Title: Evolution of Carbon Isotope Fractionation in Cyanobacteria

Authors: Renée Z. Wang*[1], Robert J. Nichols[2], Albert K. Liu[3,4], Avi I. Flamholz[5], Doug M. Banda[6], David F. Savage[2,7], John M. Eiler[1], Patrick M. Shih[3,8],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, Davis, Department of Plant Biology. (7) Howard Hughes Medical Institute, University of California, Berkeley, California 94720. (8) University of California, Berkeley, Department of Plant and Microbial Biology.

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

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

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 interrogated 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, violating 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 modified model indicated that powered inorganic carbon uptake systems contribute to ε_p, and this effect is exacerbated in ANC. These data suggested that understanding the evolution of both the CCM and rubisco is critical for interpreting the carbon isotope record, and that large fluctuations in the record may reflect the evolving efficiency of carbon fixing metabolisms as well as changes in atmospheric CO_2.

Significance Statement

Fossils record the past, but so too do modern organisms via comparative biology. Rubisco is the most abundant protein on the planet, and is a keystone enzyme in photosynthesis. To understand how this process has co-evolved with changes in the abundance of atmospheric carbon dioxide, we reconstructed an ancestral rubisco (>one billion years old), and generated a mutant Cyanobacteria strain that must rely on this ancient protein for growth. By measuring the carbon isotope fractionation *in vitro* and *in vivo* we found that prevailing models of carbon flow in Cyanobacteria could be corrected by accounting for known aspects of cyanobacterial physiology. This highlighted the value of considering both evolution and physiology for comparative biological approaches to understanding Earth history.

Main Text

Introduction

Throughout Earth's history, autotrophic cells have had to take external, oxidized inorganic carbon (C_i) and 'fix' it into reduced organic carbon (C_o) to create biomass and grow. The challenge is particularly acute for aquatic autotrophs because the diffusion constant of CO_2 in air is ~10,000-fold greater than it is in water. In addition, in waters at pH 7-8, HCO_3^- is 10-100 times more abundant than CO_2 but HCO_3^- is much less membrane permeable (1). Therefore, the chemistry of C_i imposes inherent constraints that carbon-fixing organisms must work within.

Today, and for much of Earth's history, 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 (k_{cat}), 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 the evolutionary history of this enzyme 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. Rubisco is also thought to have been the primary carboxylating enzyme of global photosynthesis since the Great Oxygenation Event, and potentially far prior (8).

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. Currently, all known Cyanobacteria have CCMs, as do many bacterial chemolithoautotrophs, many aquatic algae, and some plants (9). The bacterial CCM has two main components: i) an C_i pump producing high cytosolic HCO_3^-, and ii) co-encapsulation of carbonic anhydrase (CA) and rubisco inside proteinaceous organelles known as carboxysomes (Figure 1A) (10, 11). The timing of carboxysome CCM evolution is uncertain, but it likely arose sometime during the Proterozoic Eon (9). Therefore, for roughly half of Earth's history, bacterial rubiscos have functioned in concert with a system that pumps C_i into and around the cell.

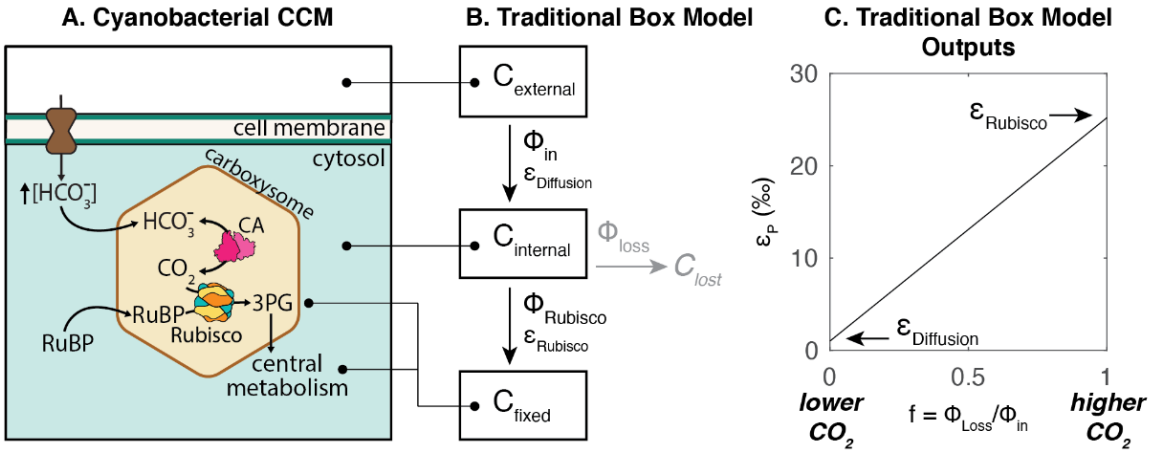

Figure 1: Cyanobacterial CO_2 Concentrating Mechanism (CCM) compared to traditional box model architecture and outputs. A) Cyanobacterial CCMs rely on i) active HCO_3^- uptake into the cell; ii) co-encapsulation of carbonic anhydrase (CA) and rubisco within the carboxysome. These components together produce a high carboxysomal CO_2 concentration that enhances CO_2 fixation by rubisco and suppresses oxygenation. B) Architecture of the traditional box model based on (12–15) mapped onto a cyanobacterial CCM; 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, or fluxes via powered inorganic carbon uptake. C) Model solution is $\varepsilon_P = (1-f)*\varepsilon_{Diffusion} + f*\varepsilon_{Rubisco}$, where ε_P is defined as the difference in $\delta^{13}C$ of $C_{external}$ and C_{fixed}, and f is defined as the ratio of Φ_{loss}/Φ_{in}. For this illustration, $\varepsilon_{Rubisco}$ = 25‰ and $\varepsilon_{Diffusion}$ = 1‰ is assumed. When $f = 0$, $\varepsilon_P = \varepsilon_{Diffusion}$, and when $f = 1$, $\varepsilon_P = \varepsilon_{Rubisco}$.

Over geological timescales, trends in the environmental CO_2 concentrations and the evolutionary history of carbon fixation and its role in the carbon cycle have been captured, in some noisy manner, in the carbon isotope record. This record is composed of measurements of the relative ratios of ^{13}C to ^{12}C isotopes in C-bearing phases in sedimentary rocks over time (reported using the delta notation ($\delta^{13}C$) and expressed in per mil (‰); see Methods for detail). The carbon isotope record established a time series of touchpoints between biological and geological processes that—if interpreted correctly—might be read to understand the evolution of carbon-fixing processes through time. One key to reading the isotope record is the observation that rubisco displays a kinetic isotope effect (KIE) where it preferentially fixes $^{12}CO_2$ over $^{13}CO_2$ due to the k_{cat} being slightly faster for $^{12}CO_2$ than $^{13}CO_2$ (16). This results in the reaction product, 3-phosphoglycerate (3-PGA), being relatively depleted in ^{13}C by several percent (tens of per mil ‰) compared to the overall isotopic composition of the initial CO_2 substrate. The difference in $\delta^{13}C$ of the CO_2 substrate and the 3-PGA product is typically reported as $\varepsilon_{Rubisco}$, and has been measured to vary between 18-30‰ for most extant rubiscos (17, 18), with the exception of rubisco from the coccolithophore *Emiliania huxleyi* at 11‰ (19). Because all resulting biomass is synthesized from 3-PGA in autotrophs utilizing the CBB cycle, biomass is depleted in ^{13}C compared to external C_i pools. The magnitude of this difference is called ε_P. C_i pools are preserved in the rock record in the form of carbonate salts (in limestones and dolomites), while biomass and C_o pools are preserved in organic phases (typically kerogen) in a myriad of lithologies and are measured as rock total organic carbon (TOC). 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 is termed ε_{TOC} and varies slightly from ε_P (20). Overall, if one can accurately convert from ε_{TOC} to ε_P, and then from ε_P to $\varepsilon_{Rubisco}$, one may be able to learn about photosynthetic physiology over time from the rock record.

Much work has been done to address two main observations seen in the carbon isotope record: i) there is variation in ε_{TOC} despite its general consistency around 25‰ (18), and ii) organic matter is more ^{13}C-depleted further back in time compared to today, particularly in Precambrian-age rocks (20–23). If one accepts that these trends are not due to chemical alteration of sediment post-deposition (diagenesis) or other preservation effects (which are expected to have the opposite sign (24)), then one looks for mechanisms that might explain such a change. To date, those efforts have focused on environmental changes (e.g. temperature, pH, pCO_2) that might have led to larger carbon isotope fractionations. Typically, model autotrophs are grown in different conditions, and then a proxy is calibrated that relates carbon isotope fractionation to the chosen environmental condition. Doing so, the community has found that in both lab cultures of photosynthetic algae (25, 26), and from field data of marine algae (13, 27, 28), that increased levels of dissolved CO_2 in solution corresponded to an increased value of ε_P. This general observation has been modified and honed by later studies to account for environmental parameters like the effects of temperature and pH on dissolved CO_2 concentrations, and physiological characteristics like growth rate, cell geometry/type, and

species/strain-type (25, 26, 29). However, even when taking these factors into account, it is still expected that increased concentrations of CO_2 will cause increased values of ε_P (see (18) for review of factors affecting ε_P).

This supposition undergirds the traditional box model of carbon isotope fractionation in algae (12, 13), which was based on a model of carbon isotope fractionation in C3 plants (14, 15), which have no CCM (Figure 1B). This is often referred to as a "diffusion-based" model because CO_2 is assumed to passively diffuse in and out of aquatic cells (like gas exchange through leaf stomata) due to a concentration gradient of external vs. internal CO_2. A "leakiness" term, *f*, can be defined as the ratio of fluxes (Φ) of carbon exiting or entering the cell based on this concentration gradient ($f = \Phi_{out}/\Phi_{in} = [C_{int}]/[C_{ext}]$). In this simplified model, ε_p is determined by the isotopic effect of two distinct steps: i) the diffusion of CO_2 into the cell ($\varepsilon_{Diffusion}$; <1‰ in water at steady state (30)); and ii) the carbon fixation step catalyzed by rubisco ($\varepsilon_{Rubisco}$; typically 18-30‰). The model is usually solved by assuming steady state, which results in a linear relationship between ε_p and *f* so that experimentally measured values of ε_p can be used to solve for *f*. This model also sets the minimum and maximum ε_p values ($\varepsilon_{Diffusion}$ and $\varepsilon_{Rubisco}$ respectively, Fig. 1C) with corresponding physiological interpretations: 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 ($f \approx 1$), rubisco can "choose" between ^{12}C and ^{13}C substrates, and the KIE of rubisco can be fully expressed. 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 (see Supplemental and (13)).

One reason that Earth scientists are interested in models of biological carbon isotope fractionation is that it may help constrain historical pCO_2 concentrations. The ice core record, which provides direct observations of the Earth's atmosphere, extends back only ≈1 million years (31), so, for the remaining four and half billion years of Earth history, model-driven proxies are used (32, 33). However, under the simplified diffusional model, ε_p cannot exceed $\varepsilon_{Rubsico}$ (Figure 1C), yet the largest ε_P values observed in the Archaean Eon exceed 30‰ (21, 34) and also exceed all existing measurements of $\varepsilon_{Rubsico}$. In addition, recent studies in dinoflagellates that showed that ε_p can regularly exceed $\varepsilon_{Rubsico}$ under certain growth conditions argue that ε_p reflects additional isotopic fractionations that may occur with carbon uptake processes. This has motivated an updated model of Eukaryotic algae accounting for the estimated isotopic fractionations of different C_i uptake mechanisms (17).

In addition to taking modern physiology into account, it is important to understand how the evolution of rubisco and the CCM may have also affected the carbon isotope composition of biomass. Recent studies have addressed this issue directly by testing model organisms that may better resemble an ancestral counterpart, including a cyanobacterial strain that lacks a CCM (22), a cyanobacterial strain that overexpresses rubisco (35), and a cyanobacterial strain expressing an inferred ancestral rubisco dating from ≈1-3 Ga in age (36, 37).

Here, we measured the ε_p of a control strain of *S. elongatus* PCC 7942 expressing the wild-type rubisco (NS2-KanR, referred to as 'WT'), as well as a mutant expressing an inferred ancestral Form 1B rubisco dating to >1 Ga (referred to as 'ANC') (38) in varied CO_2 and light

conditions. In addition, we measured the KIEs of the present-day and ancestral rubiscos ($\varepsilon_{Rubsico}$) *in vitro*. We observed that: i) ε_p is greater for ANC than for its WT counterpart for all conditions, 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 higher light while WT ε_p increases with higher 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 conditions. ANC ε_p exceeding $\varepsilon_{Rubsico}$ implies that the traditional box model is incomplete and additional processes and fractionation factors are needed. In addition, other 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 strain grows at ambient CO_2 concentrations

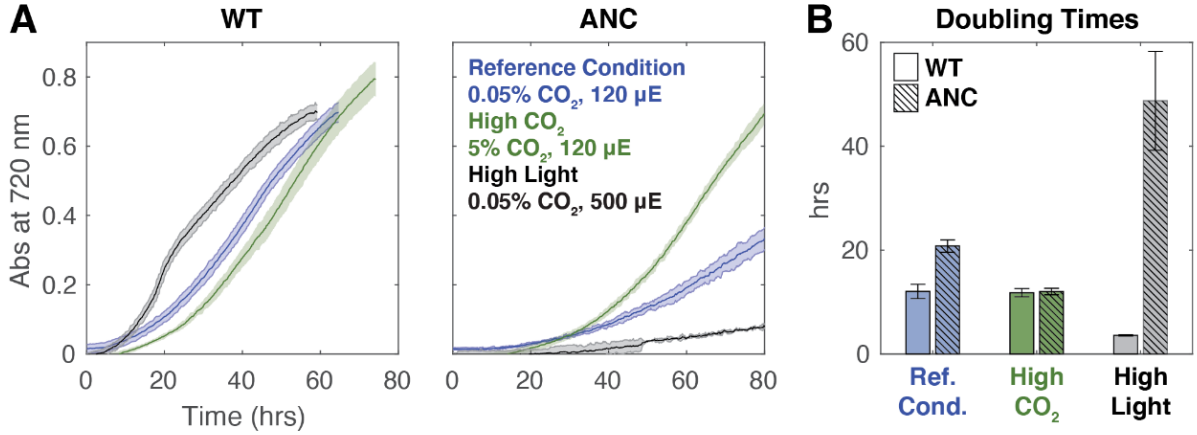

Figure 2: Growth curves for WT and ANC strains across experimental conditions. A) Averaged growth curves shown for WT and ANC strains to 80 hours, colored by growth condition as indicated in figure. Data was smoothed with a rolling median (Methods); see full ANC growth curves in Supplemental. B) Average doubling times with standard deviation for growth curves on the left. Doubling times were calculated using a Markov Chain Monte Carlo (MCMC) approach; see Supplemental for details. ANC displayed a growth defect relative to the WT at the reference condition, which was rescued at high CO_2. Across all conditions, ANC grew the slowest at high light, while WT grew the fastest.

Working in *S.elongatus* PCC 7942, we produced a mutant strain lacking the native Form 1B rubisco, expressing instead an ancestral Form 1B rubisco produced by computational ancestral sequence reconstruction (Methods). This putative ancestral rubisco was previously purified and characterized *in vitro* (38). We compared the growth of this mutant strain, which we termed 'ANC', to the parent strain (wild-type or 'WT') across several light fluxes and CO_2 concentrations: i) Reference condition (ambient pCO_2 (0.05% (v/v)), standard light flux (120 µE)); ii) High CO_2 (5% pCO_2 (v/v), 120 µE); iii) High light (0.05% pCO_2 (v/v), 500 µE). Remarkably, the ANC strain managed to grow at ambient pCO_2 and standard light conditions (Figure 2A), even though the ancestral rubisco has a carboxylation rate (V_C) roughly half that of

WT (Table 1). This result is in contrast to another historical analogue strain (lacking CCM genes) that only grew in elevated pCO_2 (22), but is similar to another ancestral analogue (with a predicted Precambrian Form 1B rubisco) that did grow in under ambient conditions (36).

The difference in V_C between the ancestral and modern rubiscos was mirrored in the doubling times of WT and ANC strains (Figure 2B, Table S3), where the doubling time of ANC was roughly twice that of WT in the reference condition (20.8 ± 1.2 vs. 12.0 ± 1.4 hours respectively (avg. ± s.d.). In addition, the carboxylation rate normalized to the Michaelis-Menten constant for CO_2 in air (V_C/K_C^{Air}) for the ancestral Form 1B rubisco, which measures the enzyme's ability to function at low CO_2 concentrations, is roughly half that of the modern Form 1B rubisco (Table 1; see (38) for full characterization of enzyme kinetics). This suggested that ANC's growth was limited by its ability to fix CO_2 in a manner proportional to rubisco's carboxylation rate. This growth defect was ameliorated at 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) in contrast to ANC, whose doubling times did change in response to elevated CO_2. This suggested that CO_2 availability is a growth-limiting factor for ANC but not WT, and further suggested that the differences in their growth kinetics are related to the efficacy of their CCMs. Consistent with our results, a similar ancestral Form 1B analogue displayed total carboxylase activity roughly half that of the modern Form 1B (39).

We observed the greatest differences in doubling times between ANC and WT when the strains were grown at high light (Figure 2, 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. S9), suggesting that ANC cultures were cannibalizing their photosynthetic antennae via a known starvation pathway to reduce the cell's capacity for light harvesting and photochemical electron transport (40). It is likely that ANC could not match its rate of CO_2 fixation to the rate of light harvesting, and hence invoked this regulatory pathway. WT, in contrast, grew rapidly in the high light condition, ostensibly because its rate of CO_2 fixation could match the higher light-harvesting rate.

Ancestral rubisco enzyme fractionates less than WT rubisco enzyme

We measured the *in vitro* fractionations of the WT and ANC rubisco using the substrate depletion method ((41–44); see Methods and Supplemental for more details). Previous work on rubisco isotope kinetics 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 (45). We therefore expected ANC and WT $\varepsilon_{Rubsico}$ values to be the same within uncertainty because of their similar $S_{C/O}$ values, but we found that ANC $\varepsilon_{Rubsico}$ (17.23 ± 0.61‰) fractionated carbon isotopes during carboxylation about 8‰ less than that of 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. Relevant enzyme characteristics for the ancestral vs. modern Form 1B rubisco. Starred values (*) for the modern Form 1B were measured in rubiscos purified from *Synechococcus* sp. PCC 6301, a close relative of our working WT strain, *Synechococcus* sp. PCC 7942. Kinetic isotope effect ($\varepsilon_{Rubisco}$, avg. ± s.e.) was measured using the substrate depletion method (41–44); see Methods and Supplemental for more detail. Rate of carboxylation (V_C), Michaelis-Menten constant for CO_2 in ambient air at modern levels of atmospheric O_2 (K_C^{Air}), rate of carboxylation normalized to the Michaelis-Menten constant (V_C/K_C^{Air}), and specificity for CO_2 vs. O_2 ($S_{C/O}$) are from (38); additional details on these values are within. $S_{C/O}$ is a unitless measure of the relative preference for CO_2 over O_2, and is calculated as $(V_C/K_C)/(V_{,O}/K_O)$ where K_C and K_O are the Michaelis-Menten constants for CO_2 and O_2 concentrations respectively, and V_C and V_O are the carboxylation and oxygenation turnover rate under substrate-saturated conditions. V_C/K_C^{Air} measures the ability of rubisco to function at low CO_2 concentrations because it describes the initial response rate of carboxylation to the CO_2 concentration.

ANC strain fractionates more than WT strain

Counter to expectations based on enzyme KIEs (Table 1), larger ε_p values were observed for ANC than WT in all CO_2 and light conditions tested (Figure 3). This was consistent with results from a similar ancestral analog, where larger ε_p values exceeded WT at ambient and elevated CO_2 levels (36). The highest ANC ε_p values were observed for cultures grown at 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 (Figure 3A). Compared to the reference condition, ANC ε_P values did not increase in high CO_2 and only increased in high light. In contrast, WT ε_P values were indifferent to high light and only increased in high CO_2. This result contrasted with the ancestral strain in (36) where ε_P values increased by ≈10‰ at 2% CO_2.

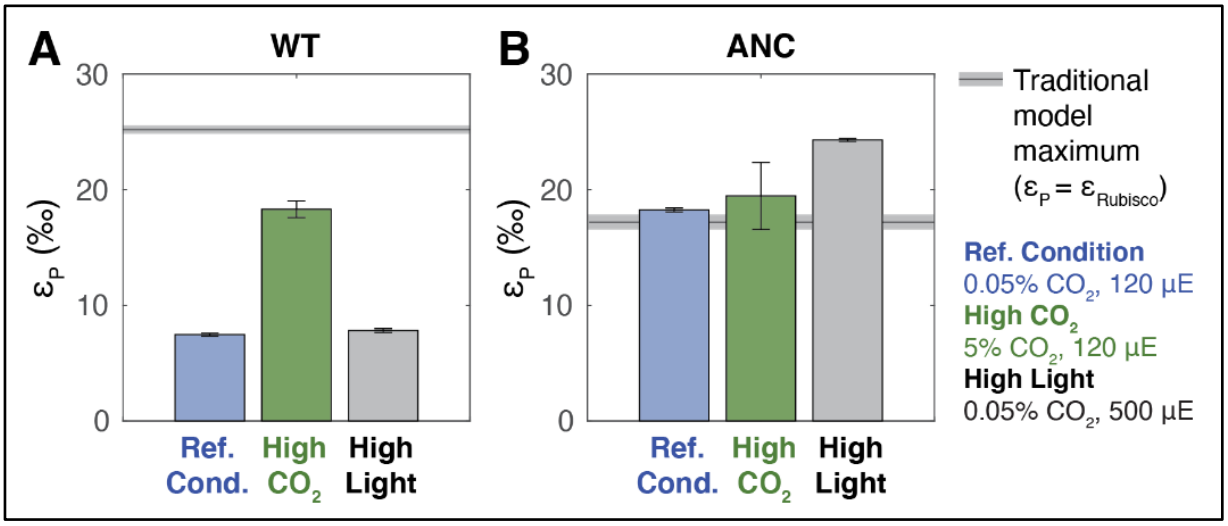

Figure 3: Whole cell carbon isotope fractionation by WT and ANC strains. ε_P (‰) values (avg. ± s.e.) for A) WT, and B) ANC strains across growth conditions.. For each strain, the maximum ε_P possible based on the traditional model ($\varepsilon_P = \varepsilon_{Rubisco}$) is shown as a gray line (avg. ± s.e.). Most measured ANC ε_P values exceed the theoretical limit ($\varepsilon_P > \varepsilon_{Rubisco}$) , while all 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 Supplemental for full results.

In addition, the traditional box model described above cannot accommodate ε_p values in excess of $\varepsilon_{Rubsico}$ (Figure 1). However, average ANC ε_P values exceeded ANC $\varepsilon_{Rubsico}$ in all growth conditions (Figure 3), 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 (Figure 1), which cannot accommodate the ANC ε_P observations.

Proposed influence of a light-powered carbonic-anhydrase

The data showed two primary observations that could not be reconciled by the traditional box model: i) ANC ε_P often exceeds $\varepsilon_{Rubsico}$, and ii) ANC ε_P varies primarily with light and not CO_2. Based on independent knowledge of carbon acquisition strategies in Cyanobacteria, we augmented the traditional box model to include an additional isotope fractionation step, and a parameter affected by light. A good candidate that fulfills both these criteria would be a process that uses energy derived from light to catalyze the energy-coupled unidirectional hydration of CO_2 to HCO_3^-—a “powered” carbonic anhydrase (CA).

Cyanobacteria have been shown to have two modes of active C_i uptake: uptake of hydrated C_i (predominantly H_2CO_3 and HCO_3^-) and uptake of CO_2 (46). In order for the CCM to function, either mode would need to produce a high, non-equilibrium concentration of HCO_3^- in the cytoplasm (9, 11). This is thought to be achieved by coupling CA to an energy source (e.g. light or an ion gradient) that drives the one-way hydration of CO_2 to HCO_3^- in the cytoplasm (47). There is now excellent data supporting this hypothesis in Cyanobacteria, where accessory proteins that bind to the cyanobacterial homologue of Complex I NADH dehydrogenase (the inducible NDH-1MS or constitutive NDH-1MS’) are known to mediate CO_2 uptake specifically (48–50). Additionally, one of these accessory proteins, CupA in NDH-1MS and CupB in NDH-

329 1MS', is homologous to a carbonic anhydrase and contains a telltale zinc active site situated
330 near a proton channel in a membrane subunit (51). This powered carbonic anhydrase Complex
331 I variant in Cyanobacteria therefore couples inorganic carbon uptake directly to the
332 photochemical electron transport chain. Moreover, a similar protein complex has been
333 described in proteobacterial chemoautotrophs, suggesting that energy-coupled CO_2 hydration is
334 widespread (52).
335 A unidirectional CA would affect ε_p for two reasons. First, CO_2 and HCO_3^- are isotopically
336 distinct. At equilibrium in standard conditions, HCO_3^- is ≈8‰ more enriched in ^{13}C than CO_2 (53,
337 54). Therefore, if a cyanobacterium is predominantly taking up CO_2, then the internal C_i pool
338 from which biomass is formed is isotopically lighter (^{13}C-depleted) than if HCO_3^- is the dominant
339 source of C_i. Second, unidirectional CO_2 hydration is expected to impart a substantial isotope
340 effect, with calculated values ranging from ≈19 to 32‰ (53, 55–58). Therefore, there are two
341 mechanistic reasons that ε_P could exceed $\varepsilon_{Rubisco}$ in conditions where energized CO_2 uptake and
342 hydration is active. Indeed, unidirectional CO_2 hydration has also been proposed to contribute to
343 algal ε_p values that exceed known algal $\varepsilon_{Rubisco}$ values (17).
344 The activity of a powered CA could modulate ε_p independent of external C_i
345 concentrations, counter to the traditional model which proposes ε_P as a direct correlate of
346 external pCO_2 (12, 13). Becausehis powered CO_2 uptake and hydration is driven by light
347 energy, e.g. via cyclic electron flow around photosystem I (51), this may explain why the
348 highest ANC ε_p values, those exceeding ANC $\varepsilon_{Rubisco}$, were observed under high light conditions.
349 Furthermore, on short timescales (≈minutes) cyanobacterial C_i uptake can be modulated by light
350 intensity alone, fully independent of external C_i concentrations (59), and CO_2 uptake can occur
351 in the absence of carbon fixation (60, 61). Based on these physiological and isotopic
352 observations, a powered CA is likely responsible for ε_p values in excess of $\varepsilon_{Rubisco}$ in ANC.

353 **Proposed model for carbon isotope fractionation**

354 As discussed above, the traditional box model cannot produce $\varepsilon_p > \varepsilon_{rubisco}$ (Figure 1).
355 Based on our isotopic data and understanding of cyanobacterial physiology, we proposed a
356 simple modification of the traditional model that can rationalize our measurements of growth and
357 ε_p for both WT and ANC strains (Figure 4). The traditional box model was able to rationalize our
358 data from WT in the sense that the values of the CO_2 leakage term, f, fit from ε_p are compatible
359 with growth (i.e. $f < 1$). In this model, $f = 1$ implies that all carbon uptake leaks out of the cell.
360 However, using the traditional model on the ANC strain data yields values of f inconsistent with
361 growth ($f > 1$; Figure 4A), which clearly violates growth curve data that shows ANC was able to
362 grow in all conditions (Figure 2).
363

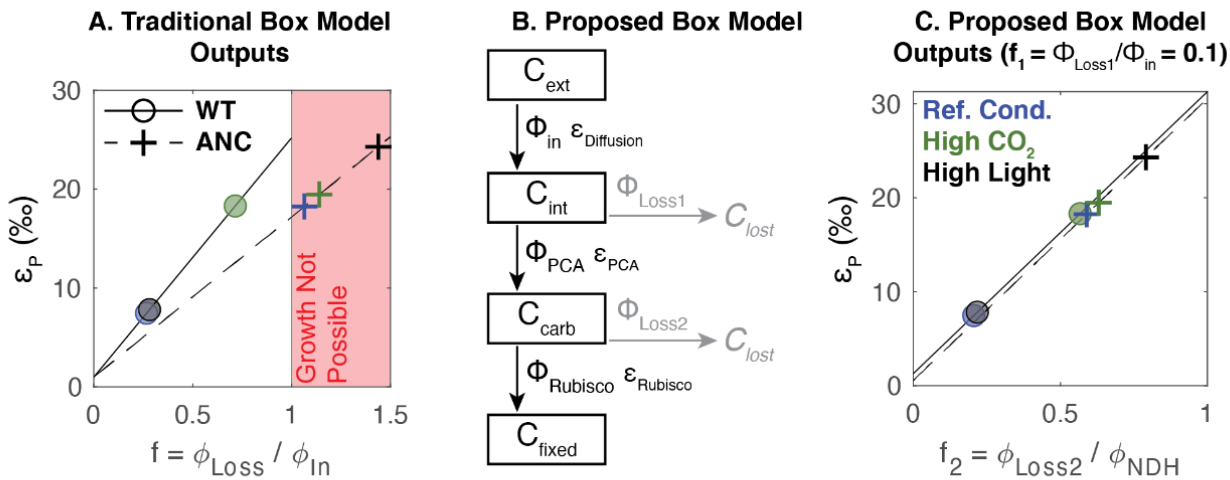

Figure 4: Proposed box model based on experimental results. A) Experimental results (circles and crosses) plotted onto traditional box model outputs (dashed and solid lines) for WT and ANC respectively. Uncertainties are smaller than data points. Colors indicate growth conditions: blue = reference condition (0.05% pCO_2 (v/v), 120 µE); green = high CO_2 (5% pCO_2 (v/v), 120 µE); black = 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) Proposed box model architecture, with main carbon pools of interest in boxes. Subscripts indicate external (*ext*), internal (*int*), carboxysome (*carb*), and fixed (*fixed*) carbon pools. Fluxes are denoted by Φ where subscripts indicate fluxes into the cell (*in*), out of the cell (*Loss1*, *Loss2*), into the carboxysome (*PCA* for *P*owered *C*arbonic *A*nhydrase), and into fixed biomass *(Rubisco*), each with a corresponding isotopic fractionation denoted with ε. Loss fluxes were assumed to have no isotopic fractionation. In this proposed model, f_1 is defined as Φ_{Loss1}/Φ_{in}, and f_2 is defined as Φ_{Loss2}/Φ_{PCA}. See text for model assumptions. C) Experimental results plotted onto proposed box model outputs for f_1 = 0.1; colors and 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 Supplemental for full solution and results; only results for f_1 = 0.1 are shown. All analyses were performed using MATLAB and Statistics Toolbox (vR2020b).

We therefore modified the traditional box model by making a distinction between carbon in the cytosol (C_{int}) and carbon in the carboxysome (C_{carb}), and by adding an additional path whereby carbon can be lost from the carboxysome (Φ_{Loss2}, Figure 4B). In this modified model, external C_i enters the cell (flux Φ_{in}) where it can either leak out (Φ_{Loss1}) or undergo active hydration (flux Φ_{PCA}, where *PCA* denotes *P*owered *C*arbonic *A*nhydrase). Intracellular C_i can then enter the carboxysome, where it is either fixed (flux $\Phi_{Rubisco}$) or ultimately leaks out of the cell (flux Φ_{Loss2}). We made similar simplifying assumptions as the traditional box model: i) an infinite supply of external carbon, ii) no isotopic fractionation for carbon lost from the cell, iii) Φ_{in} has the isotopic fractionation associated with $\varepsilon_{Diffusion}$, and iv) the system is at steady state. We did not add an explicit term for light energy used to power C_i uptake. Instead, the model included an energized CA (denoted PCA) and its associated isotopic fractionation as free parameters. In modeling each strain, we used the appropriate $\varepsilon_{Rubisco}$ measurements (Table 1). We do not know the true value for ε_{PCA}, but used a value of 30‰ similar to recent models of a one-way CA in eukaryotic algae (62). For comparison with the traditional model, we plotted Figure 4C with f_1 = 0.1 so that it could be represented in two dimensions. However, in this updated model, each value of ε_p corresponds to a set of feasible f_1 and f_2 values that fall along a line (Figure S7 and S8 for WT and ANC respectively). Therefore, our model constrains but does

not uniquely determine f_1 and f_2, nor does it allow for estimation of external C_i levels precisely because many pairs of f_1 and f_2 values can produce the same ε_p value.

With the addition of a powered CA and an additional loss term, the model was able to rationalize our experimental data of $\varepsilon_p > \varepsilon_{Rubisco}$ with leakage values compatible with cell growth ($f_2 < 1$) (Figure 4C). Our model results implied that, overall, ANC lost more carbon than WT at the branch point before rubisco (Φ_{Loss2}); i.e. even though carbon was present in the cell, it could not be fixed by the ancestral Form 1B rubisco because of its slower carboxylation rate. The excess amount of CO_2 available then allowed rubisco's kinetic isotope effect ($\varepsilon_{Rubisco}$) to be expressed. In addition, the model indicated that this effect would be exacerbated at high light (Figure 4C), which is consistent with the low ANC doubling rates seen in this condition (Figure 2). These results implied that at high light, the powered CA was delivering high amounts of CO_2 to both the WT and ANC rubisco. WT was able to keep up with this flux, which was reflected in its fast growth rate (Figure 2) and no change in ε_p vs. the reference condition (Figure 3). However, ANC was not, which led to its slowest growth rate (Figure 2), and highest ε_p values across all conditions.

These results showed that other processes relevant to the CCM, in addition to rubisco, can play an important role in ε_p values. While our approach is highly idealized and relies on a minimum set of fractionating processes associated with carbon fixation in Cyanobacteria (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 could 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, as much as or more than it informs ambient environmental CO_2 concentrations.

Consequences for understanding the evolution of carbon-fixing metabolism

The traditional box model used to describe ε_p values observed in the biomass of oxygenic photoautotrophs can produce ε_p values ranging from $\varepsilon_{diffusion}$ to $\varepsilon_{Rubisco}$ (Figure 1). However, we and others have observed anomalous ε_p values exceeding $\varepsilon_{Rubisco}$ in a variety of modern and synthetic organisms.Over Earth history, ε_p values inferred from the carbon isotope record sometimes exceed the largest measured modern $\varepsilon_{Rubisco}$ (21, 34). Explaining these diverse observations has moved us and others to posit the presence of additional fractionating processes beyond rubisco carboxylation contributing to measured ε_p values. One reasonable explanation is the presence of a one-way CA reaction, which carries a large kinetic isotope effect of ≈30‰ (53, 55–58).

This notion was first proposed more than 20 years ago based on measurements of cyanobacterial cultures with ε_p values exceeding 30‰ (63). Recent measurements of $\varepsilon_p > \varepsilon_{Rubisco}$ motivated similar conclusions regarding multiple fractionating processes in eukaryotic algae (17). In these diverse experiments, anomalous ε_p values were observed during relatively slow growth; in (63) $\varepsilon_p > \varepsilon_{Rubisco}$ occurred early in the growth curve as cells were acclimating to fresh culture media, in (17) $\varepsilon_p > \varepsilon_{Rubisco}$ 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 1) to include more physiological detail relating to the presence of a CO_2 concentrating mechanism.

Prior studies also worked to account for issues related to growth physiology—specifically growth rate, cell shape and size— to adapt the C3 plant model to unicellular algae (25). However, that landmark study measured ε_p in wild-type Cyanobacteria—the same wild-type strain studied here—and, unlike this study and (39), found ε_p to be roughly constant independent of environmental pCO_2 and growth rate. Popp and coauthors hypothesized that this observed independence stems from the large surface area to volume ratio (SA/V) of Cyanobacteria, which was taken to imply much faster passive CO_2 uptake (scaling with SA) than fixation (scaling with V). Because cyanobacterial ε_p was constant ≈17‰ and less than known cyanobacterial $\varepsilon_{Rubisco}$ values, additional fractionating factors were not needed to explain ε_p, even though some active transport processes in Cyanobacteria were known at the time; the simple linear relationship between pCO_2 and ε_p in C3 plants appeared to hold up in algae and Cyanobacteria as well.

This linear relation between pCO_2 and ε_p does not hold for our ANC strain (Figure 3). We emphasize that ANC is not a true ancestral Cyanobacteria; rather it is a chimeric construct—a modern strain saddled with a Precambrian enzyme for its carbon fixation. This reconstructed ancestral rubisco is characterized by slower carboxylation kinetics (38) and a much lower $\varepsilon_{rubisco}$ (Table 1). In ANC, we observed anomalous ε_p values exceeding $\varepsilon_{Rubisco}$ in all growth conditions, but especially in high light (Figure 3). As high light consistently slowed growth, induced chlorosis (yellowing of cultures, Figure S9) 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 (40, 64). 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 (i) the inferior kinetics of the ancestral enzyme (Table 1) and (ii) potentially a partial incompatibility with the modern CCM, e.g. sub-optimal recruitment to the carboxysome (65).

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 one-way hydration of CO_2, which has a large calculated isotope effect (53, 55–58). Cyanobacteria have been shown to take up CO_2 independently of HCO_3^- (46). In model Cyanobacteria, this activity is due to the Cup proteins (CupAS/B, also known as Chp proteins), which bind to the NADH-dehydrogenase-like complex (NDH-1) of Cyanobacteria (51, 66). The NDH-1 complex is involved in light energy capture via photosynthetic electron transport and cyclic electron flow around photosystem I (51) and, moreover, CO_2 uptake is stimulated by light alone and abrogated by inhibitors of photochemical electron transport (59). Not only has CupA been shown to carry a key Zn^{2+} in a domain resembling a carbonic anhydrase (51), but the *cupA* gene is induced under low CO_2 conditions (66). In order for CO_2 uptake to drive the CCM and promote CO_2 fixation, it would need to

482 produce a high, non-equilibrium HCO_3^- concentration in the cytoplasm (9, 11). We and others
483 therefore assumed that the complex of NDH-1 and CupAS/B (termed NDH-$1_{3/4}$ in the literature)
484 couples light energy to the one-way hydration of CO_2 to HCO_3^- at a carbonic anhydrase-like
485 active site (51).

486 It is readily apparent that $\varepsilon_{Rubisco}$ does not set an upper bound on ε_p, nor does it predict
487 which strains will have larger ε_p values *in vivo* (Figure 3). This inference was only possible
488 because we measured the isotope fractionation due to the ancestral rubisco ($\varepsilon_{Rubisco}$) and
489 compared it to ANC strain biomass (ε_p), in contrast with the study of (39), which only measured
490 ε_p. While our ANC ε_p values (≈18-24‰) fell within the range of ε_p values derived from the carbon
491 isotope record (22), they exceeded the measured $\varepsilon_{Rubisco}$ (Figure 3). As such, the relative
492 consistency of ANC ε_p values does not indicate that the traditional box model is applicable
493 across geologic time as claimed in (39). Rather, a model including some additional fractionating
494 process is required to explain our observation that $\varepsilon_p > \varepsilon_{Rubisco}$ in ANC. Attention has been paid
495 to outliers where ε_p exceeds $\varepsilon_{Rubisco}$ precisely because they violate the assumptions underlying
496 the dominant model used to interpret the carbon isotope record (17). In addition, ANC $\varepsilon_{Rubsico}$
497 (17.23 ± 0.61‰) is anomalously low; not only is it ≈8‰ less than WT $\varepsilon_{Rubsico}$ (25.18 ± 0.31‰) but
498 it is among the lowest measured rubisco KIEs. However, only thirteen unique rubisco KIEs have
499 been measured thus far (for recent review see (18)) while ≈300 distinct rubiscos have been
500 kinetically characterized (7, 67).

501 The carbon isotope record has been used to reconstruct pCO_2 measurements for the
502 vast majority of Earth's history (32, 33) because more direct observations of the past
503 atmosphere from ice only extend back ≈1 million years (31). The large carbon isotope
504 fractionation between C_i and C_o (imparted by rubisco) observed in modern environments
505 roughly matches the carbon isotope differences between carbonates and kerogen-rich rocks in
506 the geological record. Because of this correlation and prior work, it has been assumed that
507 proxies calibrated on modern oxygenic photoautotrophs over short timescales can largely be
508 applied to ancient samples to infer paleo-pCO_2 concentrations. However, our study suggests
509 that the carbon isotopic fractionations observed in both modern environments and throughout
510 the geological record reflect not just the environmental abundance of CO_2 and/or the rubisco
511 present, but also the operation of C_i uptake processes like the NDH-1 complex discussed
512 above.

513 As shown in the companion paper by Avi Flamholz and coauthors, a CCM is not
514 required in the early atmosphere, where a variety of proxies generally agree that pCO_2 was very
515 high (up to ≈0.8 bar) (68). However, Flamholz et al. found that expression of carbonic
516 anhydrases or C_i uptake systems greatly improved autotrophic growth in intermediate CO_2
517 levels (≈1% partial pressure)—levels that are thought to have been important for perhaps much
518 of Precambrian time (69). These results indicated that oxygenic photoautotrophs expressing
519 energized CO_2 uptake may have arisen relatively early in Earth's history, and rubisco has likely
520 operated in concert with some sort of CCM over Proterozoic time (9). Therefore, a uniformitarian

framework is unjustified: a model that omits all aspects of the CCM (e.g. the traditional box model) might apply early in Earth history if pCO_2 was sufficiently high, but would not necessarily apply to ancient or modern Cyanobacteria they all express complexes catalyzing light-driven, energized CO_2 uptake (or any other processes associated with large carbon isotope fraction).

A carbon isotope model that engages more fully with photosynthetic physiology, i.e. one which includes some representation of the CCM (Figure 4), 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). Notably, the model proposed here represents only a first step in this direction as it substantially simplifies the bacterial CCM (11); a similar statement applies to box models of Eukaryotic algae, which also express complex CCMs (17, 70). Future work on carbon isotope fractionation by cyanobacteria should grapple in more detail with photosynthetic physiology, including the separate uptake of C_i from external CO_2 and HCO_3^- pools, integration of both light and dark reactions, and effects of nutrient limitation. As mechanistic biochemical understanding of cyanobacterial Ci uptake improves (51), 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 the model and improve our ability to understand environmental and evolutionary changes to the carbon cycle across Earth history.

Carbon fixation was a fundamental challenge that autotrophs overcame early in the history of Earth's biosphere (8). These early processes were recorded in some fashion in the carbon isotope record, but robust interpretation of this record must take into account that the carbon cycle is an amalgam of both environmental changes and evolutionary processes, mediated by physiology. As a starting place, Earth scientists often apply uniformitarian assumptions (i.e. assuming that physical and chemical processes behave the same now as they did billions of years ago) in order to reason about the past. Such an approach is powerful but these assumptions are challenged by biological processes that undergo substantial evolution on geologic timescales. Recent work has used statistical inference to reconstruct ancestral gene sequences and resurrect ancient proteins in order to study biological evolution over geologic timescales; and was used to reconstruct the ancestral rubisco studied here (38). Here, and in the companion paper by Flamholz et al., we took a "synthetic biological" approach, constructing modern 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 (Flamholz et al.) 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. However, 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

563 Ancestral Rubisco enzyme sequences were previously reported and characterized by Shih et al.
564 (2016) (38). Briefly, for both the large subunit (LSU) and small subunit (SSU) of Rubisco,
565 encoded by *rbcL* and *rbcS* respectively, the most recent common ancestor (MRCA) for Form 1A
566 (α), 1B (β), and 1A/B (α/ β) clades were predicted from independently derived phylogenetic
567 trees for RbcL and RbcS containing a broad diversity of Form 1A and 1B Rubisco (>100
568 sequences). Maximum-likelihood algorithms were used to reconstruct the most probable
569 ancestral sequence for each clade. Ancestral sequences were then expressed in *Escherichia*
570 *coli* and purified, and enzyme kinetics were measured.
571
572 **ANC strain generation**
573 The 'ANC' strain studied here was generated by replacing the native large and small Rubisco
574 subunits (cbbL and cbbS respectively) of the parent strain (*Synechococcus elongatus* PCC
575 7942) with the reconstructed β ancestral cbbL and cbbS sequences. The NS2-KanR ('WT'
576 strain) was generated by inserting a KanR cassette into neutral site 2 (NS2) (GenBank:
577 U44761.1). *Synechococcus elongatus* PCC 7942 were transformed using the approach of
578 Golden and Sherman (1984) (71). Briefly, cultures were grown to OD750nm = 0.5. Cultures
579 were centrifuged at 18,000 x *g* for 2 minutes. Pellets were washed with 100 mM $CaCl_2$ and spun
580 again at 18,000 x *g* for 2 minutes. Pellets were resuspended in BG-11 media followed by
581 addition of plasmid and grown for 16 hours in the dark at 30°C. Transformants were then plated
582 onto BG-11 + KAN100 agar plates and placed under 100 µE of light at 30°C. Single colonies
583 were then genotyped by PCR amplification of the Rubisco locus followed by sequencing. Table
584 S1 lists plasmids and primers used in this study.
585
586 **Growth conditions**
587 For ambient CO_2 growth, NS2-KanR and β Ancestral Rubisco-KanR strains were grown in
588 quadruplicate in a photobioreactor (Photon Systems Instruments - MC 1000) at the University of
589 California, Berkeley (UC Berkeley) for four biological replicates total. Cultures were grown in
590 buffered BG-11 media with 50mM HEPES at pH 8. Cultures were inoculated at a starting
591 OD720nm = 0.015 and cultivated at 120 µE, 30°C, and bubbled with ambient air. High CO_2
592 growth was performed using the same conditions as ambient growth with the exception of
593 placing the photobioreactor in a 5% CO_2 chamber (Percival AR22L) and bubbling in air from the
594 chamber. High light growth was performed using the ambient conditions above with the
595 exception of using 500 µE for light intensity. Cells were harvested by centrifugation at 6000 x *g*
596 for 20 minutes at 4°C. Decanted pellets were then flash frozen with liquid N_2 and lyophilized
597 overnight with the Millrock Technology Model BT85A freeze dryer. Doubling time was calculated
598 by fitting the exponential phase of growth (*k*) using a Markov Chain Monte Carlo (MCMC)
599 approach, using the generic model y = a*EXP(k*x)+b. Growth curves displayed in Figure 2 were
600 smoothed with a rolling median (*n* = 12) to remove errant readings caused by bubbles advected
601 in front of the detector. See Supplemental for more information.
602
603 **Carbon isotope analysis**
604 Carbon isotope data is reported using delta notation ($\delta^{13}C$) in units of per mille (‰) where $\delta^{13}C$
605 $= [(^{13}C/^{12}C)_{sa}/(^{13}C/^{12}C)_{ref}-1]*1000$, where the subscripts 'sa' and 'ref' denote sample and
606 reference respectively. The reference used is the Vienna Pee Dee Belemnite (VPDB). $\delta^{13}C$

values of cyanobacterial cells were measured on an EA-IRMS (Elemental Analyzer Isotope Ratio Mass Spectrometer; Costech Thermo Delta-V) at the California Institute of Technology (Caltech) in Pasadena, CA. Each biological replicate was run four times with two different isotope standards – urea (-27.8‰) and sucrose (-10.45‰). A suite of urea and sucrose standards were run at the beginning, middle, and end of run for sample bracketing and to assess drift throughout the run. An average $\delta^{13}C$ and standard error were calculated and reported for each biological replicate (see Supplemental for more information). The $\delta^{13}C$ of the starting CO_2 gas was measured on the Thermo Mat 253 Ultra at Caltech. The CALT-2049C standard was used, which has a $\delta^{13}C_{VPDB}$ value of -3.62‰. CO_2 gas from high pCO_2 experiments was sourced from a CO_2 tank, while the CO_2 gas in ambient pCO_2 experiments was distilled from ambient lab air through cryogenic distillation at Caltech. ε_p, the carbon isotope fractionation between CO_2 gas and bulk cyanobacterial cells, was calculated as $(\alpha_{CO2/bio} - 1)*1000$, 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 this in contrast to other isotope literature where ε_p is calculated as $\alpha_{bio/CO2} - 1)*1000$, which would cause the positive values in this study to be negative. In this study, more positive ε_p values indicate more C-13 depleted; see Supplemental for more detail.

Rubisco KIE assay

*Syn*6301 and β-MRCA Rubisco were purified according to previous methodologies (72, 73) at University of California, Davis and then shipped on dry ice to Caltech. Clarified lysate from a BL21 DE3 Star *E. coli* culture expressing Rubisco was subjected to ammonium sulfate precipitation, at the 30-40% cut for *Syn*6301 and at the 40-50% cut for β-MRCA, followed by anion exchange chromatography and size exclusion chromatography. We then used the substrate depletion method to measure the KIE of the *Syn*6301 and β-MRCA Rubiscos ($\varepsilon_{Rubisco}$), as used previously in similar studies (41–44). Briefly, an assay mix of HCO_3^-, bovine carbonic anhydrase, Rubisco, ribulose 1,5-bisphosphate (RuBP), $MgCl_2$, bicine, and dithiothreitol (DTT) 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 (42). See Supplemental for more information.

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, 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).

References

Bibliography

1. 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).

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

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

10. 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).

11. 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).

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

13. 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).

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

15. 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).

16. 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).

17. 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.

18. A. K. Garcia, C. M. Cavanaugh, B. Kacar, The curious consistency of carbon biosignatures over billions of years of Earth-life coevolution. *ISME J.* (2021) https:/doi.org/10.1038/s41396-021-00971-5.

19. 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).

20. 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).

21. 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).

22. 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).

23. M. Schoell, F. W. Wellmer, Anomalous 13C depletion in early Precambrian graphites from Superior Province, Canada. *Nature* **290**, 696–699 (1981).

24. D. J. Des Marais, Isotopic evolution of the biogeochemical carbon cycle during the precambrian. *Reviews in Mineralogy and Geochemistry* **43**, 555–578 (2001).

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

26. 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).

27. K. H. Freeman, J. M. Hayes, Fractionation of carbon isotopes by phytoplankton and estimates of ancient CO2 levels. *Global Biogeochem. Cycles* **6**, 185–198 (1992).

28. G. H. Rau, T. Takahashi, D. J. Des Marais, Latitudinal variations in plankton delta 13C: implications for CO2 and productivity in past oceans. *Nature* **341**, 516–518 (1989).

29. 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).

30. 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).

31. 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).

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

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

34. M. Schidlowski, A 3,800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature* **333**, 313–318 (1988).

35. A. K. Garcia, *et al.*, System-level effects of CO2 and RuBisCO concentration on carbon isotope fractionation. *BioRxiv* (2021) https:/doi.org/10.1101/2021.04.20.440233.

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

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

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

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

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

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

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

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

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

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

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

47. 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).

48. 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).

49. 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).

50. 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).

51. J. M. Schuller, *et al.*, Redox-coupled proton pumping drives carbon concentration in the

photosynthetic complex I. *Nat. Commun.* **11**, 494 (2020).

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

53. 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).

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

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

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

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

58. 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).

59. 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).

60. 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).

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

62. 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).

63. 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).

64. 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.

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

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

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

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

69. P. W. Crockford, *et al.*, Triple oxygen isotope evidence for limited mid-Proterozoic primary productivity. *Nature* **559**, 613–616 (2018).

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

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

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

820 73. D. M. Banda, *et al.*, Novel bacterial clade reveals origin of form I Rubisco. *Nat. Plants* **6**, 1158–1166
821 (2020).