Main manuscript for

Trajectories for the evolution of bacterial CO_2-concentrating mechanisms

Avi I. Flamholz[1,2,3,*,†], Eli Dugan[1,*], Justin Panich[4], John J. Desmarais[1], Luke M. Oltrogge[1], Woodward W. Fischer[3,5], Steven W. Singer[4], David F. Savage[1,6,†]

[1] Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, United States
[2] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125
[3] Resnick Sustainability Institute, California Institute of Technology, Pasadena, CA 91125, USA
[4] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[5] Division of Geological & Planetary Sciences, California Institute of Technology, Pasadena, CA 91125
[6] Howard Hughes Medical Institute, University of California, Berkeley, California 94720

* Equal contribution

[†] Corresponding Authors: Avi I. Flamholz, David F. Savage
Emails: aflamhol@caltech.edu, savage@berkeley.edu

Author contributions:
Avi I. Flamholz: Conceived and designed the study with input from E.J.D. and D.F.S. Conducted preliminary *E. coli* experiments, mathematical modeling, analysis of experimental results, generated figures and wrote the paper with input from all authors.
Eli J. Dugan: Contributed to study design and paper writing. Performed *H. neapolitanus* fitness assays with J.J.D., generated *C. necator* mutants and plasmids for heterologous expression in *E. coli* and *C. necator*, performed *E. coli* growth assays.
Justin Panich: Performed *C. necator* growth experiments.
John J. Desmarais: Performed *H. neapolitanus* fitness assays with E.J.D. Assisted in analysis.
Luke M. Oltrogge: Collaborated with A.I.F. in modeling autotrophic metabolism.
Woodward Fischer: Provided guidance on geological context for this study.
Steven W. Singer: Provided expertise, lab space and materials for *C. necator* cultivation.
David F. Savage: Provided guidance in research design and analysis of experimental results, as well as lab space and materials for most experiments.

Competing interests: D.F.S. is a co-founder of Scribe Therapeutics and a scientific advisory board member of Scribe Therapeutics and Mammoth Biosciences. These companies were not involved in this work in any way.

Classification:

Keywords: physiology, evolution, photosynthesis, Earth history, synthetic biology.

This PDF file includes:

Main Text
Figures 1 to 7

Abstract

Cyanobacteria rely on CO_2 concentrating mechanisms (CCMs) that depend on ≈15 genes to produce two protein complexes: an inorganic carbon (Ci) transporter and a 100+ nm carboxysome compartment that encapsulates rubisco with a carbonic anhydrase (CA) enzyme. Mutations disrupting CCM components prohibit growth in today's atmosphere (0.04% CO_2), indicating that CCMs evolved to cope with declining environmental CO_2. Indeed, geochemical data and models indicate that atmospheric CO_2 has been generally decreasing from high concentrations over the last ≈3.5 billion years. We used a synthetic reconstitution of a bacterial CCM in *E. coli* to study the co-evolution of CCMs with atmospheric CO_2. We constructed strains expressing putative ancestors of modern CCMs — strains lacking one or more CCM components — and evaluated their growth in a variety of CO_2 concentrations. Partial forms expressing CA or Ci uptake genes grew better than controls in intermediate CO_2 levels (≈1%); we observed similar phenotypes in genetic studies of two autotrophic bacteria, *H. neapolitanus* and *C. necator*. To explain how partial CCMs improve growth, we advance a model of co-limitation of autotrophic growth by CO_2 and HCO_3^-, as both are required to produce biomass. Our model and results delineated a viable trajectory for bacterial CCM evolution where decreasing atmospheric CO_2 induces an HCO_3^- deficiency that is alleviated by acquisition of CAs or Ci uptake genes, thereby enabling the emergence of a modern CCM. This work underscores the importance of considering physiology and environmental context when studying the evolution of biological complexity.

Significance

The greenhouse gas content of the ancient atmosphere is estimated using models and measurements of geochemical proxies. Some inferred high ancient CO_2 levels using models of biological CO_2 fixation to interpret the C isotopes found in preserved organic matter. Others argued that elevated CO_2 could reconcile a faint young Sun with evidence for liquid water on Earth. We took a complementary "synthetic biological" approach to understanding the composition of the ancient atmosphere by studying present-day bacteria engineered to resemble ancient autotrophs. By showing that it is simpler to rationalize the emergence of modern bacterial autotrophs if CO_2 was once high, these investigations provided independent evidence for the view that CO_2 concentrations were significantly elevated in the atmosphere of early Earth.

Introduction

Nearly all carbon enters the biosphere through CO_2 fixation in the Calvin-Benson-Bassham (CBB) cycle. Rubisco is the carboxylating enzyme of that pathway and the most abundant enzyme on Earth (1, 2). Rubisco is often considered inefficient due to relatively slow carboxylation kinetics (3–5) and non-specific oxygenation of its five-carbon substrate ribulose 1,5-bisphosphate, or RuBP (6, 7). However, rubisco arose more than 2.5 billion years ago, when the Earth's atmosphere contained virtually no O_2 and, many argue, far more CO_2 than today (8, 9). Over geologic timescales, photosynthetic O_2 production (8) and CO_2-consuming silicate weathering reactions (10) are thought to have caused a gradual increase in atmospheric levels of O_2 (≈20% of 1 bar atmosphere today) and depletion of atmospheric CO_2 to present-day levels of a few hundred parts per million (≈280 ppm pre industrial, ≈420 ppm or ≈0.04% today). The amount of atmospheric CO_2 during the Archean Eon (4-2.5 Ga) is challenging to estimate accurately from the geological record, but is thought to have been substantially higher than today, perhaps as high as 0.1-1 bar (9, 11, 12). It is likely, therefore, that contemporary autotrophs grow on drastically lower levels of CO_2 than their ancestors did.

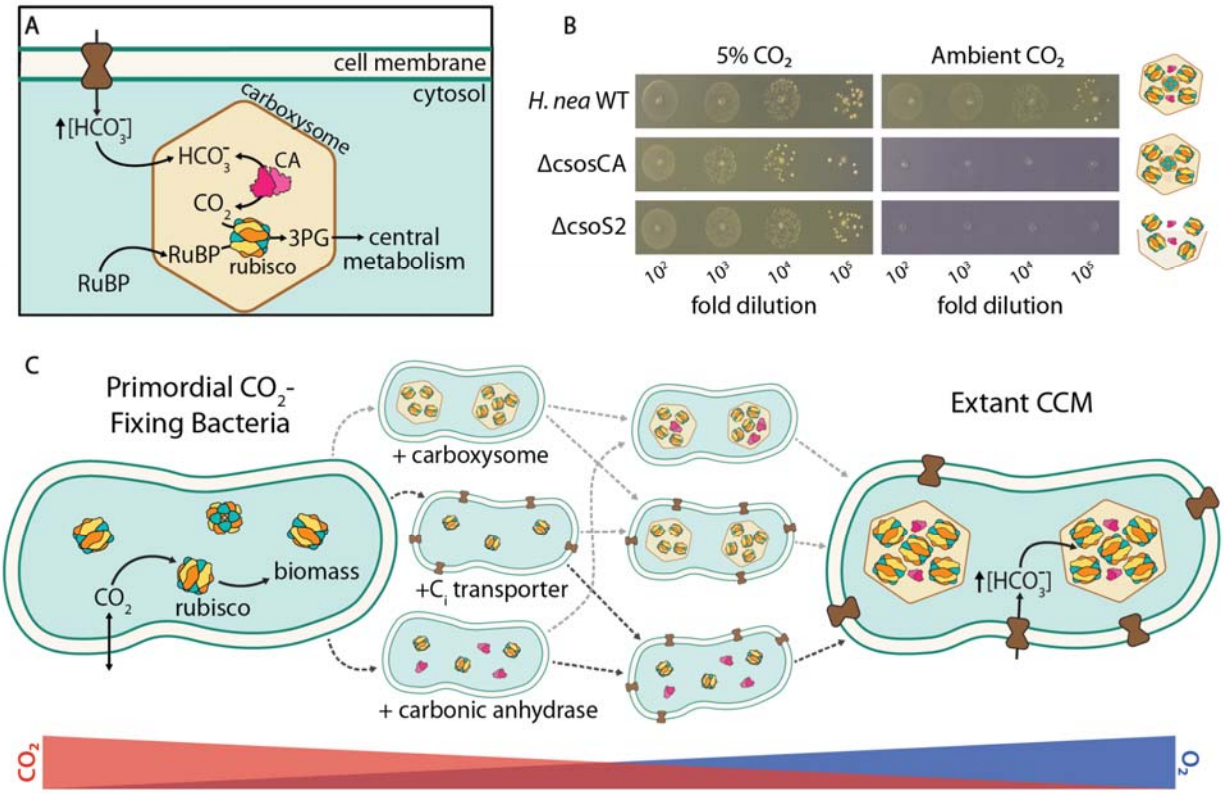

Fig. 1: Mechanism and potential routes for the evolution of the bacterial CO_2 concentrating mechanism. (A) Today the bacterial CCM functions through the concerted action of two protein complexes - an inorganic carbon (Ci) transporter at the cell membrane and carboxysome encapsulation of rubisco with carbonic anhydrase (CA). Ci uptake leads to a high intracellular HCO_3^- concentration, well above equilibrium with the external environment. Elevated HCO_3^- is converted to a high carboxysomal CO_2 concentration by CA activity located only there, which promotes carboxylation by rubisco. (B) Mutants lacking genes coding for essential CCM components grow in elevated CO_2 but fail to grow in ambient air, exemplified here by the chemoautotroph *H. neapolitanus*. Strains lacking the carboxysomal CA (Δ*csosCA*) or an unstructured protein required for carboxysome formation (Δ*csos2*) failed to grow in ambient air, but grew robustly in 5% CO_2 (> 10^8 colony forming units/ml, Figure S1). See Table S4 for description of mutant strains. (C) We consider the CCM to be composed of three functionalities beyond rubisco itself: a CA enzyme (magenta), an Ci transporter (dark brown) and carboxysome encapsulation of rubisco (light brown). Since atmospheric O_2 levels were extremely low (≈1 ppm, blue triangle) and CO_2 was likely quite high (perhaps 0.1-1 bar, red triangle) during the Archean epoch, primordial CO_2-fixing bacteria would not have needed a CCM. We sought to discriminate experimentally between the six sequential trajectories (dashed arrows) in which CCM components could have been acquired.

Many photosynthetic organisms evolved CO_2 Concentrating Mechanisms (CCMs), which help meet the challenge of fixing carbon in a low CO_2 atmosphere. CCMs concentrate CO_2 near rubisco and are found in all Cyanobacteria, some Proteobacteria, as well as many eukaryotic algae and diverse plants (13). Because CO_2 and O_2 addition occur at the same active site in rubisco (6), elevated CO_2 has the dual effects of accelerating carboxylation and suppressing oxygenation of RuBP by competitive inhibition (14–16). As shown in Figure 1A, bacterial CCMs are encoded by ≈15 genes comprising three primary

features: (i) an energy-coupled inorganic carbon (Ci) transporter at the cell membrane and (ii) a cytosolic 100+ nm protein compartment called the carboxysome found that (iii) co-encapsulates rubisco with a carbonic anhydrase (CA) enzyme (4, 17). Energized Ci transport produces a high HCO_3^- concentration in the cytosol (≈30 mM, Fig. 1A), which is converted into a high carboxysomal CO_2 concentration by carbonic anhydrase activity, localized exclusively to the carboxysome (16, 18, 19).

CCM genes are straightforward to identify as mutations disrupting essential CCM components prohibit growth in ambient air (Fig. 1B) and mutants are typically grown in 1% CO_2 or more (14, 15, 20–22). At first glance, therefore, the CCM appears to be "irreducibly complex" as all plausible recent ancestors - e.g. strains lacking individual CCM genes - are not viable in the present-day atmosphere. Irreducible complexity is incompatible with evolution by natural selection, so we and others supposed that bacterial CCMs evolved over a protracted interval of Earth history when atmospheric CO_2 concentrations were much greater than today (13, 23–26). We therefore hypothesized that ancestral forms of the bacterial CCM (i.e. those lacking some genes and complexes required today) would have improved organismal fitness in the elevated CO_2 environments that prevailed when they arose.

To test the hypothesis, we reconstructed present-day analogues of plausible CCM ancestors (henceforth "analogues of ancestral CCMs") and tested their growth across a range of CO_2 partial pressures. Our goal was to identify a stepwise pathway of gene acquisition supporting the evolutionary emergence of a bacterial CCM by improving growth in ever-decreasing CO_2 concentrations (Fig. 1C). We focused on trajectories involving sequential acquisition of genetic components because carbonic anhydrases (27), Ci transporters (22), and homologs of carboxysome shell genes (28, 29) are widespread among bacteria and could therefore be acquired horizontally.

One approach to constructing contemporary analogues of CCM ancestors is to remove CCM genes from a native host. If CCM components were acquired sequentially, some single gene knockouts would be analogous to recent ancestors, e.g. those lacking a complete carboxysome shell (30). We tested this approach by assaying a whole-genome knockout library of a □-proteobacterial chemoautotroph, *H. neapolitanus,* in five CO_2 partial pressures (22, 31). As shown in Figure 2 and elaborated below, we found that many CCM genes contribute substantially to organismal fitness even at CO_2 concentrations tenfold greater than present-day atmosphere (0.5% CO_2), supporting the view that CCM components play an important physiological role even in relatively high environmental CO_2 concentrations.

Removing single CCM genes from a native host can only produce analogues of recent ancestors, however. We recently constructed a functional CCM in an engineered *E. coli* strain called CCMB1 (4). This strain depends on rubisco carboxylation for growth and the expression of a full complement of CCM genes enabled growth in ambient air. Here we used CCMB1 to construct analogues of ancestral CCMs, including several lacking two or more essential components of modern CCMs. We assayed the growth of these putative ancestors across a range of CO_2 pressures to determine whether any ancestral forms contribute to organismal fitness — i.e. improve growth relative to a control strain expressing only rubisco — in various CO_2 pressures. In the following sections we describe the growth of these strains and discuss how it can inform our entwined understandings of bacterial physiology, CCM evolution and the CO_2 content of Earth's ancient atmosphere.

Results and Discussion

CCM genes contribute to fitness even in elevated CO_2

Using a barcoded genome-wide transposon mutagenesis screen, we previously demonstrated that a 20-gene cluster in *H. neapolitanus* contains all the genes necessary for a functional CCM (4, 22). The original screen measured gene fitness across the entire genome via batch competition assays, comparing the abundance of strains in high CO_2 (5%) and ambient air (0.04%) via high-throughput sequencing (22). This effort demonstrated that at least 12/20 genes are necessary components of the CCM as disruption of any one produced substantial and reproducible growth defects in ambient air. To mimic the changes in atmospheric CO_2 that likely occurred over Earth history, we quantified the same library in three additional CO_2 pressures to cover five CO_2 levels: ambient (≈0.04% CO_2), low (0.5%), moderate (1.5%), high (5%), and very high (10%). Replicate experiments were strongly correlated ($R > 0.85$, Figure S2), implying a high degree of reproducibility. We therefore proceeded to ask whether *H. neapolitanus* CCM genes contribute to fitness in elevated CO_2 conditions.

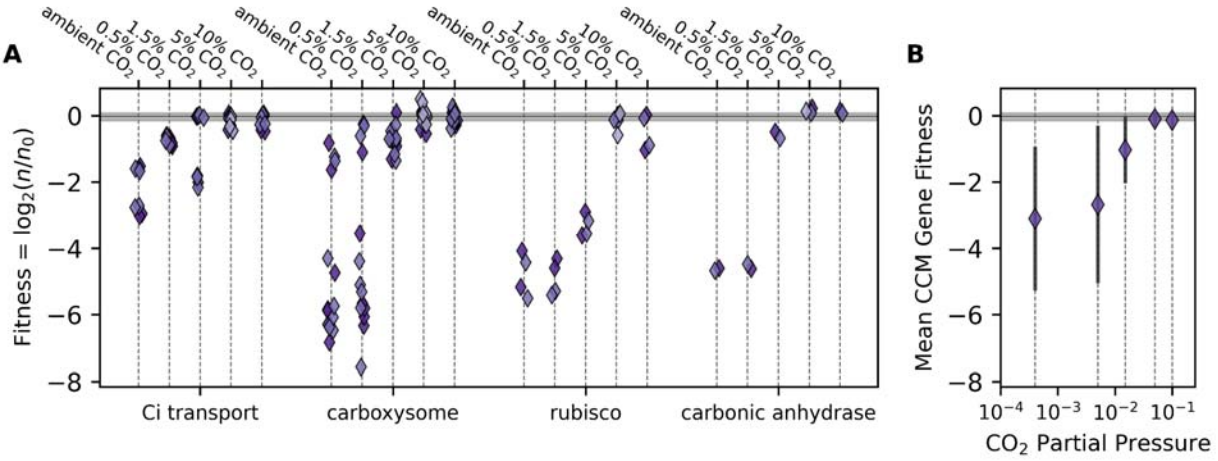

Fig. 2: The contribution of *H. neapolitanus* CCM genes to organismal fitness depends on the environmental CO_2 concentration. *H. neapolitanus* is a chemoautotroph that natively utilizes a CCM in low CO_2 environments. (A) Using a barcoded transposon library constructed in this organism (22), we profiled the contributions of CCM genes

to organismal fitness during autotrophic growth across a range of CO_2 levels (*Methods*). Fitness is estimated as the log_2 ratio of barcode counts between the endpoint sample and the pre-culture for each mutant (22, 32). Each point gives the average fitness of multiple mutants to a single CCM gene in a given CO_2 condition. The library contained 30 mutants disrupting the carboxysomal CA gene (*csosCA*) and an average of ≈40 mutants per CCM gene. A fitness of -2 implies that gene disruption was, on average, associated with a fourfold decrease in strain abundance after a defined period of growth. Biological replicates are indicated by shading, and the gray bar gives the interquartile range of fitness effects for all ≈1700 mutants across all CO_2 levels (-0.15-0.065). "Ci transport" includes 4 DAB genes in two operons (22), "carboxysome" includes 6 nonenzymatic carboxysome genes, "rubisco" denotes the two subunits of the carboxysomal rubisco, and "CA" the carboxysomal CA gene. *H. neapolitanus* also expresses a secondary rubisco, which explains why disruption of the carboxysomal rubisco is not lethal in high CO_2 (Figure S3). (B) On average, CCM genes contribute less to organismal fitness as environmental CO_2 levels rise. Error bars give the standard deviation of fitness values across all CCM genes. See Figures S2-3 for analysis of reproducibility and Tables S1-3 for detailed description of genes.

Figure 2 plots the effect of disrupting CCM genes across five CO_2 pressures, with genes grouped by their known roles in the CCM. The library contained an average of ≈35 mutants per gene in the *H. neapolitanus* genome (22), so each point in Figure 2A represents the average fitness of 5-50 individual mutants. Surprisingly, we found that many CCM genes also contributed substantially to organismal fitness in 0.5% and 1.5% CO_2, indicated by large fitness defects (negative values in Figure 2A), resulting in a negative average fitness of CCM mutants when the CO_2 pressure was 1.5% or less (Figure 2B). Carboxysome genes, for example, were critical for growth in 0.5% CO_2, while certain Ci transport genes contributed substantially to growth in 0.5% and 1.5% CO_2.

Though several *H. neapolitanus* CCM genes contributed to fitness in intermediate CO_2 levels, the CCM appeared to be dispensable in high CO_2 (5-10%, Fig. 2B). As such, the data presented in Figure 2 indicated that individual CCM components like the carboxysome, CA, or Ci transporter may improve autotrophic growth in intermediate CO_2 levels (≈1%) even in the absence of a complete and functional CCM. We considered testing this hypothesis in *H. neapolitanus* directly by constructing "ancestral-like" CCMs lacking carboxysome shell genes or Ci transporters. However, genetic manipulation of *H. neapolitanus* is cumbersome, and the native host's regulatory network could complicate interpretation — a concern highlighted by three DNA-binding proteins that emerged in our screen as likely regulators

of the CCM (Figure S3 and Table S1). We therefore decided to construct and test analogues of CCM ancestors in a non-native host, namely *E. coli*.

Evaluating putative ancestral CCMs in a rubisco-dependent *E. coli*

We recently developed an *E. coli* strain, CCMB1, that depends on rubisco carboxylation for growth in minimal medium. This strain requires elevated CO_2 for rubisco-dependent growth, but expressing the *H. neapolitanus* CCM from two plasmids enabled growth in ambient air (4). One of these plasmids, pCB', encodes the carboxysome genes along with the encapsulated rubisco and carbonic anhydrase enzymes (4). The other plasmid, pCCM', encodes the DAB1 inorganic carbon transporter (4, 22), two rubisco chaperones (22, 33, 34) and a carboxysome positioning system (35). We used this two-plasmid system to express analogues of putative CCM ancestors in CCMB1 and assayed these strains' growth in ambient air, 0.5%, 1.5% and 5% CO_2. We compared the growth of strains expressing partial CCMs to that of a reference CCMB1 strain expressing the *H. neapolitanus* rubisco on the p1A vector but no other CCM components (4). The reference strain grew robustly in 5% CO_2 but failed to grow in ambient air (Fig. 3A, 'Rubisco Alone').

Consistent with our previous work (4), expression of the full complement of CCM genes permitted growth in all CO_2 concentrations tested (Fig. 3A, 'Full CCM'). Recent modeling efforts endeavored to support the hypothesis that encapsulation of rubisco in a semi-permeable barrier could improve CO_2 fixation by generating an acidic local pH (26). This proposal does not require Ci transport or CA activity and, if correct, would therefore support an "encapsulation first" model of CCM evolution. To evaluate the effect of encapsulating rubisco alone in a protein compartment, we replaced the pCCM' plasmid with a vector control so that no Ci transporter was expressed and further deactivated the carboxysomal CA by mutating a crucial cysteine residue (pCB' CsosCA C173S, 'Encap. Rub.' in Fig. 3A). Encapsulating rubisco on its own did not improve growth over the reference strain in any CO_2 concentration tested.

However, when we left the CA active site intact (`En. Rub.+CA'), growth improved substantially in intermediate CO_2 levels (0.5% and 1.5%). These data indicated that CA plays a pivotal role in autotrophic growth in low CO_2, but did not support the hypothesis that encapsulation alone improves CO_2 fixation by rubisco (26).

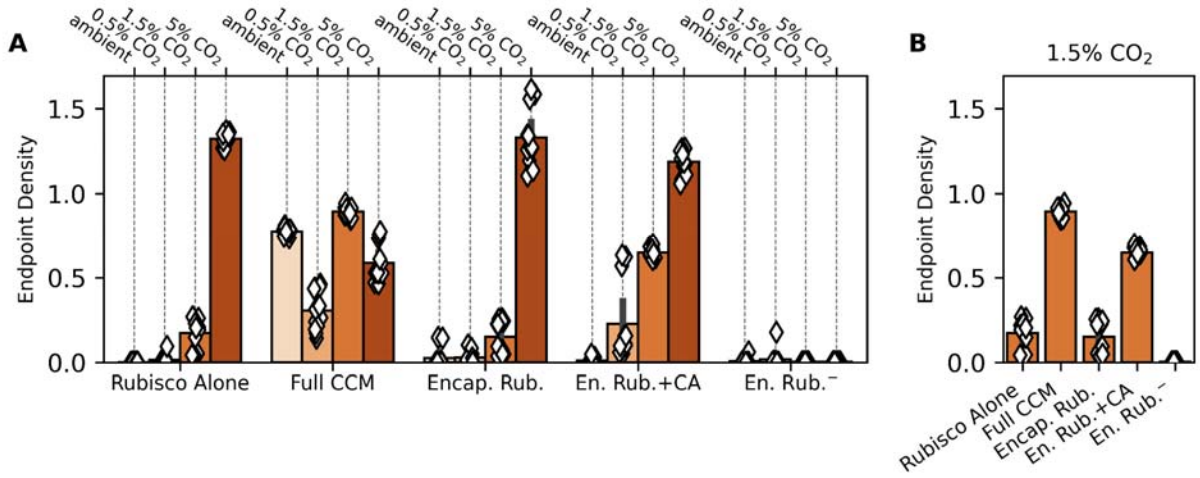

Fig. 3: Encapsulation of rubisco alone does not improve rubisco dependent growth of CCMB1 *E. coli* in any CO_2 level tested. (A) Using our reconstitution of the *H. neapolitanus* CCM in *E. coli* (4) we tested whether rubisco encapsulation has any effect on growth at four CO_2 partial pressures (Methods). Each diamond represents one technical replicate of four biological replicates. The CCMB1 strain grows in elevated CO_2 (1.5 and 5%) when rubisco is expressed ("Rubisco Alone", left). As previously reported, expressing the full complement of CCM genes from the pCB' and pCCM' plasmids ("Full CCM") enabled growth in all CO_2 levels. By replacing pCCM with a vector control and making an inactivating mutation to the carboxysomal CA (CsosCA C173S) we were able to express rubisco in a carboxysome without CA or Ci transport activities ("Encap. Rub."). This strain grew similarly to the reference "Rubisco Alone" in all conditions. When the CA active site was left intact ("En. Rub.+CA"), growth improved above the "Rubisco Alone" baseline in 0.5% and 1.5% CO_2. Finally, a negative control strain carrying inactive rubisco ("En. Rub.⁻", CbbL K194M) failed to grow in all CO_2 conditions. (B) Focusing on the growth in 1.5% CO_2 highlights the contribution of CA activity to rubisco-dependent growth. See Tables S4-5 for description of strains and plasmids, see Figures S4-5 for growth curves and analysis of statistical significance.

Carbonic anhydrase and energy-coupled inorganic carbon transport improve the growth of rubisco-dependent E. coli

Our observation that carboxysomal CA activity improves rubisco-dependent growth on its own (i.e. without also expressing a Ci transporter) motivated us to test the effects of CA and Ci transport independently of carboxysome expression. We designed plasmids that express the *H. neapolitanus* Dab2 Ci transporter (22), *E. coli*'s native CA (Can) or both. This was achieved by cloning both Dab2 and Can

into a dual expression vector and making active site mutants (DabA C539A, Can C48A, D50A) to isolate each activity. These vectors were co-transformed into CCMB1 with a constitutive version of p1A (p1Ac) so that rubisco expression would not be affected by induction of Dab2 or Can (*Methods*).

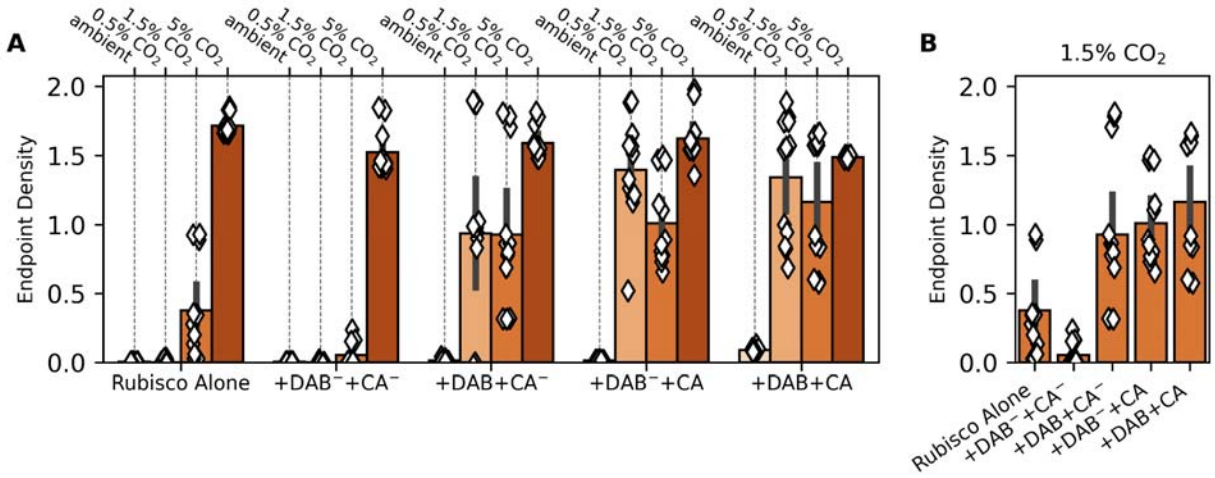

Fig. 4: Expression of carbonic anhydrase or Ci transport improves rubisco-dependent growth of CCMB1 *E. coli* in intermediate CO_2 levels even in the absence of other CCM components. As in Figure 3, the reference strain ("Rubisco Alone") grew only in elevated CO_2 (1.5% and 5%). The remaining strains expressed the *E. coli* carbonic anhydrase (Can) and the DAB2 Ci transporter from a second plasmid. These activities were isolated by means of active site mutations. The negative control "+DAB$^-$ +CA$^-$" expressed inactive Can (C48A, D50A) and DAB2 (DabA2 C539A) and grew less robustly than the reference in all conditions. If either active site was left intact ("+DAB+CA$^-$" or "+DAB$^-$+CA") we observed a sizable growth improvement in both 0.5 and 1.5% CO_2. This growth improvement remained when both active sites were left intact ("+DAB+CA"). Panel (B) emphasizes this effect clearly by focusing on growth in 1.5% CO_2. See Table S4 for strain genotypes, Figures S6-7 for growth curves and analysis of statistical significance.

Expressing active Can and Dab2, whether alone or together, improved growth substantially in 1.5% CO_2 (compare '+DAB+CA$^-$', '+DAB$^-$+CA', and '+DAB+CA' to 'Rubisco Alone' in Figure 4). This effect was even more pronounced in 0.5% CO_2, where the reference strain failed to grow (Figures 4A and S7). Similar to the reference strain, a double-negative control strain expressing inactivated versions of both Dab2 and Can ('+Dab$^-$+CA$^-$') grew poorly or not at all in 0.5% and 1.5% CO_2, implying that the observed growth improvements were due to activity and not a side-effect of heterologous gene expression.

It was not immediately obvious to us how CA and Ci uptake activities improve rubisco-dependent growth of CCMB1 *E. coli*. We found the effects of Dab2 expression especially perplexing because Ci uptake is expected to generate intracellular HCO_3^- (16–18, 22, 36) while the rubisco substrate is CO_2 (37). To confirm that these results are not a side-effect of working in an engineered *E. coli* strain, we pursued genetic experiments in a natively autotrophic proteobacterium, *C. necator*.

C. necator depends on CA for autotrophic growth at intermediate pCO_2

While all photosynthetic Cyanobacteria rely on the CBB cycle and a full complement of CCM genes (38), some chemoautotrophic bacteria depend on the CBB cycle but lack identifiable genes encoding carboxysome components or Ci transporters (39–42). As most characterized bacterial rubiscos are not CO_2-saturated in ambient air and are, in addition, substantially inhibited by atmospheric levels of O_2 (7), we expected that such organisms would require elevated CO_2 for robust growth.

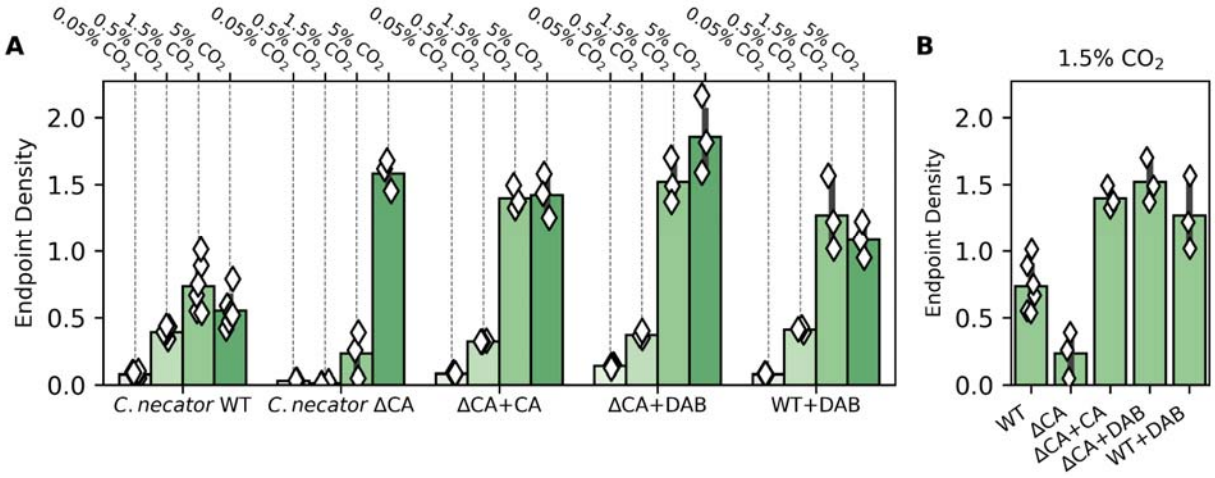

Fig. 5: *C. necator* requires carbonic anhydrase or Ci uptake for robust autotrophic growth in 0.5% and 1.5% CO_2. *C. necator* strains were grown autotrophically in minimal medium at a variety of CO_2 levels and endpoint optical density was measured after 48 hours (*Methods*). (A) Growth of the *C. necator* double carbonic anhydrase knockout (ΔCA) was greatly impaired growth in 0.5% and 1.5% CO_2. Panel (B) focuses on 1.5% CO_2. Compared to wild-type *C. necator* (WT), which grew to a final OD600 of 0.73±0.28 in 1.5% CO_2 (6 biological replicates), growth of ΔCA was greatly impaired, reaching a final OD of 0.23±0.17 (3 biological replicates). Expression of either the human carbonic anhydrase II (ΔCA+CA) or the DAB2 Ci transporter from *H. neapolitanus* (ΔCA+DAB) recovered robust growth which exceeded even the wild type, indicating that the wild type may not express saturating levels of CA. See Figure S8 for statistical analysis.

Cupriavidus necator, formerly known as *Ralstonia eutropha*, is one such bacterium (43, 44). *C. necator* is a facultative chemolithoautotroph typically found at the interface between oxic and anoxic environments where H_2 and O_2 coexist. Such "knallgas" environments include soils, sediments and geothermal sites (45) that are often characterized by elevated CO_2 (45–47). While *C. necator* is an obligate aerobe capable of chemoautotrophic growth on H_2, CO_2 and O_2 via the CBB cycle, it has no carboxysome genes and no obvious Ci transporters (41). Consistent with previous reports (48), autotrophic growth of wild-type *C. necator* was very poor in ambient CO_2 ('*C. necator* WT' in Figure 5A). We generated a double CA knockout strain, *C. necator* Δ*can* Δ*caa* (42, 49) and found that CA removal greatly attenuated autotrophic growth in 0.5% and 1.5% CO_2 ('*C. necator* ΔCA'). Consistent with our experiments in *E. coli*, this growth defect was complemented by heterologous expression of a human CA ('ΔCA+CA') or a DAB-type Ci uptake system ('ΔCA+DAB'). Moreover, as in *E. coli*, co-expression of Ci uptake with native CAs was not deleterious ('WT+DAB'). Rather, this strain grew to higher densities than wild-type in 1.5% and 5% CO_2 (e.g. Fig. 5B).

A nutritional requirement for HCO_3^- explains the observed phenotypes

We found that expression of a CA, Ci transporter, or both improved rubisco-dependent growth of *C. necator* and CCMB1 *E. coli* in intermediate CO_2 concentrations (0.5 and 1.5%, Figs. 4-5). Furthermore, to our surprise, a CCMB1 strain expressing active Dab2 and Can grew reproducibly, if slowly, in ambient air ('+Dab+CA' in Figure 6). This was remarkable because biological membranes are very permeable to CO_2, with a permeability coefficient $P_C \approx 0.1\text{-}1$ cm/s (16, 50, 51) that is several orders-of-magnitude larger than those of small charged species like acetate, phosphate and bicarbonate (50–53). Given this high permeability, we expected co-expression of energized HCO_3^- uptake with a CA to generate a deleterious "futile cycle" where energy expended pumping HCO_3^- was wasted when CO_2 produced by CA activity "leaks" back out of the cell. Instead we found that such a cycle is compatible with rubisco-dependent growth of two bacteria in relatively low CO_2 (1.5% or lower).

A naive explanation of how CA expression might improve growth is that it increases the intracellular CO_2 concentration relative to the reference strain expressing only rubisco. For this hypothesis to explain the benefit of CA expression, rubisco activity must deplete intracellular CO_2 (C_{in}) significantly below its extracellular level (C_{out}) in the reference strain. Otherwise CA can have no effect, as carbonic anhydrases are not coupled to any energy source and so cannot cause C_{in} to exceed C_{out}. However, this naive model cannot explain the growth benefits associated with expressing Ci uptake systems, which provide HCO_3^- and not CO_2. Moreover, the following calculation shows that this hypothesis is unreasonable precisely because CO_2 is so membrane-permeable that rubisco cannot deplete C_{in} much beneath C_{out}.

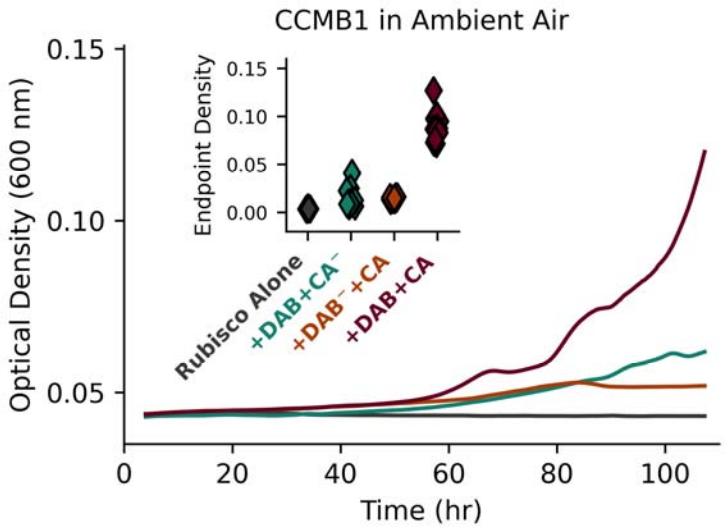

Fig. 6: Co-expression of carbonic anhydrase and Ci uptake enabled rubisco-dependent growth of CCMB1 *E. coli* in ambient air. Inspecting the ambient CO_2 growth data presented in Figure 4 revealed that coexpression of CA and Ci transport ('Rub.+DAB+CA') substantially improved rubisco-dependent growth of CCMB1 *E. coli* in ambient CO_2 concentrations. This effect was modest (≈0.1 OD units above the 'Rubisco Alone' control) but reproducible, as indicated by endpoint data plotted on the inset. Curves are colored to match labels on the inset. See Figure S9 for statistical analysis.

In a bacterium, rubisco might make up 20% of soluble protein at the very most. This amounts to a mass concentration of roughly 0.2 x 300 mg/ml ≈ 60 mg/ml rubisco (54). As each rubisco active site is attached to ≈60 kDa of protein (BNID 105007), the maximum rubisco active site concentration is ≈1 mM.

346 In this naive model, C_{in} is set by the balance of passive uptake through the membrane, with effective
347 permeability □ = P_C x SA/V ≈ 10^4 s^{-1}, and fixation by rubisco, with an effective rate constant of □ =
348 [rubisco] x k_{cat}/K_M < 10^3 s^{-1}. These values give a steady-state C_{in} of □ C_{out} / (□ + □) > 0.9 C_{out}. That is, even
349 an extreme level of rubisco activity cannot draw C_{in} beneath 90% of C_{out}. In such conditions, rubisco
350 would fix ≈10^{10} CO_2/hour, supporting a 1-2 hr doubling time (see *Supplementary Information* for full
351 calculation). So, although the CO_2 concentration is low in aqueous environments equilibrated with
352 present-day atmosphere (≈10 μM), passive CO_2 uptake can support substantial rubisco flux and,
353 therefore, CO_2 limitation of rubisco carboxylation cannot explain the observed phenotypes.
354
355 We argue that the effects of expressing CA or Ci transport can be explained by the ubiquitous
356 dependence of growth on HCO_3^- (55–57). It has been clear for at least 80 years that heterotrophs also
357 require inorganic carbon for growth. Seminal investigations in the 1930s and 1940s advanced the
358 hypothesis that this dependence is due to a specific requirement for HCO_3^- (58–61), which is the
359 substrate of several carboxylation reactions involved in lipid, nucleotide and arginine biosynthesis. This
360 explanation is now supported by detailed experiments in several heterotrophs demonstrating that
361 growth in ambient air can be supported either by CA activity (55–57) or by providing the products of
362 central-metabolic carboxylation reactions in the growth media (56, 57). The metabolic networks of
363 chemoautotrophs, Cyanobacteria and plants imply an equivalent dependence on HCO_3^- for biosynthesis
364 (62–64). We therefore advance a model of co-limitation of autotrophic growth by both CO_2 and HCO_3^--
365 dependent carboxylation fluxes, where most of biomass carbon derives from rubisco-catalyzed
366 carboxylation of CO_2 and a small minority derives from HCO_3^- (Figure 7A). We formalized this notion as
367 an idealized mathematical model fully described in *Supplementary Information*.
368
369

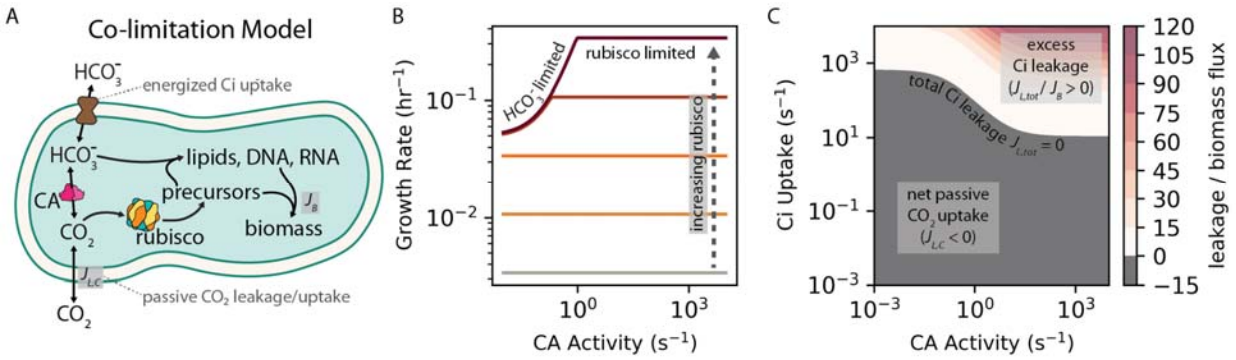

Fig. 7: Co-limitation of autotrophic growth by CO_2 and HCO_3^--dependent carboxylation reactions explains the growth improvements associated with expressing carbonic anhydrases and Ci transporters. (A) In autotrophs using the CBB cycle, nearly all biomass carbon derives from rubisco-catalyzed CO_2 fixation. However, autotrophs also require HCO_3^- for carboxylation reactions in lipid, nucleic acid, and arginine biosynthesis (55–57). We expressed this diagram as a mathematical model, which we applied to understand why CA and Ci uptake improved rubisco-dependent growth. (B) The model exhibited two regimes: one wherein growth was limited by rubisco flux another where it was limited by HCO_3^--dependent carboxylation ("bicarboxylation") flux. At low rubisco levels (lighter-colored lines), growth was rubisco-limited: increased rubisco activity produced faster growth, but the growth rate was insensitive to CA activity because slow CO_2 hydration provided sufficient HCO_3^- to keep pace with rubisco. At higher rubisco levels (maroon lines), growth was bicarboxylation-limited and increased CA activity was required for increasing rubisco activity to translate into faster growth. Increasing Ci uptake led to similar effects (Figure S11). In panel (C), color indicates the ratio of total Ci leakage ($J_{L,tot} = J_{L,C} + J_{L,H}$) to biomass production flux (J_B) at a fixed rubisco activity across a wide range of CA activities and Ci uptake rates. $J_{L,tot}$ was calculated as the sum of CO_2 and HCO_3^- leakage rates ($J_{L,C} + J_{L,H}$) with $J_{L,C} \gg J_{L,H}$ in most conditions due to the much greater membrane permeability of CO_2. So-called "futile cycling," where leakage greatly exceeds biomass production ($J_{L,tot} / J_B \gg 1$), occurs when CA and Ci uptake are co-expressed at extreme levels (redder colors). See *Supplementary Information* for detailed description of the co-limitation model.

When the HCO_3^--dependence of biomass production was included in our model, it became possible to rationalize the growth improvements associated with heterologous expression of a CA or Ci transporter, as both of these activities can supply HCO_3^- needed for biosynthesis (Figure 7A). When the modeled rubisco activity was low, growth was rubisco-limited: increasing rubisco activity (lighter colored lines in Figure 7B) or the CO_2 concentration produced faster growth, but the growth rate was insensitive to CA activity because slow CO_2 hydration (whether spontaneous or CA catalyzed) provided sufficient HCO_3^- for the anabolic carboxylations involved in biomass production (Figures S11 and S12). When rubisco activity was set to higher values (darker lines in Figure 7B) the model entered a "bicarboxylation-limited" regime where increased rubisco activity did not affect the growth rate until CA activity was increased to supply HCO_3^- (i.e. bicarbonate) for anabolic carboxylations. In bicarboxylation-limited conditions,

399 additional HCO_3^- could be supplied by CA, Ci uptake, or both (Figures S11). However, these activities are
400 not equivalent, as we discuss below.
401
402 The co-limitation model also gave insight into the nature of the supposed "futile cycle," which could
403 now be framed quantitatively by comparing the leakage of inorganic carbon from the cell (flux $J_{L,tot} = J_{L,C}$
404 + $J_{L,H}$) to biomass production by rubisco and HCO_3^--dependent carboxylases (J_B, Fig. 7A). As CO_2 is orders
405 of magnitude more membrane-permeable than HCO_3^- near neutral pH (16), CO_2 leakage ($J_{L,C}$) greatly
406 exceeded the HCO_3^- leakage flux ($J_{L,H}$) in most conditions. Low Ci leakage is desirable; when $J_{L,tot} \approx 0$ the
407 flux of Ci pumped into the cell is balanced by fixation of CO_2 and HCO_3^- such that no energy is "wasted"
408 on Ci uptake. When CA (□) and Ci uptake (□) activities were low, the model predicted slow growth and
409 net diffusive Ci uptake ($J_{L,C}$, $J_{L,H} < 0$) to balance carboxylation (Figure 7C and S14). Increasing □ could
410 cause $J_{L,tot}$ to change sign from negative (connoting passive uptake) to positive (connoting substantial
411 leakage) passing through $J_{L,tot} = 0$. If CA was also expressed - i.e. if □ was increased above a baseline
412 value - $J_{L,tot} \approx 0$ could be achieved at lower □ values due to CA-catalyzed dehydration of imported HCO_3^-
413 producing CO_2 used by rubisco. In other words, CA activity lowers the Ci uptake rate and, consequently,
414 energy expenditure, required to achieve the same rate of biomass production (Figures 7C & S14). When
415 □ and □ were both high, the model predicted substantial leakage, with $J_{L,tot} / J_B \approx 100$ in extreme cases.
416 Such extreme levels of futile cycling are likely incompatible with growth. If □ was low, in contrast, □
417 could be increased arbitrarily without incurring leakage because CAs do not pump Ci into the cell.
418
419 A second, less plausible way in which co-expression might improve growth is if CA and Ci transport are
420 fast and membrane permeability to CO_2 is far lower than typically assumed. In this case, the
421 combination of Ci uptake and CA activity might form a CO_2 pump that elevates C_{in} substantially above
422 C_{out} to accelerate rubisco carboxylation (Figure S13). As HCO_3^- is supplied sufficiently by Ci uptake in this

423 case, the co-limitation model predicted that increased rubisco fluxes would translate into faster growth.
424 For this effect to arise, however, the membrane permeability to CO_2 would have to be 100-1000 times
425 lower than the measured (50) or calculated (51), so CO_2 pumping is unlikely to explain the capacity of a
426 CCMB1 to grow in ambient air when co-expressing Dab2 and Can (Fig. 6).
427
428 Taken together, our experiments and model helped outline a plausible trajectory for the co-evolution of
429 the bacterial CCM with atmospheric CO_2 levels. Presuming an ancestral autotroph with only a rubisco-
430 driven CBB cycle and no CCM components, our data and model support a trajectory where CA and Ci
431 transport are acquired together or serially (in either order) to support growth as atmospheric CO_2 levels
432 decreased (darker arrows in Fig. 1C). The order of acquisition might depend on the environmental pH,
433 which strongly affects the extracellular HCO_3^- concentration and, thereby, the expected efficacy of HCO_3^-
434 uptake (16). The co-limitation model helped us understand the potential advantages of expressing CA
435 and Ci uptake together: modest co-expression can reduce energy expended on Ci pumping and balance
436 the supply of CO_2 and HCO_3^- with the cellular demand for rubisco and bicarboxylation products (Figure
437 7C and S14). In CCMB1 we found that co-expression of CA and Ci transport supported growth in low CO_2
438 environments (≈1%, Figure 4) and even permitted modest growth in atmosphere (Figure 6); cells
439 expressing both activities would have been "primed" for the subsequent acquisition and refinement of
440 proto-carboxysome structures that co-encapsulate rubisco and CA to enable robust growth at yet lower
441 CO_2 levels (e.g. ambient air, Figure 3). Notably, carboxysome shell proteins are structurally related to
442 two ubiquitous protein families (65) and homologous to other metabolic microcompartments (28, 65),
443 suggesting two plausible routes for the acquisition of carboxysome genes.
444
445 Results from our *E. coli* experiments indicated that such an evolutionary trajectory is "fitness positive" in
446 that each step improves organismal fitness at some environmental CO_2 level (Figs. 3, 4, and 6).

Moreover, the fitness contribution of CA and Ci transport activities was only realized at intermediate $pCO_2 \approx 1\%$ in both *E. coli* (Fig. 4) and *C. necator* (Fig. 5), supporting the view that CO_2 concentrations declining from high levels early in Earth history promoted the evolutionary emergence of bacterial CCMs.

Concluding Remarks

There is great and longstanding interest in characterizing the composition of the atmosphere over Earth history (9, 11, 66). This interest is surely justified as the contents of the atmosphere affected the temperature, climate, and the chemical conditions in which life arose, evolved, and has been maintained. In the case of reactive species like O_2, present-day measurements of old sedimentary rocks are quite informative: measurements of proxies for O_2 reactivity in geological samples demonstrate that the Archean atmosphere contained very little dioxygen, with inferred O_2 levels of 1 ppm or less (8, 9). Many posit that the Archean atmosphere also contained very high levels of CO_2. However, since CO_2 is far less reactive in Earth surface environments than O_2, available geological constraints are meager and this inference stands on shakier ground.

One reason for presuming high CO_2 levels in the early atmosphere is the paradox of the "Faint Young Sun," which would not have been sufficiently luminous to permit liquid water on Earth without a very substantial greenhouse effect (67) or some other physical mechanism warming the Earth system (68). Other approaches to estimating ancient CO_2 pressures rely indirectly on geochemical observations and arguments from the geological record. For example, some studies infer high Archean CO_2 from the presence or absence of certain carbonate minerals in ancient surface environments (9), while others rely on models of biological CO_2 fixation to estimate CO_2 concentrations from the carbon isotope ratios of organic matter in sedimentary rocks (11, 12). These diverse approaches vary in the magnitude of their

estimates, but each gives the impression that Archean CO_2 greatly exceeded present-day levels of roughly 0.04% of a 1 bar atmosphere; indeed, some estimates of Archean CO_2 pressures are as high ≈1 bar (9, 67). Silicate weathering coupled to the deposition of carbonate-bearing rocks represents a long-term crustal sink that — combined with subequal amounts of organic carbon burial — would have drawn a large reservoir of atmospheric CO_2 down to pre-industrial levels (10, 69). Furthermore, modern plants, algae, and autotrophic bacterial lineages collectively evolved four distinct classes of CCM at least 100 times indicating that autotrophy has repeatedly and convergently adapted to decreasing environmental CO_2 levels (13).

Carbon isotope methods for inferring historical changes in CO_2 rely on uniformitarian models of biological carbon fixation to calculate atmospheric CO_2 pressures from isotope ratio measurements of biogenic organic phases preserved in sediments and rocks (12). Yet autotrophy has undergone considerable evolution and diversification over the last ≈3 billion years, which complicates inference of ancient CO_2 levels with models based on contemporary autotrophs. For example, rubiscos from different lineages display variable $^{12}CO_2$ preferences (23, 70) and contemporary CCMs rely on Ci uptake systems that may fractionate C isotopes to a degree similar to rubisco (71). As demonstrated in the companion paper by Renee Wang and co-authors, the evolution of CCMs must be considered when attempting to reconstruct historical CO_2 concentrations from the C isotope record.

Here we attempted to reconstruct the evolution of a bacterial CCM by constructing analogues of plausible ancestors. This investigation was initially motivated by a basic curiosity about the evolution of physiological complexity, hoping to resolve the apparent “irreducible complexity” of bacterial CCMs (Figure 1). Irreducible complexity is incompatible with evolution by natural selection, and so we

494 hypothesized that ancestral CCMs conferred a growth advantage in the historical environments in which
495 they arose, which we and others presume were characterized by high CO_2 (5, 24–26).
496
497 Results from our experiments with rubisco-dependent *E. coli* and two bacterial autotrophs (*H.*
498 *neapolitanus* and *C. necator*) supported this rationalization of CCM evolution by showing (i) that CO_2 ≈
499 1% improves the growth of "ancestral" CCMs (Figs. 3-6) and (ii) that all CCM genes are dispensable
500 during growth in 5-10% CO_2 (Fig. 2-5). This latter result suggested to us that atmospheric CO_2
501 concentrations exceeded 1% of a 1 bar atmosphere when rubisco arose, which was likely more than 3
502 billion years ago (24, 72). If geological CO_2 sinks later brought atmospheric CO_2 to ≈1%, all organisms,
503 including autotrophs, would have begun to evolve or acquire carbonic anhydrases and/or Ci transporters
504 to provide the HCO_3^- required for biosynthesis (Figure 7). An ancestral autotroph expressing both these
505 activities may have had a growth advantage in relatively lower CO_2 pressures < 1% (Fig. 6A), and would
506 have been "primed" for the evolution of a CCM as the only missing component, the carboxysome, could
507 have evolved from oligomeric host proteins (65) or adapted from a different metabolic
508 microcompartment (28). Alternatively, it is possible that CAs (or Ci transporters) arose prior to rubisco
509 and were already widespread at the time of rubisco evolution, in which case we might expect CO_2 ⍰ 1%
510 when rubisco arose. Unfortunately, the convergent evolution of CA activity in several protein families
511 (73) makes it very challenging to constrain the timing of CA evolution with comparative biological and
512 molecular clock approaches; this issue concerns bacterial Ci transporters as well (17, 22).
513
514 It is intrinsically difficult to answer questions about Earth's deep biological history; addressing such
515 questions will surely require cooperation between scientific disciplines. Here, and in the companion
516 paper by Renee Wang et. al., we took a "synthetic biological" approach to study the molecular evolution
517 of bacterial CO_2 fixation by constructing contemporary cells intended to resemble ancient ones. In both

518 cases, our work highlighted the impact of environmental context (e.g. CO_2 concentrations) and whole
519 cell physiology (the requirement for HCO_3^-) on the evolution of CO_2 fixation. That is, neither rubisco nor
520 the CCM should be considered in isolation, but rather in the context of a metabolism that demands both
521 CO_2 and HCO_3^- to produce biomass (55–61, 71). We hope that future research advances the synthetic
522 approach to studying evolution and fully expect that this approach will enrich our understanding of
523 biological processes that have shaped the evolution of biogeochemical cycles on Earth.

524

Methods

Strains, plasmids and genomic modifications

Strains and plasmids used in this study are documented in Supplementary Tables S4 and S5. The rubisco-dependent *E. coli* strain CCMB1 was derived from *E. coli* BW25113 and has the genotype BW25113 Δ*rpiAB* Δ*edd* Δ*cynT* Δ*can*, as documented in (4). To construct CA deficient mutants of *C. necator* H16, we first knocked out the hdsR homolog *A0006* as removal of this restriction enzyme increases electroporation efficiency (41). The double CA knockout, *C. necator* H16 Δ*A0006* Δ*can* Δ*caa*, was constructed by repeated rounds of selection and counter-selection by integrating a construct encoding both kanamycin resistance and *sacB* for counter-selection. Protocols for plasmid construction and genomic modification of *C. necator* are fully described in the *Supplementary Information*.

Genome-wide fitness measurements in *H. neapolitanus*

Competitive fitness assays were performed following (22). The barcoded *H. neapolitanus* transposon library generated for that work was thawed and used to inoculate three 33 ml cultures that were grown overnight at 30 C in DSMZ-68 with 10 μg/ml Kanamycin. These pre-cultures were grown to an OD ≈ 0.07 (600 nm) in 5% CO_2. The library was subsequently back-diluted 1:64 and grown in various CO_2 concentrations (10%, 5%, 1.5%, 0.5% and ambient CO_2) on a platform shaker (New Brunswick Scientific Innova 2000, 250 RPM) in a Percival Intellus Incubator; 20 ml of pre-culture was pelleted by centrifugation (15 min at 4000g) and saved as a T_0 reference. Upon reaching 6.5-7.5 doublings, 50 mls of culture was spun down and gDNA was extracted for barcode PCRs as described in (6); barcodes were previously mapped to the genome via TnSeq (22). PCRs were purified (Zymo Research Clean and Concentrator kit) and pooled for sequencing on an Illumina MiSeq with 150 bp single end reads. We used the software pipeline from (32) to analyze barcode abundance data. Briefly, fitness of individual mutant strains was calculated as the $\log_2$ of the ratio of barcode abundance in the experimental condition over abundance in the T_0. Gene-level fitness values were then calculated considering all

549 transposon insertions expected to disrupt an individual gene. Each CO_2 concentration was tested in
550 biological duplicate except for 5% CO_2, which was assayed in biological quadruplicate.
551

552 ## *E. coli* growth conditions

553 *E. coli* strains were grown at 37 °C. 60 µg/ml kanamycin and 25 µg/ml chloramphenicol were used for
554 selection during routine cloning and propagation. For strains carrying two selectable markers, antibiotics
555 were used at half concentration (30 µg/ml kanamycin, 12.5 µg/ml chloramphenicol). All CCMB1 cultures
556 were grown in 10% CO_2 unless otherwise specified. When rubisco-independent growth was desired,
557 CCMB1 was propagated in rich LB media. CCMB1 was cultured in a rubisco-dependent manner in M9
558 minimal media supplemented with trace elements and 0.4% glycerol (v/v) as described in (4).
559 Experiments presented in Figures 3, 4 and 6 were conducted in 96-well plates in a gas-controlled plate
560 reader (Tecan Spark). For the data in Figure 3, 100 nM anhydrous tetracycline (aTc) was supplied to
561 induce expression from p1A, pCB' and pCCM' plasmids. For the data presented in Figures 4 and 6, a
562 constitutive version of p1A was used (p1Ac) and aTc was omitted from pre-cultures. These strains also
563 carried a dual-expression plasmid, pFC, for inducible expression of the DAB2 Ci transport operon and the
564 *can* carbonic anhydrase. 100nM aTc was supplied to induce DAB2 and 1 mM IPTG for *can*.
565

566 Pre-cultures were inoculated into 5 ml of M9 glycerol with 30 µg/ml kanamycin, 12.5 µg/ml
567 chloramphenicol and appropriate induction. 1 ml of each culture was transferred to a separate tube,
568 which was incubated in ambient CO_2 as a negative control; the remaining 4 ml was incubated in 10%
569 CO_2. CCMB1 strains carrying active site mutants of rubisco (cbbL K194M) were pre-cultured in LB with
570 the same antibiotic concentrations. Once cultures reached saturation they were centrifuged at 4000 g
571 for 8 min. Pellets were washed in 10 ml of M9 with no carbon source (M9 NoC) and resuspended in 5 ml
572 M9 NoC. Optical density was measured in five-fold dilution at 600 nm (Genesys 20 spectrophotometer,

Thermo Scientific) and cultures were normalized to 0.5 OD units. 96-well plates were inoculated by adding 2.5 µl of OD-normalized preculture to 247.5 µl M9 glycerol supplemented with appropriate antibiotics and induction. Kanamycin was omitted as the plasmid carrying kanamycin resistance also expresses rubisco, which is required for growth in glycerol media (4). To minimize evaporation during multi-day cultivations, 150 µl sterile water was added to the reservoirs between the wells and the plate was incubated inside of a small humidity cassette (Tecan) with 3 ml sterile water added to each reservoir. Plates were incubated with shaking for at least 4 days in a Tecan Spark plate reader configured to control the CO_2 concentration. The humidity cassette was replenished after 48 hours. All experiments were performed in biological quadruplicate and technical triplicate.

C. necator growth conditions

C. necator strains were grown in ambient CO_2 unless otherwise specified, and 200 µg/ml kanamycin was added to select for plasmid retention. When rubisco-independent heterotrophic growth was desired, *C. necator* was cultured in LB at 30 °C. For autotrophic growth experiments, strains were pre-cultured in 5 mL of LB media in 10% CO_2 in 25 mL tubes with 20 mm butyl stoppers sealed by aluminum crimping. Pre-cultures were incubated for two days at 30 °C with 200 rpm shaking, washed three times in *Cupriavidus* minimal media, and inoculated to an OD of 0.1 (550 nm) in minimal media in a 165 mL flask with a 20 mm butyl stopper sealed by aluminum crimping. *Cupriavidus* minimal growth media contained 3.24 mM $MgSO_4$, 0.42 mM $CaCl_2$, 33.5 mM NaH_2PO_4, 32.25 mM Na_2HPO_4, 2.6 mM K_2SO_4, 1 mM NaOH, 1.87 mM NH_4Cl, and 1 mL/L of a 1000x trace minerals solution containing 480 mg/L $CuSO_4{\cdot}5H_2O$, 2.4 g/L $ZnSO_4{\cdot}7H_2O$, 2.4 g/L $MnSO_4{\cdot}H_2O$, and 15 g/L $FeSO_4{\cdot}7H_2O$ (pH 6.7). The flasks were evacuated and the headspace was filled with 60% H_2, 10% O_2, and indicated concentrations of CO_2. The balance was atmospheric air. Cells were grown for 48 hours at 30° C with 200 rpm shaking, and OD_{550} values were taken using a Molecular Devices SpectraMax M2 spectrophotometer.

Acknowledgements

We dedicate this paper to the memory of Danny Salah Tawfik (z"l), a luminary to the scientific community and a dear friend and mentor to AIF. Danny's many studies of enzyme evolution taught us that evolutionary timescales are accessible in carefully-designed lab-scale experiments and his active cultivation of relationships that transcend generational, professional and cultural divides continues to inspire. We are also indebted to Arren Bar-Even (z"l) for formative conversations guiding this work. Many thanks to Darcy McRose, Elad Noor, and Renee Wang for extensive comments on the manuscript and to Cecilia Blikstad, Julia Borden, Vahe Galstyan, Josh Goldford, Ron Milo, Robert Nichols, Naiya Phillips, Noam Prywes, and Gabe Salmon for helpful input. This research was supported in part by an NSF Graduate Research Fellowship (to A.I.F), NSF Grant No. PHY-1748958, the Gordon and Betty Moore Foundation Grant No. 2919.02, and the Kavli Foundation (to A.I.F), the US Department of Energy (DE-SC00016240 to D.F.S.) and Royal Dutch Shell (Energy and Biosciences Institute project CW163755 to D.F.S. and S.S.).

References

1. S. G. Wildman, Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate carboxylase-oxygenase). *Photosynth. Res.* **73**, 243–250 (2002).

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

3. C. Bathellier, G. Tcherkez, G. H. Lorimer, G. D. Farquhar, Rubisco is not really so bad. *Plant Cell Environ.* **41**, 705–716 (2018).

4. A. I. Flamholz, *et al.*, Functional reconstitution of a bacterial CO2 concentrating mechanism in E. coli. *Elife* **9** (2020).

5. C. Iñiguez, *et al.*, Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms. *Plant J.* **101**, 897–918 (2020).

6. G. Bowes, W. L. Ogren, R. H. Hageman, Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophys. Res. Commun.* **45**, 716–722 (1971).

7. A. I. Flamholz, *et al.*, Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* **58**, 3365–3376 (2019).

8. W. W. Fischer, J. Hemp, J. E. Johnson, Evolution of Oxygenic Photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016).

9. D. C. Catling, K. J. Zahnle, The Archean atmosphere. *Sci Adv* **6**, eaax1420 (2020).

10. R. A. Berner, A. C. Lasaga, R. M. Garrels, The carbonate-silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years. *Am. J. Sci.* **283**, 641–683 (1983).

11. K. H. Freeman, J. M. Hayes, Fractionation of carbon isotopes by phytoplankton and estimates of ancient CO 2 levels. *Global Biogeochem. Cycles* **6**, 185–198 (1992).

12. C. R. Witkowski, J. W. H. Weijers, B. Blais, S. Schouten, J. S. Sinninghe Damsté, Molecular fossils from phytoplankton reveal secular Pco2 trend over the Phanerozoic. *Sci. Adv.* **4**, eaat4556 (2018).

13. A. Flamholz, P. M. Shih, Cell biology of photosynthesis over geologic time. *Curr. Biol.* **30**, R490–R494 (2020).

14. Y. Marcus, R. Schwarz, D. Friedberg, A. Kaplan, High CO2 Requiring Mutant of Anacystis nidulans R2. *Plant Physiol.* **82**, 610–612 (1986).

15. J. Pierce, T. J. Carlson, J. G. Williams, A cyanobacterial mutant requiring the expression of ribulose bisphosphate carboxylase from a photosynthetic anaerobe. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5753–5757 (1989).

16. N. M. Mangan, A. Flamholz, R. D. Hood, R. Milo, D. F. Savage, pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354–62 (2016).

17. B. D. Rae, B. M. Long, M. R. Badger, G. D. Price, Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379 (2013).

18. G. D. Price, M. R. Badger, Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO2-Requiring Phenotype Evidence for a Central Role for Carboxysomes in the CO2 Concentrating Mechanism. *Plant Physiol.* **91**, 505–513 (1989).

19. L. Whitehead, B. M. Long, G. D. Price, M. R. Badger, Comparing the in Vivo Function of α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria. *Plant Physiol.* **165**, 398–411 (2014).

20. T. Ogawa, T. Kaneda, T. Omata, A Mutant of Synechococcus PCC7942 Incapable of Adapting to Low CO2 Concentration. *Plant Physiol.* **84**, 711–715 (1987).

21. G. D. Price, M. R. Badger, Isolation and characterization of high CO2-requiring-mutants of the cyanobacterium Synechococcus PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO2 within the carboxysome. *Plant Physiol.* **91**, 514–525 (1989).

22. J. J. Desmarais, *et al.*, DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nat Microbiol* **4**, 2204–2215 (2019).

23. P. D. Tortell, Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnol. Oceanogr.* **45**, 744–750 (2000).

24. E. G. Nisbet, *et al.*, The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology* **5**, 311–335 (2007).

25. J. N. Young, R. E. M. Rickaby, M. V. Kapralov, D. A. Filatov, Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 483–492 (2012).

26. B. M. Long, B. Förster, S. B. Pulsford, G. D. Price, M. R. Badger, Rubisco proton production can drive the elevation of CO2 within condensates and carboxysomes. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

27. K. S. Smith, C. Jakubzick, T. S. Whittam, J. G. Ferry, Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 15184–15189 (1999).

28. S. D. Axen, O. Erbilgin, C. A. Kerfeld, A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).

29. M. Krupovic, V. V. Dolja, E. V. Koonin, Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* **17**, 449–458 (2019).

30. N. Mangan, M. Brenner, Systems analysis of the CO2 concentrating mechanism in cyanobacteria. *Elife*, e02043 (2014).

31. J. M. Shively, F. Ball, D. H. Brown, R. E. Saunders, Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. *Science* **182**, 584–586 (1973).

32. K. M. Wetmore, *et al.*, Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306–15 (2015).

33. N. M. Wheatley, C. D. Sundberg, S. D. Gidaniyan, D. Cascio, T. O. Yeates, Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014).

34. O. Mueller-Cajar, The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol Biosci* **4**, 31 (2017).

35. J. S. MacCready, L. Tran, J. L. Basalla, P. Hakim, A. G. Vecchiarelli, The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* **116**, 277–297 (2021).

36. K. M. Scott, *et al.*, Diversity in CO2-Concentrating Mechanisms among Chemolithoautotrophs from the Genera Hydrogenovibrio, Thiomicrorhabdus, and Thiomicrospira, Ubiquitous in Sulfidic Habitats Worldwide. *Appl. Environ. Microbiol.* **85**, 1–19 (2019).

37. T. G. Cooper, D. Filmer, The active species of "CO2" utilized by ribulose diphosphate carboxylase. *J. Biol. Chem.* **244**, 1081–1083 (1969).

38. C. A. Kerfeld, M. R. Melnicki, Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).

39. A. Pohlmann, *et al.*, Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16. *Nat. Biotechnol.* **24**, 1257–1262 (2006).

40. M. R. Badger, E. J. Bek, Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *J. Exp. Bot.* **59**, 1525–1541 (2008).

41. J. Panich, B. Fong, S. W. Singer, Metabolic Engineering of Cupriavidus necator H16 for Sustainable Biofuels from CO2. *Trends Biotechnol.* (2021).

42. C. S. Gai, J. Lu, C. J. Brigham, A. C. Bernardi, A. J. Sinskey, Insights into bacterial CO2 metabolism revealed by the characterization of four carbonic anhydrases in Ralstonia eutropha H16. *AMB Express* **4**, 2 (2014).

43. H. G. Schlegel, H. Kaltwasser, G. Gottschalk, Ein Submersverfahren zur Kultur wasserstoffoxydierender Bakterien: Wachstumsphysiologische Untersuchungen. *Arch. Microbiol.* **38**, 209–222 (1961).

44. W. Ahrens, H. G. Schlegel, Carbon dioxide requiring mutants of Hydrogenomonas eutropha strain H 16. I. Growth and Co2-fixation. *Arch. Mikrobiol.* **85**, 142–152 (1972).

45. M. Aragno, "Thermophilic, Aerobic, Hydrogen-Oxidizing (Knallgas) Bacteria" in *The Prokaryotes: A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, A. Balows, H. G. Trüper, M. Dworkin, W. Harder, K.-H. Schleifer, Eds. (Springer New York, 1992), pp. 3917–3933.

46. D. de Beer, A. Glud, E. Epping, M. Kûhl, A fast-responding CO2 microelectrode for profiling sediments, microbial mats, and biofilms. *Limnol. Oceanogr.* **42**, 1590–1600 (1997).

47. Z. S. Brecheisen, C. W. Cook, P. R. Heine, J. Ryang, D. D. Richter, Development and deployment of a field-portable soil O2 and CO2 gas analyzer and sampler. *PLoS One* **14**, e0220176 (2019).

48. N. J. Claassens, *et al.*, Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22452–22461 (2020).

49. B. Kusian, D. Sültemeyer, B. Bowien, Carbonic Anhydrase Is Essential for Growth of Ralstonia eutropha at Ambient CO2 Concentrations. *J. Bacteriol.* **184**, 5018–5026 (2002).

50. J. Gutknecht, M. A. Bisson, F. C. Tosteson, Diffusion of carbon dioxide through lipid bilayer membranes: effects of carbonic anhydrase, bicarbonate, and unstirred layers. *J. Gen. Physiol.* **69**, 779–794 (1977).

51. C. Hannesschlaeger, A. Horner, P. Pohl, Intrinsic Membrane Permeability to Small Molecules. *Chem. Rev.* **119**, 5922–5953 (2019).

52. A. C. Chakrabarti, D. W. Deamer, Permeability of lipid bilayers to amino acids and phosphate. *BBA - Biomembranes* **1111**, 171–177 (1992).

53. T.-X. Xiang, B. D. Anderson, The relationship between permeant size and permeability in lipid bilayer membranes. *J. Membr. Biol.* **140**, 111–122 (1994).

54. R. Milo, What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **35**, 1050–1055 (2013).

55. C. Merlin, M. Masters, S. McAteer, A. Coulson, Why is carbonic anhydrase essential to Escherichia coli? *J. Bacteriol.* **185**, 6415–6424 (2003).

56. J. Aguilera, J. P. Van Dijken, J. H. De Winde, J. T. Pronk, Carbonic anhydrase (Nce103p): an essential biosynthetic enzyme for growth of Saccharomyces cerevisiae at atmospheric carbon dioxide pressure. *Biochem. J.* **391**, 311–316 (2005).

57. P. Burghout, *et al.*, Carbonic anhydrase is essential for Streptococcus pneumoniae growth in environmental ambient air. *J. Bacteriol.* **192**, 4054–4062 (2010).

58. G. P. Gladstone, P. Fildes, G. M. Richardson, Carbon Dioxide as an Essential Factor in the Growth of Bacteria. *Br. J. Exp. Pathol.* **16**, 335 (1935).

59. H. G. Wood, C. H. Werkman, A. Hemingway, A. O. Nier, MECHANISM OF FIXATION OF CARBON DIOXIDE IN THE KREBS CYCLE. *J. Biol. Chem.* **139**, 483–484 (1941).

60. H. A. Krebs, Carbon Dioxide Assimilation in Heterotrophic Organisms. *Nature* **147**, 560–563 (1941).

61. A. Lwoff, J. Monod, Essai d'analyse du role de l'anhydride carbonique dans la croissance microbienne. *Ann. Inst. Pasteur* (1947).

62. J. M. Park, T. Y. Kim, S. Y. Lee, Genome-scale reconstruction and in silico analysis of the Ralstonia eutropha H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production. *BMC Syst. Biol.* **5**, 101 (2011).

63. H. Knoop, *et al.*, Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of

753 Synechocystis sp. PCC 6803. *PLoS Comput. Biol.* **9** (2013).

754 64. K. M. Hines, V. Chaudhari, K. N. Edgeworth, T. G. Owens, M. R. Hanson, Absence of carbonic
755 anhydrase in chloroplasts affects C3 plant development but not photosynthesis. *Proc. Natl. Acad.*
756 *Sci. U. S. A.* **118** (2021).

757 65. M. Krupovic, E. V. Koonin, Cellular origin of the viral capsid-like bacterial microcompartments. *Biol.*
758 *Direct* **12**, 25 (2017).

759 66. L. V. Berkner, L. C. Marshall, N.A.S. Symposium on the Evolution of the Earth's Atmosphere:
760 HISTORY OF MAJOR ATMOSPHERIC COMPONENTS. *Proc. Natl. Acad. Sci. U. S. A.* **53**, 1215 (1965).

761 67. J. F. Kasting, Faint young Sun redux. *Nature* **464**, 687–689 (2010).

762 68. M. T. Rosing, D. K. Bird, N. H. Sleep, C. J. Bjerrum, No climate paradox under the faint early Sun.
763 *Nature* **464**, 744–747 (2010).

764 69. Z. Liu, W. Dreybrodt, H. Liu, Atmospheric CO2 sink: Silicate weathering or carbonate weathering?
765 *Appl. Geochem.* **26**, S292–S294 (2011).

766 70. D. B. McNevin, *et al.*, Differences in carbon isotope discrimination of three variants of D-ribulose-
767 1,5-bisphosphate carboxylase/oxygenase reflect differences in their catalytic mechanisms. *J. Biol.*
768 *Chem.* **282**, 36068–36076 (2007).

769 71. E. B. Wilkes, A. Pearson, A general model for carbon isotopes in red-lineage phytoplankton:
770 Interplay between unidirectional processes and fractionation by RubisCO. *Geochim. Cosmochim.*
771 *Acta* (2019) https:/doi.org/10.1016/j.gca.2019.08.043.

772 72. P. M. Shih, *et al.*, Biochemical characterization of predicted Precambrian RuBisCO. *Nat. Commun.* **7**,
773 10382 (2016).

774 73. C. T. Supuran, Carbonic Anhydrases and Metabolism. *Metabolites* **8** (2018).

775