide variety of salt, buffer and solute conditions were unable to impede the loss of activity for the 4N23 preparations. Only samples frozen at -80 $^{\circ}$C showed no loss of activity. Moreover low solute concentrations (e.g. UnoQ loading buffer 20 mM Tris-HCl, 1 mM $MgCl_2$) and ammonium sulfate precipitation induced total loss of activity in samples containing 4N23 RubisCO. Due to the rapid loss of activity, a specialized protocol for partially purifying 4N23 RubisCO was developed such that samples could be placed at -80 $^{\circ}$C until they were assayed. That protocol is detailed in Materials and Methods section above. In short, however, the

protocol was optimized for speed and to minimize the use of solutions of very low or

very high osmolarity. Samples of 4N23 RubisCO eluted from the sucrose gradient

preparation were stable at -80 $^{\circ}$C indefinitely. Both recombinant metagenomic RubisCOs

were isolated at high specific activity when the above precautions and storage protocols

were followed.

The CO_2/O_2 substrate specificity factor ($\Omega = V_C K_O / V_O K_C$) was also measured for both of

the metagenomic RubisCOs. (Figure 17B) Although not providing a full kinetic

characterization, Ω is an important indicator of the physiological capabilities of these

enzymes. The specificity factor observed for 4N23 RubisCO ($\Omega = 30 \pm 5$) is a little low

for form I enzymes, as values for Ω below 30 have only been observed for *Rb. capsulatus*

form I (Horken and Tabita 1999b) ($\Omega = 26 \pm 1$) and for the two *Hydrogenovibrio*

marinus form I RubisCOs ($\Omega = 27$ and 33, respectively) (Igarashi and Kodama 1996)

The standard deviation of +/- 5, however, means that the value may not be remarkable

low. The substrate specificity factor determined for the BAC15 RubisCO ($\Omega = 35 \pm 2$) is

closer to previously reported values for type IA RubisCOs from non-cyanobacterial

bacteria (Tabita 1999) Although the specificity factors determined for the two

metagenomic RubisCOs in this study are not unusual, these are the first reports of kinetic

data for carbon fixation enzymes obtained from uncultured organisms.

A
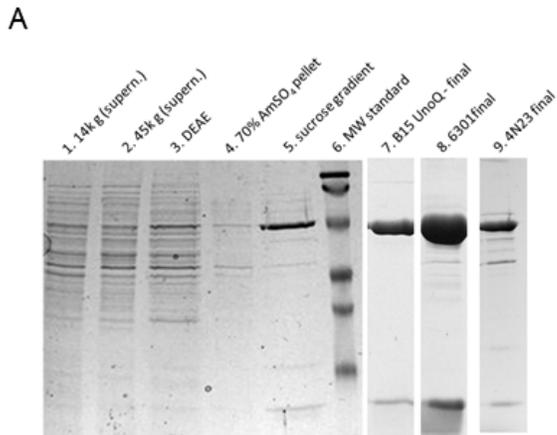
B

Enzyme	Specific activity (μmol min^{-1} mg^{-1})	Ω	K_{cat} (s^{-1})
6301	2.5 ± 0.2	36.8 ± 4.0	2.7
B15	1.7 ± 0.3	35.4 ± 2.3	1.8
4N23	0.56 ± 0.07	30.2 ± 5.4	0.62

Figure 17 - Purification of Metagenomic RubisCO

(A) SDS-PAGE gel stained with Coomassie showing purification steps for the BAC B15 RubisCO, starting from photoautotrophically-grown *Rb. capsulatus* SBI/II- complemented with plasmid pRPS-B15 (lanes 1-5). Lane 7 shows the final purified B15 protein. Lanes 8 and 9 show the final purified 6301 and BAC 4N23 proteins, respectively. Lanes 7, 8 and 9 are from different gels, but are included here to demonstrate the quality of the final products. (B) Specificity factors and k$_{cat}$ values for *Synechococcus* PCC6301, and BAC B15 and BAC 4N23 recombinant RubisCOs. All values are the result of at least two separate assays, each performed in triplicate and using two independent purifications of untagged enzyme produced in autotrophically grown *Rb. capsulatus* SBI/II-.

45

Chapter 2 – Functional RubisCO Selection

Introduction

There are significant limitations to sequence-based detection of RubisCO ORFs, as

mentioned in Chapter 1. PCR and hybridization approaches, for example, are inherently

limited in that only sequences with significant similarity to known groups can be

detected. The more degenerate bases are used in formulating the primers or the lower the

stringency of the probe, the wider the net cast, but the greater the probability of false

positives. A selection system based on functional complementation of SBI/II- by a

library of metagenomic DNA, however, would bypass many of these shortcomings.

As an indication of the diversity that could be missed by traditional techniques, Chapter 3

of this document focuses on the RubisCO from *Mc. burtonii*. In short, the *Mc. burtonii*

RubisCO seems to form a clade (with 3 other RubisCOs from methanogenic archaea –

see Figure 5, p. 6) that is intermediate between form II and III. All 4 of these sequences

were discovered only because of the recent proliferation of whole-genome sequencing.

No conserved-site PCR or probe-based hybridization methods would have discovered

such sequences with only ~30-40% identity at the amino acid level to their nearest form

II or III neighbors.

Shotgun sequencing of metagenomic samples has also uncovered a wealth of new

46

sequences despite that field being still in its infancy. Some of the possibilities and pitfalls of shotgun sequencing could be seen in the announcement that a project had identified >400 unique RubisCO sequences from the open ocean and, further, "65 of those sequences clustered with no known RubisCO sequences." (Yooseph *et al* 2007) Closer perusal of that announcement by Tabita and colleagues (2008b), however, revealed that many of those novel sequences appeared so because they were chimeras resulting from an automated curation system. Although bioinformatics tools are constantly advancing, there remain many obstacles to rapid and simple detection of full-length RubisCO sequences from environmental (uncultured) samples.

Further, simply isolating and characterizing RubisCO genes one by one would be a very time-consuming and labor-intensive procedure. Further, limits of structure-function predictive ability prevent purely molecular techniques from selecting examples of an enzyme with particular kinetic properties. Finally, it is reasonable to suppose that many of the most common, and hence easiest to find, RubisCOs have already been identified. Thus, any system that randomly selects an enzyme from the environment will mostly likely re-find the most common representatives already in culture – e.g. *Synechococcus* or *Prochlorococcus* from the open ocean, *Rhodopseudomonas* or *Rhodobacter* from anoxic fresh water, etc. Note, however, that environmentally abundant organisms are not necessarily easy to culture (e.g. the ubiquitous SAR11 marine bacterioplankton clade (Rappe *et al* 2002)), and hence the importance of culture-free methods of studying key enzymes.

47

Form III RubisCO is of particular interest in studies of how RubisCO interacts with molecular oxygen because so many of the form III enzymes are unable to perform the carboxylation reaction in the presence of even minute traces of O_2 (Finn and Tabita 2003; Finn and Tabita 2004). There is no apparent function for the oxygenation reaction in form III RubisCO – rather, it is more likely that this apparent bias is due to sampling error. Until the recent availability of fast and inexpensive whole-genome sequencing, few RubisCO genes were known from archaea. With the advent of molecular survey techniques archaea have been identified in a wide range of mesophilic and aerobic habitats such as ocean surface waters and agricultural soil (DeLong 2006; Rusch *et al* 2007; Venter *et al* 2004; Yooseph *et al* 2007). Some of these as-yet-understudied strains may well contain aerotolerant form III RubisCO, such as the *Mc. burtonii* RubisCO detailed in Chapter 3.

Some researchers suggest form III as the ancestral RubisCO, (Ashida *et al* 2005; Tabita *et al* 2007; Tabita *et al* 2008) legacies from the ancient anaerobic biosphere populated by methanogens and extremophiles. Although the evolution of RubisCO remains controversial, finding more such aerotolerant form III enzymes could shed light on the debate, as well as providing more information on the structural basis for oxygen tolerance. The deep branches (see Figure 5, p. 6) among form III RubisCOs precludes molecular techniques that would identify them, as a group, in a library of metagenomic DNA.

Although enough sequences exist to define the form IC clade (bacterial red-type), there

are far fewer representatives in the literature and databases than for other prokaryotic RubisCOs. The characterized form IC enzymes, however, appear to have the highest specificity ($\Omega = V_c K_o / V_o K_c$) among prokaryotic enzymes (reviewed in Tabita (1999)). Most discoveries of RubisCO up to this point have come from cultured organisms, and it is reasonable to suppose that many bacteria harboring IC-type RubisCO are simply recalcitrant or grow in consortia, or are simply not suited to the culture techniques to which they have been exposed. Molecular techniques have been developed that can capture portions of IC sequences from the environment, (Elsaied *et al* 2007; Spiridonova *et al* 2004) but not whole genes. More IC sequences and kinetic data will aid in broad-ranging studies correlating structure and function (c.f. Tcherkez *et al* (2006) which considered only a few bacterial enzymes and no IC enzymes). Finally, given the diverse range of habitats that can be surveyed it is reasonable to assume that some organisms will have adapted to a high-oxygen/low CO_2 environment by maximizing the efficiency of their most critical carbon-fixation enzyme. A phenotypic selection, however may find such genes where purely molecular techniques that rely on sequence homology fail.

Materials and Methods

Growth of Rb. capsulatus and E. coli strains

See Chapter 1 Materials and Methods for strains and growth conditions. Additionally, photoheterotrophic growth in *Rb. capsulatus strain* I/II- was accomplished in OM supplemented with malate or butyrate. OM-malate liquid media was made by adding filter sterilized malate to OM to a final concentration of 0.4% (v/v) in sterile 20 ml screw cap tubes without headspace. Liquid OM-butyrate media was made by adding filter

sterilized sodium butyrate to OM to a final concentration of 10 mM. Sodium bicarbonate

was added to the OM-butyrate to 10 mM final concentration. Solid OM-malate and OM-

butyrate media contained 1.5% noble agar (USB). OM-malate and OM-butyrate plates

were incubated anaerobically as described for minimal media plates in Chapter 1 –

Materials and Methods.

Enzyme Assays and Chemicals

See Chapter 1 "Materials and Methods"

Library Creation – Overview

Library creation and quality control were the major portions of this project. The sections

below go in to some detail as to the various protocols attempted. All of the methods

shared a similar outline, however:

1. Genomic DNA (gDNA) isolation
2. gDNA fragmentation and size selection
3. Insertion of gDNA fragments into primary vector
4. Transformation of *E. coli* T10 with gDNA library
5. In some protocols only: Transfer of gDNA fragments in to secondary vector via Gateway LR Recombinase
6. Transfer of library from *E. coli* T10 to *Rb. capsulatus* SBI/II- via conjugation
7. Selection of *Rb. capsulatus* complemented with gDNA library
8. Quality control via Form I RubisCO ORF PCR

For each method outlined below, certain elements were constant. Unless otherwise

noted, closed plasmids (via ligation, TOPO vectors or Gateway recombinase) were

maintained in *E. coli* TOP10 cells. Chemically competent TOP10 cells were supplied by Invitrogen, while electrocompetent cells were prepared via repeated washings in sterile ice cold double-distilled water and stored in 10% glycerol (Seidman *et al* 2001). After transformation, the cells were allowed to grow for one hour in an appropriate volume of SOC (per L of distilled water: 2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM D-glucose with the glucose filter sterilized and added after the rest of the media had been autoclaved). Chemically competent cells were obtained exclusively as kit components and transformation followed manufacturer protocols. Transformation of electrocompetent cells used 50 µl aliquots (50% cell mass, balance 10% glycerol in water) and received 1 ml of room-temperature SOC following transformation. The transformed cells were then placed in a shaking incubator at 37 ^{o}C for one hour before spread plating or further manipulation.

Site directed mutagenesis used Phusion High Fidelity DNA Polymerase (NEB) in GC buffer. The sequences of individual primers are given in the appropriate section below. The PCR reaction for each mutagenesis reaction was based on a modified version of the Stratagene QuickChange protocol. Briefly, 10 ng of template was used with 2 pmol/µl of forward and reverse primers in 50 µl reaction. The reaction proceeded over 18 cycles of 98 ^{o}C 20s, 65 ^{o}C 20s and 15s per 1 kb of template DNA. Following the reaction, the complete volume was digested with 10 U of DpnI (NEB) for one hour to remove the methylated template DNA before 1 µl of the reaction volume was used to transform electrocompetent *E. coli*. The sequence of a given mutagenized molecule was determined before the next step (i.e. subcloning) was performed.

DNA concentration was determined using a Qubit fluorometer (Invitrogen) Broad-Range

kit unless otherwise noted.

Vectors Several vectors were tested in the course of this investigation for their suitability

for building broad host range metagenomic expression libraries. The pRPS-MCS3

plasmid has been previously described (Smith and Tabita 2003). See Figure 18 for an

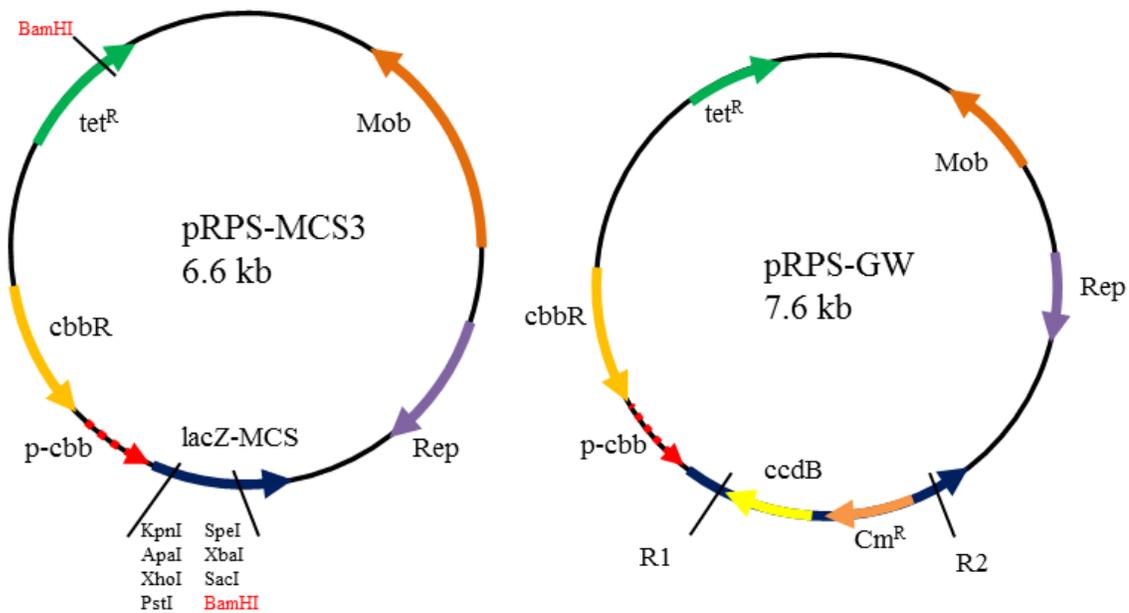illustration. The pRPS-MCSB vector was created by converting the ApaI site in pRPS-

Figure 18 - pRPS-MCS3 and pRPS-GW

pRPS-MCS3 was first described from Smith and Tabita (2003). pRPS-MCSB was created by
deleting the BamHI site in the tetracycline resistance gene (leaving the already present BamHI site
in the MCS usable for cloning). pRPS-MCSBB used pRPS-MCSB as a template, and replaced
the ApaI site in the MCS with BglII.

 pRPS-GW was created for this study. The GW designation indicates compatibility of the vector
with the Gateway cloning system (Invitrogen). pRPS-GW serves as a destination vector. R1 and
R2 designate the recombination sites that are compatible with the L1 and L2 sites in a Gateway

MCS3 to a BglII site using site directed mutagenesis. The pRPS-MCSBB plasmid is identical to pRPS-MCSB except that a BamHI site at position 5736 (in the coding region of the tetracycline resistance gene) was eliminated using site directed mutagenesis. The mutagenesis reactions employed Phusion DNA polymerase (NEB) with a cycle of 98 $^{\circ}$C 1 m, followed by 18 cycles of 98 $^{\circ}$C for 20s 63 $^{\circ}$C 20s and 72 $^{\circ}$C for 1 m 45 s. The primer sequences were GATCGGATCCATCGGTGGACTCTCCCTAGATG and the reverse complement. The pRPS-MCSBB vector thus had only one BamHI site – that in the MCS at position 3251. The mutation was silent and used codons common to both *E. coli* and *Rb. capsulatus*.

The pRPS-GW vector was created using the Gateway Vector Conversion System (Invitrogen). (See Figure 18 for a diagram). The pRPMS-MCS3 plasmid was digested with PstI and XbaI, blunted with Klenow fragment and the Gateway conversion cassette was ligated in to the vector. The Gateway conversion cassette contained a chloramphenicol gene for positive selection and a *ccdB* gene for selection against plasmids that failed to undergo recombination with target DNA. The pRPS-GW vector must be maintained in CcdB-survival strain of *E. coli* (Invitrogen) that harbors a mutation in the DNA gyrase gene that confers resistantce .

Plasmid pVK101 was used in the form described (Knauf and Nester 1982). See Figure 19 for an illustration. Preparation of pVK101 requires 100 mL of overnight culture processed with a MidiPrep Kit (Qiagen). At the beginning of this study, the sequence of pVK101 was unknown. In order to sequence inserts in pVK101, the sequence of the

cloning region needed to be determined. Thus, a 1 μg of pVK101 without insert was digested with BglII and HindIII or EcoRI and BglII (see Figure 18 for a vector map), the small fragment was gel purified and ligated in to pCDF-Duet (Novagen) in to complementary restriction sites. The 1008 and 1157 bp fragments (respectively) were sequenced. The complete sequence between the EcoRI and HindIII sites was deposited in GenBank with accession number JQ755431. Oligonucleotides pVK101_F (GGTATGAGTCAGCAACACC) and pVK101_R (CATTCGCGAGAGCCTTGAGTC) were designed to flank the BglII site in pVK101 and could then be used to initiate sequencing reactions.

Another vector used in an attempt to build a metagenomic library was pGNS-BAC (Kakirde *et al* 2011). The pGNS-BAC vector contains gentamycin and chloramphenicol resistance cassettes, as well as a multiple cloning and mobility sites. Additionally, pGNS-BAC can be induced to high copy number in *E. coli* with 1 mM arabinose which would allow for the use of minipreps for vector and library purification. A sample of pGNS-BAC was obtained from the laboratory of Dr. Mark Liles (Dept. of Biological Sciences, Auburn University) and transformed in to T10 and S17 *E. coli* via electroporation. The vector (without insert) was then transferred in to *Rb. capsulatus* I/II- and SB1003 via conjugation (as described above). Transconjugants were plated on PYE with rifampicin and 2, 6 and 12 μg/ml chloramphenicol. Samples from the same matings were streaked out on to PYE with rifampicin plus 2, 6 and 12 μg/ml gentamycin.

The chloramphenicol resistance gene from pGNS-BAC was amplified with primers

pGNS_CAT_F (GAGGTTCCAACTTTCACCATAATG) and pGNS_CAT_R

(TGACAGCTTATCATCGAATTTCTG), cloned in to pCR8/GW/TOPO and then

transferred in to pRPS-GW via the Gateway LR recombinase reaction. The T10 *E. coli*

cells with pCR8-CM or pRPS-CM were plated on LB media supplemented with 12 µg/ml

chloramphenicol to verify that the chloramphenicol gene conferred resistance. The

pRPS-CM vector was transferred to SBI/II- and SB1003 via conjugation and the

transconjugants plated on PYE supplemented with 2, 6 and 12 µg/ml chloramphenicol as

well as rifampicin.

No library construction was attempted with pGNS-BAC.

The vector pENTR3C was obtained from Invitrogen to serve as an alternate entry vector

for Gateway cloning. A derivative, pENTR3BB, was created via site directed

mutagenesis such that the EcoRV site at position 1978 was replaced with a BamHI site.

pENTR3BB thus had 4 total BamHI sites – two flanking the dual selection markers (*cam*

and *ccdB* at positions 484 and 1978) and two internal to the selection markers (positions

578 and 1275). The change was effected with primer EcoRVtoBamHI

(CGGCCGCACTCGGGATCCCTAGACCCAGC) and the reverse complement. The 1

U of Phusion DNA polymerase (NEB) was used with a program of 1 m 98 $^{\circ}$C, 18 cycles

of 20 s 98 $^{\circ}$C, 20s 63 $^{\circ}$C, 1 m 72 $^{\circ}$C followed by 5 m 72 $^{\circ}$C. For routine cloning of PCR

products produced with Taq for sequencing, pGEM EasyT vector was used as part of the
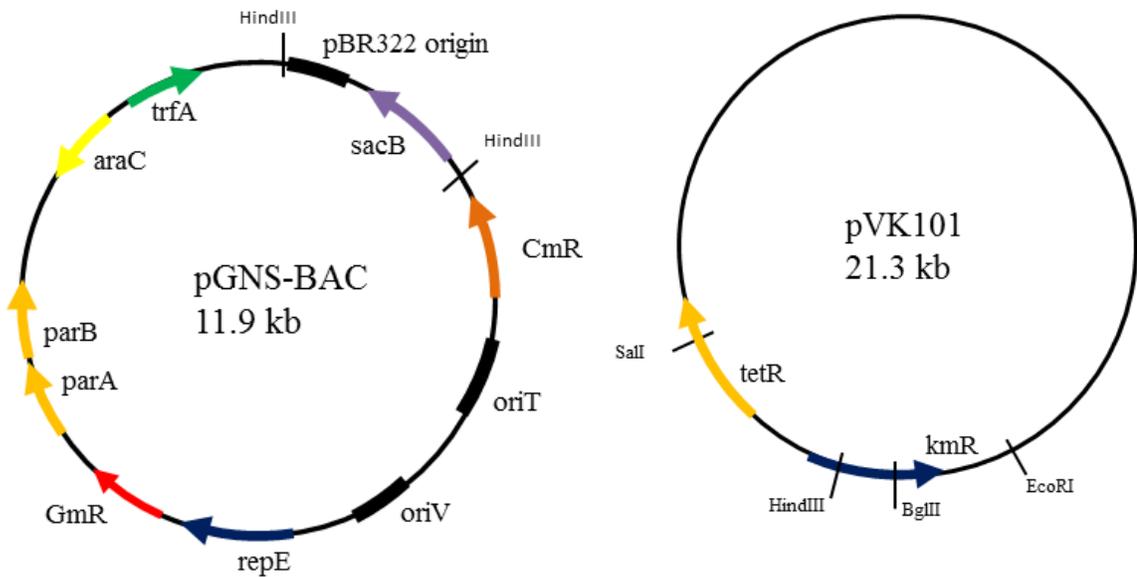
EasyT Cloning kit (Promega).

Figure 19 - pGNS-BAC and pVK101

pGNS-BAC was described in Kakirde et al (2008). pVK101 was described in Knauf and Nestor (1985). The sequence of pVK101 was unknown at the commencement of this study. Only the region from HindIII to EcoRI has been sequenced.

Library Construction. The DNA for library construction was derived from enrichment cultures of autotrophic bacteria. Samples of soil and water from the Olentangy River were added to 350 ml OM media (approximately 2% w/v or v/v, as appropriate). The bottles were then capped and exposed to light for photoautotrophic (PA) enrichment or kept in a darkened fume hood for chemoautotrophic (CA) enrichment. The PA bottles were continuously bubbled with 5% CO_2/balance H_2 and maintained at 30 ^{o}C with banks of incandescent lights for illumination. 100 µg/ml cycloheximide was added to retard growth of eukaryotic algae. The CA enrichments were bubbled with 5% CO_2/45%H_2/50% air (=10% O_2) and maintained at 30 ^{o}C in the dark. When the enrichments reached an OD_{600} of ~1.0, 2 ml was removed used as inoculum for a

subsequent 350 ml enrichment. Samples from each enrichment were centrifuged and frozen at -20 $^{\circ}$C until nucleic acid extraction.

Nucleic Acid Purification and Fragmentation. Genomic DNA was isolated from enrichments using a Genomic-tip 100/G column-based kit (Qiagen) to provide high molecular weight DNA. Elution from the columns was enhanced by warming the elution buffer to 55 $^{\circ}$C prior to application to the column. In a modification of the kit instructions, the DNA eluted from the column was precipitated with 60% isopropanol in a 15 ml Falcon tube at 4,000 g for one hour using an Eppendorf 5810R centrifuge with a swinging bucket rotor, rather than the 15 minutes at 14,000 g suggested in the instructions. The centrifuge bottles that can withstand the 14,000 g offer poor recovery of small pellets. DNA pellets were washed once with 70% ethanol, allowed to dry and then resuspended in 100 μl TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA). Yield was quantitated using a Qubit Fluorometer (Invitrogen) with the Broad Range protocol.

Purified DNA was fragmented using several alternative protocols. Sau3AI (/GATC), MspI (C/CGG), RsaI (GT/AC) were used individually or in combination. The precise ratio of restriction enzyme units to μg DNA was determined empirically with each batch of genomic DNA. A typical digest was 1 U of restriction enzyme(s) (1 U of one enzyme or 0.5 U of two different enzymes) added to 2 μg DNA in a total volume of 20 μl (1X of appropriate buffer) and incubated for 150 seconds at 37 $^{\circ}$C before the reaction was terminated with 2 μl 0.5M EDTA.

Several restriction enzymes with six nucleotide recognition sites were used as well. BglII (A/GATCT), HindIII (A/AGCTT) and BamHI (G/GATCC) were each used to digest genomic DNA to completion. A typical reaction was 10 U of enzyme with 4 μg of gDNA for 20 minutes.

DNA was also fragmented with the NEBNext dsDNA Fragmentase kit (NEB). The fragmentase kit uses a proprietary blend of enzymes that cause random single strand nicks in double stranded DNA, followed by second strand cleavage opposite the nick. The precise reaction conditions had to be empirically determined for each batch of gDNA in order to maximize of the number of fragments of desired length. A typical reaction was 1 μl of Fragmentase enzyme mix with 4 μg of gDNA in a total volume of 20 μl. The reaction was allowed to proceed for approximately one minute.

Regardless of the fragmentation method, the digested DNA was run on a 0.8% agarose gel and the region containing fragments between 1 kb and 6 kb was excised. A short (~15 minute at 100V) run was used to minimize the volume of agarose gel that required processing. The cut DNA was extracted from the gel using the Qiaquick Gel Extraction Kit (Qiagen). The concentration of the purified, fragmented DNA was determined with a Qubit Fluorometer (Invitrogen).
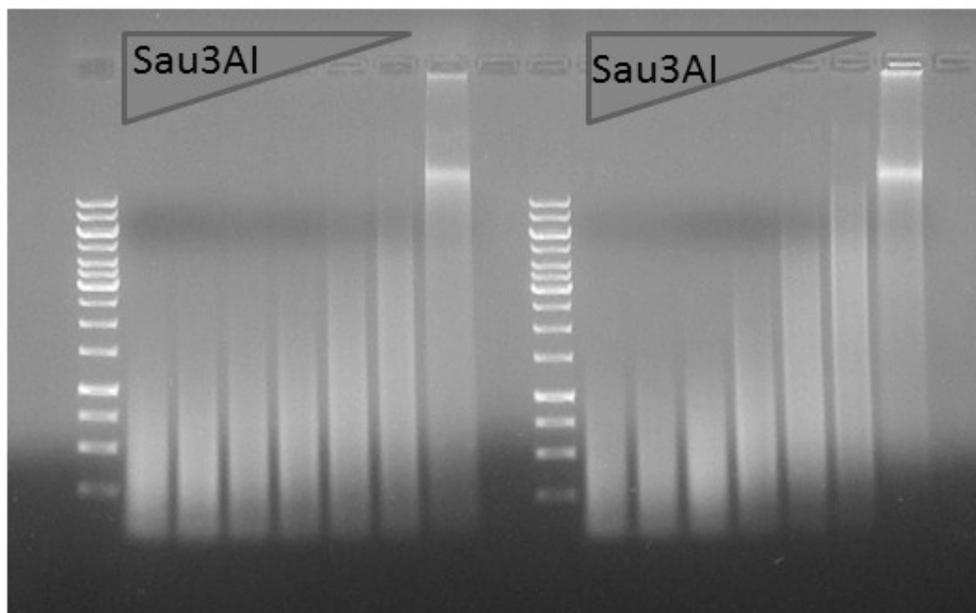
Figure 20 - Partial Digest of Genomic DNA

Genomic DNA isolated from photoautotrophically grown enrichment culture. 1 μg DNA per lane, 2 minute digest with varying concentrations of Sau3AI. Two digests were performed in parallel to illustrate repeatability. Last lane on right is undigested genomic DNA.

Library Creation – Direct Ligation in to complementation vector.

If direct ligation was to be used, the purified, fragmented DNA was mixed with digested plasmid at a 3:1 insert:vector molar ratio. The direct ligation attempts were: Sau3AI digested gDNA ligated to *Bgl*II digested pVK101; *Bgl*II digested pRPS-MCSB (see below); or BamHI digested pENTRB. The ratio was calculated assuming an average insert size of 4 kb. The destination vector varied (see Vectors section below), but each was digested with an appropriate restriction enzyme, treated with Shrimp Alkaline Phosphatase (New England Biolabs) for at least one hour after addition of SAP buffer and purified using a QiaQuick PCR CleanUp Kit (Qiagen) to remove enzyme and buffer. The concentration of the purified, digested vector was determined as previously described.

59

Ligation was performed using T4 ligase (Invitrogen or New England Biolabs) at room temperature for one hour or overnight at 14 $^{\circ}$C. Following ligation, the ligase was inactivated by heating the reaction at 65 $^{\circ}$C for 15 minutes before transformation of 50 µl
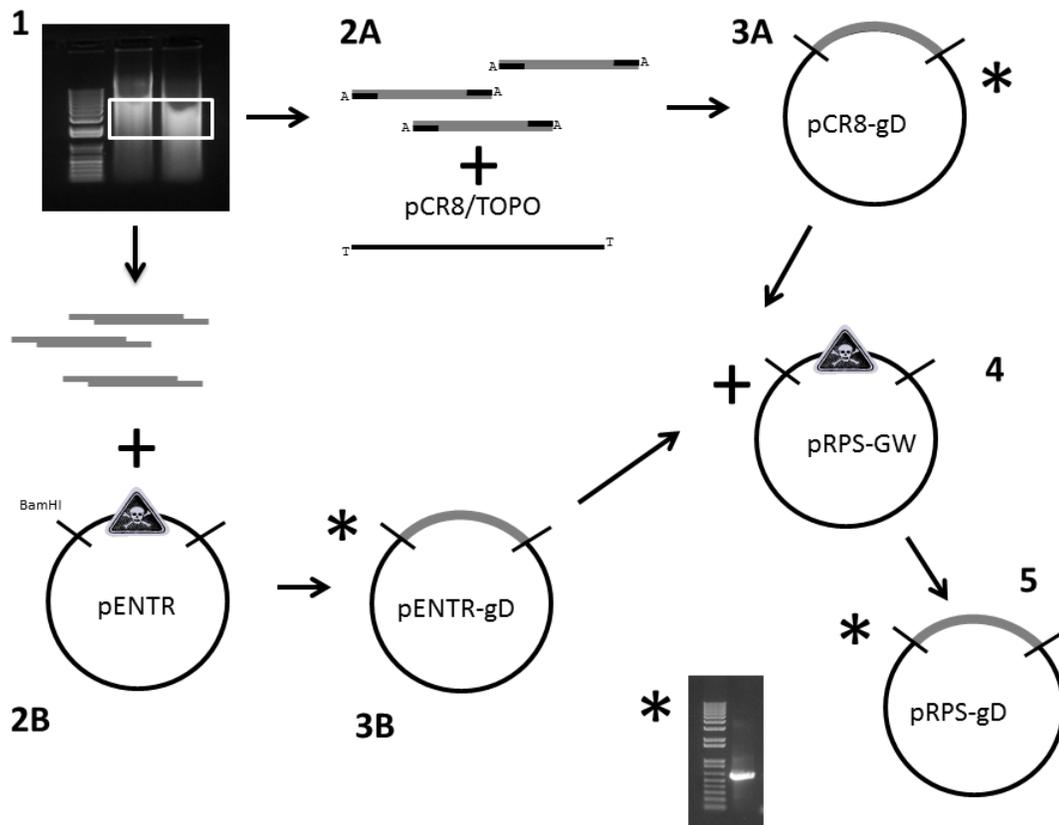

Figure 21 - Metagenomic Library with Gateway-Compatible Vectors

1. Partial digest and size selection of genomic DNA. **2A** End repair and A-tailing of gDNA before cloning with pCR8/TOPO. **3A** Library of genomic DNA in pCR8. **2B** Direct cloning of digested DNA into compatible sites in pENTR3C. Digestion of pENTR excises lethal *ccdB* cassette. **3B** Library of genomic DNA in pENTR. **4** Combine entry libraries with pRPS-GW and lambda phage LR recombinase (catalyzes recombination of L1 with R1site and L2 with R2 in the source and destination vectors, respectively). **5** After incubation, library of genomic DNA in pRPS expression vector. Successful recombination events replace the lethal *ccdB* cassette in pRPS-GW. The pRPS-gD library can then be transformed in to *E. coli* and then transferred to *Rb. capsulatus* I/II- via conjugation. * indicates stages at which the libraries are screened with form I conserved-site RubisCO primers.

60

of T10 electrocompetent cells with 1 µl of ligation reaction. The transformed *E. coli* were suspended in 1 ml of room temperature SOC before being incubated at 37 °C for one hour in a shaking incubator. 100 µl of the transformation mixture was spread on LB with suitable antibiotic for selection, and the remainder of the transformation mixture was used to inoculate 100 ml of LB containing an appropriate selective antibiotic. The 100 ml culture was grown overnight at 37 °C in order to propagate the library. Several aliquots of the overnight library culture were used to make glycerol stocks preserved at -80 °C. An appropriate volume of the remainder, depending on the copy number and size of the vector used in library construction, was used for plasmid purification.

Library Creation – TOPO cloning. Genomic DNA was cloned directly in to the PCR8/GW-TOPO vector using the PCR8/GW/TOPO-TA Cloning Kit (Invitrogen). Fragmented, size-selected and purified gDNA was end-repaired using either Klenow fragment (NEB) or the NEBNext End Repair Module (NEB) according to manufacturer instructions. The blunted DNA fragments were then purified with the QiaQuick PCR Cleanup Kit (Qiagen) before a 3' A was added in order to facilitate TA cloning. A-tailing was performed in 20 µl with 1 U of Taq (Invitrogen) and 1 mM dATP with a single 10 minute incubation at 72 °C. Alternatively, the NEBNext dA-Tailing Module (NEB) was used per manufacturer's instructions. TOPO cloning reactions used a minimum of 30 ng A-tailed DNA per 15ng TOPO vector. Reactions proceeded for one hour at room temperature. 4 µl of the cloning mixture were added to one aliquot of chemically competent *E. coli* T10 cells included with the kit. Following heat shock treatment, 250 µl of room temperature SOC were added to the 50 µl of transformed cells and the tube was

incubated for one hour in a shaking incubator at 37 °C. After that hour, 20 µl of the 300

µl transformation mixture was spread on a LB agar plate supplemented with

spectinomycin to check the efficiency of cloning. The remainder of the transformation

mix was used to inoculate 100 ml of LB - spectinomycin and grown overnight in a

shaking flask at 37 °C. 10 ml of the overnight culture was processed with a midi-prep kit

(Qiagen).

An alternative approach was modified from Schmitz *et al* (2008). See Figure 22 for an

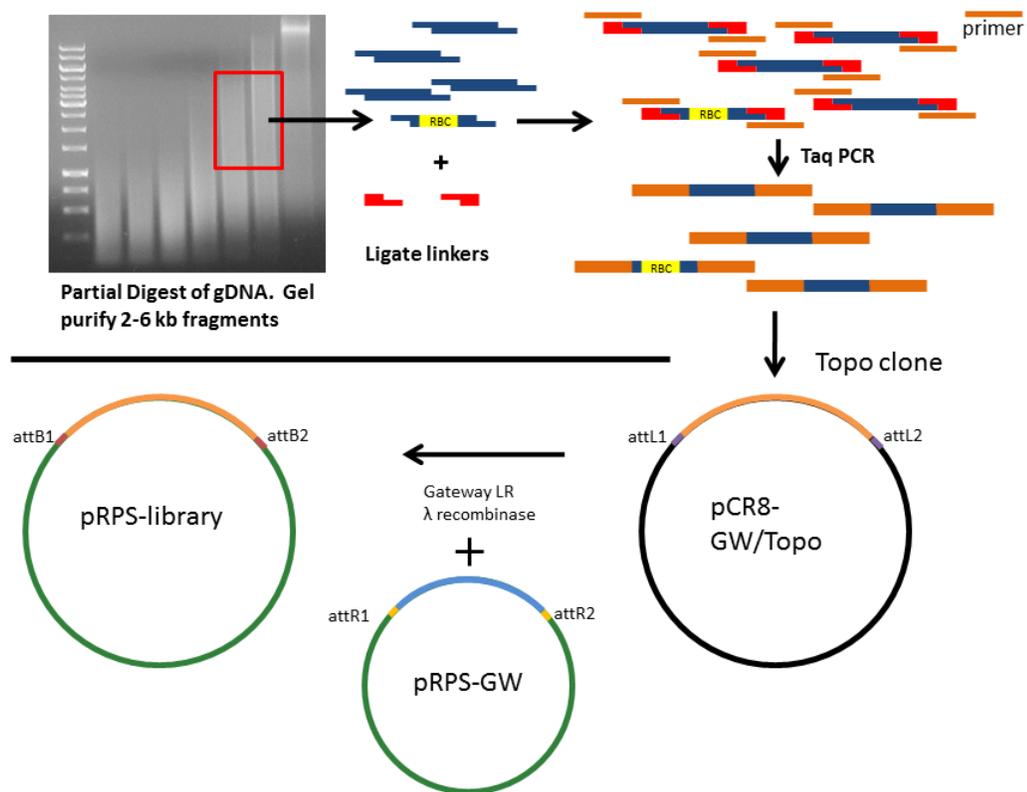**Figure 22 - Linker Ligation and PCR Library**

Schematic of the linker-based cloning method, modified from Schmitz *et al* (2008). Digested DNA is ligated to a dephosphorylated linker with "sticky" ends. The linker is then used a priming site for Taq-basedPCR and the resulting A-tailed PCR products are cloned in to the pCR8/GW/TOPO vector. pCR8 is a donor for the gateway reaction, with pRPGS-GW as the recrecipient.

illustration. Briefly, gDNA was subject to a partial digest with a restriction enzyme with a 4-base recognition site. Sau3AI and MspI were chosen because they are not affected by DNA methylation and they leave overhangs to increase ligation efficiency. The cut DNA was either ligated immediately to a dsDNA oligonucleotide linker or it was run out on an agarose gel in order to purify the fragments from 2,000-6,000 bp. The gel-purified fragments would then be quantitated and ligated to a linker, size-selected on an agarose gel, and ligated to a dsDNA oligonucleotide linker. Equimolar amounts of ssDNA oligos were mixed in 1X ligation buffer before being heated to 80 ^{o}C for 5 minutes. The tube of oligos was then placed in a beaker of 80 ^{o}C water and the beaker (and oligos) were allowed to slowly return to room temperature. 40 ng of annealed oligonucleotide linker were then added per 100 μg of digested DNA before ligation with T4 DNA ligase (NEB). The oligos had no 5' phosphate, and thus were not able to form concatomers despite having complementary sticky ends. The ligation was allowed to proceed at room temperature for 1 hour before the mixture was cleaned with the QiaQuick PCR Cleanup Kit (Qiagen). Linker NotI_R was then used as the sole primer in a PCR reaction. The reaction conditions included 63°C annealing temperature, 6 minutes extension time and 25 cycles. 4 μl of the finished PCR reaction were used for TOPO-TA cloning. The TOPO reaction was transformed as described above for the direct cloning of gDNA. Alternately, gDNA digested with Sau3AI or RsaI was end-repaired using the NEBNext End Repair Module (NEB) and an aliquot was checked on an agarose gel to determine whether the digest had produced fragments of suitable size. The blunt DNA was then ligated to the Sau3AI_NotI_F and NotI_R dsDNA linker at a ratio of 40 ng linker to 100 ng gDNA. The linker was assembled as described above before being mixed with the

gDNA and 5U of T4 DNA ligase (Invitrogen) and allowed to incubate at 14 °C overnight. This linker would ligate in an opposite orientation to the sticky-end ligation described above such that subsequent amplification would require the Sau3A_NotI_F ssDNA oligonucleotide as the PCR primer. This reaction used Taq polymerase, 63 °C annealing temperature and an extension time of 6 minutes.

Gateway cloning. A gDNA library in TOPO could be transferred in to pRPS-GW using LR Recombinase (Invitrogen) per the manufacturer's instructions. In order to maximize recombination events, the recombination reaction was allowed to proceed overnight in a thermocycler set to maintain a constant temperature of 25 °C. Following recombination, the pRPS-GW library was transformed into T10 cells. An aliquot of the transformed cells was spread on LB-tet plates to verify the efficacy of the recombination while the remainder of the transformed cells was used to inoculate 100 ml of LB-tet. The overnight library was harvested using a midi-prep kit (Qiagen) before being screened for form I RubisCO genes (see Quality Control, below).

 Conjugative DNA Transfer. The library was transferred in to SBI/II- via bi-parental mating. The protocol for this step proceeded as described in Chapter 1 – Methods section.

Quality Control. At each stage of library construction (gDNA isolation, fragmentation, size selection, entry vector library, and/or expression library), form I PCR primers were used to verify the presence of a form I RubisCO sequence. The sequences of the primers

were identical to those of Alfreider *et al* (2003). In a modification to the Alfreider *et al*

protocol, we used Invitrogen cloned Taq polymerase, 1 mM MgCl$_2$ and a cycle of 20s at

94 ^{o}C, 20s at 53 ^{o}C and 50s at 72 ^{o}C. Every QC PCR was conducted using

Synechococcus PCC 6301 *rbcL* in pUC19 as a positive control.

RubisCO context   For libraries unable to produce complementation, a novel pull-down

procedure was developed to identify the context of the RubisCO sequences detectable by

FI PCR amplification described above. First, the cloning region of a library was

amplified with Phusion High Fidelity DNA Polymerase (NEB) with GC buffer and

primers flanking the cloning region. For pCR8, the primers used were M13F

(CGCCAGGGTTTTCCCAGTCACGAC) and GW2 (GTTGCAACAAATTGATGAG-

CAATTA). In pENTR3C and pENTRBB, primers pENTR3CFlank_R

(GTGCAATGTAACATCAGAGATTTTGAGACACG) and pENTR3CFlank_F

(CAAACTCTTCCTGTTAGTTAGTTACTTAAGCTCGG) were used. In pRPS-MCS3,

the RPSF and RPSR primers were used. The complete PCR reaction was digested with

10 U of DpnI (NEB) for 1 hour at 37 ^{o}C before clean-up with QiaQuick PCR Clean Up

Kit (Qiagen). An aliquot of the reaction was run on an agarose gel to verify the success

of the reaction before 100 ng of the clean PCR product was added to 1 pmol of FI_Alf_R

primer derivatized with a 5' biotin (Invitrogen) linked to the primer with a 15 carbon

bridge. The DNA-primer mix was made in 1X PCR buffer with 0.2 mM dNTPs and 1 U

Taq. The mix was incubated at 94 ^{o}C for 1 minute, 30 s at 53 ^{o}C (FI_Alf_R annealing

temperature) and 3 minutes at 72 ^{o}C. The mix was purified again with the QiaQuick kit

before incubation with 5 µl of steptavidin-conjugated Dynabeads from the Dynabeads

KilobaseBINDER Kit (Invitrogen). The kit was used per manufacturer protocol. After the initial incubation of streptavidin-coated beads and biotinylated DNA, the beads were pulled to the side of a thin walled PCR tube and the supernatant was aspirated and saved on ice. At the conclusion of the Dynabead protocol, the beads with captured DNA fragments were resuspended in 20 µl 10 mM Tris-HCl pH 8.0.

5 µl of the beads (with DNA attached) were used as the template for a PCR reaction using 1 U Taq DNA polymerase (Invitrogen), 1 mM MgCl$_2$, 1 pmol each of M13F and GW2 primers in a total volume of 50 µl. After 25 cycles (20s at 94 ^{o}C, 20s at 55 ^{o}C and 3 minutes at 72 ^{o}C), an aliquot was run on an agarose gel. 1 µl of this PCR product was then used as the template for another round of PCR using the Form I Alfreider primers to detect the presence of Form I RubisCO large subunit genes. 1 µl of supernatant from the initial bead capture step was also used as a template to check for capture of all RubisCO-containing fragments of DNA.

The captured DNA fragments amplified with Taq (using the M13F and GW2 primers) were cloned in to pGEM EasyT vector (Promega). Individual colonies were grown overnight, the vector extracted via miniprep, and the insert sequenced in order to determine the context of RubisCO sequences.

Selection. Libraries in pVK101 or pRPS-GW that had at least one detectable RubisCO gene were transferred to I/II- via conjugation as described in Chapter 1 – Materials and Methods. After 24 hours on non-selective solid media, the mating mixture was scraped

off the plate and suspended in 1 ml sterile OM. 40 μl was spread on PYE containing

rifampicin and tetracycline to verify the success of the mating procedure. The remainder

was used to inoculate 100 ml of liquid PYE containing rifampicin and tetracycline. After

48 hours, 1 ml of this media was used to inoculate a fresh 100 ml of PYE containing

rifampicin and tetracycline to ensure no *E. coli* was present in the media.

An aliquot of the complemented I/II- was saved as a glycerol stock at -80 $^{\circ}$C. Other

aliquots were collected via centrifuge and washed in OM before being used to inoculate

OM-malate, OM-butyrate or OM liquid or solid media. Liquid media was inoculated at

an approximate starting OD660 of 0.05.

Tubes of OM-malate or OM-butyrate liquid media that showed evidence of growth after

several days were used as a source of inoculation for OM-malate, OM-butyrate and OM

plates in order to isolate individual colonies. 1 ml of photoheterotrophically grown

culture was harvested by centrifugation and washed 2X in sterile OM before being spread

on solid media. *Rb. capsulatus* I/II- complemented with the *Rs.rubrum* RubisCO gene

served as a positive control throughout.

Colonies from plates incubated under photoheterotrophic or photoautotrophic conditions

were used to inoculate selective PYE media. Plasmids were isolated and the size of the

insert was verified on an agarose gel following digestion of the purified plasmid with

appropriate restriction enzymes. The plasmids with a demonstrable insert were then

sequenced.

cDNA libraries. Whole cell RNA was isolated from actively growing enrichments using

the Trizol reagent (Invitrogen) according to the manufacturer's protocol. Cells were

lysed directly with Trizol, either after a brief centrifugation to collect a cell pellet, or by

pipetting Trizol directly over the layer of cells that had attached to the glass surface of the

bottle. For direct biofilm lysis, 8 ml of Trizol was used for the complete inner surface of

a 500 ml glass bottle. No more than 10 minutes passed between removal of the

enrichment flask from an anaerobic environment and placement of the cell sample on ice.

At the end of the RNA isolation procedure, whole cell RNA was present as dried pellets

in nuclease microcentrifuge tubes. The tubes with dried RNA were stored at -20 ^{o}C.

For the next step of library creation, a tube containing a dried pellet of whole-cell RNA

was removed from the freezer and the pellet resuspended in 100 µl ultra-pure nuclease-

free water. 80 U of RNAse OUT RNAse Inhibitor (Invitrogen) was added immediately to

the resuspension solution and the mix was incubated at 55 ^{o}C for 15 minutes. The RNA

concentration and purity was assessed using a Nanodrop Spectrophotometer. 3 µl of the

resuspension was loaded on to a non-denaturing agarose gel (0.8%, 1X TAE) in order to

check RNA quality by observing the bands corresponding to 16S and 60S rRNA.

Immediately after resuspension, the rRNA was subject to digestion with 10 U of DNAse I

(Roche) for 15 minutes. The RNA was then purified using an RNeasy Mini Kit (Qiagen).

The RNA was eluted from the Mini Kit column with two passes of 20 µl of nuclease free

water. Immediately following the second elution, 40 U RNase OUT (Invitrogen) was

added to the clean RNA. The RNA concentration was checked again using the

Nanodrop. 9 µg of RNA was then used as the starting material for the MICROBExpress

Bacterial mRNA Kit. The MICROBExpress Kit relies on affinity purification to strip

16S and 60S rRNA from solution, leaving only small RNAs. Following completion of

the kit protocol, the remaining RNA was resuspended in 20 µl of nuclease free water with

40 U of RNAseOUT and the concentration of RNA was checked again using the

Nanodrop.

An aliquot of (presumed) mRNA was incubated with 5 U *E. coli* Poly(A) Polymerase

(NEB) for 15 minutes at 37 ^{o}C in the presence of 1 mM ATP.

30 ng of RNA from several different treatments were then used as the template for

reverse transcription using the SuperScript III First Strand Synthesis System (Invitrogen).

The templates were 1) Whole-cell DNAse-treated RNA; 2) RNA with the 16S and 60S

rRNA removed; and 3) RNA with rRNA removed and 3' polyadenylated. The kit

protocol for RT was followed exactly. The primers used for RT with templates 1 and 2

above were random hexamers (N_6) supplied with the SuperScript III kit. Two RT

reactions were performed with template 3. 3A used 50 ng N_6 primers and 3B 50 ng

polyT primers (T_{10}). The RT enzyme was then inactivated by a 15 minute incubation at

75 ^{o}C.

Following the RT reaction, the RT mixtures were incubated with 10 U of RNAseH at 37

^{o}C for 15 minutes in order to destroy the RNA portion of the heteroduplex molecules. 2

μl of each reaction was then used as a template in a 50 μl PCR reaction. In addition to the 4 RT reactions detailed in the preceding paragraph (1, 2, 3A and 3B), whole cell RNA prior to DNAse I treatment was used as template 5. One set of PCR reactions was performed using the Alfreider Form I RubisCO primers, using the conditions previously described. Another set of PCR reactions used primers targeting conserved sites in bacterial 16S rDNA. The 63F (CAGGCCTAACACATGCAAGTC) and 1387R (GGGCGGWGTGTACAAGGC) primers were taken from Marchesi *et al* (1998). The 16S rDNA PCR used a 55 ^{o}C annealing temperature and an extension time of 90 s. All reactions used Taq polymerase (Invitrogen). For the Form I reaction, 10 ng of pUC19 containing *Synechococcus* 6301 RubisCO gene was the positive control. For the 16S rDNA reaction, 10 ng of a genomic DNA preparation was used as the positive control.

Following the PCR screen, 10 μl of each PCR reaction was run on an agarose gel.

Results and Discussion

Enrichment Cultures and Genomic DNA Isolation. Several of the enrichments were done years prior to DNA extraction, with small aliquots saved at -20 ^{o}C. Thus, each pool of DNA yielded at most 10 μg of DNA. This would be more than sufficient to create a library of DNA with a working protocol, but it was generally too little for the extensive troubleshooting that was required. Too often, a pool of gDNA would be exhausted before a working protocol could be completed.

The initial PCR screen of uncloned genomic DNA using form I conserved-site primers resulted in a library of 600 bp RubisCO gene fragments of moderate diversity. Of 20 fragments sequenced after being cloned in to a TOPO library, the only RubisCO sequences detected were *Hydrogenophaga pseudoflava, Rhodospirillum centenum* and *Rhodobacter blasticus*. Twelve of those twenty sequences had 90+% identity to the DNA sequence of *H. pseudoflava cbbL*. It is tempting to take this as an indication that *Hydrogenophaga* was the most abundant autotroph in the enrichment. There are many factors, however, that could account for the observed abundance of RubisCO genes from any particular strain. For one, the primers used for amplification are degenerate and it may be that the *H. pseudoflava cbbL* gene simply offered the best target for amplification. Another potential source for bias is the well-described bias (Kim and Bae 2011; Rajendhran and Gunasekaran 2011; Teske and Sorensen 2008; Zhang *et al* 2006) in amplification seen in all PCR, including Taq-based reactions. GC-rich sequences, in particular, are poorly amplified by Taq DNA polymerase and many autotrophic bacteria, including many of the purple non-sulfur bacteria, possess 65-70% GC content. Finally, the lysis used simply lysozyme and SDS to break the cells, and thus the genomic DNA used as a template could be heavily biased towards the most easily lysed cells.

Each of these factors could be controlled for if the goal of this experiment were to form a representative picture of the community of bacteria from a particular environment. That was not the goal, however. Rather, as stated in the introduction, the goal here was to demonstrate the feasibility of using a selection system to isolate RubisCO from a pool of genomic DNA. The PCR screen was used to track the progress of the selection system at

each stage, and at the most basic level, to verify that there was at least one RubisCO gene to find in the original pool of DNA. In that sense, the PCR screen succeeded admirably as several early attempts at library construction were shown to have omitted all of the RubisCO sequences detected in the original pool of gDNA.

The enrichment protocol also explains the disappointingly poor diversity of sequences the code for RubisCO that were detected. Again, the goal of this experiment was to develop a process, rather than to survey the total diversity of an environment. With that goal in mind, a less diverse pool of sequences was useful as it allowed for specific genes to be detected with sequence –specific primers. It is troubling though, that even with a pool of genomic DNA highly enriched for genomes containing RubisCO sequences there were still no positive results. There was not even verification that full-length RubisCO ORFs had been cloned from the genomic DNA.

Significantly, wild-type *Rb. capsulatus cbbL* and *cbbM* are easily amplified with the form I and form II primer sets, respectively, used for this study. It was verified that there was no PCR product obtained from a plasmid pool isolated from strain I/II- complemented with empty vector. In other words, the deletions of *cbbL* and *cbbM* were complete enough that the primers employed to amplify RubisCO did not give a false positive result.

Using the Form I RubisCO PCR as a screen, the attempts at direct cloning in to pRPS-MCS3 and pRPS-MCSB met with little success. A ligation of 100 ng digested and SAP-

treated pRPS-MCS3 with 200 ng digested gDNA often yielded only 10-20 total transformants. The only direct ligation of pRPS-MCS3 that met with any success was accidental recircularization of the digested plasmid (due to expired SAP or insufficient digestion time).

Direct ligation in to pVK101 met with slightly more success in that a form I RubisCO sequence (93% identity to *H. pseudoflava cbbL*) was detected in a pool of pVK101 library. The RubisCO sequence was detectable in libraries isolated from *E. coli* as well as PYE-grown I/II-. That library, however, did not complement phototrophic growth of I/II- on solid minimal media or in liquid minimal media with 20% CO_2. Growth was observed on butyrate-bicarbonate media and malate-minimal media, but none of the isolated colonies was able to grow autotrophically or had a RubisCO gene detectable by PCR. Further, and more significantly, none of the growth in liquid media evinced detectable RubisCO activity despite assays of between two and forty mg of protein from lysed cells. Finally, only five of ten colonies isolated from photoheterotrophic media yielded any pVK101 plasmid, and none of those had inserts. Thus, it seems apparent that I/II- rapidly develops photoheterotrophic competent (PHC) mutations. This propensity for PHC mutants rendered photoheterotrophic conditions useless as a selection mechanism.

Since there was a detectable RubisCO sequence in the pVK101 library, attempts were made to identify the context of the RubisCO fragment. It was possible that an incomplete RubisCO gene had been cloned, and thus no full-length RubisCO was being expressed. It was also possible that a full-length RubisCO gene had been cloned in to the library, but it

was not expressing due to a lack of a promoter sequence recognized by *Rb. capsulatus*.

All attempts to amplify the area near the fragment were fruitless. Both Taq and Phusion were used with varying concentrations of dimethyl sulfoxide (DMSO) to aid amplification of any GC-rich regions. Primers complementary to a region within the *H. pseudoflava cbbL* gene were matched with primers flanking the cloning region, as well as primers annealing within the kanamycin resistance gene.

Adding to the challenges of working with pVK101 library was the low copy number of the vector. Minipreps of overnight cultures yielded a scant 10 ng of vector – sufficient for PCR template, but too little for sequencing or restriction digest. Thus, each isolation of pVK101 plasmid required 100 ml cultures processed via midiprep in an expensive, hours-long process. The transfer efficiency of the plasmid during conjugation was not determined, either. Given these difficulties, the decision was made to explore alternate vectors.

The pGNS-BAC vector appeared to be a good alternate in that it was able to accept large inserts – up to several hundred kb – and it could be isolated from small culture volumes using the arabinose-inducible copy control feature. Crucially, for complementation studies in I/II-, pGNS-BAC has the RK2 origin of replication which allows plasmid replication in a broad range of hosts, including *E. coli* and *Rb. capsulatus*. Subsequent to conjugation, however, I/II- was unable to grow on PYE with as little as 2 µg/ml of chloramphenicol or gentamycin. Simultaneous mating procedures using Top10 *E. coli*

containing pRPS-MCS3 indicated that the media and helper strain were not at fault. Either the pGNS-BAC vector was failing to transfer to *Rb. capsulatus* or the vector was transferring but the selection markers did not confer resistance.

Thus, the chloramphenicol gene from pGNS-BAC was PCR amplified, along with several hundred bases upstream of the transcription start site. The sequence of the cloned region was verified subsequent to cloning in the pCR8 TOPO vector. Resistance to chloramphenicol was conferred on the host strain of *E. coli* by the pCR8-CAM vector. The pRPS-CAM vector created with pCR8-CAM as the donor also conferred chloramphenicol resistance on *E. coli*. The recipient *Rb. capsulatus*, however, was still sensitive to the antibiotic even at the lowest concentration. A thorough review of the limited literature on antibiotic resistance in *Rb. capsulatus* provided no reports of chloramphenicol- or gentamycin-based selection.

For future work, the pGNS-BAC vector could still be used with modest modifications. At a minimum, the gentamycin or chloramphenicol resistance cassettes would need to be exchanged for a tetracycline resistance cassette. There are relatively few selective agents known to be effective against *Rb. capsulatus*, and of those kanamycin and spectinomycin are already incorporated in the genome of I/II-.

Additionally, the cbb-promoter region from pRPS-MCS3 could be added just upstream of the cloning region of pGNS-BAC so that an inducible promoter would be available to drive expression of at least relatively short insertions. The promoter would be largely

irrelevant for large insertion libraries, however, since it could affect, at most, the few thousand bases closest to the cloning site.

Finally, it could be wise to determine the efficiency of conjugation with a vector as large as pGNS. Although the original description describes its use in creating a large-insert library (Kakirde *et al* 2011), there is no discussion of conjugation efficiency. Even if the vector can be made to function in *Rb. capsulatus*, poor conjugation efficiency could doom a selection scheme that depends on cloning significant portions of a genomic DNA sample.

Gateway Cloning in pRPS-GW. After repeated attempts to clone directly in to pRPS-MCS3, the Gateway destination vector pRPS-GW was created. See Figure 18 for an illustration of the vector.

Gateway technology has been described in detail elsewhere (Hartley *et al* 2000). Here, it was used to circumvent the difficult proposition of cloning directly in to pRPS-MCS3. The Gateway technology relies on site-specific recombinase derived from lambda bacteriophage. The recombinase is active in vitro where it catalyzes recombination between two sets of unique DNA sequences – L1 with L2 and R1 with R2. L1 and L2 flank the cloning site of the donor plasmid, and R1 and R2 flank the *ccdB* gene in the recipient plasmid. In a successful recombination reaction, the *ccdB* gene is replaced by the DNA between L1 and L2 in the donor plasmid. The key is that no digestion or ligation is required to prepare inserts for cloning. Unreacted recipient vector will still be

able to transform *E. coli*, but the host will simply not grow as the *ccdB* gene confers a lethal phenotype. In fact, the positive control recombination reaction provided with the LR Gateway Kit demonstrated that recombination with pRPS-GW is orders of magnitude more efficient than ligase-based cloning.

Another advantage of the Gateway approach is that the donor vector is generally high copy number and thus much more facile for molecular biology than pRPS-MCS3. The two donor vectors used for this study, pCR8/GW/TOPO and pENTR3C, possess the pUC/colE1 origin of replication and are thus present in very high copy numbers per cell. A simple miniprep from 1.5 ml of overnight culture provides sufficient material for screening a library. This is opposed to the need for a day-long midiprep procedure necessary for obtaining enough pRPS-MCS3 vector for even a simple restriction digest screen.

So far as cloning in to an entry vector went, several libraries of genomic DNA were prepared using the method of partial digest/blunting/A-tailing/TOPO-cloning in to pCR8. PCR-based screens using the form I conserved-site primers revealed that at least part of a RubisCO large subunit gene was cloned in each of several libraries. After the LR recombinase reaction, PCR screening revealed that at least one RubisCO fragment could be detected in the pRPS-GW libraries. There was, however, no evidence of complementation in SBI/II- under photoautotrophic conditions, whether on solid media or in liquid, from any of the pRPS-GW libraries prepared using this method.

As part of the quality control protocol, the product of the Form I PCR reaction was cloned from each screened library and ten clones were sequenced. In different libraries, the sequencing reaction revealed RubisCO from several autotrophic bacteria were present. Genes very close to the RubisCO from *Rs.centenum* (80% identity) and *Rhodobacter blasticus* (99% identity) and *Hydrogenophaga pseudoflava* (89% identity) were all detectable in different pRPS-GW libraries isolated from I/II-. This sequencing also ensured that the PCR screen was not simply picking up false positives or amplifying non-specific sections of genomic DNA.

Context. As noted above, we were able to amplify a fragment of the large subunit of RubisCO using the Alfreider conserved site primers in most of the libraries we constructed. The primary limitation of the PCR-based screen, however, is that it amplifies only a 600 bp region of the large subunit of form I RubisCO. Thus, it is possible for a fragment of RubisCO to have been cloned, but not the entire ~2000 bp ORF of large and small subunits. The *rbcLS* region of the published sequence for *Hydrogenophaga pseudoflava* (AC U55037 position 466 to 2372), for example, has 11 recognitions sites for Sau3AI. Since the original DNA fragmentation was either random (fragmentase) or due to a partial digest, it was expected that some RubisCO genes would have escaped intact.

It is also possible that complete RubisCO coding regions may have been cloned in one or more libraries, but did not complement growth due to poor transcription or translation. There are a number of possible reasons for lack of expression, ranging from clones being

78

in the wrong orientation relative to the promoter, secondary structure halting transcription or translation, or even codon bias in a transcript. It is also possible that one or more RubisCO ORFs possessed all the features necessary for expression, but the protein simply had very low levels of activity.

The first step, however, in resolving why no complementation has been observed is to determine the context of the RubisCO fragments in the various libraries. One key observation that was unfortunately overlooked in the early stages of this product was the small size of inserts in the TOPO libraries and the very low frequency of entry vectors with any insert at all. A survey of one library constructed of gDNA cloned directly in to pCR8 TOPO revealed just one of ten vectors contained any insert at all. Further, that single insert was only ~500 bp in size – considerably smaller than the range of DNA ostensibly selected using gel purification.

An explanation for the scarcity of inserts in the TOPO libraries is not readily apparent. As mentioned in the original description of the initial host strain, T10 *E. coli* is tolerant of diverse methylation patterns, so direct cloning of gDNA should not have been an issue. It is more likely that the blunting and/or A-tailing of the gDNA fragments is at fault. Poor blunting of the genomic DNA fragments would result in a poor substrate for TOPO-based cloning. Even if the blunting reaction worked well, inefficient A-tailing of the blunted genomic DNA would result in relatively few clones since the TOPO vector requires a single 3' A for incorporation.

One possible source of difficulty in A-tailing DNA fragments is the observed bias in the frequency with which Taq adds a 3' A depending on the existing 3' base (Brownstein *et al* 1996; Magnuson *et al* 1996). According to published experiments, greater than 90% of DNA fragments ending with a 3' G are A-tailed by Taq, while <10% ending with A received an additional non-template A. For this reason RsaI was chosen for partial digestion of genomic DNA prior to direct cloning in to pCR8-TOPO since it leaves a blunt fragment with a 3' T. There is no information available on the efficiency of A-tailing with the NEBNext dA-Tailing Module.

Testing the efficiency of the A-tailing reaction would be difficult. If [32]P dATP were used as a substrate for the A-tailing reaction, the resulting radioactivity of the gDNA fragments would give an indication of the efficacy of the various (incubation with Taq, NEBNext dA-Tailing Module, Klenow fragment) methods of adding a 3' A. The cost of [32]P dATP, however, makes this assay prohibitively expensive.

Alternatively, a control plasmid such as pENTR3BB could be digested with BamHI and the large (600 bp) fragment isolated from an agarose gel. This fragment could then be blunted and A-tailed using the protocols outlined above and samples from each stage could be sequenced using a primer at least 50 bp from the 3' end. This method would only be able to discern that some of the fragments received a 3' A, but would not indicate the overall efficiency.

The kit itself did not appear to be the source of failure as TOPO cloning reactions using

the very same kit produced abundant clones of Taq-based PCR reactions. The observed success of cloning "normal" Taq-produced PCR products suggests that the linker-based protocol should have met with greater success. The premise of this approach is that custom oligonucleotides can be ordered such that they anneal to each other and leave a "sticky" overhang that is complementary to the overhang left by a partial digestion. The protocol described in the Methods section above includes a step in which the two ssDNA oligonucleotides are mixed in equimolar amounts, heated to ensure all the DNA is single stranded and then the mix is allowed to slowly return to room temperature so that the oligonucleotides can anneal to produce a dsDNA linker.

Once the annealed dsDNA linker has been ligated to both ends of the gDNA fragments, it should be possible to use those linkers as priming sites for PCR amplification. As mentioned previously, the 3' base in DNA amplification will affect the frequency with which a non-template 3' A is added. Thus, the linker sequences were chosen such that the 3' position would be a G in order to maximize the likelihood of a non-template 3' A being added by the Taq polymerase. Once again, however, it has not been possible to obtain reliable amplification of linker-ligated gDNA, let alone cloning.

Or, to be more precise, amplification has been observed, but it is uniformly of fragments too small to contain full RubisCO large and small subunits. Even when gDNA has been size-selected for 2000-6000 bp on an agarose gel prior to linker ligation, the only visible PCR products are <1000 bp. An explanation for this amplification pattern is not readily apparent, but several possibilities suggest themselves.

One possibility is that the majority of the DNA in the PCR is high in GC. If this were the case, the polymerase would be much more efficient at amplifying small fragments. If the polymerase were failing to amplify the complete fragment, no complementary 3' priming site would be created for the subsequent round of PCR, and large fragments would thus amplify at best arithmetically, rather than exponentially. In fact, the *Rh. blasticus* and *Rs. centenum* RubisCO fragments are 64 and 67 % GC, respectively. Although this is too small a sample from which to draw significant conclusions, it is suggestive that that high GC content may be hindering amplification.

More sequencing of cloned fragments of gDNA could give a better estimate of approximate GC content of the pool of DNA. Amplification in the presence of DMSO could also be used to improve amplification, as could using a more GC-tolerant polymerase such as DeepVent or Phusion (NEB).

Another possible explanation is that the pool of gDNA prior to digestion is relatively small. For the linker ligation and PCR to work, both ends of the digested DNA must have a sticky end complementary to the sticky end of the linker. The smaller the size of the gDNA fragments, the less likely it is that there will be suitably-sized fragments after digestion with a Sau3AI cut on both ends. In other words, many of the large fragments (after digestion) will have at least one end that is the result of hydrodynamic shear during the original gDNA isolation, and is thus not amenable to ligation with the linker. Smaller fragments, conversely, would be more likely to have two sticky ends resulting from

digestion. With both ends amenable to ligation with the linker, smaller fragments would be able to amplify and would come to comprise a significant portion of the population after PCR, even if they were a minority of the template population.

There are two readily apparent approaches. One is to blunt the gDNA fragments after digestion. The linkers used have a blunt and a sticky end. Thus, the linkers could still be ligated to the gDNA fragments in a uniform orientation such that a single primer could be used to amplify the linker-ligated fragments. Using blunt end ligation could reduce the overall number of fragments with linkers, since blunt-end ligation is less efficient, but it should also eliminate the (hypothetical) problem of having a digestion site on just one end of a large fragment.

If a lack of large gDNA fragments in the initial pool is the source of the amplification problems, increasing the size of the gDNA fragments prior to PCR should increase to the number of large fragments with two restriction sites. Assessing average gDNA fragment size would best be done with pulsed field gel electrophoresis (PFGE).

As mentioned previously, many of the TOPO vectors isolated subsequent to transformation did not contain an insert. This is due, in part, to there being no counter-selection for pCR8-TOPO vectors that close without an insert. According to the manufacturer, insert-free vector should be a rare occurrence. In practice, a substantial number of plasmids – 17 out of 30 in 3 different cloning attempts – contained no insert. It is possible that this problem is due to the scarcity of suitably A-tailed inserts combined

with the long (1 hour+) reaction time used for cloning. According to the manufacturer's protocol, the long reaction time is to provide optimal conditions for cloning a diverse population (i.e. library) of DNA.

The problem of pCR8-TOPO vector without inserts is compounded by the method we used in screening the library and preparing it as a template for recombination with pRPS-GW. Our initial practice of allowing the pool of transformed cells to grow overnight in a relatively large volume of media could possibly have favored cells carrying insert-free vectors if only because of the lower metabolic cost of maintaining smaller plasmids. This hypothesis would be especially true for a pUC-based plasmid such as pCR8-TOPO which is present at 100+ copies per cell.

The pCR8-TOPO vectors without insert are still able to serve as a template for recombination with the pRPS-GW complementation vector. While pRPS-GW has extremely low background (i.e. cells die when transformed with unrecombined pRPS-GW), the pCR8-TOPO vectors without gDNA inserts still possess enough sequence between the recombinase recognition sites such that the *ccdB* gene in pCR8-GW is removed entirely. Thus does the problem of empty vector propagate into a *Rb. capsulatus* background. This is especially true since the post-conjugation I/II- was grown in selective PYE in order to remove the *E. coli* donor strains before complementation was tested. So, once again there was a selection bias for I/II- with small-insert plasmids.

The switch to the pENTR3C and pENTR3BB vectors was an attempt to reduce the

84

problem of no-insert background. pENTR3C was designed as an entry vector for the Gateway system. That is, it has the L1 and L2 recombination sites flanking a *ccdB* gene. Unlike the pCR8-TOPO, vector inserts are cloned using traditional digestion and ligation. The *ccdB* gene is removed when the plasmid is digested with flanking enzyme(s). The same *ccdB* gene prevents uncut vector from contaminating the library and thorough treatment with phosphatase acts to reduce self-ligation of cut vector. The most successful method of preventing self-ligation, i.e. digestion with two restriction enzymes that produce incompatible ends, is not practical as the inserts are the product of partial digestion and hence have identical overhangs on each end. A cut vector only control for each ligation is sufficient to measure the frequency of vector-only ligation.

The initial results of this strategy were promising, as the cut-vector-only ligation controls proved to have only one or two colonies per plate. Additionally, the vector ligated to Sau3AI-digested gDNA appeared to produce a very large number of transformants. Of ten colonies screened, all ten appeared to have inserts of varying sizes. Thus far, no successful complementation has been observed.

Summary. The results of this study have thus far been inconclusive. PCR-based screens indicate that multiple libraries, constructed using direct ligation (pVK101 and pENTRBB) as well as TOPO cloning and recombination (pRPS-GW via pCR8-TOPO and pENTRBB) contain at least one RubisCO fragment. There have not, however, been any instances of complementation under autotrophic conditions. Chapter 4 looks at future directions for this project, including an improved cloning vector.

Chapter 3 – *Methanococcoides burtonii* RubisCO

Introduction

Given the vast diversity of microbial organisms and the relatively brief time that scientists have been studying RubisCO, it is evident that much of that wild diversity remains to be explored. One of the most interesting of these RubisCOs came to light thanks to the recent proliferation of whole genome sequencing projects. Although methanogens have been known to carry and express RubisCO genes for some time, *Mc.burtonii* would probably not have been chosen as a model organism in which to study RubisCO. Not only does it lack the CBB cycle, but as a psychrophile and an obligate anaerobe, it not the easiest organism to culture. Thanks to the serendipitous deposition of the genomic sequence of *Mc. burtonii*, however, ORF Mbur2322 was pulled from GenBank as part of a survey of RubisCO sequences (Tabita *et al* 2008a).

Mc. burtonii RbcL (referred to as MBR hereinafter) is an enzyme that has little sequence identity with other well-known groups of RubisCO enzymes. Until recently it appeared to be unique in its position halfway between form II and III enzymes in a variety of phylogenetic models (see Figure 5, p. 6). This intermediate position appeared to be more profound than a simple output of the tree building algorithms. Rather, MBR shows

sequence characteristics of both neighboring forms of RubisCO, as well as at least one

entirely novel feature – a 30 residue loop near the C terminus that at the time of its

discovery had no apparent homology to any other sequence in GenBank. Since then,

three other RubisCO sequences also from methanogenic archaea have been discovered in

the course of whole genome sequencing projects. *Methanohalophilus mahii* DSM5219

(genome accession number CP001994) is, like *Methanococcoides*, a halophilic member

of the Methanosarcinacaea. *Methanosalsum zhilinae* DSM 4017 is also a halophilic

member of the Methanosarcinacaea. The species was originally described as

Methanohalophilus zhilinae (Mathrani *et al* 1988), but has since been renamed (Boone *et*

al 2001) and the genome was described as *Methanosalsum* (CP002101). Finally,

Methanosaeta concilii (Barber *et al* 2011) (CP002565) also has a RubisCO sequence with

a loop structure similar to MBR. See Figure 23 for a partial alignment.

Interestingly, deleting the thirty residues that comprise the "loop" structure from the four

"MBR-like" proteins results in the same tree topology. That is, even when the loop

structure is not considered, those four proteins are grouped together in a clade midway

between form II and form III sequences (not shown).

As mentioned initially, *Mc. burtonii* is a halophilic psychrophilic methanogenic archaea

isolated from Ace Lake in Antarctica (Franzmann *et al* 1992). Growth studies indicated

that it is an acetoclastic methanogen and is not capable of autotrophic growth. Further,

the RubisCO it possesses is not part of a functioning CBB cycle and no

phosphoribulokinase (PRK), the other enzyme needed to convert the pentose phosphate

pathway to the CBB cycle, is annotated in the genome. (Goodchild *et al* 2004a; Goodchild *et al* 2005). However, *Mc. burtonii* does possess homologs of DeoA and E2B2, the key enzymes, along with RubisCO, for the archaeal AMP-recycling pathway (Sato *et al* 2007). Whole-cell proteomics studies indicated that both RubisCO and E2B2 (Allen *et al* 2009; Goodchild *et al* 2004b) are expressed at low levels under heterotrophic growth conditions. This last observation is consistent with speculation that MBR functions in an AMP-recycling pathway, rather than one essential to carbon fixation. Finally, the organization of ORFs near MBR is consistent with other methanogenic archaea in being in close proximity to genes required for processing methylamines. In *Mc. burtonii*, a 5-ORF phage integration site (Mbur2316-2320) and a hypothetical conserved protein (Mbur2321) separates MBR (Mbur2322) from a suite of ORFs related to trimethylamine metabolism.

If the physiological role of MBR seems fairly clear, the structural and biochemical characteristics are less so. Like other archaea that possess form III RubisCO, *Mc. burtonii* is obligately anaerobic (Franzmann *et al* 1992). This would indicate that MBR should be as strongly inhibited by oxygen as all other known form III RubisCOs. A closer look makes the case less clear.

In terms of sequence, MBR is most closely related to the form II RubisCO of *Rb. capsulatus*, with 40% identity. However, MBR is also 35% identical to the form III RubisCO of *Methanocaldococcus janaschii*. The most arresting feature of MBR sequence, however, appears when sequences from a number of form I, II and III

88

RubisCOs are aligned (see Figure 23) – a 30 residue stretch near the C terminal with homology only to three closely-related enzymes from *Methanohalophilus mahii, Methanosalsum zhilinae* and *Methanosaeta concilii*. When used as the input for a pblast search, the loop sequence from MBR returns no significant hits beyond the three aforementioned RubisCO sequences. A hypothetical structure of MBR, based on the (5RUB) structure of *Rs. rubrum* (form II), was obtained from Swiss-Prot (Arnold *et al* 2006; Kiefer *et al* 2009; Schwede *et al* 2003). (See Figure 24) Here, the loop structure is shown simply extending in to space on the face of the large subunit opposite the active site(s). Other than this loop, there is good agreement between the crystal structure of *Rs. rubrum* RubisCO and the predicted structure of MBR. Predictions of biochemical properties based solely on sequence are of limited use. Thus, a series of experiments were designed to test the role of the loop structure in MBR. Here, I/II- could be used to test whether MBR can complement growth under conditions requiring RubisCO. After initial trials with the recombinant wild-type enzyme indicated lower than expected oxygen inhibition for a form III RubisCO.

Between the time these studies were begun with the first published notice that MBR was worthy of closer investigation (Tabita *et al* 2008b) and their conclusion, additional information on MBR became available (Alonso *et al* 2009). This information confirmed those initially published suspicions that MBR would be more oxygen tolerant than other then-known archaeal RubisCOs. That study, however, left the mysterious loop structure untouched. Thus, the MBR loop became the focus of this set of experiments.

Figure 23 - *Methanococcoides burtonii* **RubisCO Loop in Alignment**

A ClustalW alignment of form I, II and III RubisCO protein sequences with *Mc. burtonii* RubisCO. The sequences that comprise the "loop" is evident. These 30 residues correspond to the gold loop in the model presented in Figure 24. I, II and III refers to the formof RubisCO of the corresponding sequence.
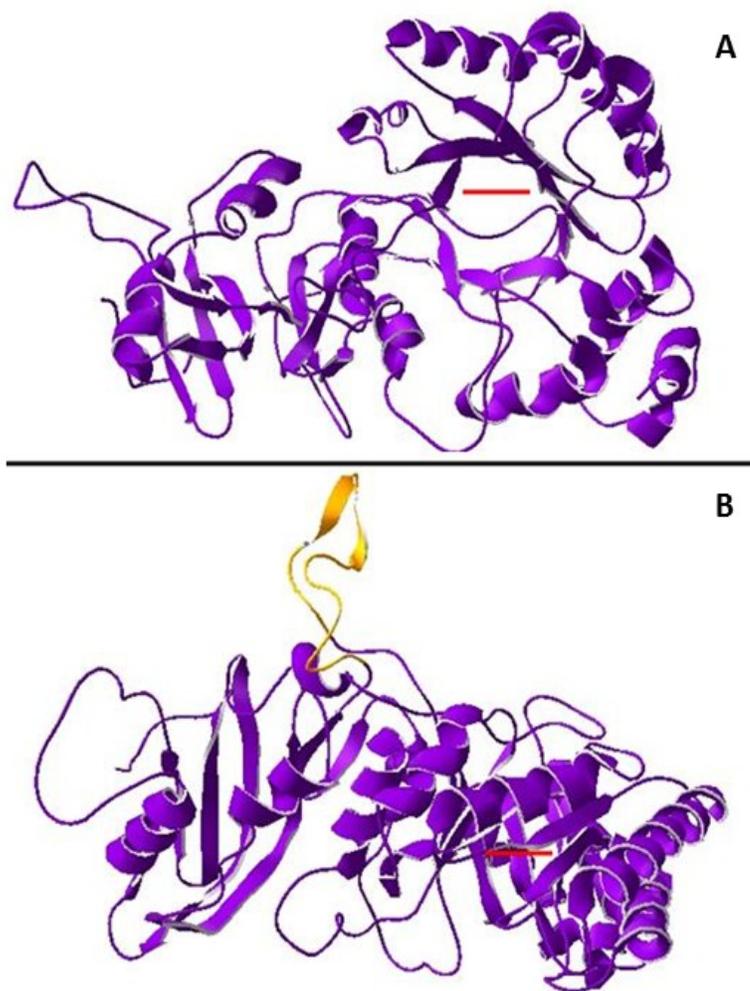
Figure 24 - *Mc. burtonii* **Model Structure**

Predicted structure of *Mc. burtonii* RbcL ABE53176, based on activated *Rhodospirillum rubrum* RubisCO crystal structure (9RUB). **A**: MBR side view. Red bar s approximate RuBP position in the active site. **B:** side view, top rotated 90° in to plane of picture. "MBR loop" in gold.

Phylogenetic tree

The phylogenetic tree presented as Figure 5 (p. 6) contains representative sequences of the four families of RubisCO. The sequences used are listed in Appendix A, with taxonomic information and gi number. The RLP sequences were set as the out-group. Both MBR and the "FIII *Methanosaeta*" group are as deeply branching as any of the other clades, although only the four proteins currently known to comprise the MBR group possess the "MBR loop" motif. Eliminating the MBR loop sequence from those four protein sequences resulted in identical topology. That is, the sequences of those proteins group together even when the loop is not taken into consideration.

The alignment upon which the tree was based was conducted in COBALT (Papadopoulos and Agarwala 2007). The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per site. The analysis involved 90 amino acid sequences. All positions containing gaps and missing data were

eliminated. There were a total of 198 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et* al 2011).

Strains and growth conditions. See chapters one and two Materials and Methods for strains and growth conditions.

Cloning. Unless otherwise noted, all restriction enzymes were supplied by NEB. Ligations were performed according to protocol using T4 DNA ligase (Invitrogen). All DNA sequencing was performed by the Plant-Molecular Genomic Facility at The Ohio State University.

Mc.burtonii genomic DNA was obtained from Dr. Kevin Sowers of the University of Maryland, Center of Marine Biotechnology. The coding region of Mbur2322 was amplified from this DNA using PCR Phusion polymerase (NEB) and primers MBRF (ATGAGTTTAATCTATGAGG) and MBRR (TTATCTATTCAAA-TAGAACTC) that correspond to the 5' and 3' regions of the MBR ORF, respectively. The product was cloned in to the pCR-BluntII-TOPO vector (Invitrogen) thereby creating pCR-MBR. The sequence of the insert was verified and the MBR ORF was then subcloned in to pET28a (Novagen) creating pET-MBR. The pET28a clone was constructed using the In-Fusion kit (Clontech) such that the N-terminal hexahistidine tag was included in the coding region. The pCR-MBR vector was used as the template to clone the gene coding for MBR and using primers MBp28_fus_F (AGCCATATGGCTAGCTTAATCTATGAGG-ACCTGGTAAAATCGC) and MBp28_fus_R (CTCGAATTCGGATCCTTATCTATT-

CAAATAGAACTCGATCGCTTC). The pET28a destination vector was prepared using

inverse PCR flanking the cloning site using the primers p28INV_fusion_F

(GGATCCGAATTCGAGCTCCGTCGA) and p28INV_fusion_R (GCTAGCCAT-

ATGGCTGCCGC). After once again verifying the sequence, BL21(DE3) was

transformed with pET-MBR alone or cotransformed with pET28-MBR plus either pG-

Tf2 (Takara) or pRARE plasmid (Novagen). The pG-Tf2 plasmid contains the

GroEL/ES and prolyl isomerase genes under a tetracycline inducible promoter. The

pRARE plasmid supplies tRNAs complementary to the rare codons AUA, AGG, AGA,

CUA, CCC, and GGA.

The gene coding for MBR was also cloned in to pRPS-MCS3 using the In-Fusion cloning

system. MBR was amplified from pCR-MBR using the Mb_noH_px_F

(TTGATATCGAATTCCTGCAGATGAGTTTAATCTATGAGGACCTGGTAAAATC

G and Mb_px_R (TGGCGGCCGCTCTAGATTATCTATTCAAATAGAACTC-

GATCGCTTCTGC). The destination vector was linearized with PstI and XbaI prior to

In-Fusion cloning reaction. After re-verification of the sequence of the insert, the pRPS-

MBR plasmid was introduced in to I/II- via conjugation using the procedure described in

Chapter 1 Materials and Methods.

Three loop deletion mutations (Figure 25) were made in pCR-MBR using inverse PCR

with Phusion polymerase and subsequent self ligation. MburMinF (GACCTTGCC-

CACCATGTCATTCTCGAAG), MburMinR (ATTAACAAGGTTGATGACATC-

CTTGTCGG); MburMedF2 (GATAGCTGGAGAGCCATGAAG) MburMedR (GGTA-

94

TCCATTATCTTTGACCATGTCTG); MburMAXF (GCCATGAAGAAGTGCT-

GCCCAATCG) and MburMAXR (TGACCATGTCTGCTCAAAGAAG). All inverse

PCR used 5' phosphorylated primers (Sigma). The resulting plasmids were named pCR-

MBRΔMax, pCR-MBRΔMid and pCR-MBRΔMin or pCR-MBR2ΔMax, pCR-

MBR2ΔMid and pCR-MBR2ΔMin, respectively. Subcloning and transformation (for

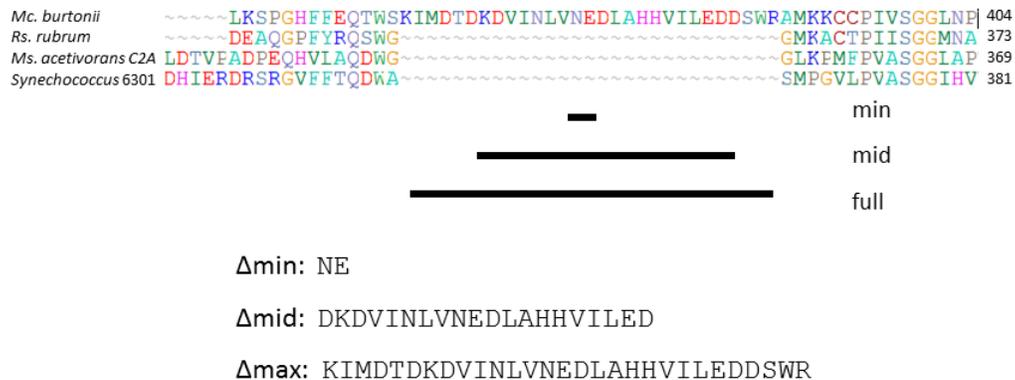pRPS clones) was performed as described for the wild-type gene.

Figure 25 - *Mc. burtonii* RubisCO Loop Deletion Mutants

An alignment is shown with form I *(Synechococcus)*, form II (*Rs. rubrum*) and form III (*Ms. acetivorans* C2A) RubisCO sequences for the sake of clarity. The sequences deleted for the three mutants are shown by the black bars below the loop. The amino acid sequence following each name refers to the residues removed from that construct.

.

Plasmids pRPS-Ma and pRPS-Rr contained the genes coding for RubisCO from

Methanosarcina acetivorans C2A and *Rs. rubrum*, respectively. Construction of both

constructs has been described previously (Finn and Tabita 2003; Singh and Tabita 2010).

Expression in <u>E. coli</u>. Initial expression tests were done in *E. coli* strain BL21(DE3).

Expression was done in LB supplemented with 10 mM $MgCl_2$ (LBM). An overnight culture in LB was used to inoculate 1 L of LBM in an 2.8 L Erlenmeyer flask. The culture was grown at 37 $^{\circ}$C and 200 rpm. Once the culture reached an OD_{600} of 0.4-0.6, IPTG was added at a concentration of 0.2 mM. The incubation temperature was lowered to ~22 $^{\circ}$C (room temperature) and shaking was reduced to 100 rpm. Expression was allowed to proceed for 6 hours. BL21(DE3) cotransformed with pRARE (Novagen) or pG-Tf2 (Takara) were induced under identical conditions, with appropriate antibiotics to maintain the plasmids, and expression levels were assessed with RubisCO activity assays of crude lysate.

Small-scale expression to obtain multiple replicates was done using the Terrific Broth Overnight Express (TBOE) autoinduction system (Novagen). 50 ml Erlenmeyer flasks of TBOE (20 ml of media each) were inoculated with individual colonies while larger batches (250 or 500 ml in 1 L or 1.8 L flasks) were inoculated with 1 ml of overnight cultures grown in LB. Cultures with the pG-Tf2 plasmid contained 5 ng/ml tetracycline in addition to the antibiotics necessary for plasmid maintenance. The flasks of TBOE were incubated with 200 rpm shaking at 37 $^{\circ}$C until visibly cloudy (OD_{600} ~0.1) when they were switched to room temperature shaking incubators. Induction was allowed to proceed overnight (~12h). Cultures were harvested by centrifugation.

Large-scale expression preparatory to Ni-affinity purification was obtained by inoculation of 500 or 1000 ml of LBM in 1.8 L or 2.8 L Erlenmeyer flasks, respectively, with 5 ml of overnight starter culture. Large-scale purification relied exclusively on *E. coli* strain

BL21(DE3) cotransformed with pRARE and pET28-MBR (or one of the loop deletion constructs). The flasks were grown at 37 °C with 200 rpm shaking until at an OD$_{600}$ of 0.4-0.5, 0.5 mM IPTG was added. After induction the temperature was reduced to room temperature (~22 °C) and shaking was reduced to 100 rpm. Expression was allowed to continue for 6 hours before the cultures were harvested by centrifugation and the pellets frozen in liquid nitrogen prior to storage at -80 °C.

Expression tests and initial activity assays. Aliquots of *E. coli* expressing MBR were harvested by centrifugation and resuspended in a range of buffers. Typically, a ~100 μg cell pellet was obtained from 1.5 ml of culture. This pellet was resuspended in 0.5 ml of buffer prior to lysis. The cell suspensions were lysed via sonication or with 10% PopCulture reagent (Novagen) supplemented with 10 U benzonase (Novagen) to degrade nucleic acids. The resulting lysate was centrifuged at 14,000 g at 4 °C for 15 minutes and the cleared supernatant removed. Specific activity assays were performed as described in Chapter 1 save for the pH and salt concentrations.

An initial set of assays was performed in the same buffer used for previously-conducted RubisCO assays (pH 8.0 in Bicine buffer), supplemented with NaCl at 0, 0.1, 0.2, 0.3, 0.4 and 0.5 M. The overall specific activity was very low (~30 nmol min^{-1} mg^{-1}), but it was sufficient to suggest that 0.1 M NaCl provided the best reaction conditions. A range of pH was then tested with 0.1 M NaCl present in each assay buffer. For pH <= 7.6, MOPS was used at 0.1 M. For pH >= 7.6, Bicine was used. The specific activity was identical for the two buffers at pH 7.6, demonstrating that the buffer composition had no effect on

activity (data not shown). Once the optimal pH was determined, the NaCl concentration assays were repeated with 0.1 M MOPS, pH 7.1 as the assay buffer. After determining the best pH and NaCl for MBR assays, a range of DTT concentrations (0.1 mM, 1 mM, 2 mM, 3 mM and 5 mM) were tested as well

Protein concentrations were measured as described in Chapter 1.

Expression levels were checked by running lysate (10 μg of protein per lane) on a 12% acrylamide SDS-PAGE gel. For comparison, 200 μl of lysis buffer was added to each pellet that remained after centrifugation to clear the lysate. 15 μl of 6X SDS-PAGE sample buffer was added to the pellets and the mixture was boiled for 10 minutes in order to denature and solubilize the pellet fraction. 5 μl of the boiled pellet mixture was loaded on to SDS-PAGE gel along with the lysate.

An/aerobic Assays of Crude Lysate. Aliquots (1.5 ml) of TBOE-grown cells were pelleted in microcentrifuge tubes. The cell pellets were transferred to an anaerobic chamber where they were resuspended in 1 ml anaerobic lysis buffer (100 mM 3-(N-morpholino)propanesulfonic acid (MOPS) pH 7.1, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT). 100 ml anaerobic PopCulture reagent (Novagen) was added to the resuspended cells along with 10 U Benzonase (Novagen) to hydrolyze nucleic acids. The lysate was transferred to 1.2 ml airtight cryogenic storage vials (Fisher Scientific) with silicone gaskets and threaded caps before being centrifuged for 10 minutes at 14,000 g. The vials were returned to the anaerobic hood and aliquots of the cleared lysate were added to 1 ml

Wheaton vials. The volume of lysate used was determined empirically and depended on the abundance of recombinant RubisCO in each lysate as well as the activity of the enzyme(s). The goal was to obtain counts at least 10-fold higher than the negative controls while consuming less than 10% of the RuBP in the reaction.

For each set of assays, two volumes of lysate were used in order to demonstrate that activity was dependent on protein concentration. Typically, the 5 or 10 μl of lysate was added to 100 μl of buffer A (100 mM MOPS-NaOH, pH 7.1, 100 mM NaCl, 12.5 mM $MgCl_2$) and each volume was prepared in triplicate. The vials were capped with butyl rubber stoppers and crimped metal seals before being removed from the anaerobic chamber. Additionally, stocks of RubisCO assay buffers, including RuBP, were prepared in Wheaton vials in the anaerobic hood using concentrated anaerobic buffers and water. Dry sodium bicarbonate for buffer "B" was transferred in to the anaerobic chamber and placed in a dry, capped Wheaton vial before removal from the chamber.

Duplicate vials were prepared aerobically from a sample of the same lysate, with all buffers being stored under standard (aerobic) lab atmosphere. The aerobic sample vials were capped with butyl rubber stoppers and crimped metal seals so that handling conditions would be identical to the anaerobic samples. The assay buffers, including bicarbonate, were prepared in a similar manner to the anaerobic buffers save that the atmosphere trapped in the vials was aerobic.

An additional set of assays for each mutant (WT, ΔMax, ΔMid, ΔMin) were prepared by

exposing uncapped assay vials containing lysate and buffer A to aerobic conditions for 15 minutes. The vials were then returned, uncapped, to the anaerobic hood (with 3 airlock cycles) before the vials were capped. This set was intended to assess the effect of oxygen exposure on the lysate before an assay under anaerobic conditions.

The vials containing lysate suspended in buffer A were partially submerged in a 30 $^{\circ}$C water bath and allowed to equilibrate for 5 minutes. During this incubation, buffer B was made by adding ice-cold buffer A (aerobic or anaerobic as appropriate) to the vial containing the dry bicarbonate. Sufficient volume was added such that the final bicarbonate concentration was 50 mM. A gas-tight Hamilton syringe was used to transfer buffers between vials. "Hot B" buffer was made by transferring an aliquot of B to another sealed vial before adding 10 µl ^{14}C bicarbonate per ml (~10 mCi/mmol) buffer B. The ^{14}C bicarbonate could not be made anaerobic as it was purchased as a liquid solution and flushing the vials with N_2 would have stripped too much CO_2. The volume of hot bicarbonate was very small, however (10 µl per ml buffer B in a total vial volume of 5 ml.) 10 µl of the hot B buffer was mixed with 1.0 ml of cold B before a 100 µl sample of the diluted hot B was added to 3 ml of EcoScint scintillation fluid (National Diagnostics). This vial was counted immediately to determine the cpm per µmol of bicarbonate in each assay.

100 µl of hot B was added to each vial with a gas tight Wheaton syringe, the vial mixed by gentle agitation before incubating in the water bath for a further 5 minutes. A gas-tight Hamilton syringe was used to add 20 µl of anaerobic RuBP (15 mM) was added to

each assay vial at 30 second intervals. Each vial was mixed by gentle agitation following RuBP addition. Exactly 6 minutes after RuBP addition, each reaction was stopped by the addition of 100 μl of propionic acid.

After the assay was complete, the stoppers were removed and vials dried in a vacuum oven over night. Once cool, the contents of the vials were resuspended in 210 μl 2N HCl. 200 μl of the resuspended assay was added to 3 ml scintillation cocktail in glass vials with screw caps, mixed by inversion, and counted on a Beckman LS-5000TD scintillation counter. Final counts were corrected for the proportion counted versus volume of HCl used to resuspend each assay.

Two negative control vials, each containing 10 μl of lysate, were treated identically to the assay vials save for the omission of RuBP. Thus, the background counts not due to RubisCO activity could be determined.

Protein concentrations in the lysate were determined via Bradford assay as described in Chapter 1.

Purification of recombinant protein. MBR and the three MBR loop deletion proteins were purified from lysed pellets of *E. coli* BL-21 large-scale expression produced as described above. Cell pellets from 500 ml or 1 L cultures were suspended in 15 ml of lysis buffer (20 mM MOPS-NaOH pH 7.1, 40 mM NaCl, 10 mM $MgCl_2$). Lysozyme was added to 1 mg/ml, as well as 40 U of RNAseA (Qiagen), 10 U of DNAse I and Triton

X100 to 0.1%. The cell suspension was incubated on ice for one hour prior to sonication. The lysate was then centrifuged for 30 minutes in a 30 ml polycarbonate bottle in a 40Ti Beckmann rotor at 4 $^{\circ}$C at 40,000 x g in an ultracentrifuge in order to clear the lysate. The lysate volume was determined and 5 M NaCl was added to bring the salt concentration to 300 mM. Finally, 0.5 M imidazole (previously adjusted to pH 7.1) was added to bring the final concentration to 20 mM.

The cleared lysate was then mixed with 1/4 volume Ni-NTA agarose (Qiagen) and incubated for one hour at 4 $^{\circ}$C with gentle agitation. Following incubation, the agarose-lysate slurry was poured in to a gravity column and allowed to drain. 40 ml of wash buffer (20 mM MOPS-NaOH pH 7.1, 300 mM NaCl, 10 mM $MgCl_2$, 40 mM imidazole) was passed over the slurry. Finally, the his-tagged protein was eluted with 5 ml elution buffer (identical to wash buffer except with 250 mM imidazole). The eluate was then transferred to 50,000 MWCO regenerated cellulose dialysis tubing (Spectra-Por) and dialyzed at 4 $^{\circ}$C against imidazole-free elution buffer supplemented with 2 mM DTT. After several hours of dialysis against the imidazole-free buffer, the tubing was moved to a beaker of assay buffer (50 mM MOPS-NaOH pH 7.1, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT) and the dialysis was allowed to continue overnight at 4 $^{\circ}$C. After dialysis, the protein was transferred a Falcon tube, glycerol was added to 20% and the protein flash frozen in aliquots in liquid nitrogen and stored at -80 $^{\circ}$C.

Samples were taken of crude lysate, cleared lysate, the loading flow-through, wash buffer after passage over the column, eluate from the Ni-affinity column and the final dialyzed

protein with glycerol. 5 µg of each protein purification was run on a 12% acrylamide

SDS-PAGE gel according to standard protocols (Coligan 1996) as an estimate of purity.

The specific activity of each preparation was determined as previously described.

*An/aerobic Assays of Purified Protein***.** Aliquots of purified protein were retrieved from -

80 ^{o}C storage and allowed to thaw on ice. The amount of protein used per assay was

determined based on the specific activity of the purified protein. The assay conditions

(temperature, gas-tight syringes, buffers, volumes of components, reaction time and

method of reading assay) were identical to the assays for lysate activity except that each

reaction vial was prepared under standard lab atmosphere conditions. The vials to be

used in anaerobic assays were kept on ice while being flushed with N_2 (USP grade) for at

least 15 min. Assays were performed using two different amounts of protein and in

triplicate for each amount of protein.

*Native PAGE gels with CABP***.** Native PAGE gels were run using a protocol modified

from (Alonso *et al* 2009), namely that the RubisCO substrate analog 2-carboxy arabinitol

bis-phosphate (CABP) would induce formation of $(L_2)_2$, $(L_2)_3$ and $(L_2)_5$ multimers. *Rs.*

rubrum RubisCO was used as a positive control for size (MW of the *Rs rubrum* dimer is

101 kDa while MW of MBR WT is 108 kDa). The three loop deletion mutants were also

run. Bi-phasic gel and buffer composition were determined from standard protocols

(Coligan 1996).

For each gel, samples were prepared with and without CABP. 80 pmol of protein (~4.4

μg of MBR WT) was added to a total volume of 10 μl. The incubation buffer was dilute loading buffer for the native PAGE gel (50 mM Tris-HCl, pH 6.8), supplemented with 20 mM $MgCl_2$ and 1 mM DTT. The CABP was procured from Dr. Sriram Satagopan and had been stored at -80 $^{\circ}$C at an acidic pH. Prior to incubation with the protein, it was converted from the lactone to the linear form by incubation at pH 9 for ~12 hours.

After addition of CABP to the appropriate tube, all the protein samples were incubated for 30 minutes at room temperature before addition of 3 μl of 4X loading buffer (50 mM Tris-HCl pH 6.8, 40% v/v glycerol, bromophenol blue). The samples were then loaded immediately onto wells in the non-denaturing gels. A constant 20 mAmps was applied across the gel for ~3 hours after which each gel was separately fixed and stained with standard Coomassie staining procedures. For the WT and each loop deletion mutant, the theoretical pI of the histidine-tagged protein was calculated using Expasy MW/pI calculator (Gasteiger *et al* 2005). WT MBR is 5.64, ΔMax = 5.87, ΔMid = 5.76, ΔMin = 5.70. For the control in this experiment, *Rs. rubrum* RubisCO, the pI is calculated as 5.75.

Growth Curves. Flasks containing 10 ml of PYE-tet were inoculated with single colonies of I/II- complemented with pRPS-MBR (WT, ΔMax, ΔMid, ΔMin), pRPS-Ma and pRPS-Rr. The flasks were incubated at 30 $^{\circ}$C with shaking for 48 hours. Samples were removed, washed with sterile OM and used to inoculate 20 ml OM in 30 ml anaerobic crimped screw-cap tubes such that the initial OD_{660} was near 0.10.

Two growth methods were used. In the bubbling method, an 18 gauge, 15 cm stainless steel luer-lock deflected point needle (Cadence Inc, Staunton, VA) was inserted through the butyl-rubber stopper. The end of the needle reached to within a few millimeters of the bottom of the growth tube. A 21-gauge needle just long enough to pierce the stopper served as a vent for the head-space. Gas mix was injected through the longer needle and allowed to bubble up through the media, providing constant mixing. The growth tubes were partially submerged in tanks of water maintained at 30 $^{\circ}$C. Banks of incandescent lights provided illumination and heat.

The static method used identical culture tubes, although only 15 ml of OM was used per tube in order to increase the headspace that served as a reservoir for the gas mix. Immediately after inoculation with starter culture and every second day for the duration of the growth curve the head space of the static culture tubes was flushed with gas mix. The flush was accomplished with a short needle connected to a gas manifold. A second needle served as the vent. The needle with the gas mixture was always inserted first and removed last such that the contents of the tube were always under positive pressure. The needles were withdrawn following each flush and each tube was shaken vigorously. Each needle was used with only one tube to prevent cross-contamination. OD_{660} was measured at regular intervals. Once the OD reached ~2.0, or no further growth was observed, the cultures were harvested by centrifugation, washed in TEMDB and frozen at -80 $^{\circ}$C.

The gas mix used was either 20% CO_2/balance H_2 or 5% CO_2/balance H_2. Each tank of gas was premixed by the Ohio State University Gas Warehouse.

Optimal expression and assay conditions.

Obtaining functional information from MBR depended on coaxing *E. coli* into expressing

functional protein. The initial tests, with pET28 in BL21(DE3), resulted in no apparent

expression (i.e. no detectable activity and no visible band on an SDS-PAGE gel), whether

in the soluble lysate or pellet fraction. The pG-Tf2 plasmid that has helped in producing

functional, soluble RubisCO in other studies (Satagopan *et al* 2009) had no discernible

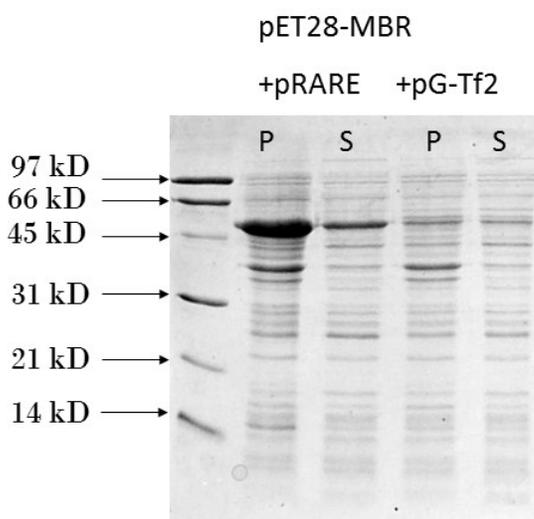effect on MBR expression, indicating that the apparent lack of protein expression after

Figure 26 - *Mc. Burtonii* RubisCO Expression in *E. coli*

12% SDS-PAGE gel of *E. coli* BL-21 lysate. MBR was expressed from the T7 promoter in pET28
with either pRARE to supply rare codons or pG-Tf2 to supply GroES/EL chaperones. Cells were
lysed with PopCulture Reagent (Novagen) and centrigud at 10,000 xG for 15 minutes. The lanes
marked "P" indicate the pellet fraction following centrifugation. The lanes marked "S" indicate the
supernatant fraction.

induction was not due to degradation of expressed but misfolded RubisCO subunits. Recombinant MBR was, however, produced in the presence of pRARE, indicating that the codon usage pattern of the archaeal protein was preventing full expression. The most likely culprits are the AGA and AGG codons for arginine. 75% (12 of 16) of the arginines in MBR use AGA or AGG, both of which are rare codons in *E. coli* (20% of arginines in *E. coli* use AGA or AGG). (See Appendix D for a chart of codon usage in *E. coli* compared to the MBR ORF.)

Further investigation was required to determine optimal assay conditions. An array of pH conditions were tested, as well as NaCl concentrations (See Figure 27). Perhaps surprisingly, given that *Mc. burtonii* growth medium contains 300 mM NaCl (Franzmann *et al* 1992), optimal activity was observed with an NaCl concentration of 100 mM. Only a slight decrease in activity was observed under low salt conditions (~10 mM Na+ due to NaOH used to adjust the buffer pH), while higher concentrations of salt resulted in a sharp decrease in activity. For 150 mM $\geq$ [NaCl] $\geq$ 10 mM, the specific activity of crude lysate in an aerobic assay was ~110 nmol CO_2 min^{-1} mg^{-1}. 200 mM and 250 mM NaCl resulted in 90 and 80 nmol CO_2 min^{-1} mg^{-1}, respectively. 500 mM NaCl resulted in a sharp decrease in activity (42 nmol CO_2 min^{-1} mg^{-1}).

As for pH, the highest activity was observed at pH 7.0-7.1 with a sharp decrease in activity under more basic conditions. Peak activity in crude lysate was 190 nmol CO_2 min^{-1} mg^{-1} at pH 7.1. Increasing the pH to just 7.2 resulted in a 60% decrease in activity (to 75 nmol CO_2 min^{-1} mg^{-1}). Activity continued to decline as pH increased, to ~ 20 nmol

CO_2 min^{-1} mg^{-1} at pH 7.9 (where the raw cpm were barely above background levels in the negative controls). The *Mc. burtonii* media is 7.0-7.2. The pH optima of other form III RubisCOs have been reported as 7.2 for *Methanocaldococcus janaschii* and *Archaeoglobus fulgidus* and 7.5 for *Methanosarcina acetivorans* (Finn and Tabita 2003). Many form II RubisCOs also show optima at near-neutral pH.

No change in activity was observed with the addition or omission of DTT (Data not shown)

An/aerobic Assays. For an/aerobic lysate assays, expression was shifted to autoinduction media. This allowed expression in small volumes, and thus coexpression of MBR in multiple flasks in the same incubator. This approach was intended to minimize variability in protein expression levels due to differences in "room temperature" during protein expression. In practice, levels of recombinant protein produced varied significantly from batch to batch. The results presented below are from the concurrent expression of 3 flasks of each recombinant protein. The results of the an/aerobic assays conducted with lysate from autoinduction cultures are presented in Figure 29
Note that MBR WT retains 40% carboxylation activity when exposed to atmospheric levels of oxygen, relative to anaerobic conditions. Since the amount of bicarbonate (50 mM) is constant under each condition, the pH of the buffers used is equal and the amount of CO_2 from standard atmosphere is negligible (~390 ppm converts to approximately 1.5 mM bicarbonate), the only difference between the two assay conditions is the availability of molecular oxygen as a substrate in the aerobic reactions.
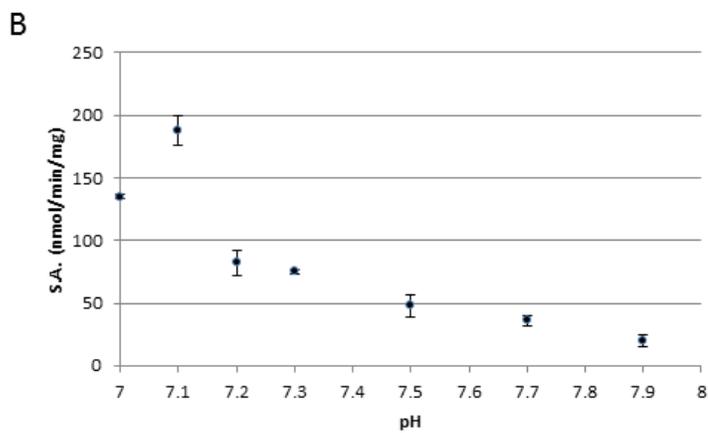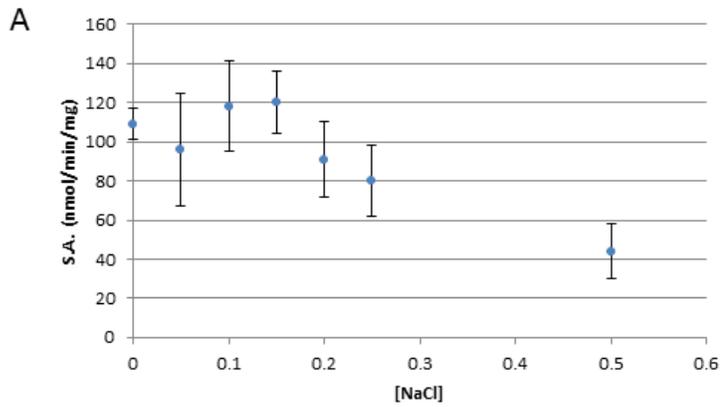
Figure 27 - pH and Salt Response of *Mc. burtonii* RubisCO

All assays were conducted with *E. coli* lysate under aerobic conditions. Each data point is the average of two independent assays, each of which was performed in triplicate. The error bars indicate standard deviation. A: Assays were performed at pH 7.1; B: Assays were performed with 0.1 M NaCl. Assay conditions were otherwise identical.
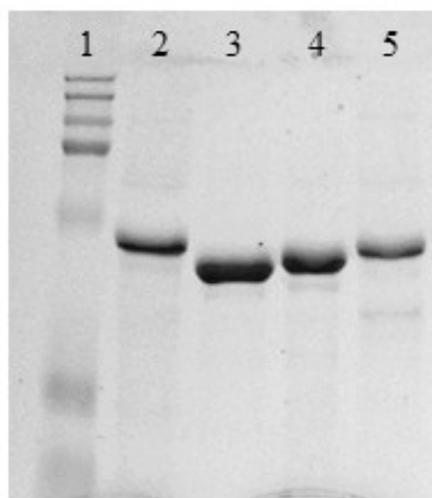
Figure 28 - Purified *Mc. Burtonii* RubisCO Protein

12% SDS-PAGE Gel, 5 µg protein per lane.
 Lane 1: Bio-Rad Kaleidoscope Prestained MW Marker
2: *Mc. burtonii* RubisCO WT
3: Max loop deletion
4: Mid-length loop deletion
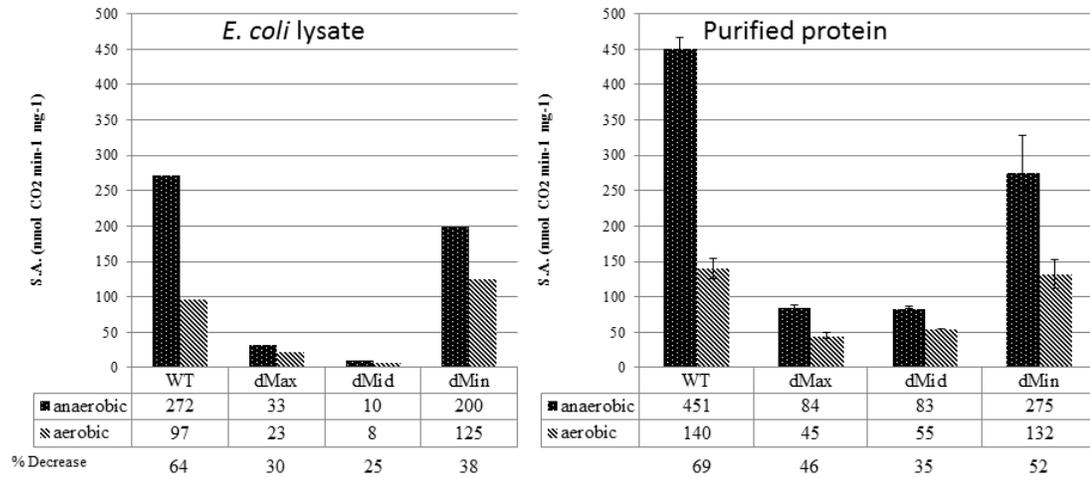5: Minimal loop deletion

Figure 29 - An/aerobic Activity of *Mc. burtonii* RubisCO

Aerobic and anaerobic assays were conducted concurrently for each mutant protein. The percent decrease refers to the change in specific activity (for carboxylation) from the anaerobic assay to the aerobic. Activities are in nmol min^{-1} mg^{-1}. Error bars indicate standard deviation.

The form II enzyme from *Rs. rubrum* retains 50% of its activity whereas, *Ms. acetivorans*, by contrast, has nearly undetectable levels of carboxylation in an aerobic assay. MBR, then, is the first RubisCO from a methanogenic archaeon to retain significant carboxylation activity in the presence of molecular oxygen.

Inhibition of MBR by oxygen does not appear to be irreversible. The set of lysate assays in which the samples were exposed to room atmosphere (21% O_2) before being made anaerobic showed no significant change in specific activity. (Data not shown.) These assays, however, were conducted without stirring the samples, and thus oxygen exposure may have been at a lower concentration than the samples used for aerobic assay (in which the lysate was exposed to standard atmosphere for a much longer time). Even for the

111

form III RubisCOs where carboxylation is strongly inhibited by molecular oxygen, however, there is no evidence of irreversible changes in the protein (Kreel and Tabita 2007).

The specific activity of the purified protein is noticeably low relative to the specific activity of the lysate (450 vs. 272 nmol min^{-1} mg^{-1}). The most likely explanation for this is instability in the enzyme in buffer without glycerol. Aliquots of enzyme removed from -80 $^{\circ}$C storage (stored in 20% glycerol) rapidly lose activity and/or noticeably precipitate if they are diluted with a glycerol-free buffer. The purification protocol required ~16 hours of dialysis, all of that conducted with glycerol-free buffer. The half-life of the enzyme with and without glycerol was not determined. Future studies could add glycerol immediately after elution of the protein from the Ni-agarose resin, and all dialysis could be done against glycerol-containing buffer.

The relatively low activity of the purified protein following purification and dialysis, as well as the observed loss of activity after incubation on ice, could also indicate that the specific activities are artificially low due to the presence of denatured or misfolded protein. If altering the dialysis conditions to include glycerol fails to stop the loss of activity, a gel filtration step immediately prior to conducting an assay could ensure that only active protein is included. Care must be taken, however, as mutations in the loop structure of MBR appear to have an effect on formation of larger-order structures (dimeric vs decameric forms – see Figure 30, below)

112

The loop structure of MBR also appears to have an effect on the degree to which the enzyme catalyzes the oxygenation reaction. All three loop deletion mutants lose less carboxylation activity under aerobic conditions than does the WT MBR protein (Figure 29, p. 111). The overall specific activity of the loop deletion mutants is lower, but the ΔMax and ΔMin deletions retain significant levels of activity. The ΔMid activity is measurable, but very low in crude lysates and the cpm recorded for ΔMid assays were typically just 2-3 fold higher than the background counts (as opposed to 10-30 fold higher for the other proteins measured).

The overall pattern seen in the lysate holds for purified protein: specific activity is WT>ΔMin>ΔMax>ΔMid. Also, the MBR WT seems to be most affected by the presence of molecular oxygen, while the three loop deletion mutants are less so. The low specific activity for the ΔMax (barely higher than for ΔMid) is more interesting in light of the growth curves for I/II- complementation presented below.

At the beginning of this study, MBR was simply a sequence in the database with a unique position between RubisCO forms II and III. The most that can be said as the study concludes is that manipulations of the loop structure do appear to affect catalysis. Surprisingly, a two residue deletion in the loop midpoint caused, relative to the wild-type protein, a very significant increase in specific activity in vitro and dramatically decreased doubling time in complemented I/II-. Deleting the full loop also resulted in an increase in specific activity and a decrease in doubling time in complemented I/II-, although the

113

effect was less pronounced. Puzzlingly, deleting just the middle half of the loop resulted in a severely compromised protein, both in terms of specific activity and doubling time in complemented I/II-. These results suggest that the loop structure that is unique to MBR and its close relatives from the Methanosarcinacaea does have a role in protein function.

The only conclusion that can be confidently drawn from these studies is the fairly trivial observation that deleting all or part of a 30-residue loop results in a less active protein. It is significant, however, that the loop does not appear to be essential for protein activity. Perhaps it is more surprising that a 30-residue interior deletion results in an enzyme with any residual activity at all, let alone being able to fold and remain in solution.

As observed in the introduction, the interaction of RubisCO with its gaseous substrates (O_2 and CO_2) appears to be influenced primarily by tertiary and quaternary structure and the subtle effects that structure imparts to the relative positioning of the active site residues. It is not difficult to imagine that deleting all or part of the MBR loop could change the interactions of previously contiguous residues. Although the loop is not close to the active site(s), it seems that the effects of these mutations might be able to propagate throughout the enzyme. It is also possible that the kinetic changes seen stem from changes in the ability of the enzyme to form higher-order structures. These changes are observed in the non-denaturing gel results presented below.

Native gels. The only extant paper on MBR (Alonso *et al* 2009) presented data showing that MBR WT formed simple dimers (L_2) in the absence of substrate, while incubation

with the RuBP analog CABP induced the formation of $(L_2)_5$ decamers. The MBR

produced for this study was isolated from an intracellular environment (*E. coli*) lacking

phosphoribulokinase (PRK), and thus without the endogenous substrate (RuBP) present

in the intracellular milieu, while the previously referenced study isolated MBR from *E.*

coli that coexpressed PRK. Further, Alonso *et al* (2009) dialyzed MBR for ~12 hours

against 20 mM EDTA prior to running their gels. Nevertheless, we find that the WT

MBR appears to form decamers with or without the presence of CABP (Figure 30).

More interestingly, the two most significant loop deletion mutants do not form decamers

at all, while the ΔMin mutant (missing only 2 residues) does.

Another noteworthy feature of this experiment is the apparently poorly defined structure

of the RubisCO loop deletion mutants incubated without CABP. For all three loop

deletions, the samples incubated in the absence of CABP showed smears, with

suggestions of dimer, possible tetramers and even hexamers. The ΔMax lane has a fairly

strong band where the dimeric form would be expected. Upon addition of CABP,

however, all three loop deletion mutants present tight, well-defined bands. Even the *Rs.*

rubrum RubisCO presented a slight change in apparent size with CABP addition. The

change in *Rs. rubrum* RubisCO is explainable by the effect that substrate binding is

known to have on RubisCO holoenzyme – namely that once substrate is bound, loop six

of the large subunit closes over the active site, blocking solvent access. The more rapidly

moving band of the plus CABP lane indicates that the protein appears to have adopted a

more closed configuration, and hence passes slightly more quickly through the gel.

It is worth noting that the attribution of certain bands to dimeric, decameric or other higher-order oligomerization is based on the PAGE gel pattern determinations made by Alonso and colleagues (2009) in which NanoESI-MS was used to interpret the oligimerization pattern of MBR incubated with RUBP. They saw a banding patter for WT MBR that is very similar to the one depicted in the upper portion of Figure 30. A gradient gel run to the pore exclusion limit would give a more precise determination of relative size of the oligomers.

Explaining the effect of CABP on the loop deletion mutants is somewhat more problematic. All that can be definitively said is that the relaxed form of the loop deletion mutants does not readily form higher-order structures following room temperature incubation. The broad smear indicates that the proteins are associating in a loose array of higher molecular weight structures. Upon addition of CABP, however, sharp bands form. For ΔMax and ΔMid, those bands are at a position that indicates simple dimers. It may be that the disarrangement of the large subunit caused by deleting all or part of the loops (as evidenced by the decrease in activity of the mutants) keeps the surfaces required for dimer formation from meeting properly. Addition of CABP would cause the shape of the enzyme to change, and in particular that of the region near the active site which sit at the interface of the large subunits.
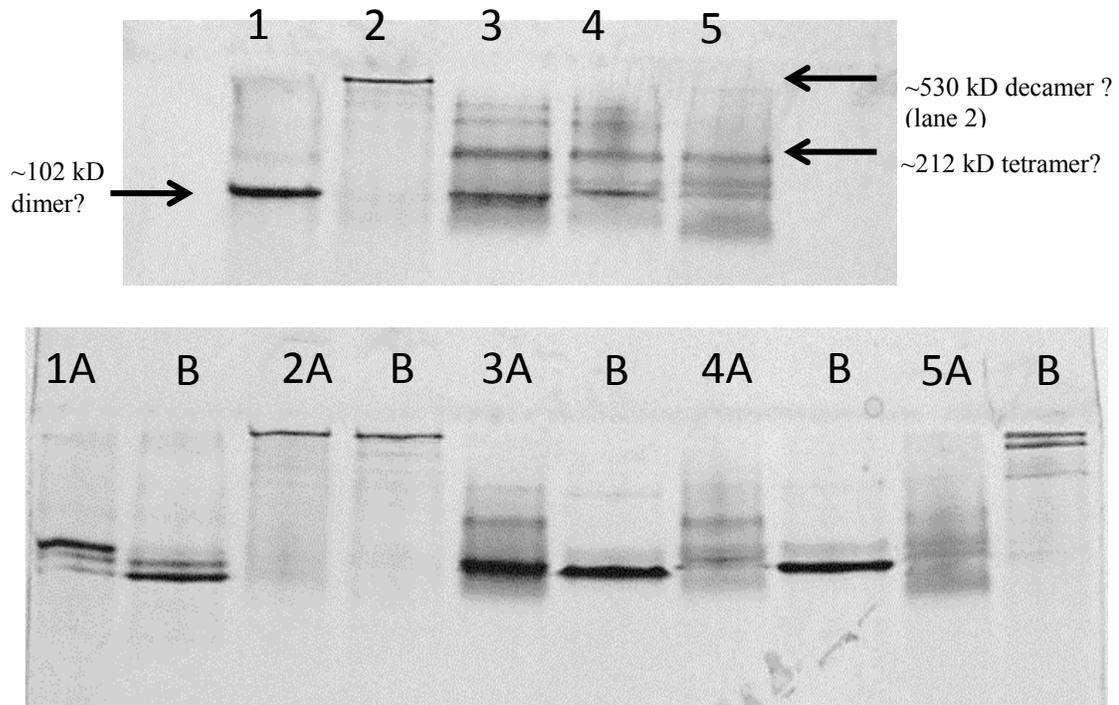
Figure 30 - Effect of Substrate Analog on *Mc. burtonii* RubisCO Structure

Upper: Samples from -80 °C glycerol stocks were thawed on ice and loaded on to gel immediately after loading buffer was added. Lane 1: *Rs. rubrum* RubisCO dimer (50.5 kDa per subunit,); 2: *Mc. burtonii* RubisCO wildtype (52.9 kDa per subunit); 3 *Mc. burton*ii RubisCO wildtype ΔMax; 4: *Mc. burtonii* RubisCO wildtype ΔMid; 5: *Mc. burtonii* RubisCO wildtype ΔMin.

Lower: 30 minute room-temperature incubation immediately prior to gel loading. Same order of proteins. Lane A was incubated with 20 mM $MgCl_2$. Lane B, each sample was incubated with 1 mM of the substrate analog 2-carboxy arabinitol bisphosphate (CABP) in addition to $MgCl_2$.

The indicated positions of multimeric forms is based on Alonso *et al* (2011). A simple native gel cannot, by itself, indicate actual molecular weights, especially of multimeric proteins. The relative positions of the bands here, however, correspond to the relationships seen in the more rigorous determination for wild-type *Mc. burtonii* RubisCO in the presence and absence of CABP.

80 pmol protein per lane (~4.2 μg *Mc. burtonii* RubisCO wildtype WT). Molecular weight differences from 6-his tags were accounted for, but resulted in very small molecular weight differences.

Both gels are two phase non-denaturing gels, 5% acrylamide stacking gel, 12% acrylamide separating gel.

In both the ΔMid and ΔMax loop deletion mutants, it appears that the ability to form decamers has been eliminated. The formation of decamers requires association of dimers on their non-active site face. The loop in MBR is projected to emerge opposite the face involved in active site formation and dimerization. In other words, it is perfectly positioned to participate in the coordination of dimers in to structures of greater complexity.The effect of CABP on the ΔMin mutant is even more puzzling since the protein changes from seemingly disordered to a decamer upon addition of CABP. The ΔMin mutant is lacking only two residues from the center of the loop, and that change does not appear sufficient to disrupt decamerization in the presence of CABP. Unlike the wild-type MBR, however, the ΔMin mutant does not form a decamer in the absence of bound substrate. Since the primary effect of CABP binding is to change the structure of the region near the active site(s) of a dimer, it would be reasonable to suspect that the lack of dimers in ΔMin – CABP is due to a disruption of the face opposite the loop. Like the other two mutants, the ΔMin mutant shows a defect in carboxylation activity relative to WT, albeit of lesser magnitude. This decrease in activity can reasonably be suspected to result from changes to active site region, and thus to the regions responsible for dimerization.

One further correlation is evident: of the WT and the three mutants, the two highest specific activities are those of forms that are able to associate as decamers (WT and ΔMin). The two mutants that do not appear to form decameric structures (ΔMid, ΔMax) have much lower specific activities. Alonso *et al* (2009) reported that there was no difference in the V_{Cmax} of dimeric versus decameric forms of the wild-type enzyme. The

difference in the assays they report, however, is just that assays were conducted with dimeric or decameric MBR, incubated for 10 minutes, and then initiated by addition of RuBP. It may be that the dimeric RubisCO they used assumed the decameric form during the incubation period, and thus all the kinetic results they report are actually for the decameric form. In any case, the loop deletion mutants described in this study are obviously different. We did not isolate dimeric MBR, and thus have no comparable data. Given the obvious structural changes in the large subunits that result from deletions in the loops, there may be a sort of synergistic disruption of activity occurring. That is, changes to the loop structure indirectly alter the structure of the active site which disrupts the ability of the protein to form stable decamers which further disrupts the structure of the active site. Of course, this is all simply speculation at this point.

Complementation. MBR WT does complement growth in *Rb. capsulatus* I/II- under photoautotrophic conditions. When bubbling gas through the cultures, the MBR WT-complemented strain supports growth of strain SBI/II- at a level intermediate between *Rs. rubrum*-complemented strains (rapid growth) and *Ms. acetivorans*-complemented strains (slow growth). (Figure 31) Although the gas bubbled through the cultures is itself anaerobic (20% CO_2/balance H_2) the bubbling method results in a periodically microaerobic environment as the culture tubes must be briefly disconnected from the gas manifold to take absorbance readings. There is no means of preventing the introduction of small volumes of air at each reading. This is relevant because, as previously observed (Finn and Tabita 2003) *Ms. acetivorans* RubisCO is strongly inhibited by molecular oxygen. Rr RubisCO, is not as strongly inhibited by molecular oxygen, although it does

119

not have a specificity factor as high as that of form I RubisCOs. In this context, then, MBR does appear to function as an intermediate between the form II and III RubisCOs in that it supports slower growth than the form II RubisCO and more rapid growth than the form III RubisCO. This pattern mirrors that of the *E. coli* lysate which showed Rr>MBR>Ma in terms of loss of activity in the presence of oxygen.

The growth curves in the "static" culture tubes present a somewhat more complicated picture. Here, the *Ms. acetivorans* RubisCO-complemented strains grow faster than the MBR-WT strains, while the Rr-complemented strains still show the fastest growth. The static tubes should have no exposure to oxygen at all, once the anaerobic gas mix has fully displaced all the initial oxygen at the beginning of the growth curve. This observation supports the hypothesis that the microaerobic environment of the bubbling tubes inhibits *Ms. acetivorans* RubisCO-complemented cells, and that it is the aerotolerant properties of MBR that allow MBR-complemented cells to grow faster than *Ms. acetivorans* RubisCO-complemented cells.

The growth curves of the loop-deletion strains are more complicated still. The one observation that can definitively be made here is that the "mystery loop" of MBR does have a role in the function of the protein. The ΔMid protein had the most significant impact in that it offers only very poor complementation for autotrophic growth. This result is in line with expectations from the purified protein studies that showed that ΔMid retained only about 13% activity relative to WT. Both the ΔMax and ΔMin mutations, however, appear to support more rapid growth than wild type MBR. This is particularly.
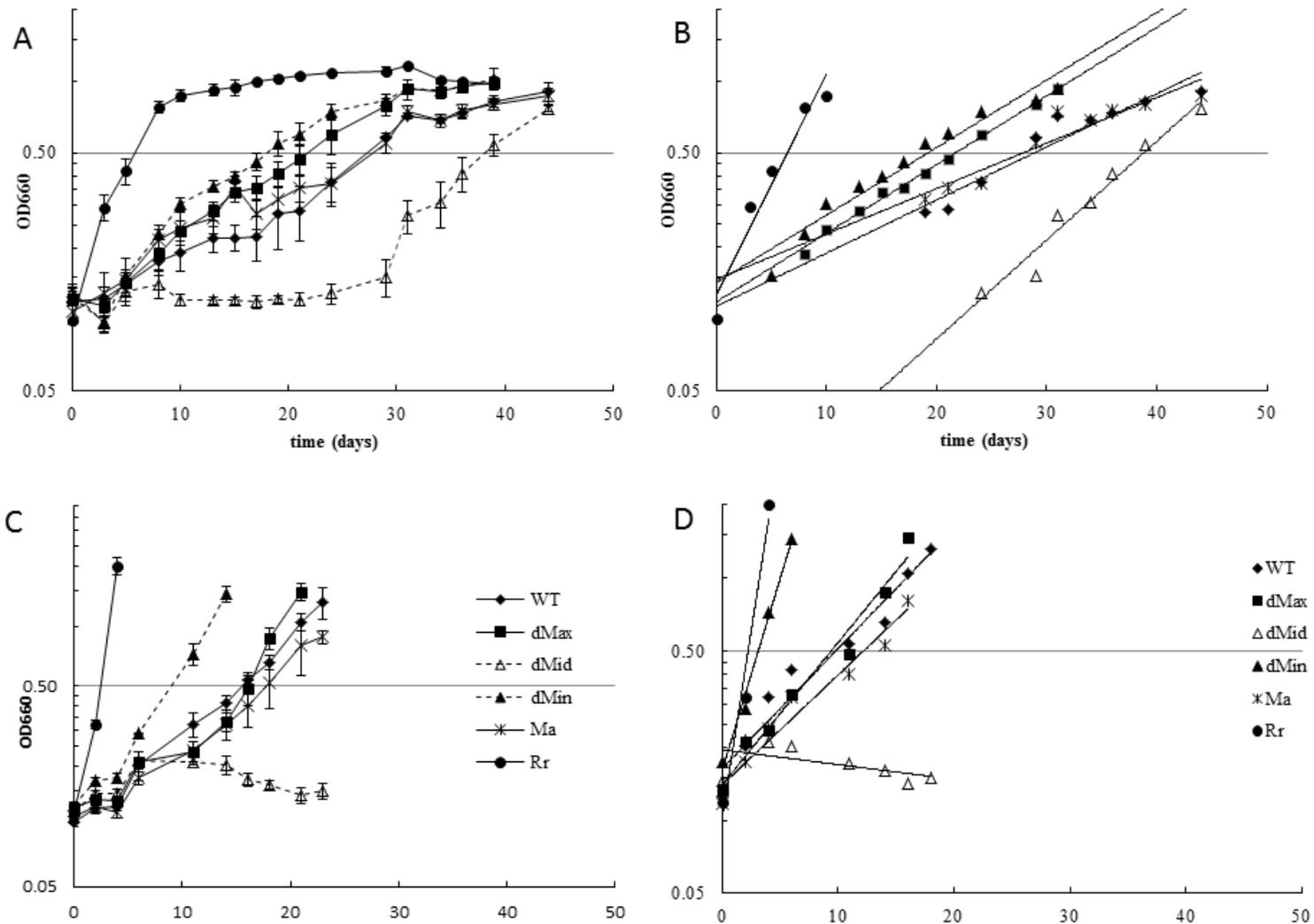
Figure 31 - *Mc. burtonii* **RubisCO Complementation in** *Rb. capsulatus* **I/II-**
Caption on next page

Figure 31 - *Mc. burtonii* **RubisCO Complementation in** *Rb. capsulatus* **I/II- (continued)**

Growth curves for SBI/II- complemented with wild type MBR (WT) and the full loop deletion (ΔMax), mid-length loop deletion (ΔMid) and minimal loop deletion (ΔMin). SBI/II- complemented with *Ms. acetivorans* C2A RubisCO (Ma) or *Rs. rubrum* RubisCO (Rr).

- A) Capped bottles with headspace of 20% CO_2/balance H_2
- B) Best-fit lines for exponential-phase growth (of A)
- C) Bottles continuously bubbled with 20% CO_2/balance H_2
- D) Best-fit lines for exponential-phase growth (of C)

All complementation was in pRPS-MCS3.

puzzling given that both mutants had significantly lower activity than wild type when purified (60% and 15% for ΔMin and ΔMax, respectively). The specific activity of the purified ΔMax mutant is nearly identical to the ΔMid mutant, but the complementation phenotype shows noticeably faster growth

Stranger still is the observation that the ΔMin mutation appears to grow more rapidly than wild-type, despite having much lower activity. No molecular oxygen is present to compete with CO_2, so the relative oxygen affinity of the WT and mutants is not an explanation for the observed phenotypes. In any case, the results of the complementation studies in I/II- were not conclusive, and could positively be termed "puzzling" given the results of the pure protein assays.

It is dangerous to go too far in speculating with so little data, but it is worth noting that the predicted structure of MBR matches that of other RubisCO structures sheet for sheet and helix for helix on both the N and C-terminal sides of the loop structure. That is, the loop appears to be a 30-residue sequence that is simply inserted into an otherwise unremarkable RubisCO protein. Removal of the entire loop results in a protein which is still capable of complementing autotrophic growth in *Rb. capsulatus* I/II-. A small change in the loop (ΔMin) likewise seems to actually improve the function of the protein. While there is, unfortunately, no clear answer to this conundrum in the present data, it would be irresponsible not to speculate.

One possible explanation lies in the presence of the hexahistidine tags that remain on the

N-terminal of the purified proteins. The complementation studies used untagged

proteins. An attempt was made to express his-tagged MBR protein in I/II-, but the

constructs were never able to support autotrophic growth. Growth on PYE-tetracycline,

however, indicated the plasmid was successfully introduced. WT MBR (without the 5'

DNA sequence coding for the hexahistidine tag) from the exact same construct (identical

sequence upstream and downstream of the coding sequence) also failed to support

growth, however. The construct that did support growth was in a slightly different

position in the vector and it remains unclear why one construct supports growth and the

other does not.

Regardless of the travails of expressing MBR in I/II-, it remains possible that the 6-his

tag affects MBR, and untagged protein may exhibit a different pattern of activity (i.e. one

that mirrors the observed complementation phenotype). This could be tested by creating

his-tagged constructs in pRPS with context identical to the untagged ORFs.

An alternate hypothesis, and one more difficult to test, is that protein stability and/or

efficiency of translation are the most important factors in determining the growth rate of

complemented I/II-. Efficiency of translation may explain the relative growth rates of

strains complemented with *Rs. rubrum* RubisCO as opposed to archaeal RubisCO. In *E.

coli*, MBR could only be expressed at all when rare codons were supplemented with the

pRARE plasmid. A quick comparison of the *Mc. burtonii* ORF codon usage (Stothard

2000) with the codon usage table for *Rb. capsulatus* (Nakamura *et al* 2000) identifies 5

codons used in MBR that are very rare in *Rhodobacter*. (See Appendix D) The rate of

transcript production in I/II-, even under autotrophic conditions, would be less than that of an ORF on a high-copy plasmid in *E. coli* with transcription driven by the T7 phage promoter. The problem of translation stalling due to a paucity of rare tRNAs would thus be less severe and some recombinant protein would be made in I/II-.

Codon bias, however, cannot explain the differences in growth among the strains complemented with mutant MBR. For this, it may make more sense to refer to the +/-CABP native gel discussed above. In photoautotrophically growing *Rb. capsulatus*, PRK will be active and hence there will be a significant intracellular concentration of RuBP. The recombinant enzyme will thus be predominantly in the forms seen in the +CABP lanes of the gel: decameric for WT and ΔMin and dimeric for ΔMax and ΔMax.

One significant difference among these proteins will be the structure of the solvent-exposed portions of the enzyme, and especially whether the loop is solvent-exposed. In the mutant with the poorest complementation phenotype (ΔMid), the holoenzyme is a dimer that leaves a damaged loop structure exposed to solvent. The ΔMax protein, by contrast, will also be dimeric, but the loop has been entirely removed such that the solvent-exposed portions of the holoenzyme will be similar to a dimeric form II enzyme such as that from *Rs. rubrum*. The WT and ΔMin holoenzyme will be assembled as decamers with the loop, presumably, protected from solvent.

This exposure of the loop is relevant because of the mechanism by which many proteins are recycled. For one, the sequence of the MBR loop has no significant homology to any

30-residue sequence in the database (aside from the loops of the three other MBR-group

archaeal RubisCOs). Although the cues by which bacterial proteins are targeted for

proteolysis are still poorly understood, one of the commonly recognized features is

denaturation (reviewed in Baneyx and Mujacic (2004); Dougan *et al* (2002)). While the

ΔMid mutant is in, as a whole, a stable fold (as evidence by its activity *in vitro*), the

damaged loop may appear to the cell's scavenging machinery as a denatured protein. It is

difficult discuss the actual role and position of the MBR loop motif without a crystal

structure to confirm the interactions.

These results, in fact, answer the implicit question posed by this fairly limited

experiment: is the loop necessary for protein function? The impetus to delete all or part

of the loop came from the observation that it was seemingly inserted in to an otherwise

ordinary protein. Certainly, the phylogenetic position of the sequence was of interest, but

loop provided a discrete and tangible focus of investigation. Deletion of the complete

loop would result in a protein structure very similar, overall, to the form II proteins which

it otherwise resembled. A complete deletion of 30 residues, however, is a radical change

and thus the two smaller deletions were included. There are quite a few possible changes

that could be made in a 30 residue stretch of amino acids. The sequences chosen were

more or less random – the mid-length deletion is 15 residues (half the loop) but taken

from the middle portion. It was possible that deleting the entire loop would result in a

"bent" protein, much like a too-tight stitch pulls fabric into a new shape rather than

simply joining edges together. The mid-length deletion, it was thought, would allow

some flexibility to remain in the position of the proximal sequences while removing the

parts of the loop, being most distant from the rest of the enzyme, that might interact with other protein moieties. The minimal loop deletion was just that – the smallest that might be imagined to have an effect, and like the mid loop deletion, taken from the center of the loop. Without a crystal structure, and with no precedent for a similar motif in other enzymes, the choice of which portion of the sequence to delete would need to be random.

The decision to delete rather than mutagenize was also due to the huge range possibilities presented by a 30-residue feature with no known function or structure. First, remove the loop to see if there is an effect, and if there is, look to narrow the search with targeted mutagenesis.

The final piece of this speculative puzzle is the personal observations of several researchers working with complementation in I/II-. There seems to be at best a weak correlation between the *in vitro* kinetic properties of a RubisCO enzyme and the growth rate of the host I/II- strain. Worse, an enzyme with a low specificity factor (*Rs. rubrum* Ω = 15) supports a faster growth rate in the presence of oxygen (i.e. under chemoautotrophic conditions) than an enzyme with a much higher specificity factor (*Synechococcus* PCC6301, Ω = 40) (Dr. Satagopan, pers. comm.).

The interior of any growing cell is a complex place and the factors that affect growth under autotrophic conditions are many and variable and poorly understood. Perhaps the most salient of these poorly understood factors is how the RubisCO supplied in *trans* interacts with the endogenous systems. It may be that the *Rs rubrum* form II RubisCO is

127

the most effective at complementation simply because it derives from a closely-related organism and hence slots in to the existing systems with the least chaos.

Consider the *Synechococcus* RubisCO, by contrast. Although both the *Synechococcus* PCC6301 RubisCO and *Rb. capsulatus* CbbLS are green-like form I enzymes, the *Synechococcus* enzyme normally occurs in a semi-crystalline array in a densely-packed intracellular structure used for CCMs (Badger and Gallagher 1987; Long *et al* 2007). *Rb. capsulatus* has no carboxysomes, so how does the 6301 RubisCO fill the interior of the cell? Without the carboxysome shell proteins, is the 6301 RubisCO randomly arrayed? Does the holoenzyme structure change subtly because it no longer has the support (i.e. inter-hexadecamer) contacts that it does in its native carboxysome? Do mutations in 6301 RubisCO that confer a growth phenotype actually interfere more with the intracellular packing of the enzyme than directly with the reaction mechanism? Or are the growth phenotypes seen in complementation due more to how "alien" the protein (or transcript!) appears to the endogenous translation and proteolysis (and/or transcription) machinery?

These questions are all beyond the scope of this project, alas. They are worth considering for future studies that attempt to infer information regarding the kinetic properties of an enzyme from the complementation phenotype in I/II-.

Chapter 4 – Future Directions

Metagenome Discovery

There are many mysteries remaining in the study of RubisCO. This document presents

what we hope are some interesting new techniques for furthering the study of RubisCO,

as well as a case study of an unusual representative. Starting with the RubisCO

complementation system in *Rb. capsulatus* I/II-, we were able to demonstrate that we

could express a RubisCO enzyme cloned from a sample of environmentally-derived

DNA.

This study has laid the groundwork for some powerful techniques for discovering novel

RubisCO genes. There are two significant, but fundamentally different, ways forward.

One relies on the abundant sequences available from metagenome shotgun sequencing

projects coupled with inexpensive DNA synthesis. The other relies on refinements of the

selection system outlined in Chapter 2 of this work.

As mentioned in the introduction to Chapter 1, the primary drawback of sequence-based

methods of finding new RubisCOs is the lack of functional information in sequence-only

detection and the time required for screening large libraries of DNA. Shotgun

sequencing of metagenomic samples offers the advantage of a high-throughput approach

to obtaining sequence information. There are, of course, many potential challenges with this approach, including the time required to assemble the sequences, the difficulty in curating the resulting sequences and identifying which sequences are worthy of deeper investigation.

The limitations of sequence assembly have been extensively covered elsewhere. The speed and quality of high-throughput sequencing is continually improving, however. Longer reads from next-generation tools offers the prospect of significantly higher quality assembled genes. The challenge of assembling long stretches of DNA without chimeric sequences should be somewhat decreased if the only goal of such a project is finding particular genes.

The importance of proper curation of genes in a library is illustrated in Tabita (Tabita *et al* 2008b) where the GOS data sets are analyzed in some detail. The original publication of this data set noted several apparent novel lineages of RubisCO genes detected in surface ocean waters. Upon closer inspection by researchers more familiar with the particulars of RubisCO genes, most of the novel lineages proved to be artifacts. The distorting effect of including very short (~60 residue) sequences in tree construction should be eliminated by focusing only on complete genes (as would be necessary if the goal is to synthesize functional RubisCO genes for further study).

The real challenge with this method would be selecting the ORFs for in vivo and in vitro

study. Although costs are continually falling for gene synthesis, a 1.4 kb form II or form III gene would cost ~$700 from commercial vendors as of 2011. While this is an order of magnitude less expensive than just a few years ago, it is still too much to allow for the synthesis of all the RubisCO genes that might be detected in a sample of seawater. Until such time as whole gene synthesis costs fall by at least an order of magnitude, researchers must find ways of prioritizing which genes of a given sequence are worth the expense of synthesis.

Determining these criteria depends, of course, on the objectives of the researcher. From a purely investigative position, one obvious criterion would be to select genes based on their position on a phylogenetic tree – i.e. to select members of a deeply divergent branch with little or no extant data. MBR is one example of this approach. Fortunately, *Mc. burtonii* had been grown in pure culture and genomic DNA was readily available. As sequencing projects expand beyond the most easily sampled environments (i.e. ocean surface waters) the diversity of RubisCO genes should increase as well.

Another interesting approach could be to screen metagenome-derived data sets for RubisCO sequences that possess particular residues at positions that have been identified as critical in previous studies. For example, in the *Synechococcus* PCC6301 RubisCO, A375 is near the active site of RubisCO, but does not directly contact the substrate(s) (Satagopan *et al* 2009). A modest change of A to V at this position was seen to have significant effects on catalysis. Although extant form I RubisCOs also possess an alanine

at the cognate position, deep sequencing of metagenomic samples could uncover form I RubisCOs with some polymorphism at this position. Of course, there would like be more than one residue that diverges from that of the 6301 RubisCO, but perhaps a pattern of substitutions could be uncovered, and the sequence data coupled to kinetic data obtained from recombinant protein obtained via expression of synthesized genes. As new generations of massively parallel sequencing technology becomes available, average read lengths are increasing and reconstruction of full genes is consequently becoming more feasible.

Looking farther into the future, structural modeling should someday be able to predict kinetic features of enzymes known only by sequence. If modeling were that nuanced, researchers could select enzymes from the database that fit a "high specificity" profile and test them in the in vivo and in vitro systems used for this study. It is the author's firm belief that we will someday reach this "holy grail" of relating sequence to structure to phenotype and the database → synthetic gene → phenotypic testing will allow for testing hypotheses relating sequence to function. In the meantime, it may be more productive to focus on extracting RubisCO genes from environments where particularly desirable properties may be expected to be found.

As long as large-scale synthesis of genes remains expensive there will be a role for phenotypic selection in discovering new RubisCOs. There are several potential means of making the methods here more efficient and effective: better vector(s) and better DNA

isolation techniques.

The most essential element for future attempts at cloning a wild RubisCO is development of a suitable vector for cloning and complementation. More time thus far has been spent in testing various vectors than in attempting complementation. As mentioned in Chapter 2, the pRPS-MCS3 vector usually used for complementation in *Rb. capsulatus* I/II-proved recalcitrant. Repeated attempts to create clone libraries in pRPS resulted in a few isolated colonies but fell far short of the thousands of clones needed to isolate a RubisCO gene from even a single genome. The pRPS-derived vectors are suitable for cloning and expression of single genes where a low rate of ligation is acceptable. Only one successful clone is required, and can be readily detected with a blue-white screen on selective/differential media. For library-efficiency cloning, pRPS-based vectors are woefully inadequate.

There are two pathways to an ideal vector, one each for screening large or small inserts. For large inserts, the best pathway forward would be a relatively straightforward modification of the pGNS-BAC vector. A simple replacement of chloramphenicol resistance with tetracycline should be sufficient to make it suitable for large-insert cloning and expression based on promoters native to the cloned DNA.
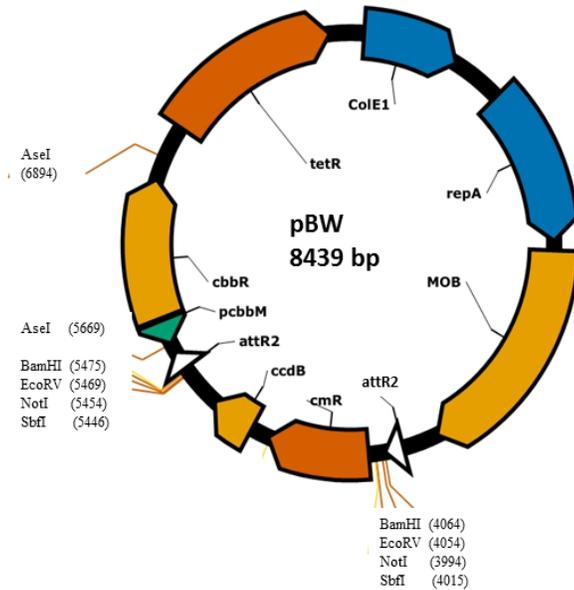
Figure 32 - Proposed Vector pBW

Clockwise from position 1: colE1 allows high copy replication in *E. coli*; repA allows replication in *Rb capsulatus*; MOB gene allows for conjugative transfer of the vector; attR1 and attR2 allow pBW to receive Gateway clones flanked by attL1 and attL2 sites; BamHI provides sites flanking the cloning region that will ligate to Sau3AI fragments; Not I and SbfI are 8-base cutters that allow screening for inserts and cloning sites for linker-enabled cloning; EcoRV provides a blunt end for blunt cloning of fragments; cm^R provide chloramphenicol resistance in *E. coli*; *ccdB* is lethal in non-resistant host strains and thus selects against vectors without inserts; *pcbb*M provides a binding site for the transcriptional regulator CbbR in an orientation that drives transcription of inserts in the cloning region; *cbbR* contains the ORF for the CbbR regulator such that transcription is derepressed in *Rb. capsulatus* only under autotrophic condition; tetR provides a selection marker for maintenance in *Rb. capsulatus*.

For smaller inserts, such as those attempted in this study, the entirely new pBW vector is proposed (Figure 32) for facile handling of small inserts. A minimal "bare bones" version is presented in Figure 33. The recombinase-based cloning system developed for this study is an unnecessary complication of what should be a straight-forward cloning system, although the Gateway sites are relatively small and do not add greatly to the complexity of the proposed vector. The pENTRBB cloning demonstrated that it is

possible to clone genomic DNA at a reasonable efficiency with low levels of background. The features this plasmid should have are an RK2 origin of replication for broad host range. A colE1 replicon adds high copy number in *E. coli* for ease of purification. A tetracycline resistance cassette would provide the necessary marker for maintenance in *Rb. capsulatus*. Alternatively, a trimethoprim cassette could provide selection even under phototrophic growth conditions. A MOB site would be needed to enable transfer via conjugation. The multiple cloning site would ideally be modeled on the pENTR MCS, with several pairs of restriction enzyme sites and sequencing primer sites flanking a *ccdB* gene to prevent uncut vector from propagating in *E. coli*. The restriction sites have been chosen to provide complementary overhangs compatible with an array of 4-base cutters. Additionally, a single 8-base cutter (SbfI) flanks all the recognition sites to enable simple screening for insert size while minimizing the probability of cutting within the inserts. Because this vector is designed for expression of small inserts, it would be useful to have the *cbb* promoter just upstream of the MCS. The *cbb* promoter allows for transcription of ORFs cloned without a native promoter, or with one that *Rb. capsulatus* does not recognize. Finally, the region between the sequencing primers and the MCS should be chosen such that no long runs of G or C will disrupt sequencing reactions.

The elements that comprise this vector were borrowed from several sources and construction would be a non-trivial undertaking using conventional means. However, *de novo* synthesis would provide an excellent opportunity to include all the elements that make for a flexible and powerful cloning platform. The most economical approach

135

would be to synthesize the "backbone" of the vector (dubbed pBWB) that would remain constant despite the host or the use intended. The backbone, in this case would extend, clockwise, from the AseI site at 6894 to the BamHI site at 4064. Additionally, the section from 5446 to 5669 (the SbfI site on the far side of the *ccdB/cmR* ORFs to the AseI site flanking the *cbbR* ORF and promoter) would need to be synthesized. Digestion of the vector with PmeI (restriction site GTTT/AAAC between the BamHI site and the SbfI site) would leave a blunt fragment for insertion of a PCR-amplified fragment containing the *ccdB* and *cmR* cassette from a Gateway destination vector. The PCR reaction could use primers ccdF CCGCATTAGGCACCCCAG and ccdR CACAACATATCCAGTCAC-TATGGTCGACC and a tm of 60 for the reaction. The sequence of pBW and pBWB are available from the author.

Likewise, the AseI site could be used as the insertion site for the *cbbR* and promoter from pRPS-MCS3, or for any other promoter desired. In the interests of keeping the vector as small as possible, the *cbbR* could be omitted entirely and replaced with the promoter region from the intended host organism, or even a constitutive promoter. For continuing studies in I/II-, the *Rb. capsulatus* SB1003 promoter sequence could be used. Transcription would then rely on the endogenous activator/repressor *cbbR*. Alternatively, the promoter sequence for the *cbb* operon from *Ralstonia eutropha* could be cloned in to the AseI site for complementation studies in that host organism. The PCR primers would vary depending on the target sequence. Each primer, though, should include a 5' overhang with the sequence of ATCGATTAAT. Adding these 10 bases to the 5' end of

the primers allows for the addition of an AseI site (A/TTAAT) to each end of the PCR

product with a four base buffer on the end to allow for a proper recognition of the site by

the restriction enzyme. It is important to remember, when using 5' overhang primers to

vary the annealing temperature in the PCR program – the first three cycles should be at

the predicted T_m for just the part of the primer that matches the template exactly. The

following ~23 cycles should use a T_m calculated based on the full length of the primer.

Since the second T_m should be higher than the first, nearly all the product should

incorporate the additional sequence. After the PCR is complete, the product can be

purified with a spin column and digested directly with AseI and then ligated in to the cut

(and phosphatased) vector. In the event of an AseI site inside the target sequence, NdeI

produces ends that are compatible with AseI. The 5' overhang would then be

ATGCCATATG. Both inserted sequences would have to be checked for the proper

orientation, since they do not use directional cloning.

Another means of improving the diversity of RubisCO genes discovered with a

phenotypic screen is to broaden the pool of DNA used to create the library. The

experiments described in Chapter 2 of this study all relied on genomic DNA obtained

from enrichments. Due to the artificial environment created, enrichment cultures

necessarily decrease the diversity of a sample. For many of the enrichments used for this

study, only a single RubisCO sequence could be detected with the conserved-site PCR

primers. This loss of diversity was acceptable as the goals of this study were to provide a

proof-of-concept for complementation with "wild" RubisCO. The long-term goal,
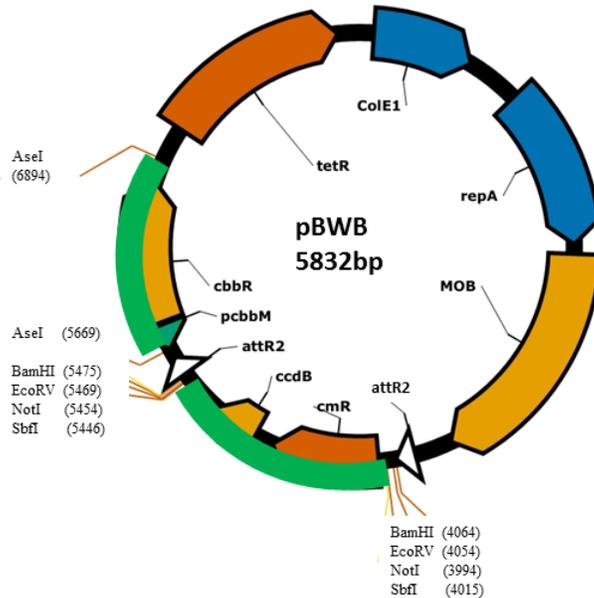
Figure 33 - Diagram of a Minimal Metagenomic Cloning Vector

The heavy green lines indicate the region that would be omitted from a minimal "bare bones" cloning vector, pBWB. This shorter version saves approximately 2600 bases. The *ccdB*, *cmR*, *cbbR* and p-*cbb* would all be omitted. The *ccdB* and *cmR* ORFs would be replaced by a PmeI restriction enzyme recognition site (GAAA/TTTC). A PCR product containing the *ccdB* and *cmR* genes could easily be inserted into the linearized vector. The cbbR and p-cbb would be replaced by a single AseI site such that any promoter sequence could be inserted upstream of the cloning region.

however, is to develop a facile means of extracting the widest possible range of functional RubisCOs from a given environment. Once the methodological kinks have been worked out, enrichments should no longer be used as the source of library DNA. Ultimately, it would be interesting to determine how effective the selection system is at identifying RubisCO genes missed by PCR screens. As in this study, the conserved-site primers could be used to screen for RubisCO genes in a pool of genomic DNA. The diversity of the sample could be determined either by extensive sequencing of PCR clones or by denaturing gradient gel electrophoresis (DGGE) run to count the diversity of PCR products. This method would probably be most effective once the

138

expression/complementation vector has been optimized. Otherwise, too few sequences will be efficiently expressed to provide a good comparison.

Methanococcoides burtonii RubisCO

The experiments described in Chapter 3 have only scratched the surface of the *Mc. burtonii* RubisCO. We have established that the loop structure in MBR indirectly affects the carboxylase activity of the protein and uncovered a strong suggestion that the loop directly impacts quaternary structure. Future investigations would address the function of the loop in greater detail as well as the role other features in MBR that allow for increased oxygen tolerance relative to other RubisCO enzymes from anaerobic archaea.

One of those features that could be targeted for further study is residue 305 in MBR. A clustalW alignment of form III RubisCOs shows that MBR305 is a valine, rather than the methionine that is conserved at the homologous position in all known form III enzymes (M295 in *Archaeoglobus fulgidus* RbcL2). M295 was recently identified as a key residue in affecting oxygen affinity in *A. fulgidus* RbcL2 (Kreel and Tabita 2007). Form II RubisCOs, by contrast, have a conserved valine or alanine at the position homologous to *Archaeoglobus* M295/MBRV305 (= *Rs. rubrum* A305). These sites are potential targets for site-directed mutagenesis that could offer corroboration of the hypotheses M295 is significant for oxygen tolerance. Changing MBR V305 to M should make the enzyme more prone to the oxygenation reaction.

The loop structure in MBR remains an open question as well. The most direct

investigation would be a simple digest with trypsin to determine whether the decameric

form of the protein protects the loop structure of MBR. The decamer form of MBR

should show slower tryptic degradation of the loop region, while the ΔMid mutant should

be rapidly cut.

Ultimately, however, the ideal would be to have a crystal structure that makes explicit the

position and interactions of the loop. With such a structure it should be possible to, first,

determine whether the loop is directly involved in forming inter-dimer contacts.

Additionally, it should be possible to mutagenize the loop such that the loop remains

intact but its (proposed) ability to direct multimerization is disrupted.

Another potential avenue for investigation would be to introduce the loop to the

analogous position in the *Rs. rubrum* form II enzyme. Since the overall structures of the

two enzymes seem to be very close, introducing the loop should have a measurable, but

not terminal, impact.

It is possible, too, that the loop could be involved in interactions with other proteins.

Determining these other interactions could be a fruitful area of further research, using a

bacterial or yeast two-hybrid system with MBR +/- "MBR loop" as bait and a library of

MBR proteins as the target. Additionally, MBR +/- "MBR loop" could be attached to a

resin and used in pull-down studies of *Mc. burtonii* lysate to see which, if any, proteins associate with MBR.

Finally, as mentioned in Chapter 3 above, there are now three other RubisCO sequences with loop structures homologous to MBR. Having these other sequences to work with gives additional points of data when attempting to determine the function of this unusual structure. Investigating this group may provide some insight, too, in to the basis for RubisCO forming higher order structures. What advantages are conferred by associations of dimers in the oxygen-free intracellular environment of the Methanosarcinacaea? What is the significance of the other deeply-branching group of RubisCO sequences – the *Methanosaeta*-group - seen in Figure 5, p. 6, there is still no example of kinetic data from any of the *Methanosaeta*-group proteins.

Although the archaeal RubisCOs must process, at best, a trivial portion of the biosphere's carbon, they do appear to be the most deeply branching groups of this critically important enzyme. Understanding the evolution of RubisCO means coming to terms with the sequence and functional diversity of what appear to be the most ancient lineages.

Appendix A: Sequences Used for Phylogenetic Trees

Table 2 - Sequences Used to Construct a Phylogenetic Tree of RubisCO

Sequence ref	Organism of origin	Abbrev. Used in Tree
gi256370937	Acidimicrobium ferrooxidans DSM 10331	Acidim_ferro
gi254169067	Aciduliprofundum boonei T469	Acidu_boon
gi114321983	Alkalilimnicola ehrlich MLHE1	Alkali_MLHE
gi288942451	Allochromatium vinosum DSM 180	Alloc_vin
gi11499228	Archaeoglobus ful gidus DSM 4304	Arch_ful
gi90420021	Aurantimonas manganoxydans SI859A1	Aurant_mangan
gi358639818	Azoarcus sp. KH32C	Azoarc_KHC
gi218511700	Bacillus subtilis	RLP_B_subt
gi157692033	Bacillus pumilus SAFR032	RLP_B_pum
gi146339298	Bradyrhizobium sp. ORS 278	Bradyr
gi339007205	Brevibacillus laterosporus LMG 15441	RLP_Brev_late
gi91777680	Burkholderia xenovorans LB400	Burk_xenov
gi129714185	Chondrophycus papillosus	Chondroph_pap
gi94310443	Cupriavidus metallidurans CH34	Cupri_meta
gi317154242	Desulfovibrio aespoeensis Aspo2	Desul_aespo
gi288931872	Ferroglobus placidus DSM 10642	Ferrog_plac
gi239826372	Geobacillus sp. WCH70	RLP_Geobac
gi292655130	Haloferax volcanii DS2	Halof_volc
gi257388805	Halomicrobium mukohataei DSM 12286	Halomi_mukoh
gi121997837	Halorhodospira halophila SL1	Halorh_halo
gi261855695	Halothiobacillus neapolitanus c2	Halot_neap
gi46359646	Hydrogenovibrio marinus	Hydr_mar
gi46359631	Hydrogenovibrio marinus	Hydr_mar
gi124027390	Hyperthermus butylicus DSM 5456	Hypert_butyl
gi305664015	Ignisphaera aggregans DSM 17230	Ignis_aggr
gi21105280	Isoetes savatieri	Iso_savat
gi83311795	Magnetospirillum magneticum AMB1	Magn_magn
gi114775651	Mariprofundus ferrooxydans PV1	Marip_ferr
gi15669420	Methanocaldococcus jannasch DSM 2661	Mcald_jann
gi282163786	Methanocella paludicola SANAE	Mcella_palu
gi91774234	Methanococcoides burtonii	Mcoc_burt
gi126179272	Methanoculleus marisnigri JR1	Mcull_maris
gi124267967	Methylibium petroleiphilum PM1	Methyl_petr
gi298674857	Methanohalobium evestigatum Z7303	Mhalob_vest
gi294495600	Methanohalophilus mahii	Mhalop_mah
gi330508790	Methanosaeta concilii	Msaeta_conc
gi357210994	Methanosaeta harundinacea 6Ac	Msaeta_harund
gi116754905	Methanosaeta thermophila PT	Msaeta_thermop
gi336476213	Methanosalsum zhilinae DSM 4017	Msals_zhil

143

gi20093339	Methanosarcina acetivorans C2A	Msarc_aceti
gi219850933	Methanosphaerula palustris E19c	Msphaer_palu
gi88603561	Methanospirillum hungatei JF1	Mspir_hung
gi76802027	Natronomonas pharaonis DSM 2160	Natronom_phar
gi11465965	Nicotiana tabacum	Nico_taba
gi92109257	Nitrobacter hamburgensis X14	Nitrob_hamb
gi88810428	Nitrococcus mobilis Nb231	Nitroc_mob
gi114330814	Nitrosomonas eutropha C91	Nitros_eut
gi334137797	Paenibacillus sp. HGF7	RLP_Paenib
gi340028807	Paracoccus sp. TRP	Paracoccus
gi414071	Picea abies	Picea
gi126695914	Prochlorococcus marinus str. MIT 9301	Proc_mari
gi14590791	Pyrococcus horikosh OT3	Pyroc_hori
gi337283520	Pyrococcus yayanos CH1	Pyroc_yaya
gi75295147	Pyropia dentata	Pyropi
gi1710033	Rhodobacter capsulatus	Rb_caps
gi77464859	Rhodobacter sphaeroides 2.4.1	Rhod_spha
gi89900184	Rhodoferax ferrireducens T118	Rhodof_ferri
gi39937699	Rhodopseudomonas palustris CGA009	Rhodops_pal
gi2978568	Riftia pachyptila endosymbiont	Rif_pachy
gi332528014	Rubrivivax benzoatilyticus JA2	Rubriv_benzo
gi150376071	Sinorhizobium medicae WSM419	Sinor_medi
gi11497536	Spinacia oleracea	Spin_ole
gi126465948	Staphylothermus marinus F1	Staphylo_mar
gi118591565	Stappia aggregata IAM 12614	Stap_aggr
gi90655413	Synechococcus PCC6301	Synech_PCC6301
gi33866250	Synechococcus sp. WH 8102	SynechWH8102
gi57642225	Thermococcus kodakarensis KOD1	Therm_kod
gi119720133	Thermofilum pendens Hrk 5	Thermof_pend
gi334144646	Thioalkalimicrobium cyclicum ALM1	Thioalk_thiocy
gi10190796	Thiobacillus denitrificans	Thiob_den
gi344342080	Thiocapsa marina 5811	Thioc_marina
gi350574213	Thiorhodovibrio sp. 970	Thiorhodov

Appendix B: pBLAST Results for ORFS on BAC15

Identity of ORFs on BAC B15 by p-blast.

ORF #	gene/putative product	organism of closest match	e value	identical residues	% Identity[a]
Carbon fixation enzymes (red)					
2	*rbcS* RubisCO small subunit	*Prochlorococcus marinus* MIT 9211	1E-44	80/107	74
		Nitrococcus mobilis	6E-41	77/109	70
3	*rbcL* RubisCO large subunit	*Synechococcus* sp. RS9917	0.0	415/469	88
		N. mobilis	0.0	408/468	87
CO$_2$ concentrating/carboxysome proteins (blue)					
5	carboxysome peptide B	*P. marinus* MIT 9303	4E-21	49/81	60
6	carboxysome peptide A	*N. mobilis*	1E-30	67/81	82
7	*csoS3* carboxysome shell polypeptide	*N. mobilis*	4E-167	291/486	59
8	*csoS2* carboxysome shell polypeptide	*N. mobilis*	0.0	359/612	58
9	*csoS1* CO2 concentrating mechanism-structural protein	*N. mobilis*	1E-42	90/93	96
Metal transport/metabolism (pink)					
14	putative arsenate reductase	*Moritella* sp	2E-36	66/112	58
Metabolic/modification enzymes (orange)					
17	succinyltransferase	*Thioalkalivibrio* sp	8E-133	242/273	88
18	desuccinylase	*Acinetobacter* sp.	3E-157	260/373	69
19	NUDIX hydrolase	*Thioalkalivibrio* sp.	1E-53	97/181	53
		N. mobilis	6E-49	95/181	52
20	LysR family transcriptional regulator	*Thioalkalivibrio* sp.	5E-106	181/314	57
		Anabaena variabilis	1E-87	153/295	51
21	cobyrinic acid diamide synthase	*Thiomicrospira crunogena*	1E-42	90/206	43
22	*tesA* esterase	*Bordetella petrii*	6E-54	102/181	56
Transport system proteins (purple)					
23	sulfate permease family protein	*Beggiatoa* sp	0.0	341/571	59
24	ABC-type Fe3+ trans. syst. periplasmic component	*Oceanospirillum*	9E-112	193/334	57
26	ABC trans. syst. ATP-binding protein	*Azoarcus* sp.	4e-104	179/353	60
27	sulfate transporter	*Arthrospira maxima*	0.0	326/544	59
28	Na+/H+ antiporter	*Saccharophagus degradans*	2e-147	269/449	59

Figure 34 - pBLAST of BAC15 ORFs

Number and color designations refer to the schematic for BAC 15 in Figure 15, p. 40

Appendix C: Strains and Vectors

Table 3 - Strains and Vectors

Vector or strain	Host Bacteria	Pertinent characteristic	Source
E. coli Top10		mcr-, recA-	Invitrogen
E. coli BL21(DE3)		expression	Novagen
E. coli ccdB		resistant to CcdB - only host for *ccdB* vectors	Invitrogen
I/II-		*Rb. capsulatus* SB1003 Δ*cbbL*/*cbbM*	(Paoli *et al* 1998)
Ch. 1			
pRPS-MCS3	Top10; I/II-	pCBB, MCS, tetR, blue-white screen, BHR	This study
pRPS-4N23	Top10; I/II-	pRPS-MCS3 with BAC4N23 *rbcLS*	This study
pRPS-B15	Top10; I/II-	pRPS-MCS3 with BAC15 *cbbLS*	This study
pRPS-6301	Top10; I/II-	pRPS-MCS3 with *Synechococcus* PCC6301 *rbcLS*	(Smith and Tabita 2003)
Ch. 2			
pGNS	Top10; I/II-	BAC, GmR, CmR, oriV, BHR	(Kakirde *et al* 2011)
pVK101	Top10; I/II-	21kb, BglII site, tetR, BHR	(Knauf and Nester 1982)
pRPS-MCSB	Top10; I/II-	pCBB, MCS w/ BglII, tetR, blue-white screen, BHR,	This study
pRPS-MCSBB	Top10; I/II-	pCBB, MCS w/ BglII and BamHI, tetR, blue-white screen, BHR	This study
pRPS-GW	*E. coli ccdB*	Gateway recipient, tetR, CmR, lethal to other strains	This study
pCDF-Duet	Top10	Cloning vector with many restriction sites	Novagen
pRPS-CAM	Top10; I/II-	pRPS with CmR from pGNS in GW site	This study
pRPS-library	Top10; I/II-	pCBB, library of inserts, BHR	This study
pCR8/GW/TOPO	TOP10	TOPO-TA cloning; CmR; Gateway entry vector	Invitrogen
pCR8-library	TOP10	CmR; post-TOPO cloning of DNA; entry vector	This study
pENTR3C	*E. coli ccdB*	kanR; NotI, EcoRI flanking ccdB; entry vector	Invitrogen
pENTRB	*E. coli ccdB*	kanR; NotI, EcoRI, BglII flanking ccdB; entry vector	This study
pENTR-library	TOP10	kanR; post-cloning; entry vector	This study
pRPS-library	Top10; I/II-	pCBB, tetR, BHR	This study
Abbriviations: BHR - Broad Host Range; pCBB - promoter for CBB operon; MCS – multiple cloning site			

Table 3 - Strains and Vectors, Continued from previous page

Vector or strain	Host Bacteria	Pertinent characteristic	Source
Ch. 3			
pUC19	Top10	ampR, MCS	Novagen
pET28a	*E. coli*	expression	Novagen
pG-Tf2	BL21	Overexpression of GroES/EL and tig, CmR	Takara
pRARE	BL21	Rare codon supplementation, CmR	Novagen
pRPS-MBR	Top10; I/II-	pRPS-MCS3 with Mc. burtonii RubisCO WT ORF	This study
pRPS-ΔMax	Top10; I/II-	pRPS-MBR with ΔMax deletion	This study
pRPS-ΔMid	Top10; I/II-	pRPS-MBR with ΔMid deletion	This study
pRPS-ΔMin	Top10; I/II-	pRPS-MBR with ΔMin deletion	This study
pRPS-Ma	Top10; I/II-	pRPS-MCS3 with *Ms. acetivorans*C2A *rbcL*	(Finn and Tabita 2003)
pRPS-Rr	Top10; I/II-	pRPS-MCS3 with *Rs. rubrum cbbM*	This study
pET28-MBR	*E. coli*	pRPS-MCS3 with *Mc. burtonii rbcL* WT	This study
pET28-ΔMax	*E. coli*	pRPS-MBR with ΔMax deletion	This study
pET28-ΔMid	*E. coli*	pRPS-MBR with ΔMid deletion	This study
pET28-ΔMin	*E. coli*	pRPS-MBR with ΔMin deletion	This study
pET28-Ma	*E. coli*	pRPS-MCS3 with *Ms. acetivorans*C2A *rbcL*	(Finn and Tabita 2003)
pET28-Rr	*E. coli*	pRPS-MCS3 with *Rs. rubrum cbbM*	This study

Appendix D: Codon Usage Charts
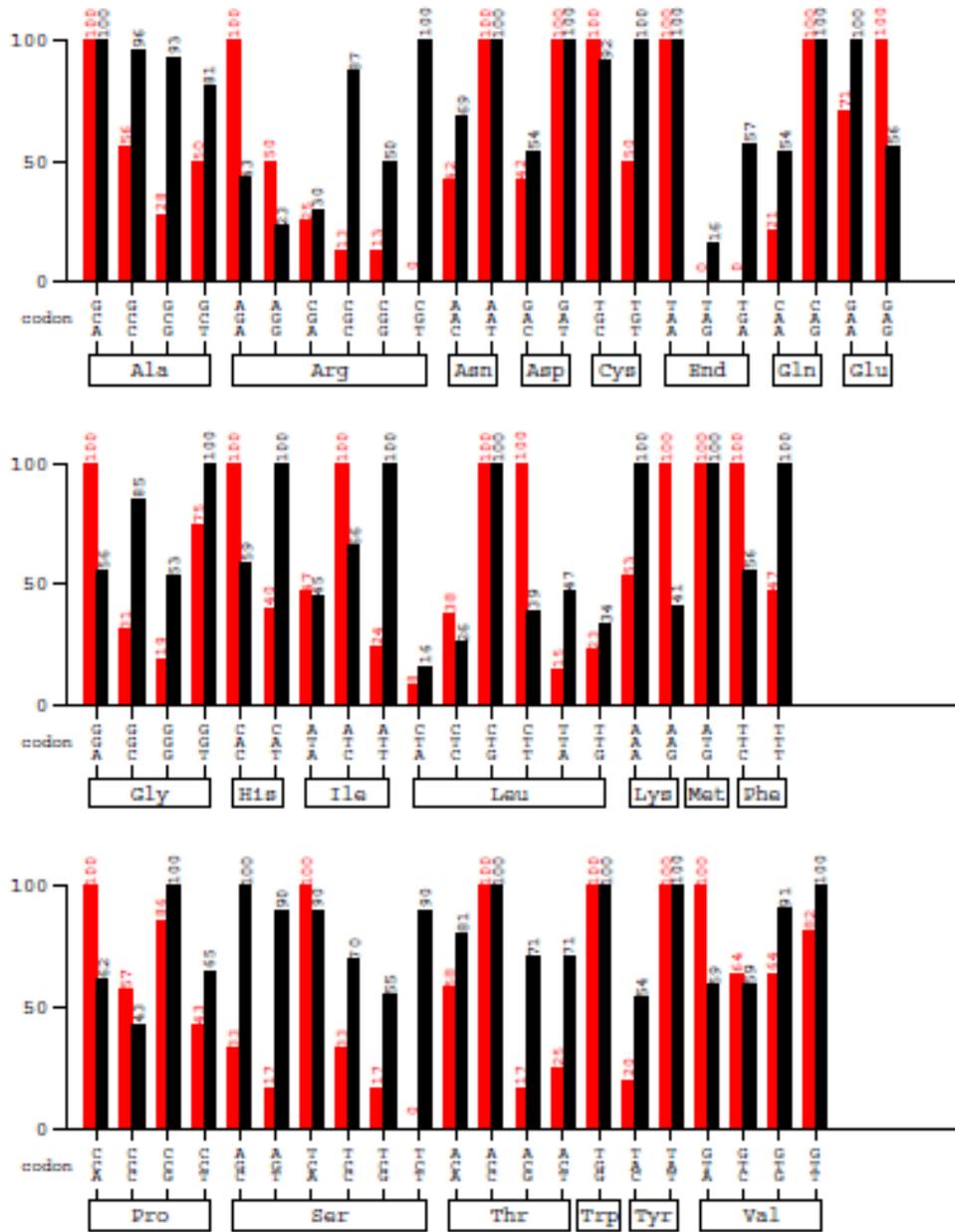
Figure 35 - Codon Usage Comparison - *E. coli* vs *Mc. burtonii*

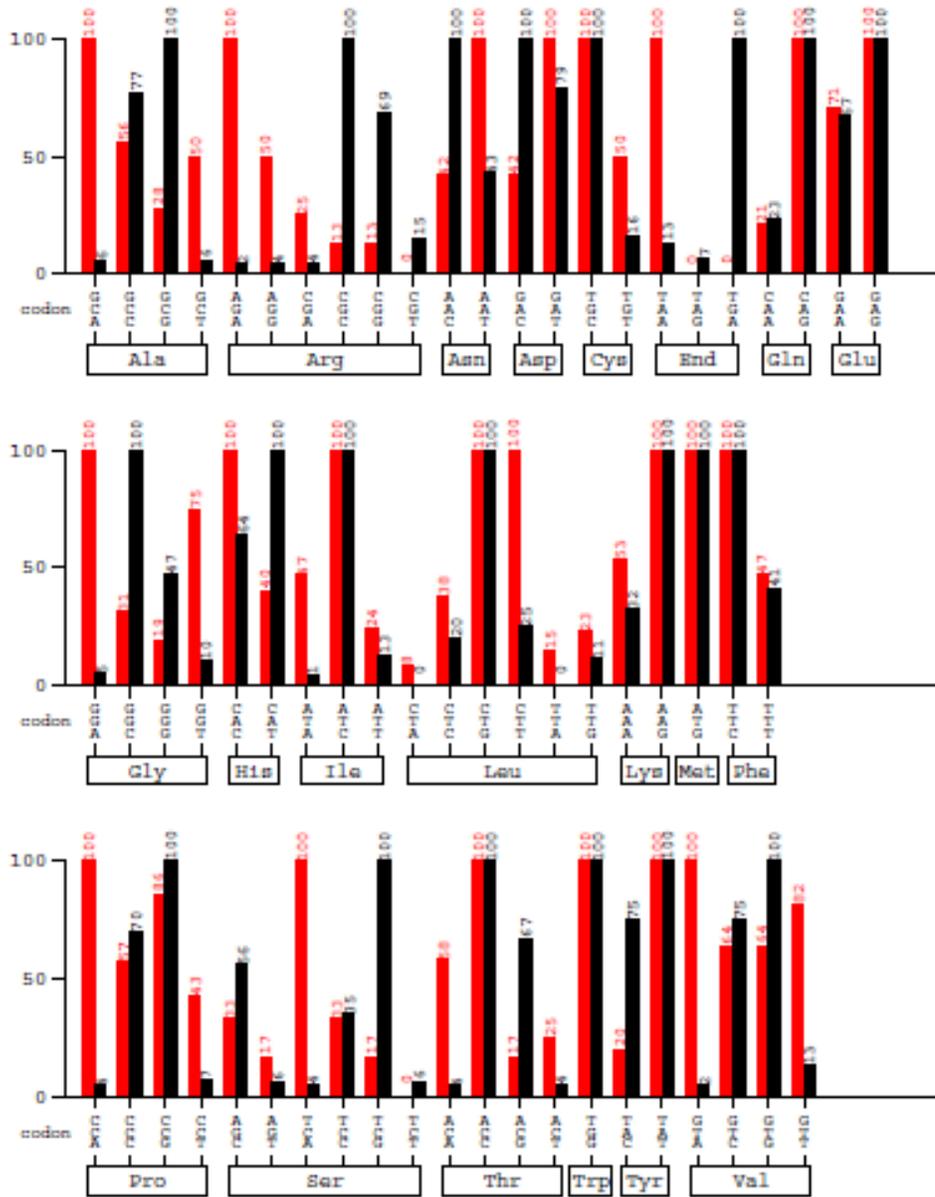Caption continues on second page.

151

Figure 36 - Codon Use Comparison - *Rb. capsulatus* vs *Mc. burtonii*

Caption continues on next page.

152

Figure 35 and Figure 36 were created using the Graphical Codon Usage Analyser, a web based application described in (Fuhrmann *et al* 2004). The program functions by comparing the frequency of codon usage in a DNA sequence to the frequency of codon usage in a host organism. For each amino acid, the frequency is set to 100, by definition, for the codon that is most frequently used for that residue. In both graphs, the codons of the sequence that codes for *Mc. burtonii* RubisCO are shown in red and the codons used by the host organism are shown in black. The conflict arises when a particular codon has a tall red bar next to a short black bar – that is, the expressed ORF codes for amino acids using codons that appear infrequently in the host organism. This data display method is derived from (Sharp and Li 1987). The host organism codon usage tables are taken from the database maintained by the Kazusa DNA Research Institute (Nakamura *et al* 2000).

Appendix E: Abbreviations

Table 4 - Abbreviations Used in the Text

2-PG	2-phosphoglycolate
3-PGA	3-phosphoglyerate
Bicine	2-(Bis(2-hydroxyethyl)amino)acetic acid/NaOH
CA	chemoautotrophic (growth conditions)
CABP	2-carboxy arabinitol bis-phosphate
CBB	Calvin-Benson-Bassham Cycle
DGGE	denaturing gradient gel electrophoresis
DK-MTP 1-P	2,3-diketo-5-methylthiopentane 1-phosphate
DMSO	dimethyl sulfoxide
FI	form I RubisCO
FII	form II RubisCO
FIII	form III RubisCO
FIV	form IV RubisCO/RLP
GAP	glyceraldehyde 3-phosphate
gDNA	genomic DNA
HK-MTP 1-P	2-hydroxy-3-keto-5-methylthiopent-1-ene 1-phosphate
K_c	Michaelis-Menten constant for carbon dioxide
K_o	Michaelis-Menten constant for oxygen
LB	Luria-Bertani media
LBM	Luria-Bertani media supplemented with 10 mM $MgCl_2$
LSU	large subunit (of RubisCO)
MBR	*Methanococcoides burtonii* RubisCO
MOPS	3-(N-morpholino)propanesulfonic acid
OM	Ormerod's media (minimal salts media for autotrophic growth in *Rb. capsulatus*)
PA	photoautotrophic (growth conditions)
PHC	photoheterotrophic (growth conditions)
PRK	phosphoribulokinase
PYE	peptone-yeast extract (chemoheterotrophic media for *Rb. capsulatus)*
RLP	RubisCO-like protein
RubisCO	ribulose 1,5-bisphosphate carboxylase/oxygenase
RuBP	1,5-ribulose bis-phosphate
SSU	small subunit (of RubisCO)
TEMDB	50 mM Tris-HCl pH 8.0, 1 mM EDTA, 10 mM $MgCl_2$, 1 mM DTT, 10 mM $NaHCO_3$
TRIS	tris(hydroxymethyl)aminomethane
V_c	maximum velocity of the carboxylation reaction
V_o	maximum velocity of the oxygenation reaction
Ω	Specificity factor

References

[Anonymous]<http://en.wikipedia.org/wiki/Observable_universe#Matter_content>.

Alfreider A, Vogt C, Geiger-Kaiser M, Psenner R. 2009. Distribution and diversity of autotrophic bacteria in groundwater systems based on the analysis of RubisCO genotypes. Syst Appl Microbiol 32(2):140-50.

Alfreider A, Vogt C, Hoffmann D, Babel W. 2003. Diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes from groundwater and aquifer microorganisms. Microb Ecol 45(4):317-28.

Allen MA, Lauro FM, Williams TJ, Burg D, Siddiqui KS, De Francisci D, Chong KW, Pilak O, Chew HH, De Maere MZ, *et al*. 2009. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: The role of genome evolution in cold adaptation. ISME J 3(9):1012-35.

Alonso H, Blayney MJ, Beck JL, Whitney SM. 2009. Substrate-induced assembly of *Methanococcoides burtonii* D-ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. J Biol Chem 284(49):33876-82.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25(17):3389-402.

Andersson I. 2008. Catalysis and regulation in rubisco. J Exp Bot 59(7):1555-68.

Andersson I and Backlund A. 2008. Structure and function of rubisco. Plant Physiol Biochem 46(3):275-91.

Andexer J, Guterl JK, Pohl M, Eggert T. 2006. A high-throughput screening assay for hydroxynitrile lyase activity. Chem Commun (Camb). (40):4201,3. Epub 2006 Aug 29.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22(2):195-201.

Ashida H, Danchin A, Yokota A. 2005. Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism? Res Microbiol 156(5-6):611-8.

Ashida H, Saito Y, Nakano T, Tandeau de Marsac N, Sekowska A, Danchin A, Yokota A. 2008. RuBisCO-like proteins as the enolase enzyme in the methionine salvage pathway: Functional and evolutionary relationships between RuBisCO-like proteins and photosynthetic RuBisCO. J Exp Bot 59(7):1543-54.

Badger MR and Gallagher A. 1987. Adaptation of photosynthetic CO_2 and HCO_3-accumulation by the cyanobacterium *Synechococcus* PCC6301 to growth at different inorganic carbon concentrations. Aust J Plant Physiol 14(2):189-201.

Baneyx F and Mujacic M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotech 22(11):1399-408.

Barber RD, Zhang L, Harnack M, Olson MV, Kaul R, Ingram-Smith C, Smith KS. 2011. Complete genome sequence of *Methanosaeta concilii*, a specialist in aceticlastic methanogenesis. J Bacteriol 193(14):3668-9.

Bar-Even A, Noor E, Lewis NE, Milo R. 2010. Design and analysis of synthetic carbon fixation pathways. Proc Natl Acad Sci U S A 107(19):8889-94.

Beloqui A, Pita M, Polaina J, Martinez-Arias A, Golyshina OV, Zumarraga M, Yakimov MM, Garcia-Arellano H, Alcalde M, Fernandez VM, *et al*. 2006. Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: Biochemical properties, structural analysis, and phylogenetic relationships. J Biol Chem. 281(32):22933,42. Epub 2006 Jun 1.

Bloom AJ, Burger M, Rubio Asensio JS, Cousins AB. 2010. Carbon dioxide enrichment inhibits nitrate assimilation in wheat and *Arabidopsis*. Science 328(5980):899-903.

Boone DR, Castenholz RW, Garrity GM. 2001. Bergey's manual of systematic bacteriology / george M. garrity, editor-in-chief. New York: Springer.

Boyer HW and Roulland-Dussoix D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J Mol Biol 41(3):459-72.

Brownstein MJ, Carpten JD, Smith JR. 1996. Modulation of non-templated nucleotide addition by taq DNA polymerase: Primer modifications that facilitate genotyping. BioTechniques 20(6):1004,6, 1008-10.

Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, Lorimer GH. 1998. Mechanism of rubisco: The carbamate as general base. Chem Rev 98(2):549-62.

Coligan JE. 1996. Current protocols in protein science. Brooklyn, N.Y.: John Wiley & Sons, Inc.

Coque TM, Oliver A, Perez-Diaz JC, Baquero F, Canton R. 2002. Genes encoding TEM-4, SHV-2, and CTX-M-10 extended-spectrum beta-lactamases are carried by multiple *Klebsiella pneumoniae* clones in a single hospital (Madrid, 1989 to 2000). Antimicrob Agents Chemother. 46(2):500-10.

DeLong EF. 2006. Archaeal mysteries of the deep revealed. Proc Natl Acad Sci U S A. 103(17):6417,8. Epub 2006 Apr 17.Delwiche CF and Palmer JD. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol Biol Evol 13(6):873-82.

Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M, Mullany P. 2003. Novel tetracycline resistance determinant from the oral metagenome. Antimicrob Agents Chemother. 47(4):1430-2.

Dougan D,A., Mogk ,A., Bukau ,B. 2002. Protein folding and degradation in bacteria: To degrade or not to degrade? That is the question. Cellular and Molecular Life Sciences (10):1607-16.

Edwards KJ. 2011. Oceanography: Carbon cycle at depth. Nature Geosci 4(1):9-11.

Elsaied HE, Kimura H, Naganuma T. 2007. Composition of archaeal, bacterial, and eukaryal RuBisCO genotypes in three western pacific arc hydrothermal vent systems. Extremophiles. 11(1):191,202. Epub 2006 Oct 6.

Estelmann S, Ramos-Vera WH, Gad'on N, Huber H, Berg IA, Fuchs G. 2011. Carbon dioxide fixation in *'Archaeoglobus lithotrophicus'*: Are there multiple autotrophic pathways? FEMS Microbiol Lett 319(1):65-72.

Evans MC, Buchanan BB, Arnon DI. 1966. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. Proc Natl Acad Sci U S A 55(4):928-34.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39(4):783-91.

Findley SD, Mormile MR, Sommer-Hurley A, Zhang XC, Tipton P, Arnett K, Porter JH, Kerley M, Stacey G. 2011. Activity-based metagenomic screening and biochemical characterization of bovine ruminal protozoan glycoside hydrolases. Appl Environ Microbiol 77(22):8106-13.

Finn MW and Tabita FR. 2004. Modified pathway to synthesize ribulose 1,5-bisphosphate in methanogenic archaea. J Bacteriol 186(19):6360-6.

Finn MW and Tabita FR. 2003. Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea. J Bacteriol 185(10):3049-59.

Flachmann R, Zhu G, Jensen RG, Bohnert HJ. 1997. Mutations in the small subunit of ribulose-1,5-bisphosphate carboxylase/ oxygenase increase the formation of the misfire product xylulose-1,5-bisphosphate. Plant Physiol 114(1):131-6.

Franzmann PD, Springer N, Ludwig W, Conway De Macario E, and Rohde M. 1992. A methanogenic archaeon from ace lake, antarctica: *Methanococcoides burtonii* sp. nov. System Appl Microbiol 15:573-81.

Fuhrmann M, Hausherr A, Ferbitz L, Schadl T, Heitzer M, Hegemann P. 2004. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. Plant Mol Biol (6):869-81.

Gao M, Wang L, Chen SF. 2011. Metagenome cloning and functional analysis of Na(+)/H (+) antiporter genes from Keke salt lake in China. Curr Microbiol.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. Walker JM, editor. In: The proteomics protocols handbook. Totowa, N.J.: Humana Press.

Goloubinoff P, Gatenby AA, Lorimer GH. 1989. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. Nature 337(6202):44-7.

Goodchild A, Raftery M, Saunders NF, Guilhaus M, Cavicchioli R. 2005. Cold adaptation of the antarctic archaeon, *Methanococcoides burtonii* assessed by proteomics using ICAT. J Proteome Res 4(2):473-80.

Goodchild A, Raftery M, Saunders NF, Guilhaus M, Cavicchioli R. 2004a. Biology of the cold adapted archaeon, *Methanococcoides burtonii* determined by proteomics using liquid chromatography-tandem mass spectrometry. J Proteome Res 3(6):1164-76.

Goodchild A, Saunders NF, Ertan H, Raftery M, Guilhaus M, Curmi PM, Cavicchioli R. 2004b. A proteomic determination of cold adaptation in the antarctic archaeon, *Methanococcoides burtonii*. Mol Microbiol 53(1):309-21.

Gutteridge S and Pierce J. 2006. A unified theory for the basis of the limitations of the primary reaction of photosynthetic CO(2) fixation: Was Dr. Pangloss right? Proc Natl Acad Sci U S A. 103(19):7203,4. Epub 2006 May 1.

Gutteridge S and Gatenby AA. 1995. Rubisco synthesis, assembly, mechanism, and regulation. Plant Cell 7(7):809-19.

Harley P,C. and Sharkey T,D. 1991. An improved model of C3 photosynthesis at high CO_2: Reversed O_2 sensitivity explained by lack of glycerate reentry into the chloroplast. Photosynthesis Res (3):169-78.

Harpel MR, Lee EH, Hartman FC. 1993. Anion-exchange analysis of ribulose-bisphosphate carboxylase/oxygenase reactions: CO_2/O_2 specificity determination and identification of side products. Anal Biochem. 209(2):367-74.

Hartley JL, Temple GF, Brasch MA. 2000. DNA cloning using in vitro site-specific recombination. Genome Res 10(11):1788-95.

Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, Santini JM, Oremland RS. 2007. *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int J Syst Evol Microbiol 57(Pt 3):504-12.

Horken KM and Tabita FR. 1999a. The "green" form I ribulose 1,5-bisphosphate carboxylase/oxygenase from the nonsulfur purple bacterium *Rhodobacter capsulatus*. J Bacteriol 181(13):3935-41.

Horken KM and Tabita FR. 1999b. Closely related form I ribulose bisphosphate carboxylase/oxygenase molecules that possess different CO_2/O_2 substrate specificities. Arch Biochem Biophys 361(2):183-94.

Horken KM. 1998. Development of new prokaryotic systems for ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco) structure function and assembly studies. Columbus, Ohio: Ohio State University.

Hu XP, Heath C, Taylor MP, Tuffin M, Cowan D. 2011. A novel, extremely alkaliphilic and cold-active esterase from antarctic desert soil. Extremophiles .

Hugler M and Sievert SM. 2011. Beyond the calvin cycle: Autotrophic carbon fixation in the ocean. Annu Rev Marine Sci 3(1):261-89.

Igarashi Y and Kodama T. 1996. Genes related to carbon dioxide fixation in *Hydrogenovibrio marinus* and *Pseudomonas hydrogenothermophila*. Lidstrom ME and Tabita FR, editors. In: Microbial growth on C1 compounds. Dordrecht ; Boston: Kluwer Academic Publishers. 363 p.

Imker HJ, Singh J, Warlick BP, Tabita FR, Gerlt JA. 2008. Mechanistic diversity in the RuBisCO superfamily: A novel isomerization reaction catalyzed by the RuBisCO-like protein from *Rhodospirillum rubrum*. Biochemistry 47(43):11171-3.

John DE, Patterson SS, Paul JH. 2007. Phytoplankton-group specific quantitative polymerase chain reaction assays for RuBisCO mRNA transcripts in seawater. Mar Biotechnol (NY) 9(6):747-59.

John DE, Wang ZA, Liu X, Byrne RH, Corredor JE, Lopez JM, Cabrera A, Bronk DA, Tabita FR, Paul JH. 2007. Phytoplankton carbon fixation gene (RuBisCO) transcripts and air-sea CO(2) flux in the mississippi river plume. ISME J 1(6):517-31.

John DE, Wawrik B, Robert Tabita F, Paul JH. 2006. Gene diversity and organization in rbcL-containing genome fragments from uncultivated *Synechococcus* in the Gulf of Mexico. Mar Ecol Prog Ser 316:23-33.

Jones DS, Albrecht HL, Dawson KS, Schaperdoth I, Freeman KH, Pi Y, Pearson A, Macalady JL. 2011. Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. ISME J .

Kahn M, Kolter R, Thomas C, Figurski D, Meyer R, Remaut E, Helinski DR. 1979. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K, and RK2. Methods Enzymol 68:268-80.

Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W, Liles MR. 2011. Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. Gene 475(2):57-62.

Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL repository and associated resources. Nucleic Acids Res 37(Database issue):D387-92.

Kim KH and Bae JW. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol 77(21):7663-8.

Kitano K, Maeda N, Fukui T, Atomi H, Imanaka T, Miki K. 2001. Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry. Structure 9(6):473-81.

Knauf VC and Nester EW. 1982. Wide host range cloning vectors: A cosmid clone bank of an *Agrobacterium* Ti plasmid. Plasmid 8(1):45-54.

Kreel NE and Tabita FR. 2007. Substitutions at methionine 295 of *Archaeoglobus fulgidus* ribulose-1,5-bisphosphate carboxylase/oxygenase affect oxygen binding and CO_2/O_2 specificity. J Biol Chem 282(2):1341-51.

Lal R. 2008. Sequestration of atmospheric CO_2 in global carbon pools. Energy Environ Sci 1(1):86-100.

LeCleir GR, Buchan A, Hollibaugh JT. 2004. Chitinase gene sequences retrieved from diverse aquatic habitats reveal environment-specific distributions. Appl Environ Microbiol. 70(12):6977-83.

Lee B and Tabita FR. 1990. Purification of recombinant ribulose-1,5-bisphosphate carboxylase/oxygenase large subunits suitable for reconstitution and assembly of active L8S8 enzyme. Biochemistry 29(40):9352-7.

Lee WT, Watson GW, Tabita FR. 1998. Chaperonins of the purple nonsulfur bacterium *Rhodobacter sphaeroides*. Methods Enzymol 290:154-61.

Lee WT, Terlesky KC, Tabita FR. 1997. Cloning and characterization of two groESL operons of *Rhodobacter sphaeroides*: Transcriptional regulation of the heat-induced groESL operon. J Bacteriol. 179(2):487-95.

Li LA and Tabita FR. 1997. Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the rbcX gene. J Bacteriol 179(11):3793-6.

Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW. 2005. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. Appl Environ Microbiol. 71(12):7768-77.

Long BM, Badger MR, Whitney SM, Price GD. 2007. Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple rubisco complexes with carboxysomal proteins CcmM and CcaA. J Biol Chem 282(40):29323-35.

Lorimer GH. 1996. A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. FASEB J 10(1):5-9.

Magnuson VL, Ally DS, Nylund SJ, Karanjawala ZE, Rayman JB, Knapp JI, Lowe AL, Ghosh S, Collins FS. 1996. Substrate nucleotide-determined non-templated addition of adenine by taq DNA polymerase: Implications for PCR-based genotyping and cloning. BioTechniques 21(4):700-9.

Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Dymock D, Wade WG. 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl Environ Microbiol 64(2):795-9.

Mathrani IM, Boone DR, Mah RA, Fox GE, Lau PP. 1988. *Methanohalophilus zhilinae* sp. nov., an alkaliphilic, halophilic, methylotrophic methanogen. Int J Syst Bacteriol 38(2):139-42.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. Nucleic Acids Res 28(1):292.

Nakano T, Ashida H, Mizohata E, Matsumura H, Yokota A. An evolutionally conserved Lys122 is essential for function in *Rhodospirillum rubrum* bona fide RuBisCO and *Bacillus subtilis* RuBisCO-like protein. Biochem Biophys Res Commun 392(2):212-6.

Nei M and Kumar S. 2000. Molecular evolution and phylogenetics. Oxford ; New York: Oxford University Press. 333 p.

Nigro LM and King GM. 2007. Disparate distributions of chemolithotrophs containing form IA or IC large subunit genes for ribulose-1,5-bisphosphate carboxylase/oxygenase in intertidal marine and littoral lake sediments. FEMS Microbiol Ecol 60(1):113-25.

Nishitani Y, Yoshida S, Fujihashi M, Kitagawa K, Doi T, Atomi H, Imanaka T, Miki K. 2010. Structure-based catalytic optimization of a type III rubisco from a hyperthermophile. J Biol Chem 285(50):39339-47.

Ormerod JG, Ormerod KS, Gest H. 1961. Light-dependent utilization of organic compounds and photoproduction of molecular hydrogen by photosynthetic bacteria; relationships with nitrogen metabolism. Arch Biochem Biophys 94:449-63.

Paoli GC, Vichivanives P, Tabita FR. 1998. Physiological control and regulation of the *Rhodobacter capsulatus* cbb operons. J Bacteriol 180(16):4258-69.

Papadopoulos JS and Agarwala R. 2007. COBALT: Constraint-based alignment tool for multiple protein sequences. Bioinformatics 23(9):1073-9.

Parry MAJ, Andralojc PJ, Mitchell RAC, Madgwick PJ, Keys AJ. 2003. Manipulation of rubisco: The amount, activity, function and regulation 10.1093/jxb/erg141. J Exp Bot 54(386):1321-33.

Paul JH, Pichard SL, Kang JB, Watson GMF, Tabita FR. 1999. Evidence for a clade-specific temporal and spatial separation in ribulose bisphosphate carboxylase gene expression in phytoplankton populations off Cape Hatteras and Bermuda. Limnol Oceanogr 44(1):12-23.

Pereto JG, Velasco AM, Becerra A, Lazcano A. 1999. Comparative biochemistry of CO_2 fixation and the evolution of autotrophy. Int Microbiol 2(1):3-10.

Rachmilevitch S, Cousins AB, Bloom AJ. 2004. Nitrate assimilation in plant shoots depends on photorespiration. Proc Natl Acad Sci U S A 101(31):11506-10.

Rajendhran J and Gunasekaran P. 2011. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. Microbiol Res 166(2):99-110.

Rappe MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418(6898):630-3.

Read BA and Tabita FR. 1992a. A hybrid ribulosebisphosphate carboxylase/oxygenase enzyme exhibiting a substantial increase in substrate specificity factor. Biochemistry 31(24):5553-60.

Read BA and Tabita FR. 1992b. Amino acid substitutions in the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase that influence catalytic activity of the holoenzyme. Biochemistry 31(2):519-25.

Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR. 2008. Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug *Elysia chlorotica*. Proceedings of the National Academy of Sciences 105(46):17867-71.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, *et al*. 2007. The sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. PLoS Biol 5(3):e77.

Saitou N and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406-25.

Sambrook J and Russell DW. 2001. Molecular cloning :A laboratory manual. 3rd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Satagopan S, Scott SS, Smith TG, Tabita FR. 2009. A rubisco mutant that confers growth under a normally "inhibitory" oxygen concentration. Biochemistry 48(38):9076-83.

Sato T, Atomi H, Imanaka T. 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science 315(5814):1003-6.

Schmitz JE, Daniel A, Collin M, Schuch R, Fischetti VA. 2008. Rapid DNA library construction for functional genomic and metagenomic screening. Appl Environ Microbiol 74(5):1649-52.

Schneider G, Lindqvist Y, Branden CI. 1992. RUBISCO: Structure and mechanism. Annu Rev Biophys Biomol Struct 21:119-43.

Schopf JW. 2011. The paleobiological record of photosynthesis. Photosynth Res 107(1):87-101.

Schopf JW. 1993. Microfossils of the early Archean apex chert: New evidence of the antiquity of life. Science 260(5108):640-6.

Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 31(13):3381-5.

Seidman CE, Struhl K, Sheen J, Jessen T. 2001. Introduction of plasmid DNA into cells. .

Sharkey TD. 1988. Estimating the rate of photorespiration in leaves. Physiol Plantarum 73(1):147-52.

Sharp PM and Li WH. 1987. The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15(3):1281-95.

Shively JM, van Keulen G, Meijer WG. 1998. Something from almost nothing: Carbon dioxide fixation in chemoautotrophs. Annu Rev Microbiol 52:191-230.

Sim SJ, Baik KS, Park SC, Choe HN, Seong CN, Shin TS, Woo HC, Cho JY, Kim D. 2011. Characterization of alginate lyase gene using a metagenomic library constructed from the gut microflora of abalone. J Ind Microbiol Biotechnol .

Singh J and Tabita FR. 2010. Roles of RubisCO and the RubisCO-like protein in 5-methylthioadenosine metabolism in the nonsulfur purple bacterium *Rhodospirillum rubrum*. J Bacteriol 192(5):1324-31.

Smith SA and Tabita FR. 2004. Glycine 176 affects catalytic properties and stability of the *Synechococcus* sp. strain PCC6301 ribulose-1,5-bisphosphate carboxylase/oxygenase. J Biol Chem 279(24):25632-7.

Smith SA and Tabita FR. 2003. Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. J Mol Biol 331(3):557-69.

Spiridonova EM, Berg IA, Kolganova TV, Ivanovskii RN, Kuznetsov BB, Turova TP. 2004. [An oligonucleotide primer system for amplification of the ribulose-1,5-bisphosphate carboxylase/oxygenase genes of bacteria of various taxonomic groups]. Mikrobiologiia. 73(3):377-87.

Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques 28(6):1102, 1104.

Strauss G and Fuchs G. 1993. Enzymes of a novel autotrophic CO_2 fixation pathway in the phototrophic bacterium C*hloroflexus aurantiacus*, the 3-hydroxypropionate cycle. Eur J Biochem 215(3):633-43.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, Reinthaler T, Poulton NJ, Masland ED, Gomez ML, *et al*. 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science 333(6047):1296-300.

Tabita FR and McFadden BA. 1974. D-ribulose 1,5-diphosphate carboxylase from *Rhodospirillum rubrum:* quaternary structure, composition, catalytic, and immunological properties. J Biol Chem. 249(11):3459-64.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008a. Distinct form I, II, III, and IV rubisco proteins from the three kingdoms of life provide clues about rubisco evolution and structure/function relationships. J Exp Bot 59(7):1515-24.

Tabita FR, Hanson TE, Satagopan S, Witte BH, Kreel NE. 2008b. 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 of London.Series B, Biological Sciences 363(1504):2629-40.

Tabita FR. 1999. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynthesis Res 60(1):1-28.

Taiz L and Zeiger E. 2010. Plant physiology. 5th ed. Sunderland, MA: Sinauer Associates. 782, [137] p.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731-9.

Tans, P and Keeling R. 2012. NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/)

Tcherkez GG, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc Natl Acad Sci U S A 103(19):7246-51.

Teske A and Sorensen KB. 2008. Uncultured archaea in deep marine subsurface sediments: Have we caught them all? ISME J 2(1):3-18.

Turova TP and Spiridonova EM. 2009. [Phylogeny and evolution of RubiCo genes in prokaryotes]. Mol Biol (Mosk) 43(5):772-88.

Uyaguari MI, Fichot EB, Scott GI, Norman RS. 2011. Characterization and quantitation of a novel beta-lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. Appl Environ Microbiol 77(23):8226-33.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, *et al.* 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304(5667):66-74.

Voget S, Steele HL, Streit WR. 2006. Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol. 126(1):26,36. Epub 2006 Apr 3.

Watson GM and Tabita FR. 1997. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A molecule for phylogenetic and enzymological investigation. FEMS Microbiol Lett 146(1):13-22.

Watson GM and Tabita FR. 1996. Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium. Plant Mol Biol 32(6):1103-15.

Wawrik B, Paul JH, Tabita FR. 2002. Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. Appl Environ Microbiol 68(8):3771-9.

Whitney SM, Houtz RL, Alonso H. 2011. Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, rubisco. Plant Physiol 155(1):27-35

Wood HG. 1991. Life with CO or CO_2 and H_2 as a source of carbon and energy. FASEB J 5(2):156-63.

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, *et al*. 2007. The sorcerer II global ocean sampling expedition: Expanding the universe of protein families. PLoS Biol 5(3):e16.

Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, Ryu S. 2004. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Appl Environ Microbiol. 70(12):7229-35.

Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, Church GM. 2006. Sequencing genomes from single cells by polymerase cloning. Nat Biotechnol. 24(6):680,6. Epub 2006 May 28.