Functional reconstitution of a bacterial CO_2 concentrating mechanism in *E. coli*

Authors: Avi I. Flamholz[1], Eli Dugan[1], Cecilia Blikstad[1], Shmuel Gleizer[2], Roee Ben-Nissan[2], Shira Amram[2], Niv Antonovsky[2,†], Sumedha Ravishankar[1,‡], Elad Noor[2,§], Arren Bar-Even[3], Ron Milo[2,*] & David F. Savage[1,*]

Affiliations:
[1] Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, United States
[2] Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[3] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany
[†] Present address: Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, NY, 10065, USA
[‡] Present address: Division of Biological Sciences, Section of Molecular Biology, University of California, San Diego, La Jolla, California, 92093, United States.
[§] Present address: Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule Zürich, Zürich CH-8093, Switzerland

*Correspondence to: ron.milo@weizmann.ac.il (R.M.) and savage@berkeley.edu (D.F.S.)

Abstract

Many photosynthetic organisms employ a CO_2 concentrating mechanism (CCM) to increase the rate of CO_2 fixation via the Calvin cycle. CCMs catalyze ≈50% of global photosynthesis, yet it remains unclear which genes and proteins are required to produce this complex adaptation. We describe the construction of a functional CCM in a non-native host, achieved by expressing genes from an autotrophic bacterium in an engineered *E. coli* strain. Expression of 20 CCM genes enabled *E. coli* to grow by fixing CO_2 from ambient air into biomass, with growth depending on CCM components. Bacterial CCMs are therefore genetically compact and readily transplanted, rationalizing their presence in diverse bacteria. Reconstitution enabled genetic experiments refining our understanding of the CCM, thereby laying the groundwork for deeper study and engineering of the cell biology supporting CO_2 assimilation in diverse organisms.

One Sentence Summary

A bacterial CO_2 concentrating mechanism enables *E. coli* to fix CO_2 from ambient air.

Introduction

Nearly all carbon in the biosphere enters by CO_2 fixation in the Calvin-Benson-Bassham cycle (Raven et al., 2017). Ribulose Bisphosphate Carboxylase/Oxygenase - commonly known as rubisco - is the CO_2 fixing enzyme in this cycle (Wildman, 2002) and likely the most abundant enzyme on Earth (Bar-On and Milo, 2019). As rubisco is abundant and central to biology, one might expect it to be an exceptional catalyst, but it is not. Photosynthetic rubiscos are modest enzymes, with carboxylation turnover numbers (k_{cat}) ranging from 1-10 s^{-1} (Flamholz et al., 2019). Moreover, all known rubiscos catalyze a competing oxygenation of the five-carbon organic substrate, ribulose 1,5-bisphosphate (Bowes and Ogren, 1972; Flamholz et al., 2019).

Rubisco arose > 2.5 billion years ago, when Earth's atmosphere contained little O_2 and abundant CO_2 (Fischer et al., 2016; Shih et al., 2016). In this environment, rubisco's eponymous oxygenase activity could not have hindered carbon fixation or the growth of CO_2-fixing organisms. Present-day atmosphere, however, poses a problem for plants and other autotrophs: their primary carbon source, CO_2, is relatively scarce (≈0.04%) while a potent competing substrate, O_2, is abundant (≈21%).

CO_2 concentrating mechanisms (CCMs) arose multiple times over the last 2 billion years (Flamholz and Shih, 2020; Raven et al., 2017) and overcome this problem by concentrating CO_2 near rubisco (Figure 1A). In elevated CO_2 environments most active sites are occupied with CO_2 and not O_2. As such, high CO_2 increases the rate of carboxylation and competitively inhibits oxygenation (Bowes and Ogren, 1972) thereby improving overall carbon assimilation (Figure 1B). Today, at least four varieties of CCMs are found in plants, algae and bacteria (Flamholz and Shih, 2020; Raven et al., 2017), organisms with CCMs are collectively responsible for ≈50% of global net photosynthesis (Raven et al., 2017), and some of the most productive human crops (e.g. maize and sugarcane) rely on CCMs.

CCMs are particularly common among autotrophic bacteria: all Cyanobacteria and many Proteobacteria have CCM genes (Kerfeld and Melnicki, 2016; Rae et al., 2013). Bacterial CCMs rely on two crucial features: (i) energy-coupled inorganic carbon uptake at the cell membrane and (ii) a 200+ MDa protein organelle called the carboxysome that encapsulates rubisco with a carbonic anhydrase enzyme (Mangan et al., 2016; McGrath and Long, 2014). In the prevailing model of the carboxysome CCM, inorganic carbon uptake produces a high, above-equilibrium cytosolic HCO_3^- concentration (≈30 mM) that diffuses into the carboxysome, where it is converted by into a high carboxysomal CO_2 concentration by carbonic anhydrase (Figure 1A).

As CCMs accelerate CO_2 fixation, there is great interest in transplanting them into crops (Ermakova et al., 2020; McGrath and Long, 2014). Carboxysome-based CCMs are especially attractive because they natively function in single cells and appear to rely on a tractable number of genes (Lin et al., 2014; Long et al., 2018; Occhialini et al., 2016; Orr et al., 2020). Modeling suggests that introducing bacterial CCM components could improve plant photosynthesis (McGrath and Long, 2014), especially if aspects of plant physiology can be modulated via genetic engineering (Wu et al., 2019). However, expressing bacterial rubiscos and carboxysome

components has, so far, uniformly resulted in transgenic plants displaying impaired growth (Lin et al., 2014; Long et al., 2018; Occhialini et al., 2016; Orr et al., 2020). More generally, as our understanding of the genes and proteins participating in the carboxysome CCM rests mostly on loss-of-function genetic experiments in native hosts (Cai et al., 2009; Desmarais et al., 2019; Marcus et al., 1986; Price and Badger, 1989a), it is possible that some genetic, biochemical and physiological aspects of CCM function remain unappreciated. We therefore sought to test whether current understanding is sufficient to reconstitute the bacterial CCM in a non-native bacterial host, namely *E. coli*.

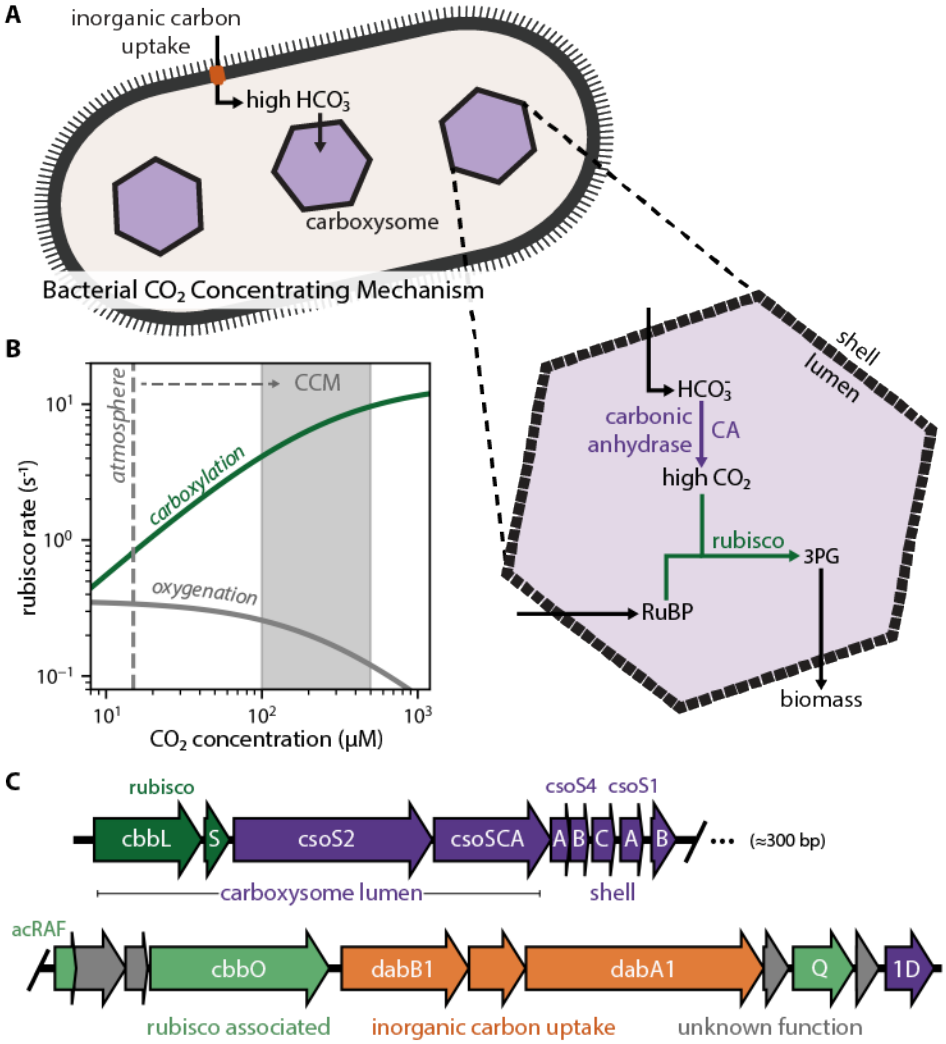


Figure 1. Twenty genes form the basis of a bacterial CCM. *(**A**) The bacterial CCM consists of at least two essential components - energy-coupled carbon uptake and carboxysome structures that encapsulate rubisco with a carbonic anhydrase (CA) enzyme (Mangan et al., 2016; McGrath and Long, 2014). Transport generates a large cytosolic HCO_3^- pool, which is rapidly converted to high carboxysomal CO_2 concentration by the carboxysomal CA. (**B**) Elevated CO_2 increases the rubisco carboxylation rate (green) and suppresses oxygenation by competitive inhibition (grey). [O_2] was set to 270 μM for rate calculations. (**C**) H. neapolitanus CCM genes are mostly contained in a 20 gene cluster (Desmarais et al., 2019) expressing rubisco and its associated chaperones (green), carboxysome structural proteins (purple), and an inorganic carbon transporter (orange).*

Using a genome-wide screen in the CO_2-fixing proteobacterium *H. neapolitanus*, we recently demonstrated that a 20-gene cluster encodes all activities required for the CCM, at least in principle (Desmarais et al., 2019). These genes include rubisco large and small subunits, the carboxysomal carbonic anhydrase, seven structural proteins of the α-carboxysome (Bonacci et al., 2012), an energy-coupled inorganic carbon transporter (Desmarais et al., 2019; Scott et al., 2019), three rubisco chaperones (Aigner et al., 2017; Mueller-Cajar, 2017; Wheatley et al., 2014), and four genes of unknown function (Figure 1C). We aimed to test whether these genes are sufficient to establish a functioning CCM in *E. coli*.

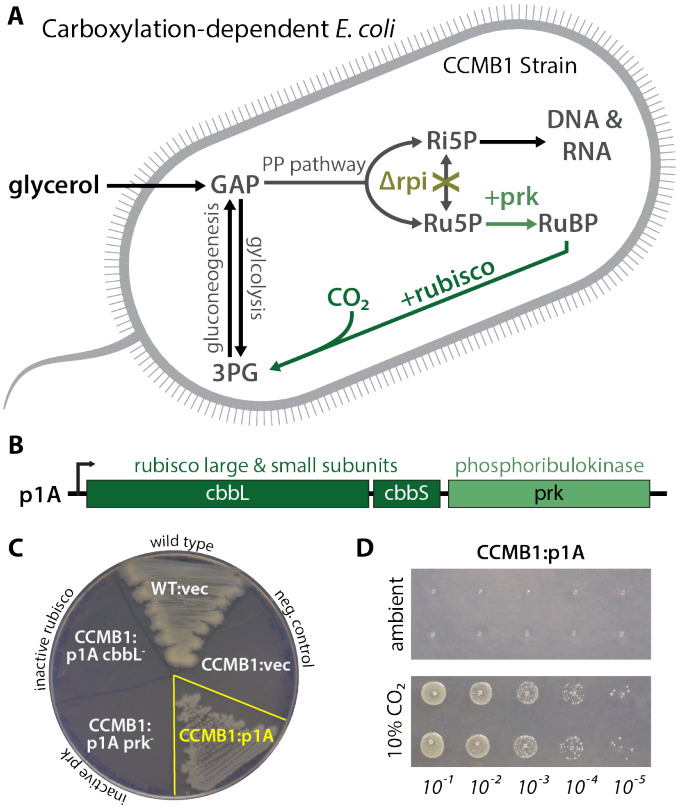
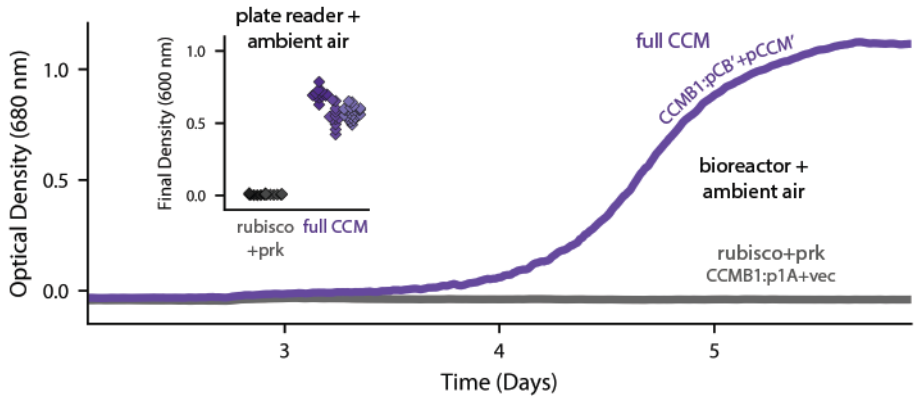


Figure 2. CCMB1 depends on rubisco carboxylation for growth on glycerol. *(**A**) Ribose-5-phosphate (Ri5P) is required for nucleotide biosynthesis. Deletion of ribose-phosphate isomerase (Δrpi) in CCMB1 blocks ribulose-5-phosphate (Ru5P) metabolism in the pentose phosphate (PP) pathway. Expression of rubisco (H. neapolitanus cbbLS) and phosphoribulokinase (S. elongatus PCC7942 prk) on the p1A plasmid (**B**) permits Ru5P metabolism, thus enabling growth on M9 glycerol media in 10% CO_2 (**C**). Mutating the rubisco active site (p1A cbbL⁻) abrogates growth, as does mutating ATP-binding residues of prk (p1A prk⁻). (**D**) CCMB1:p1A grows well under 10% CO_2, but fails to grow in ambient air. Cells grown on M9 glycerol media throughout. Acronyms: ribulose 1,5-bisphosphate (RuBP), 3-phosphoglycerate (3PG).*

Results

As *E. coli* is a heterotroph, consuming organic carbon molecules to produce energy and biomass, it does not natively rely on rubisco. Therefore, in order to evaluate the effect of heterologous CCM expression, we first designed an *E. coli* strain that depends on rubisco carboxylation for growth. To grow on glycerol as the sole carbon source, *E. coli* must synthesize ribose 5-phosphate (Ri5P) for nucleic acids. Synthesis of Ri5P via the pentose phosphate pathway forces co-production of ribulose 5-phosphate (Ru5P). Deletion of ribose 5-phosphate isomerase (*rpiAB* genes, denoted Δrpi), however, makes Ru5P a metabolic "dead-end" (Figure 2A). Expression of phosphoribulokinase (*prk*) and rubisco creates a "detour" pathway converting Ru5P and CO_2 into two units of the central metabolite 3-phosphoglycerate (3PG), enabling Ru5P metabolism and growth (Figure 2A). Additionally, cytosolic carbonic anhydrase activity is incompatible with the

141 bacterial CCM (Price and Badger, 1989b). We therefore constructed a strain, named CCMB1 for
142 "**CCM B**ackground **1**", lacking *rpiAB* and all endogenous carbonic anhydrases (Methods).
143
144 As predicted, CCMB1 required rubisco and prk for growth on glycerol minimal media in 10% CO_2
145 (Figure 2B-C). Even when expressing rubisco and *prk* on the p1A plasmid (Figure 2B), however,
146 CCMB1 failed to grow on glycerol in ambient air, presumably due to insufficient carboxylation at
147 low CO_2 (Figures 2D and S3-4). That is, CCMB1:p1A displays the "high-CO_2 requiring" phenotype
148 that is the hallmark of CCM mutants (Marcus et al., 1986; Price and Badger, 1989a).
149
150 We expected that expressing a functional CO_2-concentrating mechanism would cure CCMB1 of
151 its high-CO_2 requirement and permit growth in ambient air. We therefore generated two plasmids,
152 pCB and pCCM, that together express all 20 genes from the *H. neapolitanus* CCM cluster (Figures
153 1C and S5A). pCB encodes ten carboxysome genes (Bonacci et al., 2012), including rubisco large
154 and small subunits, along with *prk*. The remaining *H. neapolitanus* genes, including putative
155 rubisco chaperones (Aigner et al., 2017; Mueller-Cajar, 2017; Wheatley et al., 2014) and an
156 inorganic carbon transporter (Desmarais et al., 2019; Scott et al., 2019), were cloned into the
157 second plasmid, pCCM.
158



159
160 ***Figure 3. Expression of 20 CCM genes permits growth of CCMB1 in ambient air.*** *Time course data give*
161 *representative growth curves from a bioreactor bubbling ambient air. CCMB1:pCB' + pCCM' grows well (purple, "full*
162 *CCM"), while rubisco and prk alone are insufficient for growth in ambient air (grey, CCMB1:p1A+vec). Inset: a plate*
163 *reader experiment in biological triplicate (different shades) gave the same result. Expressing the full complement of*
164 *CCM genes led to an increase in culture density (optical density at 600 nm) of ≈0.6 units after 80 hours of cultivation.*
165 *Bootstrapping was used to calculate a 99.9% confidence interval of 0.56-0.64 OD units for the effect of expressing the*
166 *full CCM during growth in ambient air (Figure S6).*
167
168 CCMB1 co-transformed with pCB and pCCM initially failed to grow on glycerol media in ambient
169 air. We therefore conducted selection experiments, described fully in the supplement, that
170 resulted in the isolation of mutant plasmids conferring growth in ambient air. Briefly, CCMB1:pCB
171 + pCCM cultures were grown to saturation in 10% CO_2 and multiple dilutions were plated on
172 glycerol minimal media (Methods). Colonies became visible after 20 days of incubation in ambient
173 air (Figure S5). Deep-sequencing of plasmid DNA revealed mutations in regulatory sequences
174 (e.g. a promoter and transcriptional repressor) but none in sequences coding for CCM
175 components (Table S4). Individual post-selection plasmids pCB' and pCCM' were reconstructed

by PCR, resequenced, and transformed into naive CCMB1 (Methods). As shown in Figure 3, pCB' and pCCM' together enabled reproducible growth of CCMB1 in ambient air, suggesting that the 20 genes expressed are sufficient to produce a heterologous CCM without any genomic mutations.

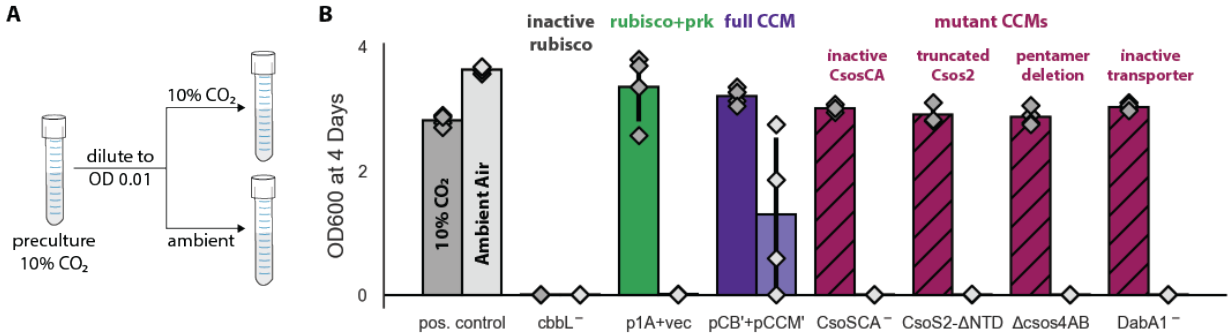


Figure 4. Growth in ambient air depends on the known components of the bacterial CCM. *We generated plasmid variants carrying inactivating mutations to known components of the CCM. (**A**) Pre-cultures were grown in 10% CO_2 and diluted into two tubes, one of which was cultured in 10% CO_2 and the other in ambient air (Methods). All strains were tested in biological quadruplicate and culture density was measured after four days. (**B**) Targeted mutations to CCM genes were found to ablate growth in ambient air while permitting growth in 10% CO_2, as expected. Each strain is described by a pair of bars. The left bar (darker color) gives the mean endpoint density in 10% CO_2. The right bar (lighter color) gives the mean in ambient air. Diamonds give individual data points and error bars give the standard deviation. From left to right: a positive control (grey, Methods) grew in both 10% CO_2 and ambient air; a negative control carrying catalytically inactive rubisco (CCMB1:pCB' cbbL⁻+pCCM') failed to grow in either condition; CCMB1 expressing rubisco and prk but no CCM genes (green, CCMB1:p1A+vec) grew only in 10% CO_2; CCMB1:pCB'+pCCM' grew in 10% CO_2 and ambient air, recapitulating results presented in Figure 3. The following four pairs of maroon bars give growth data for strains carrying targeted mutations to CCM genes: an inactivating mutation to carboxysomal carbonic anhydrase (CCMB1:pCB' CsoSCA⁻+pCCM'), deletion of the N-terminal domain of CsoS2 responsible for recruiting rubisco to the carboxysome (CCMB1:pCB' CsoS2 ΔNTD +pCCM'), deletion of both genes encoding pentameric vertex proteins (CCMB1:pCB' ΔcsoS4AB + pCCM'), and inactivating mutations to the DAB carbon uptake system (CCMB1:pCB' DabA1⁻ + pCCM'). All tested mutations to the CCM abrogated growth in air while permitting growth in 10% CO_2. The positive control is the CAfree strain expressing human carbonic anhydrase II (Methods). Detailed description of all plasmid and mutation abbreviations is given in Table S2.*

To verify that growth in ambient air depends on the CCM, we generated plasmids carrying targeted mutations to known CCM components (Figure 4). An inactivating mutation to the carboxysomal rubisco (CbbL K194M) prohibited growth entirely. Mutations targeting the CCM, rather than rubisco itself, should ablate growth in ambient air while permitting growth in high CO_2 (Desmarais et al., 2019; Mangan et al., 2016; Marcus et al., 1986; Price and Badger, 1989a; Rae et al., 2013). Consistent with this understanding, an inactive mutant of the carboxysomal carbonic anhydrase (CsoSCA C173S) required high-CO_2 for growth. Similarly, disruption of carboxysome formation by removal of the pentameric shell proteins or the N-terminal domain of CsoS2 also eliminated growth in ambient air. Removing the pentameric proteins CsoS4AB disrupts the permeability barrier at the carboxysome shell (Cai et al., 2009), while truncating CsoS2 prohibits carboxysome formation entirely (Oltrogge et al., 2020). Finally, an inactivating mutation to the inorganic carbon transporter also eliminated growth in ambient air (Desmarais et al., 2019).

These experiments demonstrate that pCB' and pCCM' enable CCMB1 to grow in ambient air in a manner that depends on the known components of the bacterial CCM. To confirm that these cells produce carboxysome structures, we performed thin section electron microscopy. Regular polyhedral inclusions of ≈100 nm diameter were visible in micrographs (Figure 5A), implying production of morphologically-normal carboxysomes.

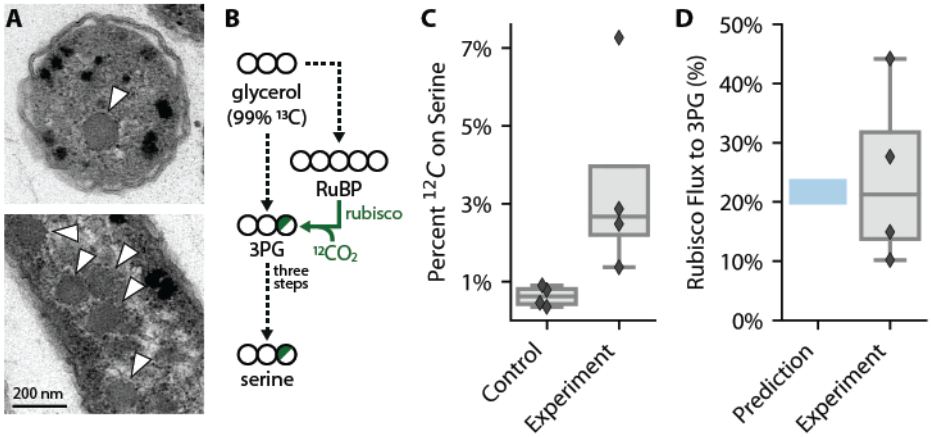


Figure 5. CCMB1:pCB'+pCCM' produces morphologically normal carboxysomes and fixes CO_2 from ambient air into biomass. *(**A**) Polyhedral bodies resembling carboxysomes are evident in electron micrographs of CCMB1:pCB'+pCCM' cells grown in air. (**B**) Cells were grown in ambient air with 99% ^{13}C glycerol as the sole organic carbon source so that $^{12}CO_2$ in air is the sole source of ^{12}C. As serine is a direct metabolic product of 3PG, we expect ^{12}C enrichment on serine when rubisco is active. 3PG also derives from glycolytic metabolism of glycerol, so complete ^{12}C labeling of serine was not expected. (**C**) The ^{12}C composition of serine from CCMB1:pCB' + pCCM' ("Experiment") is roughly threefold above the control. (**D**) The fraction of 3PG production flux due to rubisco was predicted via Flux Balance Analysis and estimated from isotopic labeling data (Methods). Estimates of the rubisco flux fraction exceed 10% for all four biological replicates and the mean estimate accords well with a ≈20% prediction.*

We next conducted isotopic labeling experiments to determine whether CCMB1:pCB' + pCCM' fixes CO_2 from ambient air into biomass. Cells were grown in minimal media with ^{13}C-labeled glycerol as the sole organic carbon source, such that CO_2 from ambient air was the dominant source of ^{12}C. The isotopic composition of amino acids in total biomass hydrolysate was analyzed via mass spectrometry (Methods). Serine is a useful sentinel of rubisco activity because *E. coli* produces serine from the rubisco product 3PG. 3PG is also an intermediate of lower glycolysis, and so the degree of ^{12}C labeling on serine reports on the balance of fluxes through rubisco and lower glycolysis (Figure 5B). We therefore expected excess ^{12}C labeling of serine when rubisco is active. Consistent with this expectation, serine from CCMB1:pCB'+pCCM' cells contained roughly threefold more ^{12}C than the rubisco-independent control (Figure 5C). We estimated the contribution of rubisco to 3PG synthesis *in vivo* by comparing labeling patterns between the rubisco-dependent experimental cultures and controls (Methods). Based on these estimates, rubisco carboxylation was responsible for at least 10% of 3PG synthesis in all four biological replicates (Figure 5D, Methods), confirming fixation of CO_2 from ambient air. As such, this work represents the first functional reconstitution of any CCM.

Reconstitution in *E. coli* enabled us to investigate which *H. neapolitanus* genes are necessary for CCM function in the absence of any regulation or genetic redundancy (i.e. genes with overlapping function) present in the native host. We focused on genes involved in rubisco proteostasis and generated plasmids lacking acRAF, a putative rubisco chaperone, or carrying targeted mutations to CbbQ, an ATPase involved in activating rubisco catalysis (Aigner et al., 2017; Mueller-Cajar, 2017; Wheatley et al., 2014). Although *acRAF* deletion had a large negative effect in *H. neapolitanus* (Desmarais et al., 2019), neither acRAF nor CbbQ were strictly required for CCMB1 to grow in ambient air. Consistent with our screen in the native host (Desmarais et al., 2019), however, *acRAF* deletion produced a substantial growth defect (Figure S7C), suggesting that the rate of rubisco complex assembly is an important determinant of carboxysome biogenesis.

Discussion

Today, CCMs catalyze about half of global photosynthesis (Raven et al., 2017), but this was not always so. Land plant CCMs, for example, arose only in the last 100 million years (Flamholz and Shih, 2020; Raven et al., 2017). Though all contemporary Cyanobacteria have CCM genes, these CCMs are found in two convergently-evolved varieties (Flamholz and Shih, 2020; Kerfeld and Melnicki, 2016; Rae et al., 2013), suggesting that the ancestor of present-day Cyanobacteria and chloroplasts did not have a CCM (Rae et al., 2013). So how did carboxysome CCMs come to dominate the cyanobacterial phylum?

Here we demonstrated that the α-carboxysome CCM from *H. neapolitanus* is readily transferred between species and confers a large growth benefit, which can explain how these CCMs became so widespread among bacteria (Kerfeld and Melnicki, 2016; Rae et al., 2013). We constructed a CCM by expressing 20 genes in an engineered *E. coli* strain, CCMB1. In accordance with its role in native autotrophic hosts (Desmarais et al., 2019; Long et al., 2018; Marcus et al., 1986; Price and Badger, 1989a), the transplanted CCM required α-carboxysomes and inorganic carbon uptake to enable CCMB1 to grow by fixing CO_2 from ambient air (Figures 3-5, S6-8). It appears, therefore, that the α-carboxysome CCM is genetically compact and “portable” between organisms. It is possible, therefore, that expressing bacterial CCMs in non-native autotrophic hosts will improve CO_2 assimilation and growth. This is a promising approach to improving plant growth (Ermakova et al., 2020; Long et al., 2016; Wu et al., 2019) and also improving microbial production of fuel, food products and commodity chemicals from CO_2 (Claassens et al., 2016; Gleizer et al., 2019).

Reconstitution also enabled us to test, via simple genetic experiments, whether particular genes play a role in the CCM (Figure S7). These experiments demonstrated that the rubisco chaperones are strictly dispensable for producing a functional bacterial CCM, though removing the *acRAF* gene produced a substantial growth defect that warrants further investigation. Further such can use our reconstituted CCM to delineate a minimal reconstitution of the bacterial CCM suitable for plant expression (Du et al., 2014; Long et al., 2018, 2016; Occhialini et al., 2016; Orr et al., 2020), test hypotheses about carboxysome biogenesis (Bonacci et al., 2012; Oltrogge et al., 2020), and probe the relationship between CCMs and host physiology (Mangan et al., 2016; Price and Badger, 1989b).

Our approach to studying CCMs by reconstitution in tractable non-native hosts can be applied to other CCMs, including β-carboxysome CCMs, the algal pyrenoid, and plausible evolutionary ancestors thereof (Flamholz and Shih, 2020). Historical trends in atmospheric CO_2 likely promoted the evolution of CCMs (Fischer et al., 2016; Flamholz and Shih, 2020), so testing the growth of plausible ancestors of bacterial CCMs (e.g. carboxysomes lacking carbonic anhydrase activity) may provide insight into paths of CCM evolution and the composition of the ancient atmosphere at the time bacterial CCMs arose. In response to these same pressures, diverse eukaryotic algae evolved CCMs relying on micron-sized rubisco aggregates called the pyrenoids (Flamholz and Shih, 2020; Wang and Jonikas, 2020). Pyrenoid CCMs are collectively responsible for perhaps 80% of oceanic photosynthesis (Mackinder et al., 2016), yet many fundamental questions remain regarding the composition and operation of algal CCMs (Wang and Jonikas, 2020). Functional reconstitution of a pyrenoid CCM is a worthy goal which, once achieved, will indicate enormous progress in our collective understanding of the genetics, cell biology, biochemistry and physical processes supporting the eukaryotic complement of oceanic photosynthesis. We hope such studies will further our principled understanding of, and capacity to engineer, the cell biology supporting CO_2 fixation in diverse organisms.

Materials and Methods

Computational design of rubisco-dependent strains

To computationally design mutant strains in which growth is coupled to rubisco carboxylation flux, we used a variant of Flux Balance Analysis (Lewis et al., 2012) called “optslope” (Antonovsky et al., 2016). Starting from a published model of *E. coli* central metabolism, the Core *Escherichia coli* Metabolic Model (Orth et al., 2010), we computationally generated all pairs of central metabolic knockouts and ignored those strains that permit growth *in silico* in the absence of rubisco and phosphoribulokinase (Prk) activities. For those knockout strains predicted to grow only when rubisco and Prk are active, we evaluated the degree of coupling between rubisco flux and biomass production during growth in nine carbon sources: glucose, fructose, gluconate, ribose, succinate, xylose, glycerate, acetate and glycerol. This approach highlighted several candidate rubisco-dependent knockout strains, including Δ*rpiAB* (Figure S1). The optslope algorithm is outlined in Figure S1 and described fully in (Antonovsky et al., 2016) and source code is available at https://gitlab.com/elad.noor/optslope.

Strains and genomic modifications

Strains used in this study are documented in Table S1. Δ*rpiAB* has the genotype Δ*rpiAB* Δ*edd* and was constructed in the *E. coli* BW25113 background by repeated rounds of P1 transduction from the KEIO collection followed by pCP20 curing of the kanamaycin selection marker (Baba et al., 2006; Datsenko and Wanner, 2000). Deletion of *edd* removes the Entner-Doudoroff pathway (Peekhaus and Conway, 1998), which forces rubisco-dependent metabolism of gluconate (Figure S2A). CCMB1 has the genotype BW25113 Δ*rpiAB* Δ*edd* Δ*cynT* Δ*can* and was constructed from Δ*rpiAB* by deleting both native carbonic anhydrases using the same methods, first transducing the KEIO Δ*cynT* and then Δ*can* from EDCM636 (Merlin and Masters, 2003), which was obtained from the Yale Coli Genetic Stock Center. Transformation was performed by electroporation (ECM 630, Harvard Biosciences) and electrocompetent stocks were prepared using standard protocols. Strain genotypes were verified by PCR, as described below.

Recombinant expression of CCM components

pFE21 are pFA31 are compatible vectors derived from pZE21 and pZA31 (Lutz and Bujard, 1997). These vectors use an anhydrotetracycline (aTc) inducible $P_{LtetO-1}$ promoter to regulate gene expression. pF plasmids were modified from parent vectors to constitutively express the tet repressor (TetR) under the P_{bla} promoter so that expression is repressed by default (Liang et al., 1999). We found that an inducible system aids in cloning problematic genes like *prk* (Wilson et al., 2018). We refer to these vectors as pFE and pFA respectively for short. The pCB plasmid is properly called pFE-CB, while pCCM is pFA-CCM. Cloning was performed by Gibson and Golden-Gate approaches as appropriate. Large plasmids (e.g. pCB, pCCM) were verified by Illumina resequencing (Harvard MGH DNA Core plasmid sequencing service) and maps were updated manually after reviewing results compiled by breseq resequencing software (Deatherage and

Barrick, 2014). Plasmids used in this study are described in Table S2 and available on Addgene at https://www.addgene.org/David_Savage/.

Growth conditions

Unless otherwise noted, cells were grown on M9 minimal media supplemented with 0.4% w/v glycerol, 0.5 ppm thiamin (10^4 dilution of 0.5% w/v stock) and a trace element mix. The trace element mix components and their final concentrations in M9 media are: 50 mg/L EDTA, 31 mM $FeCl_3$, 6.2 mM $ZnCl_2$, 0.76 mM $CuSO_4 \cdot 5H_2O$, 0.42 mM $CoCl_2 \cdot 6H_2O$, 1.62 mM H_3BO_3, 81 nM $MnCl_2 \cdot 4H_2O$. 100 nM anhydrotetracycline (aTc) was used in induced cultures. For routine cloning, 25 mg/L chloramphenicol and 60 mg/L kanamycin selection were used as appropriate. Antibiotics were reduced to half concentration (12.5 and 30 mg/L, respectively) for CCMB1 growth experiments and kanamycin was omitted when evaluating rubisco-dependence of growth as pF plasmids carrying kanamycin resistance also express rubisco. Culture densities were measured at 600 nm in a table top spectrophotometer (Genesys 20, Thermo Scientific) and turbid cultures were measured in five or tenfold dilution as appropriate in order to reach the linear regime of the spectrophotometer.

Agar plates were incubated at 37 °C in defined CO_2 pressures in a CO_2 controlled incubator (S41i, New Brunswick). For experiments in which a frozen bacterial stock was used to inoculate the culture, cells were first streaked on agar plates and incubated at 10% CO_2 to facilitate fast growth. Pre-cultures derived from colonies were grown in 2-5 mL liquid M9 glycerol media under 10% CO_2 with a matching 1 mL control in ambient air. Negative control strains unable to grow in minimal media (i.e. active site mutants of rubisco) were streaked on and pre-cultured in LB media under 10% CO_2.

Growth curves were obtained using two complementary methods: an 8-chamber bioreactor for large-volume cultivation (MC1000, PSI), and 96-well plates in a gas controlled plate reader plate (Spark, Tecan). For the 96-well format, cells were pre-cultured in the appropriate permissive media, M9 glycerol under 10% CO_2 where possible. If rich media was used, e.g. for negative controls, stationary phase cells were washed in 2x the culture volume and resuspended in 1x culture volume of M9 with no carbon source. Cultures were diluted to an OD of 1.0 (600 nm) and 250 µl cultures were inoculated by adding 5 µl of cells to 245 µl media. A humidity cassette (Tecan) was refilled daily with distilled water to mitigate evaporation during multi-day cultivation at 37 °C. Evaporation nonetheless produced irregular growth curves (e.g. Figure S3), which motivated larger volume cultivation in the bioreactor, which mixes by bubbling ambient air into each growth vessel. 80 ml bioreactor cultures were inoculated to a starting OD of 0.005 (600 nm) and grown at 37 °C to saturation. Optical density was monitored continuously at 680 nm.

Anaerobic cultivation of agar plates was accomplished using a BBL GasPak 150 jar (BD) flushed 6 times with an anoxic mix of 10% CO_2 and 90% N_2. Tenfold titers were plated on M9 glycerol media with and without 20 mM $NaNO_3$ supplementation. Because *E. coli* cannot ferment glycerol, NO_3^- was supplied as an alternative electron acceptor. Plates without NO_3^- showed no growth (Figure S4), confirming the presence of an anaerobic atmosphere in the GasPak.

387 # Phenotyping and genotyping of strains

388 As CCMB1 is a relatively slow-growing knockout strain, we occasionally observed contaminants
389 in growth experiments. We used two strategies to detect contamination by faster-growing
390 organisms (e.g. wild-type *E. coli*). As most strains grew poorly or not at all in ambient air, pre-
391 cultures grown in 10% CO_2 were accompanied by a matching 1 mL negative control in ambient
392 air. Pre-cultures showing growth in the negative control were discarded or verified by PCR
393 genotyping in cases where air-growth was plausible.
394
395 PCR genotyping was performed using primer sets documented in Table S3. Three primer pairs
396 were used to probe a control locus (*zwf*) and two target loci (*cynT* and *rpiA*). The *zwf* locus is
397 intact in all strains. *cynT* and *rpiA* probes test for the presence of the CCMB1 strain (genotype
398 BW25113 Δ*rpiAB* Δ*edd* Δ*cynT* Δ*can*). Notably, the CAfree strain (BW25113 Δ*cynT* Δ*can*) that we
399 previously used to test the activity of DAB-type transporters (Desmarais et al., 2019) is a *cynT*
400 knockout but has a wild-type *rpiA* locus, so this primer set can distinguish between wild-type,
401 CAfree and CCMB1. This was useful for some experiments where CAfree was used as a control
402 (e.g. Figures S7-8). Pooled colony PCRs were performed using Q5 polymerase (NEB), annealing
403 at 65 °C and with a 50 second extension time.

404 # Selections for growth in novel conditions

405 CCMB1 cultures carrying appropriate plasmids were first grown to saturation in rich LB media in
406 a 10% CO_2 incubator. Stationary phase cultures were pelleted by centrifugation for 10 min at
407 4,000 x g, washed in 2x the culture volume, and resuspended in 1x culture volume of M9 media
408 with no carbon source. After resuspension, multiple dilutions were plated on selective media (e.g.
409 M9 glycerol media) and incubated in the desired conditions (e.g. in ambient air) with a positive
410 control in 10% CO_2 on appropriate media. In later experiments, matching tenfold titers were plated
411 in permissive conditions (e.g. in 10% CO_2) to estimate the number of viable cells. When colonies
412 formed in restrictive conditions, they were picked into permissive media, grown to saturation,
413 washed and tested for re-growth in restrictive conditions by titer plating or streaking. Plasmid DNA
414 was isolated from verified colonies and transformed into naive CCMB1 cells to test whether
415 plasmid mutations confer improved growth (i.e. in the absence of genomic mutations). This
416 process is diagrammed in Figure S5B.

417 # Phenotyping of matched cultures in 10% CO_2 and ambient air

418 To interrogate the phenotypic effects of mutations to the CCM, we tested the growth of matched
419 biological replicate cultures of CCM mutants (e.g. disruption of carboxysome components or
420 transporter function) in M9 glycerol medium in 10% CO_2 and ambient air (Figure S7). For these
421 experiments, individual colonies were picked into a round-bottom tube with 4 mL of M9 glycerol
422 media with full strength antibiotic and 100 nM aTc. 1 mL of culture was then transferred to a
423 second tube. The 3 mL pre-culture was incubated in 10% CO_2, while the 1 mL culture was
424 incubated in ambient air as a negative control. Control strains unable to grow in minimal media
425 (e.g. those expressing inactive rubisco mutants) were pre-cultured in LB media. High-CO_2 pre-

426 cultures were grown to saturation, after which optical density (OD600) was measured in five-fold
427 dilution.
428
429 Experimental cultures were inoculated with pre-culture to a starting OD600 of 0.01 in 3 mL of M9
430 glycerol media with 12.5 mg/L chloramphenicol and 100 nM aTc. Each pre-culture was used to
431 inoculate a matched pair of experimental cultures, one incubated in 10% CO_2 and another in
432 ambient air, as diagrammed in Figure S7B. After a defined period of growth (4 days in Figure S7A
433 and 12 days in Figure S7C) all culture densities were measured at 600 nm. All experiments were
434 performed in biological quadruplicate, i.e. using four independent pre-cultures deriving from
435 distinct colonies to inoculate four pairs of matched cultures. A positive control was included in all
436 experiments to test the media composition. We used a complemented double carbonic anhydrase
437 knockout (CAfree:pFE-sfGFP+pFA-HCAII) for this purpose as its growth in air depends on the
438 expression of the human carbonic anhydrase II from the pFA-HCAII plasmid (Desmarais et al.,
439 2019).

440 Electron microscopy

441 CCMB1:pCB'+pCCM' was grown in ambient air in 3 ml of M9 glycerol medium and induced with
442 100 nM aTc. A carboxysome-negative control, CAfree:pFE-sfGFP+pFA-HCAII, was grown in the
443 same conditions. Sample preparation and sectioning were performed by the University of
444 California Berkeley Electron Microscope Laboratory. Cell pellets were fixed for 30 min at room
445 temperature in 2.5% glutaraldehyde in 0.1 M cacodylate buffer pH 7.4. Fixed cells were stabilized
446 in 1% very low melting-point agarose and cut into small cubes. Cubed sample was then rinsed
447 three times at room temperature for 10 min in 0.1 M sodium cacodylate buffer, pH 7.4 and then
448 immersed in 1% osmium tetroxide with 1.6% potassium ferricyanide in 0.1 M cacodylate buffer
449 for an hour in the dark on a rocker. Samples were later rinsed three times with a cacodylate buffer
450 and then subjected to an ascending series of acetone for 10 min each (35%, 50%, 75%, 80%,
451 90%, 100%, 100%). Samples were progressively infiltrated with Epon resin (EMS, Hatfield, PA,
452 USA) while rocking and later polymerized at 60 °C for 24 hours. 70 nm thin sections were cut
453 using an Ultracut E (Leica) and collected on 100 mesh formvar coated copper grids. The grids
454 were further stained for 5 min with 2% aqueous uranyl acetate and 4 min with Reynold's lead
455 citrate. The sections were imaged using a Tecnai 12 TEM at 120 KV (FEI) and images were
456 collected using UltraScan 1000 digital micrograph software (Gatan Inc.).

457 Sample preparation for LC-MS analysis

458 Protein-bound amino acids were analyzed in total biomass hydrolysate of 80 mL cultures grown
459 in minimal media with 99% ^{13}C glycerol (Cambridge Isotopes) as the sole organic carbon source.
460 These cultures were grown in 80 mL volumes in a bioreactor pumping ambient air (MC1000, PSI).
461 After harvesting biomass, samples were prepared and analyzed as described in (Antonovsky et
462 al., 2016). Briefly, the OD600 was recorded and 0.6 OD x mL of sample were pelleted by
463 centrifugation for 15 min at 4,000 x g. The pellet was resuspended in 1 mL of 6 N HCl and
464 incubated for 24 hours at 110 °C. The acid was subsequently evaporated under a nitrogen stream
465 using a custom-built gas manifold (Nevins et al., 2005), resulting in a dry hydrolysate. Dry

466 hydrolysates were resuspended in 0.6 mL of MilliQ water, centrifuged for 5 min at 14,000 x g, and
467 supernatant was analyzed by liquid chromatography-mass spectrometry (LC-MS). Hydrolyzed
468 amino acids were separated using ultra performance liquid chromatography (UPLC, Acquity,
469 Waters) on a C-8 column (Zorbax Eclipse XBD, Agilent) at a flow rate of 0.6 mL/min, and eluted
470 off the column using a hydrophobicity gradient. Buffers used were: A) H2O + 0.1% formic acid
471 and B) acetonitrile + 0.1% formic acid with the following gradient: 100% of A (0-3 min), 100% A
472 to 100% B (3-9 min), 100% B (9-13 min), 100% B to 100% A (13-14 min), 100% A (14-20 min).
473 The UPLC was coupled online to a triple quadrupole mass spectrometer (TQS, Waters). Data
474 were acquired using MassLynx v4.1 (Waters). Amino acids and metabolites used for analysis
475 were selected according to the following criteria: amino acids that had peaks at a distinct retention
476 time and m/z values for all isotopologues and also showed correct ^{13}C labeling fractions in control
477 samples that contained protein hydrolyzates of WT cells grown with known ratios of $^{13}C6$-glucose
478 to ^{12}C-glucose.

479 Isotopic analysis composition of biomolecules

480 The total ^{13}C fraction of each metabolite was determined as the weighted average of the fractions
481 of all the isotopologues for that metabolite:
482

483 $$f_{13_C} = \frac{\sum_{i=0}^{N} f_i \times i}{N}$$

484

485 Here N is the number of carbons in the compound (e.g. N = 3 for serine) and f_i is the relative
486 fraction of the i-th isotopologue, i.e containing *i* ^{13}C carbon atoms. Each metabolite's total ^{12}C
487 fraction was calculated as $f_{12_C} = 1 - f_{13_C}$.

488 Estimating the effective intracellular $^{12}CO_2$ fraction

489 *E. coli* cells grown in ^{13}C glycerol will simultaneously respire glycerol, producing intracellular
490 $^{13}CO_2$, and take up extracellular $^{12}CO_2$ and $H^{12}CO_3^-$. The isotopic composition of the intracellular
491 inorganic carbon (C_i) pool will therefore reflect the balance of uptake and respiration. As rubisco
492 carboxylation draws from the intracellular CO_2 pool, we must estimate the isotopic composition of
493 the C_i pool to evaluate the contribution of rubisco to metabolism. We used the carbamoyl-
494 phosphate moiety as a marker for the isotopic distribution of the intracellular C_i pool, as described
495 in (Gleizer et al., 2019). Briefly, carbamoyl-phosphate is generated by phosphorylation of
496 bicarbonate, and is condensed with ornithine in the biosynthesis of L-arginine. We compared the
497 mass isotopologue distribution of L-arginine, which contains one carbon from carbamoyl-
498 phosphate, with the mass isotopologue distribution of L-glutamate as L-glutamate is an ornithine
499 precursor.
500

501 We estimated the effective ^{13}C labeling of intracellular inorganic carbon ($f_{13_{CO_2},effective}$) as follows:

502 $$f_{13_{CO_2},effective} = \sum_{i=0}^{6} f_{arg,i} - \sum_{i=0}^{5} f_{glu,i}$$

Here $f_{13_{CO_2},effective}$ is the relative fraction of $^{13}CO_2$ out of the total CO_2 pool (or, more formally, the C_i), and $f_{arg,i}$ and $f_{glu,i}$ are the fraction of the i-th isotopologue of arginine and glutamate respectively. We assumed fast equilibration of the intracellular C_i pool because the strains used in labeling experiments express a carbonic anhydrase. An equivalent equation can be defined for the arginine-proline comparison (Gleizer et al., 2019), however proline data were of insufficient quality to use and so we report inferences based on the arginine-glutamate comparison. The effective intracellular fraction of $^{12}CO_2$ was calculated as $f_{12_{CO_2},effective} = 1 - f_{13_{CO_2},effective}$. For brevity, we refer to these fractions as $f_{12_{CO_2}}$ and $f_{13_{CO_2}}$, respectively.

Estimating the rubisco carboxylation flux *in vivo*

When CCMB1 cells are grown on glycerol, 3-phosphoglycerate (3PG) can be produced via two routes: (i) rubisco catalyzed carboxylation of RuBP and (ii) glycolytic metabolism of glycerol via dihydroxyacetone phosphate (DHAP). We denote these two fluxes as $J_{rubisco}$ and J_{pgk}, respectively, where pgk (phosphoglycerate kinase) is the glycolytic enzyme producing 3PG. Serine is a direct metabolic product of 3PG and was therefore assumed to have the same ^{12}C composition as 3PG. Rubisco carboxylates by drawing CO_2 from the intracellular inorganic carbon pool, whose ^{12}C composition $f_{12_{CO_2}}$ is inferred as described above.

Based on these assumptions, the ^{12}C composition 3PG (and therefore serine) equals a flux-weighted sum of contributions from rubisco and pgk. As such, the relative 3PG production flux that is due to rubisco, $J_{rubisco}/(J_{rubisco}+J_{pgk})$, can be inferred via the following calculation:

$$f_{ser,ctrl} = f_{3PG,ctrl} = 0 \times \frac{1}{6}(f_{12_{CO_2}} + 5 \times f_{RuBP,exp}) + 1 \times f_{DHAP,ctrl} = f_{DHAP,ctrl}$$

$$f_{ser,exp} = f_{3PG,ctrl} = \frac{J_{rubisco}}{J_{rubisco} + J_{pgk}} \times \frac{1}{6}(f_{12_{CO_2}} + 5 \times f_{RuBP,exp}) + \frac{J_{pgk}}{J_{rubisco} + J_{pgk}} \times f_{DHAP,exp}$$

Where the first equation is written for the control and the second for the experimental cells (complemented CCMB1). Here $f_{ser,ctrl}$ and $f_{ser,exp}$ are the ^{12}C composition of serine in the control and experiment respectively. Identical notation is used for RuBP and DHAP. As there are only two routes of 3PG production, the above equations can be simplified to solve for the relative flux through rubisco *in vivo*:

$$\frac{J_{pgk}}{J_{rubisco} + J_{pgk}} \equiv 1 - \frac{J_{rubisco}}{J_{rubisco} + J_{pgk}}$$

$$\frac{J_{rubisco}}{J_{rubisco} + J_{pgk}} = \frac{f_{ser,exp} - f_{DHAP,exp}}{\frac{1}{6}(f_{12_{CO_2}} + 5 \times f_{RUBP,exp}) - f_{DHAP,exp}}$$

$f_{12_{CO_2}}$ was inferred on a per-sample basis, with the mean values being $25\% \pm 4\%$ and $67\% \pm 28\%$ for the control and experiment respectively (Figure S8C). Because glycerol is converted into 3PG

539 and serine via DHAP in wild-type *E. coli* (Booth, 2005), we expect that $f_{ser,ctrl} = f_{DHAP,ctrl}$, as
540 derived in the first equation above. LC-MS measurements give $f_{ser,ctrl} = 0.6\% \pm 0.2\%$ and
541 $f_{ser,ctrl} = 3.5\% \pm 2.2\%$(Figure S8B). Valine is also a metabolic product of DHAP and was found
542 to have a similar ^{12}C fraction $f_{val,ctrl} = 0.6\% \pm 0.1\%$ in control cells. Since glycerol is immediately
543 converted to DHAP in *E. coli*, we further assumed that $f_{DHAP,ctrl} = f_{DHAP,exp}$.
544
545 RuBP is produced in CCMB1 when rubisco and prk are expressed. Since glycerol is the sole
546 carbon source and there are no carboxylation reactions between DHAP and RuBP in CCMB1, we
547 assumed $f_{RuBP,exp} = f_{DHAP,ctrl}$. This assumption is supported by LC-MS measurements of
548 histidine in control cells. Like RuBP, histidine is synthesized from a pentose-phosphate pathway
549 intermediates (Winkler and Ramos-Montañez, 2009), and measured $f_{his,ctrl} = 0.7\% \pm 0.1\%$,
550 which is very similar to $f_{ser,ctrl} = 0.6\% \pm 0.3\%$. Using mean values to illustrate the calculation
551 gives $\frac{J_{rubisco}}{J_{rubisco}+J_{pgk}} = \frac{3.5\%-0.6\%}{\frac{1}{6}(67\%+5\times0.6\%)-0.6\%} = 0.26$, implying that 26% of 3PG production is due to
552 rubisco.
553
554 10^5 random samples were drawn from the experimentally determined parameter ranges to
555 estimate a 99% confidence interval on the rubisco flux fraction. As the ^{12}C composition of
556 inorganic carbon ($f_{12_{CO_2}}$) and serine are mechanistically linked via rubisco, these values were
557 assumed to co-vary. Distributions were estimated on a per-sample basis by assuming 0.1% error
558 in direct measurement of serine and 1% error in the inference of $f_{12_{CO_2}}$. These calculations gave
559 a median flux estimate of 19% with 99% of values falling between 4.0% and 47.3%. The sample
560 with the lowest inferred rubisco flux had a median estimate of 10.2% with 99% of values falling
561 between 2.5% and 17.9%, implying that rubisco is responsible for a nonzero fraction of 3PG
562 production in all samples. Applying a wider error range of 5% to $f_{12_{CO_2}}$ did not qualitatively change
563 results, giving an overall median flux fraction of 19.1% and a 99% confidence interval 3.9-50.7%.
564 This and above calculations can be found in the following Jupyter notebook:
565 https://github.com/flamholz/carboxecoli/blob/master/notebooks/00_LCMS_calcs.ipynb.

566 Predicting rubisco carboxylation flux via Flux Balance Analysis

567 A stoichiometric model of complemented CCMB1 was generated from the Core *Escherichia coli*
568 Metabolic Model (Orth et al., 2010) by adding rubisco and prk and then deleting the rpi and edd
569 reactions. Parsimonious Flux Balance Analysis (pFBA) was applied to the resulting model to
570 calculate intracellular metabolic metabolic fluxes that maximize the rate of biomass production.
571 As many distinct flux distributions can yield the same (maximal) rate of biomass production, pFBA
572 uses the minimum sum of fluxes objective to define a unique flux solution (Holzhütter, 2004). The
573 COBRApy implementation of pFBA introduces an additional free parameter, the permissible
574 fraction of the maximal biomass production rate f_{opt} (Ebrahim et al., 2013). When $f_{opt} < 1.0$, the
575 biomass production can be less-than-optimal if this would further decrease the sum of fluxes.
576
577 pFBA was run with f_{opt} ranging from 0.8 to 1.0 in increments of 0.01 to account for the fact that
578 CCMB1 has not undergone selection to maximize biomass production with rubisco expressed.

For each resulting flux distribution the fraction of 3PG production flux due to rubisco was calculated as the fraction of 3PG molecules produced via rubisco carboxylation divided by the total flux to 3PG. These calculations predict that 19.5%-21.5% of 3PG production is due to rubisco. The model was rerun after removing all possibility for product secretion by deleting all carbon-containing exchange reactions other than glycerol and CO_2 exchange. This modification should give an upper bound on the fraction of 3PG production due to rubisco, as carbon cannot be shunted away from biomass production to overflow products. The "no overflow" model predicted that 23.9% of 3PG production is due to rubisco independent of f_{opt}. The range of predictions from 19.5-23.9% is plotted Figure S8D. All calculations were done using Python and COBRApy (Ebrahim et al., 2013), and can be found in this Jupyter notebook: https://github.com/flamholz/carboxecoli/blob/master/notebooks/01_FBA_rubisco_flux_prediction.ipynb.

Acknowledgements: We thank Matt Davis for P1 transduction materials and advice, Hernan Garcia and Han Lim for supplying pZ plasmids, Maggie Stoeva, Anna Engelbrektson, Anchal Mehra, Sophia Ewens and Tyler Barnum for help with anaerobic growth, Reena Zalpuri and Danielle Jorgens at the University of California Berkeley Electron Microscope Laboratory for advice and assistance with electron microscopy, and Rob Egbert and Adam Arkin for KEIO strains. We are grateful to Eric Estrin, Woody Fischer, Darcy McRose, Dipti Nayak, Sabeeha Merchant, Luke Oltrogge, and Naiya Phillips, who gave detailed comments on the manuscript, and to Dan Arlow, Yinon Bar-On, Dan Davidi, Jack Desmarais, Hernan Garcia, Oliver Mueller-Cajar, Rob Nichols, Kris Niyogi, Dan Portnoy, Morgan Price, Noam Prywes, Jeremy Roop, Rachel Shipps, Patrick Shih, and Dan Tawfik, for support, advice and helpful discussions throughout.

Funding: This work was supported by a National Science Foundation Graduate Research Fellowship (to A.I.F.), grants from the US Department of Energy (no. DE-SC00016240) and Royal Dutch Shell (Energy Biosciences Institute project CW163755) to D.F.S., and from the European Research Council (Project NOVCARBFIX 646827) to R.M. R.M. is the Charles and Louise Gartner Professional Chair.

Author contributions: A.I.F. conceived of and designed all experiments with mentorship from R.M. and D.F.S. and support from all authors. S.A., N.A., E.N., A.B-E. and R.M. designed and constructed the Δ*rpiAB* strain from which A.I.F., E.J.D, and S.R. constructed CCMB1. A.I.F., E.J.D, and S.R. designed and constructed all other strains and plasmids. A.I.F and E.J.D. performed growth and selection experiments. C.B. performed electron microscopy. S.G. and R.B-N. performed LC-MS analysis on biomass hydrolysate prepared by A.I.F. and E.J.D. A.I.F., S.G., R.B-N., and E.N. analysed isotopic labeling data. A.I.F and E.N. designed and executed Flux Balance Analysis. A.I.F. wrote the manuscript with input from all authors.

Competing interests: D.F.S. is a co-founder of Scribe Therapeutics and a scientific advisory board member of Scribe Therapeutics and Mammoth Biosciences. A.B.-E. is co-founder of b.fab. These companies were not involved in this research in any way. All other authors declare no competing interests.

Data and materials availability: All data associated with this work is available in the main text, supplementary materials and the repository at https://github.com/flamholz/carboxecoli. Plasmids available on Addgene at https://www.addgene.org/David_Savage/, strains distributed on request.

References

Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, Hartl FU, Hayer-Hartl M. 2017. Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science* **358**:1272–1278.

Antonovsky N, Gleizer S, Noor E, Zohar Y, Herz E, Barenholz U, Zelcbuch L, Amram S, Wides A, Tepper N, Davidi D, Bar-On Y, Bareia T, Wernick DG, Shani I, Malitsky S, Jona G, Bar-Even A, Milo R. 2016. Sugar Synthesis from CO2 in Escherichia coli. *Cell* **166**:115–125.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko K a., Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**:2006.0008.

Bar-On YM, Milo R. 2019. The global mass and average rate of rubisco. *Proc Natl Acad Sci U S A* **116**:4738–4743.

Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver P a., Savage DF. 2012. Modularity of a carbon-fixing protein organelle. *Proc Natl Acad Sci U S A* **109**:478–483.

Booth IR. 2005. Glycerol and Methylglyoxal Metabolism. *EcoSal Plus* **1**:1–8.

Bowes G, Ogren WL. 1972. Oxygen inhibition and other properties of soybean ribulose 1,5-diphosphate carboxylase. *J Biol Chem* **247**:2171–2176.

Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S. 2009. The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One* **4**:e7521.

Claassens NJ, Sousa DZ, Dos Santos VAPM, de Vos WM, van der Oost J. 2016. Harnessing the power of microbial autotrophy. *Nat Rev Microbiol* **14**:692–706.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**:6640–6645.

Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* **1151**:165–188.

Desmarais JJ, Flamholz AI, Blikstad C, Dugan EJ, Laughlin TG, Oltrogge LM, Chen AW, Wetmore K, Diamond S, Wang JY, Savage DF. 2019. DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nat Microbiol* **4**:2204–2215.

Du J, Förster B, Rourke L, Howitt SM, Price GD. 2014. Characterisation of Cyanobacterial Bicarbonate Transporters in E. coli Shows that SbtA Homologs Are Functional in This Heterologous Expression System. *PLoS One* **9**:e115905.

Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. 2013. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* **7**:74.

Ermakova M, Danila FR, Furbank RT, von Caemmerer S. 2020. On the road to C4 rice: advances and perspectives. *Plant J* **101**:940–950.

Fischer WW, Hemp J, Johnson JE. 2016. Evolution of Oxygenic Photosynthesis. *Annu Rev Earth Planet Sci* **44**:647–683.

Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, Oltrogge LM, Alves R, Savage D, Milo R. 2019. Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* **58**:3365–3376.

Flamholz A, Shih PM. 2020. Cell biology of photosynthesis over geologic time. *Curr Biol* **30**:R490–R494.

Gleizer S, Ben-Nissan R, Bar-On YM, Antonovsky N, Noor E, Zohar Y, Jona G, Krieger E, Shamshoum M, Bar-Even A, Milo R. 2019. Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell* **179**:1255–1263.e12.

Holzhütter H-G. 2004. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem* **271**:2905–2922.

Kerfeld CA, Melnicki MR. 2016. Assembly, function and evolution of cyanobacterial

683 carboxysomes. *Curr Opin Plant Biol* **31**:66–75.
684 Lewis NE, Nagarajan H, Palsson BO. 2012. Constraining the metabolic genotype–phenotype
685 relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* **10**:291–305.
686 Liang S, Bipatnath M, Xu Y, Chen S, Dennis P, Ehrenberg M, Bremer H. 1999. Activities of
687 constitutive promoters in Escherichia coli. *J Mol Biol* **292**:19–37.
688 Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, Hanson MR. 2014. A faster Rubisco with
689 potential to increase photosynthesis in crops. *Nature* **513**:547–550.
690 Long BM, Hee WY, Sharwood RE, Rae BD, Kaines S, Lim Y-L, Nguyen ND, Massey B, Bala S,
691 von Caemmerer S, Badger MR, Price GD. 2018. Carboxysome encapsulation of the CO2-
692 fixing enzyme Rubisco in tobacco chloroplasts. *Nat Commun* **9**:3570.
693 Long BM, Rae BD, Rolland V, Förster B, Price GD. 2016. Cyanobacterial CO2-concentrating
694 mechanism components: function and prospects for plant metabolic engineering. *Curr Opin*
695 *Plant Biol* **31**:1–8.
696 Lutz R, Bujard H. 1997. Independent and tight regulation of transcriptional units in Escherichia
697 coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*
698 **25**:1203–1210.
699 Mackinder LCM, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman
700 Rosenzweig ES, Pallesen L, Reeves G, Itakura A, Roth R, Sommer F, Geimer S, Mühlhaus
701 T, Schroda M, Goodenough U, Stitt M, Griffiths H, Jonikas MC. 2016. A repeat protein links
702 Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc Natl Acad Sci U S A*
703 **113**:5958–5963.
704 Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF. 2016. pH determines the energetic
705 efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc Natl Acad Sci U S A*
706 **113**:E5354–62.
707 Marcus Y, Schwarz R, Friedberg D, Kaplan A. 1986. High CO2 Requiring Mutant of Anacystis
708 nidulans R2. *Plant Physiol* **82**:610–612.
709 McGrath JM, Long SP. 2014. Can the cyanobacterial carbon-concentrating mechanism increase
710 photosynthesis in crop species? A theoretical analysis. *Plant Physiol* **164**:2247–2261.
711 Merlin C, Masters M. 2003. Why is carbonic anhydrase essential to Escherichia coli? *J Bacteriol*
712 **185**. doi:10.1128/JB.185.21.6415
713 Mueller-Cajar O. 2017. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites.
714 *Front Mol Biosci* **4**:31.
715 Nevins CP, Vierck JL, Bogachus LD, Velotta NS, Castro-Munozledo F, Dodson MV. 2005. An
716 Inexpensive Method for Applying Nitrogen Evaporation to Hexane-containing 24- or 96-well
717 Plates. *Cytotechnology* **49**:71–75.
718 Occhialini A, Lin MT, Andralojc PJ, Hanson MR, Parry MAJ. 2016. Transgenic tobacco plants
719 with improved cyanobacterial Rubisco expression but no extra assembly factors grow at
720 near wild-type rates if provided with elevated CO2. *Plant J* **85**:148–160.
721 Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF. 2020.
722 Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation.
723 *Nat Struct Mol Biol* **27**:281–287.
724 Orr DJ, Worrall D, Lin MT, Carmo-Silva E, Hanson MR, Parry MAJ. 2020. Hybrid
725 Cyanobacterial-Tobacco Rubisco Supports Autotrophic Growth and Procarboxysomal
726 Aggregation. *Plant Physiol* **182**:807–818.
727 Orth JD, Fleming RMT, Palsson BØ. 2010. Reconstruction and Use of Microbial Metabolic
728 Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal Plus*
729 **4**:1–47.
730 Peekhaus N, Conway T. 1998. What's for dinner?: Entner-Doudoroff metabolism in Escherichia
731 coli. *J Bacteriol* **180**:3495.
732 Price GD, Badger MR. 1989a. Isolation and characterization of high CO2-requiring-mutants of
733 the cyanobacterium Synechococcus PCC7942: two phenotypes that accumulate inorganic

734 carbon but are apparently unable to generate CO2 within the carboxysome. *Plant Physiol*
735 **91**:514–525.
736 Price GD, Badger MR. 1989b. Expression of Human Carbonic Anhydrase in the
737 Cyanobacterium Synechococcus PCC7942 Creates a High CO2-Requiring Phenotype
738 Evidence for a Central Role for Carboxysomes in the CO2 Concentrating Mechanism. *Plant*
739 *Physiol* **91**:505–513.
740 Rae BD, Long BM, Badger MR, Price GD. 2013. Functions, compositions, and evolution of the
741 two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in
742 cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* **77**:357–379.
743 Raven JA, Beardall J, Sánchez-Baracaldo P. 2017. The possible evolution and future of CO2-
744 concentrating mechanisms. *J Exp Bot* **68**:3701–3716.
745 Scott KM, Leonard JM, Boden R, Chaput D, Dennison C, Haller E, Harmer TL, Anderson A,
746 Arnold T, Budenstein S, Brown R, Brand J, Byers J, Calarco J, Campbell T, Carter E,
747 Chase M, Cole M, Dwyer D, Grasham J, Hanni C, Hazle A, Johnson C, Johnson R, Kirby B,
748 Lewis K, Neumann B, Nguyen T, Nino Charari J, Morakinyo O, Olsson B, Roundtree S,
749 Skjerve E, Ubaldini A, Whittaker R. 2019. Diversity in CO2-Concentrating Mechanisms
750 among Chemolithoautotrophs from the Genera Hydrogenovibrio, Thiomicrorhabdus, and
751 Thiomicrospira, Ubiquitous in Sulfidic Habitats Worldwide. *Appl Environ Microbiol* **85**:1–19.
752 Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA. 2016. Biochemical
753 characterization of predicted Precambrian RuBisCO. *Nat Commun* **7**:10382.
754 Wang L, Jonikas MC. 2020. The pyrenoid. *Curr Biol* **30**:R456–R458.
755 Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D, Yeates TO. 2014. Structure and
756 identification of a pterin dehydratase-like protein as a ribulose-bisphosphate
757 carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. *J Biol Chem*
758 **289**:7973–7981.
759 Wildman SG. 2002. Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate
760 carboxylase-oxygenase). *Photosynth Res* **73**:243–250.
761 Wilson RH, Martin-Avila E, Conlan C, Whitney SM. 2018. An improved Escherichia coli screen
762 for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics. *J*
763 *Biol Chem* **293**:18–27.
764 Winkler ME, Ramos-Montañez S. 2009. Biosynthesis of Histidine. *EcoSal Plus* **3**:1–33.
765 Wu A, Hammer GL, Doherty A, von Caemmerer S, Farquhar GD. 2019. Quantifying impacts of
766 enhancing photosynthesis on crop yield. *Nat Plants* **5**:380–388.

767