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. 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, laying the groundwork for deeper study and engineering of mechanisms enhancing CO_2 assimilation.

One Sentence Summary

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

Main Text

Nearly all carbon in the biosphere enters by CO_2 fixation in the Calvin-Benson-Bassham cycle (*1*). Ribulose Bisphosphate Carboxylase/Oxygenase - commonly known as rubisco - is the CO_2 fixing enzyme in this cycle (*2*) and likely the most abundant enzyme on Earth (*3*). 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} (*4*). Moreover, all known rubiscos catalyze a competing oxygenation of the five-carbon organic substrate, ribulose 1,5-bisphosphate (*4*, *5*).

Rubisco arose > 2.5 billion years ago, when Earth's atmosphere contained little O_2 and abundant CO_2 (*6*, *7*). 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 (*1*, *8*) and overcome this problem by concentrating CO_2 near rubisco (Fig. 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 (*5*) thereby improving overall carbon assimilation (Figure 1B). Today, at least four varieties of CCMs are found in plants, algae and bacteria (*1*, *8*), organisms with CCMs are collectively responsible for ≈50% of global net photosynthesis (*1*), 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 (*9*, *10*). 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 (*11*, *12*). 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 (Fig. 1A).

As CCMs accelerate CO_2 fixation, there is great interest in transplanting them into crops (*11*, *13*). Carboxysome-based CCMs are especially attractive because they natively function in single cells and appear to rely on a tractable number of genes (*14*–*17*). Modeling suggests that introducing bacterial CCM components could improve plant photosynthesis (*11*), especially if aspects of plant physiology can be modulated via genetic engineering (*18*). However, expressing bacterial rubiscos and carboxysome components has, so far, uniformly resulted in transgenic plants displaying impaired growth (*14*–*17*). 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 (*19*–*22*), 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*.

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 (*22*). These genes include rubisco large and small subunits, the carboxysomal carbonic anhydrase, seven structural proteins of the α-carboxysome (*23*), an energy-coupled inorganic

carbon transporter (*22*, *24*), three rubisco chaperones (*25*–*27*), and four genes of unknown function (Fig. 1C). We aimed to test whether these genes are sufficient to establish a functioning CCM in *E. coli*.

For this purpose, we 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), however, makes Ru5P a metabolic "dead-end" (Fig. 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), thereby enabling Ru5P metabolism and growth (Fig. 2A). Additionally, cytosolic carbonic anhydrase activity is incompatible with the bacterial CCM (*28*). We therefore constructed a strain, named CCMB1 for "**CCM B**ackground **1**", lacking *rpiAB* and all endogenous carbonic anhydrases (Methods).

As predicted, CCMB1 required rubisco and prk for growth on glycerol minimal media in 10% CO_2 (Fig. 2B-C). Even when expressing rubisco and *prk* on the p1A plasmid (Fig. 2B), however, CCMB1 failed to grow on glycerol in ambient air (Figs. 2D and S3-4) presumably due to insufficient carboxylation at low CO_2. That is, CCMB1:p1A displays the "high-CO_2 requiring" phenotype that is the hallmark of CCM mutants (*19*, *20*).

We expected that a functional CO_2-concentrating mechanism would cure CCMB1 of its high-CO_2 requirement and permit growth in ambient air. We therefore generated two plasmids, pCB and pCCM, that together express all 20 genes from the *H. neapolitanus* CCM cluster (Figs. 1C and S5A). pCB encodes ten carboxysome genes (*23*), including rubisco large and small subunits,

along with *prk*. The remaining *H. neapolitanus* genes, including putative rubisco chaperones (*25*–*27*) and an inorganic carbon transporter (*22*, *24*), were cloned into the second plasmid, pCCM.

CCMB1 co-transformed with pCB and pCCM initially failed to grow on glycerol media in ambient air. We therefore conducted selection experiments, described fully in the supplement, that resulted in the isolation of mutant plasmids conferring growth in ambient air. Briefly, CCMB1:pCB + pCCM cultures were grown to saturation in 10% CO_2 and multiple dilutions were plated on glycerol minimal media (Methods). Colonies became visible after 20 days of incubation in ambient air (Fig. S5). Deep-sequencing of plasmid DNA revealed mutations in regulatory sequences (e.g. a promoter and transcriptional repressor) but none in sequences coding for CCM components (Table S4). Individual post-selection plasmids pCB' and pCCM' were reconstructed by PCR, resequenced, and transformed into naive CCMB1 (Methods). As shown in Figs. 3A and S6, 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.

To verify that growth in ambient air depends on the CCM, we generated plasmids carrying targeted mutations (Fig. S7). 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. 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 (*21*), while truncating CsoS2 prohibits carboxysome formation entirely (*29*). Finally, an inactivating mutation to the inorganic carbon transporter also eliminated growth in ambient air (*22*).

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 (Figs. 3B), implying production of morphologically-normal carboxysomes.

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). As the rubisco product 3PG is a precursor for serine biosynthesis, we expected excess ^{12}C on serine when rubisco is active (Fig. 3C). Serine from CCMB1:pCB'+pCCM' cells contained roughly threefold more ^{12}C than a rubisco-independent control (Figs. 3D). We estimated the contribution of rubisco to 3PG synthesis *in vivo* by comparing labeling patterns between experiment and control (Methods). Rubisco carboxylation was responsible for at least 10% of 3PG synthesis in all four biological replicates (Fig. 3E, Methods), confirming fixation of CO_2 from ambient air. As such, this represents the first functional reconstitution of any CCM.

Reconstitution enabled us to investigate which *H. neapolitanus* genes are necessary for CCM function. 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 (*25*–*27*). Although *acRAF* deletion had a large negative

159 effect in *H. neapolitanus* (*22*), neither acRAF nor CbbQ were required for CCMB1 to grow in
160 ambient air. Consistent with a screen in the native host (*22*), however, *acRAF* deletion produced
161 a substantial growth defect (Fig. S7C). Further such experiments can test hypotheses about
162 carboxysome biogenesis (*23*, *29*), probe the relationship between CCMs and host physiology (*12*,
163 *28*), and define a minimal reconstitution of the bacterial CCM suitable for plant expression (*15*–
164 *17*, *30*).
165
166 Today, CCMs catalyze about half of global photosynthesis (*1*), but this was not always so. Land
167 plant CCMs, for example, arose only in the last 100 million years (*1*, *8*). Though all contemporary
168 Cyanobacteria have CCM genes, these CCMs are found in two convergently-evolved varieties
169 (*8*–*10*), suggesting that the ancestor of present-day Cyanobacteria and chloroplasts did not have
170 a CCM (*9*). So how did carboxysome CCMs come to dominate the cyanobacterial phylum?
171
172 Here we demonstrated that the α-carboxysome CCM from *H. neapolitanus* is readily transferred
173 between species and confers a large growth benefit, which can explain how these CCMs became
174 so widespread among bacteria (*9*, *10*). We constructed a CCM by expressing 20 genes in an
175 engineered *E. coli* strain, CCMB1. In accordance with its role in native autotrophic hosts (*16*, *19*,
176 *20*, *22*), the transplanted CCM required α-carboxysomes and inorganic carbon uptake to enable
177 CCMB1 to grow by fixing CO_2 from ambient air (Fig. 3 and S6-8). It appears, therefore, that the
178 α-carboxysome CCM is genetically compact and "portable." As such, it is possible that expressing
179 bacterial CCMs in non-native autotrophic hosts will improve CO_2 assimilation and growth. Our
180 approach to studying CCMs by reconstitution in tractable non-native hosts can also be applied to
181 study other CCMs, including β-carboxysome CCMs, the algal pyrenoid, and plausible evolutionary
182 ancestors (*8*). We hope such studies will further our principled understanding of, and capacity to
183 engineer, the cell biology supporting CO_2 fixation in diverse organisms.
184

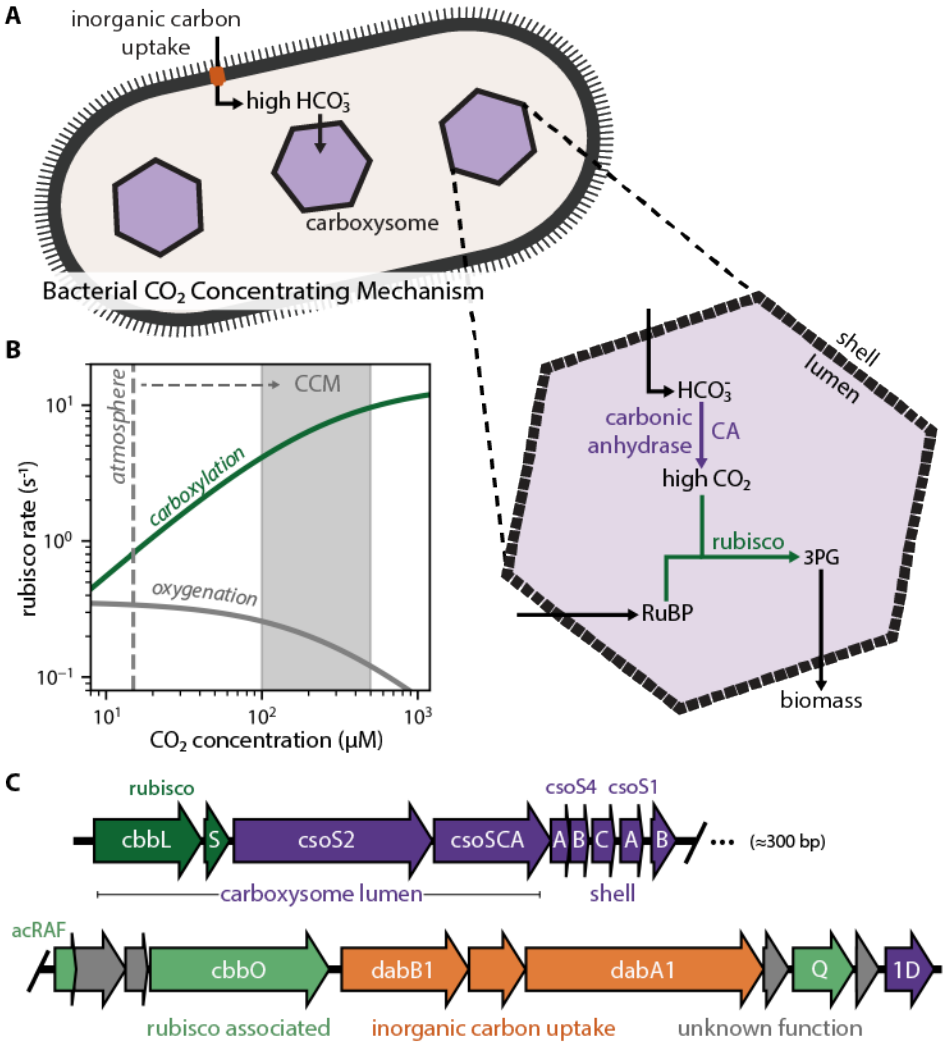

Fig. 1. Twenty genes form the basis of a bacterial CCM. *(**A**) The bacterial CCM consists of at least two essential components - energy-coupled carbon uptake and carboxysome structures that encapsulate rubisco with a carbonic anhydrase (CA) enzyme (11, 12). Transport generates a large cytosolic HCO_3^- 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 (22) expressing rubisco and its associated chaperones (green), carboxysome structural proteins (purple), and an inorganic carbon transporter (orange).*

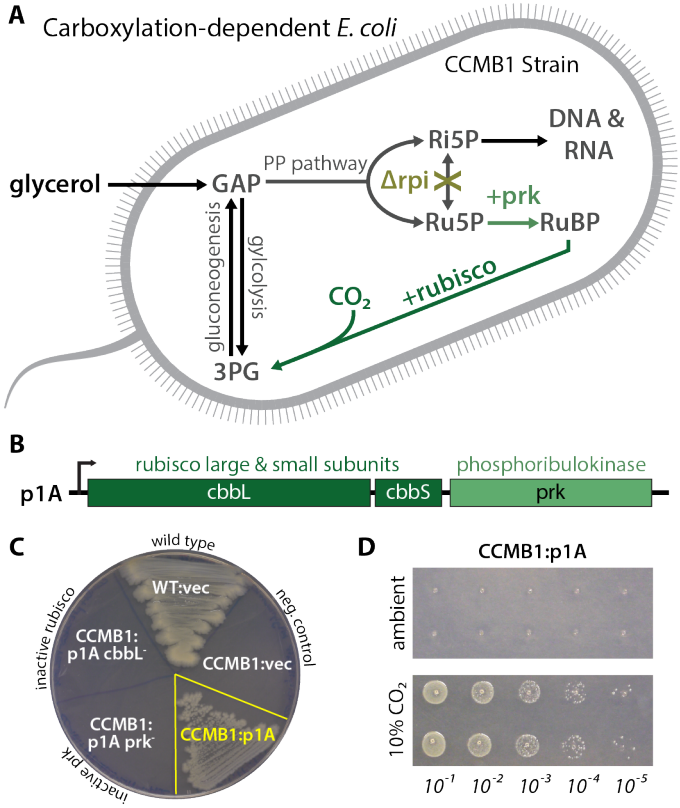

Fig. 2. CCMB1 depends on rubisco carboxylation for growth on glycerol. *(**A**) Ribose-5-phosphate (Ri5P) is required for nucleotide biosynthesis. Deletion of ribose-phosphate isomerase (Δrpi) in CCMB1 blocks ribulose-5-phosphate (Ru5P) metabolism in the pentose phosphate (PP) pathway. Expression of 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).*

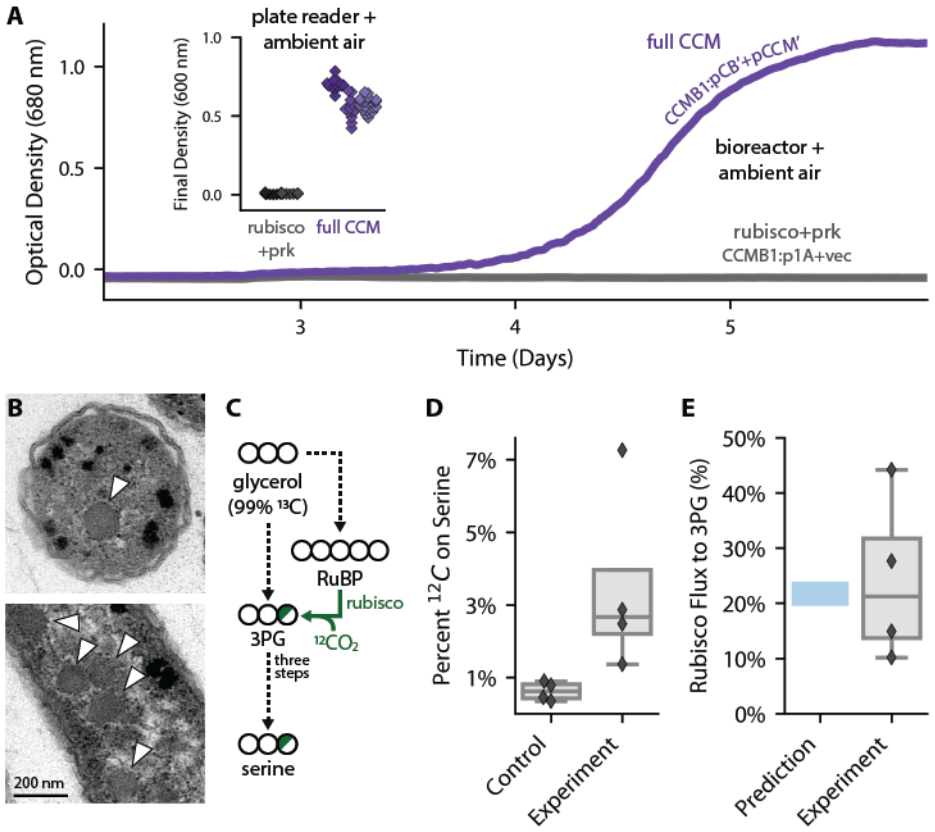

Fig. 3. Expression of 20 CCM genes permits growth of CCMB1 in ambient air. *(**A**) Representative growth curves from a bioreactor bubbling ambient air. CCMB1:pCB' + pCCM' grows well (purple, "full CCM"), while rubisco and prk alone are insufficient for growth in air (grey, CCMB1:p1A+vec). Inset: a plate reader experiment in biological triplicate (different shades) showed the same result (SI). (**B**) Polyhedral bodies resembling carboxysomes are evident in electron micrographs of CCMB1:pCB'+pCCM' cells grown in air. (**C**) 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. (**D**) The ^{12}C composition of serine from CCMB1:pCB' + pCCM' ("Experiment") is roughly threefold above the control. (**E**) 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.*

References

1. J. A. Raven, J. Beardall, P. Sánchez-Baracaldo, *J. Exp. Bot.* **68**, 3701–3716 (2017).
2. S. G. Wildman, *Photosynth. Res.* **73**, 243–250 (2002).
3. Y. M. Bar-On, R. Milo, *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4738–4743 (2019).
4. A. I. Flamholz *et al.*, *Biochemistry*. **58**, 3365–3376 (2019).
5. G. Bowes, W. L. Ogren, *J. Biol. Chem.* **247**, 2171–2176 (1972).
6. P. M. Shih *et al.*, *Nat. Commun.* **7**, 10382 (2016).
7. W. W. Fischer, J. Hemp, J. E. Johnson, *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016).
8. A. Flamholz, P. M. Shih, *Curr. Biol.* **30**, R490–R494 (2020).
9. B. D. Rae, B. M. Long, M. R. Badger, G. D. Price, *Microbiol. Mol. Biol. Rev.* **77**, 357–379 (2013).
10. C. A. Kerfeld, M. R. Melnicki, *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
11. J. M. McGrath, S. P. Long, *Plant Physiol.* **164**, 2247–2261 (2014).
12. N. M. Mangan, A. Flamholz, R. D. Hood, R. Milo, D. F. Savage, *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354–62 (2016).
13. M. Ermakova, F. R. Danila, R. T. Furbank, S. von Caemmerer, *Plant J.* **101**, 940–950 (2020).
14. M. T. Lin, A. Occhialini, P. J. Andralojc, M. A. J. Parry, M. R. Hanson, *Nature*. **513**, 547–550 (2014).
15. A. Occhialini, M. T. Lin, P. J. Andralojc, M. R. Hanson, M. A. J. Parry, *Plant J.* **85**, 148–160 (2016).
16. B. M. Long *et al.*, *Nat. Commun.* **9**, 3570 (2018).
17. D. J. Orr *et al.*, *Plant Physiol.* **182**, 807–818 (2020).
18. A. Wu, G. L. Hammer, A. Doherty, S. von Caemmerer, G. D. Farquhar, *Nat Plants*. **5**, 380–388 (2019).
19. Y. Marcus, R. Schwarz, D. Friedberg, A. Kaplan, *Plant Physiol.* **82**, 610–612 (1986).
20. G. D. Price, M. R. Badger, *Plant Physiol.* **91**, 514–525 (1989).
21. F. Cai *et al.*, *PLoS One*. **4**, e7521 (2009).
22. J. J. Desmarais *et al.*, *Nat Microbiol*. **4**, 2204–2215 (2019).
23. W. Bonacci *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 478–483 (2012).

24. K. M. Scott *et al.*, *Appl. Environ. Microbiol.* **85**, 1–19 (2019).

25. N. M. Wheatley, C. D. Sundberg, S. D. Gidaniyan, D. Cascio, T. O. Yeates, *J. Biol. Chem.* **289**, 7973–7981 (2014).

26. O. Mueller-Cajar, *Front Mol Biosci*. **4**, 31 (2017).

27. H. Aigner *et al.*, *Science*. **358**, 1272–1278 (2017).

28. G. D. Price, M. R. Badger, *Plant Physiol.* **91**, 505–513 (1989).

29. L. M. Oltrogge *et al.*, *Nat. Struct. Mol. Biol.* **27**, 281–287 (2020).

30. J. Du, B. Förster, L. Rourke, S. M. Howitt, G. D. Price, *PLoS One*. **9**, e115905 (2014).

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.