Title: Carbon isotope fractionation by an ancestral rubisco suggests biological proxies for CO_2 through geologic time should be re-evaluated

Authors: Renée Z. Wang*[1], Robert J. Nichols[2], Albert K. Liu[3,4], Avi I. Flamholz[5], Juliana Artier[6], Doug M. Banda[7], David F. Savage[2,8], John M. Eiler[1], Patrick M. Shih[3,6], Woodward W. Fischer[1]

Author affiliations: (1) California Institute of Technology, Division of Geological & Planetary Sciences. (2) University of California, Berkeley, Department of Molecular and Cell Biology. (3) Lawrence Berkeley National Lab, Joint Bioenergy Institute. (4) University of California, Davis, Biochemistry, Molecular, Cellular and Developmental Biology Graduate Group. (5) California Institute of Technology, Division of Biology and Biological Engineering. (6) University of California, Berkeley, Department of Plant and Microbial Biology. (7) University of California, Davis, Department of Plant Biology. (8) Howard Hughes Medical Institute, University of California, Berkeley, California 94720.

Corresponding author: *Renée Z. Wang
Email: rzwang@caltech.edu

ORCID:
R.Z.W.: 0000-0003-3994-3244
R.J.N.: 0000-0002-8476-0554
A.K.L: 0000-0001-9500-0449
A.I.F: 0000-0002-9278-5479
D.F.S: 0000-0003-0042-2257
J.A.: 0000-0001-5519-6627

Competing Interest Statement: Authors have no competing interests.

Classification: Major: Physical Sciences. Minor: Earth, Atmospheric, and Planetary Sciences.

Keywords: Evolution, Carbon Isotopes, Rubisco, Cyanobacteria, Precambrian

Abstract

247 words

The history of Earth's carbon cycle reflects trends in atmospheric composition convolved with the evolution of photosynthesis. Fortunately, key parts of the carbon cycle have been recorded in the carbon isotope ratios of sedimentary rocks. The dominant model used to interpret this record as a proxy for ancient atmospheric CO_2 is based on carbon isotope fractionations of modern photoautotrophs, and longstanding questions remain about how their evolution might have impacted the record. We tested the intersection of environment and evolution by measuring both biomass (ε_p) and enzymatic ($\varepsilon_{Rubisco}$) carbon isotope fractionations of a cyanobacterial strain (*Synechococcus elongatus* PCC 7942) solely expressing a putative ancestral Form 1B rubisco dating to ≫1 Ga. This strain, nicknamed ANC, grows in ambient pCO_2 and displays larger ε_p values than WT, despite having a much smaller $\varepsilon_{Rubisco}$ (17.23 ± 0.61‰ vs. 25.18 ± 0.31‰ respectively). Measuring both enzymatic and biomass fractionation revealed a surprising result—ANC ε_p exceeded ANC $\varepsilon_{Rubisco}$ in all conditions tested, contradicting prevailing models of cyanobacterial carbon isotope fractionation. However, these models were corrected by accounting for cyanobacterial physiology, notably the CO_2 concentrating mechanism (CCM). Our model suggested that additional fractionating processes like powered inorganic carbon uptake systems contribute to ε_p, and this effect is exacerbated in ANC. Understanding the evolution of rubisco and the CCM is therefore critical for interpreting the carbon isotope record. Large fluctuations in that record may reflect the evolving efficiency of carbon fixing metabolisms in addition to changes in atmospheric CO_2.

Significance Statement

116 words

Earth scientists rely on chemical fossils like the carbon isotope record to derive ancient atmospheric CO_2 concentrations, but interpretation of this record is calibrated using modern organisms. We tested this assumption by measuring the carbon isotope fractionation of a reconstructed ancestral rubisco enzyme (>1 billion years old) *in vivo* and *in vitro*. Our results contradicted prevailing models of carbon flow in Cyanobacteria, but our data could be rationalized if light-driven uptake of CO_2 is taken into account. Our study showed that the carbon isotope record tracks both the evolution of photosynthesis physiology as well as changes in atmospheric CO_2, highlighting the value of considering

both evolution and physiology for comparative biological approaches to understanding Earth's history.

Main Text

Introduction

Photoautotrophs have evolved over geologic time to harness energy from the sun in order to 'fix' external, inorganic carbon (C_i) into reduced, organic carbon (C_o), thereby creating biomass for growth. Today, and likely for much of Earth's history (1), the most widespread strategy for carbon fixation is the Calvin-Benson-Bassham (CBB) Cycle, where the key carbon fixation step is catalyzed by ribulose-1,5-bisphosphate (RuBP) carboxylase/oxygenase (rubisco) (2, 3). But rubisco's central role in the CBB cycle and oxygenic photosynthesis poses a conundrum because it is usually considered to be a non-specific and slow enzyme. The first issue concerns rubisco's dual carboxylase and oxygenase activities: the RuBP intermediate (enediolate) is susceptible to both O_2 and CO_2 attack (4). Consequently, instead of fixing a CO_2 molecule during photosynthesis, rubisco can instead assimilate O_2 to yield 2-phosphoglycolate (2-PG), which is not part of the CBB cycle and therefore must be salvaged through photorespiratory pathways that consume ATP, reducing power, and carbon (5). The second issue concerns rubisco's maximum carboxylation rate (V_C), which is ≈7-10 times slower than other central metabolic enzymes (6), and displays very limited variation across large phylogenetic distances (7).

Both issues—its dual carboxylase / oxygenase activity and limited maximum carboxylation rate—are typically rationalized by considering its evolutionary history in the context of long-term changes in environmental CO_2 and O_2 concentrations. Rubisco is thought to have evolved at a time when there was trace O_2 and much higher CO_2 concentrations in the atmosphere, in contrast to the modern atmosphere where O_2 is roughly 20% while CO_2 is only about 0.04% by partial pressure (1). In addition, rubisco is thought to have been the primary carboxylating enzyme of global photosynthesis since the Great Oxygenation Event, and potentially far prior (1).

Likely in response to these changing environmental concentrations, many aquatic photoautotrophs have evolved CO_2 concentrating mechanisms (CCMs) that concentrate CO_2 around rubisco in order to enhance carboxylation and suppress oxygenation. Even with CCMs, the effective *in vivo* rates of extant rubiscos are estimated to be much lower (≈1% for terrestrial and ≈15% for marine rubiscos) than their maximal catalytic rates observed in lab at 25°C, likely due to rubisco not working at night and the lower temperature of marine environments (2). However, all known Cyanobacteria today have CCMs, as do many bacterial chemolithoautotrophs, many aquatic algae, and some plants (8). The bacterial CCM has two main components: i) C_i pumps producing high cytosolic HCO_3^- concentrations, and ii) co-encapsulation of carbonic anhydrase (CA) and rubisco inside proteinaceous organelles known as carboxysomes (Figure 1A) (9–11). These powered C_i pumps include BCT1 (ATP-dependent

powered HCO_3^- transporter), SbtA (Na^+/HCO_3^- symporters), BicA (Na^+-dependent HCO_3^- transporter), NDH-1MS and NDH-1MS' (NADPH-dependent powered CO_2 uptake; see (12) for review). There are competing arguments in the literature for when the CCM evolved, ranging from the Proterozoic to the Phanerozoic Eon (8, 13). Therefore, for up to half of Earth's history, bacterial rubiscos have functioned in concert with a system that pumps C_i into and around the cell.

However, rationalization of rubisco's evolutionary history is highly dependent on our understanding of past environments, e.g. atmospheric CO_2 and O_2 concentrations. For the vast majority of Earth's history, we must rely on chemical fossils like the carbon isotope record for this purpose. The carbon isotope record is composed of measurements of the relative ratios of ^{13}C to ^{12}C isotopes in C-bearing phases of sedimentary rocks over time. Contemporaneous C_i pools are preserved as carbonate salts (like limestones and dolomites), while contemporaneous biomass and C_o pools are preserved in the organic-rich components (typically kerogen) of many different lithologies and are measured as rock total organic carbon (TOC) (14). The difference in C-isotope ratios between organic samples and an inorganic reference is typically reported using delta notation ($\delta^{13}C$) and expressed in per mil (‰, see Methods). Currently, carbon isotope data has been assembled globally from myriad environments to cover ≈3.8 billion years (Ga) of Earth's 4.5 Ga history; this data is viewed as a record of both inorganic and organic carbon cycle processes over geologic time (15).

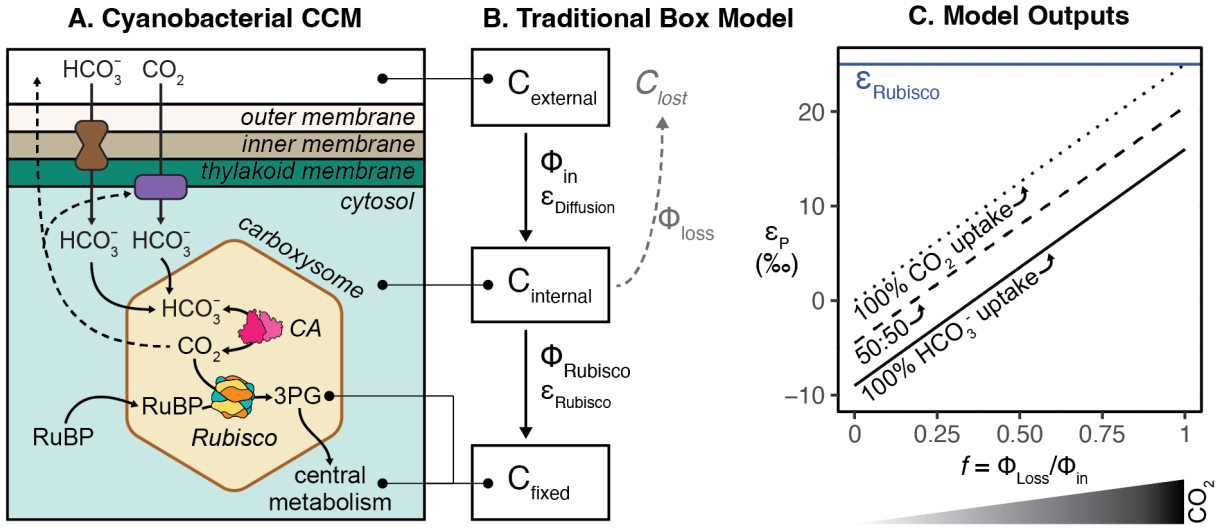


Figure 1: Comparing the Cyanobacterial CO_2 Concentrating Mechanism (CCM) to the traditional box model of photosynthetic C isotope discrimination. A) Cyanobacterial CCMs rely on i) active C_i uptake into the cell, and ii) co-encapsulation of carbonic anhydrase (CA) and rubisco within the carboxysome. Independent, powered transporters for HCO_3^- and CO_2 are shown in brown and purple; both work to increase cytosolic concentrations of HCO_3^- (see (12) for review). All CCM components work to produce a high carboxysomal CO_2 concentration that enhances CO_2 fixation by rubisco and suppresses oxygenation. Limited CO_2 escapes from the carboxysome – some is scavenged by CO_2 pumps while the rest leaves the cell. B) Architecture of the traditional box model based on (16–19); see Supplemental for full discussion of this model. Boxes denote carbon pools of interest, and fluxes between boxes are denoted by Φ. Each flux has its own isotopic fractionation denoted by ε; no fractionation is

assumed for Φ_{loss}. Model assumes an infinitely large external carbon pool, that carbon not fixed by rubisco (C_{lost}) returns to this pool, and that fluxes are at steady state. Note that this architecture does not include a box for the carboxysome. C) Model solution for the traditional model is $\varepsilon_P = a^*\varepsilon_{equil} + f^*\varepsilon_{Rubisco}$, where where ε_P is defined as the difference in $\delta^{13}C$ of $C_{external}$ and C_{fixed}, f is defined as the fraction of C_i lost (Φ_{loss}/Φ_{in}), and a is the fractional contribution of HCO_3^- to total C_i uptake. When a = 0, all C_i uptake is as CO_2 (dotted line); when a = 1, all C_i uptake is as HCO_3^- (solid line). This model is presented in (20), which is a generalization of (21) that accounts for the fact that C_i uptake (Φ_{in} in Panel B) ranges in composition between CO_2 and HCO_3^- based on which C_i uptake system is used. Values of $\varepsilon_{Rubisco}$ = 25‰ and ε_{equil} = -9‰ were used for this illustration (22). Model outputs indicate that at high external CO_2 concentrations (dark wedge under graph) there is excess CO_2 that rubisco cannot use, causing net C_i leakage (larger f values) from the cell.

The carbon isotope record is particularly important for constraining ancient atmospheric pCO_2 (23, 24) because direct observations of the past atmosphere from ice cores only extends back ≈1 million years (25). One notable feature of the record from ≈3.8 Ga to the present is that the $\delta^{13}C$ of C_o is depleted in ^{13}C by ≈25‰ compared to C_i (14, 15, 26), and this offset roughly matches the carbon isotopic fractionation of known carbon-fixing metabolisms and enzymes in the modern. (Note that the convention for reporting ε in this field is the opposite of other geochemistry fields – here, a negative value indicates a relative ^{13}C enrichment, in contrast to other fields where negative values mean ^{13}C depletion.) Rubisco displays a kinetic isotope effect (KIE) where it preferentially fixes $^{12}CO_2$ over $^{13}CO_2$ due in part to the V_C being slightly faster for $^{12}CO_2$ than $^{13}CO_2$ (27), which causes the reaction product, 3-phosphoglycerate (3-PG), to be relatively depleted in ^{13}C by several percent (tens of ‰) relative to the isotopic composition of the initial CO_2 substrate. The difference in $\delta^{13}C$ of the CO_2 substrate and the 3-PG product is typically reported as $\varepsilon_{Rubisco}$ and varies between 18-30‰ for several extant rubiscos (26, 28), with the exception of the coccolithophore *Emiliania huxleyi* with $\varepsilon_{Rubisco}$ ≈ 11‰ (29). Because all biomass is synthesized from 3-PG in autotrophs utilizing the CBB cycle, biomass is depleted in ^{13}C compared to external C_i pools and the magnitude of this difference is called ε_P. There is an additional fractionation factor associated with the preservation of biomass and C_i as rocks, so the magnitude of fractionation between C_i and C_o pools measured from the rock record is termed ε_{TOC} and varies slightly from ε_P (30). Therefore, if one can accurately derive ε_P from the rock record (ε_{TOC}) and pair it with some model relating ε_P to pCO_2, one could learn about both the evolution of photosynthetic physiology and abiotic changes in the carbon cycle over geologic time.

The dominant model used today for this purpose ("C Isotope Record Model," Figure S7) is:

$$\epsilon_p = \epsilon_f - \frac{b}{[CO_2(aq)]} \quad \text{(Equation 1)}$$

ε_f is the maximum possible isotopic fractionation of photosynthetic carbon fixation and is typically set to $\varepsilon_{Rubisco}$. The term b (‰ kg μM^{-1}) is empirically fit from pure culture experiments of eukaryotic and bacterial algae, and encompasses all physiologic effects that may affect isotopic fractionation like the CCM, growth rate, cell size and geometry, membrane permeability, growth media composition (e.g. pH, salinity, limiting nutrient), strain genetics and physiological state

(31–35). [CO_2(aq)] is the concentration of dissolved CO_2 in solution around the cells, and in the limit of high [CO_2(aq)], the term b/[CO_2(aq)] goes to zero and $\varepsilon_P = \varepsilon_f$, which is assumed to be $\varepsilon_{Rubisco}$. Therefore, the maximum value of ε_P is $\varepsilon_{Rubisco}$, and the term b sets how quickly ε_P approaches the limit of $\varepsilon_{Rubisco}$.

The C Isotope Record model has such a limit because the laboratory studies of plants and algae that it is based on ("Traditional model," Figure 1) shows such a limit. The traditional model was originally based on studies of carbon isotope fractionation in plants (dotted line in "Traditional model" in Figure 1C; all C_i uptake is as CO_2 for plants) and was later adapted to eukaryotic and bacterial algae. The primary architecture of this model stems from a seminal study by Park and Epstein (18) who proposed a "two step model" to explain ε_P of tomato plants grown in varied CO_2 concentrations and light levels. In this model, carbon can be viewed as residing in one of three pools, or "boxes" (Figure 1B) - C_i outside the cell (C_{ext}), C_i inside the cell ($C_{internal}$), or C_o as biomass (C_{fixed}). A "leakiness" term, f, is defined as the ratio of fluxes (Φ) of C_i exiting or entering the plant, where all of the C_i that entered the cell is lost when f=1. In this simplified model, ε_p is determined by the isotopic effect of two distinct steps: i) the diffusion of CO_2 into the plant ($\varepsilon_{Diffusion}$; <1‰ across a diaphragm cell in water at 25°C (36)); and ii) the carbon fixation step catalyzed by rubisco ($\varepsilon_{Rubisco}$; ≈18-30‰). Notably, Park and Epstein proposed that the isotopic fractionations of these two steps are not additive *in vivo* (i.e. $\varepsilon_p \neq \varepsilon_{Diffusion} + \varepsilon_{Rubisco}$) but instead reflects the process by which photosynthesis is limited, either diffusion of CO_2 into the cell ($\varepsilon_p = \varepsilon_{Diffusion}$) or CO_2 fixation by rubisco ($\varepsilon_p = \varepsilon_{Rubisco}$) (18).

This physiological interpretation results from the model solution, which is usually solved by assuming steady state and results in a linear relationship between ε_p and f where the minimum and maximum ε_p values are $\varepsilon_{Diffusion}$ and $\varepsilon_{Rubisco}$ respectively (Figure 1C). This allows experimentally measured values of ε_p to then be used to solve for CO_2 leakage (f, Figure 1C). Therefore, the corresponding physiological interpretations at the minimum and maximum model limits are when $\varepsilon_p \approx \varepsilon_{Diffusion}$, nearly all carbon entering the cell is used and with this mass balance constraint rubisco's ^{12}C preference is not "expressed"; conversely, when $\varepsilon_p \approx \varepsilon_{Rubisco}$, very little of the carbon entering the cell is fixed ($f \approx 1$, i.e. nearly all of the carbon leaks from the cell) and rubisco can "choose" between ^{12}C and ^{13}C substrates so that rubisco's KIE is fully expressed. Farquhar et al. (19) later derived a relationship between ε_p and the ratio of external vs. intracellular CO_2 partial pressures, allowing CO_2 concentrations at the site of rubisco to be estimated from ε_p. Therefore, given the assumption that C_i is taken up passively, it is possible to derive an increasing relationship between C_{ext} and ε_P from this model, where large ε_P suggests that high external CO_2 concentrations are creating excess CO_2 at rubisco and ultimately causing more CO_2 to leak out of the cell than can be fixed (see Supplemental and (17)).

This model was later adapted to algae to account for CCMs – mainly active uptake of C_i as HCO_3^- and/or CO_2 – and physiological parameters including growth rate and cell geometry (21, 31, 32, 37, 38). These studies grew eukaryotic and bacterial algae in a range of pCO_2 and culturing conditions to test if the linear relationship between ε_p and [CO_2] observed in plants still holds. Interestingly, cyanobacterial ε_p was found to be roughly constant independent of environmental pCO_2 and growth rate (31). In addition, because measured cyanobacterial ε_p

values were less than known cyanobacterial $\varepsilon_{Rubisco}$ values, additional isotopic fractionation factors were not needed to explain ε_p, even though some active C_i transport processes, which may fractionate carbon isotopes, were known in Cyanobacteria at the time (39–41). Therefore, though different versions of this model exist, all variations essentially modified the plant model by shifting the y-intercept of Fig. 1C to account for uptake of HCO_3^- in addition to CO_2. If C_i entering the cell is primarily CO_2, the model effectively represents plants (dotted line in Fig. 1C). If C_i entering the cell is primarily HCO_3^- through active uptake, as in many algae, all values are shifted to lower ε_p values (solid line in Fig. 1C) because of the equilibrium isotopic effect (ε_{equil}) between CO_2 and HCO_3^- (≈ -9‰ (22)). In Figure 1C, we plot the traditional model as derived in Eichner et al. (20), which is an adaptation of (21) :

$$\epsilon_P = f * \epsilon_{Rubisco} + a * \epsilon_{equil} \quad \text{(Equation 2)}$$

Here ε_{equil} is the equilibrium isotope effect, and *a* is the fraction of C_i entering the cell as CO_2 (*a*=0) or HCO_3^- (*a*=1); the diffusion isotope effect ($\varepsilon_{Diffusion}$) is considered negligible. See Supplementary Information for further discussion of the "traditional" model.

Overall, current models relating pCO_2 and autotrophic carbon isotope fractionation have a limit where ε_p cannot exceed $\varepsilon_{Rubisco}$ (Figure 1C). Yet, the largest ε_P values observed in the Archaean Eon exceed 30‰ (14, 15) and also exceed all current measurements of $\varepsilon_{Rubsico}$ (for recent compilation see (26)). In addition, recent studies in dinoflagellates have shown that ε_p can regularly exceed $\varepsilon_{Rubsico}$ under certain growth conditions (for review see (28)), and detailed studies of Cyanobacteria imply that leakage estimates derived from ε_p are not physiologically possible (20). These studies have therefore motivated updated models of carbon isotope fractionation in algae that account for the isotopic fractionations associated with different C_i uptake mechanisms (20, 28).

In addition to taking modern physiology into account, it is also important to understand how the evolution of rubisco and the CCM may have affected the carbon isotope composition of biomass and therefore $\delta^{13}C$ values of C_o preserved in the rock record. Recent studies have addressed this issue directly by testing model organisms that may better resemble an ancestral counterpart, including a cyanobacterial strain lacking a CCM (42), a cyanobacterial strain that overexpresses rubisco (43), and a cyanobacterial strain expressing an inferred ancestral rubisco dating from ≈1-3 Ga (44, 45).

Here, we measured the ε_p of a control strain of *Synechococcus elongatus* PCC 7942 expressing the wild-type rubisco (NS2-KanR, referred to as 'WT', see Methods), as well as a strain, 'ANC', engineered to express an inferred ancestral Form 1B enzyme (dating to >1 Ga) as its sole rubisco (46) in varied CO_2 and light conditions. This putative ancestral rubisco was previously purified and its kinetics were characterized *in vitro*. Its sequence was then inserted into the genome of a modern cyanobacterium, though the genome of the strain in that study contained both extant and ancestral rubisco sequences (46). Here we study a strain where the extant sequence has been fully deleted and replaced with the reconstructed ancestral sequence. In addition, we measured $\varepsilon_{Rubsico}$ of the present-day and ancestral rubiscos *in vitro*. We observed that: i) biomass ε_p is greater for ANC than for its WT counterpart for all conditions tested, even though ANC $\varepsilon_{Rubsico}$ (17.23 ± 0.61‰) is considerably less than WT $\varepsilon_{Rubsico}$ (25.18 ± 0.31‰); ii) ANC ε_p exceeds $\varepsilon_{Rubsico}$ in all tested conditions even though the traditional model sets the maximum possible $\varepsilon_p = \varepsilon_{Rubsico}$; iii) ANC ε_p increases with light levels while WT ε_p increases

with CO_2; iv) ANC displays a growth defect at ambient pCO_2 that is rescued at high pCO_2; and v) ANC growth is severely inhibited in high light. Consistent with recent studies of eukaryotic algae (20, 28), ANC ε_p exceeding $\varepsilon_{Rubsico}$ implies that the traditional box model is incomplete and additional isotopic fractionations are needed. In addition, modulation of ε_p with light suggests that aspects of cyanobacterial physiology beyond the CBB cycle must be taken into account to explain how ε_p can vary independently of CO_2. We posit additional factors related to C_i uptake that might explain fractionation measurements that deviate from box model predictions in both extant and ancient organisms.

Results & Discussion

Ancestral rubisco enzyme fractionates less than WT rubisco enzyme

We measured the carbon isotope fractionations of WT and ANC rubiscos *in vitro* using the substrate depletion method ((47–50); see Methods). The kinetics of this putative ancestral rubisco were previously characterized *in vitro* and are summarized in Table 1 (46). Previous work on rubisco isotope discrimination predicted that $\varepsilon_{Rubsico}$ should correlate positively with specificity ($S_{C/O}$), a unitless measure of the relative preference for CO_2 over O_2 (51). We therefore expected ANC and WT $\varepsilon_{Rubsico}$ values to be the same within uncertainty because of their similar $S_{C/O}$ values, but found that ANC $\varepsilon_{Rubsico}$ (17.23 ± 0.61‰) was about 8‰ less than WT $\varepsilon_{Rubsico}$ (25.18 ± 0.31‰, Table 1).

Rubisco	$\varepsilon_{Rubisco}$ (‰)	V_C (s^{-1})	K_C^{Air} (μM)	V_C/K_C^{Air} ($s^{-1}mM^{-1}$)	$S_{C/O}$
Ancestral Form 1B	17.23 ± 0.61	4.72 ± 0.14	168.7	28	49.6 ± 1.8
Modern Form 1B	25.18 ± 0.31*	9.78 ± 0.48*	184.1*	53.1*	50.3 ± 2.0*

Table 1: Rubisco characteristics. Starred values (*) for the modern Form 1B were measured in rubiscos purified from *Synechococcus* sp. PCC 6301, which has the same small and large subunit (*RbcS*, *RbcL*) sequences as our working WT strain, *Synechococcus* sp. PCC 7942 (46). Kinetic isotope effect ($\varepsilon_{Rubisco}$, avg. ± s.e.) was measured in this study using the substrate depletion method (47–50). Carboxylation turnover rate under substrate-saturated conditions (V_C); Michaelis constant for CO_2 in ambient levels of O_2 (K_C^{Air}); the catalytic efficiency towards CO_2 in ambient air (V_C/K_C^{Air}); and specificity, a unitless measure of the relative preference for CO_2 over O_2; ($S_{C/O}$) are from (46).

The ANC strain fractionates more than WT

Working in *S.elongatus* PCC 7942, we produced a mutant strain lacking the native Form 1B rubisco and expressing instead an ancestral Form 1B rubisco produced by computational ancestral sequence reconstruction (46) as its sole rubisco enzyme. We then grew this strain, termed ANC, and a control strain, termed wild-type or 'WT' (see Methods), in a variety of light and CO_2 levels: i) A reference condition (ambient pCO_2 of 0.04% v/v, standard light flux (120 µE)); ii) High CO_2 (5% pCO_2, 120 µE); iii) High light (0.04% pCO_2, 500 µE). The CO_2 gas at ambient and high CO_2 conditions had $\delta^{13}C$ values of -12.46‰ and -36.84‰ respectively.

Counter to expectations based on $\varepsilon_{Rubisco}$ (Table 1), ANC ε_p was as large or larger than WT ε_p in all conditions tested (Figure 2). This was consistent with recent results from a similar ancestral analog, where the ancestral analog's ε_p values exceeded WT in ambient and elevated CO_2 levels (44). In this study, the highest ANC ε_p values were observed for cultures grown in high light, where growth was comparatively slow (doubling time ≈ 50 hours, Figure 3 and Table S3). ANC ε_p values were also modulated by light and CO_2 differently than WT. Compared to the reference condition, WT ε_P values were indifferent to high light and only increased in high CO_2 (Figure 2A). In contrast, ANC ε_P values did not increase in high CO_2 and only increased in high light (Figure 2B). This result contrasted with the ancestral analogue in (44) where ε_P values increased by ≈10‰ at 2% CO_2.

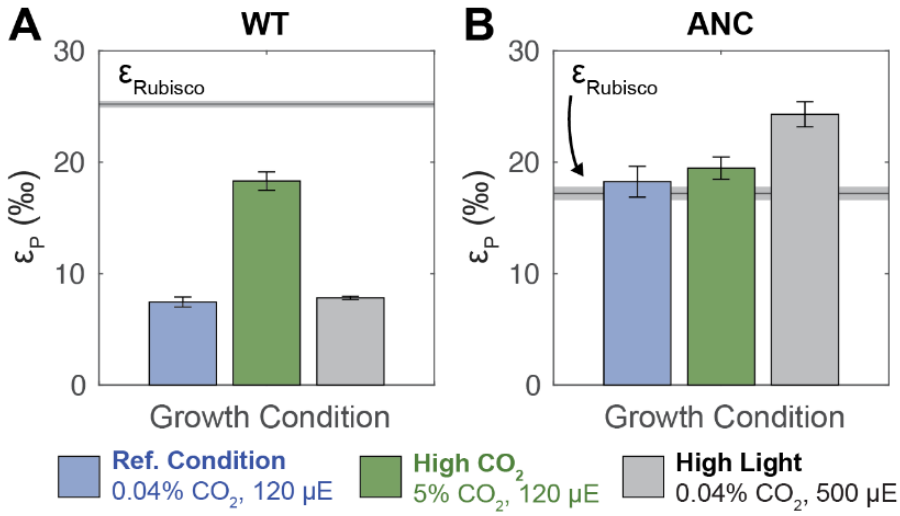


Figure 2: Whole cell carbon isotope fractionation by WT and ANC strains. ε_P (‰) values (avg. ± s.e.) for A) WT and B) ANC strains across growth conditions.. For each strain, the maximum ε_P possible based on the traditional model ($\varepsilon_P = \varepsilon_{Rubisco}$) is shown as a gray line (avg. ± s.e.). Most measured ANC ε_P values exceed the theoretical limit ($\varepsilon_P > \varepsilon_{Rubisco}$ + s.e.), while WT ε_P values do not ($\varepsilon_P < \varepsilon_{Rubisco}$). WT ε_P values increase in response to elevated CO_2 concentrations, while ANC ε_P values increase in response to elevated light flux. See Table S3 for full results.

In addition, the traditional box model described above cannot accommodate ε_p values in excess of $\varepsilon_{Rubsico}$ (Figure 1C). However, average ANC ε_P values exceeded ANC $\varepsilon_{Rubsico}$ in all growth conditions (Figure 2), particularly under high light conditions where the largest difference was seen (ε_p = 24.30 ± 0.12‰ vs $\varepsilon_{Rubsico}$ = 17.23 ± 0.61‰). The traditional box model also states that ε_P values are solely modulated by changing external pCO_2 concentrations, which cannot accommodate the ANC ε_P observations.

Ancestral rubisco strain grows at ambient CO_2 concentrations

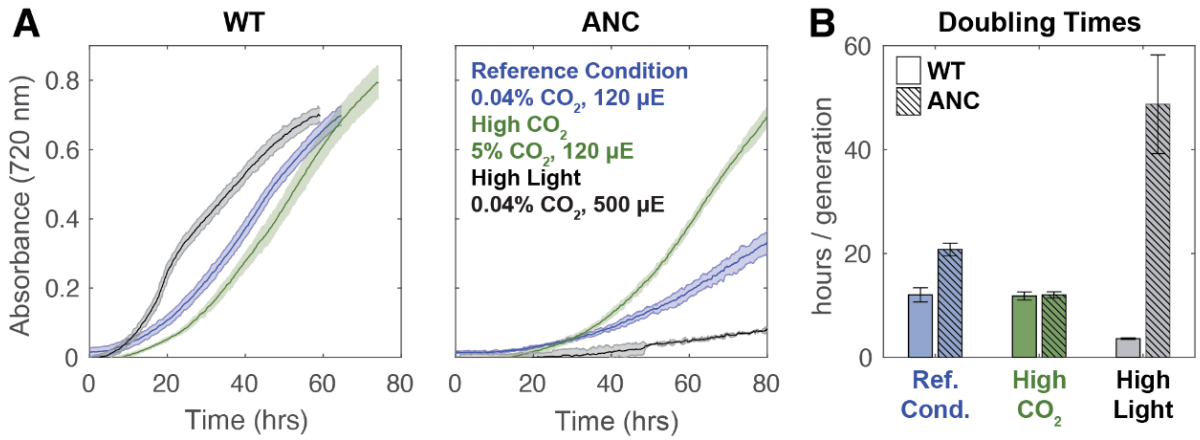


Figure 3: Growth curves for WT and ANC strains across experimental conditions. A) Averaged growth curves shown for WT and ANC strains to 80 hours, colored by growth condition as indicated in figure. Data was smoothed with a rolling median (Methods); see full ANC growth curves in Supplemental Fig. S12. B) Average doubling times with standard deviations. See Supplemental for details of doubling time calculation. ANC displayed a growth defect relative to the WT at the reference condition, which was rescued by high CO_2. ANC grew slowest in high light, while WT grew fastest in that condition.

Remarkably, the ANC strain managed to grow in ambient pCO_2 and standard light conditions (Figure 3A), even though the ancestral rubisco has a V_C roughly half that of WT (Table 1). This implies that its rubisco enzyme is properly encapsulated in the carboxysome, since it is well established that improper carboxysome formation greatly inhibits growth in ambient pCO_2 (52, 53). Mutant strains that are unable to form carboxysomes cannot grow in ambient air (53). Indeed, electron micrographs of WT and ANC cells grown in ambient CO_2 and light conditions (Methods) show multiple carboxysomes per cell in both strains (Figure 4 and Figure S13). Rubisco density can be seen within some of the carboxysomes (Figure 4C). In addition, the rubisco residues necessary for protein interactions mediating $\boldsymbol{\beta}$-carboxysome encapsulation were recently identified (54). We aligned WT and ANC rubisco sequences and found that fourteen of the sixteen residues are conserved in the ancestral sequence (Tables S8-9, Figure S14). In addition, WT and ANC strains harvested during exponential growth in the reference condition exhibit similar photosystem stoichiometry, as indicated by absorbance spectra (Figure S15). Taken together, these data indicated that carboxysomes form in ANC and the ancestral rubisco is encapsulated within these structures. Further strengthening our inference that the ancestral sequence is compatible with β-carboxysome formation, a similar ancestral analogue was also found to grow in ambient air (44).

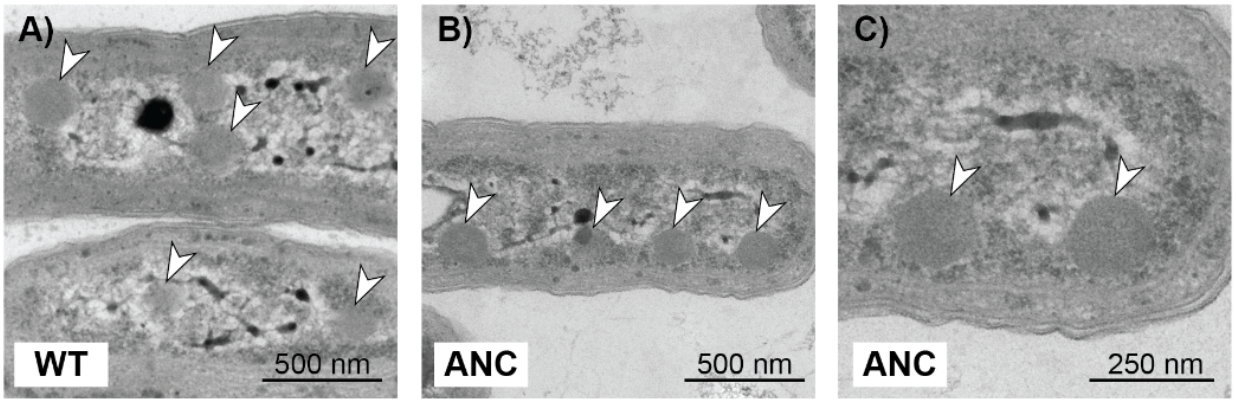


Figure 4: WT and ANC strains both produce carboxysomes at ambient pCO_2. Transmission electron micrographs of WT (A) and ANC (B,C) strains that were harvested during exponential growth in the reference condition (ambient pCO_2, standard light flux). Both strains show multiple carboxysomes per cell, as indicated by white arrows, and carboxysomes exhibit the typical hexagonal shape (53). C) is the same image as in B) but magnified to show that rubisco density seen can be within the carboxysomes of ANC. The dark internal body in A) is likely a polyphosphate body (55). See Figure S13 for additional images.

In addition, the difference in V_C between the ancestral and modern rubiscos was mirrored in the doubling times of WT and ANC strains (Figure 3B, Table S3), where ANC doubling times were roughly twice that of WT in the reference condition (20.8 ± 1.2 vs. 12.0 ± 1.4 hours respectively). In addition, the carboxylation efficiency in ambient air (V_C/K_C^{Air}) for the ancestral Form 1B rubisco, which measures the enzyme's ability to carboxylate in conditions with low CO_2 and relatively high O_2, is roughly half that of the modern Form 1B rubisco as well (Table 1). This suggested that ANC's growth was limited by its ability to fix CO_2 from ambient air. This growth defect was ameliorated by high pCO_2, where doubling times for both strains were the same within uncertainty (WT 11.8 ± 0.8 hours; ANC 12.0 ± 0.6 hours), though we observed a longer lag phase for ANC. WT doubling times were the same within uncertainty for the reference and high CO_2 conditions (12.0 ± 1.4 vs. 11.8 ± 0.8 hours respectively), consistent with previous studies where increased pCO_2 did not affect growth rate (56). In contrast, elevated CO_2 greatly accelerated the growth of ANC, reducing its doubling time from ≈21 to ≈12 hrs, supporting our inference that CO_2 availability is limiting the growth of ANC and implicating the CCM in its growth defect. Consistent with our results, a similar ancestral Form 1B analogue displayed total carboxylase activity roughly half that of the modern Form 1B (57).

We observed the greatest differences in doubling times between ANC and WT when the strains were grown at high light (Figure 3, Table S3). In these conditions, WT cultures were a dark, blue-green color typical of healthy cyanobacterial cells while ANC cultures were yellow-green (Fig. S11), suggesting degradation of phycobilisomes via a known starvation pathway to reduce the cell's capacity for light harvesting and photochemical electron transport (58). We therefore infer that ANC could not fix CO_2 at a rate matching its light harvesting capability, and hence invoked this regulatory pathway to decrease light harvesting capacity. WT, in contrast, grew rapidly in the high light condition.

385 **Proposed influence of a light-powered, vectoral carbonic-anhydrase**

386 As discussed above, recent studies in extant bacterial and eukaryotic algae have shown
387 that ε_p can regularly exceed $\varepsilon_{Rubsico}$ under certain growth conditions (for review see (28)),
388 motivating updated models of carbon isotope fractionation in both eukaryotic and bacterial algae
389 (20, 28, 59). These models suggested that ε_P values could only be explained if another enzyme
390 acting as a CA catalyzing an energy-coupled vectoral hydration of intracellular CO_2 to HCO_3^-
391 was taken into account, since this reaction is calculated to have a large isotopic effect and
392 would therefore allow ε_p to exceed $\varepsilon_{Rubisco}$ (20, 28, 59). Though the cell does not "know" where
393 CO_2 in the cytosol came from, these models primarily invoke such an enzyme for internal C_i
394 recycling, where CO_2 lost from the carboxysome could be converted to HCO_3^- so that it could
395 remain in the cell (20, 28, 59). However, energy-coupled CAs can also serve as CO_2 uptake
396 systems by converting extracellular CO_2 that passively translocates the membrane to
397 intracellular HCO_3^- (Figure 1A), which is advantageous in conditions (e.g. acidic pH) where CO_2
398 is the dominant form of extracellular C_i (10, 60, 61).

399 In general, Cyanobacteria have been shown to have two modes of active C_i uptake:
400 uptake of hydrated C_i (predominantly H_2CO_3 and HCO_3^-) and uptake of CO_2 (61). In order for the
401 CCM to function, either mode would need to produce a high, non-equilibrium concentration of
402 HCO_3^- in the cytoplasm (8, 10). This is thought to be achieved by coupling CA to an energy
403 source (e.g. light or an ion gradient) that drives the vectoral hydration of CO_2 to HCO_3^- in the
404 cytoplasm (62). There is now excellent data supporting this hypothesis in Cyanobacteria, where
405 accessory proteins that bind to the NDH complex, the cyanobacterial homolog of the respiratory
406 Complex I NADH-dehydrogenase, are known to mediate CO_2 uptake specifically (63–65).
407 Additionally, one of these accessory proteins, CupA/B, is reminiscent of a CA and contains a
408 telltale zinc active site situated near a proton channel in a membrane subunit (66). The
409 prevailing understanding of these data is, therefore, that these complexes couple inorganic
410 carbon uptake to energy supplied by photochemical electron transport. Moreover, a similar
411 protein complex has been described in proteobacterial chemoautotrophs, suggesting that
412 energy-coupled CO_2 hydration is widespread (60).

413 A vectoral CA would affect ε_p for two reasons. First, CO_2 and HCO_3^- are isotopically
414 distinct. At equilibrium in standard conditions, HCO_3^- is ≈8‰ more enriched in ^{13}C than CO_2 (67,
415 68). Therefore, if a cyanobacterium is predominantly taking up CO_2, the internal C_i pool from
416 which biomass is formed would be isotopically lighter (^{13}C-depleted) than if HCO_3^- is the
417 dominant source of C_i. Second, unidirectional CO_2 hydration is expected to impart a substantial
418 isotope effect, with calculated values ranging from ≈19 to 32‰ (67, 69–72). Therefore, there are
419 two mechanistic reasons that ε_P could exceed $\varepsilon_{Rubisco}$ in conditions where energized CO_2 uptake
420 and hydration is active. Indeed, a recent model of C-isotope fractionation in Cyanobacteria
421 specifically invoked the NDH complex to rationalize ε_p values that exceed $\varepsilon_{Rubisco}$ (20).

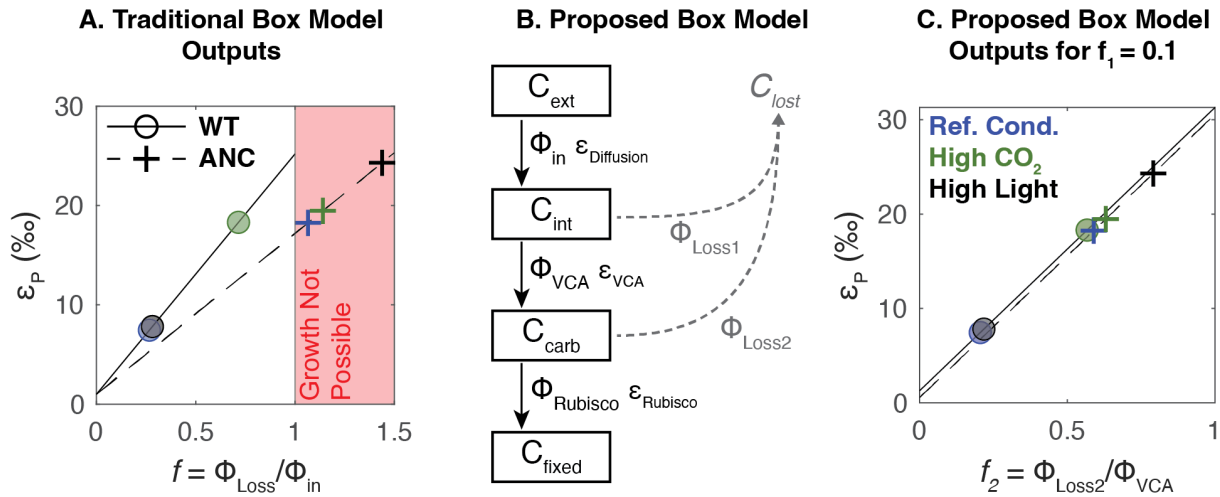
422 Because this energy-coupled CO_2 uptake and hydration by the NDH complex is driven
423 by light energy, e.g. via cyclic electron flow around photosystem I (66), and because the
424 vectoral hydration of CO_2 to HCO_3^- is thought to have a large carbon isotope fractionation, we
425 hypothesized that ε_p would increase with light intensity. Indeed, we observed the largest ANC
426 ε_P values, far exceeding ANC $\varepsilon_{Rubisco}$, in the high light condition and found that ANC ε_P varies

427 primarily with light and not CO_2 (Figure 3). This observation is counter to the traditional model
428 which proposes ε_P as a direct correlate of external pCO_2 (16, 17). Furthermore, on short
429 timescales (≈minutes) cyanobacterial C_i uptake can be modulated by light intensity alone, fully
430 independent of external C_i concentrations (73), and CO_2 uptake can occur in the absence of
431 carbon fixation (74, 75). Based on these physiological and isotopic observations, our study also
432 supports the hypothesis that a powered, vectoral CA like the NDH complex is likely active in
433 Cyanobacteria, and is likely responsible for $\varepsilon_p > \varepsilon_{Rubisco}$ in ANC.

434 **Proposed model for carbon isotope fractionation in Cyanobacteria**

435 As discussed above, the traditional box model cannot produce $\varepsilon_p > \varepsilon_{rubisco}$ (Figure 1C). In
436 this model, the C_i leakage term (*f*) is fit from measured ε_p values and $f = 1$ implies that all carbon
437 uptake leaks out of the cell. Though the traditional box model can accommodate both CO_2 and
438 HCO_3^- uptake, which differ in their equilibrium isotopic composition, even modeling 100% CO_2
439 uptake gave $f > 1$ for ANC in all conditions (Figure 5A and S8). Yet, ANC grew reproducibly in
440 all conditions tested (Figure 3). We also encountered challenges using the traditional model to
441 rationalize WT data: fitting the model gave $f < 1$ in ambient pCO_2 conditions, but high CO_2
442 conditions yielded $f > 1$ unless all C_i uptake was assumed to be as HCO_3^- (see Figure S8 and
443 Supplemental Information for discussion). Therefore, to rationalize our results for both WT and
444 ANC, we developed a box model that represents a small modification of the traditional model.
445 We modified the traditional model (Figure 1B,C) by adding an additional isotopic
446 fractionation so that ε_p can exceed $\varepsilon_{rubisco}$. As discussed above, we hypothesize that this
447 additional fractionation is due to a vectoral, powered CA like the NDH complex. In the modified
448 model, we distinguish between carbon in the cytosol (C_{int}) and carbon in the carboxysome
449 (C_{carb}), and allow a flux for carbon to be lost from the carboxysome (Φ_{Loss2}, Figure 5B).
450 Therefore, external C_i enters the cell (flux Φ_{in}) where it can either leak out (flux Φ_{Loss1}) or
451 undergo active hydration (flux Φ_{VCA}, where *VCA* denotes *V*ectoral *C*arbonic *A*nhydrase).
452 Intracellular C_i can then enter the carboxysome, where it is either fixed (flux $\Phi_{Rubisco}$) or
453 ultimately leaks out of the cell (flux Φ_{Loss2}).
454 We made similar simplifying assumptions as the traditional box model: i) an infinite
455 supply of external carbon, ii) no isotopic fractionation for carbon lost from the cell, iii) Φ_{in} has the
456 isotopic fractionation associated with $\varepsilon_{Diffusion}$, and iv) the system is at steady state. We did not
457 add an explicit term for light energy used to power C_i uptake. Instead, the model included an
458 energized CA (denoted VCA) and its associated isotopic fractionation as free parameters. In
459 modeling each strain, we used the appropriate $\varepsilon_{Rubisco}$ measurements (Table 1). We do not know
460 the true value for ε_{VCA}, but used a value of 30‰ similar to a recent model that explicitly invoked
461 the NDH complex in Cyanobacteria (20). For comparison with the traditional model, we plotted
462 Figure 5C with $f_1 = 0.1$ so that it could be represented in two dimensions; see Figure S10 for full
463 model outputs. In this updated model, each value of ε_p corresponds to a set of feasible f_1 and f_2
464 values that fall along a line (Figure S10). Therefore, our model constrains but does not uniquely
465 determine f_1 and f_2, nor does it allow for estimation of external C_i levels because many pairs of f_1
466 and f_2 values can produce the same ε_p. In addition, we focused only on C_i uptake as CO_2
467 because we are interested in a model that could achieve large ε_p values (indicating ^{13}C-depleted
468 biomass) to account for at least an additional ~8‰ of fractionation in ε_p (maximum of ~25‰ in

469 the high light condition) *greater* than $\varepsilon_{Rubsico}$ (~17‰) in ANC. HCO_3^- uptake through bicarbonate
470 pumps would not achieve this effect because it would shift all ε_p values to be 9‰ more negative
471 (^{13}C-enriched biomass, see Figure 1C).
472



473
474 **Figure 5: Proposed box model based on experimental results.** A) Experimental results (circles and
475 crosses) plotted onto traditional box model outputs (dashed and solid lines) for WT and ANC respectively
476 if C_i uptake is all CO_2. Uncertainties are smaller than data points. Colors indicate growth conditions: blue
477 = reference condition (0.05% pCO_2 (v/v), 120 μE); green = high CO_2 (5% pCO_2 (v/v), 120 μE); black =
478 high light (0.05% pCO_2 (v/v), 500 μE). *f* is as defined in Figure 1; region where *f* > 1 is shaded in red. B)
479 Proposed box model architecture, with main carbon pools of interest in boxes. Subscripts indicate
480 external (*ext*), internal (*int*), carboxysome (*carb*), and fixed (*fixed*) carbon pools. Fluxes are denoted by Φ
481 where subscripts indicate fluxes into the cell (*in*), out of the cell (*Loss1*, *Loss2*), into the carboxysome
482 (*VCA* for *V*ectoral *C*arbonic *A*nhydrase), and into fixed biomass *(Rubisco*), each with a corresponding
483 isotopic fractionation denoted with ε. Loss fluxes were assumed to have no isotopic fractionation. In this
484 proposed model, f_1 is defined as Φ_{Loss1}/Φ_{in}, and f_2 is defined as Φ_{Loss2}/Φ_{VCA}. See text for model
485 assumptions. C) Experimental results plotted onto proposed box model outputs for f_1 = 0.1; colors and
486 symbols are the same as Panel A. ε_P is defined as the difference in $\delta^{13}C$ between C_{ext} and C_{fixed}. See
487 Supplemental Figure S10 for full results; only results for f_1 = 0.1 are shown.
488
489 With the addition of a powered, vectoral CA and an additional loss term, the model was
490 able to rationalize our experimental data of $\varepsilon_p > \varepsilon_{Rubisco}$ with leakage values compatible with cell
491 growth (f_2 < 1, Figure 5C). In addition, it helped us understand why the high light condition gave
492 such varied results between WT and ANC. Our model results implied that ANC lost more carbon
493 than WT at the branch point before rubisco (Φ_{Loss2}); i.e. even though carbon was present in the
494 cell, it could not be fixed by the ancestral Form 1B rubisco, perhaps because of its lower V_C
495 (Table 1). Excess CO_2 allowed rubisco's KIE ($\varepsilon_{Rubisco}$) to be expressed in ε_P. These results
496 implied that, in high light, the powered CA was delivering high amounts of CO_2 to both the WT
497 and ANC rubiscos. The faster WT rubisco was able to match this flux, which was reflected in its
498 fast growth rate (Figure 3) and no change in ε_p vs. the reference condition (Figure 2). However,
499 the slower ANC rubisco was not, which led to its slowest growth rate (Figure 2), and highest ε_p
500 values across all conditions. Conditions where ε_p exceeded $\varepsilon_{Rubisco}$ in ANC suggested that, in

addition to Φ_{Loss2} being large (allowing $\varepsilon_{Rubisco}$ to be expressed), Φ_{Loss1} was high as well, which allows ε_{VCA} to be expressed. That is, the slower ANC rubisco created a "backup" where leakage increased all along the carbon fixing pathway, and this effect was exaggerated at high light.

In addition, we fit our data to other models that are aware of active C_i uptake as part of general algal (21) or cyanobacterial (20, 59) CCMs (Figure S9). Cyanobacterial models that incorporated an explicitly one-way, "CA-like" enzyme (59) or the NDH complex specifically (20) were mostly able to rationalize our data as well. The poorest model fits for our data were when C_i uptake was mostly as HCO_3^- (Figure S9). Overall, our model and theirs (20, 59) show that adding an additional carbon isotope fractionation step produces a model capable of rationalizing our data by enabling $\varepsilon_p > \varepsilon_{Rubisco}$ with leakage values less than 1.

We also note that our use of the term "vectoral" CO_2 hydration connotes a net flux that is dominantly in the direction of CO_2 hydration, rather than implying that the flux of HCO_3^- dehydration is zero. As such, there is likely some bidirectional activity of the NDH complex. It is difficult to experimentally measure the isotope effect associated with the hydration reaction ($CO_2 \Rightarrow HCO_3^-$), but transition state theory and quantum chemical modeling (67, 68, 71) suggest that the value is large (roughly 25‰, see (28) for review). HCO_3^- dehydration, and equilibration in general, would tend to reduce the isotopic fractionation (67). Our results here do not require a larger isotopic effect, however. Rather, a smaller value of ≈10‰ would have allowed us to rationalize our measurements, as we need only account for an additional ≈8‰ of fractionation in ε_p (maximum of ≈25‰) above $\varepsilon_{Rubsico}$ (≈17‰) in ANC.

Overall, our measurements and analyses indicated that, in addition to rubisco, processes relevant to the CCM can play an important role in setting ε_p values. While our model is highly idealized and relies on a minimum set of fractionating processes associated with carbon fixation in Cyanobacteria, i.e. adding only one additional fractionation factor and one additional leakage point, the results demonstrated that a simple addition to the traditional model accounting for a known mode of energized CO_2 uptake can explain our experimental results. Moreover, one useful implication of this model is that carbon isotope values may measure the efficiency of the CCM and carbon fixation in Cyanobacteria, in addition to ambient environmental CO_2 concentrations.

Consequences for understanding the evolution of carbon-fixing metabolisms

The goal of this study was to test if prevailing models of carbon fixation and isotopic fractionation held up in an ancestral analogue strain that may be more relevant to understanding the carbon cycle over geologic time. We did so by measuring the isotopic fractionation of a reconstructed ancestral rubisco both inside and outside a living Cyanobacterium. We emphasize that ANC is not a true ancestral Cyanobacteria; rather it is a chimeric construct—a modern strain saddled with a predicted Precambrian enzyme. This reconstructed ancestral rubisco is characterized by slower carboxylation kinetics (46) and a much lower $\varepsilon_{rubisco}$ than the modern strain's native enzyme (Table 1).

Recent studies in extant bacterial (20) and eukaryotic algae (for review see (28)) have already motivated updated models of C isotope fractionation in cells; these models address

observations that: i) ε_p can exceed $\varepsilon_{Rubisco}$ in certain conditions; ii) factors other than pCO_2 can modulate ε_p. We observed similar phenomena in our ANC strain, where ε_p exceeded $\varepsilon_{Rubisco}$ in all conditions tested, and increased light intensity led to greater ε_p values. To date, such anomalous ε_p values have been observed during relatively slow growth; in (76) $\varepsilon_p > \varepsilon_{Rubisco}$ occurred early in the growth curve as cells were acclimating to fresh culture media, in (28) $\varepsilon_p > \varepsilon_{Rubisco}$ occurred during nitrogen and phosphorus limitation, and in this study $\varepsilon_p > \varepsilon_{Rubisco}$ was observed in a mutant strain growing slowly while expressing a reconstructed ancestral rubisco. These observations indicated that growth physiology affects isotopic fractionation by photosynthetic algae and, in all cases, motivated a rethinking of the traditional box model (Figure 1B,C) to include more physiological detail relating to the presence of a CCM.

Here we observed $\varepsilon_p > \varepsilon_{Rubisco}$ in all growth conditions for ANC, and especially in high light (Figure 2). As high light consistently slowed growth, induced chlorosis (yellowing of cultures, Figure S11) and increased ε_p, we were motivated to consider the effects of light-related physiology on ε_p. The yellowing of ANC cultures in high light was consistent with the well-described phycobilisome degradation pathway, which is typically induced in nutrient starvation conditions and taken to indicate that light levels exceeded the downstream capacity for CO_2 fixation (58, 77). We interpreted these observations as indicating that the replacement of the native rubisco with a reconstructed ancestor decreased the cellular capacity for CO_2 fixation, potentially due to the inferior carboxylation rate of the ancestral enzyme (Table 1).

Low CO_2 fixation capacity would not, on its own, explain anomalously high ε_p values, however. An additional fractionating process is required to explain ε_p values in excess of $\varepsilon_{Rubisco}$, which we assumed is due to light-coupled vectoral hydration of CO_2, which has a large calculated isotope effect (67, 69–72). Cyanobacteria have been shown to take up CO_2 independently of HCO_3^- (61). In model Cyanobacteria, this activity is due to the Cup proteins (CupAS/B, also known as Chp proteins), which bind to the NDH complex of Cyanobacteria (66, 78). The NDH complex is involved in light energy capture via photosynthetic electron transport and cyclic electron flow around photosystem I (66) and, moreover, CO_2 uptake is stimulated by light alone and abrogated by inhibitors of photochemical electron transport (73). Not only has CupA been shown to carry a key Zn^{2+} in a domain resembling a carbonic anhydrase (66), but the *cupA* gene is induced under low CO_2 conditions (78). In order for CO_2 uptake to drive the CCM and promote CO_2 fixation, it would need to produce a high, non-equilibrium HCO_3^- concentration in the cytoplasm (8, 10). We and others therefore assumed that the complex of NDH-1 and CupAS/B couples light energy to the one-way hydration of CO_2 to HCO_3^- at a carbonic anhydrase-like active site (66).

It is readily apparent from our experiments that $\varepsilon_{Rubisco}$ does not set an upper bound on ε_p, nor does it predict which strains will have larger ε_p values *in vivo* (Figure 2). This inference was only possible because we measured the isotope fractionation due to the ancestral rubisco ($\varepsilon_{Rubisco}$) and compared it to ANC strain biomass (ε_p), in contrast with the study of (57), which measured ε_p but not $\varepsilon_{Rubisco}$. While our ANC ε_p values (≈18-24‰) fell within the range of ε_p values derived from the carbon isotope record (42), they exceeded its measured $\varepsilon_{Rubisco}$ (Figure 2). As such, the relative consistency of ANC ε_p values with extant ε_p values does not indicate that the traditional box model is applicable across geologic time as claimed in (57). Rather, a model including some additional fractionating process is required to explain our observation that $\varepsilon_p > \varepsilon_{Rubisco}$ in ANC. Attention has been paid to outliers in the carbon isotope record where ε_p

exceeds $\varepsilon_{Rubisco}$ precisely because they violate the assumptions underlying the dominant model (Equation 1) used to interpret the record (28). In addition, ANC $\varepsilon_{Rubsico}$ (17.23 ± 0.61‰) is anomalously low; not only is it ≈8‰ less than WT $\varepsilon_{Rubsico}$ (25.18 ± 0.31‰) but it is among the lowest measured rubisco KIEs. However, only thirteen unique rubisco KIEs have been measured thus far (for recent review see (26)) while ≈300 distinct rubiscos have been kinetically characterized (7, 79), suggesting that measuring the isotopic effects of several well-chosen rubisco variants is worthwhile.

Our study in an ancestral analogue strain suggests that the carbon isotopic fractionations observed in both modern environments and in the geological record reflect not just the environmental abundance of CO_2 and/or the rubisco present, but also the operation of C_i uptake processes, like the NDH complex, that operate as part of the CCM. Both these processes can lead to a highly variable range of carbon isotope fractionations. Our study supports the conclusion of prior studies that a carbon isotope model that engages more fully with photosynthetic physiology, like the CCM, is required to describe ε_p values and more accurately constrain environmental CO_2 concentrations from environmental context (e.g. light and nutrient levels) and physiological parameters (e.g. $\varepsilon_{Rubisco}$, photosynthetic capacity, growth rate). The model proposed here was written to mathematically validate a hypothesis – that ε_p can only exceed $\varepsilon_{Rubisco}$ if another fractionating process was considered. In addition, it represents only a first step in this direction as it substantially simplifies the bacterial CCM (10); a similar statement applies to box models of Eukaryotic algae which also express complex CCMs (28, 80). Future work on carbon isotope fractionation by Cyanobacteria should grapple in more detail with photosynthetic physiology, including uptake of CO_2 and HCO_3^- by independent systems, integration of both light and dark reactions, and effects of nutrient limitation on growth rate. As mechanistic biochemical understanding of cyanobacterial C_i uptake improves (66), it may also become feasible to directly measure or better constrain the isotopic fractionation associated with these processes. Coupling such a model with experiments in natural and engineered organisms will help validate these models and improve our ability to understand environmental and evolutionary changes of the carbon cycle over Earth history.

In addition, this study and other recent work (42, 57) have raised a greater question for the Earth Sciences: what is uniformitarianism for biology? Earth scientists often apply uniformitarian assumptions – assuming that physical and chemical processes behave the same now as they did billions of years ago – in order to reason about the past. This approach is powerful, but these assumptions are challenged by biological processes that undergo substantial evolution on geologic timescales. Ongoing discoveries of novel metabolisms have supported some principles like ‘the principle of microbial infallibility’ – that microbes will always find a way to take advantage of available energy sources (for recent review see 81) – but it is not clear what principles apply to the details of metabolism. Take rubisco, for example – most extant autotrophs use rubisco to fix carbon, but rubisco sits within a variety of broader metabolisms (i.e. C3, C4, CAM in plants) that temper the effect of $\varepsilon_{Rubisco}$ on ε_p (for recent review see (26)). We are far from having a clear answer to this question, but recent work at the interface of molecular biology and isotope geochemistry show that these ideas can be tested in the lab. Here and in other recent papers (42, 57, 82), we used synthetic biology to construct

organisms with ancestral components so that specific aspects of ancient organisms can be isolated and tested. These “ancestral-like” organisms helped sharpen our understanding of the physiological and environmental factors determining growth (82) and isotopic fractionation (this work) in both ancient and modern autotrophs, and showed that models rigidly based on modern taxa are likely not universally applicable across geologic timescales.

Overall, carbon fixation was a fundamental challenge that autotrophs overcame early in the history of Earth’s biosphere (1). These early processes were recorded in some fashion in the carbon isotope record, but robust interpretation of this record must grapple with the fact that the carbon cycle is an amalgam of both environmental changes and evolutionary processes, mediated by physiology. We now have synthetic biological approaches that offer a way to probe these long timescale co-evolutionary problems by producing ancient process analogs of carbon fixation in the laboratory.

Materials and Methods

Ancestral enzyme reconstruction

Ancestral Rubisco enzyme sequences were previously reported and characterized by Shih et al. (2016) (46). Briefly, for both the large subunit (LSU) and small subunit (SSU) of Rubisco, encoded by *rbcL* and *rbcS* respectively, the most recent common ancestor (MRCA) for Form 1A (α), 1B (β), and 1A/B (α/ β) clades were predicted from independently derived phylogenetic trees for RbcL and RbcS containing a broad diversity of Form 1A and 1B Rubisco (>100 sequences). Maximum-likelihood algorithms were used to reconstruct the most probable ancestral sequence for each clade. Ancestral sequences were then expressed in *Escherichia coli* and purified, and enzyme kinetics were measured.

ANC strain generation

The ‘ANC’ strain studied here was generated by replacing the native large and small Rubisco subunits (cbbL and cbbS respectively) of the parent strain (*Synechococcus elongatus* PCC 7942) with the reconstructed β ancestral cbbL and cbbS sequences. The NS2-KanR (‘WT’ strain) was generated by inserting a KanR cassette into neutral site 2 (NS2) (GenBank: U44761.1). This was done as a control for having the KanR in the neutral site. *Synechococcus elongatus* PCC 7942 were transformed from the WT strain using the approach of Golden and Sherman (1984) (83). Briefly, cultures were grown to OD750nm = 0.5. Cultures were centrifuged at 18,000 x *g* for 2 minutes. Pellets were washed with 100 mM $CaCl_2$ and spun again at 18,000 x *g* for 2 minutes. Pellets were resuspended in BG-11 media followed by addition of plasmid and grown for 16 hours in the dark at 30°C. Transformants were then plated onto BG-11 + KAN100 agar plates and placed under 100 µE of light at 30°C. Single colonies were then genotyped by PCR amplification of the Rubisco locus followed by sequencing. Table S1 lists plasmids and primers used in this study.

Growth conditions

For ambient CO_2 growth, NS2-KanR and β Ancestral Rubisco-KanR strains were grown in quadruplicate in a photobioreactor (Photon Systems Instruments - MC 1000) at the University of California, Berkeley (UC Berkeley) for four biological replicates total. Cultures were grown in

669 buffered BG-11 media with 50mM HEPES at pH 8. Cultures were inoculated at a starting
670 OD720nm = 0.015 and cultivated at 120 µE, 30°C, and bubbled with ambient air. High CO_2
671 growth was performed using the same conditions as ambient growth with the exception of
672 placing the photobioreactor in a 5% CO_2 chamber (Percival AR22L) and bubbling in air from the
673 chamber. High light growth was performed using the ambient conditions above with the
674 exception of using 500 µE for light intensity. Cells were harvested by centrifugation at 6000 x *g*
675 for 20 minutes at 4°C. Decanted pellets were then flash frozen with liquid N_2 and lyophilized
676 overnight with the Millrock Technology Model BT85A freeze dryer. Doubling time was calculated
677 by fitting the exponential phase of growth (*k*) using a Markov Chain Monte Carlo (MCMC)
678 approach, using the generic model y = a*EXP(k*x)+b. Growth curves displayed in Figure 3 were
679 smoothed with a rolling median (*n* = 12) to remove errant readings caused by bubbles advected
680 in front of the detector. See Supplemental for more information.
681
682 **Carbon isotope analysis**
683 Carbon isotope data is reported using delta notation ($\delta^{13}C$) in units of per mille (‰) where $\delta^{13}C$
684 = $[(^{13}C/^{12}C)_{sa}/(^{13}C/^{12}C)_{ref}-1]*1000$, where the subscripts 'sa' and 'ref' denote sample and
685 reference respectively. The reference used is the Vienna Pee Dee Belemnite (VPDB). $\delta^{13}C$
686 values of cyanobacterial cells were measured on an EA-IRMS (Elemental Analyzer Isotope
687 Ratio Mass Spectrometer; Costech Thermo Delta-V) at the California Institute of Technology
688 (Caltech) in Pasadena, CA. Each biological replicate was run four times with two different
689 isotope standards – urea (-27.8‰) and sucrose (-10.45‰). A suite of urea and sucrose
690 standards were run at the beginning, middle, and end of run for sample bracketing and to
691 assess drift throughout the run. An average $\delta^{13}C$ and standard error were calculated and
692 reported for each biological replicate (see Supplemental for more information). The $\delta^{13}C$ of the
693 starting CO_2 gas was measured on the Thermo Mat 253 Ultra at Caltech; the CALT-2049C
694 standard was used, which has a $\delta^{13}C_{VPDB}$ value of -3.62‰. CO_2 gas from high pCO_2
695 experiments was sourced from a CO_2 tank, while the CO_2 gas in ambient pCO_2 experiments
696 was distilled from ambient lab air through cryogenic distillation at Caltech. ε_p, the carbon isotope
697 fractionation between CO_2 gas and bulk cyanobacterial cells, was calculated as ($\alpha_{CO2/bio}$ -
698 1)*1000, where $\alpha_{CO2/bio} = {}^{13}R_{CO2}/{}^{13}R_{bio}$, where ^{13}R is the ratio of ^{13}C to ^{12}C in the analyte. We note
699 this in contrast to other isotope literature where ε_p is calculated as $\alpha_{bio/CO2}$ - 1)*1000, which would
700 cause the positive values in this study to be negative. In this study, more positive ε_p values
701 indicate more ^{13}C-depleted; see Supplemental for more detail.
702
703 **Rubisco KIE assay**
704 *Syn*6301 and β-MRCA Rubisco were purified according to previous methodologies (84, 85) at
705 University of California, Davis and then shipped on dry ice to Caltech. Clarified lysate from a
706 BL21 DE3 Star *E. coli* culture expressing Rubisco was subjected to ammonium sulfate
707 precipitation, at the 30-40% cut for *Syn*6301 and at the 40-50% cut for β-MRCA, followed by
708 anion exchange chromatography and size exclusion chromatography. We then used the
709 substrate depletion method to measure the KIE of the *Syn*6301 and β-MRCA Rubiscos ($\varepsilon_{Rubisco}$),
710 as used previously in similar studies (47–50). Briefly, an assay mix of HCO_3^-, bovine carbonic
711 anhydrase, rubisco, ribulose 1,5-bisphosphate (RuBP), $MgCl_2$, bicine, and dithiothreitol (DTT)
712 was prepared. As the reaction progressed to completion, aliquots of that assay mix were

injected into pre-filled exetainers containing phosphoric acid that both stopped the reaction and converted all inorganic carbon species to gaseous CO_2. The $\delta^{13}C$ of these CO_2 aliquots was then measured on a Delta-V Advantage with Gas Bench and Costech elemental analyzer at Caltech. Here, instead of RuBP being given in excess, CO_2 was given in excess. In addition, instead of determining the fraction of CO_2 (*f*) consumed independently to create a Rayleigh plot, we fit the curvature of the $\delta^{13}C$ results to find *f* before converting to a Rayleigh plot to calculate $\varepsilon_{Rubisco}$, similar to previous studies (48). See Supplemental for more information.

Transmission Electron Microscopy (TEM) Imaging of Whole Cells

WT and ANC strains were grown in the reference condition as stated above (buffered BG-11 media, shaking at 250 rpm, with white cool fluorescent light at 120 µE, 30°C, and bubbled with ambient air (0.04% CO_2 (v/v)). WT and ANC cells were collected at mid-log (40 and 80 h, respectively) at OD730=0.4 and pelleted by centrifugation (10,000 x *g* for 10 min). Pelleted cells were then resuspended in 1 mL of cold solution 2.5% Glutaraldehyde in 0.1M Sodium Cacodylate Buffer, pH 7.4 (Electron Microscopy Sciences) and stored in the fixative solution at 4°C until imaging. Sample preparation and sectioning were performed in the Electron Microscope Laboratory core facility at the University of California Berkeley. Briefly, samples were stabilized in 1% low melting-point agarose, cut into small cubes, and then washed at room temperature with 0.1 M sodium cacodylate buffer, pH 7. Samples were then mixed with 1% osmium tetroxide, 1.6% potassium ferricyanide and 0.1 M cacodylate buffer pH 7.2 for an hour in the dark with rotation. These were washed again with a cacodylate buffer pH 7.2, then DI water, and subjected to an 1 h incubation with uranyl acetate 0.5% solution. After a new wash with DI water, samples were dehydrated by an ascending series of acetone concentration (35%, 50%, 75%, 80%, 90%, 100%, 100%). Later, samples were progressively infiltrated in resin (Epon solution: Eponate 12, DDSA NMA and BDMA (Electron Microscopy Sciences) with rotation, followed by a final step at 60˚C until polymerized. Thin sections (70 nm) were cut using a Reichert Ultracut E (Leica Microsystems) and collected on 100 mesh formvar coated copper grids. Sections were post-stained using 2% uranyl acetate in 70% methanol and followed with Reynold's lead citrate. The sections were imaged using a FEI Tecnai 12 transmission electron microscope operated at 120 kV (FEI). Images were collected using UltraScan 1000 digital micrograph software (Gatan Inc).

Acknowledgments

We thank Newton Nguyen for valuable guidance in the MCMC model used to calculate doubling times from growth curve data. We thank Victoria Orphan and Alex Sessions for access to lab space and analytical instruments, as well as lab managers Stephanie A. Connon, Fenfang Wu, and Nami Kitchen for assistance. This research was supported by the David and Lucille Packard Foundation (12540178), Simons Foundation (554187), NASA Exobiology (00010652), and the Schwartz-Reisman Collaborative Science Program (12520057). R.Z.W. was supported by a National Science Foundation Graduate Research Fellowship. Work in the lab of D.F.S. was supported by the US Department of Energy (DE-SC00016240). Work in the lab of P.M.S. was supported by a Society in Science–Branco Weiss fellowship from ETH Zürich and a Packard

Fellowship from the David Lucile Packard Foundation. We thank Danielle Jorgens and Reena Zalpuri at the University of California Berkeley Electron Microscope Laboratory for advice and assistance in electron microscopy sample preparation and data collection.

References

Bibliography

1. W. W. Fischer, J. Hemp, J. E. Johnson, Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016).

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

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

4. G. H. Lorimer, T. J. Andrews, Plant photorespiration—an inevitable consequence of the existence of atmospheric oxygen. *Nature* **243**, 359–360 (1973).

5. T. J. Andrews, G. H. Lorimer, *The Biochemistry of Plants: A Comprehensive Treatise, Vol. 10, Photosynthesis*, M. D. Hatch, N. K. Boardman, Eds. (1987).

6. A. Bar-Even, *et al.*, The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).

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

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

9. 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).

10. 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 USA* **113**, E5354-62 (2016).

11. J. A. Raven, J. Beardall, CO 2 concentrating mechanisms and environmental change. *Aquatic Botany* **118**, 24–37 (2014).

12. G. D. Price, *et al.*, The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. *J. Exp. Bot.* **64**, 753–768 (2013).

13. R. Riding, Cyanobacterial calcification, carbon dioxide concentrating mechanisms, and Proterozoic?Cambrian changes in atmospheric composition. *Geobiology* **4**, 299–316 (2006).

14. M. Schidlowski, A 3,800-million-year isotopic record of life from carbon in sedimentary

rocks. *Nature* **333**, 313–318 (1988).

15. J. Krissansen-Totton, R. Buick, D. C. Catling, A statistical analysis of the carbon isotope record from the Archean to Phanerozoic and implications for the rise of oxygen. *Am. J. Sci.* **315**, 275–316 (2015).

16. J. M. Hayes, Factors controlling 13C contents of sedimentary organic compounds: Principles and evidence. *Mar. Geol.* **113**, 111–125 (1993).

17. R. Francois, *et al.*, Changes in the δ^{13} C of surface water particulate organic matter across the subtropical convergence in the SW Indian Ocean. *Global Biogeochem. Cycles* **7**, 627–644 (1993).

18. R. Park, S. Epstein, Carbon isotope fractionation during photosynthesis. *Geochim. Cosmochim. Acta* **21**, 110–126 (1960).

19. G. D. Farquhar, M. H. O'Leary, J. A. Berry, On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Aust. J. Plant Physiol.* **9**, 121 (1982).

20. M. Eichner, S. Thoms, S. A. Kranz, B. Rost, Cellular inorganic carbon fluxes in Trichodesmium: a combined approach using measurements and modelling. *J. Exp. Bot.* **66**, 749–759 (2015).

21. T. D. Sharkey, J. A. Berry, "Carbon Isotope Fractionation of Algae as Influenced by an Inducible CO2 Concentrating Mechanism" in *Inorganic Carbon Uptake by Aquatic Photosynthetic Organisms*, W. J. Lucas, J. A. Berry, Eds. (The American Society of Plant Physiologists, 1985), pp. 389–401.

22. W. G. Mook, J. C. Bommerson, W. H. Staverman, Carbon isotope fractionation between dissolved bicarbonate and gaseous carbon dioxide. *Earth and Planetary Science Letters* **22**, 169–176 (1974).

23. J. P. Jasper, J. M. Hayes, A carbon isotope record of CO2 levels during the late Quaternary. *Nature* **347**, 462–464 (1990).

24. M. Pagani, *et al.*, The role of carbon dioxide during the onset of Antarctic glaciation. *Science* **334**, 1261–1264 (2011).

25. J. A. Higgins, *et al.*, Atmospheric composition 1 million years ago from blue ice in the Allan Hills, Antarctica. *Proc Natl Acad Sci USA* **112**, 6887–6891 (2015).

26. A. K. Garcia, C. M. Cavanaugh, B. Kacar, The curious consistency of carbon biosignatures over billions of years of Earth-life coevolution. *ISME J.* **15**, 2183–2194 (2021).

27. G. D. Farquhar, J. R. Ehleringer, K. T. Hubick, Carbon Isotope Discrimination and Photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 503–537 (1989).

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

29. A. J. Boller, P. J. Thomas, C. M. Cavanaugh, K. M. Scott, Low stable carbon isotope

fractionation by coccolithophore RubisCO. *Geochim. Cosmochim. Acta* **75**, 7200–7207 (2011).

30. J. M. Hayes, H. Strauss, A. J. Kaufman, The abundance of 13C in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125 (1999).

31. B. N. Popp, *et al.*, Effect of phytoplankton cell geometry on carbon isotopic fractionation. *Geochim. Cosmochim. Acta* **62**, 69–77 (1998).

32. E. A. Laws, B. N. Popp, R. R. Bidigare, M. C. Kennicutt, S. A. Macko, Dependence of phytoplankton carbon isotopic composition on growth rate and [CO2)aq: Theoretical considerations and experimental results. *Geochim. Cosmochim. Acta* **59**, 1131–1138 (1995).

33. R. R. Bidigare, *et al.*, Consistent fractionation of^{13} C in nature and in the laboratory: Growth-rate effects in some haptophyte algae. *Global Biogeochem. Cycles* **11**, 279–292 (1997).

34. Y. G. Zhang, J. Henderiks, X. Liu, Refining the alkenone-pCO2 method II: Towards resolving the physiological parameter 'b.' *Geochim. Cosmochim. Acta* **281**, 118–134 (2020).

35. G. H. Rau, U. Riebesell, D. Wolf-Gladrow, A model of photosynthetic 13C fractionation by marine phytoplankton based on diffusive molecular CO2 uptake. *Mar. Ecol. Prog. Ser.* **133**, 275–285 (1996).

36. M. H. O'Leary, Measurement of the isotope fractionation associated with diffusion of carbon dioxide in aqueous solution. *J. Phys. Chem.* **88**, 823–825 (1984).

37. N. Cassar, E. A. Laws, B. N. Popp, Carbon isotopic fractionation by the marine diatom Phaeodactylum tricornutum under nutrient- and light-limited growth conditions. *Geochim. Cosmochim. Acta* **70**, 5323–5335 (2006).

38. J. A. Berry, "Studies of mechanisms affecting the fractionation of carbon isotopes in photosynthesis" in *Stable Isotopes in Ecological Research*, Ecological Studies., P. W. Rundel, J. R. Ehleringer, K. A. Nagy, Eds. (Springer New York, 1989), pp. 82–94.

39. H. Gimmler, C. Weiss, M. Baier, W. Hartung, The conductance of the plasmalemma for co_2. *J. Exp. Bot.* **41**, 785–795 (1990).

40. C. Rotatore, R. R. Lew, B. Colman, Active uptake of CO2 during photosynthesis in the green alga Eremosphaera viridis is mediated by a CO2-ATPase. *Planta* **188**, 539–545 (1992).

41. D. S□ltemeyer, K. Biehler, HeinrichP. Fock, Evidence for the contribution of pseudocyclic photophosphorylation to the energy requirement of the mechanism for concentrating inorganic carbon in Chlamydomonas. *Planta* **189** (1993).

42. S. J. Hurley, B. A. Wing, C. E. Jasper, N. C. Hill, J. C. Cameron, Carbon isotope evidence for the global physiology of Proterozoic cyanobacteria. *Sci. Adv.* **7** (2021).

43. A. K. Garcia, *et al.*, System-level effects of CO2 and RuBisCO concentration on carbon

870 isotope fractionation. *BioRxiv* (2021) https:/doi.org/10.1101/2021.04.20.440233.

871 44. M. Kędzior, *et al.*, Molecular foundations of Precambrian uniformitarianism. *BioRxiv*
872 (2021) https:/doi.org/10.1101/2021.05.31.446354.

873 45. B. Kacar, V. Hanson-Smith, Z. R. Adam, N. Boekelheide, Constraining the timing of the
874 Great Oxidation Event within the Rubisco phylogenetic tree. *Geobiology* **15**, 628–640
875 (2017).

876 46. P. M. Shih, *et al.*, Biochemical characterization of predicted Precambrian RuBisCO. *Nat.*
877 *Commun.* **7**, 10382 (2016).

878 47. R. D. Guy, M. L. Fogel, J. A. Berry, Photosynthetic fractionation of the stable isotopes of
879 oxygen and carbon. *Plant Physiol.* **101**, 37–47 (1993).

880 48. D. B. McNevin, M. R. Badger, H. J. Kane, G. D. Farquhar, Measurement of (carbon)
881 kinetic isotope effect by Rayleigh fractionation using membrane inlet mass spectrometry
882 for CO2-consuming reactions. *Funct. Plant Biol.* **33**, 1115 (2006).

883 49. K. M. Scott, J. Schwedock, D. P. Schrag, C. M. Cavanaugh, Influence of form IA RubisCO
884 and environmental dissolved inorganic carbon on the delta13C of the clam-
885 chemoautotroph symbiosis Solemya velum. *Environ. Microbiol.* **6**, 1210–1219 (2004).

886 50. P. J. Thomas, *et al.*, Isotope discrimination by form IC RubisCO from Ralstonia eutropha
887 and Rhodobacter sphaeroides, metabolically versatile members of “Proteobacteria” from
888 aquatic and soil habitats. *Environ. Microbiol.* (2018) https:/doi.org/10.1111/1462-
889 2920.14423.

890 51. G. G. B. Tcherkez, G. D. Farquhar, T. J. Andrews, Despite slow catalysis and confused
891 substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly
892 optimized. *Proc Natl Acad Sci USA* **103**, 7246–7251 (2006).

893 52. C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial
894 microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).

895 53. G. D. Price, M. R. Badger, Isolation and Characterization of High CO(2)-Requiring-
896 Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes that
897 Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO(2) within the
898 Carboxysome. *Plant Physiol.* **91**, 514–525 (1989).

899 54. H. Wang, *et al.*, Rubisco condensate formation by CcmM in β-carboxysome biogenesis.
900 *Nature* **566**, 131–135 (2019).

901 55. T. E. Jensen, Electron microscopy of polyphosphate bodies in a blue-green alga, Nostoc
902 pruniforme. *Archiv. Mikrobiol.* **62**, 144–152 (1968).

903 56. S. A. Kranz, S. Dieter, K.-U. Richter, B. Rost, Carbon acquisition by Trichodesmium: the
904 effect of pCO2 and diurnal changes. *Limnol. Oceanogr.* **54**, 548–559 (2009).

905 57. M. Kędzior, *et al.*, Resurrected Rubisco suggests uniform carbon isotope signatures over
906 geologic time. *Cell Rep.* **39**, 110726 (2022).

907 58. J. L. Collier, A. R. Grossman, Chlorosis induced by nutrient deprivation in Synechococcus
908 sp. strain PCC 7942: not all bleaching is the same. *J. Bacteriol.* **174**, 4718–4726 (1992).

59. J. Erez, A. Bouevitch, A. Kaplan, Carbon isotope fractionation by photosynthetic aquatic microorganisms: experiments with*Synechococcus* PCC7942, and a simple carbon flux model. *Can. J. Bot.* **76**, 1109–1118 (1998).

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

61. T. Ogawa, A. Kaplan, Inorganic carbon acquisition systems in cyanobacteria. *Photosyn. Res.* **77**, 105–115 (2003).

62. M. Volokita, D. Zenvirth, A. Kaplan, L. Reinhold, Nature of the Inorganic Carbon Species Actively Taken Up by the Cyanobacterium Anabaena variabilis. *Plant Physiol.* **76**, 599–602 (1984).

63. G. D. Price, S. Maeda, T. Omata, M. R. Badger, Modes of active inorganic carbon uptake in the cyanobacterium, Synechococcus sp. PCC7942. *Functional Plant Biol.* **29**, 131 (2002).

64. S. Maeda, M. R. Badger, G. D. Price, Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO2 hydration in the cyanobacterium, Synechococcus sp. PCC7942. *Mol. Microbiol.* **43**, 425–435 (2002).

65. B. Klughammer, D. Sültemeyer, M. R. Badger, G. D. Price, The involvement of NAD(P)H dehydrogenase subunits, NdhD3 and NdhF3, in high-affinity CO2 uptake in Synechococcus sp. PCC7002 gives evidence for multiple NDH-1 complexes with specific roles in cyanobacteria. *Mol. Microbiol.* **32**, 1305–1315 (1999).

66. J. M. Schuller, *et al.*, Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I. *Nat. Commun.* **11**, 494 (2020).

67. Z. Sade, I. Halevy, New constraints on kinetic isotope effects during CO 2(aq) hydration and hydroxylation: Revisiting theoretical and experimental data. *Geochim. Cosmochim. Acta* **214**, 246–265 (2017).

68. R. E. Zeebe, D. Wolf-Gladrow, *CO2 in seawater: Equilibrium, kinetics, isotopes* (Elsevier, 2001).

69. I. D. Clark, B. Lauriol, Kinetic enrichment of stable isotopes in cryogenic calcites. *Chem. Geol.* **102**, 217–228 (1992).

70. W. Guo, “CARBONATE CLUMPED ISOTOPE THERMOMETRY: APPLICATION TO CARBONAECOUS CHONDRITES & EFFECTS OF KINETIC ISOTOPE FRACTIONATION,” California Institute of Technology. (2009).

71. R. E. Zeebe, Kinetic fractionation of carbon and oxygen isotopes during hydration of carbon dioxide. *Geochim. Cosmochim. Acta* **139**, 540–552 (2014).

72. J. D. Boettger, J. D. Kubicki, Equilibrium and kinetic isotopic fractionation in the CO2 hydration and hydroxylation reactions: Analysis of the role of hydrogen-bonding via quantum mechanical calculations. *Geochim. Cosmochim. Acta* **292**, 37–63 (2021).

73. D. Tchernov, *et al.*, Passive entry of CO2 and its energy-dependent intracellular conversion to HCO3- in cyanobacteria are driven by a photosystem I-generated

deltamuH+. *J. Biol. Chem.* **276**, 23450–23455 (2001).

74. G. S. Espie, A. G. Miller, D. T. Canvin, High affinity transport of CO(2) in the cyanobacterium synechococcus UTEX 625. *Plant Physiol.* **97**, 943–953 (1991).

75. A. Kaplan, L. Reinhold, Co2 concentrating mechanisms in photosynthetic microorganisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 539–570 (1999).

76. J. Erez, A. Bouevitch, A. Kaplan, Carbon isotope fractionation by photosynthetic aquatic microorganisms: experiments with *Synechococcus* PCC7942, and a simple carbon flux model. *Can. J. Bot.* **76**, 1109–1118 (1998).

77. N. Adir, M. Dines, M. Klartag, A. McGregor, M. Melamed-Frank, “Assembly and disassembly of phycobilisomes” in *Complex Intracellular Structures in Prokaryotes*, Microbiology Monographs., J. M. Shively, Ed. (Springer Berlin Heidelberg, 2006), pp. 47–77.

78. N. Battchikova, M. Eisenhut, E.-M. Aro, Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim. Biophys. Acta* **1807**, 935–944 (2011).

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

80. C. Fei, A. T. Wilson, N. M. Mangan, N. S. Wingreen, M. C. Jonikas, Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nat. Plants* **8**, 583–595 (2022).

81. M. A. O’Malley, D. A. Walsh, Rethinking microbial infallibility in the metagenomics era. *FEMS Microbiol. Ecol.* **97** (2021).

82. A. I. Flamholz, *et al.*, Trajectories for the evolution of bacterial CO2-concentrating mechanisms. *Proc Natl Acad Sci USA* **119**, e2210539119 (2022).

83. S. S. Golden, L. A. Sherman, Optimal conditions for genetic transformation of the cyanobacterium Anacystis nidulans R2. *J. Bacteriol.* **158**, 36–42 (1984).

84. S. Saschenbrecker, *et al.*, Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).

85. D. M. Banda, *et al.*, Novel bacterial clade reveals origin of form I Rubisco. *Nat. Plants* **6**, 1158–1166 (2020).