Title

Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria

Permalink

https://escholarship.org/uc/item/9t4950ck

Journal

Environmental Microbiology, 18(4)

ISSN

1462-2912

Authors

Varaljay, Vanessa A
Satagopan, Sriram
North, Justin A
et al.

Publication Date

2016-04-01

DOI

10.1111/1462-2920.13138

Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria

Vanessa A. Varaljay

Sriram Satagopan

Justin A. North

Brian Witte

Manuella N. Dourado
Karthik Anantharaman

Mark A. Arbing

Shelley Hoeft McCann

Ronald S. Oremland
… See all authors

First published: 30 November 2015

https://doi.org/10.1111/1462-2920.13138

Cited by: 11

UC-eLinks
SECTIONS
PDF
TOOLS

SHARE

Summary

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) is a critical yet severely inefficient enzyme that catalyses the fixation of virtually all of the carbon found on Earth. Here, we report a functional metagenomic selection that recovers physiologically active RubisCO molecules directly from uncultivated and largely unknown members of natural microbial communities. Selection is based on CO_2-dependent growth in a host strain capable of expressing environmental deoxyribonucleic acid (DNA), precluding the need for pure cultures or screening of recombinant clones for enzymatic activity. Seventeen functional RubisCO-encoded sequences were selected using DNA extracted from soil and river autotrophic enrichments, a photosynthetic biofilm and a subsurface groundwater aquifer. Notably, three related form II RubisCOs were recovered which share high sequence similarity with metagenomic scaffolds from uncultivated members of the *Gallionellaceae* family. One of the *Gallionellaceae* RubisCOs was purified and shown to possess CO_2/O_2 specificity typical of form II enzymes. X-ray crystallography determined that this enzyme is a hexamer, only the second form II multimer ever solved and the first RubisCO structure obtained from an uncultivated bacterium. Functional metagenomic selection leverages natural biological diversity and billions of years of evolution inherent in environmental communities, providing a new window into the discovery of CO_2-fixing enzymes not previously characterized.

Introduction

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) is the most abundant protein on earth and catalyses CO_2 fixation onto the enzyme-bound enediolate of ribulose 1, 5-bisphosphate (RuBP) to produce organic carbon in plants, algae and autotrophic bacteria (Tabita *et al.*, 2008). However, RubisCO is a sluggish catalyst and its efficiency is further limited by its poor ability to discriminate between the gaseous substrates CO_2 and O_2(Spreitzer and Salvucci, 2002; Andersson, 2008), with the oxygen fixation reaction leading to energetically wasteful metabolism. Because of RubisCO's importance in CO_2 bioconversions, a number of studies have endeavoured to increase carbon capture through various artificial evolution and bioengineering methods to produce RubisCO enzymes with increased activity or higher CO_2 specificity (Spreitzer and Salvucci, 2002; Smith and Tabita, 2003; Parikh *et al.*, 2006; Yoshida *et al.*, 2007; Mueller-Cajar and Whitney, 2008; Satagopan *et al.*, 2009; 2014; Cai *et al.*, 2014; Lin *et al.*, 2014). While these and recent studies to manipulate the properties of RubisCO are especially promising (Lin *et al.*, 2014; Hauser *et al.*, 2015), it is clear that sequence or structural analyses alone are not accurate predictors of enzyme function. The molecular mechanisms underpinning the vast differences in catalytic properties between highly related RubisCO protein sequences and structural homologues from different taxa remain enigmatic (Tabita, 1999; Tabita *et al.*, 2008).

Most of what is known regarding bacterial RubisCO function is derived from pure culture analyses and relatively few model enzymes. However, recent massive deoxyribonucleic acid (DNA) sequencing efforts have shown the abundant presence of diverse and novel RubisCO-encoding genes in environmental communities of largely uncultivated bacteria and new candidate phyla, the so-called ‘microbial dark matter’ (Wrighton *et al.*, 2012; submitted; Campbell *et al.*, 2013; Guo *et al.*, 2013; Castelle *et al.*, 2015; Tebo *et al.*, 2015). Although the physiological significance of these novel RubisCO enzymes is unquantified, these organisms likely have evolved significant structural and functional adaptations to allow CO_2 to be metabolized in diverse environments such as in marine surface waters, hydrothermal vents and the terrestrial subsurface (Witte *et al*., 2010; Wrighton *et al*., 2012; submitted; Böhnke and Perner, 2015; Castelle *et al*., 2015). Sequencing, cloning and activity screening of purified recombinant proteins is possible on a small scale (Böhnke and Perner, 2015). However, it is not always possible to discern physiological function by sequence-gazing alone, nor is it practical to produce and individually screen and characterize recombinant proteins from the large number of novel RubisCO-encoding genes recently observed (Wrighton *et al*., 2012; submitted; Campbell *et al*., 2013; Guo *et al*., 2013; Castelle *et al*., 2015; Tebo *et al*., 2015). Whereas a previous study demonstrated the feasibility of studying environmentally derived RubisCO genes in a recombinant photosynthetic host (Witte *et al*., 2010), a directed functional metagenomic selection approach is necessary for high-throughput retrieval of desirable new RubisCO enzymes from uncultivated organisms. The isolation and analysis of such proteins can ultimately provide additional biochemical and structural insights with respect to key aspects of catalysis (Satagopan *et al*., 2009; 2014). In this study, we implemented such a streamlined functional metagenomic approach to isolate diverse RubisCOs from three environments and demonstrate that the approach is applicable for gaining novel biochemical and structural properties from uncultivated ‘microbial dark matter’.

Results and discussion

Metagenomic functional selection approach and validation

The feasibility of isolating novel functional enzymes encoded by metagenomic DNA from soil and other environmental samples has been previously demonstrated (Rondon *et al*., 2000; Handelsman, 2004; Leis *et al*., 2015; Ufarté *et al*., 2015). In order to capture functional RubisCOs directly from uncultivated members in microbial communities, we developed a specific functional metagenomics approach that capitalizes on previously described autotrophic CO_2-dependent growth selection using a host bacterium that expresses exogenous RubisCO genes under aerobic and anaerobic conditions (Smith and Tabita, 2003; Satagopan *et al*., 2009; 2014). DNA was extracted from microbial community biomass obtained from various environments, enzymatically digested to partial completion, and subsequently, high-performance liquid chromatography (HPLC) purified and size fractionated (Fig. 1A and B). The purified DNA was cloned into the dual promoter expression vector p83 (Fig. 1C), allowing the expression of cloned sequences from either orientation, and then transformed into *Escherichia coli*. The cloned sequences from the *E. coli* library were conjugated *en masse* into the host RubisCO-deletion strain SB I/II^{-} of *Rhodobacter capsulatus* (Smith and Tabita, 2003; see *Experimental procedures*).

This system has been used previously to successfully support autotrophic photosynthetic and chemoautotrophic CO$_2$-dependent growth when complemented with RubisCO genes from diverse microbes, including *Proteobacteria*, *Cyanobacteria* and *Archaea* under aerobic or anaerobic conditions with varying CO$_2$/O$_2$ concentrations (Finn and Tabita, 2003; Smith and Tabita, 2003; Satagopan *et al*., 2009; 2014; Witte *et al*., 2010). Because the host strain cannot grow on CO$_2$unless complemented with an exogenous and active RubisCO, clones encoding physiologically functional RubisCOs are automatically selected from the environmental DNA pool (Fig. 1D). Additionally, unlike C*yanobacteria* and green algae that are known to have carbon concentrating mechanisms (Giordano *et al*., 2005), the relative ability of the complemented RubisCO to support growth in strain SB I/II^{-} is directly related to the properties of the enzyme and response to supplied external CO$_2$ and O$_2$ concentrations.

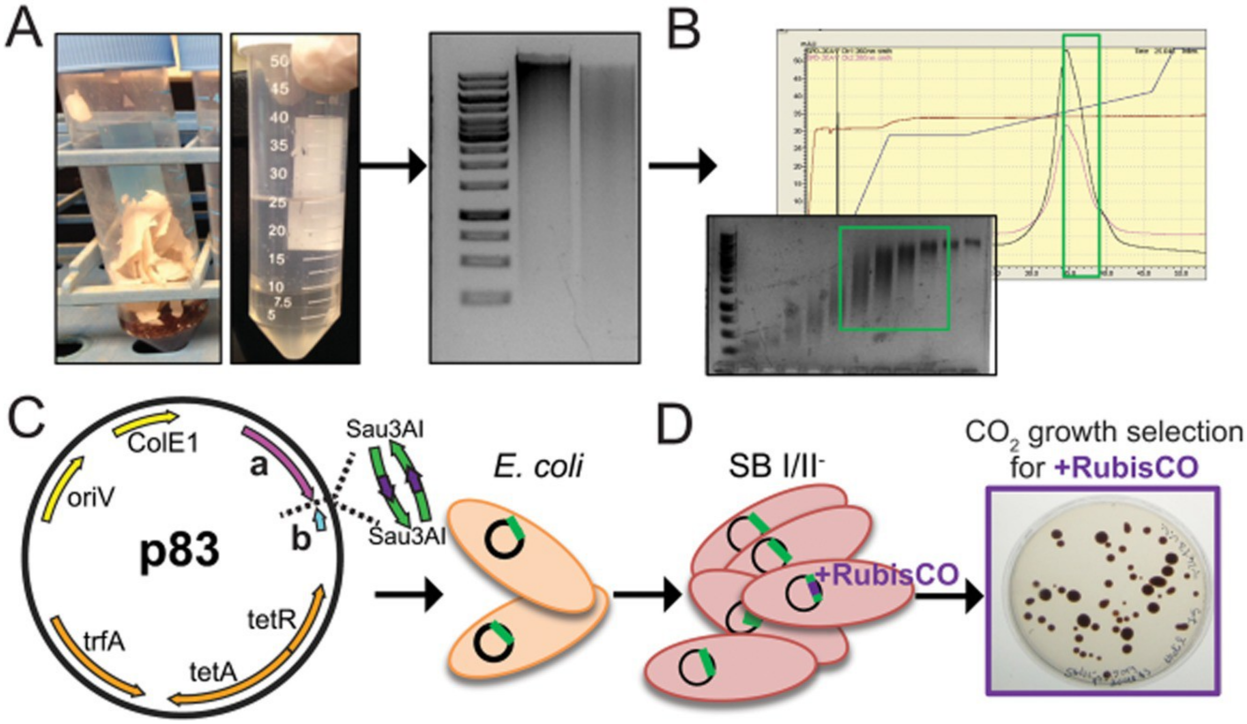


Figure 1

Open in figure viewerPowerPoint

Selection approach to recover functional metagenomic RubisCO sequences.

A. Metagenomic DNA was extracted from environmental samples and partially digested with Sau3AI.

B. Digested DNA was size fractionated (∼2–8 Kb) and purified via HPLC using a Waters Gen-Pak FAX DNA column. Selected fractions (green boxes) were combined and concentrated.

C. HPLC purified DNA was cloned into a dual *cbb* operon promoter vector, p83, containing a *R. rubrum cbbR* and *cbbM* promoter (Smith and Tabita, 2003) (A) in one orientation and the *R. capsulatus* SB1003 *cbbL* promoter (Vichivanives *et al*., 2000) (B) in the opposite orientation, ensuring that recovered RubisCO genes are expressed regardless of their orientation. Plasmid p83 cloned DNA was transformed into *E. coli* to create a metagenomic library.

D. The *E. coli* library was conjugated directly *en masse* into *R. capsulatus* strain SB I/II$^-$ and plated onto minimal medium under a CO_2/H_2 atmosphere. SB I/II$^-$colonies expressing a functional RubisCO gene on plasmid p83 supported growth using CO_2 as the sole carbon source.

The functional selection protocol was first vetted by interrogating extracted genomic DNA from five laboratory-cultured organisms (*Rhodobacter capsulatus* SB1003 [form I and II], *Rhodospirillum rubrum* ATCC 11170 [form II], *Rhodopseudomonas palustris* 010 [form I and II], *Rhodobacter sphaeroides* 2.4.1 [form I and II] and *Ralstonia eutropha* H16 [two form I]) in the presence and absence of oxygen, demonstrating the successful recovery of divergent RubisCOs (Fig. 2). From two selection experiments, eight of the nine possible form I and form II RubisCO sequences from the five organisms were recovered after cloning into the p83 dual promoter plasmid. The only RubisCO sequence not recovered was the megaplasmid-borne *R. eutropha* gene, which is basically identical [99% amino acid (aa) identity] in sequence to the chromosomal copy of this gene from this organism. Growth complementation occurred in strain SB I/II$^-$ on plates flushed with either 5% CO_2/H_2 for anaerobic photoautotrophic (PA) growth or flushed with 2.5–5% CO_2/H_2 mixed with an equal proportion of air (∼10.5% O_2) for aerobic chemolithoautotrophic (CA) growth. As expected, these experiments verified with certainty that our autotrophic selection approach successfully recovered both forms of RubisCO enzymes from a mixed pool of organisms with no false positives. Subsequently, the functional selection was employed with DNA from diverse natural microbial communities that showed promise for obtaining uncharacterized enzymes from RubisCO-harbouring organisms based on prior PCR and shot-gun metagenomic sequencing surveys (Kulp *et al*., 2008; Hoeft *et al*., 2010; Dourado-Ribeiro, 2013; Wrighton *et al*., 2012; submitted). These environments included a freshwater river and soil autotrophic enrichment, a photosynthetic biofilm derived from a hot spring within Mono Lake, CA, and a groundwater subsurface aquifer (see *Experimental procedures* for details). Selections were performed initially under PA growth conditions, allowing for fast growth and rapid analysis of any RubisCO that complements strain SB I/II$^-$ for functionality (Fig. 1). Once PA growth was achieved, selected colonies were further tested for the ability to complement under CA growth conditions.

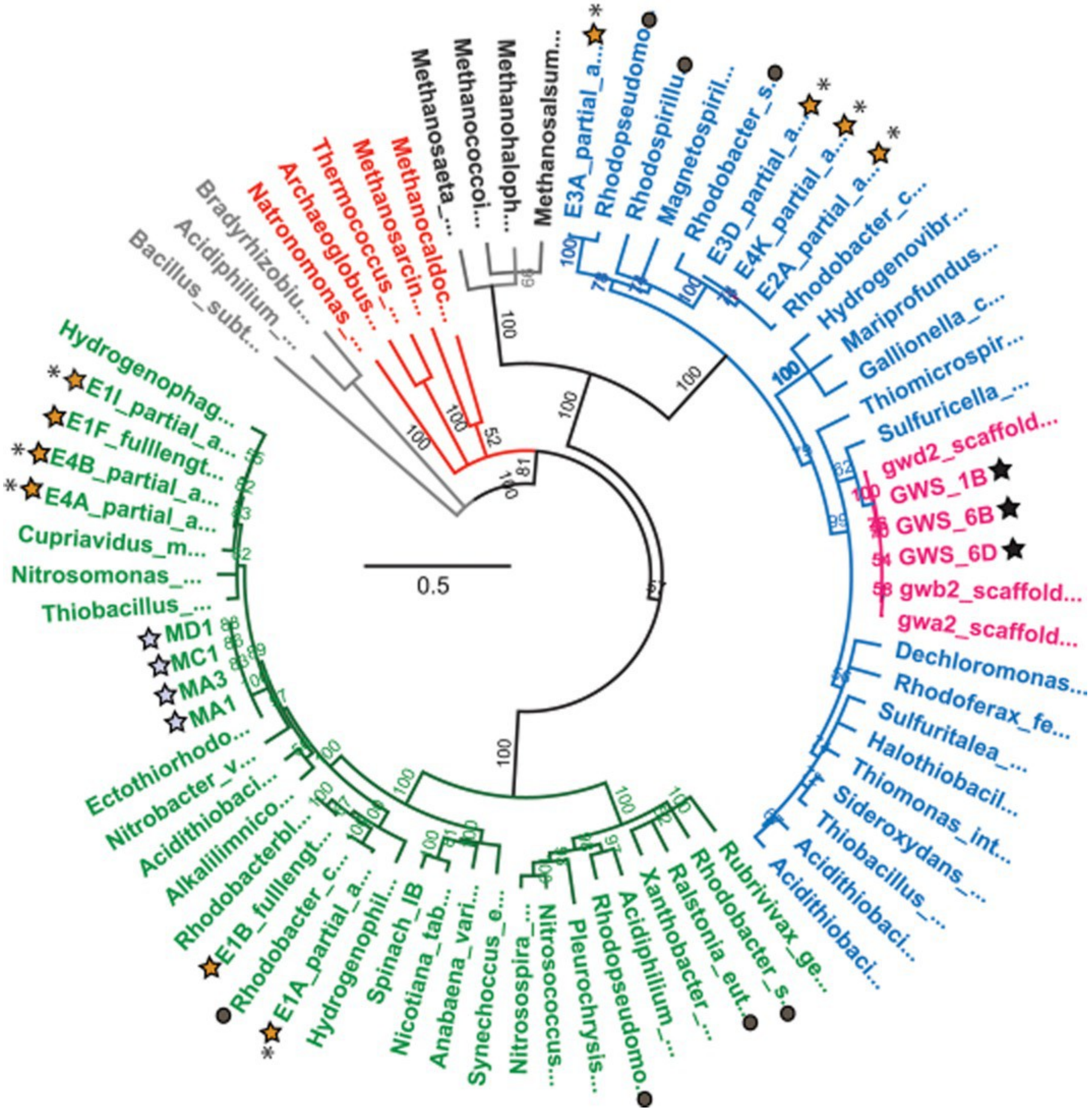


Figure 2

Open in figure viewerPowerPoint

RubisCO amino-acid tree of functionally selected RubisCOs. The following colour scheme was used: RubisCO form I sequences (green), form II (blue), form II *Gallionellaceae* (pink), hybrid form II/III (black), form III (red), and form IV (gray). Sequences retrieved as part of this study are marked with a star filled with various colours identifying the nature of the strain: Enrichment (orange), Biofilm (gray), and Groundwater aquifer (black). RubisCO sequences employed as part of the five-organism control experiment are indicated by gray-filled circles. Enrichment RubisCO sequences noted with asterisks are partial sequences encompassing aa 130-458 according to *R. rubrum* numbering for the form II and aa 80–410 according to *Synechococcus* PCC6301 numbering for the form I sequences. The rest of the sequences are full-length RubisCO sequences. Representative full-length RubisCO sequences were obtained from the NCBI ref seq database (Table S1). The distance-based RubisCO amino acid tree was made in GENEIOUS using the Jukes-Cantor genetic distance model and the neighbour-joining tree building method. *Bacillus subtilis* form IV RubisCO served as the outgroup. The tree was re-sampled using bootstrapping with 100 replicates.

Environmental RubisCO selection

Using the procedure described above, the selections resulted in 10 functional RubisCO sequences from the river and soil enrichments, four from the hot spring photosynthetic biofilm and three from the groundwater aquifer (Fig. 2). Several of the 17 recovered RubisCO sequences were identical or nearly identical to each other at the level of nucleotide (nt) or the encoded aa sequence but were isolated as unique clones with varying fragment lengths or intergenic regions (Table 1); the results suggested that these RubisCOs were derived from either the same or highly similar strains of bacteria. We selected for colonies able to complement strain SB I/II^{-} under both anaerobic PA and aerobic CA growth conditions without *a priori* knowledge of the sequences or laborious screening of multiple clones. Functional RubisCOs of the form I type (CbbLCbbS) from the biofilm (Fig. 3A and C) and of the form II type (CbbM) from the groundwater aquifer sample (Fig. 3B and D) also complemented for RubisCO-dependent growth in liquid cultures under both PA and CA growth conditions; moreover, extracts from these cells exhibited measurable levels of RubisCO activity (Fig. 3E and F) under aerobic conditions (21% O_2). The liquid growth complementation experiments showed that these selected RubisCOs, as compared with a positive control using a *Rhodosprillum rubrum* RubisCO construct, did not have unusually long lag times to achieve exponential growth and all reached similar maximum optical densities (ODs) (1.0–2.0 at OD_{660nm}).

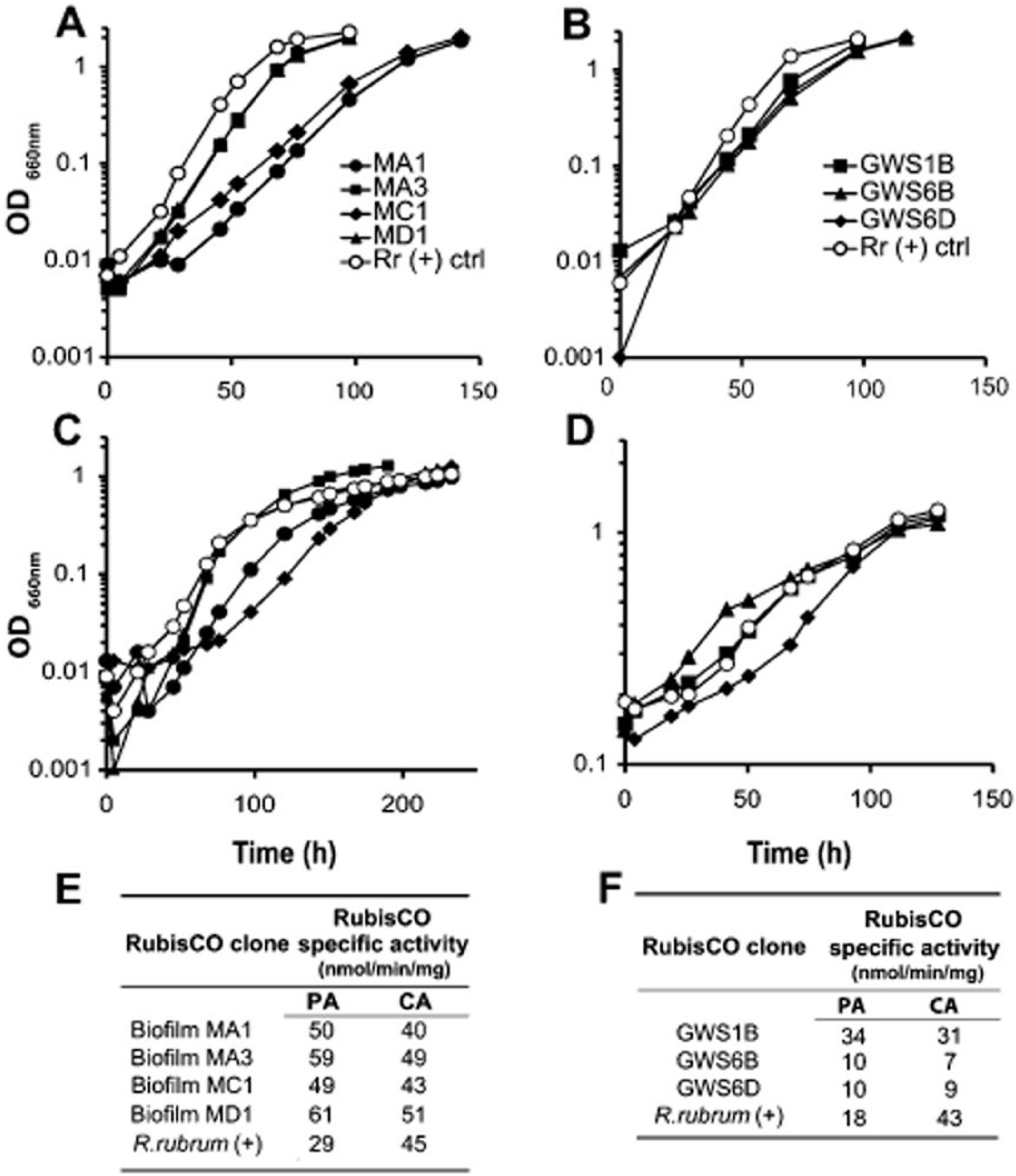
Table 1. Functional RubisCO-encoded metagenomic clones recovered from three environmental samples

Clone name	Fragment clone size (Kb)[a]	Form	Large subunit RubisCO isolate similarity (NCBI BLASTX)	No. of identical aa residues/total no.	% aa identity
Soil and river autotrophic enrichment cultures[b]					

Clone name	Fragment clone size (Kb)[a]	Form	Large subunit RubisCO isolate similarity (NCBI BLASTX)	No. of identical aa residues/total no.	% aa identity
E1A	7	I	*Rhodobacter capsulatus*	315/317	99
E1B	5.5	I	*Rhodobacter blasticus*	318/321[c, d]	99
E1F	4.5	I	*Hydrogenophaga pseudoflava*	452/473[c]	96
E1I	5.5	I	*Hydrogenophaga pseudoflava*	315/327	96
E4A	5.5	I	*Hydrogenophaga pseudoflava*	316/329	96
E4B	5.6	I	*Hydrogenophaga pseudoflava*	305/316	97
E2A	3.5	II	*Rhodobacter capsulatus*	331/331	100
E3D	2.2	II	*Rhodobacter capsulatus*	327/327	100
E4K	3.4	II	*Rhodobacter capsulatus*	330/333	99
E3A	2.8	II	*Rhodopseudomonas palustris*	331/334	99
Mono Lake photosynthetic biofilm culture from hot spring[b]					

Clone name	Fragment clone size (Kb)[a]	Form	Large subunit RubisCO isolate similarity (NCBI BLASTX)	No. of identical aa residues/total no.	% aa identity
MA1	2.8	I	*Ectothiorhodospira* PHS-1 sp.	473/473[c]	100
MA3	2.0	I	*Ectothiorhodospira* PHS-1 sp.	473/473[c]	100
MC1	2.4	I	*Ectothiorhodospira* PHS-1 sp.	473/473[c]	100
MD1	2.5	I	*Ectothiorhodospira* PHS-1 sp.	473/473[c]	100
Groundwater aquifer[b]					
GWS1B	3.1	II	*Sulfuricella dentrificans*	408/459[b]	89
GWS6B	4.5	II	*Sulfuricella dentrificans*	406/459[b]	88
GWS6D	6	II	*Sulfuricella dentrificans*	406/459[b]	88

- [a] Enrichment cloned fragment lengths were approximated via restriction digest and agarose gel analysis. The entire cloned fragments were not sequenced.
- [b] All clones were able to complement for both PA and CA growth in strain SB I/II^{-}.
- [c] Full-length large subunit sequence and alignment.
- [d] The full-length sequence was recovered via functional selection but the alignment was with the partial sequence of *R. blasticus* DSM 2131 RubisCO available in the NCBI database. There is no genome sequence available yet for *R. blasticus* sp. (as of 8/31/2015).



E

RubisCO clone	RubisCO specific activity (nmol/min/mg)	
	PA	CA
Biofilm MA1	50	40
Biofilm MA3	59	49
Biofilm MC1	49	43
Biofilm MD1	61	51
R.rubrum (+)	29	45

F

RubisCO clone	RubisCO specific activity (nmol/min/mg)	
	PA	CA
GWS1B	34	31
GWS6B	10	7
GWS6D	10	9
R.rubrum (+)	18	43

Figure 3

Open in figure viewerPowerPoint

Autotrophic growth phenotypes and RubisCO activities of enzymes encoded by metagenomic DNA fragments. Anaerobic PA growth (bubbled with 5% CO_2/95% H_2) of: (A) Mono Lake biofilm form I *cbbLS* fragments; and (B) Rifle groundwater aquifer form II *cbbM* fragments, each complemented in strain SB I/II$^-$.

C. Aerobic CA growth (2.5% CO_2/47.5% H_2/air) of Mono Lake biofilm form I *cbbLS*fragments.

D. Rifle groundwater aquifer form II *cbbM* fragments complemented in strain SB I/II- (bubbled with 5% CO_2/45% H_2/50% air). *Rhodospirillum rubrum cbbM* was cloned into plasmid p83 and served as a positive control for complementation of strain SB I/II$^-$. Empty vector (p83) served as a negative control (data not shown). RubisCO activities of crude protein lysates of cells harvested from liquid autotrophic cultures that reached an OD of > 1.0 at 660 nm E. strain SB I/II$^-$expressing Mono Lake biofilm form I RubisCO (encoded by metagenomic *cbbLcbbS* genes) and (F) strain SB I/II expressing Rifle groundwater aquifer form II (encoded by metagenomic *cbbM* gene). The data represent the averages of numbers obtained from two enzyme assay determinations, which both gave similar values.

The river and soil enrichment and biofilm samples identified RubisCOs most closely related to form I and form II sequences from the *Alphaproteobacteria, Betaproteobacteri* and *Gammaproteobacteria* (Fig. 2 and Table 1). The enrichment selection and subsequent DNA sequencing provided the first full-length open reading frame (ORF) encoding a form I RubisCO that is highly similar to that of *Rhodobacter blasticus*; which is ∼97% identical to the partial *R. blasticus* RubisCO sequence available in the National Center for Biotechnology Information (NCBI) database (985 bp) (Uchino and Yokota, 2003). The selection also identified a RubisCO with high similarity of that from a *Hydrogenophaga pseudoflava* species, which has only been partially characterized based on limited *in vitro* assays (Kim *et al*., 1997); the present study represents the first physiological complementation of a RubisCO-encoding sequence from this organism. The biofilm selection provided preliminary evidence that a form I RubisCO 100% identical to that of a moderately halophilic organism, *Ectothiorhodospira* sp. PHS1 (Kulp *et al*., 2008; Hoeft *et al*., 2010), could wholly complement the non-halophilic host strain SB I/II$^-$ in a medium with low salt (∼50 mM); however, follow-up experiments would be needed to determine the salt requirements for this RubisCO.

The groundwater-aquifer metagenomic sample provided DNA that was not derived from autotrophic enrichments. Selection with this sample resulted in the isolation of three RubisCO-encoding sequences harboured on 3 Kb (GWS1B), 4.5 Kb (GWS6B) and 6 Kb (GWS6D) metagenomic fragments all able to complement for CA growth (Fig. 3); in addition, cell extracts showed activity under full aerobic conditions (21% O_2), indicating that the respective enzymes are likely oxygen tolerant. Each of these groundwater aquifer metagenomic sequences consists of a full-length *cbbM* (form II RubisCO) gene, *cbbR*transcriptional regulator gene and potential CbbR binding sites upstream of the *cbbM* gene (Fig. S1); GWS6B also includes ∼800 nt of a partial sequence encoding a putative CbbQ RubisCO activation protein (Fig. 4). GWS6D and

GWS6B are 100% identical fragments, but the GWS6D gene fragment has an additional ∼2 Kb, which includes a partial sequence for the *groEL* gene (Fig. 4). GWS1B and GWS6B/D fragments are 94% identical across an overlapping ∼3 Kb region; the 6% difference mostly resides in the intergenic regions upstream and downstream of *cbbM*, suggesting that GWS1B and GWS6D/B might be from two closely related bacterial strains. The translated amino-acid sequences of CbbM (form II RubisCO) (99.5%) and CbbR (98.6%) proteins are highly similar between GWS1B and GWS6D/B. The best matching NCBI BLAST hits show that the recovered form II RubisCO sequences are most similar to *Sulfuricella dentrificans* (∼88–89% identity; Table 1), whereas the CbbR sequences are most similar to *Gallionella capsiferriformans* ES-2 (∼84% identity).

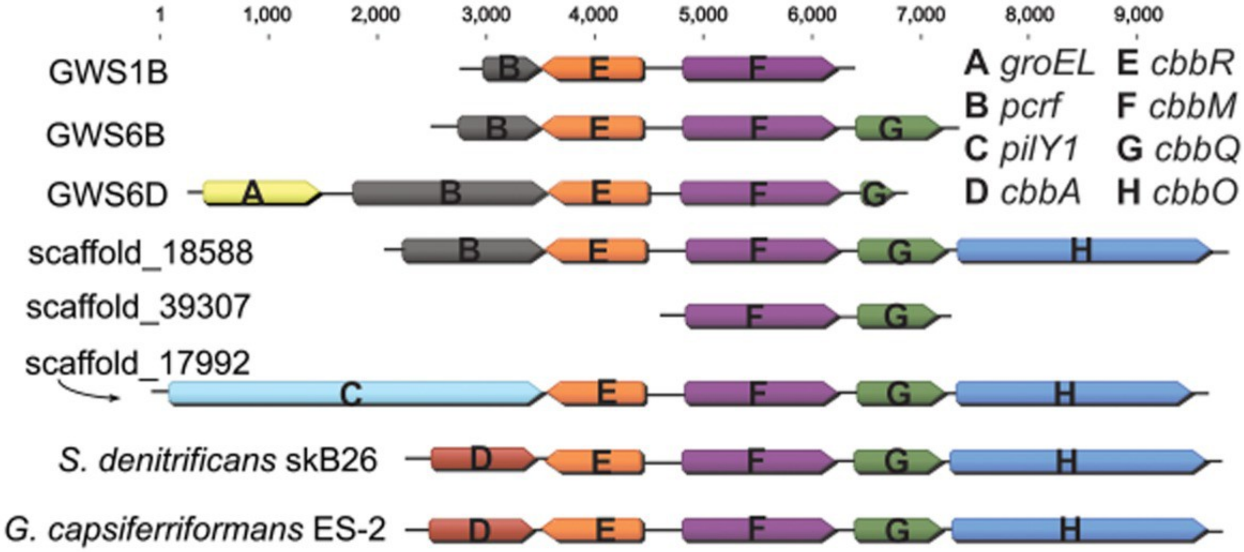


Figure 4

Open in figure viewerPowerPoint

Operon structure alignments of the recovered metagenomic fragments from the groundwater aquifer. Sequences are from the Rifle groundwater aquifer, along with closest RubisCO hits and related sequences available in the NCBI database. The acronym pcrf stands for peptide chain release factor.

Uncultivated *Gallionellaceae* sp. RubisCO analyses

Sampled time points near the selected groundwater aquifer site were previously subjected to deep sequencing, assembly and metagenomic reconstruction (Wrighton *et al*., 2012; Brown *et al*., 2015; Castelle *et al*., 2015); thus we queried the functionally selected RubisCO sequences against the binned genomic sequenced data in order to obtain the genomic context. All three recovered RubisCO fragments had best hits to metagenomic scaffolds with 83–95% nt identities across aligned regions (Table 2) and with similar gene organization (Fig. 4). Using a combination of tetranucleotide frequency (Dick *et al*., 2009) and data series abundance patterns (Brown *et al*., 2015), these RubisCO scaffolds were confidently assigned to two genomic bins (GWA2_Gallionellales_59_43, GWB2_Gallionellales_58_32). Phylogenetic analyses of concatenated ribosomal proteins extracted from these genomic bins demonstrate that these constitute distinct species within the family '*Gallionellaceae*' in the order *Betaproteobacteria* (Fig. S2). The RubisCO sequences are nearly identical at the aa level (∼98% identity; Table 2) and cluster together within the form II RubisCOs (Fig. 2). The family *Gallionellaceae* has only five sequenced genomes in NCBI, and the genomic scaffolds recovered here are derived from uncultivated lineages that lack representation in existing genomic databases.

Table 2. Sequence comparisons of groundwater aquifer RubisCO sequences obtained from functional selection with RubisCO scaffolds obtained from deep sequencing

	Scaffold GWB2_18588		**Scaffold GWD2_39307**		**Scaffold GWA2_17992**	
	(7.1 Kb)		**(2.3 Kb)**		**(9 Kb)**	
	No. of identical residues/total no.	**% identity**	**No. of identical residues/total no.**	**% identity**	**No. of identical residues/total no.**	**% identity**
Entire fragment[a]						
GWS1B 3 Kb	3054/3239	94.3	1629/1689	96.4	2229/2654[c]	84
GWS6B 4.5 Kb	4093/4324	94.7	2301/2563	90	2973/3571[c]	83
GWS6D 6	4088/4336	94.3	1881/2138	88	2533/3096[c]	82

	Scaffold GWB2_18588		Scaffold GWD2_39307		Scaffold GWA2_17992	
	(7.1 Kb)		(2.3 Kb)		(9 Kb)	
	No. of identical residues/total no.	**% identity**	**No. of identical residues/total no.**	**% identity**	**No. of identical residues/total no.**	**% identity**
Kb						
ORFs[b]						
GWS1B CbbM	458/460	99.6	442/460	96.3	455/460	99.1
GWS6B/D CbbM	458/460	99.6	440/460	96	455/460	99.1
GWS1B CbbR	301/306	98.4	n/a[d]	n/a[d]	270/306	90.6
GWS6B/D CbbR	303/306	99	n/a[d]	n/a[d]	268/306	89.9

- a nt sequence comparisons.
- b aa sequence comparisons; note that GWS6B and 6D are identical aa sequences for CbbM and CbbR.

- c Excluding PilY1 signal peptide and peptide chain release factor nt regions in alignment comparisons.
- d Scaffold ends ∼200 bp upstream of the *cbbM* gene and thus does not have a *cbbR* gene for comparison.

Because the uncultivated *Gallionellaceae* RubisCO sequences represent a poorly characterized group of form II RubisCOs, the GWS1B CbbM structural and biochemical properties were further analysed. Recombinant hexahistidine-tagged GWS1B RubisCO was purified from *E. coli* (Fig. S3A), and its enzymatic properties were evaluated *in vitro*. The specific activity was 1 (±0.1) μmol/min/mg at pH 8.0 and 2.8 (±0.3) μmol/min/mg at pH 7.2, and the substrate specificity factor (Ω) was determined to be 12 (±0.3) (Table S2). These values are highly similar to those obtained for the well-studied form II RubisCO enzymes from *Rhodopseudomonas palustris* 010 (71% aa identity) and *Rhodospirillum rubrum* ATCC 11170 (66% aa identity) (Satagopan *et al*., 2014).

Non-denaturing gel electrophoresis indicated that the GWS1B enzyme had a marked propensity to oligomerize into larger molecular weight moieties, which was not observed for *R. rubrum* or *R. palustris* enzymes under similar electrophoretic conditions (Fig. S3B). The oligomeric state of GWS1B was further evaluated by size exclusion chromatography at pH 8.0 (Fig. S4) and with size-exclusion chromatography with multi-angle light scattering (SEC-MALS) at pH 7.2 (Fig. S5). Contrary to the non-denaturing gel electrophoresis results, it could be concluded from these experiments that the enzyme assembles as an unusual hexameric structure with a molecular mass of ∼320 kDa, similar to the recently studied *R. palustris* form II enzyme (Satagopan *et al*., 2014).

The hexameric structure was further substantiated when the crystal structure of the enzyme was solved both in the ‘apo’ form (i.e. with no ligands) and the activated form complexed with 2-carboxyarabinitol 1,5-bisphosphate (CABP), a transition-state substrate analogue (Fig. 5 and Table S3). An alignment of the apo and liganded monomer backbone traces of the GWS1B crystal structures indicated that they are similar (Fig. 5A). Consistent with the relatively high-observed levels of sequence identity between them (> 65% aa identity), the tertiary structure of a monomeric GWS1B subunit aligns well with a subunit from either the *R. palustris* or the *R. rubrum* enzyme (Fig. 5B and C). The strong structural similarity between the CABP-bound forms of GWS1B and *R. palustris* CbbM is reflected in the ‘closed’ conformation for loop 6 (Fig. 5C), an important structural element in catalysis, and a virtually identical conformation for the form II-specific C-terminal domain. As with the *R. palustris*enzyme, the carboxy terminus points away from the loop 6 region (Fig. 5C), and the active-site geometry is virtually identical in both the CABP-bound *R. palustris* and GWS1B enzymes (Fig. 5D). Thus, despite apparent differences in amino acid sequence identity (∼29%) and non-denaturing gel molecular weight sizes, our further biochemical analyses support that the GWS1B and *R. palustris* RubisCOs are similar in activity levels, specificity values and structure. Overall, the discovery of a second hexameric form II enzyme suggests that hexameric assemblies of form II enzymes may be more prevalent than originally thought.

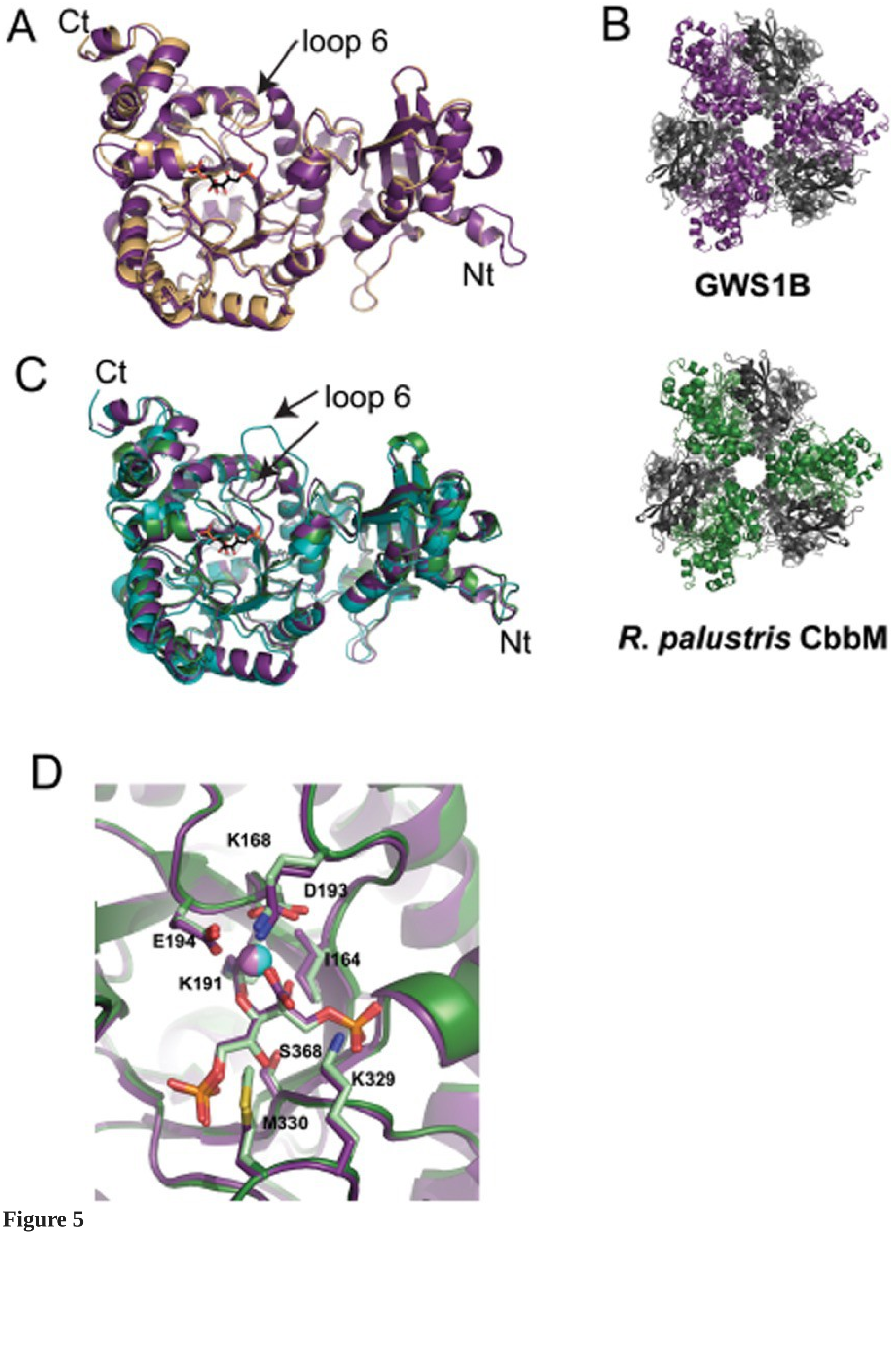

A
Ct
loop 6
Nt
B
GWS1B
R. palustris CbbM
C
Ct
loop 6
Nt
D
K168
D193
E194
I164
K191
S368
K329
M330

Figure 5

Open in figure viewerPowerPoint

Structural characteristics of form II RubisCO (GWS1B) from uncultivated *Gallionellaceae* species.

A. Superposition of activated apo (PDB code 5C2C; orange) and CABP-bound (PDB code 5C2G; purple) GWS1B form II RubisCO. The N-, C-termini and loop 6 are labelled.

B. Top views down the threefold symmetry axis of activated, CABP-bound hexameric form II RubisCO from uncultivated *Gallionellaceae* sp. (PDB code 5C2G, alternating subunits in purple and grey) and *R. palustris* (PDB code 4LF1, alternating subunits in green and grey).

C. Superimposition of a catalytic large subunit from activated, CABP-bound, x-ray crystal structures of form II CbbM RubisCO from *R. rubrum* (PDB code, 9RUB; teal), *R. palustris* (PDB code, 4LF1; dark green) and GWS1B enzyme (PDB code 5C2G, purple) characterized in this study. For clarity, CABP, the transition-state substrate analogue, bound in the active site of GWS1B is shown in stick representation in black. The loop 6, N- and C-termini of GWS1B are labelled for comparing of their positions in the other two structures.

D. Stereo view of the superimposed active sites of activated, CABP-bound GWS1B (purple) and *R. palustris* (green) form II RubisCOs. Active-site residues and CABP are shown in stick representation in lighter shades. Numbering of active site residues is according to the GWS1B amino acid sequence. Magnesium ion is shown as a sphere and coloured pink (GWS1B) or cyan (*R. palustris*). Residues from both subunits in a RubisCO dimer contribute to the active site but, for clarity, only residues from one of the subunits are shown in each structure.

Conclusions

Functional metagenomics selection can be used to recover catalytically active RubisCOs from multiple microbial communities adapted to a wide range of environmental conditions. We recovered several new RubisCO variants and characterized one such enzyme in some detail, shedding valuable structure-function insights into currently under-sampled microbial functional diversity. Since the bulk of microbial diversity in the environment is uncultivated and unknown, the present results and other metagenomic studies (Wrighton *et al*., 2012; Campbell *et al*., 2013; Castelle *et al*., 2015) suggest it is likely that much of the diversity pertaining to microbial RubisCO structure and function is still yet to be discovered. Functional selection offers a straightforward way to access and obtain enzyme variants from uncultivated bacteria that have been subjected to billions of years of evolution. This study represents an important first step in selecting specific enzymes adapted to function under highly oxic environments since most microbial RubisCO enzymes are either sequestered and protected from oxygen inhibition or possess low tolerance to oxygen due to optimally functioning under anaerobic or semi-anaerobic conditions. Selections could be refined using RubisCO-deletion host strains under specific CO_2 and O_2 regimes. The isolation of such enzymes will complement and enhance current studies for understanding RubisCO catalysis and ultimately provide advantageous catalysts for biosynthetic pathway optimization for CO_2 gas removal and conversion to useful end products.

Experimental procedures

Plasmids, strains and growth conditions

The *E. coli* One Shot Top10 strain (Invitrogen) was used for all cloning and transformation steps, the *E. coli* pRK2013 helper plasmid was used for tri-parental matings and the *E. coli*BL21 (DE3) strain for gene expression. *Escherichia coli* strains were grown on lysogeny broth (LB) broth or agar plates at 37°C supplemented with the appropriate antibiotics. Laboratory culture strains of wild-type *R. capsulatus* SB1003, *R rubrum* ATCC 11170, *R. palustris* 010, *R. sphaeroides* 2.4.1 and *R. eutropha* H16 were used for test DNA extractions and simultaneous recovery of functional RubisCO genes from these five organisms. These organisms were grown with shaking at 30°C under chemoheterotrophic conditions on liquid peptone yeast extract (PYE), LB broth, or on Ormerod's defined minimal medium (Ormerod *et al*., 1961) supplemented with malate.

The selection host, strain SB I/II$^-$, is a RubisCO deletion mutant strain of *R. capsulatus* SB1003, a purple non-sulfur photosynthetic organism capable of heterotrophic and autotrophic growth strategies. In this mutant strain, both endogenous form I and form II RubisCO genes have been inactivated with partial deletions and insertion of kanamycin and spectinomycin-resistant cassettes (Smith and Tabita, 2003).

Chemoheterotrophic growth of strain SB I/II$^-$ was carried out at 30°C in the dark either on PYE plates or shaking in liquid PYE with 3 mM KCl and 10 mM NaCl supplemented with kanamycin (25 μg/ml) and spectinomycin (25 μg/ml) as described previously (Smith and Tabita, 2003; Satagopan *et al*., 2009; 2014). For PA growth complementation, Ormerod's minimal medium plates with no organic carbon source were incubated in illuminated anaerobic sealed jars, which were periodically flushed with a gas mixture of 5% CO_2/H_2. Liquid Ormerod's minimal medium was used for anaerobic PA cultures, which were bubbled continuously with 5% CO_2/H_2 for growth analysis and harvested for RubisCO activity assays, as described below. For aerobic CA growth complementation, sealed jars were periodically flushed with a gas mixture of 2.5% or 5% CO_2/47.5% or 45% H_2, mixed with an equal proportion of air and grown in the dark.

The cloning vector, p83 (Fig. 1C), was constructed as part of this study from a modified version of the broad-host range plasmid 3716 (courtesy of Oliver Lenz and Bärbel Friedrich). The p83 plasmid has the *R. rubrum* form II *cbbM* promoter region and the *cbbR* gene (Smith and Tabita, 2003) on one side of the multiple cloning site and the form I RubisCO (*cbbLS)*promoter from *R. capsulatus* SB1003 (Vichivanives *et al*., 2000) on the other side of the multiple cloning site (Fig. 1C). Vector p83 also harbours a tetracycline antibiotic resistance cassette.

Environmental and enrichment samples

For autotrophic enrichment samples, water and soil samples were collected from the Olentangy River and embankment on The Ohio State University campus in Columbus, OH in October and December 2011 as described (Dourado-Ribeiro, 2013). Enrichment cultures were grown similarly as described above with the following modifications. Samples were inoculated into 1 L bottles of Ormerod's minimal medium and subjected to PA growth under an atmosphere of 20% CO_2/H_2 in the light or CA growth in the dark under an atmosphere of 10% CO_2/H_2 mixed with an equal proportion of air. When enrichment cultures reached an OD_{660nm} of > 1.0, the cultures were saved as frozen stocks. Enrichment cultures were resurrected from frozen stocks, inoculated into fresh medium and incubated under similar PA and CA growth conditions. These cultures were then harvested at an OD_{660nm} of > 1.0 for DNA extractions and combined in equal concentrations for downstream analyses.

The photosynthetic biofilm sample was collected from direct scrapings of rocks and cobble from Paoha Island hot springs located in Mono Lake, CA, an alkaline, hypersaline lake (Kulp *et al*., 2008; Hoeft *et al*., 2010) and the scrapings stored in a liquid mineral salts medium (Kulp *et al*., 2008) under N_2 in a sealed vial.

Groundwater aquifer samples were collected from the Rifle Integrated Field Research Challenge site in Rifle, CO as described (Wrighton *et al*., 2012; Brown *et al*., 2015; Castelle *et al*., 2015). Approximately 100 L of water were concentrated onto 0.2 μM Supor filters (Pall), using a 1.2 μM pre-filter, at sequential time points over a 4-month acetate-amended experiment from August to November 2011. The time point and filter used for this present study was the 0.2 μM filter from 9 September 2011, 13 days after the start of the acetate amendment when biomass was high.

Construction of metagenomic libraries

Genomic DNA extractions were conducted according to previous extraction methods (Zhou*et al*., 1996; Crump *et al*., 2003) with the following modifications. Five to 10 ml of pelleted cells or

∼1.5–2.0 g of filter were re-suspended in a ‘modified DNA Extraction Buffer’ [mDEB, 100 mM Tris-Cl (pH 7.5), 50 mM Na_2EDTA (pH 8.0) and 50 mM sodium phosphate buffer (pH 7.5)]. Samples were amended with lysozyme (final concentration 2 mg/ml) and freeze/thawed (80°C/37°C) three times followed by several enzymatic lysis incubations (Zhou *et al*., 1996; Crump *et al*., 2003). Following these cell lysis steps, DNA was further extracted twice with equal volumes of phenol : chloroform : isoamyl alcohol and twice with chloroform : isoamyl alcohol. DNA was precipitated overnight at room temperature in 0.6 volumes of isopropanol and linear polyacrylamide carrier. Precipitated DNA was washed twice in 70% ethanol, after which the pellet was dried and re-suspended in 0.2 x TE buffer pH 8. Extracted DNA samples were verified for high MW composition via agarose ethidium bromide gel electrophoresis and quantified on a Nanodrop (ThermoScientific) and a Qubit flourometer (Life Technologies). Approximately 2 μg of DNA were used per Sau3AI (NEB) partial restriction digest with 1–2U of enzyme for a 7 min 37 °C incubation.

In order to remove contaminants and small DNA fragments (< 1 Kb) from the Sau3AI partial digests, between 8 μg and 20 μg of the digested DNA was purified and size fractionated on a Shimadzu HPLC using a Waters Gen-Pak FAX anion-exchange column per manufacturer's instructions. Eluted fractions were collected and the size distribution in each of the fractions analysed via ethidium bromide or SYBR Gold agarose gel electrophoresis. Selected fractions representing the desired size range (> 1 Kb and < 8 Kb) were then combined and concentrated using an Amicon 30 kDa (Millipore) column.

The HPLC/Amicon-concentrated Sau3AI-restricted DNA was ligated into the BamHI (NEB) digested and phosphatased (NEB Antarctic phosphatase) dual promoter p83 vector. Ligations were carried out in 10–40 μl volumes using T4 DNA ligase and buffer (NEB) and incubated overnight at 16°C. Typically, insert::vector ratios of 2:1 were used (assuming an average insert size of 3 Kb). Ligations were desalted using a DNA clean and concentrator kit (Zymo).

Functional growth selection

Desalted plasmid p83 ligated DNA was transformed into electrocompetent Top10 (Invitrogen) *E. coli* cells using an Eppendorf Electroporator 2510 (Eppendorf). Transformations were incubated for 1 h at 37°C shaking in super optimal broth with catabolite repression (SOC) medium without antibiotics prior to a 4–6 h tetracycline (tet) (12.5 μg/ml) enrichment; this ensures enrichment of *E. coli* cells with transformed p83 tet resistant plasmid prior to mating with strain SB I/II$^-$. A small aliquot of every transformation and post-enrichment transformation mixtures were plated onto LB-tet plates (12.5 μg/ml) to verify transformation efficiencies and tet enrichment efficiency (Table S4).

Following the tet enrichment and for each experiment, the ∼100 000 electrocompetent *E. coli*member library clone pool (Table S4) was conjugated in a tri-parental mating, as previously described (Smith and Tabita, 2003), with strain SB I/II$^-$ and the *E. coli* strain carrying the pRK2013-helper plasmid kept overnight at 30°C on PYE plates. The following day, matings were re-suspended and washed in Ormerod's minimal medium, and a small aliquot was plated onto PYE supplemented with tetracycline [2 μg/ml] to verify that the mating was successful; all SB I/II$^-$ cells that received the tet plasmid grew under these conditions. The rest of the re-suspended mating was plated onto Ormerod's minimal medium plates and placed into sealed jars for either photoautotrophic or chemoauototrophic growth selection. Only colonies with a cloned fragment containing a functional RubisCO gene allowed for growth under these selection conditions. For all experiments, the p83::*R. rubrum cbbM*construct served as a positive control, and the p83 empty vector served as a negative control. Colonies that arose on plates were re-streaked to gain enough material to re-isolate the plasmid for DNA sequencing and to obtain individual colonies for starting liquid autotrophic growth.

Bioinformatic analyses

Plasmid encoding functional RubisCOs were sequenced using primer walking off the ends of the dual promoter vector into the fragment inserts with Sanger sequencing at The Ohio State Plant-Microbe Genomics Facility and analysed using NCBI BLAST for homology, BioEditfor ClustalW multiple sequence alignments and SnapGene Viewer 2.7.1 or Geneious version 7 for manual ORF annotations. For some recombinant clones recovered from the enrichment samples, only the partial RubisCO gene sequences were obtained using either a conserved form I forward primer (Alfreider *et al*., 2003) or form II forward primer (Kato *et al*., 2012).

For the Rifle groundwater aquifer sample selection, recovered RubisCO fragments were compared with assembled genomic scaffolds; information regarding the Illumina sequencing was described previously (Brown *et al*., 2015). Similar scaffolds were retrieved via in-house BLASTN or BLASTP searches using the recovered RubisCO fragment as the query and subsequently analysed in Geneious for alignments and manual ORF annotations. The RubisCO amino acid tree was made in Geneious using the Jukes–Cantor genetic distance model and the neighbour-joining tree building method. *Bacillus subtilis* form IV RubisCO served as the outgroup. The tree was re-sampled using bootstrapping with 100 replicates.

Phylogenetic analysis of the *Gallionellaceae* bins (GWA2_Gallionellales_59_43, GWB2_Gallionellales_58_32) was performed using a syntenic block of 16 universal ribosomal proteins (RP) (L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24 and S3, S8, S10, S17, S19). Each ribosomal protein (amino acid) was aligned along with 44 reference sequences using MUSCLE(Edgar, 2004) with default parameters. Alignments were manually curated to trim start and end gaps and remove ambiguously aligned regions. Individual ribosomal protein alignments were concatenated in GENEIOUS version 7 (Kearse *et al*., 2012). In total, the alignment of 46 sequences spanned 13 087 columns. Phylogenetic analysis of RP was inferred by RAXML(Stamatakis, 2014) using the PROTGAMMAGTR algorithm with a total of 100 bootstraps. RAXML was called as follows:raxmlHPC-PTHREADS -f a -s input -n result -m PROTGAMMAGTR -x 777 -# 100 -p 333. *Brevundimonas subvibrioides* ATCC 15264 (*Alphaproteobacteria*) (Genbank: CP002102) was included as the root for the tree. The entire ribosomal protein alignment used in phylogenetic analysis is provided in fasta format as Supplementary Data File S1.

RubisCO gene expression and purification of recombinant proteins

Hexa-histidine purified proteins were obtained as previously described (Satagopan *et al*., 2009; 2014). Briefly, genes were expressed via IPTG induction using pET-28a constructs, with the induction initiated at an OD_{600nm} between 0.5 and 0.7. After overnight growth at 25°C post-induction, the cells were rinsed and re-suspended in 50 mM bicine-NaOH, 10 mM $MgCl_2$, at pH 8.0, supplemented with 2 mM $NaHCO_3$, 1 mM DTT and 15 mM imidazole and lysed using a French-pressure cell at 1000 psi. The proteins were clarified using high-speed centrifugation and purified by one-step nickle-affinity gravity-flow chromatography (Ni-NTA Agarose, Qiagen). The GWS1B enzyme was further purified and size fractionated on a Superose 6 (GE Healthcare) gel filtration or a Q sepharose HP ion exchange column. Eluted fractions with peak 280 nm absorbance were pooled, concentrated and tested for RubisCO activity. All assay determinations are the average of two different enzyme preparations done in technical duplicates or triplicates. The *R. palustris* CbbM, *R. rubrum* CbbM and *R. eutropha*CbbLS purified recombinant hexahistidine purified enzymes served as positive controls for downstream activity, specificity and gel filtration assays. Assays were run in technical duplicates or triplicates as necessary.

SEC-MALS

A 1 mg sample (100 μL sample at 10 mg/ml) of the GWS1B RubisCO protein was analysed by size-exclusion chromatography with inline static light-scattering and refractive index detectors.

The protein sample was injected onto a Superdex 200 10/300 GL analytical size-exclusion column (GE Healthcare) equilibrated in 20 mM Tris-Cl, pH 7.2, 300 mM NaCl, 10% glycerol, 10 mM $MgCl_2$, 20 mM $NaHCO_3$ using an AKTA purifier (GE Healthcare) and a flow rate of 0.3 ml/min. Static light scattering was measured with a miniDAWN TREOS detector, and refractive index was measured using an Optilab T-rEX detector (both from Wyatt Technology, Santa Barbara, CA). The eluted protein peak was analysed using ASTRA software (Wyatt Technology) to determine molecular mass.

Crystallization and structure determination of GWS1B RubisCO

For crystallization screening, hexahistidine-tagged protein was purified as described above with the exception that the size exclusion chromatography step used a HiLoad 16/600 Superdex 200 column (GE Healthcare) equilibrated in crystallization buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 10% glycerol, 10 mM $MgCl_2$, 20 mM $NaHCO_3$). The protein was concentrated to ∼21 mg/ml and subjected to crystallization screening at 18°C using the hanging-drop vapour diffusion method and a variety of commercially available crystallization screens. Full details of the crystallization and structural determination methodology of the GWS1B RubisCO can be found in the Supporting Information.

RubisCO enzyme assays

RubisCO activity and substrate specificity assays were performed as previously described (Satagopan *et al*., 2009; 2014). For specific activities, cell pellets or purified proteins were re-suspended in 50 mM bicine-NaOH, 10 mM $MgCl_2$, at pH 8.0 or 50 mM Tris-Cl, 10 mM $MgCl_2$, 1 mM EDTA, at pH 7.2, both supplemented with 2 mM $NaHCO_3$ and 1 mM DTT. Crude protein lysates were obtained via sonication. Assay reaction mixtures contained 50 mM $NaHCO_3$, 2 μCi $NaH^{14}CO_{3,}$ 10 mM $MgCl_2$ and 0.8 mM RuBP in 50 mM bicine-NaOH buffer, pH 8.0. Reactions were initiated with the addition of RuBP at 30°C in 5 min. time course reactions terminated with acid. Assays with enzyme but without the addition of RuBP served as negative controls. Assays were run in either duplicate or triplicate; exceptions are noted. For substrate specificity assays (Ω), ∼50 μg of purified proteins were used under saturating O_2concentrations (1170 μM) and in the presence of 400 μM [1-^{3}H] RuBP. Product areas were fractionated and integrated via a MonoQ ion exchange column (GE Healthcare) using HPLC (Shimadzu). Peaks corresponding to 3-PGA and 2-PG were identified using β-ram detection (Lab Logic), and the peak areas were obtained after integrating the peaks using a radiochromatography data collection and analysis LAURA software (Lab Logic). These integrated peak areas were used to calculate substrate specificity factor values as described previously (Smith and Tabita, 2003; Satagopan *et al*., 2009; 2014).

Acknowledgements

We wish to thank O. Lenz and B. Friedrich for plasmid 3716; R. Daly for DNA extraction advice; A. Dangel for CbbR binding analysis (see Supplementary section); A. Shin for assistance with protein purification; B. Amer for assistance with SEC-MALS; M. Collazo (UCLA Crystallization Core Facility) for screening crystallization conditions; D. Cascio for expert advice on x-ray crystallography; and D. Eisenberg for his generous support for the project. This work was supported by grant GM095742 to F.R. Tabita from the National Institutes of Health. J. North was supported by the National Institutes of Health under the Ruth L. Kirschstein National Research Service Award (F32GM109547) from the National Institute of General Medical Sciences. Structural studies of GWS1B RubisCO were performed at the UCLA-DOE Protein Expression Technology Center and UCLA-DOE X-ray Crystallography Core Facility, both supported by the Department of Energy Grant DE-FC02-02ER63421.

Data accessibility

RubisCO sequences obtained from this study were deposited in NCBI with accession numbers (KT749899 – KT749915). Protein Data Bank Accession numbers: 5C2C (GWS1B apo) and 5C2G (GWS1B CABP bound). Genomes are also available through ggKbase: GWA2_Gallionellales_59_43 (http://ggkbase.berkeley.edu/organisms/1596), GWB2_Gallionellales_58_32 (http://ggkbase.berkeley.edu/organisms/2084). ggKbase is a 'live' site, thus genomes may be improved after publication.