From Photosynthesis to Detoxification: Microbial Metabolisms Shape Earth's Surface Chemistry

Thesis by
Renée Zurui Wang

In Partial Fulfillment of the Requirements
for the Degree of
Ph.D. in Geochemistry

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
(Defended September 29, 2023)

© 2023

Renée Zurui Wang
ORCID: 0000-0003-3994-3244

ACKNOWLEDGEMENTS

So much has happened over the six years of my Ph.D. When I first started at Caltech in July 2017, I reasonably expected to leave with a doctorate… but I couldn't have guessed that I would leave with a husband and son as well! These past six years have been critical to my growth and development as a scientist, as well as my personal maturation as a young adult. Therefore, many acknowledgements are in order.

The most impactful lessons that I learned at Caltech were: i) You do not have to be a genius to be a scientist; often what distinguishes the great scientist from the average is steady, consistent work; and ii) There are many ways to be a successful scientist – there are myriad ways to ask and investigate questions, each with their own strengths and weaknesses. I am deeply grateful to those who taught me these lessons, modeled through their own behavior. These include my advisors, Dianne Newman and John Eiler, and Woody Fischer, who I also worked with extensively. Dianne's child-like curiosity and boundless enthusiasm for science have inspired me to approach long-standing geochemistry questions from a more expansive perspective, fostering novel approaches and modes of inquiry. John's hands-off approach gave me the autonomy to explore and develop my research independently while also providing invaluable guidance and support when needed. Woody's unwavering enthusiasm and nonjudgmental attitude towards my scientific ideas encouraged both my intellectual independence and the continual generation of new ideas. I am also incredibly grateful for their thoughtful and dedicated advising through two historic events that my Ph.D. overlapped with, the COVID-19 pandemic and the Black Lives Matter movement – they went above and beyond the typical expectations of a Ph.D. advisor during this historic time. I am also grateful to Alex Sessions, who has always gamely tried to address the (seemingly) random scientific questions I barge into his office with; I am particularly grateful for his insight on data I collected during my internship at Los Alamos National Labs (LANL). I am also grateful to George Rossman, who gave me plates and other household goods when I first moved to Pasadena, and whose door is always open to talk about anything interesting at all. I am also grateful to other P.I.s that I worked with during my PhD, including Patrick Shih at UC Berkeley, and Marie Kroeger and Michaeline Albright at LANL – they were excellent collaborators and encouraging mentors despite the physical distance.

In addition, I could not have come this far without the dedicated, hands-on advising of two postdoc mentors – Avi Flamholz and Zach Lonergan. There is so much one needs to learn for the actual, day-to-day task of being a scientist, and Chapters 2 and 4 could not have happened without their respective mentoring. I am also grateful for other postdocs and staff scientists who have helped me along the way, including Peter Jorth, Will DePas, Megan Bergkessel, Melanie Spero, Lisa Racki, Chelsey VanDrisse, Darcy McRose, Daniel Dar, Elise Wilkes, Amy Hofmann, Nami Kitchen, Fenfang Wu, Nithya Thiagarajan, Danielle Monteverde Potocek, and Miquela Ingalls. The same goes for fellow graduate students who have provided both friendship and intellectually stimulating conversations, including Hao Xie, Guannan Dong, Madison Douglas, Will Palfey, Newton Nguyen, Yayaati Chachan, Aida Behmard, Usha Lingappa, Grayson Chadwick, David Basta, Scott Saunders, Elena Perry, Lucas Meirelles, Lev Tsypin, John Ciemniecki, Steve Wilbert, Elise Tookmanian, Shaelynn Silverman, Elliott Mueller, and Sarah Zeichner.

Looking back, I am incredibly grateful for my undergraduate experience and the mentors there who first started cultivating the skills I would need to succeed at Caltech. This includes Josh West, who gave me the freedom to explore my intellectual ideas (by allowing me to tinker in his lab in my free time and treating me as an equal group-member during his weekly lab meetings), but also pushed me to think critically about my work and present my ideas to a scientific audience (i.e. poster presentations at AGU and turning my senior thesis into a manuscript for submission). I am also incredibly grateful to Will Berelson, who really took me under his wing and took seriously my ideas on making the department better, including starting an alumnus mentoring program and allowing me to turn a forgotten basement room into a hang-out lounge for the Earth Science undergrads. In addition, there was no way I could have learned any basic lab or research skills without the dedicated, one-on-one mentoring of a few amazing graduate students, including Paulina Piñedo-Gonzalez, Joyce Yager, and Kirstin Washington. I also have to thank the USC Earth Science graduate students as a whole, who took an incredibly eager undergraduate Renée under their wing by inviting me to their social events, helping me when I was in lab on the weekends (because I thought lab work was more fun than other things one could do as a USC undergrad) and making me feel comfortable asking questions at department seminars and in class. I am also extremely grateful to the USC Earth Science department as a whole – it was a wonderfully welcoming and friendly place to be an undergraduate, and I would not have pursued a Ph.D. if I did not have such a great undergraduate academic and research experience.

Acknowledgements are also due for those outside of research who have supported me. These include those I swam with on the Caltech Master's Swim Team, including Alice Wang, Suzy Dodd, Linda Simons, Nikki Takarabi, Alexandra Greenbaum, and Emma Dodd. I especially want to thank Linda Simons, whose own marathon ocean swimming inspired my own. I also want to thank the Los Alamos Master's Swim Team, which provided a home-away-from-home while I was there for my short internship. I also want to thank John Grotzinger, Mike Brown and Mike Brewer, who helped advised the Caltech Triathlon Club, and graduate students Soichi Hirokawa, Alistair Hayden, and Newton Nguyen who put so much effort into starting and organizing the club. I also am incredibly grateful to those at the CCID that I worked with for the Asian Pacific Islander Desi American (APIDA+) club that I co-founded with undergrad Abigail Jiang, including Alison Tominaga, Monique Thomas, Taso Dimitriadis and Yazmin Gonzalez, as well as those in the Office of Student Experience that I worked with as a Residential Associate including Isabel Peng and Felicia Hunt. I am also grateful for the Pasadena Buddhist Temple, particularly Reverend Gregory Gibbs and his wife, Kyoko, who helped make my husband and I feel welcome once in-person services resumed after the pandemic.

Finally, I am incredibly indebted to all the people who have helped me get through the final hurdle of my PhD – pregnancy, birth and postpartum. These include Melanie Spero, Rachel Shipps, Claire Bucholz, Annelise Trujillo Reny, Nami Kitchen, my advisors, my high school friends who are also having children, and the USC Arcadia Hospital lactation support group. I am incredibly grateful for all their advice on how to juggle a career with pregnancy, birthing, breastfeeding, and parenting, as well as being able to commiserate with them on the less-than-fun aspects of all those things. And, of course, deep thanks are due to family members, who have always been there to help.

Finally, I must thank my life partner Charles Guan, who I met here at Caltech, married on April 2, 2022, and with whom I welcomed a baby boy, Cohota Hongyi Guan, into the world on May 2, 2023. I could write an entire thesis on how much Charles means to me, so for brevity I will just say that I could not have completed this Ph.D. without his support and encouragement, not only as a scientist but as a husband and father as well. I know we will one day be able to look back at this crazy period of our lives and laugh about how we thought it was a good idea to finish our Ph.Ds., get pregnant, give birth, take care of an infant, and start new jobs all at the same time. I wouldn't have done it with anyone else!

ABSTRACT

Earth's chemistry, through geologic time and in the present, is inextricably linked with biologically mediated reactions. All major elemental cycles on Earth's surface have arisen from two competing processes – life shaping its chemical environment through the evolution of key biochemical pathways, and the environment constraining metabolism by dictating which reactions will occur. Understanding this complicated interplay motivates the research presented in this thesis, which studies this phenomenon over two major elemental cycles – the modern Nitrogen (N) and ancient Carbon (C) cycle.

Chapters One and Two focus on the evolution of ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco), the enzyme that catalyzes the key carbon fixation step in modern oxygenic photosynthesis. This reaction also imparts a large kinetic isotope effect (KIE) that causes the fixed carbon to be relatively depleted in natural abundance ^{13}C compared to its substrate; this isotopic fingerprint can be seen in both the modern C cycle and in rock records recording the ancient C cycle. Therefore, this KIE has been used both *in vitro* (outside the cell) by biochemical models to rationalize rubisco's reaction mechanism, and *in vivo* (in the cell) as a proxy for environmental CO_2 concentrations in the past and present. However, both the *in vitro* and *in vivo* measurements are calibrated using modern organisms even though rubisco and oxygenic photosynthesis have undergone profound evolution over geologic time. Therefore, we measured the KIE *in vitro* and *in vivo* of a reconstructed ancestral Form IB rubisco dating to >> 1 Ga, and the KIE *in vitro* of a recently discovered Form I' rubisco that presents a modern analogue to ancestral Form I rubiscos prior to the evolution of the small subunit. Overall, we find that the KIEs of both rubiscos are smaller than their modern counterparts, which is surprising given that the rock record indicates overall carbon isotope fractionations *in vivo* are larger in the past. In addition, we find that models strictly based on modern organisms may not apply to the past, questioning the basic assumption that uniformitarianism can be readily applied to biological processes. However, these models can be rescued by accounting for other aspects of cell physiology.

Chapter Three focuses on disentangling the source of key metabolites, like nitrous oxide (N_2O) in the modern N cycle. Like Chapters 1 and 2, an isotopic fingerprint that measures the 'preference' of ^{15}N for the central or outer nitrogen site in N_2O ("Site Preference" or "SP") has primarily been calibrated using dissimilatory, or energy-generating, nitric oxide (NO) reductases (NORs). However, there exists a much larger and phylogenetically widespread class of NO-detoxifying enzymes; in particular, flavohemoglobin proteins (Fhp/Hmp) produce N_2O as a strategy to neutralize damaging NO-radicals in anoxic conditions. This enzyme, which generates N_2O in non-growing and anoxic conditions, may be more relevant to natural environments where N_2O production has been detected. Surprisingly, we found that Fhp imparts a distinct SP on N_2O that differs from both bacterial and eukaryotic NORs, and that this value better aligns with existing *in situ* measurements of N_2O from soils. In addition, we find that in strains with both Fhp and NOR, the Fhp signal dominates when cells are first exposed to high concentrations of NO in oxic conditions while growing before being shifted to an anoxic, non-growing state. Therefore, in addition to telling us 'Who's there,' the SP fingerprint may also be able to tell us something about cell physiology *in vivo*. We propose a new framework for interpreting the source of N_2O based on SP values.

PUBLISHED CONTENT AND CONTRIBUTIONS

Wang, Renée Z, Robert J Nichols, Albert K Liu, Avi I Flamholz, Juliana Artier, Doug M Banda, David F Savage, John M Eiler, Patrick M Shih, and Woodward W Fischer. 2023. "Carbon Isotope Fractionation by an Ancestral Rubisco Suggests That Biological Proxies for CO2 through Geologic Time Should Be Reevaluated." Proceedings of the National Academy of Sciences of the United States of America 120 (20): e2300466120. https://doi.org/10.1073/pnas.2300466120.
- R.Z.W. designed research; performed research; analyzed data; wrote paper.
- Adapted here as Chapter 2

Wang, Renée Z., Albert K. Liu, Douglas M. Banda, Woodward W. Fischer, and Patrick M. Shih. 2023. "A Bacterial Form I' Rubisco Has a Smaller Carbon Isotope Fractionation than Its Form I Counterpart." Biomolecules 13 (4), p.596. https://doi.org/10.3390/biom13040596
- R.Z.W. developed method; performed formal analysis; writing – original draft preparation, review and editing.
- Adapted here as Chapter 3

All contents are reprinted with the permission of the National Academy of Sciences and the Multidisciplinary Digital Publishing Institute (MDPI).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1 Motivation: Geochemistry necessitates biochemistry

A true understanding of Earth's chemistry necessitates a deep understanding of biologically mediated reactions. The evolution of a few, key biochemical pathways, catalyzed by keystone enzymes, has profoundly changed Earth's chemistry over time by greatly increasing the rate of reactions, thereby introducing new geochemical fluxes over time. Concurrently, changes in environmental conditions have shaped the evolution and prevalence of these biochemical pathways as well, constraining which reactions are enzymatically feasible and which pathways dominate in certain environments. Understanding this complicated yet profound interplay is critical for a mechanistic understanding of biogeochemical cycles in both the past and the present. Therefore, this thesis seeks to deepen our understanding of Earth's biochemistry by studying two systems using related isotopic and microbiology techniques – the evolution of ribulose-1,5-bisphosphate (RuBP) carboxylase / oxygenase (rubisco) in the context of the Carbon cycle, and modern nitric oxide reductases in the context of the Nitrogen cycle.

1.2 Approach: Pairing isotope geochemistry with microbiology and biochemistry

Geochemistry as a field is often conflated with isotope geochemistry, a subdiscipline that leverages high-precision measurements of natural-abundance isotopes in the environment. This is an analytical chemistry technique driven by necessity – due to the intersecting nature of Earth's chemistry, many reactions can generate and consume common metabolites, like CO_2. But, by looking past the nominal molecular formula and looking closely at its isotopic composition ($^{13}C^{16}O_2$ vs. $^{12}C^{16}O_2$ vs. $^{12}C^{16}O^{18}O$, etc.), another dimension of information can be added that may help us disentangle these overlapping pathways. Due to small differences in potential energy that isotopic substitution confers, isotopes can be used to track reaction pathways as well as provenance. Asymmetric molecules like N_2O allow us to glean even more information by looking at the where in the molecule the rare isotope sits, like $^{14}N^{15}N^{16}O$ vs. $^{15}N^{14}N^{16}O$ or the "Site Preference" (SP) of ^{15}N in N_2O.

However, though isotope geochemistry presents a powerful set of analytical tools that gives us additional information of natural environments, piecing together this information into a larger understanding of the environment requires a deep understanding of the processes that lead to isotopic enrichments, particularly for biochemically mediated reactions like the ones studied here. Engineered strains, like a Cyanobacteria with an 'ancestral' rubisco ((Wang et al. 2023); Chapter 1) allow us to systematically test specific hypotheses, and *in vitro* enzyme assays adopted from biochemistry allow us to cleanly measure the KIE of enzymes

((Renée Z Wang et al. 2023; Renée Z. Wang et al. 2023); Chapters 1-2). Understanding the genetic regulation of these enzymes then allows us to manipulate wild-type strains in experimental conditions relevant to natural environments (Chapter 3). Therefore, this thesis worked at the intersection of isotope geochemistry, biochemistry, and microbial physiology to enable a deeper understanding of Earth's chemistry.

1.3 Chapters 1-2: The most abundant enzyme on Earth

Most biochemical reactions happen far too slowly in the temperature and pressure conditions that are compatible with life; therefore, enzymes have evolved to accelerate these reactions by well over a million-fold (Cooper 2000), facilitating new geochemical fluxes in the process. Nowhere is this more apparent than in the evolution of rubisco, which catalyzes the key carbon fixation step in the Calvin Benson Bassham (CBB) Cycle as part of modern oxygenic photosynthesis. Due to the ecological dominance of the CBB cycle, rubisco is the most abundant protein on Earth today (Bar-On and Milo 2019) and gross primary productivity (GPP), largely catalyzed by rubisco, represents the single most massive organic carbon flux in the modern carbon cycle with ≈120 Gt C yr^{-1} in terrestrial (Beer et al. 2010) and ≈100 Gt C yr^{-1} in marine environments (Bar-On and Milo 2019; Field et al. 1998).

Earth scientists also leverage the biochemistry of rubisco to study past environments – rubisco imparts a large kinetic isotope effect (KIE) where the fixed carbon is relatively depleted in ^{13}C compared to the CO_2 substrate (Farquhar et al. 1989). This KIE is inherited in biomass by photoautotrophs utilizing rubisco, like Cyanobacteria, and it is then preserved in the organic-rich fractions of sediments over geologic time (Schidlowski 1988). Experiments based on modern Cyanobacteria and eukaryotic algae have led to the development of a proxy where variations in the carbon isotope composition of sedimentary organic matter can be interpreted as changes atmospheric CO_2 over geologic time (Freeman and Hayes 1992) and in the present (Francois et al. 1993). These inferred paleo-CO_2 concentrations and correlation of rubisco's KIE with other biochemical parameters have then been used to rationalize the evolutionary history of rubisco, which is then used to inform bioengineering approaches to create a 'better' rubisco (Cummins et al. 2018; Savir et al. 2010; Spreitzer and Salvucci 2002; Tcherkez et al. 2006).

However, these models were based on measurements of modern organisms even though rubisco and the physiology of carbon fixation undergoes significant evolution over geological time scales. Therefore, we tested if these models held in an engineered strain that may better resemble its ancient counterparts – a Cyanobacteria with an inferred 'ancestral' rubisco dating to >>1 Ga (Chapter 1; (Wang et al. 2023)). We found that models based strictly on modern organisms did not hold up, but that these models may be rescued by accounting for Cyanobacterial physiology. In addition, our understanding of past organisms is necessarily biased by our understanding of modern ones. Though knowing rubisco's KIE is central to these models, very few measurements of rubisco KIEs exist and they cover a

limited phylogeny (for recent review see (Garcia et al. 2023)). However, novel clades of rubisco have recently been discovered (Banda et al. 2020), calling into question how accurately we can model into the past when we don't fully understand the diversity of the present. Therefore, we measured the KIE of a novel Form I' rubisco (Banda et al. 2020), which lacks the small subunit, to better characterize the variation in KIE of modern rubiscos and to help understand which characteristics of rubisco affect its KIE (Chapter 2; (Wang et al. 2023)). Surprisingly, the Form I' rubisco fractionated less than its Form I counterpart, even though the small subunit that Form I' lacks does not contain the active site; our result suggest that prior models rationalizing rubisco KIEs may not be sufficient (Tcherkez et al. 2006) and suggests novel avenues of inquiry regarding enzymes and their KIEs.

1.4 Chapter 3: Dealing with stress instead of growing

The evolution of keystone enzymes were foundational for the proliferation of microbial life on Earth – doing so allowed microbes to overcome fundamental metabolic 'chokepoints' of natural environments, like the availability of 'fixed', organic carbon as discussed above. By evolving clever biological machinery that lowers the activation energy barrier of these critical reactions, microbes have been able to work through these metabolic chokepoints. In contrast, humans use energy gained from burning fossil fuels to bypass these energy barriers instead. Perturbation of the modern nitrogen cycle clearly illustrates this point; nitrogen fixation by microbes utilizing the enzyme nitrogenase are typically the only source of 'reactive' nitrogen – nitrogen that can be utilized by organisms – in natural environments (~400 Tg yr^{-1} (Fowler et al. 2013)), but creation of reactive nitrogen through industrial processes is now on the same order of magnitude (~200 Tg yr^{-1} (Fowler et al. 2013)) with potentially far reaching consequences that are still not fully understood (Gruber and Galloway 2008).

One observation is that atmospheric concentrations of nitrous oxide (N_2O), a potent greenhouse gas (GHG), have increased by more than 20% compared to preindustrial levels (Tian et al. 2020). N_2O is unique as a GHG because its primary production and consumption pathways are biochemical (Stein 2020) – therefore, a deep understanding of the interplay between environment and metabolism is necessary for tracking and mitigating anthropogenic N_2O emissions. Currently, it is thought that excess anthropogenic reactive nitrogen, in the form of nitrate or ammonia, is stimulating dissimilatory microbial nitrogen cycle processes (for review see (Tian et al. 2020)). Specifically, microbial denitrification (an anaerobic respiratory pathway where nitrate and nitrite are sequentially reduced to dinitrogen) and nitrifier-denitrification (where ammonia is first oxidized to nitrate or nitrite before being reduced) are usually pointed to as the primary culprits of N_2O production. Nitric oxide reductases (NORs), which reduce nitric oxide (NO) to N_2O as part of these pathways, are therefore fingered as the enzymatic culprit for increased atmospheric N_2O concentrations.

However, these studies ignore the direct substrate of NORs – NO, the reactive small molecule that plays multiple roles in the context of cell physiology. Not only is it used as a signaling molecule in bacteria and eukaryotes, but it is also employed by eukaryotes in the arsenal of reactive nitrogen species (RNS) to defend against bacterial pathogens (Davis et al. 2001). In addition, NO can be generated both by specific enzymes (i.e. inducible nitric oxide synthase, iNos, in eukaryotes) and indiscriminately as a by-product of respiration (Davis et al. 2001). Therefore, cells have come up with multiple ways to confront this cosmopolitan small molecule. In other words – at times, it is much more important for cells to find a way to neutralize the oxidative threat of NO than it is to find a way to utilize it for energy-generation.

Understanding the microbial physiology of NO enabled us to look beyond NORs and think more broadly about enzymes that utilize NO – this led us to think about NO-detoxifying enzymes. We became particularly interested in flavohemoglobin proteins (Fhp/Hmp/Yhb), phylogenetically widespread proteins that protect against nitrosative stress in bacteria and yeast (Poole and Hughes 2000), reducing NO to N_2O under anoxic conditions (Bonamore and Boffi 2008; Poole and Hughes 2000). Given their phylogenetic abundance, we hypothesized that Fhp may be playing a role in generating environmental N_2O, and that this may be seen using isotopic tracers – i.e. the isotopic signature of *in situ* measurements of environmental N_2O may better match Fhp than NORs. Therefore, we made a novel isotopic measurement of N_2O produced by Fhp from three bacterial strains, found that it differed significantly from NOR, and saw that the isotopic fingerprint of Fhp better matched literature data of *in situ* soil N_2O measurements (Chapter 3; Wang et al., in prep). We also found that – in a strain that had both Fhp and NOR – which enzyme ultimately produced N_2O was stimulated in a dose-dependent manner by NO, potentially expanding the utility of this isotopic fingerprint from just telling us "Who's there" to saying something about the environment.

1.5 References

Banda, Douglas M, Jose H Pereira, Albert K Liu, Douglas J Orr, Michal Hammel, Christine He, Martin A J Parry, et al. 2020. "Novel Bacterial Clade Reveals Origin of Form I Rubisco." Nature Plants 6 (9): 1158–66. https://doi.org/10.1038/s41477-020-00762-4.

Bar-On, Yinon M, and Ron Milo. 2019. "The Global Mass and Average Rate of Rubisco." Proceedings of the National Academy of Sciences of the United States of America 116 (10): 4738–43. https://doi.org/10.1073/pnas.1816654116.

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon

Dioxide Uptake: Global Distribution and Covariation with Climate." Science 329 (5993): 834–38. https://doi.org/10.1126/science.1184984.

Bonamore, Alessandra, and Alberto Boffi. 2008. "Flavohemoglobin: Structure and Reactivity." IUBMB Life 60 (1): 19–28. https://doi.org/10.1002/iub.9.

Cooper, Geoffrey M. 2000. "The Central Role of Enzymes as Biological Catalysts." In The Cell: A Molecular Approach., 2nd ed. Sunderland, MA: Sinauer Associates. https://www.ncbi.nlm.nih.gov/books/NBK9921/.

Cummins, Peter L, Babu Kannappan, and Jill E Gready. 2018. "Directions for Optimization of Photosynthetic Carbon Fixation: Rubisco's Efficiency May Not Be so Constrained after All." Frontiers in Plant Science 9 (March): 183. https://doi.org/10.3389/fpls.2018.00183.

Davis, K L, E Martin, I V Turko, and F Murad. 2001. "Novel Effects of Nitric Oxide." Annual Review of Pharmacology and Toxicology 41: 203–36. https://doi.org/10.1146/annurev.pharmtox.41.1.203.

Farquhar, G D, J R Ehleringer, and K T Hubick. 1989. "Carbon Isotope Discrimination and Photosynthesis." Annual Review of Plant Physiology and Plant Molecular Biology 40 (1): 503–37. https://doi.org/10.1146/annurev.pp.40.060189.002443.

Field, C B, M J Behrenfeld, J T Randerson, and P Falkowski. 1998. "Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components." Science 281 (5374): 237–40. https://doi.org/10.1126/science.281.5374.237.

Fowler, David, Mhairi Coyle, Ute Skiba, Mark A Sutton, J Neil Cape, Stefan Reis, Lucy J Sheppard, et al. 2013. "The Global Nitrogen Cycle in the Twenty-First Century." Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 368 (1621): 20130164. https://doi.org/10.1098/rstb.2013.0164.

Francois, Roger, Mark A. Altabet, Ralf Goericke, Daniel C. McCorkle, Christian Brunet, and Alain Poisson. 1993. "Changes in the δ13 C of Surface Water Particulate Organic Matter across the Subtropical Convergence in the SW Indian Ocean." Global Biogeochemical Cycles 7 (3): 627–44. https://doi.org/10.1029/93GB01277.

Freeman, K H, and J M Hayes. 1992. "Fractionation of Carbon Isotopes by Phytoplankton and Estimates of Ancient CO2 Levels." Global Biogeochemical Cycles 6 (2): 185–98. https://doi.org/10.1029/92GB00190.

Garcia, Amanda K, Mateusz Kędzior, Arnaud Taton, Meng Li, Jodi N Young, and Betül Kaçar. 2023. "Effects of RuBisCO and CO2 Concentration on Cyanobacterial Growth and Carbon Isotope Fractionation." Geobiology, January. https://doi.org/10.1111/gbi.12543.

Gruber, Nicolas, and James N Galloway. 2008. "An Earth-System Perspective of the Global Nitrogen Cycle." Nature 451 (7176): 293–96. https://doi.org/10.1038/nature06592.

Poole, R K, and M N Hughes. 2000. "New Functions for the Ancient Globin Family: Bacterial Responses to Nitric Oxide and Nitrosative Stress." Molecular Microbiology 36 (4): 775–83.

Savir, Yonatan, Elad Noor, Ron Milo, and Tsvi Tlusty. 2010. "Cross-Species Analysis Traces Adaptation of Rubisco toward Optimality in a Low-Dimensional Landscape." Proceedings of the National Academy of Sciences of the United States of America 107 (8): 3475–80. https://doi.org/10.1073/pnas.0911663107.

Schidlowski, Manfred. 1988. "A 3,800-Million-Year Isotopic Record of Life from Carbon in Sedimentary Rocks." Nature 333 (6171): 313–18. https://doi.org/10.1038/333313a0.

Spreitzer, Robert J, and Michael E Salvucci. 2002. "Rubisco: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme." Annual Review of Plant Biology 53: 449–75. https://doi.org/10.1146/annurev.arplant.53.100301.135233.

Stein, Lisa Y. 2020. "The Long-Term Relationship between Microbial Metabolism and Greenhouse Gases." Trends in Microbiology 28 (6): 500–511. https://doi.org/10.1016/j.tim.2020.01.006.

Tcherkez, Guillaume G B, Graham D Farquhar, and T John Andrews. 2006. "Despite Slow Catalysis and Confused Substrate Specificity, All Ribulose Bisphosphate Carboxylases May Be Nearly Perfectly Optimized." Proceedings of the National Academy of Sciences of the United States of America 103 (19): 7246–51. https://doi.org/10.1073/pnas.0600605103.

Tian, Hanqin, Rongting Xu, Josep G Canadell, Rona L Thompson, Wilfried Winiwarter, Parvadha Suntharalingam, Eric A Davidson, et al. 2020. "A Comprehensive Quantification of Global Nitrous Oxide Sources and Sinks." Nature 586 (7828): 248–56. https://doi.org/10.1038/s41586-020-2780-0.

Wang, Renée Z, Robert J Nichols, Albert K Liu, Avi I Flamholz, Juliana Artier, Doug M Banda, David F Savage, John M Eiler, Patrick M Shih, and Woodward W Fischer. 2023. "Carbon Isotope Fractionation by an Ancestral Rubisco Suggests That Biological Proxies for CO2 through Geologic Time Should Be Reevaluated." Proceedings of the National Academy of Sciences of the United States of America 120 (20): e2300466120. https://doi.org/10.1073/pnas.2300466120.

Wang, Renée Z., Albert K. Liu, Douglas M. Banda, Woodward W. Fischer, and Patrick M. Shih. 2023. "A Bacterial Form I' Rubisco Has a Smaller Carbon Isotope Fractionation than Its Form I Counterpart." Biomolecules, April.

2. Carbon isotope fractionation by an ancestral rubisco suggests that biological proxies for CO_2 through geologic time should be reevaluated

Wang, Renée Z, Robert J Nichols, Albert K Liu, Avi I Flamholz, Juliana Artier, Doug M Banda, David F Savage, John M Eiler, Patrick M Shih, and Woodward W Fischer. 2023. "Carbon Isotope Fractionation by an Ancestral Rubisco Suggests That Biological Proxies for CO_2 through Geologic Time Should Be Reevaluated." Proceedings of the National Academy of Sciences of the United States of America 120 (20): e2300466120. https://doi.org/10.1073/pnas.2300466120.

2.1 Significance

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 considered. Our study suggests that the carbon isotope record tracks both the evolution of photosynthetic physiology as well as changes in atmospheric CO_2, highlighting the importance of considering both evolution and physiology for comparative biological approaches to understanding Earth's history.

2.2 Summary

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. Therefore, we measured 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). Surprisingly, ANC ε_p exceeded ANC $\varepsilon_{Rubisco}$ in all conditions tested, contradicting prevailing models of cyanobacterial carbon isotope fractionation. Such models can be rectified by introducing additional isotopic fractionation associated with powered inorganic carbon uptake mechanisms present in Cyanobacteria, but this amendment hinders the ability to accurately estimate historical pCO_2 from geological data.

Understanding the evolution of rubisco and the CO_2 concentrating mechanism is therefore critical for interpreting the carbon isotope record, and fluctuations in the record may reflect the evolving efficiency of carbon fixing metabolisms in addition to changes in atmospheric CO_2.

2.3 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 and energy storage. Today, and likely for much of Earth's history (Fischer et al. 2016), 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) (Bar-On and Milo 2019; Wildman 2002). But rubisco's central role in the CBB cycle and oxygenic photosynthesis poses a conundrum because it is usually considered to be a nonspecific 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 attacks (Lorimer and Andrews 1973). 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 adenosine triphosphate (ATP), reducing power, and carbon (Andrews and Lorimer 1987). The second issue concerns rubisco's maximum carboxylation rate (V_C), which is ≈7 to 10 times slower than other central metabolic enzymes (Bar-Even et al. 2011), and displays very limited variation across large phylogenetic distances (Flamholz et al. 2019).

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 been the primary carboxylating enzyme of global photosynthesis since the Great Oxygenation Event and potentially far prior (Fischer et al. 2016). It is also thought to have evolved 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 (Fischer et al. 2016).

Likely in response to these changing environmental concentrations, many aquatic photoautotrophs evolved CO_2 concentrating mechanisms (CCMs) that enhance carboxylation and suppress oxygenation by immersing rubisco in a high-CO_2 environment. Even with CCMs, the effective in vivo rates of extant rubiscos are estimated to be lower (≈1% for terrestrial and ≈15% for marine rubiscos) than the maximal catalytic rates measured at 25 °C (Bar-On and Milo 2019). Today, all known Cyanobacteria have CCMs, as do many bacterial chemolithoautotrophs, many aquatic algae and some plants (Flamholz and Shih 2020). The bacterial CCM has two main components: i) C_i pumps producing high cytosolic HCO_3^- concentrations,

and ii) coencapsulation of carbonic anhydrase (CA) and rubisco inside proteinaceous organelles known as carboxysomes (Fig. 2.1A) (Rae et al. 2013; Mangan et al. 2016; Raven and Beardall 2014). These 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 (Price et al. 2013) for review). It is unclear exactly when the bacterial CCM arose, with proposals ranging from the Proterozoic to the Phanerozoic Eon (Flamholz and Shih 2020; Riding 2006). Therefore, for up to half of Earth's history, cyanobacterial rubiscos have functioned in concert with a system that pumps C_i into and around the cell.
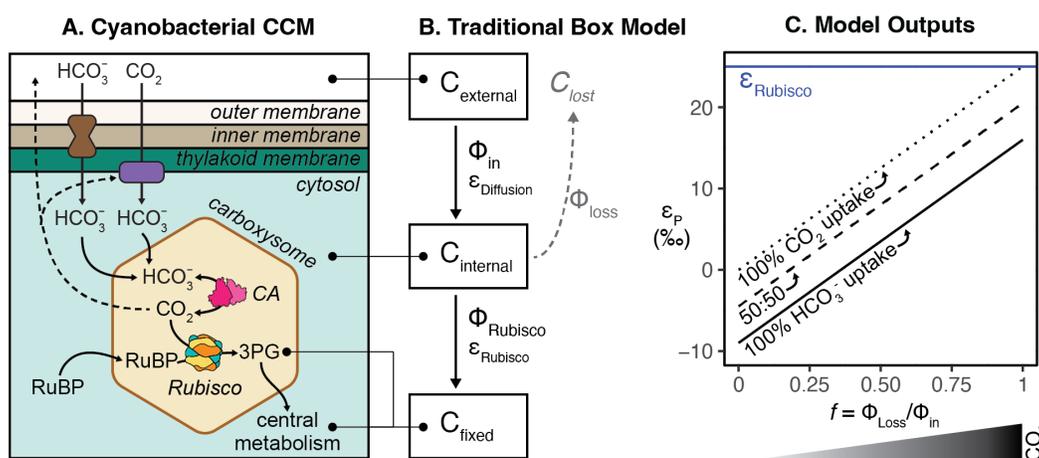
Figure 2.1 Comparing the cyanobacterial CO_2 concentrating mechanism (CCM) to the traditional box model of photosynthetic C isotope discrimination.
(A) Cyanobacterial CCMs rely on i) active C_i uptake into the cell, and ii) coencapsulation of carbonic anhydrase (CA) and rubisco within the carboxysome. 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 (Price et al. 2013) 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 (Hayes 1993; Francois et al. 1993; Park and Epstein 1960; Farquhar et al. 1982); see SI Appendix 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 $\varepsilon_P = a*\varepsilon_{equil} + f*\varepsilon_{Rubisco}$ (Eqn. 2.2), 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 (Eichner et al. 2015), which is a generalization of (Sharkey and Berry 1985) 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 $\varepsilon_{equil} = -9‰$ were used for this illustration (Mook et al. 1974). Model outputs indicate that at high external CO_2 concentrations (dark wedge under graph), there is greater net C_i leakage (larger f values) from the cell.

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 $^{13}CO_2$ (Farquhar et al. 1989), leading 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 CO_2 substrate. This effect is typically reported in delta ($\delta^{13}C$) and epsilon (ε) notation in units of per mille (‰), where $\delta^{13}C$ = [$^{13}R_{sa}/^{13}R_{ref}$ - 1]*1000 and ^{13}R is the ratio of $^{13}C/^{12}C$ in the sample or reference, respectively; see *Materials and Methods*. The difference in $\delta^{13}C$ of the CO_2 substrate and the 3-PG product is reported as $\varepsilon_{Rubisco}$ and varies between 18 and 30‰ for extant rubiscos (Wilkes and Pearson 2019; Garcia et al. 2021), with the exception of the coccolithophore *E. huxleyi* with $\varepsilon_{Rubisco} \approx$ 11‰ (Boller et al. 2011). Because autotrophs utilizing the CBB cycle synthesize biomass from 3-PG, biomass is ^{13}C-depleted compared to external C_i pools—the magnitude of this difference is called ε_p.

The KIE of rubisco, along with other and more minor processes that affect carbon isotope ratios, is recorded in the carbon isotope record, which is comprised of measurements of the relative ratios of ^{13}C to ^{12}C isotopes in C-bearing phases of sedimentary rocks over time (Schidlowski 1988). Carbon isotope data have been assembled globally from myriad of ancient environments to cover $\approx$3.8 billion years (Ga) of Earth's 4.5 Ga history (Krissansen-Totton et al. 2015). Contemporaneous C_i pools are preserved as carbonate salts (e.g., limestones and dolomites), while contemporaneous biomass and C_o pools are preserved in the organic-rich components (e.g., kerogen) of many different lithologies and are measured as rock total organic carbon (TOC) (Schidlowski 1988). There are additional fractionations associated with the preservation of biomass and C_i as rocks, so the magnitude of fractionation between rock C_i and C_o is termed ε_{TOC} and differs slightly from ε_p (Hayes et al. 1999). Therefore, if one can derive ε_p from the rock record (ε_{TOC}) and pair it with a model relating ε_p to pCO2, in principle one can infer the history of atmospheric pCO2 from the carbon isotope record.

The carbon isotope record is particularly important for constraining ancient atmospheric pCO2 (Jasper and Hayes 1990; Pagani et al. 2011) because direct observations of the past atmosphere from trapped gas in ice cores only extends back $\approx$1 million years (Higgins et al. 2015). One notable feature of the record from $\approx$3.8 Ga to the present is that rock C_o is depleted in ^{13}C by $\approx$25‰ compared to C_i (Garcia et al. 2021; Schidlowski 1988; Krissansen-Totton et al. 2015), and this offset roughly matches the KIE of extant rubiscos (Schidlowski 1988). The dominant model used to derive ancient atmospheric CO_2 from the geological record (referred to as the "C Isotope Record Model" here; *SI Appendix* and Fig. 2.11; Eqn. 2.1) reflects this observation by fixing the maximum possible fractionation of biomass to be that of rubisco:

$$\varepsilon_p = \varepsilon_f - \frac{b}{[CO_2(aq)]}$$

Equation 2.1

where ε_f is the maximum isotopic fractionation for carbon fixation and is typically set to equal $\varepsilon_{Rubisco}$, [CO$_2$(aq)] is the concentration of dissolved CO$_2$ in solution around the cells, and *b* is a fitted parameter derived from experiments (Bidigare et al. 1997). This physiological factor, *b* (‰ kg μM^{-1}), is fit from pure culture experiments of eukaryotic and bacterial algae, and encompasses all physiological effects that may affect cellular isotopic fractionation including the CCM, growth rate, cell size and geometry, membrane permeability, growth medium composition (e.g., pH, salinity, limiting nutrient), strain genetics, and physiological state (Bidigare et al. 1997; Popp et al. 1998; Laws et al. 1995; Zhang et al. 2020; Rau et al. 1996). 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 equal $\varepsilon_{Rubisco}$. Therefore, with this model framework 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 term *b* and the assumption that $\varepsilon_{Rubisco}$ sets the upper limit of ε_P directly follows from the "traditional model" (Fig. 2.1B and C and Eqn. 2.2) that was developed by measuring ε_P of plants and algae while parameters like pCO$_2$ were varied:

$$\varepsilon_p = f * \varepsilon_{Rubisco} + a * \varepsilon_{equil}$$

Equation 2.2

where *f* is a ratio describing how much C$_i$ exits vs. enters the organism (*f* = 1 is all C$_i$ that enters is lost), 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) (Eichner et al. 2015; Sharkey and Berry 1985). The diffusion isotope effect ($\varepsilon_{Diffusion}$) is considered negligible. This model (Fig. 2.1B and C and Eqn. 2.2) is therefore the physiological underpinning Eqn. 2.1 and subsequent interpretations of the C isotope record; both show a limit where the maximum ε_P is $\varepsilon_{Rubisco}$.

This traditional model was originally developed from studies of C isotope fractionation in plants (dotted line in Fig. 2.1C; all C$_i$ uptake is as CO$_2$ for plants) and was later adapted to eukaryotic and bacterial algae. The primary architecture of the traditional model stems from a seminal study by (Park and Epstein 1960) 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" (Fig. 2.1B) – 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 (O'Leary 1984)]; and ii) the carbon fixation step catalyzed by rubisco ($\varepsilon_{Rubisco}$; ≈18 to 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 entry of CO_2 into the cell ($\varepsilon_p = \varepsilon_{Diffusion}$) or CO_2 fixation by rubisco ($\varepsilon_p = \varepsilon_{Rubisco}$) (Park and Epstein 1960).

Solving the traditional model at steady state 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 (Fig. 2.1C). This allows experimentally measured values of ε_p to be used to solve for CO_2 leakage (f, Fig. 2.1C). When $\varepsilon_p \approx \varepsilon_{Diffusion}$, nearly all carbon entering the cell is used ($f \approx 0$) and 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 ($f \approx 1$, nearly all 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. (Farquhar et al. 1982) 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 roughly 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 indicates that high external CO_2 concentrations generate excess CO_2 at rubisco and ultimately cause more CO_2 to leak out of the cell than can be fixed [see *SI Appendix* and (Francois et al. 1993)].

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 (Sharkey and Berry 1985; Popp et al. 1998; Laws et al. 1995; Cassar et al. 2006; Berry 1989). 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 pCO_2 observed in plants still held. Interestingly, cyanobacterial ε_p was found to be roughly constant independent of environmental pCO_2 and growth rate (Popp et al. 1998). Because cyanobacterial ε_p values were less than known corresponding $\varepsilon_{Rubisco}$ 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 (Gimmler et al. 1990; Rotatore et al. 1992; Sültemeyer et al. 1993). Therefore, though different versions of this "traditional model" exist, all variations essentially modified the plant model by shifting the y-intercept of Fig. 2.1C 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. 2.1C). If C_i is taken up primarily as HCO_3^-, as in many algae, ε_p is shifted to lower values (solid line in Fig. 2.1C) because of the equilibrium isotopic effect (ε_{equil}) between CO_2 and HCO_3^- [$\approx -9‰$ (Mook et al. 1974)]. In Fig. 2.1C, we plot the traditional model as derived in (Eichner et al. 2015), which is an adaptation of (Sharkey and Berry 1985).

The C Isotope Record Model (Eqn. 2.1 and *SI Appendix*, Fig. 2.11) and the traditional model (Eqn. 2.2 and Fig. 2.1C) have a limit where ε_p cannot exceed $\varepsilon_{Rubisco}$. Yet, the largest ε_p values observed in the Archaean Eon exceed 30‰ (Schidlowski 1988; Krissansen-Totton et al. 2015) and also exceed all current measurements of $\varepsilon_{Rubisco}$ (Garcia et al. 2021). In addition, recent studies in dinoflagellates have shown that ε_p can regularly exceed $\varepsilon_{Rubisco}$ under certain growth conditions (Wilkes and Pearson 2019), and detailed studies of Cyanobacteria imply

that leakage estimates derived from ε_p are not physiologically possible (Eichner et al. 2015). These studies motivated updated models of algal carbon isotope fractionation that account for the isotopic fractionations associated with different C_i uptake mechanisms in order to rationalize anomalous ε_p values (Eichner et al. 2015; Wilkes and Pearson 2019).

These experiments made clear that the physiology of algae and Cyanobacteria—e.g., how they take up C_i as CO_2 or HCO_3^- and by which mechanism—affects the C isotopic content of biomass, ε_p. Further, these C_i transporters and other integral components of modern CCMs were once absent from ancient autotrophs, who used various forms of rubisco alone to grow in Archaean or Proterozoic atmospheres (Flamholz et al. 2022). Efforts to draw inferences about the ancient Earth from the C isotope record must, therefore, include some understanding of the physiology and evolution of CCMs in Cyanobacteria and eukaryotic algae (Wilkes and Pearson 2019; Flamholz et al. 2022; Hurley et al. 2021). Recent studies have attempted to address this issue by characterizing model organisms that may better resemble an ancestral counterpart, including a cyanobacterial strain lacking a CCM (Hurley et al. 2021), a strain that overexpresses rubisco (Garcia et al. 2023), and a strain expressing an inferred ancestral rubisco dating from ≈1 to 3 Ga (Kacar et al. 2017; Kędzior et al. 2022).

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" for "wild-type", see *Materials and Methods*), as well as a strain, nicknamed "ANC" for "ancestral", engineered to express an inferred ancestral Form 1B enzyme (dating to >1 Ga) as its sole rubisco (Shih et al. 2016) 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 (Shih et al. 2016). Here we study a strain where the extant rubisco was fully removed and replaced with the reconstructed ancestor. In contrast to (Kędzior et al. 2022), we also measured $\varepsilon_{Rubisco}$ of the present-day and ancestral rubiscos in vitro. We observed that: i) biomass ε_p is greater for ANC than WT for all conditions tested, even though ANC $\varepsilon_{Rubisco}$ (17.23 ± 0.61‰) is considerably smaller than WT $\varepsilon_{Rubisco}$ (25.18 ± 0.31‰); ii) ANC ε_p increases with light levels while WT ε_p increases with CO_2; iii) ANC displays a growth defect at ambient pCO_2 that is rescued at high pCO_2; and iv) ANC growth is severely inhibited in high light. Consistent with recent studies of eukaryotic algae (Eichner et al. 2015; Wilkes and Pearson 2019), ANC ε_p exceeding $\varepsilon_{Rubisco}$ in all conditions implies that the traditional box model is incomplete and additional isotope fractionations are needed to rationalize measured ε_p. In addition, modulation of ANC ε_p with light suggests that some light-powered component of the CCM is responsible for excess fractionation beyond $\varepsilon_{Rubisco}$. We posit that fractionation due to C_i uptake might explain isotopic measurements that deviate from traditional model predictions in both extant and ancient organisms.

2.4 Results and Discussion

2.4.1 Ancestral rubisco enzyme fractionates less than the modern rubisco

We measured the carbon isotope fractionations of WT and ANC rubiscos in vitro using the substrate depletion method (Guy et al. 1993; McNevin et al. 2006; Scott et al. 2004; Thomas et al. 2018). Note that there exists experimental variation in $\varepsilon_{Rubisco}$ measurements, both within and across studies, and its cause remains uncertain at present [see *SI Appendix*, section 2.9.4.1 and (Wang et al. 2023)]; so we employed the same general approach as others (the substrate depletion method) to be consistent with prior literature. 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 (Tcherkez et al 2006). We therefore expected ANC and WT $\varepsilon_{Rubisco}$ values to be the same within uncertainty because of their similar $S_{C/O}$ values (previously measured in (Shih et al. 2016)), but found that the fractionation factor ($\varepsilon_{Rubisco}$) of the ancestral rubisco (17.23 ± 0.61‰) was about 8‰ lower than that of the extant rubisco (25.18 ± 0.31‰, Table 2.1).

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 IB	17.23 ± 0.61	4.72 ± 0.14	168.7	28	49.6 ± 1.8
Modern Form IB	25.18 ± 0.31*	9.78 ± 0.48*	184.1*	53.1*	50.3 ± 2.0*

Table 2.1 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 (Shih et al. 2016). Kinetic isotope effect ($\varepsilon_{Rubisco}$, avg. ± SE) was measured in this study using the substrate depletion method (Guy et al. 1993; McNevin et al. 2006; Scott et al. 2004; Thomas et al. 2018). Carboxylation turnover under substrate-saturated conditions (V_C); Michaelis constant for CO_2 in ambient levels of O_2 (K_C^{Air}); the catalytic efficiency toward 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 (Shih et al. 2016).

2.4.2 Ancestral rubisco strain grows at ambient CO_2 concentrations

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 (Shih et al. 2016) as its sole rubisco enzyme. We then grew this strain, termed ANC, and a control strain, termed wild-type or "WT" (*Materials and 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); and 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.

Remarkably, as in (Kędzior et al. 2022), the ANC strain managed to grow in ambient pCO_2 and standard light conditions (Fig. 2.2), even though the ancestral rubisco has a V_C roughly half that of WT (Table 2.1). This implies that its rubisco enzyme is properly encapsulated in the carboxysome, since improper carboxysome formation prohibits growth in ambient air (Kerfeld et al. 2018; Price and Badger 1989). Additional characterization of the physiology of the ANC could be valuable, but our inference of proper carboxysome encapsulation is supported by several experiments and analyses as follows. First, electron micrographs of WT and ANC cells grown in ambient CO_2 and light conditions (*Materials and Methods*) showed multiple carboxysomes per cell in both strains (Fig. 2.3 and *SI Appendix*, Fig. 2.17). Rubisco density can be seen within some of the carboxysomes (Fig. 2.3C). Second, the rubisco amino acid residues necessary for protein interactions mediating β-carboxysome encapsulation were recently identified (Wang et al. 2019), and the ANC sequence retains fourteen of the sixteen residues involved (*SI Appendix*, Tables 2.8 and 2.9 and Fig. 2.18). In addition, WT and ANC strains harvested during exponential growth in the reference condition exhibit similar photosystem stoichiometry, as indicated by absorbance spectra (*SI Appendix*, Fig. 2.19). Taken together, these data indicated that carboxysomes form in ANC and the ancestral rubisco is encapsulated within these structures.
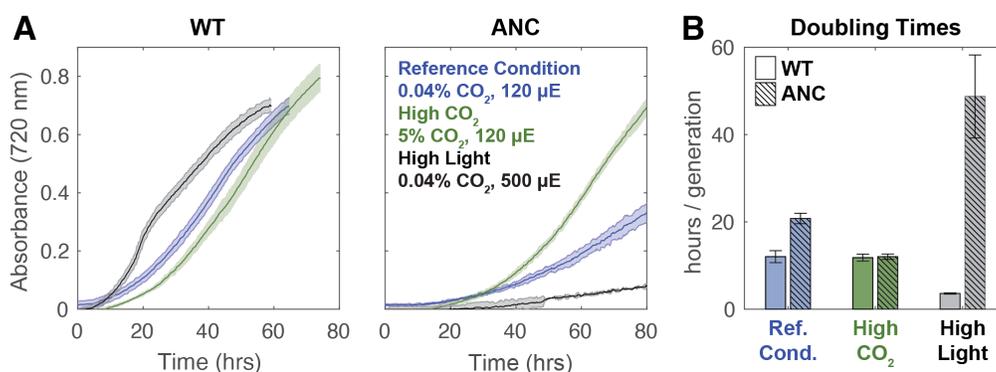
Figure 2.2 Growth curves for WT and ANC strains across experimental conditions.
(A) Averaged growth curves shown for WT and ANC strains to 80 h, colored by growth condition as indicated in figure. Data were smoothed with a rolling median (Materials and Methods); see full ANC growth curves in *SI Appendix*, Fig. 2.16. (B) Average doubling times with SDs. See SI Appendix 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.
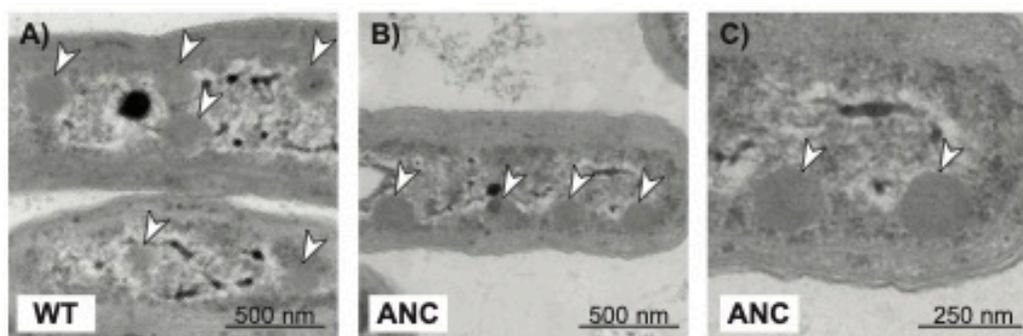
Figure 2.3 WT and ANC strains both produce carboxysomes at ambient pCO₂.
Transmission electron micrographs of WT (A) and ANC (B and C) strains that were harvested during exponential growth in the reference condition (ambient pCO₂, standard light flux). Both strains show multiple carboxysomes per cell, as indicated by white arrows, and carboxysomes exhibit the typical hexagonal shape (Price and Badger 1989). (C) is the same image as in (B) but enlarged to show that rubisco density seen can be within the carboxysomes of ANC. The dark internal body in (A) is likely a polyphosphate body (Jensen 1968). See *SI Appendix*, Fig. 2.17 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 (Fig. 2.2B and *SI Appendix*, Table 2.3), where ANC doubling times were roughly twice that of WT in the reference condition (20.8 ± 1.2 vs. 12.0 ± 1.4 h, respectively). 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 pCO₂, where doubling times for both strains were the same within uncertainty (WT 11.8 ± 0.8 h; ANC 12.0 ± 0.6 h). In contrast to WT, elevated CO_2 greatly accelerated the growth of ANC, reducing its doubling time from ≈ 21 to ≈ 12 h (Fig. 2.2B), supporting our inference that CO_2 availability limits the growth of ANC in ambient air, implicating the CCM in its growth defect. Similar results were found in (Kędzior et al. 2022).

We observed the greatest differences in doubling times between ANC and WT when the strains were grown in high light (500 µmol photons m^{-2} s^{-1}, Fig. 2.2 and *SI Appendix*, Table 2.3). In these conditions, WT cultures were a dark, blue-green color typical of healthy cyanobacterial cells while ANC cultures were yellow-green (*SI Appendix*, Fig. 2.15), suggesting degradation of phycobilisomes via a known starvation pathway to reduce the cell's capacity for light harvesting and photochemical electron transport (Śliwińska-Wilczewska et al. 2020; Grébert et al. 2018). Note that this is a very high light intensity for Cyanobacteria and may induce a severe photoinhibitory response (Richardson et al. 1983). We therefore inferred that ANC could not fix CO_2 at a rate matching its light harvesting capability, and hence expressed this regulatory pathway to decrease light harvesting capacity. WT, in contrast, grew rapidly in high light.

2.4.3 The ANC strain fractionates more than WT

Counter to expectations based on $\varepsilon_{Rubisco}$ (Table 2.1), ANC ε_p was as large or larger than WT ε_p in all conditions tested (Fig. 2.4). This was consistent with recent results from a similar ancestral mutant, where that mutant's ε_p values exceeded WT in ambient and elevated CO_2 levels (Kędzior et al. 2022). In this study, the highest ANC ε_p values were observed for cultures grown in high light, where growth was significantly slower than the WT (doubling time $\approx$ 50 vs. 4 h, respectively, Fig. 2.3 and *SI Appendix*, Table 2.3). ANC ε_p values were also modulated differently by light and CO_2 compared to WT. Compared to the reference condition, WT ε_p values were indifferent to high light and only increased in high CO_2 (Fig. 2.4A). In contrast, ANC ε_p values did not increase in high CO_2 and only increased in high light (Fig. 2.4B). This result contrasted with the ancestral mutant in (Kędzior et al. 2022) where ε_p values increased by $\approx$10‰ at 2% CO_2.
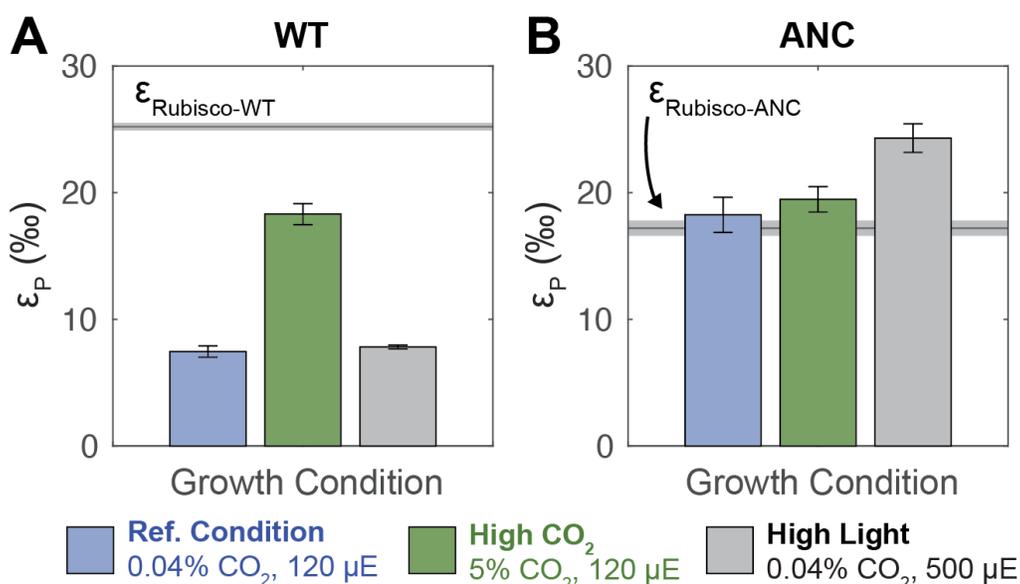
Figure 2.4 Whole-cell carbon isotope fractionation by WT and ANC strains.
ε_p (‰) values (avg. $\pm$ SE) for (A) WT and (B) ANC strains across growth conditions. For each strain, the maximum ε_p possible based on the traditional model ($\varepsilon_p = \varepsilon_{Rubisco}$) is shown as a gray line (avg. $\pm$ SE). Most measured ANC ε_p values exceed the theoretical limit ($\varepsilon_p > \varepsilon_{Rubisco-ANC}$ + SE), while WT ε_p values do not ($\varepsilon_p < \varepsilon_{Rubisco-WT}$ + SE). 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 *SI Appendix*, Table 2.4 for full results.

As discussed above, the traditional box model cannot accommodate ε_p values in excess of $\varepsilon_{Rubisco}$ (Fig. 2.1C). However, average ANC ε_p values exceeded ANC $\varepsilon_{Rubisco}$ in all growth conditions (Fig. 2.4), particularly under high-light conditions where the largest difference was seen (ε_p = 24.30 $\pm$ 0.12‰ vs. $\varepsilon_{Rubisco}$ = 17.23 $\pm$ 0.61‰). The traditional box model also states that ε_p values are

solely modulated by changing external pCO_2 concentrations, which is plainly contradicted by Fig. 2.4B.

2.4.4 Proposed influence of a light-powered, vectoral carbonic anhydrase

Recent studies in extant bacterial and eukaryotic algae have shown that ε_p can regularly exceed $\varepsilon_{Rubisco}$ under certain growth conditions (Wilkes and Pearson 2019), motivating updated models of carbon isotope fractionation in both eukaryotic and bacterial algae (Eichner et al. 2015; Wilkes and Pearson 2019; Erez et al. 1998). Taken together, these studies indicated that observed ε_p values could only be rationalized if an additional fractionation factor was present. Several studies argued that this factor is an energy-coupled CA catalyzing the vectoral hydration of intracellular CO_2 to HCO_3^-, as this reaction is calculated to have a large isotopic effect and would allow ε_p to exceed $\varepsilon_{Rubisco}$ (Eichner et al. 2015; Wilkes and Pearson 2019; Erez et al. 1998). Energy-coupled CAs can facilitate CO_2 uptake by converting extracellular CO_2 that passively translocates the membrane to intracellular HCO_3^- (Fig. 2.1A), which is advantageous in acidic conditions where CO_2 is the dominant form of extracellular C_i (Mangan et al. 2016; Desmarais et al. 2019; Ogawa and Kaplan 2003). Vectoral CAs are also thought to potentially "recycle" CO_2 that leaks from the carboxysome by converting it to HCO_3^- (Price et al. 2013).

Cyanobacteria and eukaryotic algae have two general modes of active C_i uptake: uptake of hydrated C_i (predominantly H_2CO_3 and HCO_3^-) and of CO_2 (Ogawa and Kaplan 2003). In order for the CCM to function, either mode must produce a high, nonequilibrium concentration of HCO_3^- in the cytoplasm (Flamholz and Shih 2020; Mangan et al. 2016). This is thought to be achieved by coupling CA to an energy source (e.g., light or an ion gradient) that drives the vectoral hydration of CO_2 to HCO_3^- in the cytoplasm (Volokita et al. 1984). There is now excellent data supporting this hypothesis in Cyanobacteria, where accessory proteins that bind to the NDH complex, the cyanobacterial homolog of the respiratory Complex I NADH-dehydrogenase, are known to mediate CO_2 uptake specifically (Price et al. 2002; Maeda et al. 2002; Klughammer et al. 1999). Additionally, one of these accessory proteins, CupA/B, is reminiscent of a CA and contains a telltale zinc active site situated near a proton channel in a membrane subunit (Schuller et al. 2020). The prevailing understanding of these data is, therefore, that these complexes couple C_i uptake to energy supplied by photochemical electron transport (Schuller et al. 2020; Artier et al. 2022). Moreover, a similar protein complex has been described in proteobacterial chemoautotrophs, suggesting that energy-coupled CO_2 hydration is widespread (Desmarais et al. 2019).

A vectoral CA would affect ε_p for two reasons. First, CO_2 and HCO_3^- are isotopically distinct. At equilibrium in standard conditions, HCO_3^- is $\approx 9‰$ more enriched in ^{13}C than CO_2 (Mook et al. 1974; Sade and Halevy 2017; Zeebe and Wolf-Gladrow 2001). Therefore, if a cyanobacterium is predominantly taking up CO_2, the internal C_i pool from which biomass is formed would be isotopically lighter (^{13}C-depleted) than if HCO_3^- is the dominant source of C_i. We focused only on

C_i uptake as CO_2 because we were interested in a modification to the traditional model that could achieve large ε_p values (indicating ^{13}C-depleted biomass) to account for at least an additional ~8‰ of fractionation in ε_p (maximum of ~25‰ in the high-light condition) *greater* than $\varepsilon_{Rubisco}$ (~17‰) in ANC. Though HCO_3^- uptake through bicarbonate transporters (e.g., SbtA) was likely occurring under our experimental conditions (Price et al. 2002), isotopically it would not help us achieve the measured large ε_p values because it would shift all ε_p values to be maximally 9‰ more negative (i.e., ^{13}C-enriched biomass, Fig. 2.1C) when we seek to explain values that are ~8‰ more positive. Second, unidirectional CO_2 hydration ($CO_2 + H_2O \rightarrow HCO_3^- + H^+$) is expected to impart a substantial KIE, with calculated values ranging from ≈19 to 32‰ (Sade and Halevy 2017; Clark and Lauriol 1992; Guo 2009; Zeebe 2014; Boettger and Kubicki 2021). Therefore, there are two mechanistic reasons (CO_2 vs. HCO_3^- uptake; unidirectional CO_2 hydration) that ε_p could exceed $\varepsilon_{Rubisco}$ in conditions where energized CO_2 uptake and hydration is active. Indeed, a recent model of C-isotope fractionation in Cyanobacteria specifically invoked the NDH complex to rationalize ε_p values that exceed $\varepsilon_{Rubisco}$ (Eichner et al. 2015).

Because energy-coupled CO_2 uptake and hydration by the NDH complex is driven by light energy, e.g., via cyclic electron flow around photosystem I (Schuller et al. 2020), and because the vectoral hydration of CO_2 to HCO_3^- is thought to have a large carbon isotope fractionation (Sade and Halevy 2017; Clark and Lauriol 1992; Guo 2009; Zeebe 2014; Boettger and Kubicki 2021), ε_p should increase with light intensity. Indeed, we observed the largest ANC ε_p values in the high-light condition and found that ANC ε_p varies primarily with light and not CO_2 (Fig. 2.4). This observation is counter to the traditional model, which proposes ε_p as a direct correlate of external pCO_2 (Hayes 1993; Francois et al. 1993). Furthermore, on short timescales (≈minutes) cyanobacterial C_i uptake can be modulated by light intensity alone, fully independent of external C_i concentrations (Tchernov et al. 2001), and CO_2 uptake can occur in the absence of carbon fixation (Espie et al. 1991; Kaplan and Reinhold 1999). Based on these physiological and isotopic observations, our study also supports the hypothesis that an energy-coupled vectoral CA like the NDH complex permits $\varepsilon_p > \varepsilon_{Rubisco}$, as observed here for ANC in all growth conditions.

2.4.5 Conceptual model for carbon isotope fractionation in Cyanobacteria

As discussed above, the traditional box model cannot produce $\varepsilon_p > \varepsilon_{Rubisco}$ (Fig. 2.1C). In this model, the C_i leakage term (f) is fit from measured ε_p values and $f = 1$ implies that all carbon uptake leaks out of the cell. Though the traditional box model can accommodate both CO_2 and HCO_3^- uptake, which differ in their equilibrium isotopic composition, it does not account for the isotopic effect of vectoral CO_2 hydration. As such, even modeling 100% CO_2 uptake gave physiologically infeasible values of $f > 1$ for ANC in all conditions (Fig. 2.5A and *SI Appendix*, Fig. 2.13), yet ANC grew reproducibly in all conditions tested (Fig. 2.2). We also encountered challenges using the traditional model to rationalize WT data: fitting the model gave $f < 1$ in ambient pCO_2 conditions, but high-

CO_2 conditions yielded $f > 1$ unless all C_i uptake was assumed to be as HCO_3^- (see *SI Appendix*, Fig. 2.13 for discussion). Therefore, to rationalize our results, we developed a simple modified box model that permits $\varepsilon_p > \varepsilon_{Rubisco}$ by including fractionation due to C_i uptake through vectoral CAs.
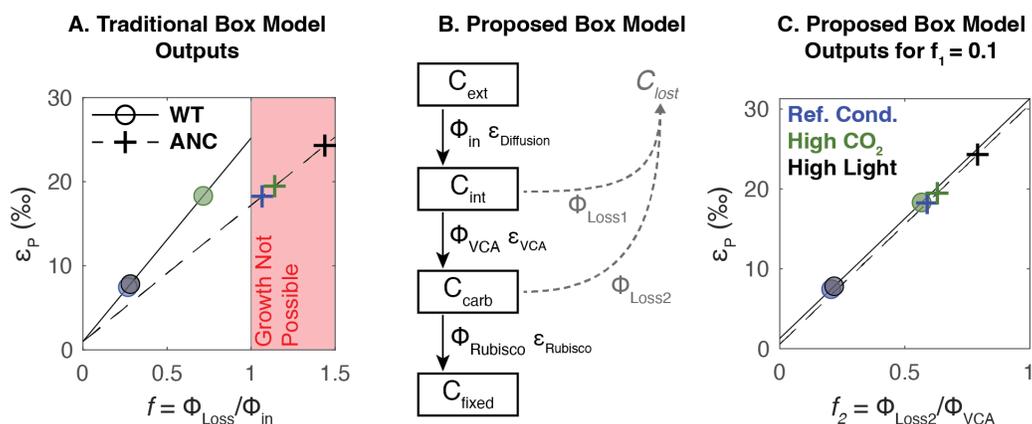
Figure 2.5 Proposed box model based on experimental results.
(A) Experimental results (circles and crosses) plotted onto traditional box model outputs (solid and dashed lines) for WT and ANC, respectively, if C_i uptake is all CO_2. See *SI Appendix*, Table 2.11 for quantification of uncertainty. Colors indicate growth conditions as in Fig. 2.2. The red shaded region demarcates the physiologically infeasible region where $f > 1$. (B) Our proposed box model architecture. 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 (VCA for Vectoral Carbonic Anhydrase), 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 model, f_1 is defined as Φ_{Loss1}/Φ_{in}, and f_2 is defined as Φ_{Loss2}/Φ_{VCA}. (C) Experimental results plotted onto proposed box model outputs for $f_1 = 0.1$; colors and symbols are the same as Panel A; see *SI Appendix*, Table 2.11 for quantification of uncertainty. ε_p is defined as the difference in $\delta^{13}C$ between C_{ext} and C_{fixed}. Here only results for $f_1 = 0.1$ are shown; see <u>SI Appendix, Supplementary Text</u> for full description of model assumptions and results.

In this modified model, we explicitly represent the CCM by distinguishing between carbon in the cytosol (C_{int}) and carbon in the carboxysome (C_{carb}), allowing carbon to be lost from the carboxysome (Φ_{Loss2}, Fig. 2.5B). Therefore, external C_i enters the cell (flux Φ_{in}) where it can either leak out (flux Φ_{Loss1}) or undergo active hydration (flux Φ_{VCA}, where *VCA* denotes *V*ectoral *CA*). 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 VCA) and its associated isotopic fractionation as free parameters. In modeling each strain, we used the appropriate $\varepsilon_{Rubisco}$ measurements (Table 2.1). We do not know the true

value for ε_{VCA}, but used a value of 30‰ similar to a recent model that explicitly invoked the NDH complex in Cyanobacteria (Eichner et al. 2015). For comparison with the traditional model, we plotted Fig. 2.5C with $f_1 = 0.1$ so that it could be represented in two dimensions; see *SI Appendix*, Fig. 2.12 for full model outputs. 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 (*SI Appendix*, Fig. 2.12). 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 because many pairs of f_1 and f_2 values can produce the same ε_p.

The modified model was able to rationalize our experimental data of $\varepsilon_p > \varepsilon_{Rubisco}$ with leakage values compatible with cell growth ($f_2 < 1$, Fig. 2.5C). It may also explain why ANC and WT responded so differently to high light. Our model results implied that 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 rubisco, perhaps due to its lower V_C (Table 2.1). Excess CO_2 allowed rubisco's KIE ($\varepsilon_{Rubisco}$) to be expressed in ε_p. These results indicated that, in high light, the vectoral CA was delivering high amounts of CO_2 to both the WT and ANC rubiscos. The faster WT rubisco was able to match this flux, which was reflected in its fast growth rate (Fig. 2.2) and no change in ε_p vs. the reference condition (Fig. 2.4). However, the slower ANC rubisco was not, which led to its slowest growth rate (Fig. 2.2), and highest ε_p values across all conditions (Fig. 2.4). 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. However, since we could not independently determine Φ_{Loss1} and Φ_{Loss2} —i.e., what proportion of ε_p reflects the contribution of $\varepsilon_{Rubisco}$ vs. ε_{VCA} —we could only conclude that overall the slower ANC rubisco created a "backup" where leakage increased all along the CO_2 fixation pathway and that this effect was exaggerated at high light.

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 ($CO_2 + H_2O \rightarrow HCO_3^- + H^+$), rather than implying that the flux of HCO_3^- dehydration ($HCO_3^- + H^+ \rightarrow CO_2 + H_2O$) is zero. As such, there is likely some bidirectional activity ($CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$) of the NDH complex. It is difficult to experimentally measure the isotope effect associated with the CO_2 hydration reaction, but transition state theory and quantum chemical modeling (Sade and Halevy 2017; Zeebe and Wolf-Gladrow 2001; Zeebe 2014) suggest that the value is large (roughly 25‰, see (Wilkes and Pearson 2019) for review). HCO_3^- dehydration, and equilibration in general, would tend to reduce the isotopic fractionation (Sade and Halevy 2017). Our results here do not require a larger isotopic effect, however. Rather, a smaller value of $\varepsilon_{VCA} =$ 10‰ (*SI Appendix*, Fig. 2.12) 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. See *SI Appendix*, Fig. 2.12 for further discussion.

In addition, ours is not the only model structure that can permit $\varepsilon_p > \varepsilon_{Rubisco}$. We tested other models by fitting our data to them (*SI Appendix*, Fig. 2.14). Models that incorporated an explicitly one-way, "CA-like" enzyme (Erez et al. 1998) or the

NDH complex specifically (Eichner et al. 2015) were mostly able to rationalize our data as well. The poorest fits are when C_i uptake was mostly as HCO_3^- (*SI Appendix*, Fig. 2.14) which is not surprising since we need more positive ε_p values and HCO_3^- uptake would shift all ε_p values to be 9‰ more negative (Fig. 2.1C). Altogether, model fitting indicates adding an additional carbon isotope fractionation step produces a model capable of rationalizing our data by enabling $\varepsilon_p >$ $\varepsilon_{Rubisco}$ with plausible leakage values $f < 1$.

2.4.6 Consequences for understanding the evolution of carbon-fixing metabolisms

Our goal was to test if prevailing models of carbon fixation and isotopic fractionation apply to an ancestral analogue strain that may be 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 (Shih et al. 2016) and a much lower $\varepsilon_{rubisco}$ than the modern strain's native enzyme (17.23 ± 0.61‰ vs. 25.18 ± 0.31‰, Table 2.1).

Recent studies in extant bacterial (Eichner et al. 2015) and eukaryotic algae (Wilkes and Pearson 2019) have 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. Our results emphasize that similar caveats apply to Cyanobacteria, where ANC ε_p exceeded $\varepsilon_{Rubisco}$ in all conditions tested. Inference of Archaean and early Proterozoic pCO_2 from the C isotopic record relies intimately on models of cyanobacterial physiology due to their distinction as the oldest oxygenic photoautotrophs (Fischer et al. 2016). Yet, our results show that the traditional form of these models is not generally reliable.

To date, such anomalous ε_p values have been observed during relatively slow growth; in (Erez et al. 1998) $\varepsilon_p > \varepsilon_{Rubisco}$ occurred early in the growth curve as cells were acclimating to fresh culture media, in (Wilkes and Pearson 2019) $\varepsilon_p > \varepsilon_{Rubisco}$ 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 (Fig. 2.1B and C) to include more physiological detail relating to the presence of a CCM.

As high light consistently slowed growth of ANC, induced chlorosis (yellowing of cultures, *SI Appendix*, Fig. 2.15), 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 starvation and taken to indicate that light levels exceeded the downstream capacity for CO_2 fixation (Collier and Grossman 1992; Adir et al. 2006). We interpreted these observations as indicating that the

replacement of the native rubisco with a slower enzyme decreased capacity for CO_2 fixation (Table 2.1).

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 assume is due to light-coupled vectoral hydration of CO_2, which has a large calculated isotope effect (Sade and Halevy 2017; Clark and Lauriol 1992; Guo 2009; Zeebe 2014; Boettger and Kubicki 2021). It is well established that modern Cyanobacteria have light-coupled CO_2 uptake systems (Price et al. 2002; Schuller et al. 2020) and in model Cyanobacteria, this activity is due to the Cup proteins (CupAS/B, also known as Chp proteins), which bind the NDH complex (Schuller et al. 2020; Battchikova et al. 2011). In order for CO_2 uptake to drive the CCM and promote CO_2 fixation, it would need to produce a high, nonequilibrium HCO_3^- concentration in the cytoplasm (Flamholz and Shih 2020; Mangan et al. 2016). We and others therefore assumed that the complex of NDH-1 and CupAS/B couples light energy to the vectoral hydration of CO_2 to HCO_3^- at a CA-like active site (Schuller et al. 2020). Disruption of a Cup protein by point mutation was also shown to largely affect cell growth (Artier et al. 2022), suggesting that the energy-induced directionality is important for Cyanobacteria.

It is 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 (Fig. 2.4). This was only apparent because we measured the isotopic fractionation due to the ancestral rubisco enzyme ($\varepsilon_{Rubisco}$) and compared it to ANC strain biomass (ε_p), in contrast with (Kędzior et al. 2022), which measured ε_p but not $\varepsilon_{Rubisco}$. While our ANC ε_p values ($\approx$18 to 24‰) fell within the range of ε_p values derived from the carbon isotope record (Hurley et al. 2021), they exceeded ANC $\varepsilon_{Rubisco}$ (Fig. 2.2). 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 models Eqns. 2.1 and 2.2) used to interpret the record (Wilkes and Pearson 2019). In addition, ANC $\varepsilon_{Rubsico}$ (17.23 $\pm$ 0.61‰) is anomalously low; not only is it $\approx$8‰ less than WT $\varepsilon_{Rubisco}$ (25.18 $\pm$ 0.31‰) but it is among the lowest measured rubisco KIEs. However, only thirteen unique rubisco KIEs have been measured thus far (Garcia et al. 2021) while $\approx$300 distinct rubiscos have been kinetically characterized (Flamholz et al. 2019; Iñiguez et al. 2020), suggesting that measuring the isotopic effects of several well-chosen rubisco variants is worthwhile.

Turning to trends in carbon isotope data from the geological record, our results suggested there are at least two nonunique ways to achieve the large ε_p values observed earlier in Earth history: i) High external concentrations of C_i, or ii) Active CO_2 uptake driven by photochemical electron transport. Our proposed model (an idealized extension of the traditional model, Eqn. 2.2) cannot be applied readily to the C Isotope Record Model (Eqn. 2.1). Doing so currently gives nonsensical values of b because ANC $\varepsilon_p > \varepsilon_{Rubisco}$ (see *SI Appendix*, section 2.9.6 and Fig. 2.20 for further discussion), and because we cannot independently constrain the extra degree of freedom introduced (two loss fluxes, Φ_{Loss1} and Φ_{Loss2}, instead of one, f). In addition, these parameters could vary over evolutionary history as the CCM and the

efficiency of carbon fixation evolves. Additional measurements that constrain these parameters (i.e., Φ_{Loss1} and Φ_{Loss2}) could enable pCO_2 to be back-calculated from ε_p, but further work must be done to then adapt those observations to the C Isotope Record Model (Eqn. 2.2). Importantly, the modified model framework proposed here is not the only approach to producing $\varepsilon_p > \varepsilon_{Rubisco}$ with physiologically feasible leakage fluxes. Rather than advocating for our specific model, we offer it as an example form of a solution – showing that ε_p can only exceed $\varepsilon_{Rubisco}$ if additional fractionating process is considered. As shown in *SI Appendix*, Fig. 2.14, several approaches to extending the traditional box model can accommodate $\varepsilon_p > \varepsilon_{Rubisco}$ (Eichner et al. 2015; Erez et al. 1998), yet all of these models represent substantial simplifications of bacterial and algal CCMs. Overall, our study supports the conclusion of prior studies (Eichner et al. 2015; Wilkes and Pearson 2019) that a modified traditional model that engages more fully with photosynthetic physiology, like the CCM, is required to more accurately constrain environmental C_i concentrations from ε_p.

In addition, this study and other recent work (Hurley et al. 2021; Kędzior et al. 2022) 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 (O'Malley and Walsh 2021)—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 physiologies—e.g., C3, C4, CAM in plants—that temper the effect of $\varepsilon_{Rubisco}$ on ε_p (Garcia et al. 2021). 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 (Flamholz et al. 2022; Hurley et al. 2021; Kędzior et al. 2022), 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 (Flamholz et al. 2022) and isotopic fractionation (this work) in both ancient and modern autotrophs, showing that models rigidly based on modern taxa are likely not universally applicable across geologic time.

Overall, carbon fixation was a fundamental challenge that autotrophs overcame early in the history of Earth's biosphere (Fischer et al. 2016). 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 coevolutionary problems by producing ancient process analogs of carbon fixation in the laboratory. Utilizing these tools will enable us to better understand how the evolution of key metabolisms have shaped Earth's chemistry over time.

2.5 Materials and Methods

2.5.1 Ancestral enzyme reconstruction

Ancestral Rubisco enzyme sequences were previously reported and characterized by (Shih et al. 2016). Briefly, for both the large subunit and small subunit 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.

2.5.2 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). Briefly, cultures were grown to OD750 nm = 0.5. Cultures were centrifuged at 18,000 x g for 2 min. Pellets were washed with 100 mM $CaCl_2$ and spun again at 18,000 x g for 2 min. Pellets were resuspended in BG-11 media followed by addition of plasmid and grown for 16 h 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 selected in media with antibiotic until segregation and then genotyped by PCR amplification of the rubisco locus followed by sequencing to confirm homoplasmic ANC strain rubisco sequence. *SI Appendix,* Table 2.2 lists plasmids and primers used in this study.

2.5.3 Growth Conditions

For ambient CO_2 growth, NS2-KanR ("WT") and β Ancestral Rubisco-KanR ("ANC") 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 buffered BG-

11 media with 50mM HEPES at pH 8. Cultures were inoculated at a starting OD720 nm = 0.015 and cultivated at 120 µmol photons m^{-2} s^{-1}, 30 °C, and bubbled with ambient air. High CO_2 growth was performed using the same conditions as ambient growth with the exception of placing the photobioreactor in a 5% CO_2 chamber (Percival AR22L) and bubbling in air from the chamber. High-light growth was performed using the ambient conditions above with the exception of using 500 µmol photons m^{-2} s^{-1} for light intensity. Cells were harvested by centrifugation at 6000 x g for 20 min at 4 °C. Decanted pellets were then flash frozen with liquid N_2 and lyophilized overnight with the Millrock Technology Model BT85A freeze dryer. Doubling time was calculated by fitting the exponential phase of growth (k) using a Markov Chain Monte Carlo (MCMC) approach, using the generic model y = a*EXP(k*x)+b. Growth curves displayed in Fig. 2.3 were smoothed with a rolling median (n = 12) to remove errant readings caused by bubbles advected in front of the detector. See *SI Appendix* for more information.

2.5.4 Carbon isotope analysis

Carbon isotope data are reported using delta notation ($\delta^{13}C$) in units of per mille (‰) where $\delta^{13}C = [(^{13}C/^{12}C)_{sa}/(^{13}C/^{12}C)_{ref}-1]*1000$, where the subscripts "sa" and "ref" denote sample and 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 SE were calculated and reported for each biological replicate (see *SI Appendix* 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. In addition, we labored to keep gas pressures approximately constant during our experiments (i.e., equilibrating to ambient pressure by bubbling) because of potential unwanted isotopic pressure effects. 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 ^{13}C-depleted; see *SI Appendix* for more detail.

2.5.5 Rubisco KIE assay

*Syn*6301 and β-MRCA rubisco were purified according to previous methodologies (Saschenbrecker et al. 2007; Banda et al. 2020) 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 to 40% cut for *Syn*6301 and at the 40 to 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 (Guy et al. 1993; McNevin et al. 2006; Scott et al. 2004; Thomas et al. 2018). Briefly, an assay mix of HCO_3^-, bovine CA, 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 prefilled 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 (McNevin et al. 2006). See *SI Appendix* for more information.

2.5.6 Transmission electron microscopy imaging of whole cells

WT and ANC strains were grown in the reference condition—buffered BG-11 media, shaking at 250 rpm, with white cool fluorescent light at 120 µE, 30 °C, ambient air (0.04% CO_2 (v/v). WT and ANC cells were collected at mid-log (40 and 80 h, respectively) at OD730 nm = 0.4 and pelleted by centrifugation (10,000 x g for 10 min). Pelleted cells were then resuspended in 1 mL 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 a 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 poststained 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).

2.5.7 Data, Materials, and Software Availability

All study data are included in the article and/or *SI Appendix*.

2.6 Acknowledgments

2.7 Author contributions

R.Z.W., R.J.N., A.K.L., D.F.S., J.M.E., P.M.S., and W.W.F. designed research; R.Z.W., R.J.N., A.K.L., J.A., and D.M.B. performed research; R.Z.W., R.J.N., A.K.L., A.I.F., and J.A. analyzed data; D.F.S., J.M.E., P.M.S., and W.W.F. advised on project; and R.Z.W. and A.I.F. wrote the paper.

2.8 Competing interests

The authors declare no competing interest.

2.9 Supplementary Materials

Name	Sequence	Notes
	tcaccaataaataacgcccggcggcaaccgagcgttctgaac aaatccagatggagttctgaggtcattactggatctatcaacag gagtccaagcgagctcgatatcaaattacgccccgccctgcc actcatcgcagtactgttgtaattcattaagcattctgccgacat ggaagccatcacaaacggcatgatgaacctgaatcgccagc ggcatcagcaccttgtcgccttgcgtataatatttgcccatggtg aaaacgggggcgaagaagttgtccatattggccacgtttaaat caaaactggtgaaactcacccagggattggctgaaacgaaaa acatattctcaataaaccctttagggaaataggccaggttttcac cgtaacacgccacatcttgcgaatatatgtgtagaaactgccg gaaatcgtcgtggtattcactccagagcgatgaaaacgtttca gtttgctcatggaaaacggtgtaacaagggtgaacactatccc atatcaccagctcaccgtctttcattgccatacgaaattccggat gagcattcatcaggcgggcaagaatgtgaataaaggccggat aaaacttgtgcttatttttctttacggtctttaaaaaggccgtaata tccagctgaacggtctggttataggtacattgagcaactgactg aaatgcctcaaaatgttctttacgatgccattgggatatatcaac ggtggtatatccagtgatttttttctccattttagcttccttagctcc tgaaaatctcgataactcaaaaaatacgcccggtagtgatctta tttcattatggtgaaagttggaacctcttacgtgcccgatcaatc atgaccaaaatcccttaacgtgagttttcgttccactgagcgtca gaccccgtagaaaagatcaaaggatcttcttgagatcctttttt ctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctac cagcggtggtttgtttgccggatcaagagctaccaactctttttc cgaaggtaactggcttcagcagagcgcagataccaaatactg ttcttctagtgtagccgtagttaggccaccacttcaagaactctg tagcaccgcctacatacctcgctctgctaatcctgttaccagtg gctgctgccagtggcgataagtcgtgtcttaccgggttggact caagacgatagttaccggataaggcgcagcggtcgggctga acggggggttcgtgcacacagcccagcttggagcgaacgac ctacaccgaactgagatacctacagcgtgagctatgagaaag cgccacgcttcccgaagggagaaaggcggacaggtatccg gtaagcggcagggtcggaacaggagagcgcacgagggag cttccaggggggaaacgcctggtatctttatagtcctgtcgggttt cgccacctctgacttgagcgtcgatttttgtgatgctcgtcagg ggggcggagcctatggaaaaacgccagcaacgcggccttttt acggttcctggccttttgctggccttttgctcacatgttctttcctg cgttatcccctgattctgtggataaccgtagggcgcgcctgca ggcggccgcgaattggtcctgtactgcgatcgtgcaaggcac ggtttctaatgtgaccgttgcggtcgaagccgggatgtatgcc gctgagcggatcggccagctcaacgcaatcatggtcattccc	Plasmid used to generate β ancestral RuBisCO strain-KanR
pAncRubisco-KanR		

agaccgctagacgacttgatggacagcttgcctgagccgcag
tcggatagcgaagcagcccagccactccaattaccgctgcgg
gttcgcgaaaaacaaccgctgttggagctaccggaactcgaa
cggcagccgatcgcgatcgaagcaccgcgacttttagcaga
agagcgacagtctgcgttggaattggctcaagagacaccgct
cgccgagcccttagagctccccaatcctcgtgatgatcagtga
tggaaaaagcactgtaattcccttggttttttggctgaaagtttcg
gactcagtagacctaagtacagagtgatgtcaacgccttcaag
ctagacgggaggcggcttttgccatggttcagcgatcgctcct
catcttcaataagcagggcatgagccagcgttaagcaaatcaa
atcaaatctcgcttctgggcttcaataaatggttccgattgatgat
aggttgattcatgaggaatctaaggcttaattctccacaaaaga
attaagcgtccgtcgcaacggaatgctccgctggacttgcgct
gtgggactgcagctttacaggctccccctgccagaaatcctga
atcgtcgagcatatctgacatatctctagggagagacgacatg
accaaaacccagagcgccgccggttacaaagccggtgtgaa
ggattatcgcctgacctactataccccccgattacaccccgaag
gataccgatctgctggctgcctttcgcgtgaccccgcagcccg
gtgtgccgcccgaagaggccggcgcggctgtggcggccga
aagcagcacgggcacctggaccaccgtctggaccgatctgc
tgaccgatatggatcgctacaaaggccgctgctatcacattga
acccgtgccgggcgaggataacagctactttgccttcatcgcg
tatccccctggatctgttcgaagagggcagcgtcaccaacatcc
tgaccagcattgtgggcaatgtctttggcttcaaagctctgcgc
gccctgcgcctggaagatattcgctttccggtggcctatgtcaa
gacctttcaaggcccgcccccacggcatccaagtggagcgcg
ataaactgaacaagtatggccgcccccctgctgggctgcacca
ttaaaccgaagctgggcctgagcgccaaaaattacggccgc
gctgtctatgaatgcctgcgcggcggcctggattttaccaagg
atgatgagaacatcaatagccagcccttccaacgctggcgcg
atcgctttctgttcgtggcggatgctattcacaaagcccaggcg
gaaaccggcgagatcaagggccattacctgaacgtcaccgct
cccacctgcgaagagatgatgaaacgcgcggaatttgctaag
gagctgggcatgccgatcattatgcacgattttctgaccgcggg
gcttcaccgctaacaccaccctggccaaatggtgccgcgata
atggcgtgctgctgcacatccatcgcgccatgcatgcggtcat
tgatcgccagaaaaatcacggcatccattttcgcgtgctggcta
agtgcctgcgcctgagcggcggcgatcacctgcatacggggt
acggtggtgggtaaactggaaggcgatcgcgccagcaccct
gggttttgtcgatctgctgcgcgaagattatattgaggctgatc
gcagccgcggcatctttttcacccaagattgggctagcatgcc
gggtgtgatggctgtggctagcggcggcatccacgtctggca
tatgccggccctggtcgaaatctttggcgatgatagcgtgctg
cagtttggtggcggcacccctgggtcacccatggggtaatgctc

| | ccggcgccacggctaatcgcgtggctctggaagcctgcgtcc aagctcgcaatgagggtcgcgatctgatgcgcgagggcggc gatattctgcgcgaagctgcgaagtggagcccggagctggc ggctgccctggaactgtggaaagagatcaagtttgaatttgaa accgtcgataagctgtaaggagcctctgactatcgctggggg agtgagcgttgctgcgtaaagctttctccccagcctttcgactta acctttcaggatttctgaatcatgcaagtgtggacccccgcgaa gaacaagaagtacgaaaccttcagctacctgcccccgctgag cgatgagcagatcgctaagcagatccaatacattctgagccaa ggctgggtgccctgcgtcgaatttaacgaggatagccacccg gaaaatcgctattggaccatgtggaaactgccgctgtttggtgc tcaggatgcggcccaagtgctgagcgaggtccaagcttgcc gcaaagcctttccgaactgctacatccgcgtggtcggcttcga taatgtgaagcagtgccaatgcatgagcttcattgtccatcgcc cggcgtaaagcctgatttgtcttgatagctgctcctgcctttggg caggggctttttctgtctgccattcttgaggaagtaagcttagat cgacctgcaggggggggggggggaaagccacgttgtgtctcaa aatctctgatgttacattgcacaagataaaaatatatcatcatga acaataaaactgtctgcttacataaacagtaatacaagggggtgt tatgagccatattcaacgggaaacgtcttgctcgaggccgcga ttaaattccaacatggatgctgatttatatgggtataaatgggct cgcgataatgtcgggcaatcaggtgcgacaatctatcgattgt atgggaagcccgatgcgccagagttgtttctgaaacatggca aaggtagcgttgccaatgatgttacagatgagatggtcagact aaactggctgacggaatttatgcctcttccgaccatcaagcattt tatccgtactcctgatgatgcatggttactcaccactgcgatccc cgggaaaacagcattccaggtattagaagaatatcctgattca ggtgaaaatattgttgatgcgctggcagtgttcctgcgccggtt gcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtat ttcgactcgctcaggcgcaatcacgaatgaataacggtttggtt gatgcgagtgattttgatgacgagcgtaatggctggcctgttga acaagtctggaaagaaatgcataagcttttgccattctcaccgg attcagtcgtcactcatggtgatttctcacttgataaccttatttttg acgaggggaaattaataggttgtattgatgttggacgagtcgg aatcgcagaccgataccaggatcttgccatcctatggaactgc ctcggtgagttttctccttcattacagaaacggctttttcaaaaat atggtattgataatcctgatatgaataaattgcagtttcatttgatg ctcgatgagtttttctaatcagaattggttaattggttgtaacactg gcagagcattacgctgacttgacgggacggcggctttgttgaa taaatcgaacttttgctgagttgaaggatcagatcacgcatcttc ccgacaacgcagaccgttccgtggcaaagcaaaagttcaaa atcaccaactggtccacctacaacaaagctctcatcaaccgtg gctccctcactttctggctggatgatggggcgattcaggcctg gtatgagtcagcaacaccttcttcacgaggcagacctcagcg | |

	cccccccccccctgcaggtctggcggactctttcccttttgctct acgcccatgaatgcgatcgcagtctcccctgtccagcacgttg gagtgattggtggtggccagttagcttggatgctggcaccagc agcgcaacagttggggatgtcgctgcacgttcaaacacccaa tgatcacgacccagcagtagcgatcgcggatcaaaccgtatt agcagcagttgctgacgctgcagcgactgcgaaattggctca agcctgtgacgtcatcacattcgaaaatgagtttgttgatctgcc ggctttgaccgagctggaggaaactggtgtccggtttcgcccc cgtccagcggcgatcgcctccctgctcgacaaacttgatcag cgacaactattgactcgtctgggattgccaaccccacgcttttta gcgatcgcggcagcaaccgcaacagagtcggagctaacag ccttgggctttccggtggtgctgaagcaacgccgccatggcta cgacggcaagggaacacaggttttgcgatcgctagcagaact tcaacaagccttgcagtcttatggcgatacgccactactcctcg aagagttcattccctttgagcaggaattagcggtgatggttgcc cgtagtcagagtggggcgatcgcgactttccctgtggttcaga cccatcagcagaatcaggtctgtcgttgggtcgttgctcctgct gccatcccaggcgcgttgcaaaaagccgttgctgcgatcgcc cgaaccctcgttgagacggtcgattatgttggcgtcgcgggca ttgaactctttcagcagggcgatcgcctctgggtgaacgaaatt gcgccccgcacccacaactcaggacactacagcttggacgc ctgccagacttcgcagtttgaacagcagttgcgagcgatcgct gatctgcctttgggatcgacagcattgcagtggcccggtgcct taatggttaatctcctcggcttcgaggatcaccgctggatctcct gcaggcggccgcggcgcgcc	
pNS2-KanR	cacaccacgtctcacccttcacacaggaaacagaccatgtcc agagaccattaatgcagctggcacgacaggtttcccgactgg aaagcgggcagtgagcgcaacgcaattaatgtgagttagctc actcattaggcaccccaggctttacactttatgcttccggctcgt atgttgtgtggaattgtgagcggataacaatttcacacaggaaa cagctatgaccatgattacgccaagcttgcatgcctgcaggtc gactctagaggatccccgggtaccgagctcgaattcactggc cgtcgttttacaacgtcgtgactgggaaaaccctggcgttacc caacttaatcgccttgcagcacatcccccttcgccagctggc gtaatagcgaagaggcccgcaccgatcgcccttcccaacagt tgcgcagcctgaatggcgaatggcgcctgatgcggtattttct ccttacgcatctgtgcggtatttcacaccgcatatggtgcactct cagtacaatctgctctgatgccgcatagttaagccagccccga cacccgccaacacccgctgacgcgccctgacgggcttgtctg ctcccggcatccgcttacagacaagctggtctctagcggttaa gagaagattttcagcctgatacagattaaatcagaacgcagaa gcggtctgataaaacagaatttgcctggcggcagtagcgcgg	Plasmid used to generate NS2-KanR strain

tggtcccacctgaccccatgccgaactcagaagtgaaacgcc
gtagcgccgatggtagtgtggggtatccccatgcgagagtag
ggaactgccaggcatcaaataaaacgaaaggctcagtcgaa
agactgggcctttcgttttatctgttgtttgtcggtgaacgctctc
ctgagtaggacaaatccgccgggagcggatttgaacgttgcg
aagcaacggcccggagggtggcgggcaggacgcccgcca
taaactgccaggcatcaaattaagcagaaggccatcctgacg
gatggccttttagtaagcttagatcgacctgcagggggggggg
gggaaagccacgttgtgtctcaaaatctctgatgttacattgca
caagataaaaatatatcatcatgaacaataaaactgtctgcttac
ataaacagtaatacaagggtgttatgagccatattcaacggg
aaacgtcttgctcgaggccgcgattaaattccaacatggatgct
gatttatatgggtataaatgggctcgcgataatgtcgggcaatc
aggtgcgacaatctatcgattgtatgggaagcccgatgcgcc
agagttgtttctgaaacatggcaaaggtagcgttgccaatgatg
ttacagatgagatggtcagactaaactggctgacggaatttatg
cctcttccgaccatcaagcattttatccgtactcctgatgatgcat
ggttactcaccactgcgatccccgggaaaacagcattccagg
tattagaagaatatcctgattcaggtgaaaatattgttgatgcgc
tggcagtgttcctgcgccggttgcattcgattcctgtttgtaattg
tcctttaacagcgatcgcgtatttcgactcgctcaggcgcaat
cacgaatgaataacggtttggttgatgcgagtgattttgatgac
gagcgtaatggctggcctgttgaacaagtctggaaagaaatg
cataagcttttgccattctcaccggattcagtcgtcactcatggt
gatttctcacttgataaccttattttttgacgaggggaaattaatag
gttgtattgatgttggacgagtcggaatcgcagaccgatacca
ggatcttgccatcctatggaactgcctcggtgagttttctccttc
attacagaaacggcttttttcaaaaatatggtattgataatcctgat
atgaataaattgcagtttcatttgatgctcgatgagtttttctaatc
agaattggttaattggttgtaacactggcagagcattacgctga
cttgacgggacggcggctttgttgaataaatcgaacttttgctg
agttgaaggatcagatcacgcatcttcccgacaacgcagacc
gttccgtggcaaagcaaaagttcaaaatcaccaactggtccac
ctacaacaaagctctcatcaaccgtggctccctcactttctggc
tggatgatggggcgattcaggcctggtatgagtcagcaacac
cttcttcacgaggcagacctcagcgcccccccccccctgcag
gtcgatctggtaaccccagcgcggttgctaccaagtagtgacc
cgcttcgtgatgcaaaatccgctgacgatattcgggcgatcgc
tgctgaatgccatcgagcagtaacgtggcaccccgcccctgc
caagtcaccgcatccagactgaacagcaccaagaggctaaa
acccaatcccgccggtagcagcggagaactacccagcattg
gtcccaccaaagctaatgccgtcgtggtaaaaatcgcgatcg
ccgtcagactcaagcccagttcgctcatgcttcctcatctaggt
cacagtcttcggcgatcgcatcgatctgatgctgcagcaagcg

ttttccataccggcgatcgcgccgtcgccctttcgctgccgtgg
cccgcttacgagctcgtttatcgaccacgatcgcatccaaatcc
gcgatcgcttcccagtccggcaattcagtctggggcgtccgttt
cattaatcctgatcaggcacgaaattgctgtgcgtagtatcgcg
catagcggccagcctctgccaacagcgcatcgtgattgcctg
cctcaacaatctggccgcgctccatcaccaagatgcggctgg
cattacgaaccgtagccagacggtgagcaatgataaagaccg
tccgtccctgcatcacccgttctagggcctcttgcaccaaggtt
tcggactcggaatcaagcgccgaagtcgcctcatccagaatt
aaaatgcgtggatctagccgcgttgctggcgtttttccataggc
tccgcccccctgacgagcatcacaaaaatcgacgctcaagtc
agaggtggcgaaacccgacaggactataaagataccaggcg
tttccccctggaagctccctcgtgcgctctcctgttccgaccct
gccgcttaccggatacctgtccgcctttctcccttcgggaagc
gtggcgctttctcatagctcacgctgtaggtatctcagttcggtg
taggtcgttcgctccaagctgggctgtgtgcacgaacccccc
gttcagcccgaccgctgcgccttatccggtaactatcgtcttga
gtccaacccggtaagacacgacttatcgccactggcagcagc
cactggtaacaggattagcagagcgaggtatgtaggcggtgc
tacagagttcttgaagtggtggcctaactacggctacactaga
agaacagtatttggtatctgcgctctgctgaagccagttaccttc
ggaaaaagagttggtagctcttgatccggcaaacaaaccacc
gctggtagcggtggtttttttgtttgcaagcagcagattacgcg
cagaaaaaaaggatctcaagaagatcctttgatctgtccagctt
gtcatctgccggatgaggcaaaaccctgcctacggcgcgatt
acatcgtcccagcgcgatcgctcttactgttgatggctcgtgctt
aaaaacaatgcaaacttcaccgtttcagctggtgattttcgactg
tgatggtgtgcttgttgatagcggaacgcatcactaatcgcgtc
tttgcagacatgctcaatgaactgggtctgttggtgactttggat
gacatgtttgagcagtttgtgggtcattccatggctgactgtctc
aaactaattgagcgacggttaggcaatcctccaccccctgact
ttgttcagcactatcaacgccgtacccgtatcgcgttagaaacg
catctacaagccgttcctggggttgaagaggctttggatgctct
tgaattgccctactgtgttgcgtccagtggtgatcatcaaaagat
gcgaaccacactgagcctgacgaagctctggccacgatttga
gggacgaatcttcagcgtgactgaagtacctcgcggcaagcc
atttcccgatgtctttttgttggccgccgatcgcttcggggttaat
cctacggcctgcgctgtgatcgaagacacccccttgggagta
gcggcaggcgtggcggcaggaatgcaagtgtttggctacgc
gggttccatgcccgcttggcgtctgcaagaagccggtgccca
tctcattttgacgatatgcgactgctgcccagtctgctccaatc
gtcgccaaaagataactccacagcattgcccaatccctaaccc
ctgctcgcgccgcaactacacactaaaccgttcctgcgcgatc
gctcttactgttgatggctcgtgcttaaaaacaatgcaaccctaa

	ccgtttcagctggtgattttcggacgatttggcttacagggataactgagagtcaacagcctctgtccgtcattgcacacccatccatgcactggggacttgactcatgctgaatcacatttcccttgtccattgggcgagagggggaggggaatcttctggactcttcactaagcggcgatcgcaggttcttctacccaagcagtggcgatcgcttgattgcagtcttcaatgctggcctctgcagccatcgccgccaccaaagcatcgtaggcgggacgttgttgctccagtaaagtcttcgcccgtaacaatccccagcgactgcgtaaatccgcttcggcaggattgcgatcgagttgccgccacagttgtttccactgggcgcgatcgtcagctcccccttccacgttgccgtagaccagttgctctgccgctgcaccggccatcaacacctgacaccactgttccagcgatcgctgactgagttgcccctgtgcggcttcggcttctagcgcagctgcttggaactgcacacccccgcgaccaggttgtccttggcgcagcgcttcccacgctgagagggtgtagcccgtcacgagcgcttacagacaagctgtgaccgcctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaacctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggttggctgagacgtggtgtg	
Primer RJN610	agggcatgagccagcgttaa	Anneals upstream of RbcLS locus
Primer RJN611	ggtggtgttggcggtgaaac	Anneals to WT RbcL locus
Primer RJN612	cacgcgaaaatggatgccg	Anneals to mutant ancestral RbcL
Primer RJN613	gcaatcccagacgagtcaatagtt	Anneals downstream of RbcLS locus

Table 2.2. List of primers and plasmids used in this study.
The NS2-KanR strain is referred to as 'WT' in this study, while the β ancestral rubisco strain-KanR is referred to as 'ANC.'

2.9.1 Calculating doubling times from growth curves
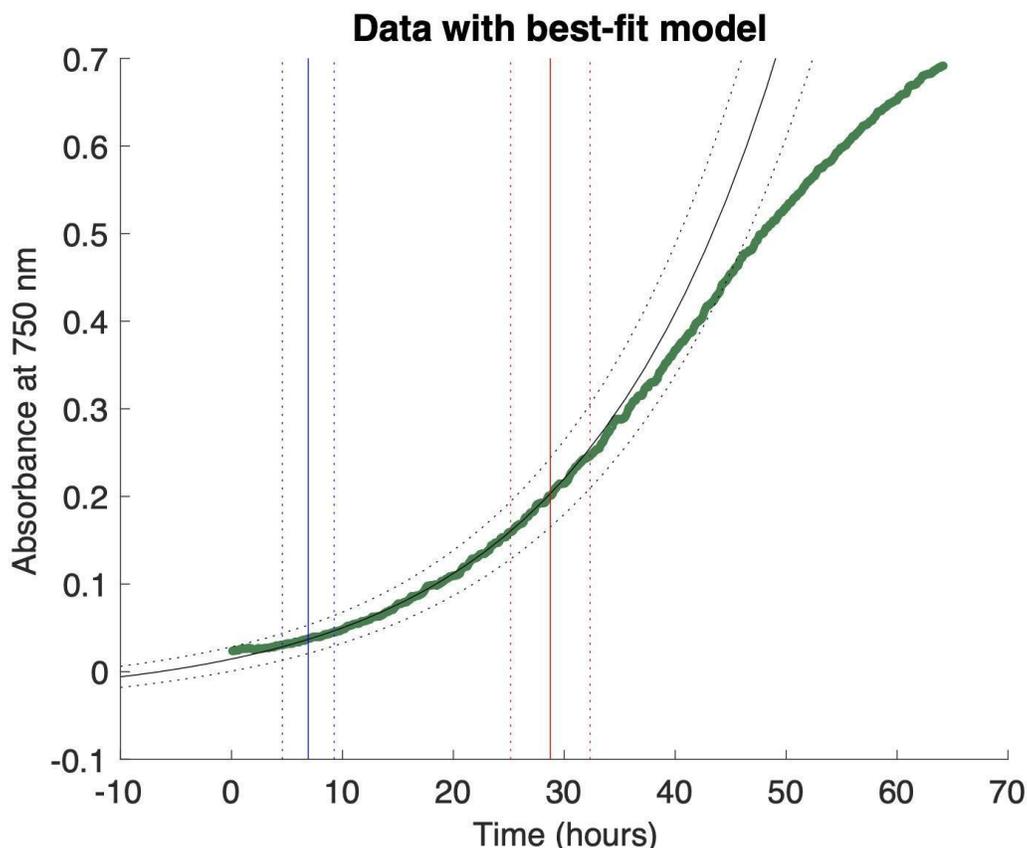
Figure 2.6. Best-fit model for calculating growth constant, *k*, for one growth curve.
Black solid and dotted lines indicate best fit for the exponential section of the growth curve. Blue solid and dotted lines indicate best fit left bound. Red solid and dotted lines indicate best fit right bound. Analyses were performed using MATLAB and Statistics Toolbox (vR2020b).

Growth constants, k (hr^{-1}), were fit using a custom Markov Chain Monte Carlo (MCMC) approach, written using MATLAB and Statistics Toolbox (vR2020b). Code can be found at https://github.com/reneezwang/ancestral-rubisco-cyano. We used this approach to limit human-based error on assessing when the exponential phase ended, and therefore left this as a free parameter for the MCMC.

To fit the exponential phase of growth, we created a model with five free parameters, and used an MCMC approach to find the best-fit values for each parameter. The model we fit follows an equation for exponential growth:

$$y = a * e^{k*x} + b$$

Equation 2.3
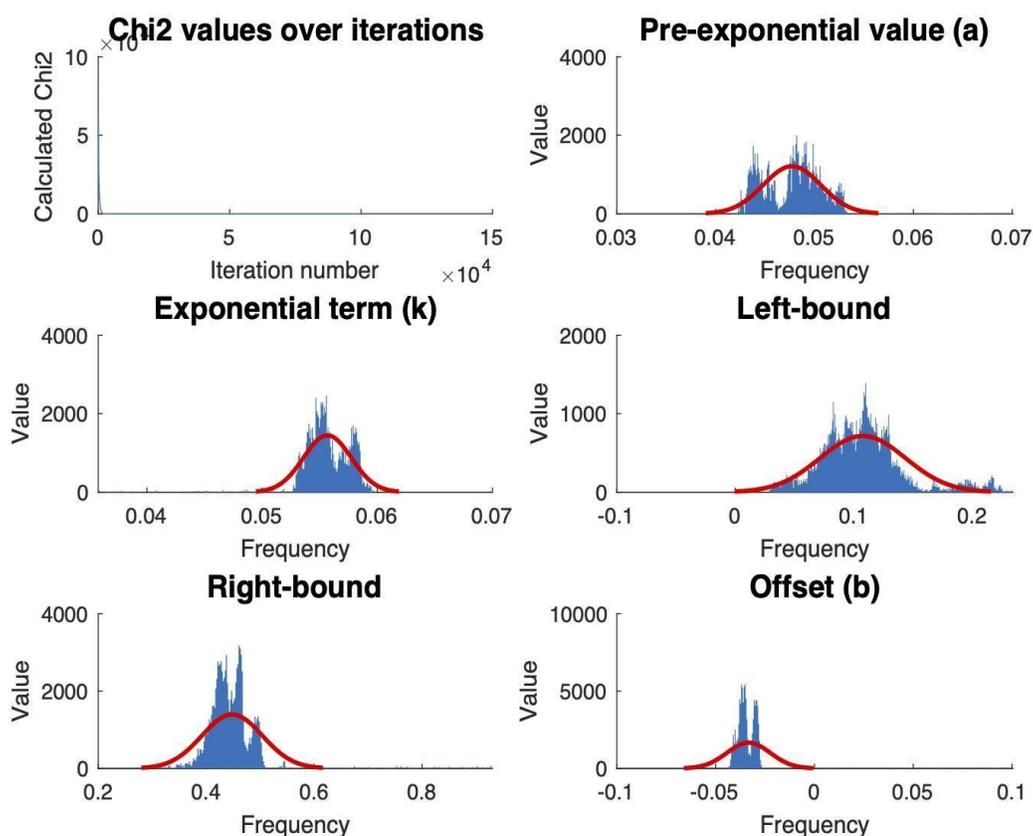
Figure 2.7. Outputs for parameters used in MCMC to calculate the growth constant, *k*.
A histogram of each output is shown in blue, and a probability density function fit to the data is
shown in red. Analyses were performed using MATLAB and Statistics Toolbox (vR2020b).

We then fit this model between a left-bound, *L*, and a right-bound, *R*, around
the phase of exponential growth, so that only the exponential phase is fit. These
bounds were left unconstrained so that they could be optimized by the MCMC. In
total, we fit five parameters: 1) *a,* the pre-exponential factor (units of absorbance at
750 nm); 2) *k*, the growth constant (units of 1/hr); 3) *b*, the offset (units of
absorbance at 750 nm); 4) *L*, the left bound (percentage of the length of data for
each curve); 5) *R*, the right bound (percentage of the length of data for each curve).
The MCMC found the best parameter by minimizing the χ^2 value, and 100,000 to
1,000,000 steps were run for each curve. Fig. 2.6 shows the best-fit model for one
growth curve. In black is the best-fit curve for the exponential phase of growth,
with 1 sigma error shown in the black dotted lines. The best-fit left-bound, *L*, is
shown in blue, with its 1 sigma error in blue dotted lines. The best-fit right-bound,
R, is shown in red, with its 1 sigma error shown in red dotted lines. The
corresponding parameter outputs are shown in Fig. 2.7. χ^2 is quickly minimized,
and a Gaussian curve is fit to each parameter to find the best fit value and 1 sigma
error. The fitted *k* constants for each growth curve are summarized in Table 2.3.

The doubling time was then calculated as $\ln(2)/k$. Full growth curves are shown in Fig. 2.17.

Strain	Replicate	Condition	Growth constant (k) (1/hr)	Doubling Time (hrs)
WT	1	Reference Condition	0.0557 ± 0.0021	12.4 ± 0.5
WT	2	Reference Condition	0.0563 ± 0.0026	12.3 ± 0.6
WT	3	Reference Condition	0.0521 ± 0.0012	13.3 ± 0.3
WT	4	Reference Condition	0.0687 ± 0.0052	10.1 ± 0.8
ANC	1	Reference Condition	0.0342 ± 0.0020	20.3 ± 1.2
ANC	2	Reference Condition	0.0313 ± 0.0046	22.1 ± 3.3
ANC	3	Reference Condition	0.0348 ± 0.0029	19.9 ± 1.7
WT	1	High CO_2	0.0535 ± 0.0029	13.0 ± 0.7
WT	2	High CO_2	0.0606 ± 0.0037	11.4 ± 0.7
WT	3	High CO_2	0.0618 ± 0.0025	11.2 ± 0.5
WT	4	High CO_2	0.0598 ± 0.0060	11.6 ± 1.2
ANC	1	High CO_2	0.0553 ± 0.0045	12.5 ± 1.0
ANC	2	High CO_2	0.0614 ± 0.0026	11.3 ± 0.5
ANC	3	High CO_2	0.0591 ± 0.0069	11.7 ± 1.4
ANC	4	High CO_2	0.0553 ± 0.0102	12.5 ± 2.3
WT	1	High Light	0.1980 ± 0.0188	3.5 ± 0.3
WT	2	High Light	0.1874 ± 0.0144	3.7 ± 0.3
ANC	1	High Light	0.0165 ± 0.0015	42.0 ± 3.8
ANC	2	High Light	0.0125 ± 0.0019	55.5 ± 8.4

Table 2.3: Fitted growth constants and doubling times for growth curves.
Outputs from MCMC approach for fitting exponential phase of growth phase (avg. ± s.d.). Doubling time was calculated as $\ln(2)/k$.

2.9.2 Carbon isotope measurements

2.9.2.1 Delta notation ($\delta^{13}C$)

 Carbon isotope data were reported using delta notation ($\delta^{13}C$) in units of per mille (‰) where $\delta^{13}C = [(^{13}C/^{12}C)_{sa}/(^{13}C/^{12}C)_{ref}-1]*1000$, where the subscripts 'sa' and 'ref' denote sample and reference respectively. All values in this study were reported relative to the Vienna Pee Dee Belemnite (VPDB) reference.

2.9.2.2 CO_2 substrate

 Two different CO_2 substrates were used. For strains grown at ambient CO_2 concentrations (Reference Condition and High Light condition), ambient air was bubbled into the photobioreactor. Ambient air from the Savage lab at UC Berkeley was sampled into two 500 mL pre-evacuated glass bottles. Bottles were delivered by car to Caltech, where the contents were distilled on a vacuum line to separate and concentrate CO_2. Ambient air was cycled repeatedly as follows: 1) Sample was run over two traps filled with 3 mm diameter glass beads and immersed in liquid nitrogen in order to condensate H_2O and CO_2; 2) H_2O was then removed using a dry ice / ethanol slurry. For the High CO_2 condition, the CO_2 was sourced from a CO_2 tank so an aliquot was taken. The purified CO_2 from ambient air and the tank CO_2 were then both analyzed in triplicate on a Thermo MAT 253 at Caltech to measure its $\delta^{13}C$ value. The CALT-2049C standard, which has a $\delta^{13}C_{VPDB}$ value of -3.62‰, was used to correct measured lab values to the international Vienna Pee Dee Belemnite (VPDB) carbon isotope standard. Measured values can be found in Table S3.

2.9.2.3 Bulk cyanobacterial cells

 As stated in the main text, cells were grown in a photobioreactor in the Savage Lab at UC Berkeley in each condition. Cells were then harvested by centrifugation at 6000 x *g* for 20 minutes at 4°C. Decanted pellets were then flash frozen with liquid N_2 and lyophilized overnight with the Millrock Technology freeze dryer (Model BT85A). Pelleted cells were then shipped on dry ice overnight to Caltech, where they were measured on a Delta-V Advantage with Gas Bench and Costech Elemental Analyzer (EA) at the California Institute of Technology. Each sample was measured 4 times on the EA. Each biological replicate was run four times with two different isotope standards – urea (-27.8‰) and sucrose (-10.45‰), so that carbon isotope values could be reported relative to VPDB. The uncertainties from correcting samples to this standard curve were smaller than the analytical replicate uncertainties, and so were ignored moving forward. 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. See Table 2.4 for finalized, drift-corrected values reported relative to VPDB.

2.9.2.4 Error on reported $\delta^{13}C$ values

For each condition, multiple biological replicates were grown (see Table 2.4 for number of replicates). Each biological replicate was then analyzed 4 times on the EA. The average of 4 technical replicates was taken to represent each biological replicate. The standard deviation (s.d.) was calculated from these values, and the standard error (s.e.) was calculated as s.d./(n^0.5), where n is the number of technical replicates.

Strain	Rep.	Condition	$\delta^{13}C$ of CO_2 (‰)	$\delta^{13}C$ of bulk cells (‰)	ε_P (CO_2/bio) (‰)
WT	1	Reference Condition	-12.455 ± 0.005	-19.371 ± 0.043	7.053 ± 0.045
WT	2	Reference Condition	-12.455 ± 0.005	-19.850 ± 0.046	7.544 ± 0.048
WT	3	Reference Condition	-12.455 ± 0.005	-19.480 ± 0.053	7.165 ± 0.055
WT	4	Reference Condition	-12.455 ± 0.005	-20.343 ± 0.087	8.052 ± 0.090
ANC	1	Reference Condition	-12.455 ± 0.005	-31.482 ± 0.088	19.646 ± 0.093
ANC	2	Reference Condition	-12.455 ± 0.005	-30.129 ± 0.089	18.223 ± 0.094
ANC	3	Reference Condition	-12.455 ± 0.005	-28.841 ± 0.102	16.873 ± 0.107
WT	1	High CO_2	-36.839 ± 0.021	-54.247 ± 0.298	18.407 ± 0.322
WT	2	High CO_2	-36.839 ± 0.021	-54.162 ± 0.097	18.315 ± 0.108
WT	3	High CO_2	-36.839 ± 0.021	-55.037 ± 0.572	19.258 ± 0.618
WT	4	High CO_2	-36.839 ± 0.021	-53.160 ± 0.133	17.237 ± 0.146
ANC	1	High CO_2	-36.839 ± 0.021	-53.924 ± 1.002	18.059 ± 1.079
ANC	2	High CO_2	-36.839 ± 0.021	-55.750 ± 1.382	20.027 ± 1.494
ANC	3	High CO_2	-36.839 ± 0.021	-56.029 ± 1.307	20.329 ± 1.413
ANC	4	High CO_2	-36.839 ± 0.021	-55.216 ± 1.605	19.451 ± 1.732
WT	1	High Light	-12.455 ± 0.005	-20.213 ± 0.102	7.918 ± 0.105
WT	2	High Light	-12.455 ± 0.005	-20.007 ± 0.132	7.706 ± 0.136
ANC	1	High Light	-12.455 ± 0.005	-36.632 ± 0.082	25.097 ± 0.088
ANC	2	High Light	-12.455 ± 0.005	-35.131 ± 0.073	23.501 ± 0.077

Table 2.4: Measured carbon isotope values ($\delta^{13}C$) and calculated ε_P values.
Values (avg. ± s.e.) are reported relative to VPDB.

2.9.2.5 Calculating ε_P (CO_2/bio) and its error

ε_P, the vectorial isotopic fractionation between the inorganic carbon pool (CO_2) and bulk biomass (bio) can be calculated in one of two ways: i) From CO_2 to bulk biomass, or ii) From bulk biomass to CO_2. We calculated this value to be consistent with existing literature (i.e. (Popp et al. 1998)) in the fashion that follows. In this notation, a more positive ε_P value means reaction products were more depleted in ^{13}C.

We first calculated isotope fractionations as alpha values ($\alpha_{CO2/bio}$). $\alpha_{CO2/bio}$ is the relative difference between the $^{13}C/^{12}C$ ratios of two materials. This first requires converting $\delta^{13}C$ values to ratios of $^{13}C/^{12}C$ relative to the VPDB standard ($^{13}R_{VPDB}$; R denotes 'ratio'):

$$^{13}R_{sa(VPDB)} = \left(\frac{\delta^{13}C_{sa(VPDB)}}{1000} + 1 \right) \times \ ^{13}R_{std(VPDB)}$$

Equation 2.4

Where $^{13}R_{sa(VPDB)}$ or $^{13}R_{std(VPDB)}$ is the ^{13}R ratio of the sample or standard vs. the VPDB international scale, and $^{13}R_{std(VPDB)} = 0.01107828$ as reported in Meija et al.(2) $\alpha_{CO2/bio}$ is then calculated as:

$$\alpha_{CO_2/bio} = \frac{13R_{CO_2(VPDB)}}{13R_{cells(VPDB)}}$$

Equation 2.5

Then, alpha values were converted to $\varepsilon_{CO2/bio}$ values as:

$$\varepsilon_{CO_2/bio} = \left(\alpha_{CO_2/bio} - 1 \right) \times 1000$$

Equation 2.6

This $\varepsilon_{CO2/bio}$ value is the ε_P value referred to in the text. A summary of all the calculated $\varepsilon_{CO2/bio}$ values are shown in 2.4, and the values used in the text are in Table 2.5.

Strain	Condition	CO_2 Concentration (%)	Light intensity (μE)	ε_P (CO_2/bio) (‰)
WT	Reference Condition	0.04	120	7.453 ± 0.124
ANC	Reference Condition	0.04	120	18.247 ± 0.170
WT	High CO_2	5	120	18.304 ± 0.720
ANC	High CO_2	5	120	19.467 ± 2.897
WT	High Light	0.04	500	7.812 ± 0.172
ANC	High Light	0.04	500	24.299 ± 0.117

Table 2.5: ε_P values used for Figure 2 in main text.
Values (avg. $\pm$ s.e.) are reported relative to VPDB.

2.9.3 Kinetic isotope effect of rubisco

2.9.3.1 Explanation of rubisco assay
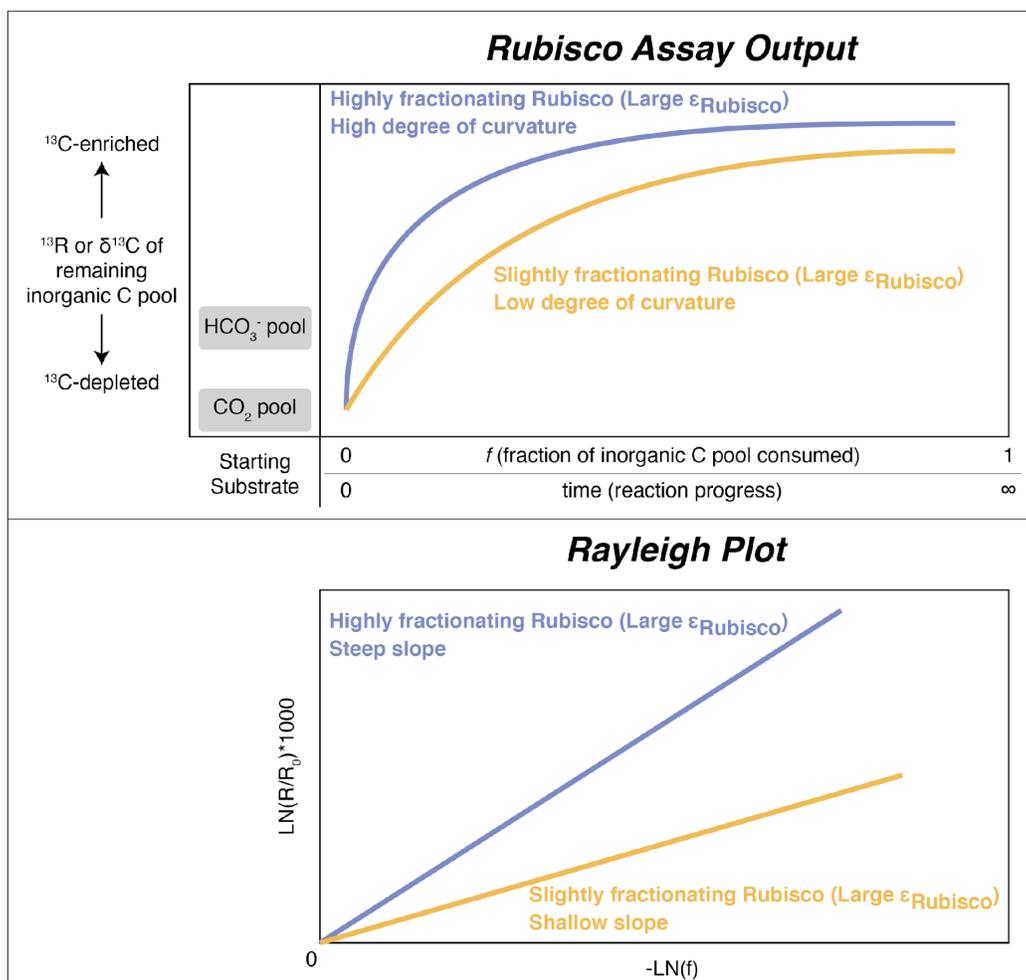
Figure 2.7. Cartoon showing expected results of Rubisco assay for strongly vs. slightly fractionating Rubisco.
Top panel shows measured outputs of $\delta^{13}C$ or ^{13}R values vs. reaction progress or fraction of inorganic C pool consumed (f). Bottom panel shows the log-log version of that plot, which is called a Rayleigh plot. R/R_0 is the ^{13}R ratio of the sample at a given time point vs. the initial ^{13}R ratio of the starting inorganic C pool.

We used the substrate depletion method to measure the kinetic isotope effect catalyzed by Rubisco ($\varepsilon_{Rubisco}$), as used previously in similar studies (Guy et al. 1993; Thomas et al. 2018; McNevin et al. 2006; Scott et al. 2004). In this method, instead of directly measuring the difference in $\delta^{13}C$ of the reactants (i.e. 1 mol CO_2 and 1 mol RuBP) and products (i.e. 2 mol 3PGA), the $\delta^{13}C$ of one of the reactants (CO_2) is measured over time as the reaction goes to completion. One of the reactants is given in excess while the other is limited so that the $\delta^{13}C$ of the reactant pool

eventually asymptotes to a final number as the reaction completes. In previous experiments, RuBP was given in excess while the CO_2 was limited. Finally, $\varepsilon_{Rubisco}$ is calculated by fitting the curvature of the results. This is often done by converting the results to a log-log plot, called a Rayleigh Plot, for ease of fitting. The curvature of this line, or its steepness in log-log space, is proportional to $\varepsilon_{Rubisco}$ - i.e. a rubisco with a large $\varepsilon_{Rubisco}$ will have a high degree of curvature and a larger slope in log-log space, and vice versa (Fig. 2.8).

The assay mix we used is based on previous similar studies. In this set-up, CO_2 is supplied in the form of HCO_3^- which is converted to CO_2 by a carbonic anhydrase, typically derived from bovines. At equilibrium, this would cause the CO_2 pool to be lighter in $\delta^{13}C$ than the HCO_3^- pool (Fig. 2.8). CO_2 and RuBP is then catalyzed by Rubisco to create 3PGA. Therefore, our reaction mixture contains carbonic anhydrase, rubisco, HCO_3^-, and RuBP to create the full reaction, and additional reagents such as: i) $MgCl_2$ to ensure the Rubisco active site is fully and correctly metallated, ii) bicine as a buffer, iii) dithiothreitol (DTT) to prevent oxidizing conditions that can inhibit rubisco activity and stimulate its degradation (Marcus et al. 2003).

In our experiment, instead of limiting CO_2, we limited the other reactant, RuBP. In addition, f (the proportion of CO_2 remaining) must be known from an external measurement. Previous experiments have generally done so by taking a separate aliquot to measure the concentration of CO_2 directly (Guy et al. 1993; Scott et al. 2004). In our experiment, we converted sampling time to f by fitting to the model y = a*EXP(-b*x)+c, based on the fact that the $\delta^{13}C$ of the reactant pool with increase during the reaction and then asymptote to a fixed value as the reaction ceases (i.e. no further carbon isotope discrimination can occur because Rubisco can no longer pull from the CO_2 pool as RuBP runs out). In essence, we are purely looking at the curvature of this line, similar to previous rubisco assays where the $\delta^{13}C$ of the reaction vessel headspace was monitored continually on a membrane inlet mass spectrometer (McNevin et al. 2006) instead of traditional methods where discrete aliquots are taken (Guy et al. 1993). See Section 2.9.3.4 for further explanation.

The rubiscos used here were purified by the Shih lab according to previous methodologies (Saschenbrecker et al. 2007; Banda et al. 2020), and had their kinetics characterized previously (Shih et al. 2016). Briefly, as stated in the main text, clarified lysate from a BL21 DE3 Star *E. coli* culture expressing Rubisco (either the WT *Syn*6301 or β-MRCA) 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. The enzyme was then shipped on dry ice to Caltech, where the rubisco kinetic isotope effect (KIE) assay was performed.

2.9.3.2 Assay preparation and execution

Prior to running the Rubisco KIE assay, the activity of bovine erythrocytes carbonic anhydrase (CA) ordered from Sigma Aldrich (C3934) was checked following the Sigma protocol titled "Enzymatic Assay of Carbonic Anhydrase for Wilbur-Anderson [W-A] Units (EC 4.2.1.1)" (Anon n.d.). We found a value of 3,368 W-A units / mg protein, which exceeded the product stated value of ≥2,000 W-A units / mg protein, and proceeded to use this active CA enzyme prep in the rubisco KIE assay.

First, for the rubisco KIE assay, three external standards were prepared by weighing out Carrara marble standards (CIT_CM2013, $\delta^{13}C$ = +2.0 ± 0.1 (‰VPDB)) into three exetainer vials. Standards were then sealed within each tube, purged with He gas for 5 minutes, and then acidified by needle injection with concentrated phosphoric acid (42% v/v).

Next, three substrate exetainers were prepared. Three exetainer containers were purged with He gas for 5 minutes, and then injected with the substrate (HCO_3^- dissolved in DI water). They were then acidified by needle injection with concentrated phosphoric acid (42% v/v) to convert HCO_3^- to CO_2, and placed in a 70°C water bath for at least 20 minutes to help the reaction go to completion.

Then, 22 exetainer sampling vials were prepared for the WT and ANC rubisco assays (11 each). All tubes were first purged with He gas for 5 minutes, and then injected with ~1 mL of phosphoric acid. The phosphoric acid will both stop the reaction, and convert all C species into CO_2 for analysis.

Next, the reaction assay for each rubisco was prepared. First, a carbonic anhydrase (CA) stock solution was made by dissolving carbonic anhydrase from bovine erythrocytes from Sigma Aldrich (C3934) into DI water. Next, a ribulose 1,5-bisphosphate (RuBP) stock solution was made by dissolving D-Ribulose 1,5-bisphosphate sodium salt hydrate from Sigma Aldrich (R0878) in DI water. Then, one drop of concentrated hydrochloric acid (38% v/v) was added to 20 mL of autoclaved DI water while it was simultaneously stirred with a stir bar and purged with N_2 gas from a tube inserted into the solution. This was all done to remove any residual HCO_3^- or CO_2 in the solution. The solution was purged for 10 minutes. Then, while N_2 gas was blown over the surface of the solution, reagents were added to create a final concentration of 100 mM bicine, 30 mM $MgCl_2$, and 1 mM dithiothreitol (DTT). $NaHCO_3$ from a pre-prepared stock solution was added, and pH was adjusted to 8.5 with NaOH and HCl. CA from the CA stock was added, and then either the WT or ANC rubisco was added to the solution. The solution was gently bubbled with N_2 gas for 10 minutes while rubisco 'activated.'

Next, the syringes used for each WT and ANC assay were cleaned with ethanol and water. We used two separate 25 mL gas-tight syringes with a sample-locking needle from the Hamilton Company for each Rubisco (Ref #86326, Model 1025 SL SYR).

Then, RuBP from the RuBP stock was added to each reaction assay, mixed through pipetting and swirling, and then quickly transferred to their respective gas-

tight syringes. The first time point (t=0 min) was immediately taken after transfer. To sample, ~1 mL of the reaction assay was injected into the pre-prepared sampling exetainer vial so that the phosphoric acid in the vial would stop the reaction and convert all remaining HCO_3^- to CO_2. Each assay was sampled 11 times over 429 minutes.

A control was run in a separate experiment, where all the assay components were mixed together with the exception of a rubisco enzyme. Its isotopic content was monitored through time. The $\delta^{13}C$ of the measured headspace did not change appreciably during this time period, with $\delta^{13}C$ = -0.42 ± 0.03 (‰VPDB) at t = 0 minutes, and $\delta^{13}C$ = -0.55 ± 0.03 (‰VPDB) at t = 277 minutes. The absolute values of these measurements reflect the $\delta^{13}C$ of the substrate used on that experimental day and cannot be related to the WT and ANC data shown here.

2.9.3.3 Isotopic measurement
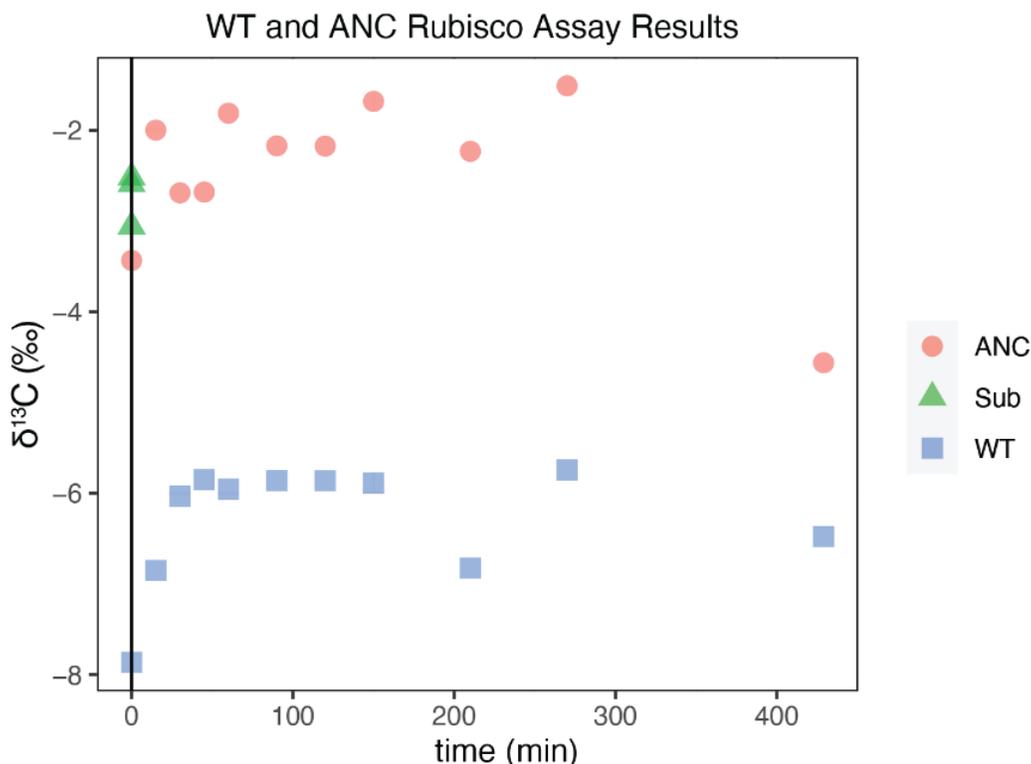
Figure 2.9 Rubisco Assay Results.
Results of WT (blue squares) and ANC (red circles) rubisco assays, shown as $\delta^{13}C$ (‰) vs. time (minutes). Substrate (green triangles) indicates acidified HCO_3^- substrate; it is plotted at t=0 for ease of viewing. Figure was produced using the ggplot2 package (v3.3.6; Wickham, 2016) in R Statistical Software (v4.1.0; R Core Team 2021). Analyses were performed using MATLAB and Statistics Toolbox (vR2020b).

The $\delta^{13}C$ of CO_2 in the headspace of each exetainer was measured on a Delta-V Advantage with Gas Bench and Costech elemental analyzer. Before

measuring samples, two tests were performed to ensure the instrument was functioning normally: i) An 'on/off' test where an internal CO_2 standard was opened and closed to ensure instrument sensitivity and to establish a baseline intensity at a 'zero' CO_2 concentration, and ii) A linearity test where the concentration of CO_2 was increased linearly within the designated sensitivity range of the instrument to ensure that a linear increase in CO_2 concentration corresponds to a linear increase in electrical signal on the collector cups. We measured at three masses (44-46 amu). The instrument was also tuned to ensure that each mass was measured at the center of its mass peak.

The headspace of each sample and standard was measured ten times, with an internal CO_2 reference run before and after each suite of measurements. Each sample, with its ten measurement repetitions, were visually inspected to ensure the sample was being measured within the correct sensitivity range of the instrument (i.e. of similar intensity and pressure as the internal CO_2 reference). Peaks that did not meet this requirement were to be discarded, though no peaks were discarded for this particular assay. The 'raw' $\delta^{13}C$ values were then corrected relative to VPDB using the three standards run. The results of the WT and ANC rubisco assays can be seen in Table 2.6 and Fig. 2.9.

ID	Rep	time (min)	Avg δ¹³C	Std dev δ¹³C	Std err δ¹³C	Avg R	Std dev R	Std err R
Sub	1	0	-3.06	0.19	0.06	0.0111628	1.46E-06	4.60E-07
Sub	2	0	-2.59	0.09	0.03	0.0111664	7.21E-07	2.28E-07
Sub	3	0	-2.52	0.12	0.04	0.0111670	9.15E-07	2.89E-07
ANC	1	0	-3.43	0.36	0.11	0.0111598	2.83E-06	8.96E-07
ANC	2	15	-2.00	0.33	0.10	0.0111711	2.60E-06	8.22E-07
ANC	3	30	-2.69	0.22	0.07	0.0111657	1.72E-06	5.43E-07
ANC	4	45	-2.68	0.19	0.06	0.0111658	1.51E-06	4.79E-07
ANC	5	60	-1.81	0.31	0.10	0.0111726	2.46E-06	7.78E-07
ANC	6	90	-2.17	0.28	0.09	0.0111697	2.21E-06	6.97E-07
ANC	7	120	-2.17	0.19	0.06	0.0111697	1.53E-06	4.85E-07
ANC	8	150	-1.68	0.30	0.09	0.0111736	2.33E-06	7.38E-07
ANC	9	210	-2.23	0.22	0.07	0.0111693	1.75E-06	5.53E-07
ANC	10	270	-1.51	0.29	0.09	0.0111750	2.25E-06	7.11E-07
ANC	11	429	-4.56	0.09	0.03	0.0111509	6.64E-07	2.10E-07
WT	1	0	-7.87	0.23	0.07	0.0111249	1.80E-06	5.68E-07
WT	2	15	-6.85	0.16	0.05	0.0111329	1.28E-06	4.04E-07
WT	3	30	-6.03	0.27	0.08	0.0111394	2.12E-06	6.70E-07
WT	4	45	-5.85	0.23	0.07	0.0111408	1.82E-06	5.75E-07
WT	5	60	-5.96	0.28	0.09	0.0111400	2.23E-06	7.05E-07
WT	6	90	-5.86	0.22	0.07	0.0111407	1.78E-06	5.62E-07
WT	7	120	-5.86	0.23	0.07	0.0111407	1.82E-06	5.76E-07
WT	8	150	-5.89	0.20	0.06	0.0111405	1.56E-06	4.94E-07
WT	9	210	-6.83	0.13	0.04	0.0111331	1.04E-06	3.30E-07
WT	10	270	-5.74	0.23	0.07	0.0111417	1.84E-06	5.82E-07
WT	11	429	-6.48	0.14	0.04	0.0111359	1.10E-06	3.49E-07

Table 2.6. Results of the WT and ANC Rubisco assays.
Avg $\delta^{13}C$ refers to the average of 10 measurement repetitions. Standard deviation (Std dev) and standard error (Std err) are calculated as described.

2.9.3.4 Calculating $\epsilon_{Rubisco}$ and its error

There are two sources of uncertainty that needed to be assessed in the Rayleigh plot; these sources are: 1) The spread in $\delta^{13}C$ or ^{13}R in the final few data

points of the assay; 2) The $\delta^{13}C$ or ^{13}R of the $t = 0$ time point for both assays are different.

  The spread in the last few points of our assay may be due to a variety of reasons, including: 1) Ambient CO_2 contaminating the exetainer containers as they are left out after the reaction; 2) Re-equilibration of the aqueous and gaseous inorganic carbon pools; 3) Instrument error. Since we expect the points to follow an exponential curve that eventually reaches an asymptote, we would therefore expect the points to fall along a straight line in a log-log plot. So, we converted our points from a linear space to a log-log space, systematically fitted lines through different sets of points in this space, and calculated the resulting error. The ^{13}R value for these fits consistently decreased for the ANC assay after data point 10, and after data point 8 for the WT assay. Therefore, we proceeded using data points 1-10 for the ANC assay, and data points 1-8 for the WT assay.

  The other issue in our data is that the $\delta^{13}C$ or ^{13}R of the $t = 0$ time point for both assays are different. We expect them to be similar, since both were given the same inorganic carbon pool to start with. However, the WT assay results are depleted in $\delta^{13}C$ relative to the substrate (Fig. 2.8) even though the remaining inorganic pool should become heavier as Rubisco preferentially uses $^{12}CO_2$ over $^{13}CO_2$ (so that our assay outputs, which sample this remaining pool, gets heavier). It appears that the initial substrate pool is contaminated with isotopically light HCO_3^- or CO_2. Therefore, in order to treat both data sets equally, we did not use the $\delta^{13}C$ values of the HCO_3^- substrate pool, as has been done previously to correct for the fractionation factor between HCO_3^- or CO_2 (Guy et al. 1993) and instead derived the KIE from the curvature of the line (or slope in log-log space) as discussed in Section 3a and as done previously in (McNevin et al. 2006). Therefore, we used $t = 0$ as the initial R_0 value for our starting substrate.

  We converted time to f, the fraction of the inorganic C pool consumed. Since RuBP was the limiting substrate, we could calculate the moles of CO_2 consumed if we assume: i) A 1:1 ratio of RuBP to CO_2 was utilized by Rubisco, and ii) Full consumption of the RuBP pool. In this experiment, 5.47% of the initial CO_2 pool was consumed, or $f = 0.9543$. We then assume that $f = 1$ at $t = 0$, and $f = 0.9543$ at the upper bound of the fit.

  A general model of $y = a*EXP(-b*x)+c$ was used. The model $y = a*EXP(-b*(x-d))+c$ was also tried, but no improvement to the fit occurred so we are only showing the best-fit model to the data. The model was fit three times using non-linear regression using MATLAB's *cftool* interface. The resulting fits and errors of those fits are shown in Table 2.6.

Strain	Fit	Model Fits			Goodness of Fit				
		a	b	c	sse	rsquare	dfe	adjrsquare	rmse
ANC	1	-1.03E-05	0.66312	0.011170	1.19E-10	0.34286	7	0.15511	4.12E-06
ANC	2	-1.03E-05	0.07952	0.011171	8.57E-11	0.52600	7	0.39058	3.50E-06
ANC	3	-9.97E-06	0.03655	0.011171	1.08E-10	0.40244	7	0.23171	3.93E-06
WT	1	-1.54E-05	0.68629	0.011140	2.72E-11	0.88130	5	0.83381	2.33E-06
WT	2	-1.62E-05	0.04995	0.011141	7.11E-12	0.96901	5	0.95662	1.19E-06
WT	3	-1.63E-05	0.05983	0.011141	5.78E-12	0.97484	5	0.96477	1.07E-06

Table 2.7. Model fits for the general model y = a*EXP(-b*x)+c.
sse = Sum of Squares Due to Error or summed square of residuals. rsquare = R-Square value. dfe = Residual Degrees of Freedom. adjrsquare = Degrees of Freedom Adjusted R-Square. rmse = Root Mean Squared Error.

Time was then converted to f using the equation:

$$f = 1 - \left(\frac{R_i - R_1}{R_{upper} - R_1} \times (1 - F) \right)$$

Equation 2.7

Where R_1 is the first measured R value in each set of data, R_{upper} is the fitted value 'c' from the general model y = a*EXP(-b*x)+c, and F = 0.9543, which is calculated from the amount of RuBP added to the assay.

Next, the values were converted to log space so that a Rayleigh plot could be made. We used the equation outlined in (Guy et al. 1993) to transform the R values:

$$y = ln(R/R_0) \times 1000$$

Equation 2.8

Where R_0 is the first R value measured in each series. The f values were transformed by taking the negative natural log. The values were then fit with the model y = m*x + b, and the coefficient 'm' was taken as $\varepsilon_{Rubisco}$. Results and the Rayleigh Plot are shown in Fig. 2.9 and Table 2.7. The average and standard deviation was calculated by averaging the three different 'm' coefficients that came from the three different fits. The standard error was calculated by dividing the standard deviation by the square root of n. The uncertainty in the 95% confidence interval was less than that of the standard deviation, and was therefore ignored for error propagation.

Strain	Fit	m	b
ANC	1	16.23 [16.05, 16.42]	0.009900 [-0.000475, 0.020275]
ANC	2	17.12 [16.94, 17.30]	0.009391 [-0.000460, 0.019243]
ANC	3	18.33 [18.15, 18.52]	0.008776 [-0.000441, 0.017994]
WT	1	24.56 [24.39, 24.72]	0.004023 [-0.004336, 0.012382]
WT	2	25.53 [25.36, 25.70]	0.003873 [-0.004178, 0.011923]
WT	3	25.46 [25.29, 25.62]	0.003884 [-0.004189, 0.011956]

Table 2.8. Fit results of Rayleigh curve.
m and b are the constants in the model y = m*x+b. Values inside brackets indicate 95% confidence interval.

We found the WT (*Syn*6301 Rubisco) $\varepsilon_{Rubisco}$ value to be 25.18 ± 0.31‰ (avg ± s.e.), which was consistent with a previous measurement of a highly similar Form 1B Rubisco from *Synechococcus elongatus* 6301 by (Guy et al. 1993), which found a value of 22.0‰. It is also consistent with other Form IB Rubiscos previously measured: i) 28.2 - 30.3‰ for *Spinacia oleracea* (Guy et al. 1993; Scott et al. 2004; Roeske and O'Leary 1985), and ii) 27.4‰ for *Nicotiana tabacum* (McNevin et al. 2007). See (Thomas et al. 2018; Garcia et al. 2021; Tcherkez et al. 2006) for excellent review and discussion of all currently known and measured Rubisco KIEs. We then found the ANC $\varepsilon_{Rubisco}$ value to be 17.23 ± 0.61‰ (avg ± s.e.).
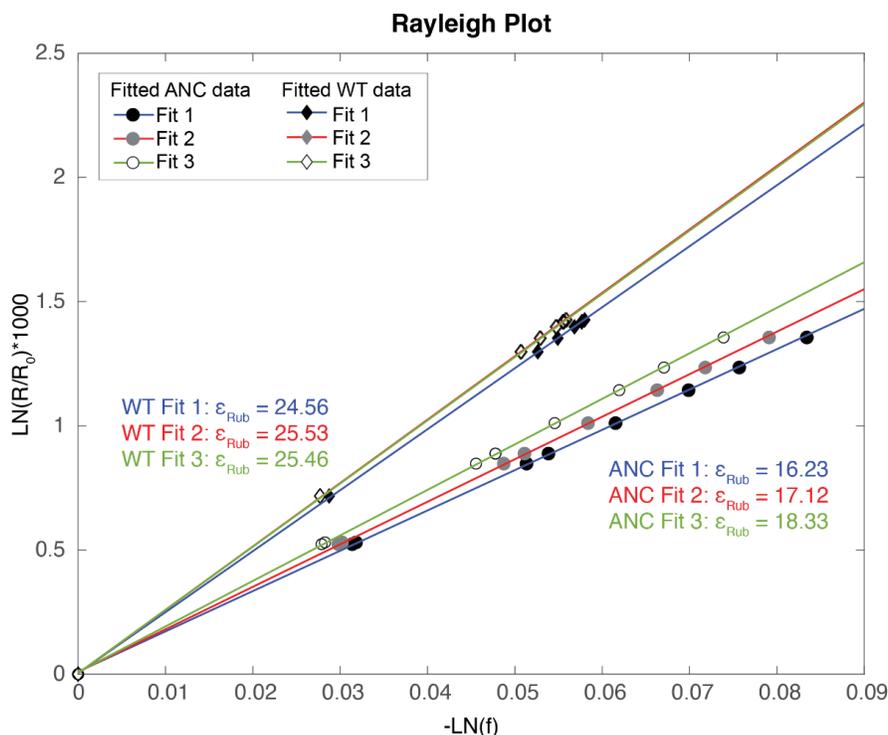
Figure 2.10. Rayleigh plot for WT and ANC Rubisco assays.
ANC data shown in circles; WT data shown in diamonds. Three different fits were done for each strain (Fit 1: blue line, filled black shapes; Fit 2: red line; filled gray shapes; Fit 3: green line; filled white shapes). Fit 2 and 3 overlap for WT and may be hard to see. Analyses were performed using MATLAB and Statistics Toolbox (vR2020b).

2.9.4 Cyanobacterial box models

2.9.4.1 Traditional box model

The "traditional box model" described in the text is a simplified version of the model commonly used to relate ε_P and CO_2 concentrations. We note that this is a dynamic area of research, and that many versions of this model topology exist with minor modifications. In this paper, we present a simplified version that is both accessible to those who are not isotope geochemists, and illustrates the primary relationship of interest – that as ε_P increases, the external concentrations of CO_2 increase as well. The full history of this field cannot be covered here, but we give a brief summary to rationalize the traditional box model presented in the main text, and to give an introductory history to those who are not isotope geochemists.

A. Traditional Box Model Architecture

$C_{external}$
$\delta C_{external}$

Returns to external carbon pool

Φ_{in}
ε_{in}
δ_{in}

$C_{internal}$
$\delta C_{internal}$

Φ_{loss}
ε_{loss}
δ_{loss}

C_{lost}
δC_{lost}

$\Phi_{Rubisco}$
$\varepsilon_{Rubisco}$
$\delta_{Rubisco}$

C_{fixed}
δC_{fixed}

$f = \Phi_{Loss}/\Phi_{in}$

B. Proposed Box Model Architecture

$C_{external}$
$\delta C_{external}$

Returns to external carbon pool

Φ_{in}
ε_{in}
δ_{in}

$C_{internal}$
$\delta C_{internal}$

Φ_{Loss1}
ε_{Loss1}
δ_{Loss1}

C_{lost}
δC_{lost}

Φ_{PCA}
ε_{PCA}
δ_{PCA}

Returns to external carbon pool

$C_{carboxysome}$
$\delta C_{carboxysome}$

Φ_{Loss2}
ε_{Loss2}
δ_{Loss2}

C_{lost}
δC_{lost}

$\Phi_{Rubisco}$
$\varepsilon_{Rubisco}$
$\delta_{Rubisco}$

C_{fixed}
δC_{fixed}

$f_1 = \Phi_{Loss1}/\Phi_{in}$
$f_2 = \Phi_{Loss2}/\Phi_{NDH}$

Figure 2.11: Box Model Architectures.
Model architecture for: A) The traditional box model, and B) Our proposed box model. PCA = Powered Carbonic Anhydrase.

The history of studying and modeling the carbon isotope fractionation of autotrophs (i.e. plants, algae, Cyanobacteria) tracks the birth and maturation of the field of isotope geochemistry. It began with the creation of the first modern, high-resolution mass spectrometers – the fundamental analytical tool that has enabled the field of modern isotope geochemistry – by the American physicist, Alfred O. Nier. Soon after Nier made the first isotopic measurements on a modern, high-resolution sector mass spectrometer (Nier 1936; Nier 1937), his attention soon turned towards the isotopic composition of the natural world. In a seminal paper, Nier and Gulbransen noted the natural variation in carbon isotope ratios among igneous rocks, limestones, plants (in the form of anthracite coal and a modern pine tree), and "unclassified" samples like the air and a modern clam (Nier and Gulbransen 1939). Because of the advanced instrumentation, Nier and Gulbransen were able to improve upon previous studies by showing that these variations were not due to measurement error. Doing so, Nier and Gulbransen made the critical observation that plants tend to "concentrate the light isotope [^{12}C]" in comparison to air.

Later, more systematic measurements of plants and algae were carried out, which resulted in different theories of carbon isotope fractionation by autotrophs, notably a disagreement over if the CO_2 the plant was fixing was solely derived from the atmosphere, or potentially also from CO_2 originating from soils (either

produced by microbial respiration of soil organic matter, or dissolved from limestone substrates) (Wickman 1952; Craig 1953).

The model that has come to dominate the field originated from a seminal study by (Park and Epstein 1960). They measured the carbon isotope ratios of tomato plants at varied CO_2 concentrations and light levels, as well as the carbon isotope fractionation associated with the rubisco enzyme itself. This key measurement allowed the construction of a "two step model" that could explain existing plant and algae data. Their model concluded that the first limiting step was "absorption of the CO_2 from the atmosphere by the leaf," and the second was the "enzymatic conversion of 'dissolved CO_2' in the cytoplasm to carbohydrates." They proposed that the isotopic fractionations of rubisco and diffusion are not additive *in vivo* – instead, they proposed that the net isotopic fractionation *in vivo* (bulk biomass carbon isotope composition) reflects the process by which photosynthesis is being limited. Therefore, if photosynthesis were exclusively limited by diffusion, the bulk biomass fractionation (ε_P) would reflect only the diffusive process ($\varepsilon_P = \varepsilon_{Diffusion}$). And if diffusion did not limit photosynthesis, the bulk fractionation would instead reflect rubisco ($\varepsilon_P = \varepsilon_{Rubisco}$). Finally, they noted that though "[t]he model presented here is necessarily in its simplest form and as such, does not define in detail mechanisms responsible for the C^{13}/C^{12} fractionation in CO_2 fixation," they were still able to explain both their experimental & literature data based on the "two step model."

Farquhar et al. 1982 built upon this key assumption from Park and Epstein – that the isotopic fractionations of diffusion and rubisco are *not* additive *in vivo* (Farquhar et al. 1982). While Farquhar et al. acknowledged that other factors in addition to diffusion and rubisco may affect isotopic fractionation during photosynthesis, their goal was to reconcile most of the differences between observed and expected fractionations, and to create a model so that "measurements of gas exchange physiology and isotopic fractionation" could be made. Importantly, they derived a relationship between the ratio of the partial pressures of atmospheric vs. intercellular CO_2 and the bulk carbon isotope fractionation. This allowed their model to be used to predict changes in plant water use efficiency in photosynthesis & carbon isotope fractionation, since both are tied to opening / closing the stomata (where CO_2 diffuses into the plant). It also allows the CO_2 concentration at the site of rubisco to be estimated from the measured isotopic fractionation.

Interestingly, it was debated in the literature at the time if the isotopic fractionation of each rubisco enzyme *itself* varied. This would be a way to explain variations in ε_p. (Farquhar et al. 1982) does note that (Whelan et al. 1973) found that rubisco fractionation changes with temperature, but that (Christeller et al. 1976) does not. Farquhar et al. state that "it is likely that much of the variation presently evident in the literature reflects experimental uncertainties rather than intrinsic variations in the capacity of the enzyme to fractionate carbon isotopes" (Farquhar et al. 1982). Therefore, the current isotope models built upon & after (Farquhar et al. 1982) all make the assumption that the isotopic fractionation of rubisco is constant.

This "two-step model," largely based on Park and Epstein, can be derived for plants as follows. In this model architecture, Fig 2.10A and Main Text Figure 2.1B, carbon can be: i) external to the cell ($C_{external}$ or C_{ext}), ii) inside the cell ($C_{internal}$ or C_{int}), or iii) fixed by the cell into biomass (C_{fixed}). Carbon that enters the cell but does not get fixed by Rubisco is assumed to eventually be lost by the cell, and return to the external carbon pool (C_{lost}).

We used the classic Hayes isotope flux model system to evaluate our results (Hayes 2001). In this approach, each flux has its own isotopic fractionation (ε), as well as carbon isotope composition (δ). For the carbon pools, this δ refers to the isotopic composition of the pool. For the fluxes, δ refers to the instantaneous isotopic composition of that flux (see (Hayes 2001) for a detailed review). We also made a set of simplifying assumptions: i) The system is at steady state, ii) The external carbon pool is infinitely large compared to the cell (i.e. its carbon isotope composition does not change). We first defined the isotopic relationships for each flux in our system:

$$\delta_{in} = \delta C_{ext} + \epsilon_{in}$$

Equation 2.9

$$\delta_{loss} = \delta C_{int} + \epsilon_{loss}$$

Equation 2.10

$$\delta_{Rubisco} = \delta C_{int} + \epsilon_{Rubisco}$$

Equation 2.11

We will also define ε_P as the difference in $\delta^{13}C$ of the external vs. fixed carbon pools, i.e.:

$$\varepsilon_P = \delta C_{ext} - \delta C_{fixed}$$

Equation 2.12

Most of these models are solved with the assumption of steady state, which we will assume as well. We can then define the mass balance relationships with φ denoting fluxes; φ_{in} is the flux of carbon into the cell, φ_{loss} is carbon loss from the cell, and $\varphi_{Rubisco}$ is carbon that is fixed by rubisco:

$$\varphi_{in} = \varphi_{loss} + \varphi_{Rubisco}$$

Equation 2.13

The traditional model assumes that the amount of carbon entering the cell is inversely proportional to a concentration gradient of pCO_2 inside vs. outside of the cell, or that $\Phi_{out}/\Phi_{in} = [C_{int}]/[C_{ext}]$. So, we can then define a loss fraction:

$$f = \frac{\varphi_{loss}}{\varphi_{in}}$$

<div align="right">Equation 2.14</div>

The isotopic relationships and mass balance equations were combined to create an isotope mass balance equations:

$$\varphi_{in}\delta_{in} = \varphi_{loss}\delta_{loss} + \varphi_{Rubisco}\delta_{Rubisco}$$

<div align="right">Equation 2.15</div>

These sets of equations can be solved symbolically to arrive at the solution:

$$\varepsilon_P = (1-f)(\varepsilon_{in}) + f(\varepsilon_{Rubisco})$$

<div align="right">Equation 2.16</div>

This solution is plotted as the green line in Figure 2.11A and referred to as the 'plant-based' model.

Much work was done after this to adapt the plant-based model to algae. The main modification done was to account for active C_i uptake in the form of HCO_3^- or CO_2 (Sharkey and Berry 1985). The (Sharkey and Berry 1985) model is very similar to the plant-based model in that: 1) A linear relationship exists between ε_P and inorganic carbon (C_i) leakage out of the cell (defined as F_3/F_1 in Sharkey & Berry (1985), and defined as $f=\Phi_{loss}/\Phi_{in}$ in this study); 2) ε_P cannot exceed $\varepsilon_{Rubisco}$. We have plotted the plant-based model vs. the Sharkey & Berry (1985) in Figure 2.11A below – the slope of both lines is set by $\varepsilon_{Rubisco}$, and the models only differ by their y-intercept. This is because active C_i uptake was a known part of the CCM, and (Sharkey and Berry 1985) took this into account by assuming all C_i entering was HCO_3^- (flux F_1 in (Sharkey and Berry 1985) Figure 2.4). This causes the C_i pool inside the cell to be ≈8‰ enriched in ^{13}C, which causes the y-intercept to be more negative (in this community's framework, a positive ε_P value means ^{13}C-depletion while a more negative ε_P value means ^{13}C-enrichment). This is plotted as the blue line in Figure 2.11A and referred to as the Sharkey & Berry model.

(Popp et al. 1998) and (Laws et al. 1995) also made key contributions by extending this plant-based model to algae. Popp et al. 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. Interestingly, they found cyanobacterial ε_p to be roughly constant independent of environmental pCO_2 and growth rate. (This is in contrast to contemporaneous studies in Cyanobacteria at the time that *did* find cyanobacterial ε_p varies with pCO_2 (Erez et al. 1998).) They hypothesized that this invariance 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 related to light were known in Cyanobacteria at the time (Gimmler et al. 1990; Rotatore et al. 1992; S ltemeyer et al. 1993). They note, "Although results of our experiments suggest that CO_2(aq) does not cross the plasmalemma by passive diffusion alone, but rather is supplemented by an active transport mechanism, the inescapable conclusion is that ε_p nonetheless varies as a linear function of growth rate, [CO_2(aq)] and the cellular-carbon-to-surface-area ratio under most natural conditions." In other words, the simple linear relationship between pCO_2 and ε_p in C3 plants appeared to hold up in algae and Cyanobacteria as well.
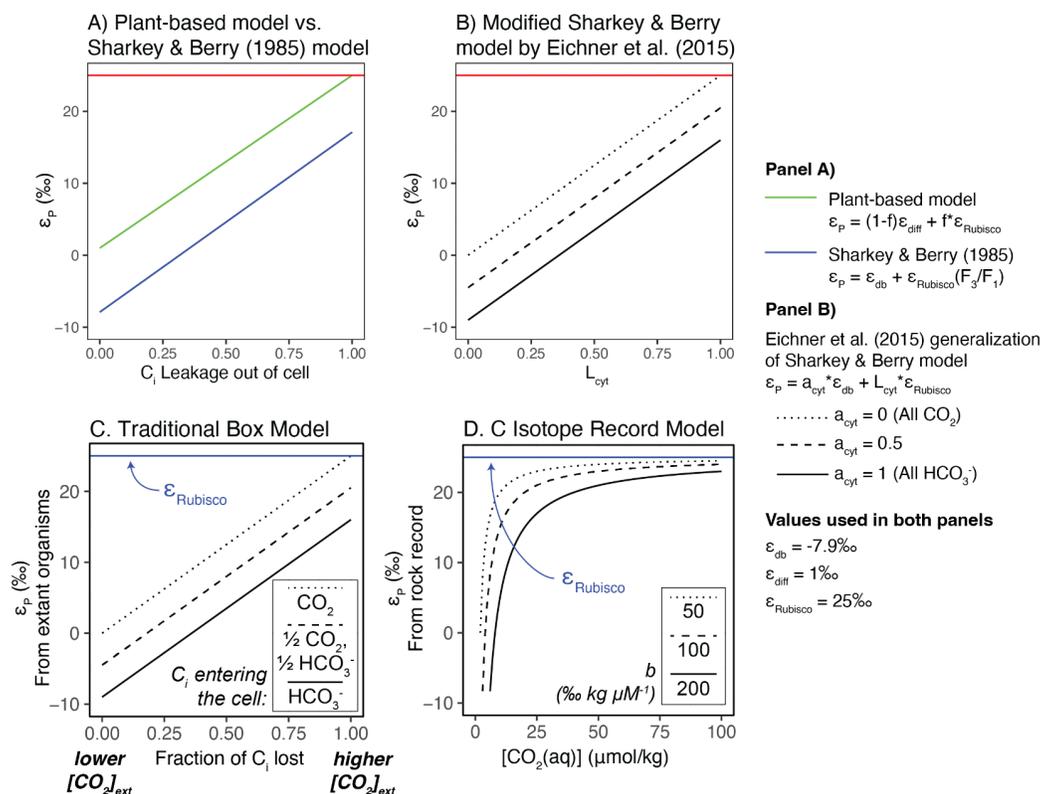
Figure 2.12: Variations on the "Traditional Model" and the "C Isotope Record Model."
A) The plant-based model we derived in the supplemental is shown in green, while the model proposed by Sharkey and Berry (1985) for algae generally is shown in blue. Both models have the slope of $\varepsilon_{Rubisco}$ (25‰ is used as an example here). They are offset by the equilibrium fractionation of $CO_2 \leftrightarrow HCO_3^-$, where HCO_3^- is ^{13}C-enriched relative to CO_2 (in this field's reference frame, a more negative isotopic value). The equations for each model are given in the right panel of the figure; for simplicity, we label the x-axis as "C_i leakage out of the cell" because it is named differently in each model (f in our derivation; F_3/F_1 for (Sharkey and Berry 1985) Figure 2.4). B) The Eichner et al. (2015) generalization of the Sharkey and Berry model. (Eichner et al. 2015) derives a two-compartment cyanobacterial model that can be generalized to the Sharkey and Berry model, as well as the plant-based model. The equation is shown in the right-most panel, and results in a line with a slope of $\varepsilon_{Rubisco}$ and a y-intercept set by the term $a_{cyt}*\varepsilon_{db}$ to show if the total C_i uptake is primarily CO_2 ($a_{cyt}=0$) or primarily HCO_3^- ($a_{cyt}=1$). When $a_{cyt}=0$, you effectively get the plant-based model in Panel A), and when $a_{cyt}=1$, you get the Sharkey & Berry model in Panel A). Other values used are $\varepsilon_{Rubisco}$ (fractionation of the enzyme rubisco) = 25‰, ε_{diff} (fractionation of CO_2 diffusing into the cell) = 1‰. For ε_{db}, the fractionation of the $CO_2 \leftrightarrow HCO_3^-$ equilibrium, Sharkey & Berry (1985) used a value of -7.9 while Eichner et al. (2015) uses a value of -9. All analyses were performed using R Statistical Software (v4.1.0; R Core Team 2021) and figures were produced using the ggplot2 package (v3.3.6; Wickham, 2016). C) The traditional box model as shown in the main text. ε_P values are measured from extant organisms in the lab. D) The C Isotope record model. ε_P values are derived from the rock record. Both C) and D) have an upper limit where $\varepsilon_P = \varepsilon_{Rubisco}$.

Many versions of this traditional model exist. (Eichner et al. 2015) presents a nice version of the traditional model that is stated in their study as a generalization of the Sharkey & Berry (1985) model (Equation 15 in Eichner et al.) that we are citing and presenting as the "traditional" model in Figure 1 in the main text. It

relates the plant-based model and the Sharkey & Berry model by introducing the term a_{cyt}, which varies the proportion of CO_2 vs. HCO_3^- in total C_i uptake (Figure 2.1; Figure 2.11B). We use the Eichner et al. ε_{db} value of -9‰ instead of Sharkey & Berry (1985) ε_{db} value of -7.9‰. Essentially, in the Eichner et al. version of the Sharkey & Berry model, when $a_{cyt} = 0$, all C_i uptake is CO_2 and you get the plant-based model. When $a_{cyt} = 1$, all C_i uptake is HCO_3^- and you get the Sharkey & Berry (1985) model solution. We note that all of these models have the key limitation that ε_P cannot exceed $\varepsilon_{Rubisco}$.

The final step was to extend this model to environments both modern and ancient. Francois et al. 1993 and Rau et al. 1989 both found, from measuring the carbon isotope composition of particulate organic matter (POM) or phytoplankton from ocean surface waters, that concentrations of dissolved CO_2 were correlated with ε_P values (Francois et al. 1993; Rau et al. 1989). These studies were notable because they showed the prior model calibrated in the lab could potentially be extended to the field, and that a model calibrated in plants even seemed to hold in algae. In addition, Hayes formalized the above model into an isotope flux model that is the dominant mathematical form used to model autotrophic carbon isotope fractionation today (Hayes 1993). Hayes also increased the model's detail by predicting the isotopic composition of specific metabolic intermediates, and by extending this model to new metabolic systems like eukaryotic lipid biosynthesis. He also noted that values of ε_P derived from the carbon isotope record may "provide information about the nature of the primary producer organisms and their environment" like "CO_2 paleobarometry."

(Popp et al. 1989) had previously determined the isotopic compositions of sedimentary porphyrins, but did not estimate paleo-CO_2 levels because their model had empirically fit parameters (i.e. "b") that they could not determine for ancient environments and materials. This b term is an empirically fit slope that "quantifies the rate at which ε_P decreases as concentrations of CO_2 become smaller," and is related to ε_P by the relationship $\varepsilon_P = \varepsilon_f - b/C_e$, where ε_f is the isotopic fractionation of all carbon-fixation reactions active in the cell but mainly rubisco, and C_e is the concentration of dissolved CO_2 (Bidigare et al. 1997). The term b effectively sets how quickly ε_P approaches the limit of $\varepsilon_{Rubisco}$ (Figure 2.11D). (Freeman and Hayes 1992) subsequently showed that, indeed, they could calculate ancient CO_2 levels up to 100 Ma after calibrating the empirical "b" value from (Popp et al. 1989). Much work continues today empirically calibrating this model so that it can be applied to geologic time (Zhang et al. 2020). Common values used for b are on the order of magnitude ≈ 100‰ kg μM^{-1} (Witkowski et al. 2018). Overall, both the C Isotope Record Model and the Traditional Model have a limit where ε_P cannot exceed $\varepsilon_{Rubisco}$ (Figure 2.11C,D).

We refer to this as the "C Isotope Record Model" in the main text. It is derived from work based on model organisms in the lab (i.e. the Traditional Box Model shown in Figure 2.1 and 2.11C) because the parameter b is derived from bench-top lab experiments.

2.9.4.2 Proposed box model

Our proposed model incorporated two more boxes and an additional isotope fractionation factor (Figure 2.10B). Therefore, the four main reservoirs are: i) Carbon that is external to the cyanobacterial cell (C_{ext}); ii) Carbon inside the cell (C_{int}); iii) Carbon in the carboxysome (C_{carb}); and iv) Carbon that is fixed into biomass (C_{fixed}). The three isotope effects are: i) Diffusion into the cell (ε_{in}); ii) Fractionation by the a powered carbonic anhydrase which catalyzes the unidirectional hydration of CO_2 to HCO_3^- (ε_{PCA}); iii) Fractionation by rubisco during carbon fixation ($\varepsilon_{Rubisco}$). For ε_{in} a value of 1(‰VPDB) was used based on the diffusion of CO_2 in water (O'Leary 1984). For ε_{PCA}, a wide range of values exist in the literature based on both lab experiments and *ab initio* calculations using transition state theory, but they range from 13-39(‰VPDB) as shown in Wilkes and Pearson (2019), which offers an excellent discussion on the topic, and we direct the reader to that paper for further reading (Wilkes and Pearson 2019). We used a value of 30 (‰VPDB) based on a previous study by Eichner et al. (2015), who used this value to model C isotope fractionation by the NDH-1$_4$ complex in their model organism *Trichodesmium erythraeum* IMS101 (Eichner et al. 2015). For $\varepsilon_{Rubisco}$, two different values were used for either the WT or ANC strain, based on *in vivo* measurements done for this paper. $\varepsilon_{Rubisco} = 17.23 \pm 0.61$ (‰VPDB) for ANC, and $\varepsilon_{Rubisco} = 25.18 \pm 0.31$ (‰VPDB) for WT. These values were derived as detailed in Section 2.9.3.

Finally, we then permitted two pathways for loss in our system. The first flux is for C that diffuses into the cell, but then exits the cell and does not continue into the carboxysome (φ_{loss1}). The second flux is for C that enters the carboxysome but is not fixed by Rubisco, and then exits the cell (φ_{loss2}).

We again use the classic (Hayes 2001) isotope model to model our system. This model assumes that the system is at steady state. We defined the isotopic relationships for each box and flux in our system:

$$\delta_{in} = \delta C_{ext} + \varepsilon_{in}$$

Equation 2.17

$$\delta_{loss1} = \delta C_{int} + \varepsilon_{loss1}$$

Equation 2.18

$$\delta_{PCA} = \delta C_{int} + \varepsilon_{PCA}$$

Equation 2.19

$$\delta_{loss2} = \delta C_{carb} + \varepsilon_{loss2}$$

Equation 2.20

$$\delta_{Rubisco} = \delta C_{carb} + \varepsilon_{Rubisco}$$

Equation 2.21

We then defined the bulk cyanobacterial fractionation, ε_P, as:

$$\varepsilon_P = \delta C_{ext} - \delta C_{fixed}$$

Equation 2.22

Since there is only one path for the last flux into the C_{fixed} box,

$$\delta_{Rubisco} = \delta C_{fixed}$$

Equation 2.23

So:

$$\varepsilon_P = \delta C_{ext} - \delta_{Rubisco}$$

Equation 2.24

As in the prior section, we then defined the mass balance relationships with φ denoting fluxes; φ_{in} is the flux of carbon into the cell, φ_{loss1} and φ_{loss2} are carbon loss from the cell, and φ_{PCA} is carbon that goes through a hypothetical powered carbonic anhydrase (PCA), and $\varphi_{Rubisco}$ is carbon that is fixed by rubisco:

$$\varphi_{in} = \varphi_{loss1} + \varphi_{PCA}$$

Equation 2.25

$$\varphi_{PCA} = \varphi_{loss2} + \varphi_{Rubisco}$$

Equation 2.26

We also defined the two loss fractions, f_1 and f_2:

$$f_1 = \frac{\varphi_{loss1}}{\varphi_{in}}$$

Equation 2.27

$$f_1 = \frac{\varphi_{loss2}}{\varphi_{PCA}}$$

Equation 2.28

The isotope relationships and mass balance equations were combined to create the isotope mass balance equations:

$$\varphi_{in}\delta_{in} = \varphi_{loss1}\delta_{loss1} + \varphi_{PCA}\delta_{PCA}$$

Equation 2.29

$$\varphi_{PCA}\delta_{PCA} = \varphi_{loss2}\delta_{loss2} + \varphi_{Rubisco}\delta_{Rubisco}$$

Equation 2.30

These set of equations was solved symbolically to arrive at the solution:

$$\varepsilon_P = \varepsilon_{loss2} - \varepsilon_{loss1} - \varepsilon_{in} + f_1(\varepsilon_{PCA} - \varepsilon_{loss1}) + f_2(\varepsilon_{Rubisco} - \varepsilon_{loss2})$$

Equation 2.31

Equation 2.31 was solved analytically as described in the section above, except two different f vectors were inputted: f_1 and f_2. See GitHub for code for plotting and solving at https://github.com/reneezwang/ancestral-rubisco-cyano. Full model results are shown in Fig. 2.12, Panel B. Figure 2.4C in the main text shows solutions for $f, = 0.1$, which is denoted as shown in Fig. 2.12 Panel B.

In addition, we focused only on C_i uptake as CO_2 because we are interested in a model that could achieve more negative ε_p values (^{13}C-depleted biomass), and HCO_3^- uptake (i.e. through bicarbonate pumps like BicA, SbtA, or BCT1 (Price et al. 2013)) would not help us because it would shift all εp values to be ≈8‰ more positive (^{13}C-enriched biomass).

Model outputs are discussed in the main text, and we note that our model is *highly* idealized – we tried to modify the traditional model as little as possible to explain our data, which was to achieve $\varepsilon_p > \varepsilon_{Rubisco}$ with physiologic consequences that make sense. We wanted to demonstrate with our simple, proposed model that just slight modifications to the traditional model can start to harmonize our experimental results with model outputs. This may allow for future modeling avenues that can continue to augment our understanding of carbon isotope fractionation within bacterial autotrophs.

In addition, as discussed in the main text, using 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. This is shown in Figure 2.11, panel C. This is due to uncertainty in how 'one-way' the CO_2 hydration reaction is in the isotopic equilibrium reaction $^{12}CO_2 + H^{13}CO_3^- \rightleftharpoons {}^{13}CO_2 + H^{12}CO_3^-$. The full chemical reactions are shown in Figure 2.11 panel A, with the CO_2 hydration denoted with the reaction constant k_+, and the dehydration reaction denoted by k_- per similar notation used by (Zeebe and Wolf-Gladrow 2001). The carbon isotope reactions for CO_2 hydration and dehydration are denoted as shown in Figure 2.11B. A separate reaction and rate constant can be calculated or measured for each isotopic species, i.e. $^{12}k_+$ and $^{13}k_+$ for $^{12}CO_2$ and $^{13}CO_2$ respectively. The isotope effect is then calculated and reported either in the alpha (α) or epsilon (ε) notation as shown. (We note that in this field, the convention is to calculate α by taking the ratio of k's with ^{12}k value in the numerator. Putting ^{13}k in the numerator would just cause the epsilon value to be negative instead of positive.) The equilibrium isotope effect (ε_{Equil}) is the difference between that of the forward reaction (ε_{HYD}) and the reverse reaction (ε_{DEHYD}). We use a value of $\varepsilon_{Equil} = -9‰$, indicating that in the reaction $^{12}CO_2 + H^{13}CO_3^- \rightleftharpoons {}^{13}CO_2$

+ $H^{12}CO_3^-$, ^{13}C slightly prefers to partition to HCO_3^-. This also means that ε_{HYD} and ε_{DEHYD} must be offset by 9‰, so if $\varepsilon_{HYD} = 30$‰ then $\varepsilon_{DEHYD} = 39$‰.

As noted above, the values for ε_{HYD} and ε_{DEHYD} are debated. This is partially due to mass balance, as described in (Zeebe and Wolf-Gladrow 2001): "If the reactant is completely transformed into product, then the final isotope ratio of the product will be identical to the initial isotope ratio of the reactant, irrespective of whether the reaction rate is sensitive to the mass of the reacting species or not. This is a result of conservation of mass: just as in a pipeline, everything that goes in - including neutrons - will eventually have to come out (Hayes 1982). Thus, for a kinetic isotope effect to be expressed, an incomplete reaction is required." In addition, the values for ε_{HYD} and ε_{DEHYD} are debated because the strictly one-way reaction of hydration or dehydration is difficult to measure experimentally, so this isotope effect is typically calculated based on transition state theory models. As noted above, (Wilkes and Pearson 2019) gave a good discussion on this in their Section 2.4 and their supplemental Table S4 summarizes their review. They recommend using a value of $\varepsilon_{HYD} = 25$‰, which would set $\varepsilon_{DEHYD} = 34$‰. We used $\varepsilon_{HYD} = 30$‰ (referred to as ε_{VCA} in our paper) to be consistent with (Eichner et al. 2015).

A) CO_2 Hydration / Dehydration Reactions

Net Chemical Reaction $CO_2 + H_2O \; \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \; HCO_3^- + H^+$

CO_2 Hydration
(Forward Reaction)

$^{12}CO_2 + H_2O \xrightarrow{^{12}k_+} H^{12}CO_3^- + H^+$

$^{13}CO_2 + H_2O \xrightarrow{^{13}k_+} H^{13}CO_3^- + H^+$

$\alpha_{HYD} = {}^{12}k_+ / {}^{13}k_+ = 1.030$
$\varepsilon_{HYD} = (\alpha_{HYD} - 1)*1000 = 30‰$

CO_2 Dehydration
(Reverse Reaction)

$H^{12}CO_3^- + H^+ \xrightarrow{^{12}k_-} {}^{12}CO_2 + H_2O$

$H^{13}CO_3^- + H^+ \xrightarrow{^{13}k_-} {}^{13}CO_2 + H_2O$

$\alpha_{DEHYD} = {}^{12}k_- / {}^{13}k_- = 1.039$
$\varepsilon_{HYD} = (\alpha_{DEHYD} - 1)*1000 = 39‰$
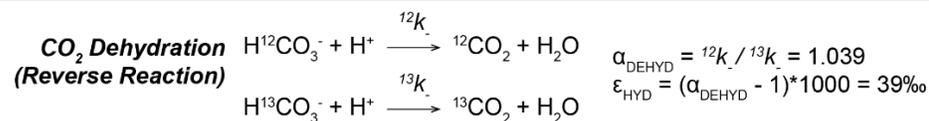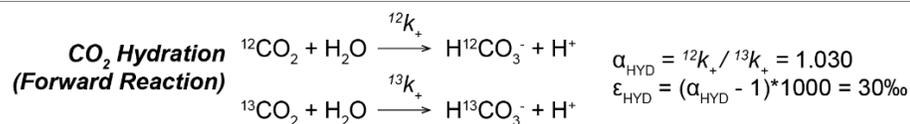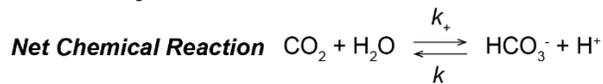
B) Solely CO_2 hydration *(k_+ only; ε_{VCA} = 30‰)*

Ref. Condition – –
High CO_2 ·····
High Light ——

C) Primarily CO_2 hydration *($k_+ > k_-$; ε_{VCA} = 10‰)*

Ref. Condition – –
High CO_2 ·····
High Light ——
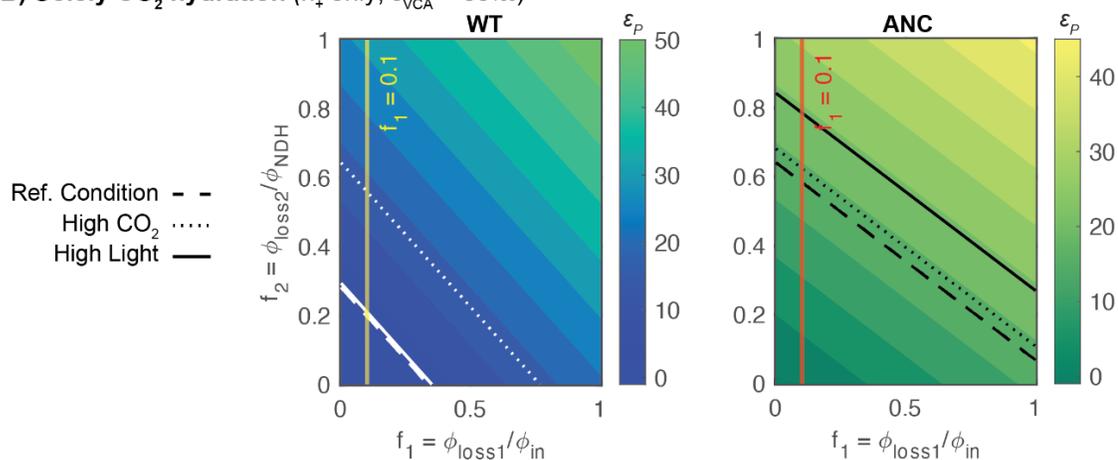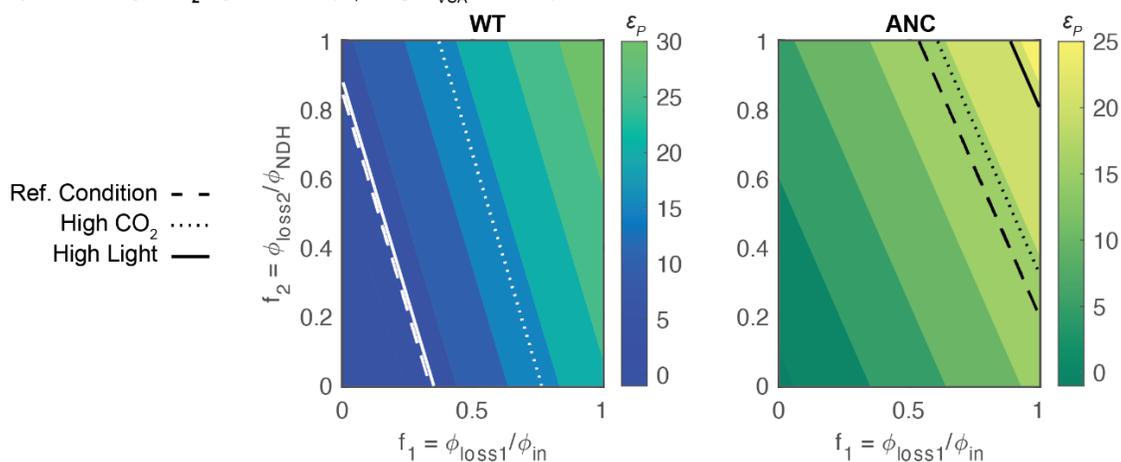
Figure 2.13: Full model outputs for the proposed box model.
A) CO_2 hydration / dehydration reactions. A separate rate constant, k_+ and k_-, is defined for the forward and reverse reactions respectively per similar notation used by (Zeebe and Wolf-Gladrow 2001). A separate reaction and rate constant can then be defined for each isotopic species, i.e. $^{12}k_+$ and $^{13}k_+$ for $^{12}CO_2$ and $^{13}CO_2$ respectively. The isotope effect is then calculated and reported either in the alpha (α) or epsilon (ε) notation as shown. The equilibrium isotope effect (ε_{Equil}) is the difference between that of the forward reaction (ε_{HYD}) and the reverse reaction (ε_{DEHYD}). We use a value of $\varepsilon_{Equil} = -9‰$, indicating that in the reaction $^{12}CO_2 + H^{13}CO_3^- \rightleftharpoons {}^{13}CO_2 + H^{12}CO_3^-$, ^{13}C slightly prefers to partition to HCO_3^-. This also means that ε_{HYD} and ε_{DEHYD} must be offset by 9‰, so if ε_{HYD} = 30‰ then ε_{DEHYD} = 39‰. See section 4b for further discussion on picking exact values for ε_{HYD} (referred to as ε_{VCA} in our paper). B) Model outputs assuming ε_{VCA} = 30‰; i.e. reaction is solely CO_2 hydration. $f_l = 0.1$ is denoted with either a yellow or red solid line for WT or ANC respectively. Mean experimental ε_P values for each condition are shown as diagonal lines as follows: 1) Dashed line is the reference condition; 2) Dotted line is the high CO_2 condition; 3) Solid line is the high light condition. C) Model outputs assuming ε_{VCA} = 30‰; i.e. reaction is primarily CO_2 hydration and $k_+ > k_-$. Analyses and visualization were performed using MATLAB and Statistics Toolbox (vR2020b).

2.9.4.3 Fitting our data with other models
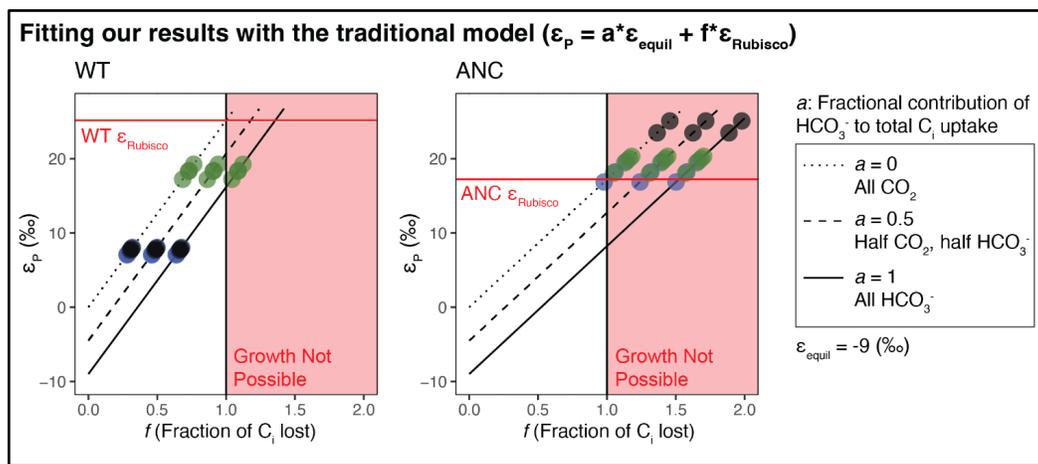
Figure 2.14: ANC data cannot be rationalized with the traditional model.
Measured ε_P values for each strain (circles) were fit with the traditional model at varying C_i uptake compositions (lines). Blue circles indicate reference condition (ambient pCO_2 (0.05% (v/v)), standard light flux (120 µE)); Green circles indicate high CO_2 condition (5% pCO_2 (v/v), 120 µE); Black circles indicate high light condition (0.05% pCO_2 (v/v), 500 µE). Dotted lines shows traditional model solution with C_i uptake as 100% CO_2; solid line shows C_i uptake as 100% HCO_3^-; dashed line shows C_i uptake as 50% CO_2, 50% HCO_3^-. The $\varepsilon_{Rubisco}$ values used for WT and ANC were 25.18‰ and 17.23‰ respectively. Solid red line indicates where $\varepsilon_P = \varepsilon_{Rubisco}$. We use the same ε_{equi} value of -9‰ as used in (Eichner et al. 2015). All analyses were performed using R Statistical Software (v4.1.0; R Core Team 2021) and figures were produced using the ggplot2 package (v3.3.6; Wickham, 2016).

We fit our data with three other algal carbon isotope models to see if they could rationalize our results – the (Sharkey and Berry 1985) model, the (Erez et al. 1998) model, and the (Eichner et al. 2015) model.

Sharkey and Berry measured the carbon isotope fractionation of plants and eukaryotic algae, *Chlamydomonas reinhardtii*, grown at varied pCO_2 conditions

and derived a model for carbon isotope fractionation by algae that accounts for the algal CCM (see Figure 4 and Equation 2 in (Sharkey and Berry 1985); re-written in Figure 2.11A). This model accounted for the CCM by taking into account active C_i uptake, and it assumed that all C_i entering the cell was in the form of HCO_3^- and that all C_i lost from the cell is as CO_2. They defined the loss of C_i from the cell as the ratio of two relative fluxes, F_3 and F_1, which are plotted on the x-axis in Figure S9A. We plotted our measured ε_P values (colored circles) using this model and got C_i leakage values (F_3/F_1) that exceeded 1 for all ANC data, and for WT High CO_2 data. Leakage values greater than 1 imply that the cell is not fixing any carbon, which is incompatible with our growth curve data (i.e. ANC grew in all conditions, and was therefore fixing carbon).

(Erez et al. 1998) grew batch cultures of the cyanobacterium *Synechococcus* sp. PCC7942 (the same parent strain used in this study) bubbled with ambient lab air and found ε_p values up to 33‰, greater than $\varepsilon_{Rubisco}$ values known at the time (28 or 22‰). This result is in contrast to Popp et al. who found using *Synechococcus* sp. CCMP838 that cyanobacterial ε_p values do not vary with growth rate or changing CO_2 concentrations or exceed known cyanobacterial $\varepsilon_{Rubisco}$ values (Popp et al. 1998). Therefore, Erez et al. also need an additional fractionation factor to explain their data, and presented a model in their Equation 4 that modifies the (Sharkey and Berry 1985) model by adding a separate compartment for the carboxysome. They also invoke a "CA-like" enzyme that catalyzes the one-way hydration of CO_2, which both scavenges CO_2 lost from the carboxysome and introduces an additional isotopic fractionation factor since the isotopic fractionation of this reaction is thought to be large (they test 12‰ and 15‰ as potential values). We are interested in the relationship between ε_p and C_i lost, which is the difference in C_i lost (F_3 in their Figure 6) versus C_i uptaken (F_1 in their Figure 6). So, we rearranged Equation 4 in Erez et al. using Equation 1 in Erez et al. to derive the equation:

$$\epsilon_P = X\epsilon_{equil} + \epsilon\left(\frac{F_3}{F_1}\right)$$

Equation 2.32

Where X is the fraction of CO_2 to total C_i uptake (X=1 is all CO_2, X=0 is all HCO_3^-). The modification to the Sharkey and Berry model is the addition of this term, X. The Erez model was able to largely rationalize ANC ε_p data (F_3/F_1<1), but only if all C_i uptake is CO_2 (X=1), and it gives extremely high leakage values for the high light condition (0.99 and 0.90) (Figure 2.14B). In addition, if X=1 for WT, then implausible negative values for leakage (F_3/F_1) are calculated for three of the four reference condition replicates (-0.04, -0.02, -0.03) and all the high light replicates (-0.003, -0.01). Overall, the Erez model implies that C_i leakage is overall higher for ANC vs. WT. In addition, their model only fits ANC ε_p values in the unlikely scenario that all C_i uptake by ANC is CO_2.

A) Sharkey & Berry (1985): $\varepsilon_P = \varepsilon_{equil} + \varepsilon_{Rubisco}(F_3/F_1)$

Sharkey Model: WT

Sharkey Model: ANC

WT $\varepsilon_{Rubisco}$

Growth Not Possible

ANC $\varepsilon_{Rubisco}$

Growth Not Possible

ε_P (‰)

ε_P (‰)

F_3/F_1

F_3/F_1

Condition

● High CO2 (green)
● High Light (black)
● Reference Condition (blue)

ε_{equil} = -7.9 (‰)
WT $\varepsilon_{Rubisco}$ = 25.18 (‰)
ANC $\varepsilon_{Rubisco}$ = 17.23 (‰)

B) Erez et al. (1998): $\varepsilon_P = X\varepsilon_{equil} + \varepsilon_{Rubisco}(F_3/F_1)$

Erez Model: WT

Erez Model: ANC

WT $\varepsilon_{Rubisco}$

Growth Not Possible

ANC $\varepsilon_{Rubisco}$

Growth Not Possible

ε_P (‰)

ε_P (‰)

F_3/F_1

F_3/F_1

X = Fraction of CO_2 to total C_i uptake

····· $X = 1$
All CO_2

-- - $X = 0.5$
Half CO_2, half HCO_3^-

—— $X = 0$
All HCO_3^-

ε_{equil} = +8 (‰)
WT $\varepsilon_{Rubisco}$ = 25.18 (‰)
ANC $\varepsilon_{Rubisco}$ = 17.23 (‰)

C) Eichner et al. (2015): $\varepsilon_P = a_{cyt}\varepsilon_{equil} + L_{cyt}(a_{carb}\varepsilon_{cyt} + L_{carb}\varepsilon_{Rubisco})$

Eichner model: WT

Eichner model: ANC

WT $\varepsilon_{Rubisco}$

Growth Not Possible

ANC $\varepsilon_{Rubisco}$

Growth Not Possible

ε_P (‰)

ε_P (‰)

L_{cyt}

L_{cyt}

L_{carb}: Fractional leakage of C_i out of the carboxysome

····· $L_{carb} = 1$
All C_i lost

-- - $L_{carb} = 0.5$
Half of C_i lost

—— $L_{carb} = 0$
No C_i lost

ε_{equil} = -9 (‰), ε_{cyt} = 30 (‰)
a_{carb} = 1, a_{cyt} = 0.8
WT $\varepsilon_{Rubisco}$ = 25.18 (‰)
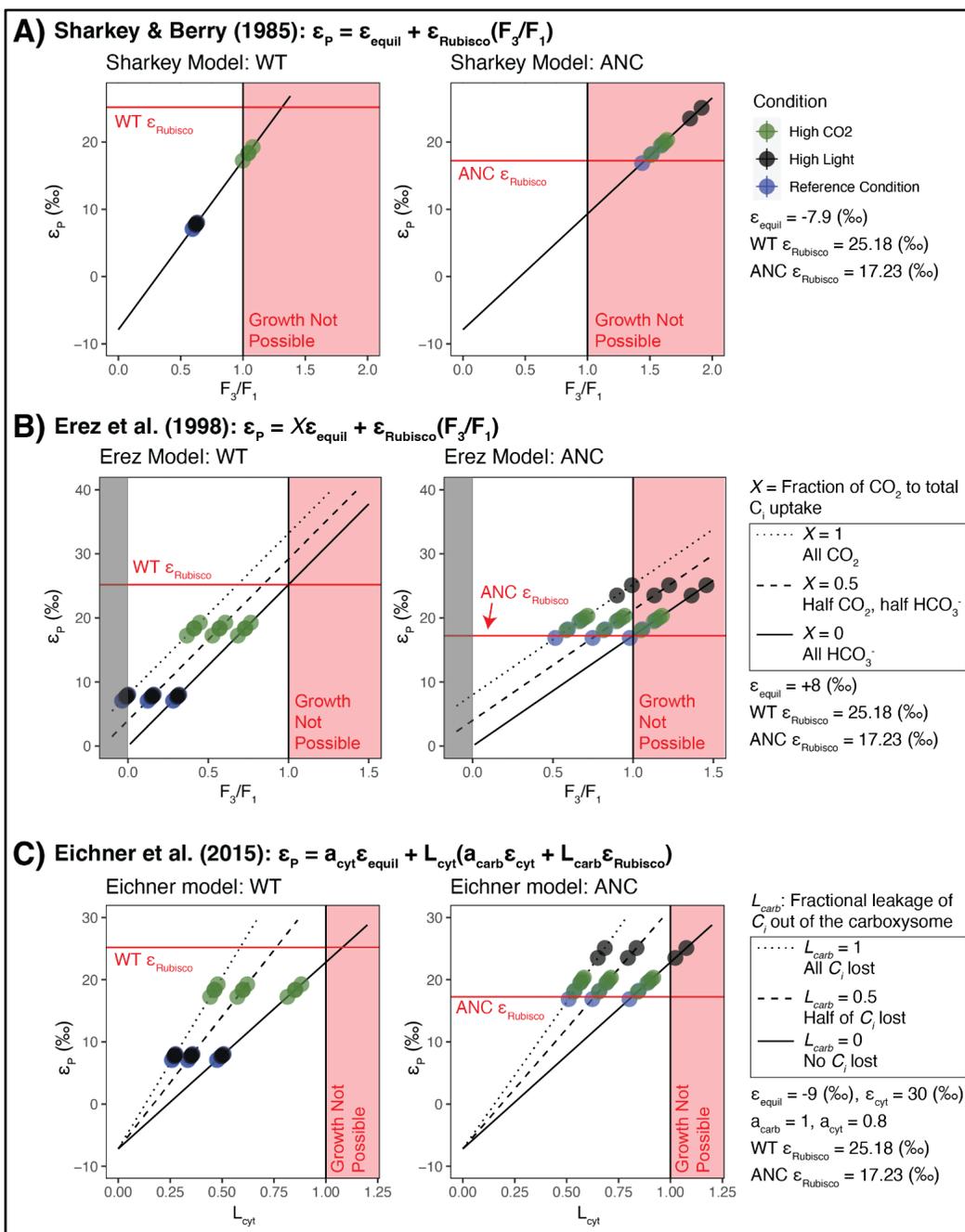ANC $\varepsilon_{Rubisco}$ = 17.23 (‰)

Figure 2.15: WT and ANC data fit with other models.
Measured ε_P values for each strain (circles) were fit with the A) (Sharkey and Berry 1985), B) (Erez et al. 1998), and C) (Eichner et al. 2015). For all models, the $\varepsilon_{Rubisco}$ values used for WT and ANC were 25.18‰ and 17.23‰ respectively, and the solid red line indicates where $\varepsilon_P = \varepsilon_{Rubisco}$. For all models, Blue circles indicate reference condition (ambient pCO_2 (0.05% (v/v)), standard light flux (120 µE)); Green circles indicate high CO_2 condition (5% pCO_2 (v/v), 120 µE); Black circles indicate high light condition (0.05% pCO_2 (v/v), 500 µE). For all models, the red-shaded zone indicates leakage values >1. A) We used the same ε_{equil} value (-7.9‰) used in (Sharkey and Berry 1985). F_3/F_1 indicates leakage of C_i from cell. B) We used the same ε_{equil} value (+8‰) used in (Erez et al. 1998). Negative leakage values are shaded in gray. Dotted lines shows solution with C_i uptake as 100% CO_2; solid line shows C_i uptake as 100% HCO_3^-; dashed line shows C_i uptake as 50% CO_2, 50% HCO_3^-. C) We use the same ε_{equil} value (-9‰) used in (Eichner et al. 2015). The values chosen for ε_{cyt}, a_{carb}, and a_{cyt} are from Scenario 5 in Table 2 in (Eichner et al. 2015); see text for discussion. Dotted lines shows solution where all C_i taken into the carboxysome leaks out; solid line shows solution where all C_i taken into the carboxysome is fixed by rubisco; dashed line shows where half of C_i taken into the carboxysome is fixed. L_{cyt}, on the x-axis, is leakage of C_i from the cell. All analyses were performed using R Statistical Software (v4.1.0; R Core Team 2021) and figures were produced using the ggplot2 package (v3.3.6; Wickham, 2016).

(Eichner et al. 2015) grew the cyanobacterium *Trichodesmium erythraeum* IMS101 with varied nitrogen sources at varied pCO_2 concentrations and compared leakage estimates derived from ε_p with an independent, non-isotopic method of membrane inlet mass spectrometry (MIMS). We note that they used a diazotrophic cyanobacterium while we did not. Similar to our study, they found that isotopic leakage estimates derived using the (Sharkey and Berry 1985) model regularly exceeded 1, while MIMS estimates gave more reasonable values (see their Figure 3). Similar to (Erez et al. 1998), they needed an additional isotopic fractionation factor, so they modified the Sharkey and Berry model by adding a compartment for the carboxysome and called upon the NDH complex specifically, which results in Equation 14 and 15 of their paper, re-written as:

$$\epsilon_P = a_{cyt}\epsilon_{equil} + L_{cyt}(a_{carb}\epsilon_{cyt} + L_{carb}\epsilon_{Rubisco})$$

Equation 2.33

Where the fractionational contribution of HCO_3^- to total C_i uptake into the cytosol or carboxysome is a_{cyt} or a_{carb} respectively (a=1 is all HCO_3^-); the relative proportion of C_i leaking out of versus entering the cytosol or carboxysome is L_{cyt} or L_{carb} respectively; ε_{cyt} is the isotopic fractionation of the NDH-1_4 complex.

Because of the independent MIMS method used in (Eichner et al. 2015), they were able to independently constrain parameters that we could not (i.e. a_{cyt}). Therefore, we use the values they found most likely to explain their results, which is Scenario 5 in their Table 2 (a_{carb}=1, a_{cyt}=0.8, ε_{cyt}=30) and varied L_{carb} from 0 to 1. They note that although an ε_{cyt} value less than 30‰ could explain their data if the other parameters were varied, "In a scenario assuming an upper estimate for ε_{cyt} of +30‰ (scenario 5, Table 2), which is within the range of fractionation measured in other enzymes such as RubisCO, our MIMS-measured data can be reproduced even for the high pCO$_2$ treatment."

Using the Eichner model, we are able to rationalize all of our WT and ANC data, though only for $L_{carb} > \approx 0.2$ for ANC (Figure S9C). This is consistent with their results, which suggests an L_{carb} value of 0.9. However, we note that they invoke the NDH complex for internal C_i recycling, to convert CO_2 lost from the carboxysome back to HCO_3^- for re-entry in the carboxysome. We invoke the NDH complex for light-powered CO_2 uptake. Regardless, both the Eichner model and ours are able to rationalize ε_p data by calling an additional fractionation factor that allows $\varepsilon_{Rubisco}$ to exceed ε_p (i.e. derived leakage values are less than 1).

For all models, we solved analytically for values of ε_P, given the experimentally measured values of $\varepsilon_{Rubisco}$, and inputting values of f ranging from 0 to 1. We then plotted our experimental ε_P values onto the model output, which gave us a value of f. After doing so, we noticed—perhaps unsurprisingly—that ANC ε_P values could not be plotted onto the model outputs, as described in the main text and shown in Figure 2.4A. The code for plotting and solving can be found on GitHub at https://github.com/reneezwang/ancestral-rubisco-cyano.

2.9.5 Emplacement of rubisco into the carboxysome

Figure 2.16. Chlorosis of ANC strain in high light.
Photo showing WT strain (left) and ANC strain (right) at the end of Condition 3 growth conditions. Note yellow-green color indicative of chlorosis.
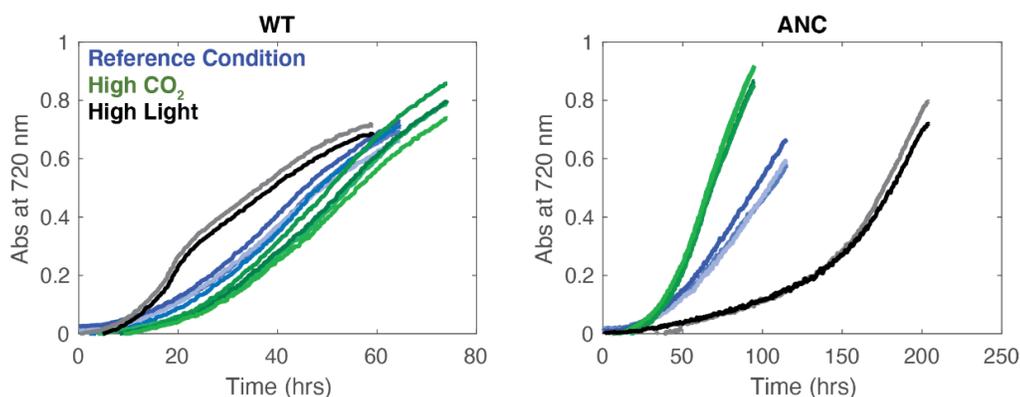
Figure 2.17. Full growth curves for WT and ANC strains.
Analyses were performed using MATLAB and Statistics Toolbox (vR2020b).

ANC strain growth at ambient pCO_2 supports the conclusion that the CCM is functioning properly, as it is well established that CCM deletions / mutations prevent cyanobacterial growth at ambient CO_2 (see (Rae et al. 2013; Kerfeld et al. 2018) for review). In addition, the carboxylation rate (V_C) for the ancestral rubisco is roughly half that of the extant rubisco (4.72 ± 0.14 s^{-1} vs. 9.78 ± 0.48 s^{-1} respectively), so the CCM has to be working for it to be able to grow at ambient. Consistent with these past results, a recent paper utilizing an ancestral analogue strain, (Hurley et al. 2021), deleted the CCM and found that their strain does not grow at CO_2 levels of 1, 18, and 30x PAL (present atmospheric levels) but was able to grow at 36 and 107x PAL at pH 7.3-8.1. Therefore, ANC strain growth at ambient pCO_2 supports the conclusion that the CCM is functional.

In addition, (Shih et al. 2016) shows rubisco emplacement using fluorescence microscopy with tags for RbcL (rubisco large subunit) and CcmN (carboxysomal subunit) in their Figure 8. Though that strain expresses both the extant and ancestral rubisco RbcS and RbcL sequence, there is no rubisco fluorescence seen external of the carboxysome.

For even further due diligence, we wanted to ensure that the ancestral rubisco emplaces properly into the carboxysome, and that doing the full swap of the extant for the ancestral rubisco sequence does not have any unintended physiologic effects on other aspects of the CCM. We performed two additional analyses: 1) Transmission electron microscopy (TEM) imaging of carboxysomes, and 2) Searching for residues shown to be required for successful rubisco emplacement into the carboxysome.

2.9.5.1 Additional TEM images

Additional TEM images are shown in Figure 2.17. Briefly, WT and ANC cells were grown in the reference condition (ambient pCO_2, normal light flux) and harvested at mid-log. Cells were sectioned and prepared for TEM imaging with the help of University of California Berkeley Electron Microscopy Lab. See Methods for full sample preparation and sectioning details.

73
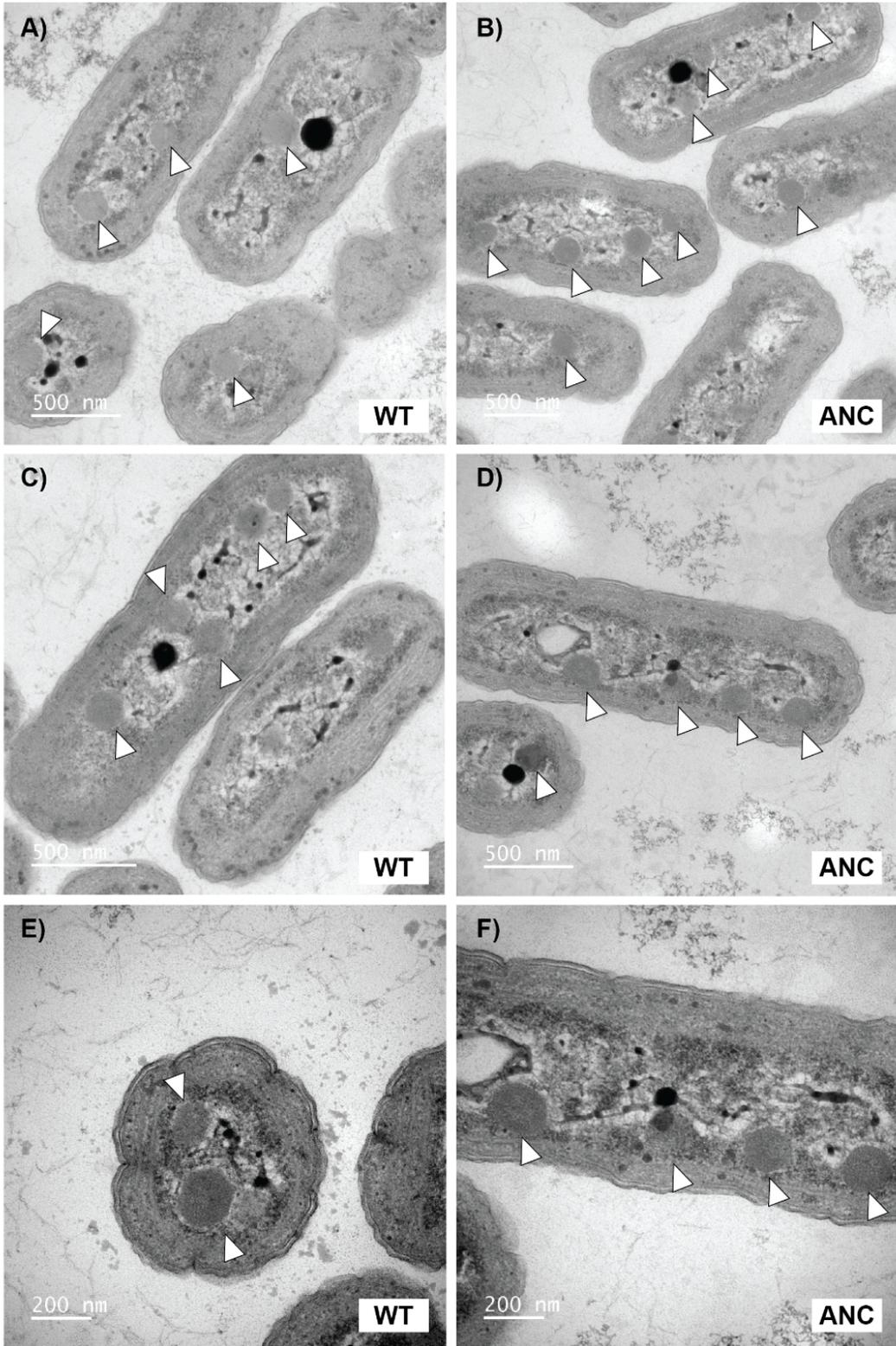
Figure 2.18: Additional TEM Images of WT and ANC strains showing carboxysomes and similar cell shape and size.
Transmission electron microscopy (TEM) images show WT (A,C,E) and ANC (B,D,F) strains that were harvested mid-log phase while growing at ambient pCO_2 and normal light conditions (see Methods). Both strains show multiple carboxysomes per cell, as indicated by white arrows, and carboxysomes exhibit classic hexagon shape (Price and Badger 1989). The dark internal bodies are likely polyphosphate bodies (Jensen 1968). WT Image C) is main text Figure 2.4A; ANC Image D is main text Figure 2.4B,C.

2.9.5.2 Reconstructed ancestral rubisco residue analysis

In Cyanobacteria, rubisco and carbonic anhydrase (CA) proteins are packed tightly within the carboxysome as liquid condensates (Rae et al. 2013). Successful formation of β-carboxysomes involves aggregation of rubisco by the scaffolding protein CcmM. It has recently been shown in *Synechococcus elongatus* PCC 7942, the same model organism used in this study, that cysteine residues in the small subunit-like (SSUL) module of CcmM is key for this process, and that disulfide bond formation in the SSUL is required for carboxysome formation *in vivo* (Wang et al. 2019).

In addition, Wang et al. show that SSUL interacts with rubisco at two interfaces, Interface I and Interface II. The structural features of these two interfaces are shown in Figure 4c and 4d of their manuscript with the contact residues specified (Wang et al. 2019). We performed an alignment of the WT and reconstructed ancestral rubisco sequence using Clustal Omega (Goujon et al. 2010; Sievers et al. 2011) and looked for these residues. We found that eight of the ten residues were conserved for Interface I, and all residues were conserved in the ancestral sequence for Interface II (Tables 2.8 and 2.9, and Figure 2.18). This, in addition to the TEM imaging and growth of ANC at ambient pCO_2, gives us confidence that substituting the extant rubisco sequence with the reconstructed ancestral sequence does not affect carboxysome function, and that the ancestral rubisco emplaces within the carboxysome.

Interface I Amino Acids			
Rubisco subunit	Wang et al. (2019) reported residue number	Residue number with offset	Present in reconstructed ancestral rubisco?
RbcL	Asp76 / D76	Asp73 / D73	Yes
RbcL	Arg79 / R79	Arg76 / R76	Yes
RbcL	Glu351 / E351	Glu348 / E348	Yes
RbcL	His353 / H353	His350 / H350	No
RbcL	Glu355 / E355	Glu352 / E352	Yes
RbcS	Gln36 / Q36	*N/A*	Yes
RbcS	Gly37 / G37	*N/A*	Yes
RbcS	Asp93 / D93	*N/A*	Yes
RbcS	Asn94 / N94	*N/A*	Yes
RbcS	Ile95 / I95	*N/A*	No

Table 2.9: Contact residues between RbcL, RbcS, and SSUL at Interface I in *Synechococcus elongatus* PCC 7942.
Interface I involves both the large (RbcL) and small (RbcS) subunits of rubisco. Numbered amino acids are taken from Figure 4c of (Wang et al. 2019) There is an offset of -3 between the numbering of Wang et al. and our WT sequence for RbcL. There is no offset for RbcS. We first converted the reported residue number to the offset number before looking for the residue in our sequence.

Interface II Amino Acids			
Rubisco subunit	Wang et al. (2019) reported residue number	Residue number with offset	Present in reconstructed ancestral rubisco?
RbcL	Tyr29 / Y29	Tyr26 / Y26	Yes
RbcL	Thr30 / T30	Thr27 / T27	Yes
RbcL	Pro31 / P31	Pro28 / P28	Yes
RbcL	Lys32 / K32	Lys29 / K29	Yes
RbcL	Tyr85 / Y85	Tyr82 / Y82	Yes
RbcL	His86 / H86	His83 / H83	Yes

Table 2.10: Contact residues between RbcL and SSUL at Interface II in *Synechococcus elongatus* PCC 7942.
Interface II only involves the large (RbcL) subunit of rubisco. Numbered amino acids are taken from Figure 4c of (Wang et al. 2019). There is an offset of -3 between the numbering of Wang et al. and our WT sequence for RbcL. We first converted the reported residue number to the offset number before looking for the residue in our sequence.

```
┌────────────────────────────────────────────────────────────────────────────────┐
│ Rubisco Large Subunit                                                            │
│                                                                                  │
│  WTRbcL    MPKTQSAAGYKAGVKDYKLTYYTPDYTPKDTDLLAAFRFSPQPGVPADEAGAAIAAESST    60    │
│  ANCRbcL   MTKTQSAAGYKAGVKDYRLTYYTPDYTPKDTDLLAAFRVTPQPGVPPEEAGAAVAAESST    60    │
│            * *************:********:**********.:****** :*****:******             │
│                                                                                  │
│  WTRbcL    GTWTTVWTDLLTDMDRYKGKCYHIEPVQGEENSYFAFIAYPLDLFEEGSVTNILTSIVGN   120    │
│  ANCRbcL   GTWTTVWTDLLTDMDRYKGRCYHIEPVPGEDNSYFAFIAYPLDLFEEGSVTNILTSIVGN   120    │
│            ************:***:***:* **:*******************************             │
│                                                                                  │
│  WTRbcL    VFGFKAIRSLRLEDIRFPVALVKTFQGPPHGIQVERDLLNKYGRPMLGCTIKPKLGLSAK   180    │
│  ANCRbcL   VFGFKALRALRLEDIRFPVAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAK   180    │
│            ******:*:*********** ************* *******:*************             │
│                                                                                  │
│  WTRbcL    NYGRAVYECLRGGLDFTKDDENINSQPFQRWRDRFLFVADAIHKSQAETGEIKGHYLNVT   240    │
│  ANCRbcL   NYGRAVYECLRGGLDFTKDDENINSQPFQRWRDRFLFVADAIHKAQAETGEIKGHYLNVT   240    │
│            *******************************************:***************             │
│                                                                                  │
│  WTRbcL    APTCEEMMKRAEFAKELGMPIIMHDFLTAGFTANTTLAKWCRDNGVLLHIHRAMHAVIDR   300    │
│  ANCRbcL   APTCEEMMKRAEFAKELGMPIIMHDFLTAGFTANTTLAKWCRDNGVLLHIHRAMHAVIDR   300    │
│            ************************************************************             │
│                                                                                  │
│  WTRbcL    QRNHGIHFRVLAKCLRLSGGDHLHSGTVVGKLEGDKASTLGFVDLMREDHIEADRSRGVF   360    │
│  ANCRbcL   QKNHGIHFRVLAKCLRLSGGDHLHTGTVVGKLEGDRASTLGFVDLLREDYIEADRSRGIF   360    │
│            *:**********************:***********:*********.***.*:********:*             │
│                                                                                  │
│  WTRbcL    FTQDWASMPGVLPVASGGIHVWHMPALVEIFGDDSVLQFGGGTLGHPWGNAPGATANRVA   420    │
│  ANCRbcL   FTQDWASMPGVMAVASGGIHVWHMPALVEIFGDDSVLQFGGGTLGHPWGNAPGATANRVA   420    │
│            ***********: ************************************************             │
│                                                                                  │
│  WTRbcL    LEACVQARNEGRDLYREGGDILREAGKWSPELAAALDLWKEIKFEFETMDKL*    472         │
│  ANCRbcL   LEACVQARNEGRDLMREGGDILREAAKWSPELAAALELWKEIKFEFETVDKL*    472         │
│            ************** **********.**********:************:****               │
│                                                                                  │
│ Rubisco Small Subunit                                                            │
│                                                                                  │
│  WTRbcS    MSMKTLPKERRFETFSYLPPLSDRQIAAQIEYMIEQGFHPLIEFNEHSNPEEFYWTMWKL    60    │
│  ANCRbcS   MQVWTPAKNKKYETFSYLPPLSDEQIAKQIQYILSQGWVPCVEFNEDSHPENRYWTMWKL    60    │
│            *.: *   *:::::***********.*** **:*::.**: * :****.*:**: *******        │
│                                                                                  │
│  WTRbcS    PLFDCKSPQQVLDEVRECRSEYGDCYIRVAGFDNIKQCQTVSFIVHRPGRY*    111          │
│  ANCRbcS   PLFGAQDAAQVLSEVQACRKAFPNCYIRVVGFDNVKQCQCMSFIVHRPA*--    109          │
│            ***..:.  ***.**: **. : :*****.****:**** :*******.                     │
│                                                                                  │
│              ┌─────────────────────────────────┐                                │
│              │ Interface I                      │                                │
│              │   █ Conserved in ANC             │                                │
│              │   █ Not conserved in ANC         │                                │
│              │ Interface II                     │                                │
│              │   █ Conserved in ANC             │                                │
│              └─────────────────────────────────┘                                │
└────────────────────────────────────────────────────────────────────────────────┘
```

Figure 2.19: Sequence alignment of RbcL and RbcS for the WT, extant rubisco sequence and the reconstructed ancestral rubisco sequence.
Alignment was performed using Clustal Omega (Goujon et al. 2010; Sievers et al. 2011) for the large and small subunits of the extant, WT rubisco sequence (WTRbcL and WTRbcS respectively), and for the large and small subunits of the reconstructed ancestral rubisco (ANCRbcL and ANCRbcS respectively). Residues that are conserved in ANC in Interface I are shown in green, and residues that are not conserved are shown in red. All ANC residues were conserved for Interface II and are shown in blue. Residue numbering is shown with -3 offset from (Wang et al. 2019) and black diamonds are placed every ten residues to help with counting. Asterisk (*) indicates fully conserved residue between WT and ANC sequences; colon (:) indicates conservation between amino acids of strongly similar properties as; period (.) indicates conservation between groups of weakly similar properties as below; blank space indicates no conservation. A dash (-) indicates absence of amino acid. For more information on alignment calculations, see (Goujon et al. 2010; Sievers et al. 2011) for more information.

2.9.5.3 Spectroscopy

In order to compare the pigment composition displayed by wild type versus ANC mutant, we performed room temperature absorbance spectra measurement between 400-800 nm for cultures with similar density (OD730=0.4). WT and ANC strains were grown in the reference condition – 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. Samples with OD730 = 0.4 (NanoDrop OneC Microvolume UV-Vis, Thermo Scientific) were obtained and absorbance spectra were measured with a UV–Vis Scanning Spectrophotometer (UV-2101PC, Shimadzu, Japan) in the range of 400-800 nm. Data was normalized to emission at 800 nm. Results can be seen in Figure 2.19. Absorbance measurements confirmed the chlorosis phenotype observed for the ANC strain. The WT and ANC strains were normalized to the same optical density at 800nm, however, the ANC strain demonstrated lower relative absorbance values at 620nm where phycocyanin, the major pigment of phycobilisomes is known to absorb.
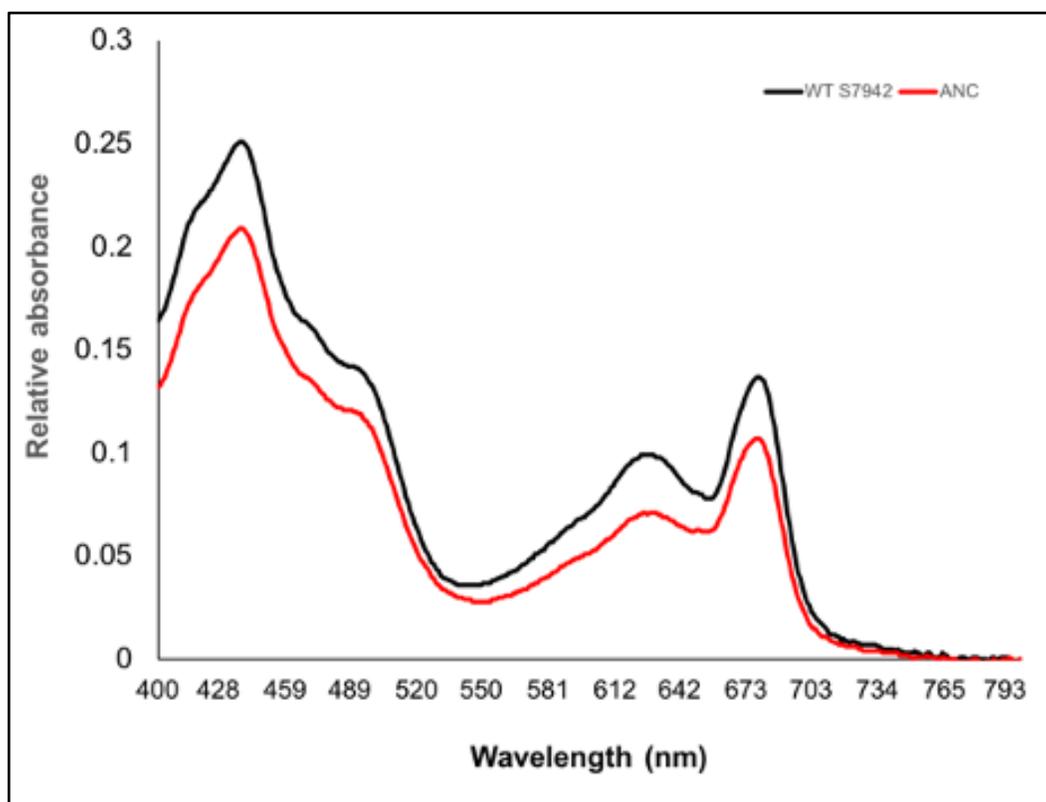
Figure 2.20: Absorption spectra of WT and ANC.
Absorption spectra of *Synechococcus elongatus* PCC 7942 wild type (black line) versus ANC mutant (red line). Absorption curves are representative of two replicates and data was normalized to values at 800 nm. Absorbance at 620nm is lower for the ANC strain indicating lower levels of phycocyanin, the major pigment of the phycobilisome, per cell compared to the wild type strain.

2.9.6 C isotope record model

We get nonsensical results when applying our results to the C Isotope Record model (main text Equation 2.1; Figure 2.11) because both that model and the organismal models it is based on (main text Equation 2.2; Figure 2.1B,C) are based on the fundamental limit that ε_P cannot exceed $\varepsilon_{Rubisco}$.

First, one must calculate *b*, the parameter that sets how quickly ε_P approaches $\varepsilon_{Rubisco}$ as the concentration of CO_2(aq) changes. *b* can be calculated in two similar ways: 1) By solving for *b* directly by re-arrangement of the relationship $\varepsilon_P = \varepsilon_f - b/C_e$, so $b = (\varepsilon_f - \varepsilon_P)*C_e$ as shown in Table 3 of (Bidigare et al. 1997); 2) By plotting all ε_P vs. μ/C_e of a given strain across various conditions (traditionally, varied pCO2) and calculating the slope through linear regression as shown in Figure 1 of (Bidigare et al. 1997). Calculated *b* values using the first method are shown in the Table S10.

Strain	Rep	Condition	% CO_2	$CO_{2(aq)}$ (µmol kg^{-1})	ε_P (‰)	b (‰ µmol kg^{-1})
WT	1	Reference Condition	0.04	10	7.1	181.3
WT	2	Reference Condition	0.04	10	7.5	176.4
WT	3	Reference Condition	0.04	10	7.2	180.2
WT	4	Reference Condition	0.04	10	8.1	171.3
ANC	1	Reference Condition	0.04	10	19.6	-24.2
ANC	2	Reference Condition	0.04	10	18.2	-9.9
ANC	3	Reference Condition	0.04	10	16.9	3.6
WT	1	High CO_2	5	1255	18.4	8500.4
WT	2	High CO_2	5	1255	18.3	8615.0
WT	3	High CO_2	5	1255	19.3	7432.2
WT	4	High CO_2	5	1255	17.2	9968.0
ANC	1	High CO_2	5	1255	18.1	-1040.3
ANC	2	High CO_2	5	1255	20.0	-3510.7
ANC	3	High CO_2	5	1255	20.3	-3889.4
ANC	4	High CO_2	5	1255	19.5	-2787.3
WT	1	High Light	0.04	10	7.9	172.6
WT	2	High Light	0.04	10	7.7	174.7
ANC	1	High Light	0.04	10	25.1	-78.7
ANC	2	High Light	0.04	10	23.5	-62.7

Table 2.11: Calculated b values for this study.
We used the R package *seacarb* to calculate concentrations of CO_2(aq) (Gattuso et al. 2015), similar to (Hurley et al. 2021) who write in their supplemental, "the headspace pCO₂ were used to calculate dissolved CO_2 via the csys program adapted for the R statistical computing environment." For ambient conditions, 0.04% CO_2, (Hurley et al. 2021) get 7 µmol/kg and for 3% CO_2 they get 538 µmol/kg. In this study, for ambient conditions, 0.04% CO_2, we get 10 µmol/kg and for 5% CO_2 we get 1255 µmol/kg. b is calculated as $b = (\varepsilon_f - \varepsilon_P)*C_e$; we used a value of $\varepsilon_f = 25.18$ for WT and $\varepsilon_f = 17.23$ for ANC per our *in vitro* KIE measurements.

Typical values of b are roughly on the order of 100 (i.e. (Bidigare et al. 1997)), but those are based on measurements of algae taken around ambient CO_2 air concentrations. We can see that we get values within that range for WT in the reference condition and high light condition (i.e. when CO_2 is at ambient concentrations, 0.04% CO_2), but not at our high CO_2 levels (5%) which are CO_2 concentrations that were not originally tested when this model was proposed (Table 2.10). In addition, the most aberrant values are for ANC across all conditions where negative values are achieved. This is because ANC ε_P values exceed $\varepsilon_{Rubisco}$ in most conditions tested, violating a central tenant that the C isotope record model was based on – that ε_P cannot exceed $\varepsilon_{Rubisco}$. In other words, the equation $b = (\varepsilon_f - \varepsilon_P)*C_e$ assumes that $\varepsilon_P < \varepsilon_f$ so b is always a positive number. This can be more clearly seen if we calculate b through the second method – plotting ε_P vs. μ/C_e and calculating the slope – shown in Figure 2.20.

In Figure 1 of (Bidigare et al. 1997), they did not know the specific $\varepsilon_{Rubisco}$ of their strains, but instead note that the intercept (24.6) is "representative of the maximum isotopic fractionation (ε_f) and is similar to the 'consensus value' of 25‰ that emerges from a variety of recent investigations [Hayes, 1993; Laws et al., 1995]." Therefore, they could fit for b based on this model. Figure S16A, however, clearly shows that ANC ε_P values exceed $\varepsilon_{Rubisco}$ in most conditions tested. So, though we can mathematically fit a value for b for ANC (14.9 ± 26.4), the value is nonsensical. This is clearly illustrated in Figure 2.20B, which shows that the only way we can fit our ANC data to the C Isotope Record model ($\varepsilon_P = \varepsilon_f - b/[CO_{2(aq)}]$; Equation 1 in main text) is if we use a value of $b = 14.9 \pm 26.4$. Because most of our ANC data lies above the theoretical limit of $\varepsilon_P = \varepsilon_{Rubisco}$, only a negative b value that lies within the fitted uncertainty of b (or, $14.9 - 26.4 = -11.5$) can create a model that accommodates our data (i.e. that allows the curve to lie above the $\varepsilon_P = \varepsilon_{Rubisco}$ limit).

Taken another way – our ANC results imply that the current model for interpreting the C isotope record ($\varepsilon_P = \varepsilon_f - b/[CO_{2(aq)}]$; Equation 2.1 in main text) may not be the right tool for the job. Substantively proposing a new model lies outside the scope of this study since the primary 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. In addition, we only grew ANC and WT at a few conditions so we only have a few data points to fit a curve to, leaving the shape of this curve pretty unconstrained. Therefore, we do not feel confident offering a new model for interpreting the C isotope record but can instead only say that the ANC strain violates a key assumption of this model, that the maximum ε_P value cannot exceed $\varepsilon_{Rubisco}$.

A) Fitting for *b* in C Isotope Record Model
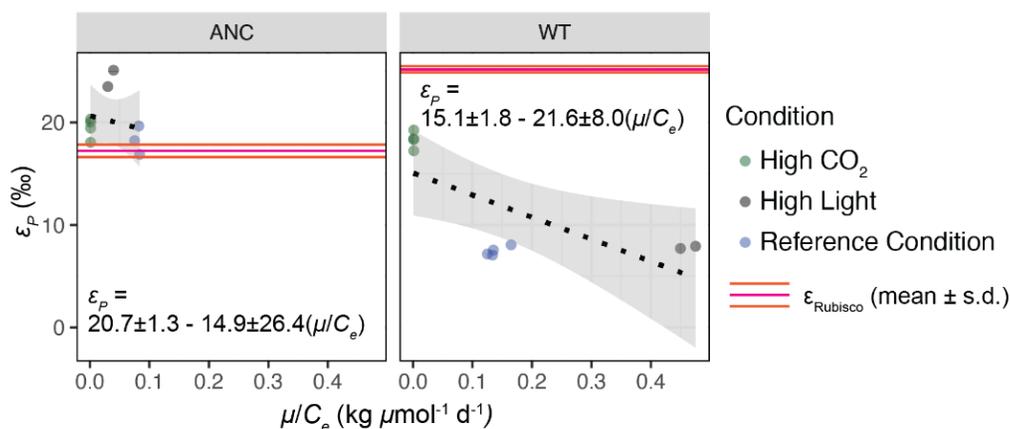
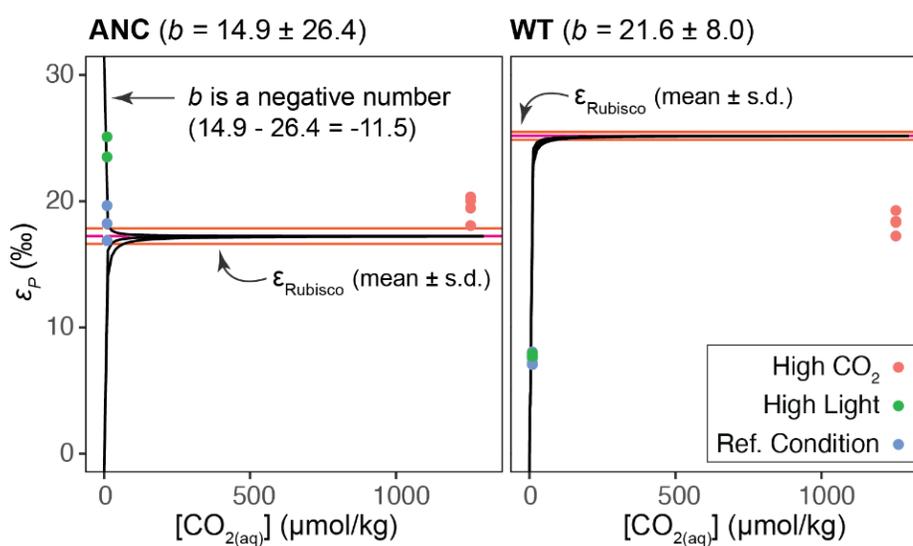## B) C Isotope Record Model for WT and ANC

Figure 2.20: C Isotope Record Model fails for ANC.
A) Fitting ANC and WT across all conditions tested for *b*, similar to Figure 1 in (Bidigare et al. 1997). Linear regression was fitted using the non-linear least squares function (call: *nls*(); R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021)). Fitted *b* values (mean ± s.d.) are 14.9 ± 26.4 for ANC and 21.6 ± 8.0 for WT. $\varepsilon_{Rubisco}$ (mean ± s.d.) shown as horizontal pink and orange lines. B) C Isotope Record Model ($\varepsilon_P = \varepsilon_f - b/[CO_{2(aq)}]$; Equation 1 in main text) using fitted *b* values from panel A. The three different model fits show the mean, mean - s.d., and mean + s.d. calculated from *b*. The only way we can fit ANC data is if we use a negative *b* value within the uncertainty of the fitted value (i.e. *b* = 14.9 - 26.4 = -11.5). $\varepsilon_{Rubisco}$ (mean ± s.d.) shown as horizontal pink and orange lines. Data visualization in both panels was performed using the ggplot2 package (v3.3.6; (Wickham et al. 2016)).

Condition	ε_P (‰)	f (traditional model)	f_2 (proposed model; assume f_1=0.1)
Reference Condition	7.453 ± 0.124	0.267 ± 0.005	0.205 ± 0.004
High CO_2	18.304 ± 0.720	0.715 ± 0.030	0.567 ± 0.024
High Light	7.812 ± 0.172	0.282 ± 0.007	0.217 ± 0.006
Reference Condition	18.247 ± 0.170	1.065 ± 0.011	0.589 ± 0.006
High CO_2	19.467 ± 2.897	1.140 ± 0.179	0.630 ± 0.097
High Light	24.299 ± 0.117	1.438 ± 0.007	0.791 ± 0.004

Table 2.12: Model outputs plotted in Figure 2.5.
Measured ε_P values were used to calculate f values using the traditional box model (main text Equation 2). Uncertainty is smaller than the markers used in main text Figure 5A; they are reported here instead. Similarly, f_2 values were calculated assuming f_1=0.1 using the proposed box model (Figure 5B; Supplemental Equation S29). Uncertainty is also smaller than the markers used in main text Figure 5C so uncertainty is reported here instead.

2.10 Appendix A: Clarification of CO_2 vs. DIC pools

We want to clarify what the carbonate and organic carbon pools in the rock record are recording, and how it relates to our study. In the interest of presenting a simplified narrative, $CO_{2(aq)}$ and DIC (HCO_3^- and H_2CO_3) were referred to as being similar in the text. However, we recognize that there is a large (roughly 10‰) isotopic difference between these difference carbon species, and that this difference is temperature dependent. For review, ε_P is the difference in $\delta^{13}C$ between $CO_{2(aq)}$ and bulk biomass. We did not account for the small fractionation (roughly 1‰) between $CO_{2(g)}$ and $CO_{2(aq)}$, where $CO_{2(aq)}$ is lighter than $CO_{2(g)}$. If we had done so, our results would have been more magnified – i.e. ε_P exceeds $\varepsilon_{Rubisco}$ to an even greater degree. Therefore, the primary conclusions of our study still hold.

Turning to the rock record, $CO_{2(aq)}$ is significantly lighter than DIC (8.2‰ at 30°C (Hayes et al. 1999)). This heavier DIC pool is what is recorded in the rock record as carbonate rocks. In addition, there is an additional small fractionation as DIC is recorded in carbonate rocks. Calcium carbonates are roughly 1.2‰ lighter than the DIC they form from (Hayes et al. 1999).

In addition, there is a slight difference in $\delta^{13}C$ between primary biomass and sedimentary TOC recorded in the rock record, due to secondary biological processes. This difference is expected to be small and positive when the secondary processes are dominated by respiratory remineralization, but can be negative when non-photosynthetic organisms are involved (see (Hayes et al. 1999) for review).

Finally, ε_{TOC} summarizes the difference in $\delta^{13}C$ between carbonates and sedimentary TOC. Therefore, ε_{TOC} reflects at least four processes: i) C-fixation by rubisco *in vivo* to create ε_P; ii) The isotopic fractionation between $CO_{2(aq)}$ and DIC,

which is temperature and pH dependent; iii) The fractionation between DIC and carbonate minerals; iv) Secondary biological processes that cause isotopic fractionation between primary biomass and TOC that is eventually recorded in the sedimentary record. See Figure 5 in (Hayes et al. 1999) for review. Therefore, all four processes must be accounted for when 'reading' the rock record; our study only addresses the first factor.

2.11 Appendix B: Proposed model

We want to note that much work has been done to create a model where ε_P can exceed $\varepsilon_{Rubisco}$ and that the goal of this study was not to create a definitive new model; rather, the goal was to test if these models held up in a strain, ANC, that may better resemble their ancestral counterparts. We relied upon the work of (Wilkes and Pearson 2019), which developed such a model in eukaryotic algae, and had fruitful discussions with the authors about their model that were extremely helpful and instructive with regards to rationalizing our experimental results. However, we engaged more fully with the models from (Erez et al. 1998) and (Eichner et al. 2015) because their models were developed from and written for Cyanobacteria, the bacterial clade that we worked within. Eukaryotic algae indeed have a version of the NDH complex that can cause ε_P values to exceed $\varepsilon_{Rubisco}$, as noted in (Wilkes and Pearson 2019), and this model would have been able to rationalize our results as well.

2.12 Appendix C: Clarification of Rubisco and Bulk Biomass Measurements

We want to clarify the methods used in this study. For the rubisco KIE assays, the Gas Bench with the Delta-V Advantage was used to measure the $\delta^{13}C$ of $CO_{2(g)}$ within the headspace of the exetainers. We did 10 replicate injections of each exetainer. The Costech Elemental Analyzer (EA) with the Delta-V Advantage was used to measure the $\delta^{13}C$ of bulk biomass. For the bulk biomass assays, data were corrected for blank capsule contributions and linearity. In addition, for both assays, though we measured at Masses 44-46, we only peak-centered on Mass 45. For all measurements, a ^{17}O correction based on Mass 46 was performed as in (Santrock et al. 1985).

For the rubisco KIE assays, exetainers were measured as soon as possible after the rubisco assay was quenched. However, each isotopic measurement took longer than the time between sampling points. Therefore, the lag time between rubisco quenching and measurement was short for the initial time points (i.e. immediately or ~1 hour), but the lag time was longer for the later points (i.e. more than 3 hours). Therefore, it is likely that the negative spread in $\delta^{13}C$ values seen in the later timepoints reflects contamination from ambient air – i.e. the seals on the exetainers were allowing ambient $CO_{2(g)}$, which has a negative $\delta^{13}C$ value, to enter the exetainer. However, we pre-processed the data to remove these errant

values (see Supplemental) so our fitted rubisco KIE values are not affected by this.

In addition, we attempted to use the total amount of CO_2 measured in the headspace to calculate f in our rubisco KIE assays. However, the error on these values was large, likely because we only drew down the total CO_2 pool to a small amount (i.e. uncertainty was on the order of magnitude of the effect we were trying to measure). Therefore, we calculated f from the time sampling points and the change in $\delta^{13}C$ (see Supplemental).

2.13 References

Adir, Noam, Monica Dines, Merav Klartag, Ailie McGregor, and Meira Melamed-Frank. 2006. "Assembly and Disassembly of Phycobilisomes." In Complex Intracellular Structures in Prokaryotes, edited by Jessup M. Shively, 2:47–77. Microbiology Monographs. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/7171_020.

Andrews, T J, and G H Lorimer. 1987. The Biochemistry of Plants: A Comprehensive Treatise, Vol. 10, Photosynthesis. Edited by M D Hatch and N K Boardman. Vol. 10.

Artier, Juliana, Ross M Walker, Neil T Miller, Minquan Zhang, G Dean Price, and Robert L Burnap. 2022. "Modeling and Mutagenesis of Amino Acid Residues Critical for CO2 Hydration by Specialized NDH-1 Complexes in Cyanobacteria." Biochimica et Biophysica Acta. Bioenergetics 1863 (1): 148503. https://doi.org/10.1016/j.bbabio.2021.148503.

Banda, Douglas M, Jose H Pereira, Albert K Liu, Douglas J Orr, Michal Hammel, Christine He, Martin A J Parry, et al. 2020. "Novel Bacterial Clade Reveals Origin of Form I Rubisco." Nature Plants 6 (9): 1158–66. https://doi.org/10.1038/s41477-020-00762-4.

Bar-Even, Arren, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S Tawfik, and Ron Milo. 2011. "The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters." Biochemistry 50 (21): 4402–10. https://doi.org/10.1021/bi2002289.

Bar-On, Yinon M, and Ron Milo. 2019. "The Global Mass and Average Rate of Rubisco." Proceedings of the National Academy of Sciences of the United States of America 116 (10): 4738–43. https://doi.org/10.1073/pnas.1816654116.

Battchikova, Natalia, Marion Eisenhut, and Eva-Mari Aro. 2011. "Cyanobacterial NDH-1 Complexes: Novel Insights and Remaining Puzzles." Biochimica et Biophysica Acta 1807 (8): 935–44. https://doi.org/10.1016/j.bbabio.2010.10.017.

Berry, J A. 1989. "Studies of Mechanisms Affecting the Fractionation of Carbon Isotopes in Photosynthesis." In Stable Isotopes in Ecological Research, edited by P W Rundel, J R Ehleringer, and K A Nagy, 68:82–94. Ecological Studies. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4612-3498-2_6.

Bidigare, Robert R., Arnim Fluegge, Katherine H. Freeman, Kristi L. Hanson, John M. Hayes, David Hollander, John P. Jasper, et al. 1997. "Consistent Fractionation of 13 C in Nature and in the Laboratory: Growth-Rate Effects in Some Haptophyte Algae." Global Biogeochemical Cycles 11 (2): 279–92. https://doi.org/10.1029/96GB03939.

Boettger, Jason D., and James D. Kubicki. 2021. "Equilibrium and Kinetic Isotopic Fractionation in the CO2 Hydration and Hydroxylation Reactions: Analysis of the Role of Hydrogen-Bonding via Quantum Mechanical Calculations." Geochimica et Cosmochimica Acta 292 (January): 37–63. https://doi.org/10.1016/j.gca.2020.09.019.

Boller, Amanda J., Phaedra J. Thomas, Colleen M. Cavanaugh, and Kathleen M. Scott. 2011. "Low Stable Carbon Isotope Fractionation by Coccolithophore RubisCO." Geochimica et Cosmochimica Acta 75 (22): 7200–7207. https://doi.org/10.1016/j.gca.2011.08.031.

Cassar, Nicolas, Edward A. Laws, and Brian N. Popp. 2006. "Carbon Isotopic Fractionation by the Marine Diatom Phaeodactylum Tricornutum under Nutrient- and Light-Limited Growth Conditions." Geochimica et Cosmochimica Acta 70 (21): 5323–35. https://doi.org/10.1016/j.gca.2006.08.024.

Christeller, John T., William A. Laing, and John H. Troughton. 1976. "Isotope Discrimination by Ribulose 1,5-Diphosphate Carboxylase." Plant Physiology 57 (4): 580–82. https://doi.org/10.1104/pp.57.4.580.

Clark, Ian D., and Bernard Lauriol. 1992. "Kinetic Enrichment of Stable Isotopes in Cryogenic Calcites." Chemical Geology 102 (1–4): 217–28. https://doi.org/10.1016/0009-2541(92)90157-Z.

Collier, J L, and A R Grossman. 1992. "Chlorosis Induced by Nutrient Deprivation in Synechococcus Sp. Strain PCC 7942: Not All Bleaching Is the Same." Journal of Bacteriology 174 (14): 4718–26. https://doi.org/10.1128/jb.174.14.4718-4726.1992.

Craig, Harmon. 1953. "The Geochemistry of the Stable Carbon Isotopes." Geochimica et Cosmochimica Acta 3 (2–3): 53–92. https://doi.org/10.1016/0016-7037(53)90001-5.

Desmarais, John J, Avi I Flamholz, Cecilia Blikstad, Eli J Dugan, Thomas G Laughlin, Luke M Oltrogge, Allen W Chen, et al. 2019. "DABs Are Inorganic Carbon Pumps Found throughout Prokaryotic Phyla." Nature Microbiology 4 (12): 2204–15. https://doi.org/10.1038/s41564-019-0520-8.

Eichner, Meri, Silke Thoms, Sven A Kranz, and Björn Rost. 2015. "Cellular Inorganic Carbon Fluxes in Trichodesmium: A Combined Approach Using Measurements and Modelling." Journal of Experimental Botany 66 (3): 749–59. https://doi.org/10.1093/jxb/eru427.

"Enzymatic Assay of Carbonic Anhydrase for Wilbur-Anderson Units (EC 4.2.1.1)." n.d. Accessed June 8, 2022. https://www.sigmaaldrich.com/US/en/technical-documents/protocol/protein-biology/enzyme-activity-assays/enzymatic-assay-of-carbonic-anhydrase.

Erez, Jonathan, Anne Bouevitch, and Aaron Kaplan. 1998a. "Carbon Isotope Fractionation by Photosynthetic Aquatic Microorganisms: Experiments with Synechococcus PCC7942, and a Simple Carbon Flux Model." Canadian Journal of Botany 76 (6): 1109–18. https://doi.org/10.1139/cjb-76-6-1109.

———. 1998b. "Carbon Isotope Fractionation by Photosynthetic Aquatic Microorganisms: Experiments withSynechococcus PCC7942, and a Simple Carbon Flux Model." Canadian Journal of Botany 76 (6): 1109–18. https://doi.org/10.1139/b98-067.

Espie, G S, A G Miller, and D T Canvin. 1991. "High Affinity Transport of CO(2) in the Cyanobacterium Synechococcus UTEX 625." Plant Physiology 97 (3): 943–53. https://doi.org/10.1104/pp.97.3.943.

Farquhar, G D, J R Ehleringer, and K T Hubick. 1989. "Carbon Isotope Discrimination and Photosynthesis." Annual Review of Plant Physiology and Plant Molecular Biology 40 (1): 503–37. https://doi.org/10.1146/annurev.pp.40.060189.002443.

Farquhar, G D, M H O'Leary, and J A Berry. 1982. "On the Relationship between Carbon Isotope Discrimination and the Intercellular Carbon Dioxide Concentration in Leaves." Australian Journal of Plant Physiology 9 (2): 121. https://doi.org/10.1071/PP9820121.

Fischer, Woodward W., James Hemp, and Jena E. Johnson. 2016. "Evolution of Oxygenic Photosynthesis." Annual Review of Earth and Planetary Sciences 44 (1): 647–83. https://doi.org/10.1146/annurev-earth-060313-054810.

Flamholz, Avi, and Patrick M Shih. 2020. "Cell Biology of Photosynthesis over Geologic Time." Current Biology 30 (10): R490–94. https://doi.org/10.1016/j.cub.2020.01.076.

Flamholz, Avi I, Eli Dugan, Justin Panich, John J Desmarais, Luke M Oltrogge, Woodward W Fischer, Steven W Singer, and David F Savage. 2022. "Trajectories for the Evolution of Bacterial CO2-Concentrating Mechanisms." Proceedings of the National Academy of Sciences of the United States of America 119 (49): e2210539119. https://doi.org/10.1073/pnas.2210539119.

Flamholz, Avi I, Noam Prywes, Uri Moran, Dan Davidi, Yinon M Bar-On, Luke M Oltrogge, Rui Alves, David Savage, and Ron Milo. 2019. "Revisiting Trade-Offs between Rubisco Kinetic Parameters." Biochemistry 58 (31): 3365–76. https://doi.org/10.1021/acs.biochem.9b00237.

Francois, Roger, Mark A. Altabet, Ralf Goericke, Daniel C. McCorkle, Christian Brunet, and Alain Poisson. 1993. "Changes in the δ13 C of Surface Water Particulate Organic Matter across the Subtropical Convergence in the SW Indian Ocean." Global Biogeochemical Cycles 7 (3): 627–44. https://doi.org/10.1029/93GB01277.

Freeman, K H, and J M Hayes. 1992. "Fractionation of Carbon Isotopes by Phytoplankton and Estimates of Ancient CO2 Levels." Global Biogeochemical Cycles 6 (2): 185–98. https://doi.org/10.1029/92GB00190.

Garcia, Amanda K, Colleen M Cavanaugh, and Betul Kacar. 2021. "The Curious Consistency of Carbon Biosignatures over Billions of Years of Earth-Life Coevolution." The ISME Journal 15 (8): 2183–94. https://doi.org/10.1038/s41396-021-00971-5.

Garcia, Amanda K, Mateusz Kędzior, Arnaud Taton, Meng Li, Jodi N Young, and Betül Kaçar. 2023. "Effects of RuBisCO and CO2 Concentration on Cyanobacterial Growth and Carbon Isotope Fractionation." Geobiology, January. https://doi.org/10.1111/gbi.12543.

Garcia, Amanda K., Mateusz Kedzior, Arnaud Taton, Meng Li, Jodi Young, and Betul Kacar. 2021. "System-Level Effects of CO2 and RuBisCO Concentration on Carbon Isotope Fractionation." BioRxiv, April. https://doi.org/10.1101/2021.04.20.440233.

Gattuso, J P, J M Epitalon, and H Lavigne. 2015. "Seacarb: Seawater Carbonate Chemistry with R. R Package Version 3.0. 6, The Comprehensive R Archive Network."

Gimmler, H, C Weiss, M Baier, and W Hartung. 1990. "The Conductance of the Plasmalemma for Co2." Journal of Experimental Botany 41 (7): 785–95. https://doi.org/10.1093/jxb/41.7.785.

Golden, S S, and L A Sherman. 1984. "Optimal Conditions for Genetic Transformation of the Cyanobacterium Anacystis Nidulans R2." Journal of Bacteriology 158 (1): 36–42. https://doi.org/10.1128/jb.158.1.36-42.1984.

Goujon, Mickael, Hamish McWilliam, Weizhong Li, Franck Valentin, Silvano Squizzato, Juri Paern, and Rodrigo Lopez. 2010. "A New Bioinformatics Analysis Tools Framework at EMBL-EBI." Nucleic Acids Research 38 (Web Server issue): W695-9. https://doi.org/10.1093/nar/gkq313.

Grébert, Théophile, Hugo Doré, Frédéric Partensky, Gregory K Farrant, Emmanuel S Boss, Marc Picheral, Lionel Guidi, et al. 2018. "Light Color Acclimation Is a Key Process in the Global Ocean Distribution of Synechococcus Cyanobacteria." Proceedings of the National Academy of Sciences of the United States of America 115 (9): E2010–19. https://doi.org/10.1073/pnas.1717069115.

Guo, Weifu. 2009. "CARBONATE CLUMPED ISOTOPE THERMOMETRY: APPLICATION TO CARBONAECOUS CHONDRITES & EFFECTS OF KINETIC ISOTOPE FRACTIONATION." Doctoral dissertation, California Institute of Technology.

Guy, R D, M L Fogel, and J A Berry. 1993. "Photosynthetic Fractionation of the Stable Isotopes of Oxygen and Carbon." Plant Physiology 101 (1): 37–47. https://doi.org/10.1104/pp.101.1.37.

Hayes, John M., Harald Strauss, and Alan J. Kaufman. 1999. "The Abundance of 13C in Marine Organic Matter and Isotopic Fractionation in the Global Biogeochemical Cycle of Carbon during the Past 800 Ma." Chemical Geology 161 (1–3): 103–25. https://doi.org/10.1016/S0009-2541(99)00083-2.

Hayes, J M. 1993. "Factors Controlling 13C Contents of Sedimentary Organic Compounds: Principles and Evidence." Marine Geology 113 (1–2): 111–25. https://doi.org/10.1016/0025-3227(93)90153-M.

———. 2001. "Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes." Reviews in Mineralogy and Geochemistry 43 (1): 225–77. https://doi.org/10.2138/gsrmg.43.1.225.

Higgins, John A, Andrei V Kurbatov, Nicole E Spaulding, Ed Brook, Douglas S Introne, Laura M Chimiak, Yuzhen Yan, Paul A Mayewski, and Michael L Bender. 2015. "Atmospheric Composition 1 Million Years Ago from Blue Ice in the Allan Hills, Antarctica." Proceedings of the National Academy of Sciences of the United States of America 112 (22): 6887–91. https://doi.org/10.1073/pnas.1420232112.

Hurley, Sarah J, Boswell A Wing, Claire E Jasper, Nicholas C Hill, and Jeffrey C Cameron. 2021. "Carbon Isotope Evidence for the Global Physiology of Proterozoic Cyanobacteria." Science Advances 7 (2). https://doi.org/10.1126/sciadv.abc8998.

Iñiguez, Concepción, Sebastià Capó-Bauçà, Ülo Niinemets, Heather Stoll, Pere Aguiló-Nicolau, and Jeroni Galmés. 2020. "Evolutionary Trends in RuBisCO Kinetics and Their Co-Evolution with CO2 Concentrating Mechanisms." The Plant Journal 101 (4): 897–918. https://doi.org/10.1111/tpj.14643.

Jasper, J P, and J M Hayes. 1990. "A Carbon Isotope Record of CO2 Levels during the Late Quaternary." Nature 347 (6292): 462–64. https://doi.org/10.1038/347462a0.

Jensen, Thomas E. 1968. "Electron Microscopy of Polyphosphate Bodies in a Blue-Green Alga, Nostoc Pruniforme." Archiv F r Mikrobiologie 62 (2): 144–52. https://doi.org/10.1007/BF00410400.

Kacar, B, V Hanson-Smith, Z R Adam, and N Boekelheide. 2017. "Constraining the Timing of the Great Oxidation Event within the Rubisco Phylogenetic Tree." Geobiology 15 (5): 628–40. https://doi.org/10.1111/gbi.12243.

Kaplan, Aaron, and Leonora Reinhold. 1999. "Co2 Concentrating Mechanisms in Photosynthetic Microorganisms." Annual Review of Plant Physiology and Plant Molecular Biology 50 (June): 539–70. https://doi.org/10.1146/annurev.arplant.50.1.539.

Kędzior, Mateusz, Amanda K Garcia, Meng Li, Arnaud Taton, Zachary R Adam, Jodi N Young, and Betül Kaçar. 2022. "Resurrected Rubisco Suggests Uniform Carbon Isotope Signatures over Geologic Time." Cell Reports 39 (4): 110726. https://doi.org/10.1016/j.celrep.2022.110726.

Kerfeld, Cheryl A, Clement Aussignargues, Jan Zarzycki, Fei Cai, and Markus Sutter. 2018. "Bacterial Microcompartments." Nature Reviews. Microbiology 16 (5): 277–90. https://doi.org/10.1038/nrmicro.2018.10.

Klughammer, B, D Sültemeyer, M R Badger, and G D Price. 1999. "The Involvement of NAD(P)H Dehydrogenase Subunits, NdhD3 and NdhF3, in High-Affinity CO_2 Uptake in Synechococcus Sp. PCC7002 Gives Evidence for Multiple NDH-1 Complexes with Specific Roles in Cyanobacteria." Molecular Microbiology 32 (6): 1305–15. https://doi.org/10.1046/j.1365-2958.1999.01457.x.

Krissansen-Totton, J, R Buick, and D C Catling. 2015. "A Statistical Analysis of the Carbon Isotope Record from the Archean to Phanerozoic and Implications for the Rise of Oxygen." American Journal of Science 315 (4): 275–316. https://doi.org/10.2475/04.2015.01.

Laws, Edward A., Brian N. Popp, Robert R. Bidigare, Mahlon C. Kennicutt, and Stephen A. Macko. 1995. "Dependence of Phytoplankton Carbon Isotopic Composition on Growth Rate and [CO2)Aq: Theoretical Considerations and Experimental Results." Geochimica et Cosmochimica Acta 59 (6): 1131–38. https://doi.org/10.1016/0016-7037(95)00030-4.

Lorimer, G H, and T J Andrews. 1973. "Plant Photorespiration—an Inevitable Consequence of the Existence of Atmospheric Oxygen." Nature 243 (5406): 359–60. https://doi.org/10.1038/243359a0.

Maeda, Shin-ichi, Murray R Badger, and G Dean Price. 2002. "Novel Gene Products Associated with NdhD3/D4-Containing NDH-1 Complexes Are Involved in Photosynthetic CO_2 Hydration in the Cyanobacterium, Synechococcus Sp. PCC7942." Molecular Microbiology 43 (2): 425–35. https://doi.org/10.1046/j.1365-2958.2002.02753.x.

Mangan, Niall M, Avi Flamholz, Rachel D Hood, Ron Milo, and David F Savage. 2016. "PH Determines the Energetic Efficiency of the Cyanobacterial CO_2 Concentrating Mechanism." Proceedings of the National Academy of Sciences of the United States of America 113 (36): E5354-62. https://doi.org/10.1073/pnas.1525145113.

Marcus, Yehouda, Hagit Altman-Gueta, Aliza Finkler, and Michael Gurevitz. 2003. "Dual Role of Cysteine 172 in Redox Regulation of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase Activity and Degradation." Journal of Bacteriology 185 (5): 1509–17. https://doi.org/10.1128/jb.185.5.1509-1517.2003.

McNevin, Dennis B, Murray R Badger, Spencer M Whitney, Susanne von Caemmerer, Guillaume G B Tcherkez, and Graham D Farquhar. 2007. "Differences in Carbon Isotope Discrimination of Three Variants of D-Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Reflect Differences in Their Catalytic Mechanisms." The Journal of Biological Chemistry 282 (49): 36068–76. https://doi.org/10.1074/jbc.M706274200.

McNevin, D B, M R Badger, H J Kane, and G D Farquhar. 2006. "Measurement of (Carbon) Kinetic Isotope Effect by Rayleigh Fractionation Using Membrane Inlet Mass Spectrometry for CO2-Consuming Reactions." Functional Plant Biology : FPB 33 (12): 1115. https://doi.org/10.1071/FP06201.

Mook, W G, J C Bommerson, and W H Staverman. 1974. "Carbon Isotope Fractionation between Dissolved Bicarbonate and Gaseous Carbon Dioxide." Earth and Planetary Science Letters 22 (2): 169–76. https://doi.org/10.1016/0012-821X(74)90078-8.

Nier, Alfred. 1937. "A Mass-Spectrographic Study of the Isotopes of Hg, Xe, Kr, Be, I, As, and Cs." Physical Review 52 (9): 933–37. https://doi.org/10.1103/PhysRev.52.933.

Nier, Alfred O., and Earl A. Gulbransen. 1939. "Variations in the Relative Abundance of the Carbon Isotopes." Journal of the American Chemical Society 61 (3): 697–98. https://doi.org/10.1021/ja01872a047.

Nier, Alfred O. 1936. "A Mass-Spectrographic Study of the Isotopes of Argon, Potassium, Rubidium, Zinc and Cadmium." Physical Review 50 (11): 1041–45. https://doi.org/10.1103/PhysRev.50.1041.

O'Leary, Marion H. 1984. "Measurement of the Isotope Fractionation Associated with Diffusion of Carbon Dioxide in Aqueous Solution." The Journal of Physical Chemistry 88 (4): 823–25. https://doi.org/10.1021/j150648a041.

O'Malley, Maureen A, and David A Walsh. 2021. "Rethinking Microbial Infallibility in the Metagenomics Era." FEMS Microbiology Ecology 97 (8). https://doi.org/10.1093/femsec/fiab092.

Ogawa, Teruo, and Aaron Kaplan. 2003. "Inorganic Carbon Acquisition Systems in Cyanobacteria." Photosynthesis Research 77 (2–3): 105–15. https://doi.org/10.1023/A:1025865500026.

Pagani, Mark, Matthew Huber, Zhonghui Liu, Steven M Bohaty, Jorijntje Henderiks, Willem Sijp, Srinath Krishnan, and Robert M DeConto. 2011. "The Role of Carbon Dioxide during the Onset of Antarctic Glaciation." Science 334 (6060): 1261–64. https://doi.org/10.1126/science.1203909.

Park, R, and S Epstein. 1960. "Carbon Isotope Fractionation during Photosynthesis." Geochimica et Cosmochimica Acta 21 (1–2): 110–26. https://doi.org/10.1016/S0016-7037(60)80006-3.

Popp, Brian N., Edward A. Laws, Robert R. Bidigare, John E. Dore, Kristi L. Hanson, and Stuart G. Wakeham. 1998. "Effect of Phytoplankton Cell Geometry on Carbon Isotopic Fractionation." Geochimica et Cosmochimica Acta 62 (1): 69–77. https://doi.org/10.1016/S0016-7037(97)00333-5.

Popp, B N, R Takigiku, J M Hayes, J W Louda, and E W Baker. 1989. "The Post-Paleozoic Chronology and Mechanism of 13C Depletion in Primary Marine Organic Matter." American Journal of Science 289 (4): 436–54. https://doi.org/10.2475/ajs.289.4.436.

Price, G. Dean, Shin-ichi Maeda, Tatsuo Omata, and Murray R. Badger. 2002. "Modes of Active Inorganic Carbon Uptake in the Cyanobacterium, Synechococcus Sp. PCC7942." Functional Plant Biology 29 (3): 131. https://doi.org/10.1071/PP01229.

Price, G D, and M R Badger. 1989. "Isolation and Characterization of High CO(2)-Requiring-Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes That Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO(2) within the Carboxysome." Plant Physiology 91 (2): 514–25. https://doi.org/10.1104/pp.91.2.514.

Price, G Dean, Jasper J L Pengelly, Britta Forster, Jiahui Du, Spencer M Whitney, Susanne von Caemmerer, Murray R Badger, Susan M Howitt, and John R Evans. 2013. "The Cyanobacterial CCM as a Source of Genes for Improving Photosynthetic CO2 Fixation in Crop Species." Journal of Experimental Botany 64 (3): 753–68. https://doi.org/10.1093/jxb/ers257.

Rae, Benjamin D, Benedict M Long, Murray R Badger, and G Dean Price. 2013. "Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria

and Some Proteobacteria." Microbiology and Molecular Biology Reviews 77 (3): 357–79. https://doi.org/10.1128/MMBR.00061-12.

Rau, G H, U Riebesell, and D Wolf-Gladrow. 1996. "A Model of Photosynthetic 13C Fractionation by Marine Phytoplankton Based on Diffusive Molecular CO2 Uptake." Marine Ecology Progress Series 133: 275–85. https://doi.org/10.3354/meps133275.

Rau, G H, T Takahashi, and D J Des Marais. 1989. "Latitudinal Variations in Plankton Delta 13C: Implications for CO2 and Productivity in Past Oceans." Nature 341 (6242): 516–18. https://doi.org/10.1038/341516a0.

Raven, J A, and J Beardall. 2014. "CO 2 Concentrating Mechanisms and Environmental Change." Aquatic Botany 118 (August): 24–37. https://doi.org/10.1016/j.aquabot.2014.05.008.

Richardson, K, J Beardall, and J A Raven. 1983. "Adaptation of Unicellular Algae to Irradiance: An Analysis of Strategies." The New Phytologist 93 (2): 157–91. https://doi.org/10.1111/j.1469-8137.1983.tb03422.x.

Riding, R. 2006. "Cyanobacterial Calcification, Carbon Dioxide Concentrating Mechanisms, and Proterozoic?Cambrian Changes in Atmospheric Composition." Geobiology 4 (4): 299–316. https://doi.org/10.1111/j.1472-4669.2006.00087.x.

Roeske, C A, and M H O'Leary. 1985. "Carbon Isotope Effect on Carboxylation of Ribulose Bisphosphate Catalyzed by Ribulosebisphosphate Carboxylase from Rhodospirillum Rubrum." Biochemistry 24 (7): 1603–7.

Rotatore, C, R R Lew, and B Colman. 1992. "Active Uptake of CO2 during Photosynthesis in the Green Alga Eremosphaera Viridis Is Mediated by a CO2-ATPase." Planta 188 (4): 539–45. https://doi.org/10.1007/BF00197046.

Sade, Ziv, and Itay Halevy. 2017. "New Constraints on Kinetic Isotope Effects during CO 2(Aq) Hydration and Hydroxylation: Revisiting Theoretical and Experimental Data." Geochimica et Cosmochimica Acta 214 (October): 246–65. https://doi.org/10.1016/j.gca.2017.07.035.

Santrock, J, S A Studley, and J M Hayes. 1985. "Isotopic Analyses Based on the Mass Spectrum of Carbon Dioxide." Analytical Chemistry 57 (7): 1444–48. https://doi.org/10.1021/ac00284a060.

Saschenbrecker, Sandra, Andreas Bracher, Karnam Vasudeva Rao, Bharathi Vasudeva Rao, F Ulrich Hartl, and Manajit Hayer-Hartl. 2007. "Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco." Cell 129 (6): 1189–1200. https://doi.org/10.1016/j.cell.2007.04.025.

Schidlowski, Manfred. 1988. "A 3,800-Million-Year Isotopic Record of Life from Carbon in Sedimentary Rocks." Nature 333 (6171): 313–18. https://doi.org/10.1038/333313a0.

Schuller, Jan M, Patricia Saura, Jacqueline Thiemann, Sandra K Schuller, Ana P Gamiz-Hernandez, Genji Kurisu, Marc M Nowaczyk, and Ville R I Kaila. 2020. "Redox-Coupled Proton Pumping Drives Carbon Concentration in the Photosynthetic Complex I." Nature Communications 11 (1): 494. https://doi.org/10.1038/s41467-020-14347-4.

Scott, Kathleen M, Julie Schwedock, Daniel P Schrag, and Colleen M Cavanaugh. 2004. "Influence of Form IA RubisCO and Environmental Dissolved Inorganic Carbon on the Delta13C of the Clam-Chemoautotroph Symbiosis Solemya Velum." Environmental Microbiology 6 (12): 1210–19. https://doi.org/10.1111/j.1462-2920.2004.00642.x.

Sharkey, Thomas D., and Joseph A. Berry. 1985. "Carbon Isotope Fractionation of Algae as Influenced by an Inducible CO2 Concentrating Mechanism." In Inorganic Carbon Uptake by Aquatic Photosynthetic Organisms, edited by W J Lucas and J A Berry, 389–401. The American Society of Plant Physiologists.

Shih, Patrick M, Alessandro Occhialini, Jeffrey C Cameron, P John Andralojc, Martin A J Parry, and Cheryl A Kerfeld. 2016. "Biochemical Characterization of Predicted Precambrian RuBisCO." Nature Communications 7 (January): 10382. https://doi.org/10.1038/ncomms10382.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. 2011. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." Molecular Systems Biology 7 (October): 539. https://doi.org/10.1038/msb.2011.75.

Śliwińska-Wilczewska, Sylwia, Zofia Konarzewska, Kinga Wiśniewska, and Marta Konik. 2020. "Photosynthetic Pigments Changes of Three Phenotypes of Picocyanobacteria Synechococcus Sp. under Different Light and Temperature Conditions." Cells 9 (9). https://doi.org/10.3390/cells9092030.

S   ltemeyer, Dieter, Klaus Biehler, and HeinrichP. Fock. 1993. "Evidence for the Contribution of Pseudocyclic Photophosphorylation to the Energy Requirement of the Mechanism for Concentrating Inorganic Carbon in Chlamydomonas." Planta 189 (2). https://doi.org/10.1007/BF00195082.

Tcherkez, Guillaume G B, Graham D Farquhar, and T John Andrews. 2006. "Despite Slow Catalysis and Confused Substrate Specificity, All Ribulose Bisphosphate Carboxylases May Be Nearly Perfectly Optimized." Proceedings of the National Academy of Sciences of the United States of America 103 (19): 7246–51. https://doi.org/10.1073/pnas.0600605103.

Tchernov, D, Y Helman, N Keren, B Luz, I Ohad, L Reinhold, T Ogawa, and A Kaplan. 2001. "Passive Entry of CO2 and Its Energy-Dependent Intracellular Conversion to HCO3- in Cyanobacteria Are Driven by a Photosystem I-Generated DeltamuH+." The Journal of Biological Chemistry 276 (26): 23450–55. https://doi.org/10.1074/jbc.M101973200.

Thomas, Phaedra J, Amanda J Boller, Sriram Satagopan, F Robert Tabita, Colleen M Cavanaugh, and Kathleen M Scott. 2018. "Isotope Discrimination by Form IC RubisCO from Ralstonia Eutropha and Rhodobacter Sphaeroides, Metabolically Versatile Members of 'Proteobacteria' from Aquatic and Soil Habitats." Environmental Microbiology, September. https://doi.org/10.1111/1462-2920.14423.

Volokita, M, D Zenvirth, A Kaplan, and L Reinhold. 1984. "Nature of the Inorganic Carbon Species Actively Taken Up by the Cyanobacterium Anabaena Variabilis." Plant Physiology 76 (3): 599–602. https://doi.org/10.1104/pp.76.3.599.

Wang, H, X Yan, H Aigner, A Bracher, N D Nguyen, W Y Hee, B M Long, G D Price, F U Hartl, and M Hayer-Hartl. 2019. "Rubisco Condensate Formation by CcmM in β-Carboxysome Biogenesis." Nature 566 (7742): 131–35. https://doi.org/10.1038/s41586-019-0880-5.

Wang, Renée Z., Albert K. Liu, Douglas M. Banda, Woodward W. Fischer, and Patrick M. Shih. 2023. "A Bacterial Form I' Rubisco Has a Smaller Carbon Isotope Fractionation than Its Form I Counterpart." Biomolecules, April.

Whelan, Tom, W M Sackett, and C R Benedict. 1973. "Enzymatic Fractionation of Carbon Isotopes by Phosphoenolpyruvate Carboxylase from C4 Plants." Plant Physiology 51 (6): 1051–54. https://doi.org/10.1104/pp.51.6.1051.

Wickham, H, W Chang, and M H Wickham. 2016. "Package 'Ggplot2'." Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2 (1): 1–189.

Wickman, Frans E. 1952. "Variations in the Relative Abundance of the Carbon Isotopes in Plants." Geochimica et Cosmochimica Acta 2 (4): 243–54. https://doi.org/10.1016/0016-7037(52)90018-5.

Wildman, Sam G. 2002. "Along the Trail from Fraction I Protein to Rubisco (Ribulose Bisphosphate Carboxylase-Oxygenase)." Photosynthesis Research 73 (1–3): 243–50. https://doi.org/10.1023/A:1020467601966.

Wilkes, Elise B., and Ann Pearson. 2019. "A General Model for Carbon Isotopes in Red-Lineage Phytoplankton: Interplay between Unidirectional Processes and Fractionation by RubisCO." Geochimica et Cosmochimica Acta, September. https://doi.org/10.1016/j.gca.2019.08.043.

Witkowski, Caitlyn R, Johan W H Weijers, Brian Blais, Stefan Schouten, and Jaap S Sinninghe Damsté. 2018. "Molecular Fossils from Phytoplankton Reveal Secular Pco2 Trend over the Phanerozoic." Science Advances 4 (11): eaat4556. https://doi.org/10.1126/sciadv.aat4556.

Zeebe, Richard E. 2014. "Kinetic Fractionation of Carbon and Oxygen Isotopes during Hydration of Carbon Dioxide." Geochimica et Cosmochimica Acta 139 (August): 540–52. https://doi.org/10.1016/j.gca.2014.05.005.

Zeebe, R E, and D Wolf-Gladrow. 2001. CO2 in Seawater: Equilibrium, Kinetics, Isotopes. Vol. 65. Elsevier Oceanography Series. Elsevier. https://doi.org/10.1016/S0422-9894(01)X8001-X.

Zhang, Yi Ge, Jorijntje Henderiks, and Xiaoqing Liu. 2020. "Refining the Alkenone-PCO2 Method II: Towards Resolving the Physiological Parameter 'b.'" Geochimica et Cosmochimica Acta 281 (July): 118–34. https://doi.org/10.1016/j.gca.2020.05.002.

3. A bacterial Form I' rubisco has a smaller carbon isotope fractionation than its Form I counterpart

Wang, Renée Z., Albert K. Liu, Douglas M. Banda, Woodward W. Fischer, and Patrick M. Shih. 2023. "A Bacterial Form I' Rubisco Has a Smaller Carbon Isotope Fractionation than Its Form I Counterpart." Biomolecules 13 (4), p.596. https://doi.org/10.3390/biom13040596

3.1 Summary

Form I rubiscos evolved in Cyanobacteria $\geq$ 2.5 billion years ago and are enzymatically unique due to the presence of small subunits (RbcS) capping both ends of an octameric large subunit (RbcL) rubisco assembly to form a hexadecameric (L_8S_8) holoenzyme. Although RbcS was previously thought to be integral to Form I rubisco stability, the recent discovery of a closely related sister clade of octameric rubiscos (Form I'; L_8) demonstrates that the L_8 complex can assemble without small subunits (Banda et al. 2020). Rubisco also displays a kinetic isotope effect (KIE) where the 3PG product is depleted in ^{13}C relative to ^{12}C. In Cyanobacteria, only two Form I KIE measurements exist, making interpretation of bacterial carbon isotope data difficult. To aid comparison, we measured in vitro the KIEs of Form I' (*Candidatus* Promineofilum breve) and Form I (*Synechococcus elongatus* PCC 6301) rubiscos and found the KIE to be smaller in the L_8 rubisco (16.25 ± 1.36‰ vs. 22.42 ± 2.37‰, respectively). Therefore, while small subunits may not be necessary for protein stability, they may affect the KIE. Our findings may provide insight into the function of RbcS and allow more refined interpretation of environmental carbon isotope data.

3.2 Introduction

Rubisco (ribulose-1,5-bisphosphate carboxylase-oxygenase) is a keystone enzyme linking Earth's inorganic and organic carbon cycles, which makes it a prime target for bioengineering associated with food systems and carbon sequestration. It is the most abundant protein on Earth today (Bar-On and Milo 2019) because it catalyzes the essential carbon fixation step in one of the most ecologically dominant carbon-fixing metabolisms, the Calvin Benson Bassham (CBB) cycle in oxygenic photosynthesis (Fischer et al. 2016). Rubisco and oxygenic photosynthesis form the basis of our food web in terrestrial and marine systems because both eukaryotic and bacterial primary producers utilize rubisco to convert inorganic carbon (CO_2 and HCO_3^-) into biomass that is then consumed by heterotrophs up the food chain. In addition, the annual flux of CO_2 fixed by rubisco is very large, representing the single most massive flux in the global carbon cycle. Gross primary productivity (GPP), which accounts for all forms of carbon fixation but is vastly dominated by oxygenic photosynthesis, is ≈120 Gt C yr^{-1} in terrestrial

(Beer et al. 2010) and ≈100 Gt C yr^{-1} in marine environments (Bar-On and Milo 2019; Field et al. 1998), compared to ≈10 Gt C yr^{-1} emitted of anthropogenic fossil CO_2 (Friedlingstein et al. 2022). Therefore, multiple efforts exist to engineer a 'better' rubisco that fixes more CO_2 in order to increase crop yields and sequester anthropogenic CO_2, among many other motivations (see (Spreitzer and Salvucci 2002) for review).

However, these bioengineering approaches are informed to a degree by our current understanding of rubisco's evolutionary history, which itself is based on our understanding of past Earth environments. These evolutionary questions largely center on the canonical paradox that, despite being a central metabolism enzyme, rubisco is: (i) 'slow,' and (ii) 'confused' because it can fix O_2 instead of CO_2 (Lorimer and Andrews 1973), which invokes a salvage pathway that costs ATP, reducing power, and carbon (Andrews and Lorimer 1987). This paradox is usually resolved by considering the atmospheric composition when rubisco first evolved more than 2.5 billion years ago, when CO_2 was much higher (potentially up to ≈20x present atmospheric levels in the Precambrian (Sheldon 2006)) and O_2 only existed at trace levels (Fischer et al. 2016). However, in a Shakespearean tragedy, once rubisco was incorporated into the greater metabolism of oxygenic photosynthesis, it poisoned the very world it came from—successful CO_2 fixation was coupled with oxygenation that permanently changed the atmosphere to one where O_2 is dominant (≈20%) and CO_2 is trace (≈0.04%). Now saddled with a rubisco evolved from a chemical world that no longer exists, diverse land plants, algae, and Cyanobacteria have independently evolved complex CO_2 concentrating mechanisms (CCMs) that effectively hyper-concentrate CO_2 at the expense of O_2 around rubisco (Flamholz and Shih 2020)—in effect, replicating the ancient atmosphere within their own cells. Those without CCMs (e.g., C3 plants) instead accommodate the low carboxylation rate by producing this enzyme at such high concentrations that up to 65% of all soluble protein in leaf extracts is just rubisco (Ellis 1979). This narrative, contingent on our understanding of the geologic carbon cycle, suggests either that rubisco is an 'accident' of evolutionary history, or that it is truly the optimal enzyme designed by evolution for a difficult task. Therefore, a better understanding of the evolutionary history of this enzyme is useful for rubisco engineering efforts.

Rubisco is also notable because it displays a large carbon kinetic isotope effect (KIE) where it preferentially fixes $^{12}CO_2$ over $^{13}CO_2$ due to the rate of carboxylation being slightly faster for $^{12}CO_2$ (Farquhar et al. 1989). This effect is typically reported in delta ($\delta^{13}C$) and epsilon (ε) notation in units of per mille (‰), where $\delta^{13}C = [^{13}R_{sa}/^{13}R_{ref} - 1]*1000$ and ^{13}R is the ratio of $^{13}C/^{12}C$ in the sample or reference, respectively. ε is roughly the difference in $\delta^{13}C$ between the product and the reactant ($\varepsilon_{Rubisco} \approx \delta^{13}C_{3PG} - \delta^{13}C_{CO2}$). Thirteen unique rubisco KIEs ($\varepsilon_{Rubisco}$ values) have been measured across a limited range of phylogenies and species, but measurements so far indicate that rubisco fractionates at roughly 20–30‰ (for a recent review see (Garcia et al. 2021)).

This KIE is useful because it allows one to track mass flux through complex systems in both modern and ancient environments (Hayes 2001), and because it may give insight into non-isotopic enzyme kinetics (Tcherkez et al. 2006). Since all biomass is ultimately synthesized from 3PG in autotrophs utilizing the CBB cycle, rubisco's KIE is inherited by bulk biomass such that organic carbon is also relatively depleted in ^{13}C relative to inorganic carbon. Therefore, when incorporated into larger metabolic models of carbon fixation, rubisco KIEs have facilitated the estimation of water use efficiency in plants (Farquhar et al. 1982), the efficiency of carbon fixation in bacterial and eukaryotic algae (Sharkey and Berry 1985), the contribution of terrestrial plants to global GPP (Lloyd and Farquhar 1994), and the proportion of C3 vs. C4 plants in mammalian diets (Cerling et al. 1997), among many other examples. Similarly, in ancient environments, it has been used to estimate paleo atmospheric CO_2 levels (Witkowski et al. 2018; Bidigare et al. 1997), track the inorganic and organic carbon cycle through time (Schidlowski 1988), and the diet of ancient mammals (Cerling and Harris 1999). In addition, rubisco KIEs have been used to support interpretation of important non-isotopic kinetic parameters such as the inverse correlation between specificity for CO_2 over O_2 ($S_{C/O}$) and rate of carboxylation (V_C) (Tcherkez et al. 2006). Therefore, knowing the KIEs of many rubiscos is valuable because it facilitates empirical measurements of mass flux in many systems, natural and engineered, where other measurements may be difficult.

However, the landscape of rubisco evolution and its effect on KIE has not been well characterized. This is particularly true in Cyanobacteria, the organism within which rubisco and oxygenic photosynthesis is thought to have evolved. Most rubisco KIEs have been measured for Form IB rubiscos from plants, and in Cyanobacteria, only one Form IA and one Form IB rubisco KIE have been measured ((Scott et al. 2007; Guy et al. 1993), for a recent review see (Garcia et al. 2021)). This is particularly important for reconstructing paleo pCO_2 levels because direct measurements of the atmosphere from ice core records only extend back ≈1 million years (Higgins et al. 2015), so for the remainder of Earth's 4.567 billion year history we must rely on indirect measurements such as the carbon isotope record: globally assembled measurements of δ^{13}C in the inorganic or organic carbon bearing phases of sedimentary rocks (Krissansen-Totton et al. 2015). Interpretation of these records relies on geochemical models, largely based on extant modern organisms, that incorporate the rubisco KIE to explain most of the offset in δ^{13}C between inorganic and organic carbon pools (see (Wilkes and Pearson 2019) for recent review of current models). These models inform our understanding of ancient atmospheres which in turn can influence our ideas of rubisco evolution in the past and engineering strategies in the present. It is therefore critical that we better understand the evolution of rubisco's KIE through time because it underlies many assumptions we make when interpreting both the past and present.

We therefore tried to address this gap in knowledge by studying one key example, a Form I rubisco that lacks the small subunit. All forms of rubisco are assembled from the basic functional building block of dimers (L_2), where two large

subunits (RbcL) are assembled head-to-tail. This is the smallest known catalytically active form of rubisco. Form I rubiscos, the most ecologically abundant form of the enzyme, are hexadecameric holoenzymes (L_8S_8) composed of four dimers with eight small subunits (RbcS) that cap both ends of the junction between adjacent dimers. The small subunit is unique to Form I rubiscos, so it has traditionally been thought that RbcS was integral to both Form I protein stability and its evolutionary history (Spreitzer 2003). However, a novel clade of rubiscos (Form I') lacking small subunits, a sister to Form I, has recently been discovered through metagenomic analyses, and a representative octameric rubisco (L_8) was successfully purified and kinetically characterized (Banda et al. 2020). Other, novel closely-related clades of L_8 rubiscos (Forms I-α and I'') have also been recently discovered in a similar fashion (West-Roberts et al. 2021). Form I' rubiscos likely diverged before the evolution of Cyanobacteria and the small subunit (Banda et al. 2020); therefore, studying rubiscos from this clade presents a unique opportunity to study the effect of evolution on rubisco KIEs. We therefore measured in vitro the KIE of an L_8S_8 Form I rubisco from *Synechococcus elongatus* PCC 6301 in comparison to the KIE of an L_8 Form I' rubisco from *Candidatus* Promineofilum breve. We found the fractionation to be smaller in the L_8 rubisco compared to the L_8S_8 rubisco ($16.25 \pm 1.36‰$ vs. $22.42 \pm 2.37‰$, respectively). Our results imply that while the presence of a small subunit is not necessary for protein function, it may affect the KIE. Our findings may help provide insight into the function of the small subunit and allow more refined interpretation of carbon isotope data in environments, past and present, where Form I' rubiscos may be unknowingly operating.

3.3 Materials and Methods

3.3.1 Delta notation ($\delta^{13}C$)

Carbon isotope data were reported using delta notation ($\delta^{13}C$) in units of per mille (‰) where $\delta^{13}C = [^{13}R_{sa}/^{13}R_{ref} - 1]*1000$, where the subscripts 'sa' and 'ref' denote sample and reference, respectively and ^{13}R is the ratio of $^{13}C/^{12}C$. All values in this study were reported relative to the Vienna Pee Dee Belemnite (VPDB) reference.

3.3.2 Rubisco purification

The rubiscos used here were purified according to previous methodologies and had their kinetics characterized previously (Banda et al. 2020; Saschenbrecker et al. 2007). Briefly, 14xHis-bdSUMO-tagged *Candidatus* P. breve rubisco and untagged *S. elongatus* PCC 6301 rubisco were expressed in BL21 DE3 Star *E. coli* cultures. P. breve enzyme was prepared by conducting Ni-NTA affinity purification on clarified lysate, followed by subsequent purification by anion exchange chromatography and size exclusion chromatography. *Syn*6301 enzyme

was prepared by subjecting clarified lysate to ammonium sulfate precipitation at the 30–40% cut, followed by subsequent purification by anion exchange chromatography and size exclusion chromatography. The enzyme was then stored on dry ice and the KIE assay performed within one week. UCSF ChimeraX (version 1.5) was used for visualization of protein models and preparation of manuscript figures (Pettersen et al. 2021; Goddard et al. 2018).

3.3.3 Rubisco KIE assay summary

We used a substrate depletion method to measure the KIE of each rubisco as used previously in similar studies (Guy et al. 1993; McNevin et al. 2006; Scott et al. 2004; Thomas et al. 2018). Briefly, this method relies on measuring the time-varying δ^{13}C value of the CO_2 pool as the reaction goes to completion instead of directly measuring the difference in δ^{13}C between the initial CO_2 and final 3PG pool. The KIE is then calculated from these data using a Rayleigh relationship, which considers the log-log transformation of the CO_2 isotope data against the fraction of substrate consumed. Linear regression of these data can then be converted to a measure of the instantaneous isotope fractionation—the empirical measure of the isotope effect associated with rubisco carboxylation. With this formulation, larger KIEs correspond to steeper slopes in a Rayleigh plot.

The assay mix we used is based on previous similar studies. In this set-up, inorganic carbon is supplied as HCO_3^- which is converted to CO_2 by a carbonic anhydrase (CA), typically derived from bovines. CO_2 and RuBP is then catalyzed by rubisco to create 3PG. Therefore, our reaction mixture contains CA, rubisco, HCO_3^-, and RuBP to yield the full reaction, and additional reagents including: (i) $MgCl_2$ (Sigma-Aldrich, St. Louis, MO, USA) to support correct rubisco active site metalation, (ii) bicine (Sigma-Aldrich, St. Louis, MO, USA) as a buffer, and (iii) dithiothreitol (DTT) (Sigma-Aldrich, St. Louis, MO, USA) to prevent rubisco oxidation and degradation (Marcus et al. 2003).

In our experiment, instead of limiting CO_2, we limited RuBP. In addition, f (the proportion of CO_2 remaining) is typically known from an external measurement. Prior experiments have labored to constrain f by taking a separate aliquot of the assay to measure CO_2 concentration directly (Guy et al. 1993; Scott et al. 2004). In our experiment, we converted sampling time to f by fitting our data to the model y = a*EXP(−b*x) + c based on the fact that the δ^{13}C of the reactant pool will increase during the reaction and then asymptote to a fixed value as the reaction ceases (i.e., no further carbon isotope discrimination can occur because rubisco can no longer pull from the CO_2 pool as RuBP runs out). In essence, we are interested in the curvature of this line, similar to prior rubisco assays where the δ^{13}C of the reaction vessel headspace was monitored continually on a membrane inlet mass spectrometer (McNevin et al. 2006) instead of traditional methods where discrete aliquots are taken (Guy et al. 1993). See below and Supplemental for further discussion.

3.3.4 Assay preparation and execution

Prior to running the KIE assay, the activity of bovine erythrocytes CA (Sigma Aldrich; St. Louis, MO, USA C3934) was checked following manufacturer guidelines (Anon n.d.). We found a value of 3368 W-A units/mg protein, which exceeded the product stated value of $\geq$2000 W-A units/mg protein, and so we proceeded to use this active CA enzyme prep in the KIE assay.

Glass sampling vials with septum tops ('Exetainer,' 12 mL, Labco, Lampeter, UK) were prepared. First, three external standards were prepared by weighing out Carrara marble standards (CIT_CM2013, $\delta^{13}C = 2.0 \pm 0.1$‰) into individual exetainers. Standards were then sealed within each tube, purged with He gas for 5 min, and then acidified by needle injection with concentrated phosphoric acid (42% *v/v*) (Sigma-Aldrich, St. Louis, MO, USA). Then, three HCO_3^- substrate exetainers were also sealed, purged with He gas, acidified by needle injection of phosphoric acid to convert HCO_3^- to CO_2, and placed in a 70 °C water bath for at least 20 min. Finally, 22 exetainer sampling vials were prepared for the rubiscos (12 for L_8, 10 for L_8S_8). All sampling tubes were first sealed and purged with He gas for 5 min, and then injected with ~1 mL of anhydrous phosphoric acid (Sigma-Aldrich, St. Louis, MO, USA). The phosphoric acid both stops the reaction progress and converts all dissolved inorganic carbon species into CO_2 for analysis.

Next, the reaction assay for each rubisco was prepared. First, a CA stock solution was made by dissolving bovine erythrocytes CA into DI water. Next, an RuBP stock solution was made by dissolving D-Ribulose 1,5-bisphosphate sodium salt hydrate (Sigma Aldrich; St. Louis, MO, USA R0878) in DI water. Then, one drop of concentrated hydrochloric acid (38% *v/v*) was added to 20 mL of autoclaved DI water while it was simultaneously stirred with a stir bar and vigorously bubbled with N_2 gas for 10 min to remove any residual HCO_3^- or CO_2. Then, while N_2 gas was blown over the surface of the solution to inhibit O_2, reagents were added to create a final concentration of 100 mM bicine, 30 mM $MgCl_2$, 1 mM dithiothreitol (DTT) (St. Louis, MO, USA), and 6.25 mM $NaHCO_3$ (St. Louis, MO, USA). pH was adjusted to 8.5 with NaOH and HCl. CA from the CA stock was added, and then either the L_8 or L_8S_8 rubisco was added to the solution. We used 0.996 mg of L_8S_8 and 1.18 mg of L_8 rubisco. The solution was gently bubbled with N_2 gas for 10 min while rubisco 'activated.' While the solution was bubbling, the syringes used for each rubisco assay were rinsed with ethanol and water. We used a separate 25 mL gas-tight syringe with a sample-locking needle for each rubisco (Ref #86326, Model 1025 SL SYR, Hamilton Company, Reno, NV, USA).

RuBP was then added to each reaction assay and mixed through pipetting and swirling. This entire solution was then quickly transferred to a gas-tight syringe. The first time point (t = 0 min) was taken as quickly as possible after transfer. To sample, ~1 mL of the reaction assay was injected into the pre-prepared sampling exetainer containing phosphoric acid. Each assay was sampled 10–12 times over 390 min.

A control was run in a separate experiment, where all the assay components were mixed together with the exception of a rubisco enzyme. The $\delta^{13}C$ of the measured headspace did not change appreciably during this time period, with $\delta^{13}C$ = −0.42 ± 0.03‰ at 0 min and $\delta^{13}C$ = −0.55 ± 0.03‰ at 277 min. The absolute values of these measurements reflect the $\delta^{13}C$ of the substrate used on that experimental day and cannot be related to the data shown here.

3.3.5 Isotopic measurement

The $\delta^{13}C$ of CO_2 in the headspace of each exetainer was measured on a Delta-V Advantage with Gas Bench and Costech elemental analyzer (Thermo Scientific, Waltham, MA, USA) at Caltech. Before measuring samples, two tests were performed to ensure the instrument was functioning normally: (i) An 'on/off' test with an internal CO_2 standard for instrument sensitivity and to establish a baseline intensity at a 'zero' CO_2 concentration, and (ii) a linearity test where the concentration of CO_2 was increased linearly within the designated sensitivity range of the instrument to ensure that a linear increase in CO_2 concentration corresponds to a linear increase in electrical signal on the collector cups. We measured the concentration of $^{12}CO_2$ at mass 44, and $^{13}CO_2$ at mass 45. The instrument was also tuned to ensure that each mass was measured at the center of its mass peak.

The headspace of each sample and standard was measured 10 times (10 analytical replicates), with an internal CO_2 reference run before and after each suite of 10 analytical replicates. Data were visually inspected to ensure the sample was being measured within the correct sensitivity range of the instrument (i.e., of similar intensity and pressure as the internal CO_2 reference). The 'raw' $\delta^{13}C$ values were then corrected relative to VPDB using the three external standards. Assay results can be seen in Table 3.2 and Fig. 3.2.

3.3.6 Calculation of KIE

We first pre-processed the data by assessing which data points to fit. We expected the $\delta^{13}C$ of CO_2 to increase following an exponential curve that eventually reaches an asymptote, but the last few data points start to decrease in $\delta^{13}C$. This may be due to a variety of reasons, including: (1) Ambient CO_2 contaminating the exetainer containers as they are left out after the reaction; (2) re-equilibration of the aqueous and gaseous inorganic carbon pools; or (3) instrument error upon needle sampling of exetainer vial. Because exponential curves are linear in a log-log space, we therefore log-transformed the data points then systematically fit a linear regression through varying sets of data and calculated the resulting error (adjusted R^2 value). The adjusted R^2 value consistently decreased after data point 9 for L_8, and after data point 8 for L_8S_8 (Fig. 3.2B,C). Therefore, we proceeded to use data points 1–9 for L_8 and 1–8 for L_8S_8.

We then converted time to f, the fraction of the inorganic C pool remaining. Since RuBP was the limiting substrate, we could calculate the moles of

CO_2 consumed if we assume: (i) A 1:1 ratio of RuBP to CO_2 was utilized by Rubisco, and (ii) full consumption of the RuBP pool. For each rubisco assay, 125 μmol of RuBP and 9.84 μmol of $NaHCO_3$ were added. Therefore, 7.87% of the initial CO_2 pool was consumed, or $F = 0.9213$. We then assume that $f = 1$ at $t = 0$, and $f = 0.9213$ at the upper bound of the fit. A general model of y = a*EXP(−b*x) + c was applied to the data, though with carbon isotope data in the ^{13}R format instead of the $\delta^{13}C$ format because ^{13}R values can be manipulated arithmetically while $\delta^{13}C$ values cannot (Hayes 1983). The model was then fitted using the non-linear least squares function (call: *nls*(); R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021))). Model outputs are shown in Table 3.2 and Fig. 3.3.

Time was then converted to f using the equation:

$$f = 1 - \left(\frac{R_i - R_0}{R_{upper} - R_0} \times (1 - F) \right)$$

Equation 3.1

where R_0 is the first measured ^{13}R value in each set of data, R_{upper} is the fitted parameter c from the model and $F = 0.9213$, which is calculated from the amount of RuBP added to the assay.

Next, a correction was done to account for the C isotope fractionation between CO_2 and HCO_3^- at equilibrium, where CO_2 is ~8‰ lighter (more negative $\delta^{13}C$ value) than HCO_3^- (Mook et al. 1974). We followed the correction outlined in (Guy et al. 1993) where the adjustment is applied before linear regression in a Rayleigh plot:

$$R/R_0 \ adj. = \frac{(fR/R_0)^C}{f}$$

Equation 3.2

where $C = (1.009 + 10\hat{\ }(pK − pH))/(1 + 10\hat{\ }(pK − pH))$. The pK is that of carbonic acid, for which we used a value of 6.4 (Zeebe and Wolf-Gladrow 2001). The pH of the L_8S_8 assay was 8.49, and the pH of the L_8 assay was 8.52.

Finally, a Rayleigh plot was made for each rubisco plotting $\ln(^{13}R/^{13}R_0)_{adj.}$*1000 vs. $-\ln(f)$ (Fig. 3.4). The best fit slope, D, was calculated using a linear regression (call: *lm*(); R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021))). D was then converted to Δ, the KIE, using the equation $\Delta = D/(1 − D/1000)$ (Guy et al. 1993). Doing so, we found the KIE of the L_8S_8 rubisco to be 22.42 ± 2.37, and 16.25 ± 1.36 for the L_8 rubisco. Results are shown in Table 3.1.

Strain	Rubisco	KIE (‰)	V_C (s^{-1})	K_C (µM)	$S_{C/O}$	V_O (s^{-1})	K_O (µM)
Synechococcus elongatus PCC6301	L_8S_8	22.42 ± 2.37	14.3 ± 0.71	235 ± 20.0	56.1 ± 1.3	1.10	983 ± 81
Candidatus Promineofilum breve	L_8	16.25 ± 1.36	2.23 ± 0.04	22.2 ± 9.7	36.1 ± 0.9	1.11	401 ± 115

Table 3.1. KIE and non-isotopic kinetic measurements from L_8 vs. L_8S_8 rubiscos.
KIEs were measured in this study using the substrate depletion method (Guy et al. 1993; McNevin et al. 2006; Scott et al. 2004; Thomas et al. 2018); see Methods for more detail. Non-isotopic kinetic measurements are from (Banda et al. 2020). V_C and V_O indicate maximum carboxylation and oxygenation rates under substrate-saturated conditions, respectively; K_C and K_O are Michaelis constants for the carboxylation and oxygenation reactions, respectively; $S_{C/O}$ indicates specificity, a unitless measure of the relative preference for CO_2 over O_2 and is calculated as $S_{C/O} = (V_C/K_C)/(V_O/K_O)$. Uncertainties on non-isotopic kinetics reflect mean ± s.e.m. from multiple experiments; see (Banda et al. 2020) for more detail. Error on KIEs reflect mean ± s.d. from model fitting uncertainty from one experiment; see Methods and Supplemental for more detail.

3.4 Results

3.4.1 L_8 rubisco has a smaller KIE than its L_8S_8 counterpart

The KIE of the L_8 rubisco is ≈5‰ less than that of the L_8S_8 rubisco (16.25 ± 1.36‰ vs. 22.42 ± 2.37‰, respectively; Table 3.1). We note that there is variation among KIE measurements of similar or the same strains. Prior measurements which we compare our data against (Fig. 3.1, Table 3.3) are bacterial (Form II, Form I') or Cyanobacterial (Form I) rubisco measurements, where a pure enzyme, substrate-depletion assay such as ours was performed on well-characterized strains where rubisco was obtained through expression and subsequent purification from *E. coli*. We also included one Form II measurement from a *Riftia pachyptila* symbiont, *Candidatus* Endoriftia Persephone (Robinson et al. 2003), where rubisco was purified from the host trophosome because at the time of the measurement the symbiont could not be cultured separately from the host, though a complete genome has recently been published (De Oliveira et al. 2022). Therefore, we did not include measurements where a non-native bacterial rubisco was expressed by another organism in vivo and KIE calculated by extrapolating ratios of intracellular to extracellular CO_2 (von Caemmerer et al. 2014), nor measurements from plants or the *Solemya velum* symbiont because it is not a member of the Cyanobacteria (Scott et al. 2004). It has been proposed and measured that rubisco KIEs vary with pH, temperature, and metal ion concentrations (Whelan et al. 1973; O'Leary 1978), yet other studies contradict this claim (Christeller and Laing 1976) and have instead proposed that much of the variation in the literature

reflects experimental uncertainties rather than intrinsic variations in KIE (Farquhar et al. 1982). This study and (Wang et al. 2023) measured an L_8S_8 rubisco KIE from *Synechococcus elongatus* PCC6301 and 7942, respectively (identical RbcL and RbcS sequences) in similar assay conditions but found values that are similar but do not overlap in uncertainty, supporting the conclusion that variations in reported KIE values are due to experimental uncertainty rather than intrinsic enzymatic variations. However, the KIEs presented in Fig. 3.1 were measured in assays that span a range of pH, temperature, and $MgCl_2$ concentrations (Table 3.3), notably with increasing $MgCl_2$ concentration corresponding with increasing KIEs measured in the Form II rubisco by (Guy et al. 1993). Because of the lack of repeated, rigorous measurements of multiple rubisco KIEs across variations relevant parameters (i.e., pH, temperature, metalation), it is difficult to conclude what is causing the variation in KIE values across studies. Therefore, we can only conclude that the L_8 rubisco KIE is less (by roughly 5‰) than its L_8S_8 counterpart measured in this study, and less than the range of L_8S_8 rubiscos measured from previous studies.
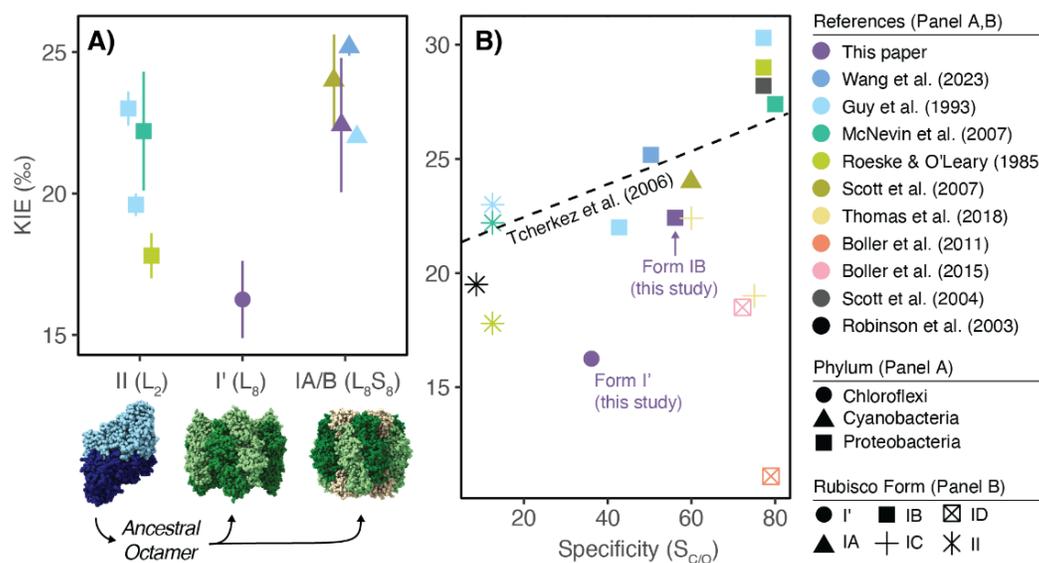
Figure 3.1. Form I' rubisco fractionates less than both Form II and Form I rubiscos, and cannot be explained by prior model relating specificity and KIE.

(**A**) KIE (‰) for relevant bacterial Form II (L_2), Form I' (L_8), and Cyanobacterial Form IA/B (L_8S_8) rubiscos with representative rubisco structures below; Protein Data Bank (PDB) codes from left to right: 5RUB, 6URA, 1RBL. Hypothesized evolutionary pathway is shown in black arrows, showing that ancestral dimers (L_2) likely evolved to a common ancestral octamer (L_8) (Schulz et al. 2022) that then speciated into either Form I' (L_8) or Form I (L_8S_8) rubiscos (Banda et al. 2020). Rubisco phylum is shown as shapes and references are shown in colors. Form II KIEs are from *Rhodospirillum rubrum* or *Candidatus* Endoriftia persephone (Guy et al. 1993; Robinson et al. 2003; McNevin et al. 2007; Roeske and O'Leary 1985), Form I' measurement is from *Candidatus* Promineofilum breve (this study), all Form IB rubiscos are from *Synechococcus elongatus* PCC6301 or 7942 (identical RbcL and RbcS sequence) (Guy et al. 1993; Wang et al. 2023) and this study, and Form IA KIE is from *Prochlorococcus marinus* MIT9313 (Scott et al. 2007). Error is reported as 95% confidence intervals for (Scott et al. 2007); as standard deviation for this study and (Wang et al. 2023; McNevin et al. 2007; Roeske and O'Leary 1985); as standard error for (Guy et al. 1993). See Table 3.3 for literature values used, notes on variation between measurements, and rationale for which data was included and excluded. For recent compilation of all measured rubisco KIEs, see (Garcia et al. 2021). (**B**) Compilation of additional KIE and specificity values in Form IC and ID rubiscos (Guy et al. 1993; Banda et al. 2020; Scott et al. 2004; Thomas et al. 2018; McNevin et al. 2007; Shih et al. 2016; Davidi et al. 2020; Horken and Tabita 1999; Badger et al. 1998; Haslam et al. 2005; Read and Tabita 1994; Kane et al. 1994; Boller et al. 2015; Boller et al. 2011; Roeske and O'Leary 1984), in addition to data shown in Fig. 3.1A. Forms shown in shapes, references shown in the same colors as in Panel A. See Tables 3.2 and 3.3 for compilation of data used. Dotted line indicates original linear regression from (Tcherkez et al. 2006). Figure was prepared with the assistance of the *ggplot2* package (v.3.3.66; (Wickham et al. 2016)).

Similarly, compared to prior Form II (L_2) rubisco KIE measurements, the Form I' (L_8) rubisco may fractionate less. Compared to Form I KIEs, there is wider variation in previously measured Form II KIEs, with the Form I' rubisco measured here overlapping in value with one Form II rubisco within uncertainty (Roeske and O'Leary 1985). We note that all the Form II data presented here are largely from one species, *Rhodospirillum rubrum*, though the specific strain is not reported for all studies. Therefore, the variations may reflect experimental uncertainty with the

exception of the measurement in (Guy et al. 1993), where $MgCl_2$ concentration was changed. Therefore, we are not confident concluding either way if the L_8 KIE is less than the L_2 KIE or not.

3.5 Discussion

3.5.1 Presence or absence of RbcS external to active site may influence KIE

Rubisco KIEs have also been used to support conclusions gleaned from non-isotopic kinetic parameters, both to better understand the reaction mechanism and to offer complementary data to traditional measurements, but our results belie an easy interpretation within that existing framework. The dominant theory in this field posits that rubisco specificity is positively correlated with the CO_2 KIE because of an observed increase in carbon isotope fractionation, but not oxygen isotope fractionation, with specificity (Tcherkez et al. 2006; Guy et al. 1993). This argument originates from studies of deuterium (D) isotope effects on enzymatic reaction rates, which have been traditionally performed because deuterium displays a much larger (and easier to measure) KIE due to the large relative mass difference between D and its major isotope, H, in comparison to other rare isotopes such as ^{13}C vs. ^{12}C or ^{15}N vs. ^{14}N (Frey and Hegeman 2007). These foundational experiments have led to the conclusion that the isotope effect is determined at the rate-limiting step at the transition state, and small asymmetries in the transition state caused by transition state structure will cause small variations in the isotope effect (Frey and Hegeman 2007; Westheimer 1961). Applied to rubisco, (Tcherkez et al. 2006) proposed that the inherent difficulty in binding a 'featureless' CO_2 vs. O_2 molecule has caused natural selection in the transition state, where rubiscos that maximize the structural difference in transition states for carboxylation vs. oxygenation are able to be more specific. That then causes a trade-off where greater resemblance to the final carboxyketone product causes the product to also be tightly bound, leading to a higher $S_{C/O}$ correlating with a lower V_C, but also a prediction that the intrinsic KIE for CO_2 addition (but not O_2 addition) should increase as the transition state becomes more product like, i.e., higher-specificity rubiscos should have higher KIEs, which is indeed what the data at the time supported (Tcherkez et al. 2006). This has also led to the conclusion that rubisco is actually perfectly optimized for the time and places where it is found today, precluding any opportunity to use rubisco engineering to achieve increased biomass yields (Tcherkez et al. 2006).

However, new CO_2 KIE measurements that do not show a correlation with specificity are empirically questioning this conclusion (Fig. 3.1B). Prior studies (Thomas et al. 2018) have pointed out that the spread in KIE data, particularly at high specificity, cannot easily be described by a simple inverse relationship or linear regression. Indeed, our Form I' measurement lies below the original regression line (dashed line in Fig. 3.1B) proposed in (Tcherkez et al. 2006); its KIE is effectively too low given what one would predict via its specificity. However, although an increasing spread in CO_2 KIE becomes apparent as more

rubiscos are measured, they cannot directly address the dominant theory because of the general dearth of O_2 KIE measurements. In addition, specificity is typically not reported in the same study with KIE (see notes in Tables 3.2 and 3.3), so some of the spread in Fig. 1B may be due to uncertainties in the true specificity for the given rubisco measured. Therefore, additional paired measurements of CO_2 and O_2 KIEs with specificity are necessary before a new theory relating isotopic and non-isotopic kinetics can be proposed; more data are needed to decide between potential theories.

In addition, this transition state optimization theory is based on the assumption that it is the active site (which binds the intermediary carboxylation or oxygenation product) that concurrently affects both specificity and KIE, so the naïve assumption is that the absence or presence of the small subunit, which does *not* contain the active site, should not affect KIE. Unexpectedly, the L_8 rubisco fractionates roughly 5‰ less than that of the L_8S_8 rubisco (16.25 ± 1.36‰ vs. 22.42 ± 2.37‰, respectively). The specificity of the L_8 rubisco is indeed less than that of the L_8S_8 (36.1 ± 0.9 vs. 56.1 ± 1.3, respectively, (Banda et al. 2020)) but this may be a coincidence because that prediction is based on a theory reliant on rubisco's active site which the small subunit does not directly impact. Our comparative study suggests the tantalizing hypothesis that the small subunit increases rubisco KIEs. However, Form I' has only recently been discovered (Banda et al. 2020) and only a limited number of sequences exist. Future work consisting of dual CO_2 and O_2 KIE measurements of other novel Form I' rubiscos compared to Form I rubiscos, across a range of assay parameters, will be needed for a more robust comparative study. Potentially, comparative studies of extant L_8 vs. L_8S_8 rubiscos could be complemented with experiments using ancestral rubiscos demonstrated to not require RbcS–RbcL interactions (Schulz et al. 2022) that would allow one to effectively strip the small subunit from an L_8S_8 rubisco and measure its effect on the KIE. Similarly, pairings of one RbcL sequence with various RbcS sequences of tobacco rubiscos (Lin et al. 2020) would allow one to test how various small subunits affect the KIE in Form I (L_8S_8) rubiscos. Alternately, it has been shown that mutations distal from the active site affecting oligomerization can affect enzyme kinetics, which is somewhat analogous to losing RbcS in that does not directly interact with the active site. KIE measurements from such rubiscos may also help shed help shed light on the relationship between RbcS, specificity, and KIE (Liu et al. 2022). Therefore, it remains an open question what structural and biochemical aspects of rubisco may also affect KIEs in addition to active site and transition state theory mechanisms.

3.5.2 Supports prior work positing that rubisco KIEs vary across phylogeny in the modern day and across time

Our work supports previous work showing that the rubisco KIE varies across phylogeny in the modern day, though with the caveats that few unique rubiscos have been measured, there is variation across experiments, and the vast

majority of measurements are from Form I rubiscos (Fig. 1B, and see (Garcia et al. 2021; Thomas et al. 2018) for recent compilation across phylogeny). A smaller KIE measured from one novel Form I' rubisco, in comparison to the bacterial Form I rubiscos, supports this observation, though more measurements across the Form I' clade are needed to quantify any potential in-clade variation.

In addition, if we view L_8 as an evolutionary 'missing link' between the evolution of L_2 and L_8S_8 rubiscos, this measurement supports the idea that rubisco KIE may have varied across evolutionary time. Prior work has explored this question by measuring the KIE of a putative Precambrian ancestral Form IB rubisco reconstructed using a combination of phylogenetic and molecular biology techniques (Shih et al. 2016); that study found the ancestral rubisco to fractionate less than its modern counterpart ($17.23 \pm 0.61‰$ vs. $25.18 \pm 0.31‰$, respectively) (Wang et al. 2023). Interestingly, the Form I' and putative ancestral Form IB rubisco have similar, lower KIE values ($16.25 \pm 1.36‰$ vs. $17.23 \pm 0.61‰$, respectively) compared to most modern Form I rubiscos (roughly 20-30‰; for recent review see (Garcia et al. 2021)). This supports prior predictions that the KIE should have varied over geologic time in response to changing pCO_2, though that prediction was based on an assumption of inverse correlation between specificity (selected for by changing CO_2/O_2 levels) and KIE (Tcherkez et al. 2006). This implies that the KIE of ancestral rubiscos may have been lower than modern rubiscos today, though this is a tentative hypothesis that, by necessity, relies on ancestral enzyme reconstruction and comparative biology techniques instead of direct measurements of 'true' ancestral enzymes.

Finally, it is hypothesized that the small subunit may have evolved in response to rising atmospheric oxygen levels roughly 2.4 billion years ago because the high V_O stabilization that RbcS offers allows simultaneous exploration of RbcS and RbcL protein space (Banda et al. 2020). Therefore, understanding the KIE of Form I' rubiscos may allow us to better understand changes in rubisco biochemistry that may have accompanied evolutionary changes and facilitate better tracking of carbon mass flux at key times in Earth's evolutionary history.

3.6 Acknowledgments

3.7 Author contributions

Conceptualization, W.W.F. and P.M.S.; Methodology, R.Z.W., A.K.L., and D.M.B.; Formal Analysis, R.Z.W.; Writing—Original Draft Preparation, R.Z.W.; Writing—Review and Editing, R.Z.W., W.W.F., A.K.L., and P.M.S.; Funding Acquisition, W.W.F. and P.M.S. All authors have read and agreed to the published version of the manuscript.

3.8 Competing interests

The authors declare no conflict of interest.

3.9 Supplementary materials

Rubisco	Time (min)	$\delta^{13}C$ (avg.)	$\delta^{13}C$ (std. err.)	^{13}R (avg.)	^{13}R (std. err.)
L_8	0	−0.592	0.010	0.0111549	$1.81*10^{-7}$
L_8	15	−0.425	0.007	0.0111580	$1.24*10^{-7}$
L_8	30	−0.129	0.021	0.0111634	$3.92*10^{-7}$
L_8	45	0.111	0.026	0.0111679	$4.80*10^{-7}$
L_8	60	0.268	0.017	0.0111708	$3.11*10^{-7}$
L_8	90	0.327	0.010	0.0111718	$1.81*10^{-7}$
L_8	120	0.506	0.007	0.0111751	$1.35*10^{-7}$
L_8	150	0.473	0.012	0.0111745	$1.35*10^{-7}$
L_8	210	0.652	0.013	0.0111778	$2.34*10^{-7}$
L_8	270	0.454	0.007	0.0111742	$1.26*10^{-7}$
L_8	330	0.399	0.006	0.0111732	$1.08*10^{-7}$
L_8	390	0.348	0.014	0.0111722	$2.55*10^{-7}$
L_8S_8	0	0.599	0.026	0.0111768	$4.68*10^{-7}$
L_8S_8	15	0.990	0.017	0.0111840	$3.14*10^{-7}$
L_8S_8	30	1.058	0.008	0.0111853	$1.53*10^{-7}$
L_8S_8	45	1.553	0.015	0.0111944	$2.69*10^{-7}$
L_8S_8	60	1.490	0.010	0.0111932	$1.84*10^{-7}$
L_8S_8	90	1.776	0.015	0.0111985	$2.82*10^{-7}$
L_8S_8	120	1.905	0.013	0.0112009	$2.33*10^{-7}$
L_8S_8	150	1.997	0.011	0.0112025	$1.92*10^{-7}$
L_8S_8	210	1.951	0.009	0.0112017	$1.64*10^{-7}$
L_8S_8	270	1.948	0.008	0.0112016	$1.44*10^{-7}$

Table 3.2. Results of rubisco KIE assay.
Experimental outputs of rubisco KIE assay; $\delta^{13}C$ vs. time is plotted in Figure S1A. Average $\delta^{13}C$ or ^{13}R ($n = 10$ analytical replicates) is reported with standard error (standard deviation divided by square root of n).

Strain	Form	Phylum	Specificity	KIE	Notes	Specificity Reference	KIE Reference
Prochlorococcus marinus MIT9313	IA	Cyano-bacteria	59.9 ± 7.0	24.0 [22.2, 25.6][a]	pH 7.5, 25 mM $MgCl_2$, 25C, expressed from *E coli*	(Shih et al. 2016)	(Scott et al. 2007)
Synechococcus elongatus PCC6301	IB	Cyano-bacteria	56.1 ± 1.3	22.42 ± 2.37[b]	pH 8.49, 30 mM $MgCl_2$, 22C, expressed from *E coli*	(Banda et al. 2020)	This paper
Synechococcus elongatus PCC6301	IB	Cyano-bacteria	50.3 ± 2.0	25.18 ± 0.31[b]	pH 8.38, 30 mM $MgCl_2$, 22C, expressed from *E coli*	(Shih et al. 2016)	(Wang et al. 2023)
Synechococcus elongatus PCC6301	IB	Cyano-bacteria	42.7 ± 2.8	22.0 ± 0.2[c]	pH 8.1, 25 mM Mg^{2+}, 25C, expressed from *E coli*	(Davidi et al. 2020)	(Guy et al. 1993)
Candidatus Promineofilum breve	I'	Chloro-flexi	36.1 ± 0.9	16.25 ± 1.36[b]	pH 8.52, 30 mM $MgCl_2$, 22C, expressed from *E coli*	(Banda et al. 2020)	This paper
Rhodospirillum rubrum	II	Proteo-bacteria	12.5 ± 0.6	23.0 ± 0.6[c]	pH 7.9, 25 mM Mg^{2+}, 25C, expressed from *E coli*	(Davidi et al. 2020)	(Guy et al. 1993)
Rhodospirillum rubrum	II	Proteo-bacteria	12.5 ± 0.6	19.6 ± 0.4[c]	pH 7.9, 2 mM Mg^{2+}, 25C, expressed from E coli	(Davidi et al. 2020)	(Guy et al. 1993)
Rhodospirillum rubrum	II	Proteo-bacteria	12.5 ± 0.6	22.2 ± 2.1[b]	pH 8.0, 20 mM $MgCl_2$, room temp?, expressed from *E coli* (XL1-blue)	(Davidi et al. 2020)	(McNevin et al. 2007)
Rhodospirillum rubrum	II	Proteo-bacteria	12.5 ± 0.6	17.8 ± 0.8[b]	pH 7.8, 10 mM $MgCl_2$, 25C, "gift from John Schloss"	(Davidi et al. 2020)	(Roeske and O'Leary 1985)

Strain	Form	Phylum	Specificity	KIE	Notes	Specificity Reference	KIE Reference
Candidatus Endorifita persephone	II	Proteo-bacteria	8.6 ± 0.9	19.5 ± 1.0^c	pH 8.0, 30C, 5 mM $MgCl_2$, purified from *R pachyptila* trophosomes	(Robinson et al. 2003)	(Robinson et al. 2003)

Table 3.3. Literature compilation of data used to make Figure 3.1A.
For KIE measurements: Each figure reports uncertainty on the measurement in a different way; superscripts indicate: *a* 95% confidence interval; *b* standard deviation; *c* standard error. Strains for *R. rubrum* not specified in (Guy et al. 1993; McNevin et al. 2007; Roeske and O'Leary 1985). We only used data where a pure enzyme, substrate-depletion assay like ours was done. In addition, we only used data from well-characterized strains where rubisco was obtained through expression in *E. coli*. Therefore, we are not including the (von Caemmerer et al. 2014) measurement because it was done in a tobacco plant mutant expressing an *R. rubrum* rubisco sequence *in vivo*, and KIE was calculated by extrapolating to a ratio of intercellular to ambient CO_2 (C_i/C_a) of 1. In addition, we are only showing Form IA/B data from Cyanobacteria and therefore do not include plants or the *Solemya velum* symbiont (Scott et al. 2004). Assay temperature was assumed to be room temperature for (McNevin et al. 2007); rubisco was assumed to be expressed from *E. coli* in (Roeske and O'Leary 1985). In addition, only the pH 7.9, 25 mM Mg^{2+} condition from (Guy et al. 1993) was plotted in Figure 1B. See Table 3 in (Garcia et al. 2021) for a recent compilation of all measured KIEs. For Specificity measurements: Most specificity values were not reported with the study, with the exception of this paper and (Robinson et al. 2003; Wang et al. 2023). Therefore, specificity values were taken from (Banda et al. 2020; Shih et al. 2016) where indicated.

Strain	Form	Specificity	KIE (‰)	Specificity Reference	KIE Reference
Ralstonia eutropha	IC	75	19 [17.5, 20.4]	(Horken and Tabita 1999)	(Thomas et al. 2018)
Rhodobacter sphaeroides	IC	60	22.4 [21.1, 24.0]	(Horken and Tabita 1999)	(Thomas et al. 2018)
Emiliania huxleyi	ID	79	11.1 [9.8, 12.6]	(Badger et al. 1998)	(Boller et al. 2015)
Skeletonema costatum	ID	72.2 ± 2.2	18.5 [17.0, 19.9]	(Haslam et al. 2005)	(Boller et al. 2011)
Spinacia oleracea	IB	77.2 ± 1.4	30.3 ± 0.8	(Read and Tabita 1994)	(Guy et al. 1993)
Spinacia oleracea	IB	77.2 ± 1.4	29 ± 1	(Read and Tabita 1994)	(Roeske and O'Leary 1984)
Spinacia oleracea	IB	77.2 ± 1.4	28.2 [26.6, 29.8]	(Read and Tabita 1994)	(Scott et al. 2004)
Nicotiana tabacum	IB	82.1 ± 0.8	27.4 ± 0.9	(Kane et al. 1994)	(McNevin et al. 2007)

Table 3.4. Additional specificity and KIE values used for Figure 2.1B.
Data compilation is similar to that used in Figure 4 from (Thomas et al. 2018). Most specificity values were measured separate from the KIE and are taken from other prior literature (Horken and Tabita 1999; Badger et al. 1998; Haslam et al. 2005; Read and Tabita 1994; Kane et al. 1994), similar to what was done by (Thomas et al. 2018). *Solemya velum* gill symbiont (Form IA, KIE = 24.4‰) from (Scott et al. 2004) was not included because the specificity could not be found. In addition, (Guy et al. 1993) gives two values at two different assay conditions for *S. oleracea*; here we use the value at pH 8.5, 20 mM MgCl$_2$ but they also report a KIE of 29.0 ± 0.3‰ at pH 7.6, 5 mM Mg^{2+}. KIE values are from (Guy et al. 1993; Scott et al. 2004; Thomas et al. 2018; McNevin et al. 2007; Boller et al. 2015; Boller et al. 2011; Roeske and O'Leary 1984). Error in brackets is reported as mean with 95% confidence intervals; otherwise error is reported as mean ± s.e.

Rubisco	Parameter	Estimate	Std. Error	t value	Pr(>|t|)	Signif. Code
L_8S_8	a	$-2.786*10^{-5}$	$2.759*10^{-6}$	-10.097	$1.63*10^{-4}$	***
L_8S_8	b	$1.640*10^{-2}$	$4.310*10^{-3}$	3.804	0.012573	*
L_8S_8	c	$1.120*10^{-2}$	$2.908*10^{-6}$	3852.671	$< 2*10^{-16}$	***
L_8	a	$-2.338*10^{-5}$	$1.369*10^{-6}$	-17.072	$2.58*10^{-6}$	***
L_8	b	$1.769*10^{-2}$	$2.767*10^{-3}$	6.392	$6.90*10^{-4}$	***
L_8	c	$1.118*10^{-2}$	$1.203*10^{-6}$	9291.937	$< 2*10^{-16}$	***

Table 3.5. Model outputs for converting time to *f*.
The nonlinear least squares function in R Statistical Software was used for calculation with initial guesses of $a = -1*10^{-5}$, $b = 0.1$, $c = 0.01$ for L_8S_8; $a = -1*10^{-4}$, $b = 0.1$, $c = 0.01$ for L_8. The parameter *c* gives R_{upper} in Equation 1, which then allows time to to be convert to *f*. For L_8S_8, the model found a residual standard error of $1.252*10^{-6}$ on 6 degrees of freedom, required 5 iterations to convergence, and achieved a convergence tolerance of $1.405*10^{-6}$. For L_8, the model found a residual standard error of $1.252*10^{-6}$ on 6 degrees of freedom, required 5 iterations to convergence, and achieved a convergence tolerance of $1.405*10^{-6}$. The significant codes indicate: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. All analyses were performed using R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021)).
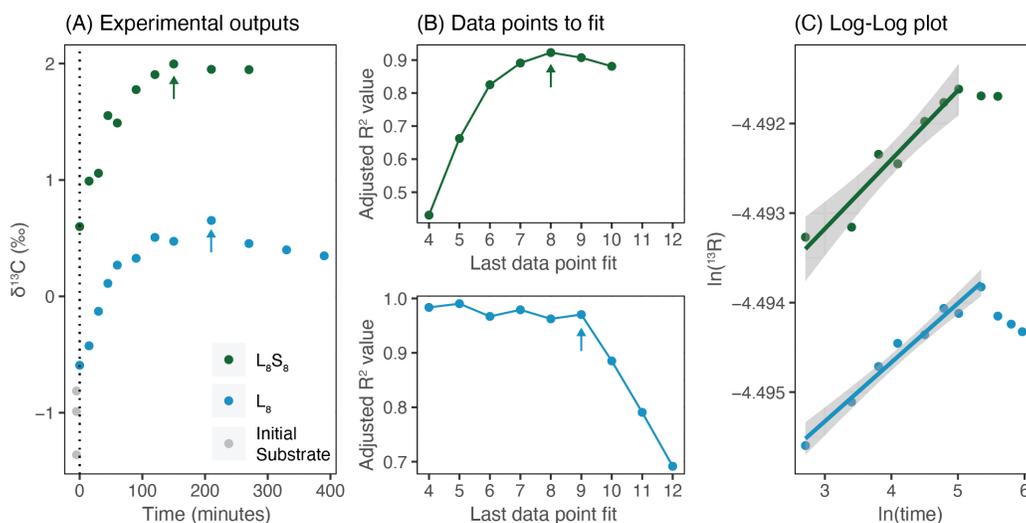
Figure 3.2. Data preprocessing step.
A) Experimental outputs of rubisco KIE assay, showing how the $\delta^{13}C$ of the CO_2 headspace evolves over the experiment. The first time point taken is shown at 0 minutes, and the initial $NaHCO_3$ substrate is shown plotted at -5 minutes for ease of comparison. The arrow indicates the final data point to fit after preprocessing. B) Subplots showing the adjusted R^2 value for the L_8S_8 (above, green) and the L_8 rubisco (below, blue) for linear regressions across different lengths of log-transformed data points. Arrows indicate where the R^2 value starts to decrease (point 8 for L_8S_8 rubisco in green; point 9 for L_8 rubisco in blue); these arrows refer to the same point in Panel A. C) Linear regression across natural log-transformed data to data point 8 for the L_8S_8 rubisco (green) and to data point 9 for the L_8 rubisco (blue). Note the isotopic data is in ^{13}R vs. $\delta^{13}C$ format. The first data point is not plotted because the natural log of zero is undefined. All analyses were performed using R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021)). Data visualization was performed using the *ggplot2* package (v3.3.6; Wickham, 2016, (Wickham et al. 2016)).
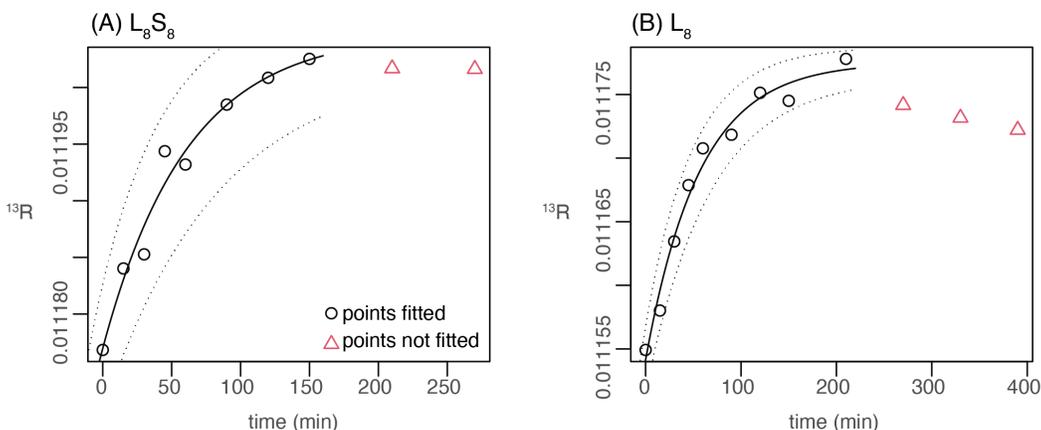
Figure 3.3. Calculating *f* from time.
Plots showing best fit exponential model for (**A**) L_8S_8 vs. (**B**) L_8 rubisco in solid black line. Dotted lines indicate model uncertainty (std. dev.). See Table S3 for best-fit model parameters. Open black circles are points fitted, as determined in Figure S1. Open red triangles are the points not fit. All analyses and data visualization were performed using R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021)).
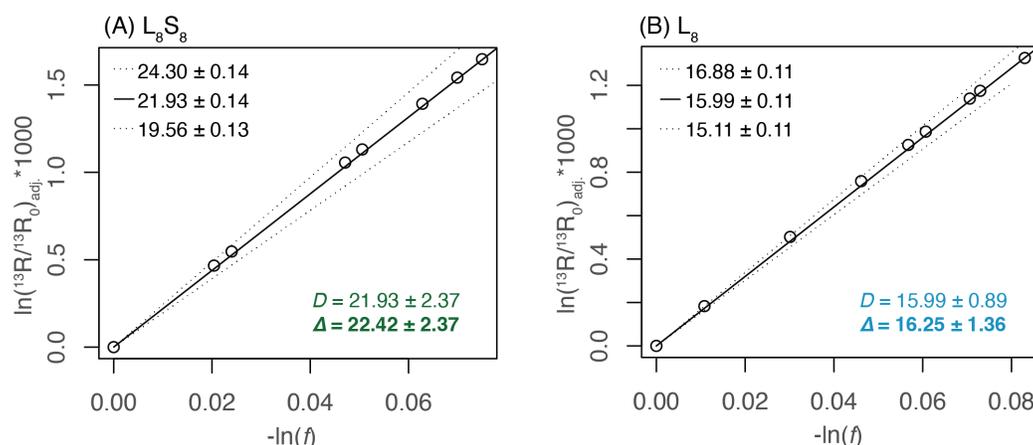
Figure 3.4. Rayleigh plots with equilibrium adjustment.
A) and B) show L_8S_8 and L_8 rubisco with equilibrium adjustment for ^{13}R values (Equation 2; (Guy et al. 1993)) before linear regression. Solid line gives best fit value using f values calculated from the best estimate for parameter c. Dotted lines give fit for f values calculated using the best estimate $\pm$ std. error for c as shown in Table S3. Slopes for each line are reported in the upper left corner (best estimate $\pm$ std. error). D is the slope of the solid, best fit line. Δ is converted from D using $\Delta = D/(1-D/1000)$ (Guy et al. 1993)). All analyses and data visualization were performed using R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021)).

3.10 Appendix A: Rubisco KIE Assay Clarification

We errantly stated that 125 μmol of RuBP and 9.84 μmol of $NaHCO_3$ were added; 9.84 mmol of $NaHCO_3$ was added instead. We also incorrectly stated that we expected the $\delta^{13}C$ of CO_2 to increase following an exponential curve that eventually reaches an asymptote; we meant a logarithmic curve instead.

Like Chapter 2, the Gas Bench with the Delta-V Advantage was used to measure the $\delta^{13}C$ of $CO_{2(g)}$ within the headspace of the exetainers. We did 10 replicate injections of each exetainer. Though we measured at Masses 44-46, we only peak-centered on Mass 45. For all measurements, a ^{17}O correction based on Mass 46 was performed as in (Santrock et al. 1985).

For the rubisco KIE assays, exetainers were measured as soon as possible after the rubisco assay was quenched. However, each isotopic measurement took longer than the time between sampling points. Therefore, the lag time between rubisco quenching and measurement was short for the initial time points (i.e. immediately or ~1 hour), but the lag time was longer for the later points (i.e. more than 3 hours). Therefore, it is likely that the negative spread in $\delta^{13}C$ values seen in the later timepoints reflects contamination from ambient air – i.e. the seals on the exetainers were allowing ambient $CO_{2(g)}$, which has a negative $\delta^{13}C$ value, to enter the exetainer. However, we pre-processed the data to remove these errant values (see Supplemental) so our fitted rubisco KIE values are not affected by this.

In addition, we attempted to use the total amount of CO_2 measured in the headspace to calculate f in our rubisco KIE assays. However, the error on these values was large, likely because we only drew down the total CO_2 pool to a small

amount (i.e. uncertainty was on the order of magnitude of the effect we were trying to measure). Therefore, we calculated *f* from the time sampling points and the change in $\delta^{13}C$ (see Supplemental).

3.11 Appendix B: Effect of Kinetics and Michaelis Constant on KIE

Does K_C affect the measured KIE of rubisco? Perhaps.

K_C alone should not affect the KIE. For an enzymatic reaction, there are two fundamental parameters: i) The maximum reaction velocity at an infinite substrate concentration (V_{max}); and ii) The Michaelis constant (K_M), which is the amount of substrate that yields a velocity of half V_{maz}. The isotope effect should only affect the rate of reaction, or V_{max}.

However, the assay used in this study (and prior studies mentioned above) are actually measuring the isotopic fractionation of V/K, the apparent first-order reaction rate constant for reaction at low substrate concentration – i.e. the initial portion of the logarithmic Michaelis curve that can be generalized to be linear. This is referred to as the 'internal competition method;' see (Cleland 1987) for review. In addition, in the assays typically performed for the rubisco KIE, only the isotopic composition of the substrate pool (CO_2) is measured. The isotope effect can then be described as $^{13}(V_C / K_C) = \log(1 - f) / \log[(1 - f)(R_S/R_0)]$ where *f* is the 'fractional reaction' (extent of reaction completeness) at the time of measurement, R_S is the isotopic ratio of the residual substrate, and R_0 is the isotopic ratio of the initial substrate. See (Cleland 1987) for review.

This can be seen in our assay and others. The $\delta^{13}C$ of the headspace quickly grows heavy until it asymptotes at a certain value; we then fit the curved portion of the graph to derive the KIE. For this study, the reaction stops because RuBP runs out. Since we added 9.84 mmol of $NaHCO_3$, we were above the K_C value for both enzymes and both were likely saturated during our assay.

Therefore, when measuring $^{13}(V_C / K_C)$, K_C may affect the overall measured value if it is large enough to 'swamp out' the effect of isotopic substitution on V_C. For example, if we assume the effect ^{13}C substitution to be similar for all rubisco V_C values, then what we may be measuring are differences in K_C that affect overall measured $^{13}(V_C / K_C)$ values. We do note that in this study, in addition to V_C being much lower for the L_8 vs. L_8S_8 rubisco (2.23 ± 0.04 vs. 14.3 ± 0.71 s^{-1} respectively; Table 3.1), the K_C value is much lower as well (22.2 ± 9.7 vs. 235 ± 20.0 μM respectively; Table 3.1). Intriguingly, the values of V_C/K_C differ by roughly twofold (~0.06 s^{-1} μM^{-1} for L_8S_8 vs. ~0.1 s^{-1} μM^{-1} for L_8), while the KIEs differ by less than twofold (22.42 ± 2.37‰ for L_8S_8 vs. 16.25 ± 1.36‰ for L8).

However, this is dependent on our knowledge of how much ^{13}C substitution affects V_C. If ^{13}C-substitution uniformly affects V_C values of ~1 and ~100 s^{-1}, then K_C would start to matter. However, if say ^{13}C-substitution affects V_C values of ~100 more than values of ~1, then K_C would matter less. Further work is needed to shed light on this issue.

Interestingly, the dominant narrative of enzyme kinetics affecting KIE (Tcherkez et al. 2006) is based around specificity, which is a unitless measure of the relative preference for CO_2 over O_2 and is calculated as $S_{C/O} = (V_C/K_C)/(V_O/K_O)$. Since the KIE is actually reflecting a measurement of $^{13}(V_C / K_C)$, relating specificity and KIE may be redundant – the true shape of this relationship may instead of shaped by the oxygen fractionation, $^{18}(V_O / K_O)$. Indeed, further work relating the oxygen and carbon isotope fractionation of rubisco may shed light on this intriguing problem.

3.12 References

Andrews, T J, and G H Lorimer. 1987. The Biochemistry of Plants: A Comprehensive Treatise, Vol. 10, Photosynthesis. Edited by M D Hatch and N K Boardman. Vol. 10.

Badger, Murray R, T John Andrews, S M Whitney, Martha Ludwig, David C Yellowlees, W Leggat, and G Dean Price. 1998. "The Diversity and Coevolution of Rubisco, Plastids, Pyrenoids, and Chloroplast-Based CO2 - Concentrating Mechanisms in Algae." Canadian Journal of Botany 76 (6): 1052–71. https://doi.org/10.1139/b98-074.

Banda, Douglas M, Jose H Pereira, Albert K Liu, Douglas J Orr, Michal Hammel, Christine He, Martin A J Parry, et al. 2020. "Novel Bacterial Clade Reveals Origin of Form I Rubisco." Nature Plants 6 (9): 1158–66. https://doi.org/10.1038/s41477-020-00762-4.

Bar-On, Yinon M, and Ron Milo. 2019. "The Global Mass and Average Rate of Rubisco." Proceedings of the National Academy of Sciences of the United States of America 116 (10): 4738–43. https://doi.org/10.1073/pnas.1816654116.

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." Science 329 (5993): 834–38. https://doi.org/10.1126/science.1184984.

Bidigare, Robert R., Arnim Fluegge, Katherine H. Freeman, Kristi L. Hanson, John M. Hayes, David Hollander, John P. Jasper, et al. 1997. "Consistent Fractionation of13 C in Nature and in the Laboratory: Growth-Rate Effects in Some Haptophyte Algae." Global Biogeochemical Cycles 11 (2): 279–92. https://doi.org/10.1029/96GB03939.

Boller, Amanda J., Phaedra J. Thomas, Colleen M. Cavanaugh, and Kathleen M. Scott. 2011. "Low Stable Carbon Isotope Fractionation by Coccolithophore

RubisCO." Geochimica et Cosmochimica Acta 75 (22): 7200–7207. https://doi.org/10.1016/j.gca.2011.08.031.

Boller, A J, P J Thomas, C M Cavanaugh, and K M Scott. 2015. "Isotopic Discrimination and Kinetic Parameters of RubisCO from the Marine Bloom-Forming Diatom, Skeletonema Costatum." Geobiology 13 (1): 33–43. https://doi.org/10.1111/gbi.12112.

Caemmerer, Susanne von, Youshi Tazoe, John R Evans, and Spencer M Whitney. 2014. "Exploiting Transplastomically Modified Rubisco to Rapidly Measure Natural Diversity in Its Carbon Isotope Discrimination Using Tuneable Diode Laser Spectroscopy." Journal of Experimental Botany 65 (13): 3759–67. https://doi.org/10.1093/jxb/eru036.

Cerling, Thure E, and John M Harris. 1999. "Carbon Isotope Fractionation between Diet and Bioapatite in Ungulate Mammals and Implications for Ecological and Paleoecological Studies." Oecologia 120 (3): 347–63. https://doi.org/10.1007/s004420050868.

Cerling, Thure E., John M. Harris, Bruce J. MacFadden, Meave G. Leakey, Jay Quade, Vera Eisenmann, and James R. Ehleringer. 1997. "Global Vegetation Change through the Miocene/Pliocene Boundary." Nature 389 (6647): 153–58. https://doi.org/10.1038/38229.

Christeller, J T, and W A Laing. 1976. "Isotope Discrimination by Ribulose 1,5-Diphosphate Carboxylase: No Effect of Temperature or HCO(3) Concentration." Plant Physiology 57 (4): 580–82.

Cleland, W.Wallace. 1987. "The Use of Isotope Effects in the Detailed Analysis of Catalytic Mechanisms of Enzymes." Bioorganic Chemistry 15 (3): 283–302. https://doi.org/10.1016/0045-2068(87)90026-5.

Davidi, Dan, Melina Shamshoum, Zhijun Guo, Yinon M Bar-On, Noam Prywes, Aia Oz, Jagoda Jablonska, et al. 2020. "Highly Active Rubiscos Discovered by Systematic Interrogation of Natural Sequence Diversity." The EMBO Journal 39 (18): e104081. https://doi.org/10.15252/embj.2019104081.

De Oliveira, André Luiz, Abhishek Srivastava, Salvador Espada-Hinojosa, and Monika Bright. 2022. "The Complete and Closed Genome of the Facultative Generalist Candidatus Endoriftia Persephone from Deep-Sea Hydrothermal Vents." Molecular Ecology Resources 22 (8): 3106–23. https://doi.org/10.1111/1755-0998.13668.

Ellis, R.John. 1979. "The Most Abundant Protein in the World." Trends in Biochemical Sciences 4 (11): 241–44. https://doi.org/10.1016/0968-0004(79)90212-3.

"Enzymatic Assay of Carbonic Anhydrase for Wilbur-Anderson Units (EC 4.2.1.1)." n.d. Accessed June 8, 2022. https://www.sigmaaldrich.com/US/en/technical-documents/protocol/protein-biology/enzyme-activity-assays/enzymatic-assay-of-carbonic-anhydrase.

Farquhar, G D, J R Ehleringer, and K T Hubick. 1989. "Carbon Isotope Discrimination and Photosynthesis." Annual Review of Plant Physiology and Plant Molecular Biology 40 (1): 503–37. https://doi.org/10.1146/annurev.pp.40.060189.002443.

Farquhar, G D, M H O'Leary, and J A Berry. 1982. "On the Relationship between Carbon Isotope Discrimination and the Intercellular Carbon Dioxide Concentration in Leaves." Australian Journal of Plant Physiology 9 (2): 121. https://doi.org/10.1071/PP9820121.

Field, C B, M J Behrenfeld, J T Randerson, and P Falkowski. 1998. "Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components." Science 281 (5374): 237–40. https://doi.org/10.1126/science.281.5374.237.

Fischer, Woodward W., James Hemp, and Jena E. Johnson. 2016. "Evolution of Oxygenic Photosynthesis." Annual Review of Earth and Planetary Sciences 44 (1): 647–83. https://doi.org/10.1146/annurev-earth-060313-054810.

Flamholz, Avi, and Patrick M Shih. 2020. "Cell Biology of Photosynthesis over Geologic Time." Current Biology 30 (10): R490–94. https://doi.org/10.1016/j.cub.2020.01.076.

Frey, Perry A., and Adrian D. Hegeman. 2007. Enzymatic Reaction Mechanisms. Oxford University Press. https://doi.org/10.1093/oso/9780195122589.001.0001.

Friedlingstein, Pierre, Michael O'Sullivan, Matthew W. Jones, Robbie M. Andrew, Luke Gregor, Judith Hauck, Corinne Le Quéré, et al. 2022. "Global Carbon Budget 2022." Earth System Science Data 14 (11): 4811–4900. https://doi.org/10.5194/essd-14-4811-2022.

Garcia, Amanda K, Colleen M Cavanaugh, and Betul Kacar. 2021. "The Curious Consistency of Carbon Biosignatures over Billions of Years of Earth-Life Coevolution." The ISME Journal 15 (8): 2183–94. https://doi.org/10.1038/s41396-021-00971-5.

Goddard, Thomas D, Conrad C Huang, Elaine C Meng, Eric F Pettersen, Gregory S Couch, John H Morris, and Thomas E Ferrin. 2018. "UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis." Protein Science 27 (1): 14–25. https://doi.org/10.1002/pro.3235.

Guy, R D, M L Fogel, and J A Berry. 1993. "Photosynthetic Fractionation of the Stable Isotopes of Oxygen and Carbon." Plant Physiology 101 (1): 37–47. https://doi.org/10.1104/pp.101.1.37.

Haslam, Richard P., Alfred J. Keys, P John Andralojc, Pippa J. Madgwick, Andersson Inger, Anette Grimsrud, Hans C. Eilertsen, and Martin A. J. Parry. 2005. "Specificity of Diatom Rubisco." In Plant Responses to Air Pollution and Global Change, edited by Kenji Omasa, Isamu Nouchi, and Luit J. De Kok, 157–64. Tokyo: Springer Japan. https://doi.org/10.1007/4-431-31014-2_18.

Hayes, J M. 1983. "Practice and Principles of Isotopic Measurements in Organic Geochemistry." Organic Geochemistry of Contemporaneous and Ancient Sediments 5.

———. 2001. "Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes." Reviews in Mineralogy and Geochemistry 43 (1): 225–77. https://doi.org/10.2138/gsrmg.43.1.225.

Higgins, John A, Andrei V Kurbatov, Nicole E Spaulding, Ed Brook, Douglas S Introne, Laura M Chimiak, Yuzhen Yan, Paul A Mayewski, and Michael L Bender. 2015. "Atmospheric Composition 1 Million Years Ago from Blue Ice in the Allan Hills, Antarctica." Proceedings of the National Academy of Sciences of the United States of America 112 (22): 6887–91. https://doi.org/10.1073/pnas.1420232112.

Horken, K M, and F R Tabita. 1999. "Closely Related Form I Ribulose Bisphosphate Carboxylase/Oxygenase Molecules That Possess Different CO2/O2 Substrate Specificities." Archives of Biochemistry and Biophysics 361 (2): 183–94. https://doi.org/10.1006/abbi.1998.0979.

Kane, H J, J Viil, B Entsch, K Paul, M K Morell, and T J Andrews. 1994. "An Improved Method for Measuring the CO2/O2 Specificity of Ribulosebisphosphate Carboxylase-Oxygenase." Functional Plant Biology 21 (4): 449. https://doi.org/10.1071/PP9940449.

Krissansen-Totton, J, R Buick, and D C Catling. 2015. "A Statistical Analysis of the Carbon Isotope Record from the Archean to Phanerozoic and Implications for the Rise of Oxygen." American Journal of Science 315 (4): 275–316. https://doi.org/10.2475/04.2015.01.

Lin, Myat T, William D Stone, Vishalsingh Chaudhari, and Maureen R Hanson. 2020. "Small Subunits Can Determine Enzyme Kinetics of Tobacco Rubisco Expressed in Escherichia Coli." Nature Plants 6 (10): 1289–99. https://doi.org/10.1038/s41477-020-00761-5.

Liu, Albert K., Jose H. Pereira, Alexander J. Kehl, Daniel J. Rosenberg, Douglas J. Orr, Simon K. S. Chu, Douglas M. Banda, et al. 2022. "Structural Plasticity Enables Evolution and Innovation of RuBisCO Assemblies." Sci. Adv. 8 (34). https://doi.org/10.1126/sciadv.adc9440.

Lloyd, Jon, and Graham D Farquhar. 1994. "13C Discrimination during CO2 Assimilation by the Terrestrial Biosphere." Oecologia 99 (3–4): 201–15. https://doi.org/10.1007/BF00627732.

Lorimer, G H, and T J Andrews. 1973. "Plant Photorespiration—an Inevitable Consequence of the Existence of Atmospheric Oxygen." Nature 243 (5406): 359–60. https://doi.org/10.1038/243359a0.

Marcus, Yehouda, Hagit Altman-Gueta, Aliza Finkler, and Michael Gurevitz. 2003. "Dual Role of Cysteine 172 in Redox Regulation of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase Activity and Degradation." Journal of Bacteriology 185 (5): 1509–17. https://doi.org/10.1128/jb.185.5.1509-1517.2003.

McNevin, Dennis B, Murray R Badger, Spencer M Whitney, Susanne von Caemmerer, Guillaume G B Tcherkez, and Graham D Farquhar. 2007. "Differences in Carbon Isotope Discrimination of Three Variants of D-Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Reflect Differences in Their Catalytic Mechanisms." The Journal of Biological Chemistry 282 (49): 36068–76. https://doi.org/10.1074/jbc.M706274200.

McNevin, D B, M R Badger, H J Kane, and G D Farquhar. 2006. "Measurement of (Carbon) Kinetic Isotope Effect by Rayleigh Fractionation Using Membrane Inlet Mass Spectrometry for CO2-Consuming Reactions." Functional Plant Biology : FPB 33 (12): 1115. https://doi.org/10.1071/FP06201.

Mook, W G, J C Bommerson, and W H Staverman. 1974. "Carbon Isotope Fractionation between Dissolved Bicarbonate and Gaseous Carbon Dioxide." Earth and Planetary Science Letters 22 (2): 169–76. https://doi.org/10.1016/0012-821X(74)90078-8.

O'Leary, M H. 1978. "Heavy Atom Isotope Effects in Enzyme-Catalyzed Reactions. In'Transition States of Biochemical Processes'.(Eds R. Gandour and RL Schowen.) Pp. 285-316.," 285.

Pettersen, Eric F, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, Tristan I Croll, John H Morris, and Thomas E Ferrin. 2021. "UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers." Protein Science 30: 70–82. https://doi.org/10.1002/pro.3943.

Read, B A, and F R Tabita. 1994. "High Substrate Specificity Factor Ribulose Bisphosphate Carboxylase/Oxygenase from Eukaryotic Marine Algae and Properties of Recombinant Cyanobacterial RubiSCO Containing 'Algal' Residue Modifications." Archives of Biochemistry and Biophysics 312 (1): 210–18. https://doi.org/10.1006/abbi.1994.1301.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (version 4.1.0). Computer software.

Robinson, Jonathan J., Kathleen M. Scott, S. Todd Swanson, Marion H. O'Leary, Kempton Horken, F. Robert Tabita, and Colleen M. Cavanaugh. 2003. "Kinetic Isotope Effect and Characterization of Form II RubisCO from the Chemoautotrophic Endosymbionts of the Hydrothermal Vent TubewormRiftia Pachyptila." Limnology and Oceanography Letters 48 (1): 48–54. https://doi.org/10.4319/lo.2003.48.1.0048.

Roeske, C A, and M H O'Leary. 1985. "Carbon Isotope Effect on Carboxylation of Ribulose Bisphosphate Catalyzed by Ribulosebisphosphate Carboxylase from Rhodospirillum Rubrum." Biochemistry 24 (7): 1603–7.

Roeske, C A, and Marion H. O'Leary. 1984. "Carbon Isotope Effects on Enzyme-Catalyzed Carboxylation of Ribulose Bisphosphate." Biochemistry 23 (25): 6275–84. https://doi.org/10.1021/bi00320a058.

Santrock, J, S A Studley, and J M Hayes. 1985. "Isotopic Analyses Based on the Mass Spectrum of Carbon Dioxide." Analytical Chemistry 57 (7): 1444–48. https://doi.org/10.1021/ac00284a060.

Saschenbrecker, Sandra, Andreas Bracher, Karnam Vasudeva Rao, Bharathi Vasudeva Rao, F Ulrich Hartl, and Manajit Hayer-Hartl. 2007. "Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco." Cell 129 (6): 1189–1200. https://doi.org/10.1016/j.cell.2007.04.025.

Schidlowski, Manfred. 1988. "A 3,800-Million-Year Isotopic Record of Life from Carbon in Sedimentary Rocks." Nature 333 (6171): 313–18. https://doi.org/10.1038/333313a0.

Schulz, Luca, Zhijun Guo, Jan Zarzycki, Wieland Steinchen, Jan M Schuller, Thomas Heimerl, Simone Prinz, Oliver Mueller-Cajar, Tobias J Erb, and Georg K A Hochberg. 2022. "Evolution of Increased Complexity and Specificity at the Dawn of Form I Rubiscos." Science 378 (6616): 155–60. https://doi.org/10.1126/science.abq1416.

Scott, Kathleen M, Julie Schwedock, Daniel P Schrag, and Colleen M Cavanaugh. 2004. "Influence of Form IA RubisCO and Environmental Dissolved Inorganic Carbon on the Delta13C of the Clam-Chemoautotroph Symbiosis Solemya Velum." Environmental Microbiology 6 (12): 1210–19. https://doi.org/10.1111/j.1462-2920.2004.00642.x.

Scott, K M, M Henn-Sax, T L Harmer, D L Longo, C H Frame, and C M Cavanaugh. 2007. "Kinetic Isotope Effect and Biochemical Characterization of Form IA RubisCO from the Marine Cyanobacterium Prochlorococcus Marinus MIT9313." Limnology and Oceanography Letters 52 (5): 2199–2204. https://doi.org/10.4319/lo.2007.52.5.2199.

Sharkey, Thomas D., and Joseph A. Berry. 1985. "Carbon Isotope Fractionation of Algae as Influenced by an Inducible CO2 Concentrating Mechanism." In Inorganic Carbon Uptake by Aquatic Photosynthetic Organisms, edited by W J Lucas and J A Berry, 389–401. The American Society of Plant Physiologists.

Sheldon, Nathan D. 2006. "Precambrian Paleosols and Atmospheric CO2 Levels." Precambrian Research 147 (1–2): 148–55. https://doi.org/10.1016/j.precamres.2006.02.004.

Shih, Patrick M, Alessandro Occhialini, Jeffrey C Cameron, P John Andralojc, Martin A J Parry, and Cheryl A Kerfeld. 2016. "Biochemical Characterization of Predicted Precambrian RuBisCO." Nature Communications 7 (January): 10382. https://doi.org/10.1038/ncomms10382.

Spreitzer, Robert J, and Michael E Salvucci. 2002. "Rubisco: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme." Annual Review of Plant Biology 53: 449–75. https://doi.org/10.1146/annurev.arplant.53.100301.135233.

Spreitzer, Robert J. 2003. "Role of the Small Subunit in Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase." Archives of Biochemistry and Biophysics 414 (2): 141–49. https://doi.org/10.1016/S0003-9861(03)00171-1.

Tcherkez, Guillaume G B, Graham D Farquhar, and T John Andrews. 2006. "Despite Slow Catalysis and Confused Substrate Specificity, All Ribulose Bisphosphate Carboxylases May Be Nearly Perfectly Optimized." Proceedings

of the National Academy of Sciences of the United States of America 103 (19): 7246–51. https://doi.org/10.1073/pnas.0600605103.

Thomas, Phaedra J, Amanda J Boller, Sriram Satagopan, F Robert Tabita, Colleen M Cavanaugh, and Kathleen M Scott. 2018. "Isotope Discrimination by Form IC RubisCO from Ralstonia Eutropha and Rhodobacter Sphaeroides, Metabolically Versatile Members of 'Proteobacteria' from Aquatic and Soil Habitats." Environmental Microbiology, September. https://doi.org/10.1111/1462-2920.14423.

Wang, Renée Z, Robert J Nichols, Albert K Liu, Avi I Flamholz, Juliana Artier, Doug M Banda, David F Savage, John M Eiler, Patrick M Shih, and Woodward W Fischer. 2023. "Carbon Isotope Fractionation by an Ancestral Rubisco Suggests That Biological Proxies for CO2 through Geologic Time Should Be Reevaluated." Proceedings of the National Academy of Sciences of the United States of America 120 (20): e2300466120. https://doi.org/10.1073/pnas.2300466120.

West-Roberts, Jacob A., Paula B. Matheus-Carnevali, Marie Charlotte Schoelmerich, Basem Al-Shayeb, Alex D. Thomas, Allison Sharrar, Christine He, et al. 2021. "The Chloroflexi Supergroup Is Metabolically Diverse and Representatives Have Novel Genes for Non-Photosynthesis Based CO2 Fixation." BioRxiv, August. https://doi.org/10.1101/2021.08.23.457424.

Westheimer, F H. 1961. "The Magnitude of the Primary Kinetic Isotope Effect for Compounds of Hydrogen and Deuterium." Chemical Reviews 61 (3): 265–73. https://doi.org/10.1021/cr60211a004.

Whelan, Tom, W M Sackett, and C R Benedict. 1973. "Enzymatic Fractionation of Carbon Isotopes by Phosphoenolpyruvate Carboxylase from C4 Plants." Plant Physiology 51 (6): 1051–54. https://doi.org/10.1104/pp.51.6.1051.

Wickham, H, W Chang, and M H Wickham. 2016. "Package 'Ggplot2'." Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2 (1): 1–189.

Wilkes, Elise B., and Ann Pearson. 2019. "A General Model for Carbon Isotopes in Red-Lineage Phytoplankton: Interplay between Unidirectional Processes and Fractionation by RubisCO." Geochimica et Cosmochimica Acta, September. https://doi.org/10.1016/j.gca.2019.08.043.

Witkowski, Caitlyn R, Johan W H Weijers, Brian Blais, Stefan Schouten, and Jaap S Sinninghe Damsté. 2018. "Molecular Fossils from Phytoplankton Reveal

Secular Pco2 Trend over the Phanerozoic." Science Advances 4 (11): eaat4556. https://doi.org/10.1126/sciadv.aat4556.

Zeebe, R E, and D Wolf-Gladrow. 2001. CO2 in Seawater: Equilibrium, Kinetics, Isotopes. Vol. 65. Elsevier Oceanography Series. Elsevier. https://doi.org/10.1016/S0422-9894(01)X8001-X.

4. Widespread detoxifying NO reductases impart a distinct isotopic fingerprint on N₂O under anoxia

4.1 Abstract

Nitrous oxide (N₂O), a potent greenhouse gas, can be generated by compositionally complex microbial populations in diverse contexts. Accurately tracking the dominant biological sources of N₂O has the potential to improve our understanding of N₂O fluxes from soils as well as inform the diagnosis of human infections. Isotopic "Site Preference" (SP) values have been used towards this end, as bacterial and fungal nitric oxide reductases produce N₂O with different isotopic fingerprints. Here we show that flavohemoglobin, a hitherto biogeochemically neglected yet widely distributed detoxifying bacterial NO reductase, imparts a distinct SP value onto N₂O under anoxic conditions that correlates with typical environmental N₂O SP measurements. We suggest a new framework to guide the attribution of N₂O biological sources in nature and disease.

4.2 Introduction

Nitrous oxide (N₂O) is a ubiquitous metabolite present in myriad environments ranging from soils, marine and freshwater systems, and the atmosphere to the human body. Because N₂O can be produced and consumed by multiple microbial nitrogen-cycling processes (Kuypers et al. 2018), tracking its fate is challenging. One motivation to do so springs from the fact that N₂O is a potent greenhouse gas, whose current atmospheric concentration is more than 20% compared to preindustrial levels (Tian et al. 2020); a better understanding of N₂O sources could help facilitate mitigation efforts. Analogously, because N₂O has been measured in chronic pulmonary infections (Kolpen et al. 2014), clarity on which pathogens are metabolically active in disease contexts could inform treatment strategies (Cook et al. 2014).

An intramolecular isotopic fingerprint called "Site Preference" (SP), which measures the relative enrichment of natural abundance ^{15}N in the central (α) versus terminal (β) nitrogen position in N₂O (Fig. 4.1A; (Toyoda and Yoshida 1999)) may be applied for such purposes. Unlike traditional natural abundance isotopic measurements that measure the total enrichment of ^{15}N in the bulk molecule (Denk et al. 2017), SP does not rely on the isotopic composition of the source substrate but instead reflects the reaction mechanism (Wang et al. 2004), making it a potentially powerful tool to disentangle N₂O sources in different contexts.

The median values of *in situ* SP measurements where microbes are present are 10.9, 20.9 and 23.0 per mille (‰) for soils, marine and freshwater systems, respectively (Fig. 4.1A). These values are bounded by the median values of *in vitro*, pure culture studies of N₂O-producing biogenic end-members like bacterial and fungal denitrifiers as well as ammonia-oxidizing bacteria (AOB; Fig. 4.1A). Bacterial and fungal denitrifiers are thought to represent two extremes of SP values

for N$_2$O producers with median SP values of -4.3 and 32.2‰ respectively (Fig. 4.1A), which are assumed to reflect the activity of dissimilatory Nitric Oxide Reductases (NOR); in AOBs, the SP varies between roughly -11 and 36‰ due to multiple dissimilatory N$_2$O formation pathways (Frame and Casciotti 2010). Because the vast majority of *in situ* observations lie between end-member values for bacterial and fungal NORs and AOBs, the SP values of biogenic N$_2$O produced in the environment has been rationalized by mixing biogenic end-members.
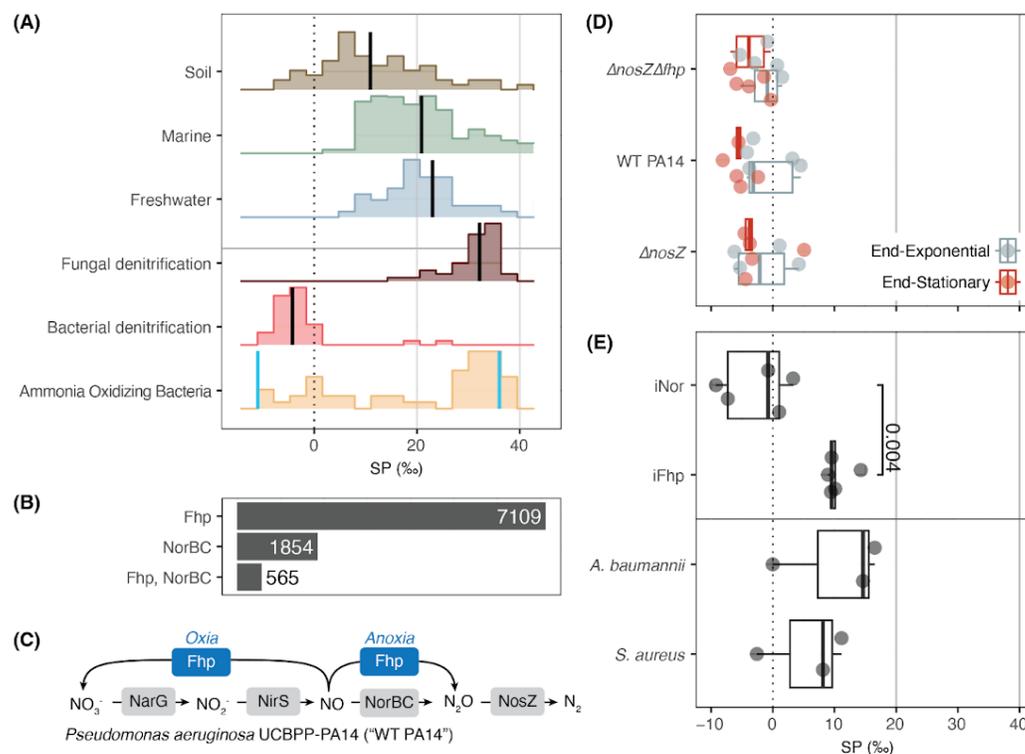
Fig. 4.1. N$_2$O production via NO detoxification under anoxic conditions may explain environmental SP values.
(A) Measured *in situ* SP values for environmental (Soil, Marine, Freshwater) vs. in vitro measurements of biogenic end-members (Bacterial and Fungal Denitrification, Ammonia Oxidizing Bacteria (AOB)); black line shows median; blue lines show end-member values for AOB (Frame and Casciotti 2010). Histogram height is normalized to each category; see Fig. 4.18 for outlier values and more detail. **(B)** Number of bacterial genomes hits at the phylum level for flavohemoglobin protein (Fhp) and nitrous oxide reductase (NorBC) alone or in combination from Annotree (Mendler et al. 2019); minimum amino acid sequence similarity of 30% was used. See Fig. 4.21, Tables 4.12-4.14 for phylogenetic distribution. **(C)** Relevant N-oxide pathways of *Pseudomonas aeruginosa* UCBPP-PA14 (*Pa*), the model organism used in this study. *Pa* possesses the full denitrification pathway as well as Fhp. **(D)** SP of N$_2$O produced by *Pa* and mutant strains with *fhp* and/or *nosZ* genes deleted (*ΔnosZΔfhp*; *ΔnosZ*) in denitrifying conditions; see Fig. S2 for more detail. **(E)** SP of *Pa* strains with rhamnose-induced expression of *norBC* (iNOR) or *fhp* (iFhp) alone as well as *Acinetobacter baumannii* and *Staphylococcus aureus*, which only have Fhp. *P* value was calculated via Welch's t-test. Each data point represents an individual biological replicate in (D) and (E).

Current practices for interpreting SP measurements in natural environments focus on catabolic pathways and assume that N_2O production or consumption is tied to microbial growth. However, an entire other class of enzymes exists that produce N_2O as a consequence of nitric oxide (NO) detoxification and not for energy-conservation (Ferousi et al. 2020). Flavohemoglobin proteins (e.g. Fhp/Hmp/Yhb–henceforth referred to as "Fhp") are phylogenetically widespread and protect against nitrosative stress in bacteria and yeast (Poole and Hughes 2000). Members of this family are roughly four times more abundant than NORs in annotated bacterial genomes (Fig. 4.1B, Fig. 4.17, Tables 4.12-4.14; 7109 vs. 1854 genome hits at the phylum level for Fhp vs. NorBC using 30% minimum amino acid sequence similarity (Mendler et al. 2019)). While their ability to oxidize NO to nitrate (NO_3^-) under oxic conditions is well known, their ability to reduce NO to N_2O under anoxic conditions has received less attention (Bonamore and Boffi 2008; Poole and Hughes 2000). Given that bacterial denitrifiers commonly possess both Fhp and NOR (Fig. 4.1B and Table 4.14), we hypothesized that Fhp might play a role in N_2O emissions and set out to determine whether it imparts a SP onto N_2O distinct from that of bacterial or fungal NORs.

4.3 Overall SP values reflect NOR during denitrification

To compare the SP of Fhp to NOR in a whole cell context (*in vivo*), we used the model bacterial denitrifier, *Pseudomonas aeruginosa* UCBPP-PA14 (*Pa*, Fig. 4.1C). Because this organism is genetically tractable, it provides a means to study the cellular processes of interest in a controlled way. To determine SP values under denitrifying conditions, *Pa*, *ΔnosZ* and *ΔnosZΔfhp* – strains with deletions of the nitrous oxide reductase (NOS) gene, *nosZ* (*PA14_20200*) and/or *fhp* (*PA14_29640*; Table 4.1) – were grown anaerobically in defined medium batch cultures and sampled at late exponential and late stationary growth phase (Table 4.2, Fig. 4.5 and Methods). N_2O was cryogenically distilled and analyzed for nitrogen and oxygen isotopes on the Thermo Scientific Ultra High-Resolution Isotope Ratio Mass Spectrometer (HR-IRMS; (Eiler et al. 2013); Methods). All isotope data is reported in the delta (δ) notation in units of per mille (‰) where $\delta^{15}N = [(^{15}N/^{14}N)_{sample} / (^{15}N/^{14}N)_{reference} - 1]*1000$ and $SP = \delta^{15}N^\alpha - \delta^{15}N^\beta$. Values are reported relative to the international reference of AIR for nitrogen; see Methods for more detail.

Name	Strain Description	Fhp?	Nor?	Source
WT *Pa*	Wild-type *Pseudomonas aeruginosa* UCBPP-PA14	Yes	Yes	Lab Collection
ΔnosZ	Deletion of nitrous oxide reductase gene (*nosZ*, *PA14_20200*) from WT *Pa*	Yes	Yes	This study
ΔnosZΔfhp	Deletion of *nosZ* and flavohemoglobin protein (*fhp*, *PA14_29640*) from WT *Pa*	Yes	No	This study
iFhp	Rhamnose-induced expression of *fhp* integrated into the chromosome of WT *Pa* with deletion of native *norBC*, *fhp*, and *nosZ*.	Yes	No	This study
iNOR	Rhamnose-induced expression of the nitric oxide reductase operon, *norBCD* (*PA14_16810*, *PA14_16830*, *PA14_06840*), integrated into the *att* neutral chromosomal site of *Pa* with deletion of native nitrate reductase (*narGHJI*; *PA14_13780-13830)*, nitrite reductase (*nirS*; *PA14_06750*), *norBC*, *nosZ*, and *fhp*.	No	Yes	This study
S. aureus	Wild-type *Staphylococcus aureus* USA300 LAC	Yes	No	Gift
A. baumannii	Wild-type *Acinetobacter baumannii* ATCC 17978	Yes	No	Gift

Table 4.1. Strains studied.
The SP of N_2O produced by five strains of *Pseudomonas aeruginosa* (WT *Pa*, *ΔnosZ*, *ΔnosZΔfhp*, *iFhp*, *iNOR*) and two wild-type strains of *Staphylococcus aureus* and *Acinetobacter baumannii* were measured. See Materials and Methods for further detail. *S. aureus* and *A. baumannii* were both kindly provided by Eric Skaar, Vanderbilt University Medical Center.

The SP of *ΔnosZΔfhp* should only reflect NOR, since all other known pathways for N_2O production and consumption were deleted. The *in vivo* SP of this strain did not vary significantly by growth phase (Welch's t-test, *P*=0.2), and its average value across all growth phases (-2.53 ± 2.90, n = 10) was consistent with prior *in vitro* measurements of NOR purified from *Paracoccus denitrificans* ATCC 35512 (-5.9 ± 2.1‰, (Yamazaki et al. 2014)). The SP of the *ΔnosZΔnorBC* strain, which only has *fhp*, was not measured because it did not grow appreciably under these conditions (Fig. 4.5) presumably due to growth suppression when NO build-up is too high (Wilbert and Newman 2022; Yoon et al. 2002).

Interestingly, WT *Pa*, which can produce N_2O through both Fhp and NOR (Fig. 4.1C), displayed SP values that did not vary significantly from ΔnosZΔfhp across all growth phases (*P*=0.7). In addition, the SP of WT *Pa* did not vary significantly by growth phase (*P*=0.07). The SP of *ΔnosZ* was also measured because prior studies showed that NOS can increase the SP of the residual N_2O pool through preferential cleavage of the $^{14}N\text{-}O$ vs. $^{15}N\text{-}O$ bond in N_2O (Casciotti et al. 2018; Ostrom et al. 2007); however, SP values of *ΔnosZ* were similar to *ΔnosZΔfhp* (*P*=0.7) and did not vary by growth phase (*P*=0.8; Fig. 4.1D). Therefore, even though Fhp was likely present in all previously measured bacterial denitrifier strains for in vitro measurements (Table 4.15), it does not affect the overall SP value

when strains are grown under denitrifying conditions, suggesting that NOR dominates the isotopic signature under these conditions. However, the potential for Fhp to impact the SP of N_2O under other conditions remained open.

4.4 Fhp has an intermediate, positive SP value compared to NOR

To distinguish the SP of Fhp and NOR, we engineered two Pa strains possessing only Fhp ("iFhp") or NOR ("iNOR") that could be induced in the presence of rhamnose; inducible Fhp and NOR functionality was validated by complementation experiments (Table 4.1, Fig. 4.4). Since these strains lack denitrification enzymes and are incapable of anaerobic growth, suspension assays were developed to culture bacteria aerobically while inducing gene expression prior to placement in non-growing, anoxic conditions. Strains were provided exogenous NO via the small molecule donor DETA NONOate ($C_4H_{13}N_5O_2$) at sub-toxic concentrations (Figure 4.22). Strains were incubated under anoxic conditions for 24 hours at 37°C before the headspace was sampled; see Table 4.2 and Methods for more detail.

Strain	Assay Type	Aerobic pre-growth	Anaerobic incubation	SP (‰)	n
iNOR	Suspension	100 mM nitrate	100 mM nitrate, 500 μM DETA NONOate, 305 μM rhamnose	-2.60 ± 5.41	5
iFhp	Suspension	100 mM nitrate	100 mM nitrate, 500 μM DETA NONOate, 305 μM rhamnose	10.45 ± 2.17	5
A. baumannii	Suspension	100 mM nitrate	100 mM nitrate, 500 μM DETA NONOate	10.38 ± 9.05	3
S. aureus	Suspension	100 mM nitrate	100 mM nitrate, 500 μM DETA NONOate	5.56 ± 7.21	3
ΔnosZ	Batch; End-exponential	100 mM nitrate	100 mM nitrate	-1.56 ± 5.04	4
	Batch; End-stationary	100 mM nitrate	100 mM nitrate	-2.21 ± 4.10	5
ΔnosZΔfhp	Batch; End-exponential	100 mM nitrate	100 mM nitrate	-1.39 ± 2.78	5
	Batch; End-stationary	100 mM nitrate	100 mM nitrate	-3.68 ± 2.81	5
WT *Pa*	Batch; End-exponential	100 mM nitrate	100 mM nitrate	-0.70 ± 4.19	5
	Batch; End-stationary	100 mM nitrate	100 mM nitrate	-5.43 ± 2.04	5
	Suspension	100 μM DETA NONOate	500 μM DETA NONOate	-2.59 ± 7.53	2
	Suspension	100 μM DETA NONOate + 100 mM nitrate	500 μM DETA NONOate	9.14 ± 3.70	2
	Suspension	100 mM nitrate	500 μM DETA NONOate + 100 mM nitrate	2.61 ± 9.31	5
	Batch; End-stationary	100 mM nitrate	500 μM DETA NONOate + 100 mM nitrate	-3.34 ± 0.83	2

Table 4.2. Culturing conditions and SP results.
All strains were grown in aerobic pre-growths before being resuspended in fresh media and anoxically incubated for headspace sampling as batch culture or suspension assays (Fig. 4.7); nitrate and/or DETA NONOate ($C_4H_{13}N_5O_2$) was supplemented to provide endogenous vs. exogenous NO respectively. See Methods for more detail. SP values (mean ± s.d.) of n biological replicates; see Supplemental for full data table.

Under these conditions, iFhp displayed SP values (10.45 ± 2.17, *n*=5) that were significantly more positive than iNOR (-2.60 ± 5.41, *n*=5; *P*=0.004; Fig. 4.1E). To validate these SP values outside Pa, two wild-type, non-denitrifying strains with only Fhp, *Staphylococcus aureus* USA300 LAC and *Acinetobacter baumannii* ATCC 17978 were also measured. Fhp from *S. aureus* shows 31.6% amino acid sequence similarity to Fhp from *P. aeruginosa*, while Fhp from *A. baumannii* shows 98.5% similarity. However, all Fhps share a common catalytic site for NO binding and reduction, a globin module with heme B (Poole and Hughes 2000), that is responsible for imparting the observed SP. The SP of *S. aureus* (5.56 ± 7.21‰, *n*=3) and *A. baumannii* (10.38 ± 9.05‰, *n*=3) were both positive and statistically indistinguishable from *Pa* iFhp (Fig. 4.1E).

iNOR values were consistent with both prior *in vitro* NOR measurements (Yamazaki et al. 2014) and our *ΔnosZΔfhp* measurements. We note that there is large variation between the biological replicates measured by (Yamazaki et al. 2014) (-5 and -9‰; *n* = 2) that is similar to the variation of NOR SP in this study . This variation neither correlates with degree of NO consumption nor N_2O production (Fig. 4.16), indicating that there may be inherent variation in SP for NOR on the order of 10‰.

4.5 Exogenous NO shifts SP values toward Fhp

Given the potential for Fhp to impart a positive SP distinct from NOR, we next sought to identify physiological conditions where it might dominate the N_2O isotopic fingerprint in the wild-type. Historically, N_2O isotopic measurements from pure cultures have been made for actively growing cells, which would be expected to amplify isotopic signatures imparted by catabolic enzymes. Yet evidence is mounting that slow, survival physiology dominates microbial existence in diverse habitats (Bodor et al. 2020; Bergkessel et al. 2016), motivating N_2O SP measurement during non-growth conditions.
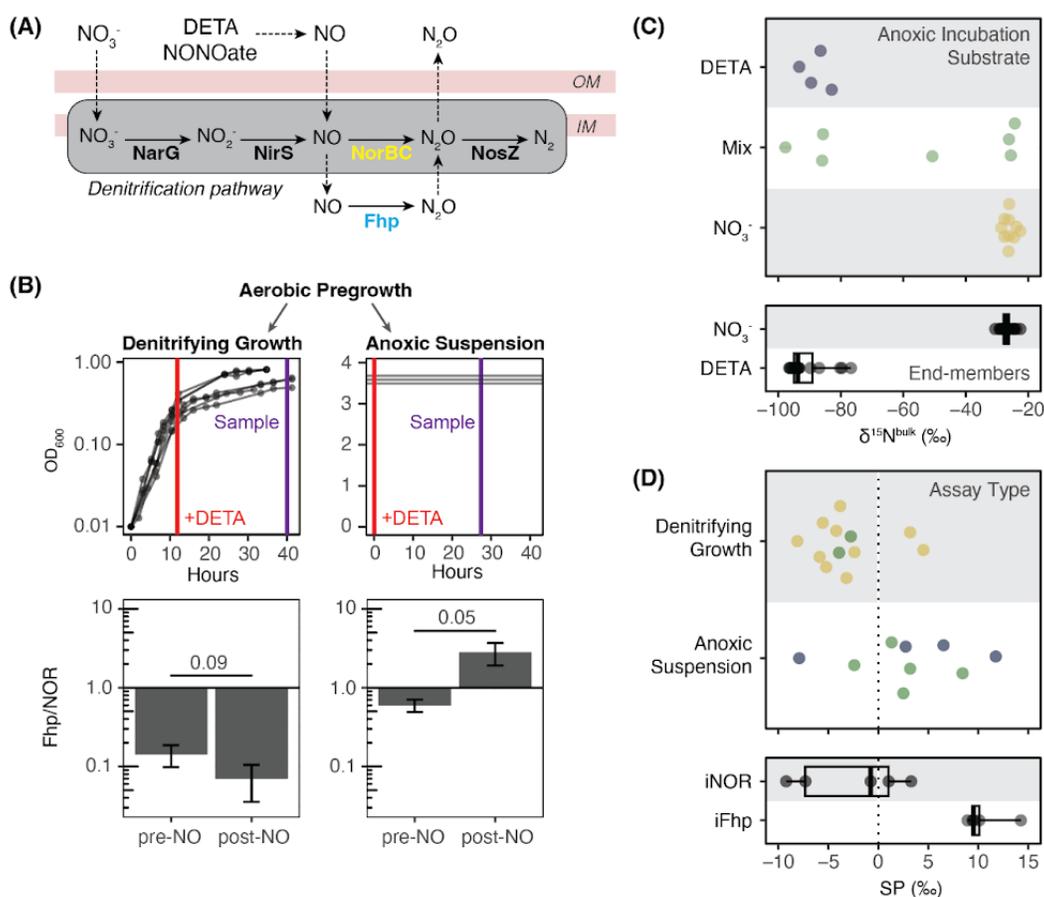
Fig. 4.2: High concentrations of NO shift SP values towards Fhp.
(A) In *Pa*, NorBC contributes to overall cell energetics as part of the denitrification pathway; Fhp does not and is primarily used for NO detoxification. **(B)** WT *Pa* was cultured anaerobically via two assay types after aerobic pre-growth with nitrate to either maximize growth via denitrification (left) or be re-suspended as non-growing cells (right). Exogenous NO was supplied through DETA NONOate (red lines) and headspace was then sampled for SP analysis (purple lines). Culture aliquots for proteomics analysis were taken immediately prior to NO addition ("pre-") or during the same time as headspace sampling ("post-NO"). Ratio of Fhp to NOR in these conditions are shown as bar charts below; see Fig. 4.19 for full results. *P* values were calculated via Welch's t-test. **(C)** $\delta^{15}N^{bulk}$ values for WT *Pa* incubated anoxically with DETA (blue), nitrate (yellow) or both (green); end-member values are from non-WT *Pa* strains incubated with only nitrate or DETA (Fig. 4.11). **(D)** SP measurements for WT *Pa* grown as denitrifying growths or anoxic suspensions, as illustrated in (B). Colors indicate anoxic incubation substrate and are the same as panel (C). iNOR and iFhp SP values are from Fig. 4.1E. For (C, D), box plots indicate median, upper and lower quartiles, and extreme values.

To test if *Pa* can produce positive SP values indicative of Fhp activity, we grew WT *Pa* in denitrifying batch cultures and non-growing, anoxic suspensions with varying combinations of nitrate (NO_3^-) and DETA NONOate to provide NO endogenously via denitrification (Fig 4.1C) and/or exogenously via small molecule-mediated NO release, which we hypothesized would promote NOR or Fhp activity, respectively. We validated the induction of NOR and Fhp using

quantitative unlabeled proteomics (Methods) and calculated the ratios of Fhp to NOR to quantify relative changes of each NO reductase. In denitrifying, batch culture conditions (Fig. 4.2B), the ratio of Fhp to NOR is less than one (~0.25) and does not significantly change upon addition of NO (P=0.09; Fig. 4.2B). By contrast, NorB, which contains the catalytic subunit of NOR, is undetectable before NO addition in the suspension assays (Fig. 4.19), which transitions from oxic pre-growth to nongrowing, anoxic conditions. Although NorB increases to detectable levels upon the addition of DETA NONOate (Fig. 4.19), Fhp is far more abundant, leading to a high ratio of Fhp to NOR (~3, Fig. 4.2C).

SP data (Fig. 4.2D) is consistent with denitrifying batch cultures favoring NOR production, and non-growing, anoxic suspension assays favoring Fhp. When WT *Pa* is grown under denitrifying conditions, SP values are more negative and within the range of iNOR. However, in suspension assays, SP values span the range from iNOR to iFhp, consistent with increased Fhp abundance in these conditions. The most positive SP values (yellow circles, Fig. 4.2D) that are within the range of iFhp are seen when *Pa* is given a high dose of both endogenous and exogenous NO following oxic pre-growth (NO_3^- and DETA NONOate).

Paired SP and $\delta^{15}N^{bulk}$ data allowed us to track which pool of NO was used by Fhp or NOR for N_2O production. When N-oxides are reduced to N_2O, $\delta^{15}N^{bulk}$ retains the isotopic signature of the original N (Sigman et al. 2001); because our NO_3^- and DETA NONOate had distinct $\delta^{15}N$ values (0.40 ± 1.28 vs. -22.95 ± 0.15‰ respectively), we could clearly distinguish between the NO source. When WT *Pa* is incubated anoxically with either NO_3^- or DETA NONOate, $\delta^{15}N^{bulk}$ values correspondingly show only one NO source (Fig. 4.2E); when given both substrates simultaneously, N_2O can be made from varying ratios of both exogenous and endogenous NO. However, the most positive SP values in WT *Pa* correspond with an exogenous NO source (yellow circles in Fig. 4.2D; Fig. 4.9) indicating that Fhp preferentially utilizes exogenous NO.

4.6 Consequences for interpreting existing SP data

Given that Fhp homologs are present in many denitrifying bacteria and AOB (Fig. 4.21, Table 4.16), it is possible that Fhp may have contributed to the SP values measured in previous pure culture studies. Notably, all prior reports of SP from bacterial denitrifiers used strains that also have Fhp (Table 4.15); given the sensitivity of enzyme abundance to the physiological state during the time of measurement, it is plausible that the positive spread in SP values observed in these studies (Toyoda et al. 2005) may reflect cryptic Fhp activity. An Fhp homologue, Yhb, exists in yeast (Poole and Hughes 2000) and is present in previously studied fungal denitrifiers as well, possibly contributing to the tail towards 10‰ observed from the literature (Fig. 4.1A).

Fhp is phylogenetically widespread and more abundant than NOR; therefore, measuring Fhp values from a representative group of diverse bacteria may illuminate the natural variation in SP values. In addition, measuring other NO-

detoxifying proteins may shed further light on the SP values of this neglected class of non-catabolic enzymes. Flavo-diiron proteins, which only operate in anoxic conditions and only reduce NO to N_2O for detoxification (Ferousi et al. 2020) present an attractive next target for SP measurements. Finally, further detailed studies of Fhp's reaction mechanism paired with SP values may help shed light on broader questions of what determines the SP of N_2O through NO reduction, for both abiotic and biotic reactions (Stanton et al. 2018; Wang et al. 2004; Yeung 2016; Schmidt et al. 2004).

Beyond helping to explain the N_2O SP variation seen in prior pure-culture studies, our finding that Fhp produces an intermediate SP value that overlaps with many natural SP measurements begs the question: How can we distinguish Fhp-generated N_2O from that produced by a mixture of other enzyme sources in complex environments such as soils or infected tissues? This is a difficult task. Though we can infer whether certain enzymes may be present and active based on knowledge of what regulates their expression, in order to predict whether they are active in any given sample, we need to know the conditions experienced by cells in situ. For example, our work indicates that Fhp dominates the SP fingerprint when cells grown under oxic conditions subsequently encounter a concentrated pulse of NO under anoxia. Intriguingly, pulses of NO and N_2O have been detected after wetting of dryland soils (Krichels et al. 2022; Homyak et al. 2016) and opportunistic pathogens are thought to experience NO bursts from different cell types in the human immune system (Kolpen et al. 2014). Yet to speculate on whether such pulses may trigger Fhp activity, we would need to be able to track NO and oxygen concentrations at the microscale – measurements that are challenging to make. Ultimately, knowledge of the relative abundance of NO reductases present in any given sample where N_2O SP is measured will be necessary to attribute sources with confidence.

4.7 Materials and Methods

4.7.1 Media and nitric oxide donors

Media was optimized to increase cell and N_2O yields. WT *P. aeruginosa* (*Pa*) was first grown in synthetic cystic fibrosis media ("Base SCFM") (Palmer et al. 2007) with varying concentrations of potassium nitrate (KNO_3; 20 to 100 mM) to test if high levels of nitrate were toxic but did not see any significant growth defects or benefits (Fig. 4.3), suggesting that other media components were limiting growth. Therefore, Base SCFM was then amended with 20 mM sodium succinate and trace metals to increase cell and N_2O yields ("SCFM Amended" or SCFM-A; Fig. 4.3). A 1000x solution of the trace metal stock (Trace element sol. SL-10; DSMZ) at a total volume of 1000 mL was comprised of: 1) 10.00 mL of HCl (25%; 7.7 M); 2) 1.50 g of $FeCl_2$ x 4 H_2O; 3) 70.00 mg of $ZnCl_2$; 4) 100.00 mg of $MnCl_2$ x 4 H_2O; 5) 6.00 mg of H_3BO_3; 5) 190.00 mg of $CoCl_2$ x 6 H_2O; 6) 2.00 mg of $CuCl_2$ x 2 H_2O; 7) 24.00 mg of $NiCl_2$ x 6 H_2O; 8) 36.00 mg of Na_2MoO_4 x 2 H_2O;

9) 990.00 mL of distilled water. All strains in this study were grown in SCFM-A media. The small molecule NO donor DETA NONOate ($C_4H_{13}N_5O_2$, #82120 Cayman Chemical Company) was used in certain experiments. It decays following first order kinetics in a pH-dependent manner to release two moles of NO per mole of DETA NONOate (half-life of 20 hours at 37°C and pH 7.4).
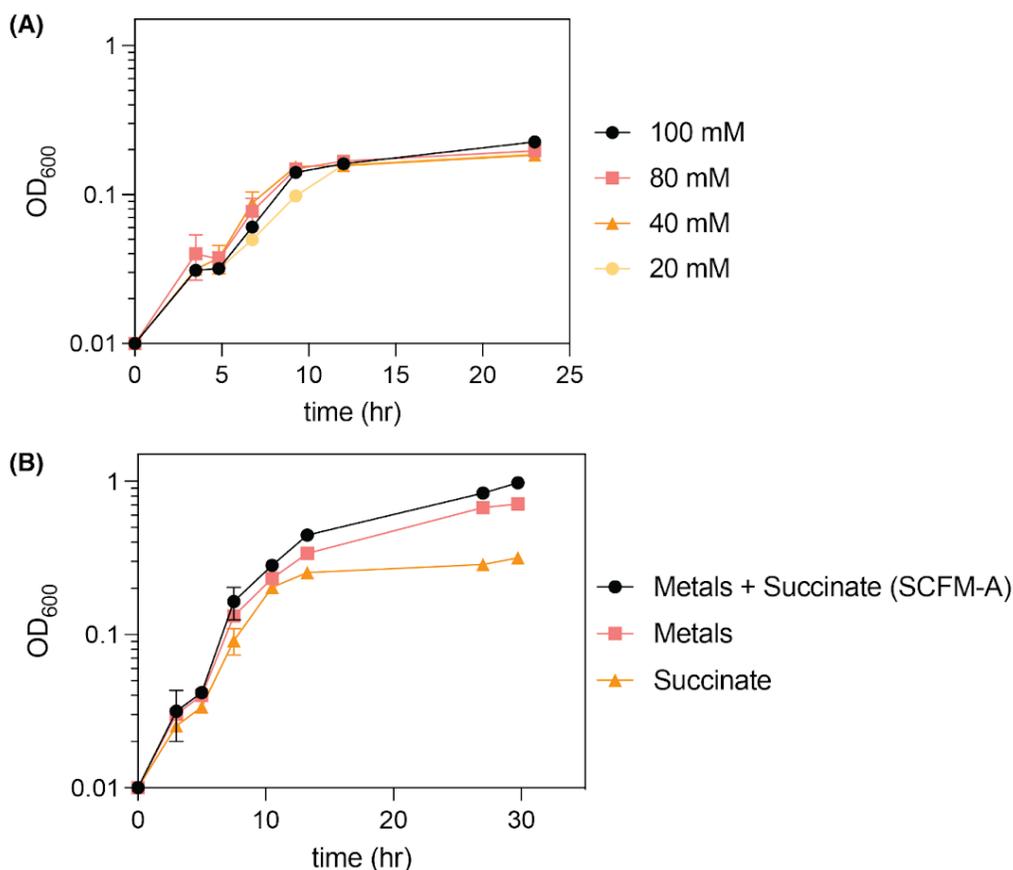

Fig. 4.3. Media optimization.
(A) Growth curves of WT *P. aeruginosa* UCBPP-PA14 grown in Base SCFM Media (1) in the vacuum sampling flasks with varying concentrations of nitrate. OD$_{600}$ is optical density at 600 nm.
(B) Growth curves of WT PA14 in Base SCFM with trace metal amendments, 20 mM sodium succinate, or both. The final media, "SCFM Amended," has both trace metal and succinate amendment.

4.7.2 Strain generation

We measured the SP of N$_2$O produced by five strains of *Pa*, and two wild-type strains of *Staphylococcus aureus* and *Acinetobacter baumannii* (Table 4.1).

Pseudomonas aeruginosa UCBPP-PA14 was the wild-type (WT) and parent strain of all genetic manipulations done in this study. Individual and combinatory mutants of *Pa* nitrate reductase (*ΔnarGHJI*; *PA14_13780-13830)*, nitrite reductase (*ΔnirS*; *PA14_06750*, nitric oxide reductase (*ΔnorBC*;

PA14_16810, *PA14_16830*) and nitrous oxide reductase (*ΔnosZ*, *PA14_20200*) were generated previously (Wilbert and Newman 2022). *ΔnosZΔfhp* has the additional deletion of *fhp*, the flavohemoglobin protein / nitric oxide dioxygenase (*PA14_29640*). Clean deletions were done using allelic exchange as previously described (Spero and Newman 2018); briefly ~1 kb fragments surrounding the gene of interest were amplified by PCR and Gibson cloned into pMQ30 (Gibson et al. 2009). Deletion constructs were introduced into *Pa* via triparental conjugation, and *E. coli* plasmid and helper strains were selected against on VBMM containing 50 µg/ml gentamicin (Choi and Schweizer 2006). Resulting Gent^R *Pa* cells were plated on 10% sucrose LB agar to isolate recombinants and screened via PCR. See Table 4.3 for primers used. Another strain, *ΔnosZΔnorBC*, was also used but it did not grow appreciably in the anaerobic, batch culture growth condition (Fig. 4.5); therefore its SP was not measured.

Two strains with inducible expression were created to increase N_2O production amounts for isotopic measurement (Fig. 4.4). Strains with inducible *fhp* ('iFhp,' to denote *P. aeruginosa ΔnosZΔfhpΔnor att*::*mTn7*(GentR,*fhp*)) and *norBCD* ('iNOR,' to denote *P. aeruginosa ΔnarΔnirΔnorΔnosZΔfhp att*::*mTn7*(GentR,*norBCD*)) were generated by, first, amplifying *fhp* or *norCBD* from *P. aeruginosa* genomic DNA. See Table 4.3 for primers used. PCR products were ligated into plasmid the miniTn7 plasmid pJM220 (Choi and Schweizer 2006) via Gibson cloning (Gibson et al. 2009) 3' of the *rhaB* promoter for rhamnose-specific expression. Plasmids were delivered to *P. aeruginosa* via triparental conjugation with *Escherichia coli* SM10(λpir) and SM10(λpir) pTNS1 (Choi and Schweizer 2006), and exconjugants were selected on LB agar supplemented with chloramphenicol (10 µg/ml) and gentamicin (20 µg/ml) and verified by PCR.

In addition, we measured the SP of N_2O produced by two wild-type, non-denitrifying bacteria with only *fhp/hmp* annotated in their genomes – *Staphylococcus aureus* USA300 LAC (putative flavohemoprotein *SAUSA300_0234*) and *Acinetobacter baumannii* ATCC 17978 (putative flavohemoprotein *A1S_3085*) (both kindly provided by Eric Skaar, Vanderbilt University Medical Center). Strains were first screened for N_2O production in the presence of NO (See "N_2O Screen" below; Fig. 4.6, Tables 4.4).
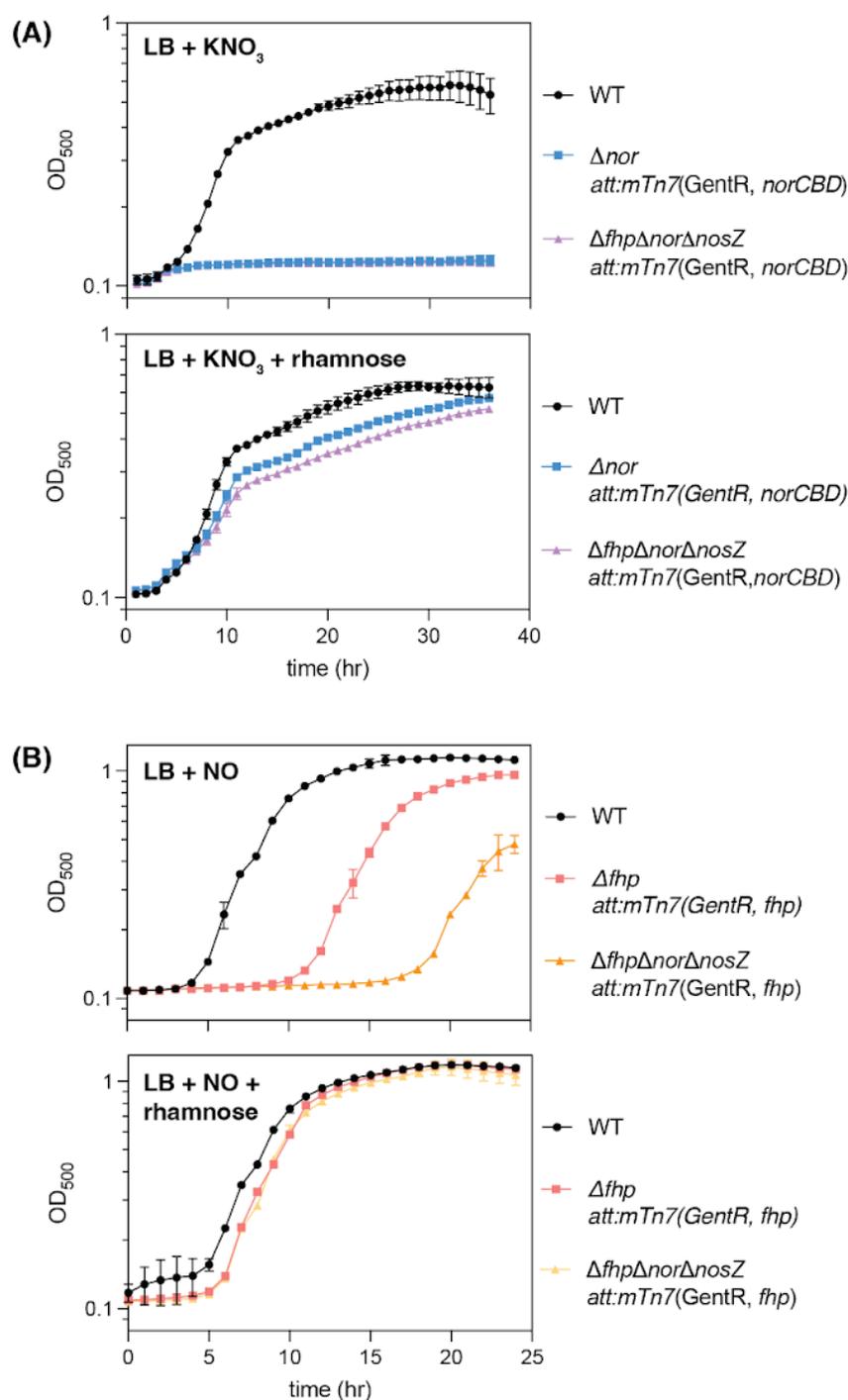
Fig. 4.4 Construction of iNOR and iFhp strains.
(A) WT *Pa* (black), *Δnor* (blue) and *ΔfhpΔnorΔnosZ* (purple) strains were grown anaerobically in Luria-Bertani (LB) media with 40 mM nitrate alone (upper panel) or 40 mM nitrate and 305 μM rhamnose (lower panel). **(B)** WT *Pa* (black), *Δfhp* (pink) and *ΔfhpΔnorΔnosZ* (yellow) strains were grown aerobically in LB media with 500 μM DETA NONOate alone (upper panel) or 500 μM DETA NONOate and 305 μM rhamnose (lower panel).
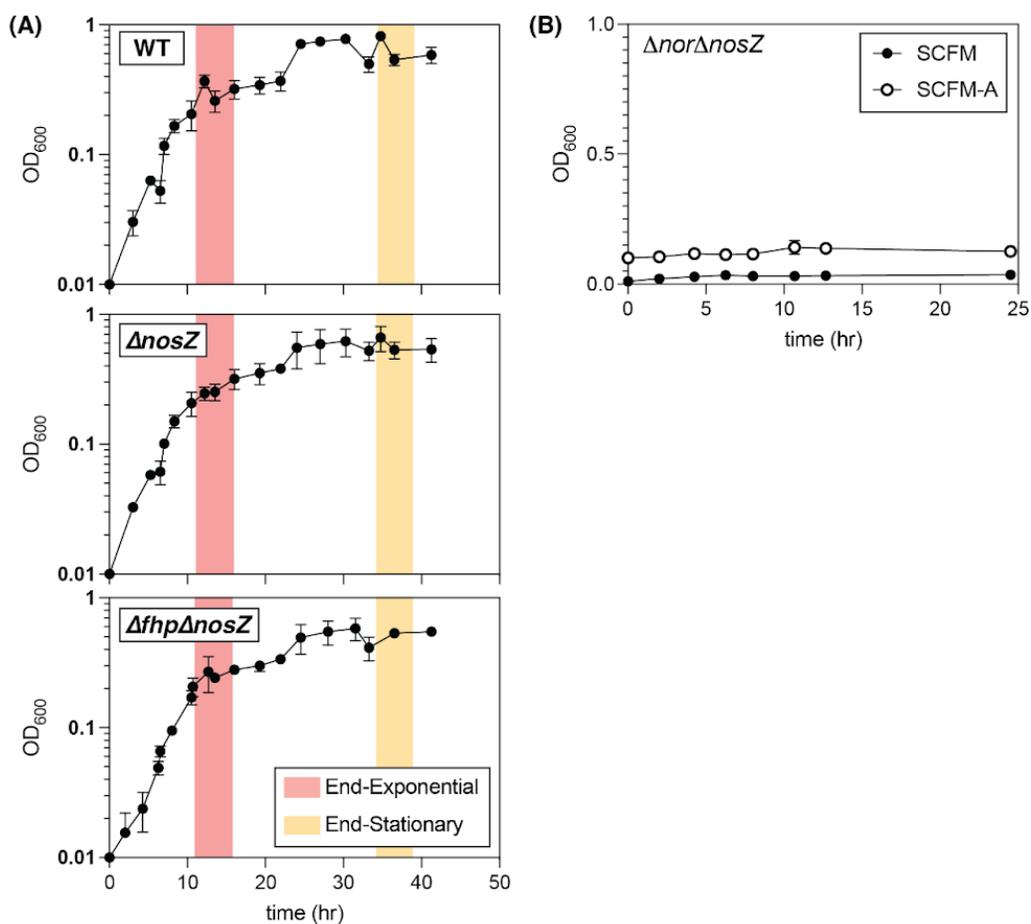
Fig. 4.5. Growth Curves.
(A) Growth curves for WT *Pa*, *ΔnosZ*, and *ΔfhpΔnosZ* grown in batch culture, denitrifying conditions with headspace sampling times for SP measurements (end-exponential, red; end-stationary, yellow). **(B)** Growth curves *ΔnorΔnosZ* strain in SCFM (black) and SCFM-A (white) media.

Primer ID	Sequence	Description
fhp_1-55	tctgcaggaattcctcgagaagcttatgttgtccaatgcccaacgtgcc	Amplify *fhp* for generation of iFhp strain, forward
fhp_1-56	gcaaggccttcgcgaggtacctcaggcgtccagcgcggc	Amplify *fhp* for generation of iFhp strain, reverse
norCBD_2-5	tctgcaggaattcctcgagaagcttatgtccgagacctttaccaaaggcatggc	Amplify *nor* for generation of iNOR strain, forward
norCBD_2-6	gcaaggccttcgcgaggtacctcagcggcgcaggcgccg	Amplify *nor* for generation of iNOR strain, reverse
fhp DN F	GCATGCGTCAGGAGTCATCTTGGACGCCTGAAGCGACGGG	
fhp DN R	CATGATTACGAATTCGAGCTAGCACGCAGCCCAGCAGGAT	
fhp UP F	ACGACGGCCAGTGCCAAGCTTGGCCGAACAATTCGCTTTC	
fhp UP R	CCCGTCGCTTCAGGCGTCCAAGATGACTCCTGACGCATGC	
fhp Genotyping F	GCAAGGGATTGGTGGTCATTTCG	
fhp Genotyping R	CATCAGCCTGGAACGATCAAGC	

Table 4.3. Primers used in this study.
Primers used for amplification and deletion of *fhp* and *norCBD* in parent strain *Pseudomonas aeruginosa* UCBPP-PA14.

Sample ID	Sample Description	Moles of N_2O sampled	OD_{600}	Corrected N_2O / OD_{600}
BLK 1	DETA NONOate only	4.7E-08	*NA*	*NA*
BLK 2	DETA NONOate only	5.6E-08	*NA*	*NA*
Sa 1	*Staphylococcus aureus*	*NA*	2.979	*NA*
Sa 2	*Staphylococcus aureus*	7.4E-07	3.986	5.80E+06
Ab 1	*Acinetobacter baumannii*	1.4E-06	2.794	2.00E+06
Ab 2	*Acinetobacter baumannii*	1.2E-06	3.546	2.99E+06

Table 4.4. Results of N_2O screen for Fhp-only strains.
See Fig. S4 for representative scans and measurement positions for peak intensities. The relative N_2O peak intensity was calculated by dividing the intensity at mass 43.663 (N_2O-only peak) by the intensity at 43.63 (N_2O and CO_2 peak). The pressure of each sample in the Ultra bellows was recorded, and total moles of gas were calculated using the ideal gas law and a bellow volume of 40 cc and a temperature of 295.15 K. Moles of total gas was then multiplied by the relative proportion of N_2O at the 44 peak to calculate total moles of N_2O. Corrected N_2O indicates that moles of N_2O sampled have been subtracted by the average moles of N_2O produced in the no-cell controls (BLK 1 and 2). A concentration of 500 μM was used for DETA NONOate.
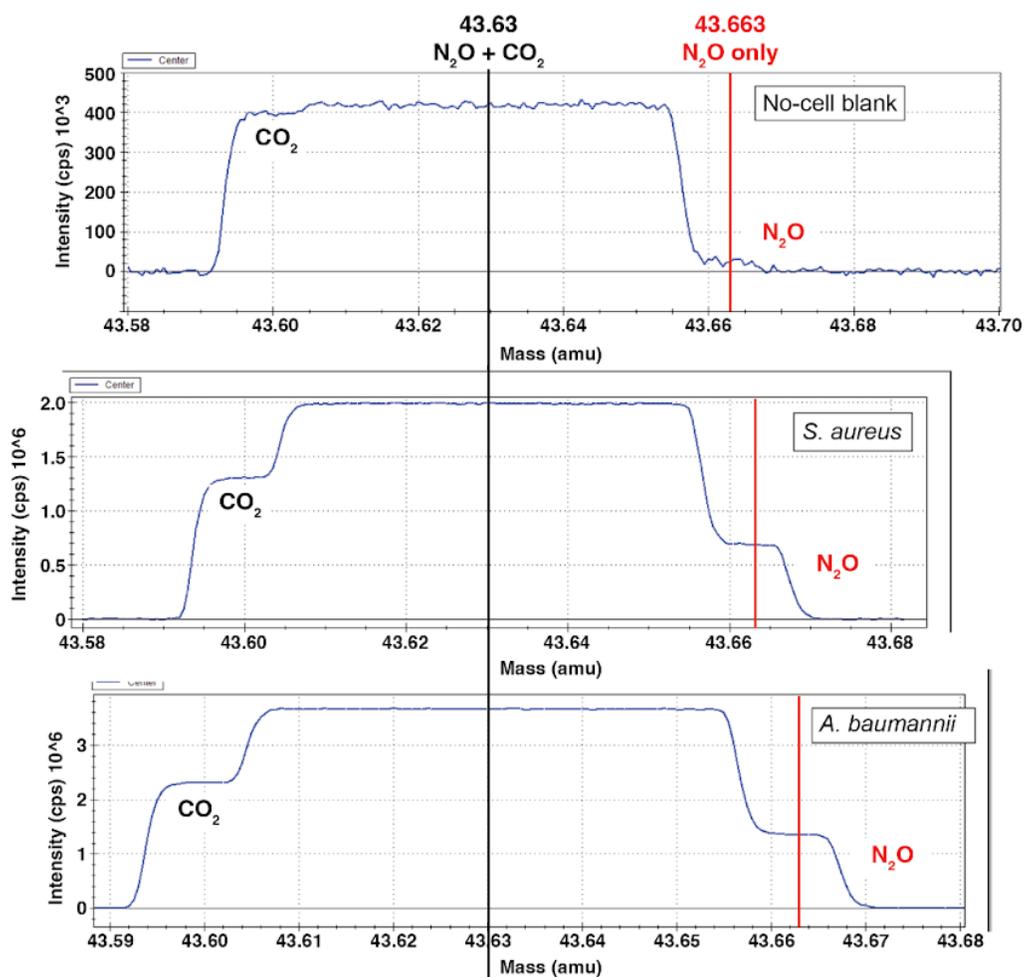
Fig. 4.6. Results of N₂O Screen for strains with only Fhp.
Representative Mass 44 peak scans on the HR-IRMS for *S. aureus*, *A. baumannii,* and a no-cell blank after a preliminary distillation of the headspace (Methods). N₂O and CO₂ peaks are annotated as above; peak location was determined using an N₂O reference gas. Vertical lines indicate peak positions used to measure total N₂O and CO₂ intensity (black line, 43.63 amu) and N₂O-only intensity (43.663 amu, red line); these intensities were used to calculate total N₂O yield.

4.7.3 Culturing conditions

iNOR, iFhp, and non-*Pseudomonas* were first screened for N₂O production before scaling up the culturing process for isotopic measurement. All strains were first grown to a high density (OD_{600} ~ 3-4) from glycerol freezer stocks in aerobic pre-growths (25 mL SCFM-A, 250 rpm shaking for 16 hours at 37°C). Cells were then pelleted and fully re-suspended into 25 mL of fresh media in sealed, glass 18 x 150 mm Balch tubes. The headspace was then purged with N_2 gas to establish anoxia, and 500 μM DETA NONOate was added. Balch tubes were incubated statically for 24 hours at 37°C. The headspace was then sampled on the vacuum line and distilled to concentrate N_2O and CO_2 in a preliminary distillation (see below for further detail).

Next, all isotopic measurements were performed on strains grown by two types of assays – suspension assays or batch culture (Fig. 4.7). All strains were grown in SCFM-A, but the NO source (KNO_3^- vs. DETA NONOate) varied per experiment. iNOR, iFhp, *A. baumannii*, and *S. aureus* were only grown as suspension assays. *ΔnosZ* and *ΔnosZΔfhp* were only grown as batch cultures. WT *Pa* was grown as both suspension assays and batch cultures. All anoxic incubations were performed in custom vacuum sampling flasks (Fig. 4.8). Vacuum flasks could not be sterilized through autoclaving because the flask cracked under high pressures. Therefore, flasks were instead sterilized with 80% ethanol, then dried overnight at 56°C and exposed to UV light in a sterile, laminar-flow hood for 10 minutes.

For suspension assays, strains were first grown in shaking, aerobic pre-growths for 16 hours at 37°C (OD_{600} ~3-4) in 150 mL SCFM-A. The aerobic pre-growths for iNOR, iFhp, *A. baumannii*, and *S. aureus* were supplemented with 100 mM KNO_3. For WT PA14 suspension assays, pre-growth was supplemented with 100 μM DETA NONOate (triangles, Fig. 4.9), 100 μM DETA NONOate and 100 mM nitrate (stars, Fig. 4.9), or 100 mM nitrate (circles, Fig. 4.9). Next, cells transferred to 50 mL conical tubes, pelleted for 15 minutes at 23°C and 6,800 xg, and resuspended in 150 mL fresh SCFM-A. 500 μM DETA NONOate was added to iNOR, iFhp, *A. baumannii*, and *S. aureus* experiments; iNor and iFhp was also supplemented with 305 μM L-rhamnose monohydrate ($C_6H_{12}O_5 \cdot H2O$ (Sigma-Aldrich R3875-25G) to promote rhamnose-inducible expression of *norBC* or *fhp*. For WT PA14 suspension assays, either 500 μM DETA NONOate (triangles and stars, Fig. 4.9) or 500 μM DETA NONOate and 100 mM nitrate (circles, Fig. 4.9) was added. Following suspension setup, vacuum flask headspace was purged with N_2 gas to establish anoxia, and flasks were incubated statically for 24 hours at 37°C before headspace sampling.

For batch culture assays, strains were first grown in aerobic pre-growths of 5 mL SCFM-A with 100 mM nitrate for 16 hours at 37°C, 250 rpm shaking (OD_{600} ~3-4). Cells were then diluted to OD_{600} = 0.01 in vacuum flasks with 150 mL of SCFM-A, For *ΔnosZ* and *ΔnosZΔfhp*, 100 mM of KNO_3 was added. For WT *Pa*, either 100 mM KNO_3 (squares, Fig. 4.9) or 500 μM DETA NONOate and 100 mM KNO_3 (circles, Fig. 4.9) was added. Vacuum flask headspace was purged with N_2 gas to establish anoxia and incubated statically at 37°C. Flasks were sampled twice: first, approximately 12 hours at end-exponential growth, and second, approximately 40 hours at end-stationary. One WT *Pa* batch culture experiment, where 500 μM DETA NONOate and 100 mM KNO_3 were added to the vacuum flask, was only sampled at ~40 hours after the DETA NONOate was added at ~12 hours (Fig. 4.7). Additional moles of nitrate were accidentally added in the Aug192021 batch for a final concentration of 233 mM nitrate (Table 4.5); however, no difference in SP was observed (Fig. 4.1D).
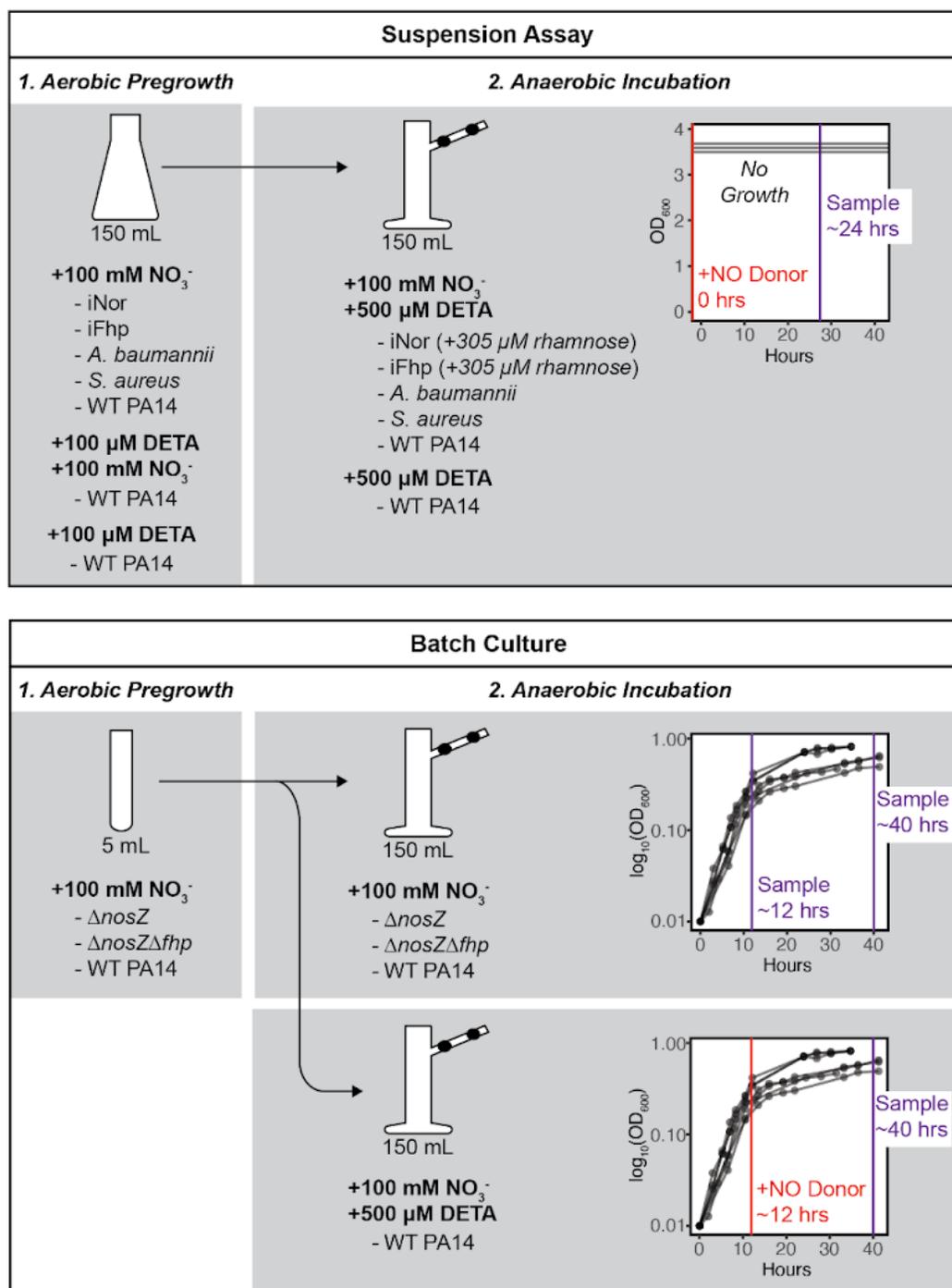
Fig. 4.7. Culturing conditions.
Suspension assays or batch culture assays were performed for this study. Nitrate and / or DETA NONOate was added at varying concentrations for each experiment – concentrations are written in bolded text while strains grown in those conditions are listed below. All strains were first grown in an aerobic pre-growth before anaerobic incubation in the vacuum sampling flask.
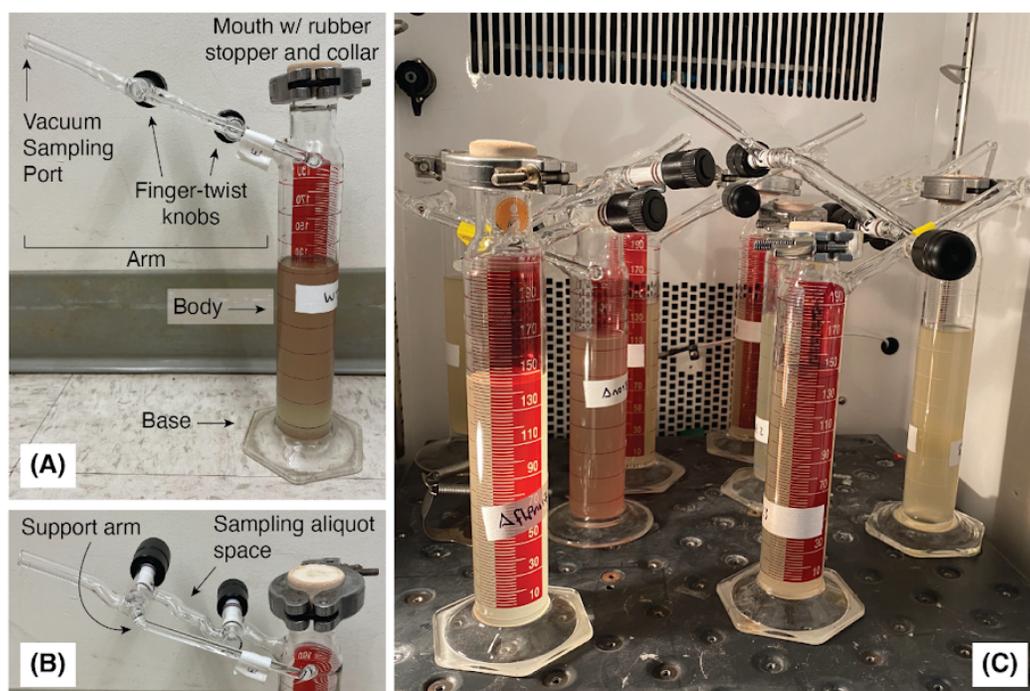
Fig. 4.8. Vacuum sampling flasks for anoxic incubation.
Custom glass vacuum flasks, building off a similar design used in (Magyar 2017; Magyar et al. 2016), were made in collaboration with the Caltech Glass Shop. **(A, B)** The mouth of a 200 mL borosilicate graduated cylinder was removed and a narrowed neck with a rubber stopper was attached. After media and cells were added, a finger-tightened metal collar was placed around the rubber stopper as an additional safeguard. ⅜" gauge glass tubes with two finger-twist knobs were added at the neck of the flask for headspace sampling on the vacuum line. A small sampling aliquot space was retained between the two knobs to isolate gas from the culturing media and the vacuum line. **(C)** Multiple flasks incubated at 37°C, as in a typical sampling workflow. Flasks were incubated in the dark; the light was turned on for the photo.
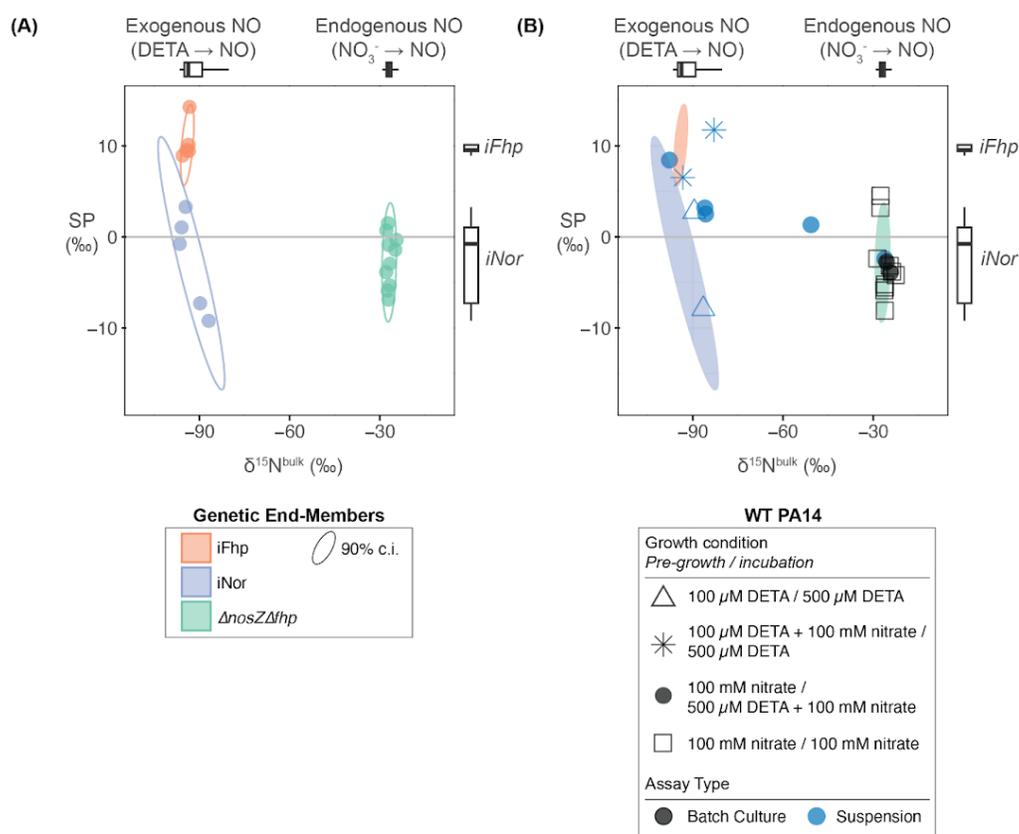
Fig 4.9: Paired SP and $\delta^{15}N^{bulk}$ data for all experiments in parent strain *P. aeruginosa*.
(A) $\delta^{15}N^{bulk}$ and SP data for all genetic end-member; i.e. strains with only NOR or Fhp (iFhp, iNor and *ΔnosZΔfhp*). Ellipses show 90% confidence intervals (c.i.). Boxplots showing expected values for exogenous vs. endogenous NO source (above plot) are taken from Fig. S15A, and boxplots showing iNor vs. iFhp values are taken from Fig. 1E. iFhp and iNor were grown as suspension assays while *ΔnosZΔfhp* was grown as batch culture. NOR can use both endogenous and exogenous NO (green vs. blue data) while Fhp can only use exogenous NO (orange data). **(B)** Overlay of 90% c.i. from (A) with experimental results from WT PA14, which has both NOR and Fhp. WT PA14 was grown as either batch culture or suspension assays (black vs. blue data points) with varying combinations of NO sources in the aerobic pre-growth vs. anoxic incubation (triangles, stars, circles, and squares as noted in the legend). Overall, batch cultures strongly cluster towards iNor values, while suspension assays range over iFhp and iNor. The most positive SP values (blue stars) only occur when WT PA14 is grown as a suspension assay and given a high dose of NO in the aerobic pre-growth (100 μM DETA + 100 mM nitrate), and incubated anoxically with exogenous NO (500 μM DETA). All analyses and data visualization were performed using R Statistical Software (v4.1.0; (R Core Team 2021)) and the ggplot2 package (v3.3.6; (Wickham et al. 2016)).

Batch	Material	$\delta^{15}N$ (‰)	Moles of N added
Feb102022	Nitrate	0.15 ± 0.26	0.015
Feb102022	SCFM Amended	-0.82 ± 0.19	0.00420
Aug302021	Nitrate	1.79 ± 0.02	0.015
Aug302021	SCFM Amended	-1.77 ± 0.13	0.00420
Aug192021	Nitrate	-0.73 ± 0.08	0.0350
Aug192021	SCFM Amended	-1.91 ± 0.14	0.00420

Table 4.5. Batch culture nitrogen isotopes.
Batch culture experiments were carried out using three batches of nitrate and SCFM Amended media – Feb102022, Aug302021 and Aug192021. $\delta^{15}N$ values (mean ± s.d.) are corrected for tin capsule blanks; moles of N indicate total how many moles of N from nitrate or SCFM Amended were added to the total 150 mL culture volume. Additional moles of nitrate were accidentally added in the Aug192021 batch.

4.7.4 Headspace sampling and N_2O distillation

N_2O was distilled from the headspace samples on an ultra-torr vacuum line prior to isotopic analysis (Fig. 4.10). First, the sample was expanded onto the left side of the line (Step 1); higher pressure samples were sampled by taking multiple aliquots while lower pressure samples were fully exposed to the line. Next, non-condensables (i.e. N_2, Ar) were removed by passing the sample over a trap in liquid nitrogen (LN_2, T2 in Fig. 4.10). Then, the sample was passed back and forth over the ascarite tube and the ethanol / dry ice slurry trap (T3, Fig. 4.10) to remove CO_2 and H_2O. Each pass lasted four minutes. The sample was isolated from the vacuum and the directionality of the sample flow was determined by either submerging T1 (clockwise flow) or T2 (counter-clockwise flow) in LN_2. The ascarite tube was re-made roughly every six samples; it consists of a ≈10" length pyrex tube of ⅜" gauge containing sodium hydroxide (Ascarite II CO_2 Absorbent, Thomas Scientific) and sealed with quartz wool on both ends. The ethanol / dry ice trap was a slurry of 100% ethanol (v/v) mixed with dry ice (solid CO_2). Finally, the sample was passed over the ethanol / dry ice slurry for a final time and flame-sealed into a pyrex glass finger until isotopic analysis.

Two vacuum distillation blanks (0100, 0101) and a no-cells vacuum flask blank (0112) were measured to test if the distillation process causes significant isotopic fractionation (Table 4.6). A total mixture of 640 μmol CO_2 and 290 μmol N_2O were expanded to a total volume of 127 cc on the vacuum line, then equilibrated with a pyrex finger of ~5 cc containing room air and ~0.1 mL DI water. Two aliquots of this mixture were taken as mock samples (0100 and 0101). The no-cells vacuum flask blank (0112) was prepared as a batch culture, but after the

headspace was purged with N_2 gas, N_2O from the reference tank was injected into the flask. This flask was then incubated at 37°C for ~12 hours and sampled at end-exponential phase (~12 hours). 0100 and 0101 showed little difference from the original N_2O gas (roughly 0.1 ± 0.5‰ difference), indicating that the distillation process does not significantly fractionate our target gas. 0112 showed a -2.25 ± 0.90‰ difference in $\delta^{18}O$; this may have been caused by exchange of O isotopes between the incubated N_2O gas and H_2O as noted in the main text (Fig. 4.11) – therefore our study relies on interpretation of the N isotopes instead.
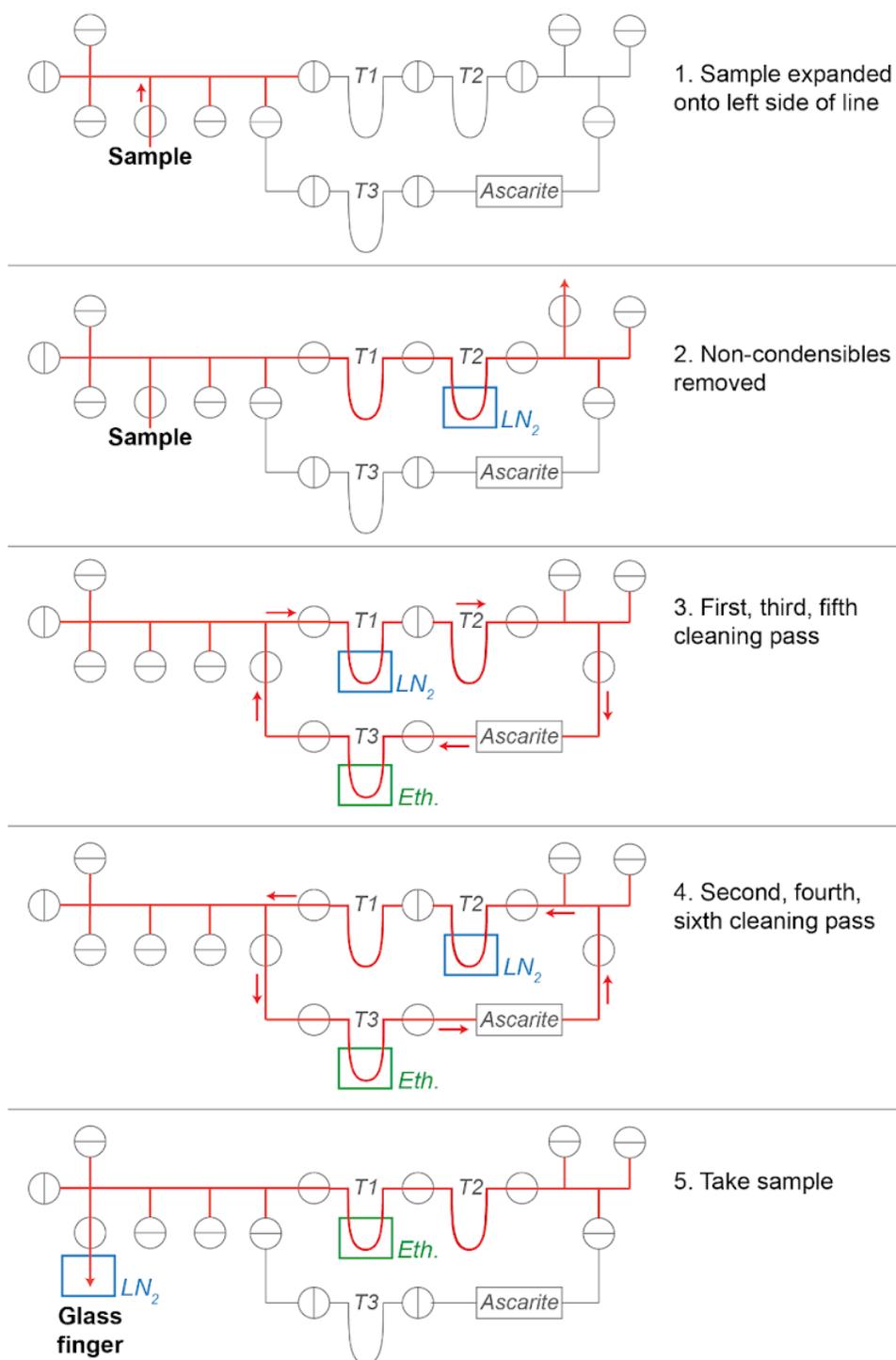
Fig. 4.10. Distillation of N₂O from headspace samples.
Diagram of ultra torr vacuum line used to distill N_2O from headspace samples. Red indicates portions of the line with sample gas. T1, T2 and T3 refer to different traps that were submerged in either liquid nitrogen (LN_2) or an ethanol / dry ice slurry (Eth.). Ascarite tube used for CO_2 removal is shown as a rectangle; valves are shown as circles with the center line indicating if the valve was closed or not; directionality of sample gas flow is shown with red lines.

	0100	**0101**	**0112**
$\delta^{15}N^{bulk}$	$0.12 \pm 0.32‰$	$0.07 \pm 0.21‰$	$0.41 \pm 0.42‰$
$\delta^{18}O$	$-0.03 \pm 0.30‰$	$0.02 \pm 0.41‰$	$-2.25 \pm 0.90‰$
$\delta^{15}N^{\alpha}$	$-0.09 \pm 0.28‰$	$-0.39 \pm 0.47‰$	$-0.34 \pm 1.17‰$

Table 4.6. N$_2$O distillation blank (0100, 0101) and no-cells vacuum flask blank (0112).
Values (mean ± s.d.) are reported relative to the working reference gas, Caltech Ref Gas. Measurement was done on the Prototype Ultra.
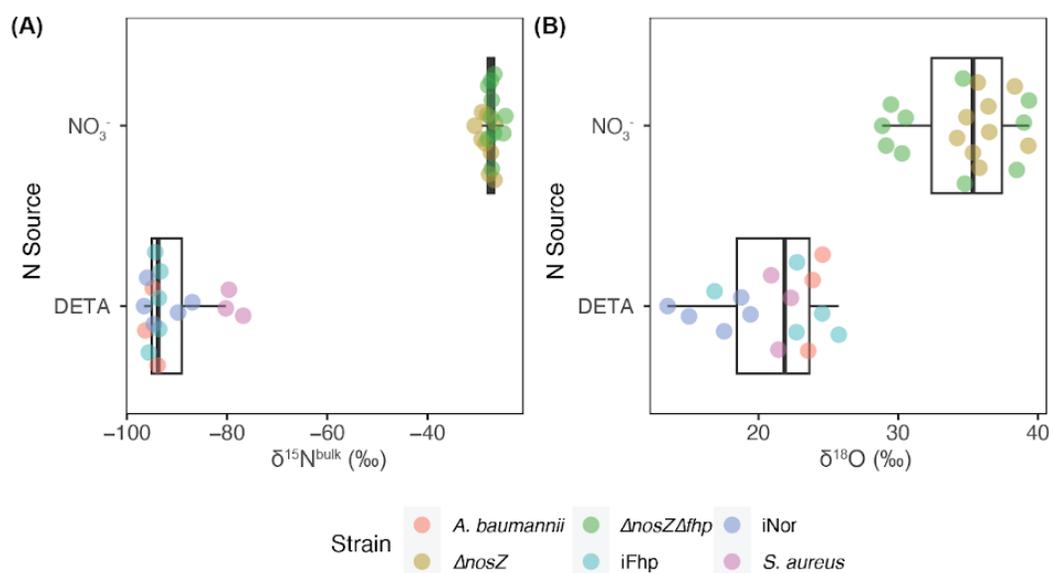
Fig 4.11: $\delta^{15}N^{bulk}$ and $\delta^{18}O$ data for all non-WT PA14 experiments.
(A) $\delta^{15}N^{bulk}$ data for all non-WT PA14 strains separated by primary NO source – endogenous NO through reduction of nitrate through the denitrifcation pathway, or exogenous NO through addition of DETA NONOate. *A. baumannii*, iNor, iFhp and *S. aureus* strains were all grown as suspension assays while *ΔnosZ* and *ΔnosZΔfhp* were grown as batch cultures. When N-oxides are reduced to N$_2$O, $\delta^{15}N^{bulk}$ retains the isotopic signature of the original N (Sigman et al. 2001) **(B)** $\delta^{18}O$ data for the same strains. $\delta^{18}O$ reflects O from both the original substrate and exchange with ^{18}O in solution; for nitrate, oxygen is also subject to isotope fractionations associated with each step of the denitrification pathway (i.e. 'branching isotope effect,' (Casciotti et al. 2007)). For both panels, boxplots show quartiles taken over all strains for each NO source. All analyses and data visualization were performed using R Statistical Software (v4.1.0; R Core Team 2021, (R Core Team 2021)) and the ggplot2 package (v3.3.6; Wickham, 2016, (Wickham et al. 2016)).

4.7.5 Site Preference measurements

4.7.5.1 Delta (δ) notation and definition of Site Preference (SP)

All isotopic measurements in this study are reported in the delta notation (δ) in units of per mil (‰) where:

$$\delta^{15}N = \left(\frac{^{15}R_{sam}}{^{15}R_{ref}} - 1\right) \times 1000$$

Equation 4.1

$$\delta^{18}O = \left(\frac{^{18}R_{sam}}{^{18}R_{ref}} - 1\right) \times 1000$$

Equation 4.2

Where ^{15}R is the ratio of $^{15}N/^{14}N$ in the sample ("sam") or reference ("ref"). All values here are reported to the international reference of Air for nitrogen.

Site Preference ($\delta^{15}N^{SP}$ or "SP" in this study) is defined as the relative, intramolecular enrichment of the rare, stable isotope ^{15}N for the central vs. terminal nitrogen in the linear, asymmetrical N_2O molecule. To be consistent with prior work, we use the designations as defined by (Yoshida and Toyoda 2000; Toyoda and Yoshida 1999) where the terminal nitrogen is labeled β, and the central nitrogen is labeled α. In this convention, the ^{15}R ratios for each site is defined as:

$$^{15}R^{\alpha} = \frac{[^{14}N^{15}N^{16}O]}{[^{14}N^{14}N^{16}O]}$$

Equation 4.3

$$^{15}R^{\beta} = \frac{[^{15}N^{14}N^{16}O]}{[^{14}N^{14}N^{16}O]}$$

Equation 4.4

Therefore, in delta notation:

$$\delta^{15}N^{\alpha} = \left(\frac{^{15}R^{\alpha}_{sam}}{^{15}R^{\alpha}_{ref}} - 1\right) \times 1000$$

Equation 4.5

$$\delta^{15}N^{\beta} = \left(\frac{^{15}R^{\beta}_{sam}}{^{15}R^{\beta}_{ref}} - 1\right) \times 1000$$

Equation 4.6

Site Preference is as defined as in (Toyoda and Yoshida 1999):

$$SP \equiv \delta^{15}N^{\alpha} - \delta^{15}N^{\beta}$$

Equation 4.7

4.7.5.2 Correction to international reference frame

The working reference gas ("Caltech Ref Gas") used in this study was previously characterized to the international working standards for nitrogen and oxygen isotopes (Air and VSMOW respectively) by Tokyo Tech (Magyar 2017; Magyar et al. 2016). These values are referred to as δ_{wg_intl} below. In those studies, Caltech Ref Gas was measured at the Tokyo Institute of Technology (Yokohama, Japan) by Sakae Toyoda and Naohiro Yoshida and reported values were $\delta^{15}N^\alpha = 7.53‰$ and $\delta^{15}N^\beta = 0.89‰$ with no uncertainty given (Table 4.9). For due diligence, Caltech Ref Gas was measured again by Stanford University (Stanford, CA, USA). They reported values of $\delta^{15}N^{bulk} = 4.64 \pm 0.24‰$, $\delta^{15}N^\alpha = 7.85 \pm 0.38‰$, $\delta^{15}N^\beta = 1.42 \pm 0.25‰$, $\delta^{15}N^{SP} = 6.43 \pm 0.43‰$ and $\delta^{18}O = 39.57 \pm 0.14‰$ (mean $\pm$ s.d.; Table 4.9). We note that the scrambling correction done by Stanford, which would affect values for $\delta^{15}N^\alpha$, $\delta^{15}N^\beta$, and $\delta^{15}N^{SP}$ differs from the one used by (Magyar 2017; Magyar et al. 2016; Toyoda and Yoshida 1999) and the Tokyo Tech group. See further discussion of the scrambling correction below.

We corrected our sample values that were measured relative to Caltech Ref Gas (δ_{sam_wg}) to be reported relative to the international reference (δ_{sam_intl}) by using the values for Caltech Ref Gas characterized to Air / VSMOW (δ_{wg_intl}). Each δ value is defined as:

$$\delta_{sam-wg} = \left(\frac{R_{sam}}{R_{wg}} - 1\right) \times 1000$$

<div align="right">Equation 4.8</div>

$$\delta_{wg-air} = \left(\frac{R_{wg}}{R_{intl}} - 1\right) \times 1000$$

<div align="right">Equation 4.9</div>

$$\delta_{sam-air} = \left(\frac{R_{sam}}{R_{intl}} - 1\right) \times 1000$$

<div align="right">Equation 4.10</div>

We would like to solve for R_{sam}/R_{intl} while keeping everything in delta (δ) notation. This is to avoid inconsistencies across labs which may be using different exact values for ^{15}R and ^{18}R, the ratio of $^{15}N/^{14}N$ or $^{18}O/^{16}O$ in Air or VSMOW respectively. Therefore, we can rearrange Equation 4.8 to get:

$$\left(\frac{\delta_{sam-wg}}{1000} + 1\right) = \frac{R_{sam}}{R_{wg}}$$

<div align="right">Equation 4.11</div>

Then rearrange Equation 4.9 to get:

$$\left(\frac{\delta_{wg-intl}}{1000} + 1\right) = \frac{R_{wg}}{R_{intl}}$$

Equation 4.12

We can then solve for R_{sam}/R_{air} by multiplying Equation 4.11 and 4.12 to cancel R_{wg}:

$$\frac{R_{sam}}{R_{intl}} = \frac{R_{sam}}{R_{wg}} \times \frac{R_{wg}}{R_{intl}}$$

Equation 4.13

Finally, we can substitute Equations 4.11 and 4.12 into Equation 4.13 to get our final values reported values:

$$\delta_{sam-intl} = \left[\left(\frac{\delta_{sam-wg}}{1000} + 1\right) \times \left(\frac{\delta_{wg-intl}}{1000} + 1\right) - 1\right] \times 1000$$

Equation 4.14

We can then measure other externally characterized working reference gasses to determine which values for δ_{wg_intl} (by Tokyo Tech or Stanford) we should proceed with. We measured aliquots of RM1B, RM3B and RM5 (Table 4.9), reference gasses that have been previously characterized as part of an effort to create a suite of isotopically characterized N_2O reference materials to be used as community standards (Mohn et al. 2022). Using our measured value for RM5 vs. Caltech Ref Gas (δ_{sam_wg}), we could then calculate δ_{wg_air} using either Tokyo Tech or Stanford's reported δ_{wg_intl} value for $\delta^{15}N^{bulk}$ and $\delta^{18}O$, which are not affected by the scrambling factor. The δ_{wg_intl} values reported by Tokyo Tech gave the most similar values for RM1B, RM3B, and RM5 as reported by (Mohn et al. 2022) (Table 4.8). Therefore, we proceeded to use the δ_{wg_air} values reported by Tokyo Tech to correct our measured sample values (δ_{sam_wg}) to Air or VSMOW (δ_{sam_intl}).

Finally, we note that if we had used the δ_{wg_intl} values reported by Stanford, it would not significantly affect the interpretation of our results. Doing so would have shifted all our $\delta^{15}N^{bulk}$ values to be more positive by roughly 0.4‰ and our $\delta^{18}O$ values to be more negative by roughly 0.4‰. If $\delta^{15}N^{\alpha}$ does not change, this would shift all SP values to be more negative by roughly 0.8‰. Therefore, the absolute values of each data point would change, but because this affects all data points, this would not change our interpretation of relative differences in the data.

	Characterized by Tokyo Tech (9,10)	Characterized by Stanford (this study; mean ± std. dev.)
$\delta^{15}N^{bulk}$	4.21‰	4.64 ± 0.24‰
$\delta^{15}N^{\alpha}$	7.53‰	7.85 ± 0.38‰
$\delta^{15}N^{\beta}$	0.89‰	1.42 ± 0.25‰
$\delta^{15}N^{SP}$	6.64‰	6.43 ± 0.43‰
$\delta^{18}O$	39.96‰	39.57 ± 0.14‰

Table 4.7. Characterization of working reference gas to international isotope standards
Values are reported relative to Air for $\delta^{15}N$ and to VSMOW for $\delta^{18}O$. Caltech Ref Gas was characterized by: i) Colleagues at Tokyo Tech and reported without error in (Magyar 2017; Magyar et al. 2016); ii) Colleagues at Stanford University with values reported as mean ± s.d. The N_2O measurement by Stanford University is a continuous-flow method optimized for N_2O in seawater; measurement method and scrambling correction done by Stanford University is detailed in (C. Kelly et al. 2023; C. L. Kelly et al. 2023). See text or more detail.

	RM1B	RM3B	RM5
$\delta^{15}N^{bulk}$	0.22 ± 0.05‰	16.08 ± 0.05‰	33.44 ± 0.05‰
$\delta^{15}N^{\alpha}$	-0.38 ± 0.91‰	15.74 ± 0.91‰	43.54 ± 0.91‰
$\delta^{15}N^{\beta}$	0.82 ± 1.29‰	16.42 ± 1.29‰	23.34 ± 1.29‰
$\delta^{15}N^{SP}$	-1.19 ± 0.91‰	-0.68 ± 0.91‰	20.2 ± 0.91‰
$\delta^{18}O$	38.86 ± 0.15‰	55.17 ± 0.15‰	39.52 ± 0.15‰

Table 4.8. External reference gasses measured.
Data is from Table 12 of (11); Error is reported as 1 s.d.

	Tokyo Tech Values

Sample ID	$\delta^{15}N^{bulk}$	Absolute offset	$\delta^{18}O$	Absolute offset
RM1B_1	0.82 ± 0.33	0.60 ± 0.33	39.05 ± 0.43	0.19 ± 0.45
RM1B_2	0.88 ± 1.29	0.66 ± 1.29	38.65 ± 0.23	0.21 ± 0.27
RM1B_3	1.01 ± 0.42	0.79 ± 0.42	38.50 ± 0.49	0.36 ± 0.51
RM3B_1	16.90 ± 0.86	0.82 ± 0.86	55.24 ± 0.27	0.07 ± 0.31
RM3B_2	17.03 ± 0.80	0.95 ± 0.81	55.13 ± 0.21	0.04 ± 0.26
RM3B_3	16.83 ± 0.57	0.75 ± 0.57	55.02 ± 0.39	0.15 ± 0.42
RM5_1	34.26 ± 0.36	0.82 ± 0.37	39.31 ± 0.36	0.21 ± 0.39
RM5_2	34.01 ± 0.53	0.57 ± 0.54	39.41 ± 0.48	0.11 ± 0.50
	Stanford Values			
Sample ID	$\delta^{15}N^{bulk}$	Absolute offset	$\delta^{18}O$	Absolute offset
RM1B_1	1.25 ± 0.33	1.03 ± 0.33	38.66 ± 0.43	0.20 ± 0.45
RM1B_2	1.31 ± 1.29	1.09 ± 1.29	38.26 ± 0.23	0.60 ± 0.27
RM1B_3	1.44 ± 0.42	1.22 ± 0.42	38.11 ± 0.49	0.75 ± 0.51
RM3B_1	17.34 ± 0.86	1.26 ± 0.86	54.84 ± 0.27	0.33 ± 0.31
RM3B_2	17.46 ± 0.80	1.38 ± 0.81	54.74 ± 0.21	0.43 ± 0.26
RM3B_3	17.26 ± 0.57	1.18 ± 0.57	54.63 ± 0.39	0.54 ± 0.42
RM5_1	34.70 ± 0.36	1.26 ± 0.37	38.92 ± 0.36	0.60 ± 0.39
RM5_2	34.45 ± 0.53	1.01 ± 0.54	39.02 ± 0.48	0.50 ± 0.50

Table 4.9. Characterization of external reference gasses using Tokyo Tech vs. Stanford values for Caltech Ref Gas.
$\delta^{15}N^{bulk}$ and $\delta^{18}O$ values of RM1B, RM3B, and RM5 were converted to the international reference frame corrected values for Caltech Ref Gas reported by Tokyo Tech or Stanford using Equation S18. Values are reported as mean ± s.d. The absolute difference for $\delta^{15}N^{bulk}$ and $\delta^{18}O$ was smallest when using the δ_{wg_intl} reported by Tokyo Tech; therefore we proceeded using the Tokyo Tech values to convert our data from vs. Caltech Ref Gas to the international reference frame (Air or VSMOW).

4.7.5.3 SP Measurement

SP measurements were performed on two versions of the Thermo Scientific Ultra High-Resolution Isotope Ratio Mass Spectrometer (HR-IRMS), the 'Prototype Ultra' (Eiler et al. 2013) and the 'Production Ultra.' Two measurements were performed on each sample – the first at Mass 30 and 31 for $\delta^{15}N^{\alpha}$, and the second at Mass 44, 45 and 46 for $\delta^{15}N^{bulk}$ and $\delta^{18}O$. Larger samples were measured on the Prototype Ultra – Mass 30 on the C4 cup using a 10^{11} Ω resistor, Mass 31

on the Center cup using a 10^{12} Ω resistor; Mass 44 on the Center cup typically with a 10^{11} Ω resistor; Mass 45 on the C6 cup using a 10^{12} Ω resistor; Mass 46 on the C7 cup using a 10^{12} Ω resistor. On the Production Ultra, Mass 30 was measured on the Center cup using an 10^{12} Ω resistor; Mass 31 on the H2 Cup using the CDD (Compact Discrete Dynode secondary electron multipliers); Mass 44 on the Center cup using an 10^{12} Ω resistor, Mass 45 on the H2 Cup using the CDD, Mass 46 on the H3 Cup using the CDD. All measurements were corrected for background ("Johnson") noise. Background correction was done before and after measurement on-peak to adjust for any pressure-related intensity changes, or other instrument changes over the course of the measurement. All measurements were performed using a settling time of 1 second, an integration time of 8.38 seconds, 10 total integrations, and then 10 repeats of the measurement block. Pressure was adjusted to the main peak (Mass 30 or 44) with an allowed error of ±1%. Bellows were automatically re-balanced by the instrument, and all sample measurements were bracketed by the reference gas.

A Mass 45 foot correction was done to correct for a $^{13}C^{16}O_2$ 'foot' that overlaps with the $^{14}N^{15}N^{16}O$ / $^{15}N^{14}N^{16}O$ measurement 'shoulder.' This 'foot' is present in both reference and sample gasses; reference gas was obtained through MATSON. The correction was done by calculating a pressure-varying ratio of the foot vs. shoulder, then applying this pressure-varying ratio over the course of the measurement block. Two foot correction observations bracketing the measurement on-peak on the 'shoulder' were done to account for any pressure-related intensity changes over the course of the measurement. Over the observation period, both the foot and shoulder signal will decay exponentially with pressure, so both signals were fit with equations for exponential decay:

$$I_{foot} = a_{foot} \times e^{(b_{foot}*t)}$$

Equation 4.16a

$$I_{shoulder} = a_{shoulder} \times e^{(b_{shoulder}*t)}$$

Equation 4.16b

Where I = signal intensity, t = time, and a and b are fitted constants. We can then take the ratio of both equations for the correction:

$$\frac{I_{foot}}{I_{shoulder}} = \frac{a_{foot}}{a_{shoulder}} \times e^{[t*(b_{foot}-b_{shoulder})]}$$

Equation 4.17

Then this correction can be applied to the raw Mass 45 signal for the corrected Mass 45 signal:

$$45_{corr} = \left[1 - \frac{a_{foot}}{a_{shoulder}} \times e^{[t*(b_{foot}-b_{shoulder})]}\right] \times 45_{raw}$$

Equation 4.18

This correction generally caused the $\delta^{15}N$ of 45/44 to become more negative by roughly 1‰.

4.7.5.4 Shot noise error and limits of precision

In addition to background noise, caused by the inherent electrical current of a conducting body (Johnson 1928; Nyquist 1928), shot noise is another inherent limit of isotope ratio measurement that results from the fact that an ion current is composed of discrete particles distributed randomly with respect to time (Hayes 1983; Schottky 1918). Isotope ratios are calculated from ratios of observed ion currents and therefore are sensitive to anything that may affect that ion current (i.e. heat, mechanical vibrations, amplifier noise, etc.). However, even if there were no additional sources of noise, the precision attainable in all measurements would still be limited by shot noise (also called counting statistics); ideally, in modern isotope ratio mass spectrometry, shot noise should be the principal source of noise and measurement precision should reach shot noise limits (Hayes 1983). Therefore, for each measurement we calculated the shot noise error and compared it to the actual observed standard deviation of the measurement to see how close we approach shot noise limits.

Figure 4.12 shows calculated shot noise vs. observed standard deviation for all measurements made for this study (samples, zero-enrichments, sample reruns; *n* = 79) across both Prototype and Production Ultras over six experimental sessions over two years. Most measurements lie between the 1:1 and 2:1 line (Fig. 4.12D), indicating most measurements were done at or twice the shot noise limit. Linear regression across all points gives a slope close to 1 (*m* = 1.14 ± 0.04), indicating that overall measurements are approaching the shot noise error. Median calculated shot noise across all measurements (0.49‰, Fig. 4.12D) is less than the median measured std. dev. (0.61‰, Fig. 4.12D), indicating that, overall, measurements were reaching shot noise limits.
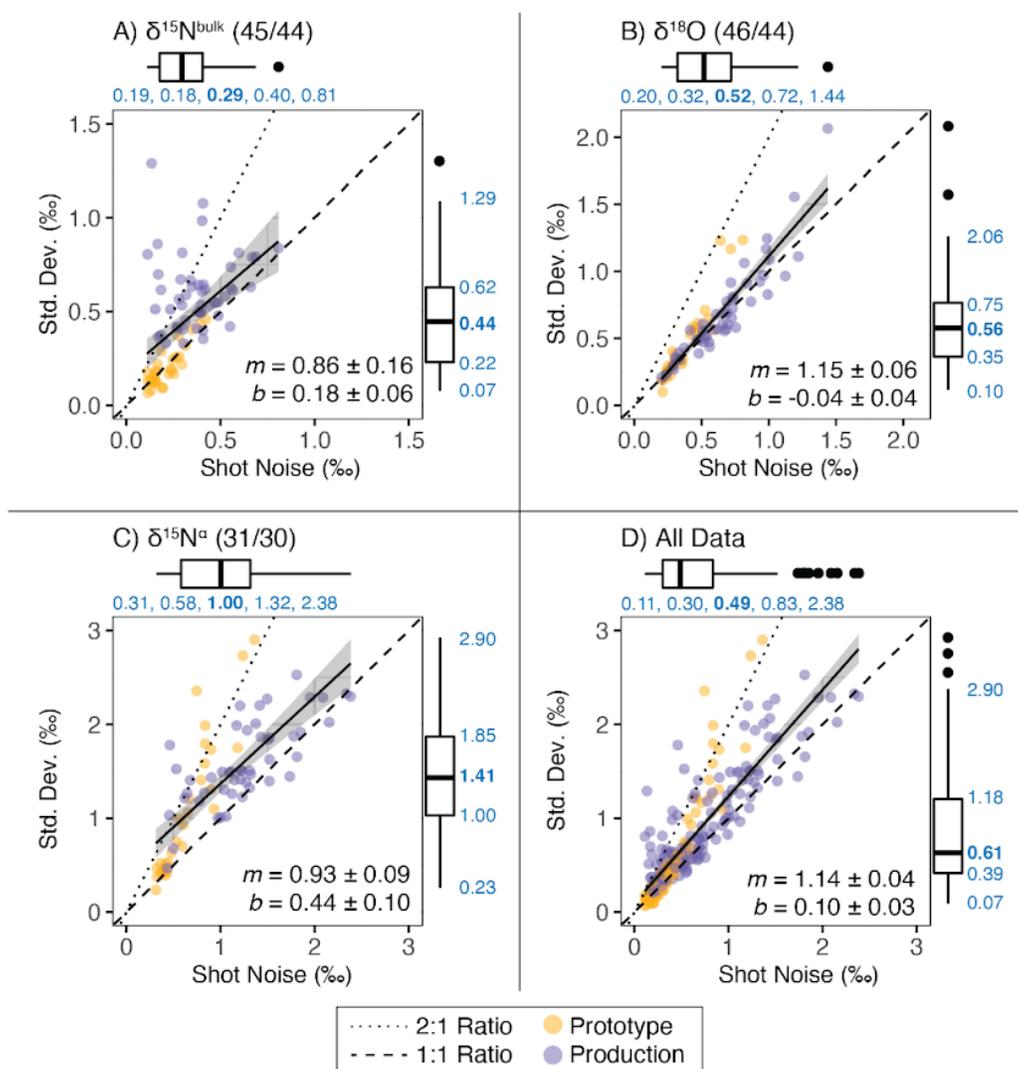
Fig. 4.12. Shot noise and limits of precision.
Calculated shot noise (x-axis) vs. observed standard deviation (y-axis) for all measurements (n=79) across the Prototype (yellow) and Production (purple) Ultra for (A) $\delta^{15}N^{bulk}$, (B) $\delta^{18}O$, (C) $\delta^{15}N^{\alpha}$, and (D) all isotopic measurements. Dashed line shows 1:1 line, which indicates when measurement precision has reached the calculated possible attainable precision (i.e. std. dev. = shot noise). Dotted line shows a 2:1 ratio of std. dev. to shot noise. Distributions for shot noise (boxplot on the top of each panel) and std. dev. (boxplot on the right of each panel) are shown with calculated quantiles in blue text, with median (50%) value in bold font [0%, 25%, **50%**, 75%, 100%]. Boxplots, quantiles and linear regression are calculated for measurements across both instruments. All analyses and data visualization were performed using R Statistical Software (v4.1.0; (R Core Team 2021)) and the ggplot2 package (v3.3.6; (Wickham et al. 2016)).

4.7.5.5 Zero enrichment tests and instrument performance

'Zero enrichment' tests where the reference gas is measured as a sample against itself were regularly performed over the course of the study to ensure that pressure balance for the sample and reference gas bellows were correctly calibrated.

If the bellows are correctly pressure calibrated, we would expect each measurement ($\delta^{15}N^{bulk}$, $\delta^{18}O$ and $\delta^{15}N^{\alpha}$) to give a value of 0‰ within uncertainty (1 s.d.).

Fig. 4.13 shows the result of eight zero enrichment tests for A) $\delta^{15}N^{bulk}$, C) $\delta^{18}O$ and E) $\delta^{15}N^{\alpha}$ run on the Prototype (pink circles) and Production Ultra (blue circles). Results are largely 0‰ within uncertainty (1 s.d.) with the exception of $\delta^{15}N^{bulk}$ and $\delta^{18}O$ for Jun. 2021, and $\delta^{15}N^{bulk}$ for Mar. 2023. However, this offset is on the order of 0.1‰, which is within the uncertainty of our shot noise error and is therefore likely due to the inherent limits of precision in our measurement. In addition, zero enrichments were run over a range of bellow pressures to gauge instrument performance across sample size (Fig. 4.13B,D,F). A proxy for bellow pressure is ion beam intensity, since increased sample volume causes increased ion beam intensity. Calculating a linear correlation (not shown on figure) with minor ion intensity as the independent variable and δ values as the dependent variable gives adjusted R^2 values of 0.09, 0.14 and -0.12 for $\delta^{15}N^{bulk}$, $\delta^{18}O$ and $\delta^{15}N^{\alpha}$ respectively (calculation performed using R Statistical Software (v4.1.0; (R Core Team 2021)) [call: *lm*()]. Therefore, there is no strong correlation between minor ion intensity (in counts per second, cps) and δ values, which means that our measurement method is accurate across a range of sample sizes, bellow pressures, and signal intensities.
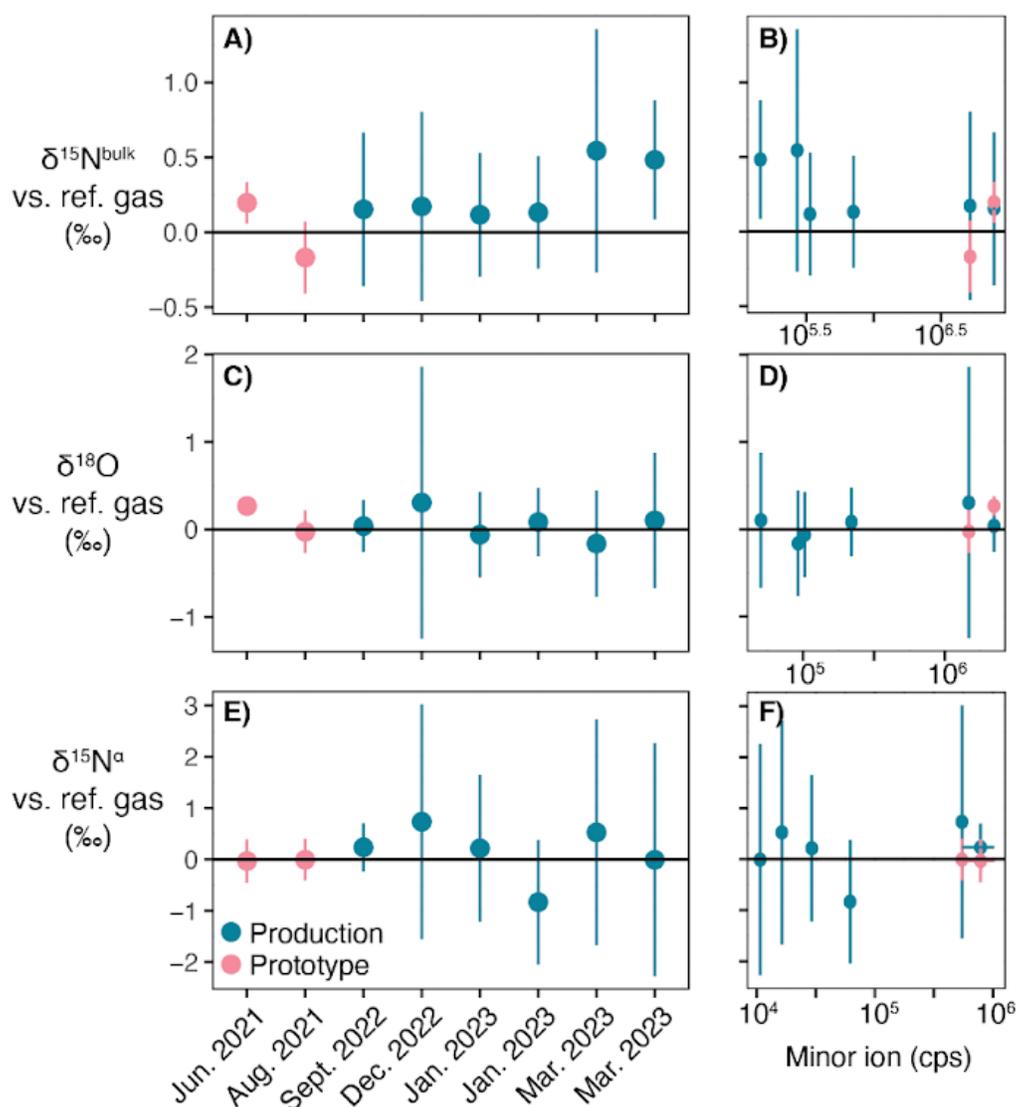
Fig. 4.13. Zero enrichment test results.
Zero enrichment test results on the Production (blue) and Prototype Ultras (pink); zero enrichment tests are where the reference gas was measured as a sample against itself. **(A)**, **(C)** and **(E)** show $\delta^{15}N^{bulk}$, $\delta^{18}O$ and $\delta^{15}N^{\alpha}$ vs. experimental session. **(B)**, **(D)** and **(F)** show the same δ values vs. minor ion intensity (cps) for Mass 45, 46 and 31 respectively. All analyses and data visualization were performed using R Statistical Software (v4.1.0; (R Core Team 2021)) and the ggplot2 package (v3.3.6; (Wickham et al. 2016)).

4.7.5.6 Measurement consistency across instruments

Two samples, 0225 and 0230, were measured on both the Prototype and Production Ultras to gauge measurement consistency across instruments (Fig. 4.14). Samples were first measured in April 2022 on the Prototype Ultra. They were then removed from the sample bellow by freezing into a small glass finger with a finger-twist valve using liquid nitrogen (LN₂). Samples were then taken to the

vacuum line, frozen into glass break-seals using LN_2, and stored in break-seals until further measurement. Samples were then re-measured on the Production Ultra on Dec. 2022. Measurements of 0225 and 0230 on both Ultras give the same value within measurement uncertainty (1 s.d.) vs. the reference gas for $\delta^{15}N^{bulk}$, $\delta^{18}O$ and $\delta^{15}N^{\alpha}$, and all measurements approach the shot noise limit where std. dev. to shot noise ratio is 1 (Fig. 4.14). Standard deviation for both 0225 and 0230 is lower on the Prototype Ultra because there was more total sample in the first measurement (i.e. some sample was consumed during the first measurement).
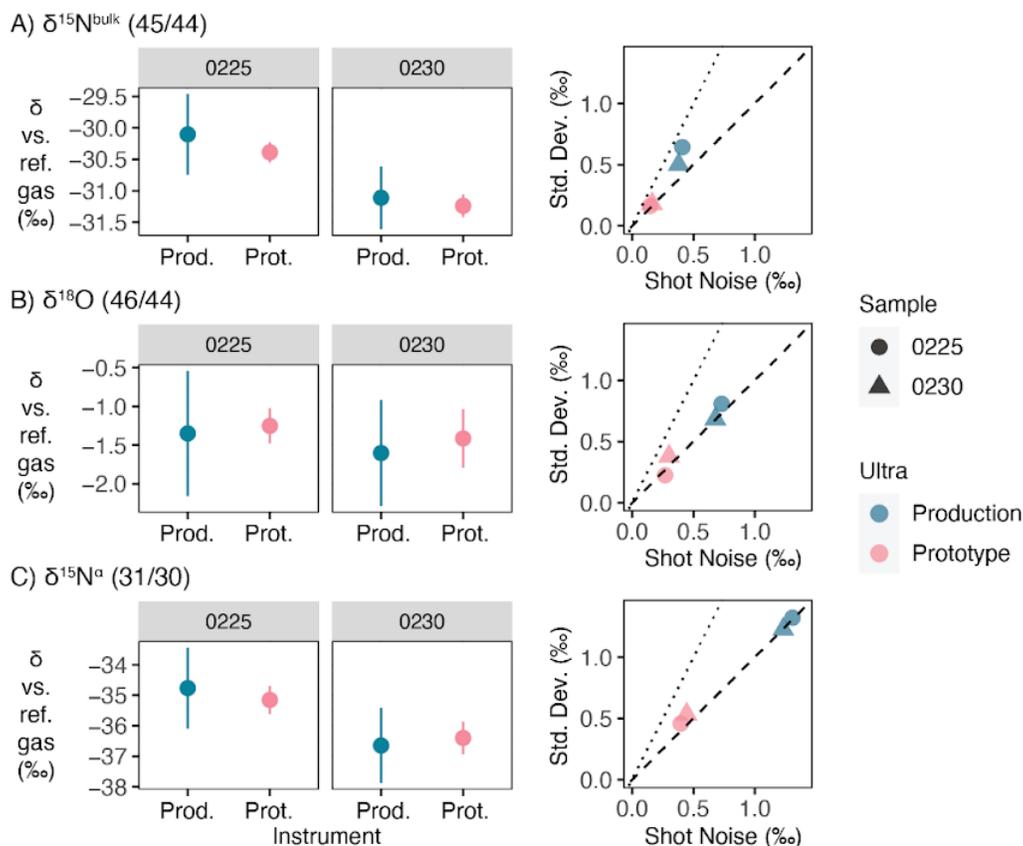
Fig. 4.14. Measurement consistency across instruments.
Two samples, 0225 (circles) and 0230 (triangles), were measured on both the Prototype (pink) and Production (blue) Ultras to gauge measurement accuracy across instruments. Measurements are reported vs. the reference gas for **(A)** $\delta^{15}N^{bulk}$, **(B)** $\delta^{18}O$ and **(C)** $\delta^{15}N^{\alpha}$. Right column shows shot noise (‰) on the x-axis and std. dev. (‰) on the y-axis; 1:1 ratio is shown as a dashed line and 2:1 ratio is shown as a dotted line.

4.7.5.7 Scrambling Correction

The SP measurement is reliant on two primary $\delta^{15}N$ measurements, that of the full molecule at Mass 44-45 ($\delta^{15}N^{bulk}$) and that of the NO fragment at Mass 30-31 ($\delta^{15}N^{\alpha}$). However, the ionization process in a gas source mass spectrometer "scrambles" all isotopologues of N_2O, causing the inner (α) nitrogen and the outer

(β) nitrogen appear to be switched (Begun and Landau 1961). This scrambling causes the variance in SP from raw, measured data to be reduced in comparison to the variance of true SP values; therefore, this scrambling must be accounted for in order to report accurate SP measurements (Begun and Landau 1961). In this study, the most important scrambling behavior to correct for is that of $^{14}N^{15}N^{16}O$ and $^{15}N^{14}N^{16}O$.

Standardizing scrambling corrections across labs is an ongoing area of research (Ostrom et al. 2018; Mohn et al. 2014; Mohn et al. 2022; Westley et al. 2007). There are largely three levels of complexity that the scrambling correction can be performed at: i) A single-factor correction (γ) that assumes the scrambling behavior of $^{14}N^{15}N^{16}O$ and $^{15}N^{14}N^{16}O$ are equal, and that the contribution of ^{17}O is negligible at Mass 31 (Toyoda and Yoshida 1999; Kaiser et al. 2004); ii) A two-factor correction (γ and κ) that accounts for the difference in scrambling between $^{14}N^{15}N^{16}O$ and $^{15}N^{14}N^{16}O$, and assumes that ^{17}O follows a mass-dependent relationship with ^{18}O (Frame and Casciotti 2010; Kelly et al. 2023); and iii) A nine-factor correction that accounts for differences in scrambling between $^{14}N^{15}N^{16}O$, $^{15}N^{14}N^{16}O$, $^{15}N^{15}N^{16}O$, $^{14}N^{14}N^{17}O$, $^{14}N^{15}N^{17}O$, and $^{15}N^{14}N^{17}O$ (Westley et al. 2007).

We used the single-factor correction following (Toyoda and Yoshida 1999; Yoshida and Toyoda 2000) because the nine-factor scrambling correction (Westley et al. 2007) requires measurement of up to nine external reference gasses, which we did not have, and because we believe the scrambling effects of $^{15}N^{15}N^{16}O$, $^{14}N^{14}N^{17}O$, $^{14}N^{15}N^{17}O$, and $^{15}N^{14}N^{17}O$ are negligible at the level of precision needed for this study – i.e. the variations in SP between NOR and Fhp are on the order of 10‰. We did not use the two-factor correction following (Frame and Casciotti 2010; Kelly et al. 2023) because we were able to mass resolve ^{17}O directly, and because that method is optimized for continuous-flow SP measurements.

We measured two replicates (RM5_1 and RM5_2) of external reference gas RM5 (Mohn et al. 2022) on the Production Ultra to calculate the scrambling factor (Table 4.10). Replicates were measured three months apart, and RM5_2 was measured at a lower sample amount. RM5 was used because it has a large, ~10‰ difference in $\delta^{15}N$ between $\delta^{15}N^{bulk}$ and $\delta^{15}N^{\alpha}$ compared to RM1B and RM3B, which had similar values of $\delta^{15}N^{bulk}$ and $\delta^{15}N^{\alpha}$ within uncertainty (Table 4.8). We measured $\delta^{15}N^{bulk}$ within uncertainty for RM5_2, and close to within uncertainty for RM5_1 (Table 4.10). We consistently measured the mean $\delta^{15}N^{\alpha}$ value to be more depleted by roughly 1‰, though all measured $\delta^{15}N^{\alpha}$ values overlapped with the reported $\delta^{15}N^{\alpha}$ value within uncertainty. This cause the SP values of RM5_1 and RM5_2 to be 'compressed' towards 0‰ compared to its reported value. $\delta^{18}O$ values for RM5_1 and RM5_2 were measured to be their reported values within uncertainty (Table 4.10).

Value	RM5 Reported	RM5_1 Measured	RM5_2 Measured
$\delta^{15}N^{bulk}$	33.44 ± 0.05‰	34.26 ± 0.36‰	34.01 ± 0.53‰
$\delta^{15}N^{\alpha}$	43.54 ± 0.91‰	42.80 ± 1.27‰	42.54 ± 1.99‰
$\delta^{15}N^{\beta}$	23.34 ± 1.29‰	25.72 ± 1.32‰	25.48 ± 2.06‰
$\delta^{15}N^{SP}$	20.2 ± 0.91‰	17.08 ± 1.83‰	17.06 ± 2.86‰
$\delta^{18}O$	39.52 ± 0.15‰	39.31 ± 0.36‰	39.41 ± 0.48‰

Table 4.10. Characterization of RM5.
Values for RM5 are taken from (Mohn et al. 2022). RM5_2 was measured three months after RM5_1 and at a lower sample amount, which caused RM5_2 to have larger measurement uncertainties overall. All values are reported as mean ± s.d. and versus AIR.

We then followed (Toyoda and Yoshida 1999) to calculate $\gamma = 0.04 \pm 0.08$ for RM5_1, and $\gamma = 0.05 \pm 0.11$ for RM5_2. We therefore use an average γ value of 0.045 ± 0.136 for samples measured on the Production Ultra. (Magyar 2017; Magyar et al. 2016) used the Prototype Ultra and measured samples in similar tuning conditions as used in this study; they used a one-factor correction of 0.110 ± 0.002. We therefore used $\gamma = 0.110 \pm 0.002$ for samples measured on the Prototype Ultra. γ was likely lower on the Production Ultra because it has a lower baseline source pressure than the Prototype Ultra (3×10^{-10} vs. 9×10^{-8} mbar respectively).

The one-factor scrambling correction was performed as follows; an example correction is shown for one measurement of iFhp (Table 4.11). First, the sample is measured vs. Caltech Ref Gas. Next, values are corrected to international standards (AIR for N, VSMOW for O) using values reported by Tokyo Tech values (Table 4.7). Finally, the scrambling-adjusted $^{15}R^{\alpha}$ value ($^{15}R^{\alpha}_{adj}$) is calculated from the measured value ($^{15}R^{\alpha}_{meas}$) and the measured bulk value (R^{bulk}_{meas}):

$$R^{\alpha}_{adj} = \frac{R^{\alpha}_{meas} - 2\gamma R^{bulk}_{meas}}{-2\gamma + 1}$$

Equation 4.19

The final reported values, with scrambling correction and reported vs. AIR, are shown in the rightmost column of Table 4.11.

Value	Vs. Caltech Ref Gas	Vs. Intl (No scrambling corr.)	Final reported values
$\delta^{15}N^{bulk}$	-97.39 ± 0.18‰	-93.59 ± 0.18‰	-93.59 ± 0.18‰
$\delta^{15}N^{\alpha}$	-95.53 ± 0.51‰	-88.72 ± 0.51‰	-88.54 ± 0.51‰
$\delta^{15}N^{\beta}$	-99.25 ± 0.54‰	-98.46 ± 0.54‰	-98.65 ± 0.54‰
$\delta^{15}N^{SP}$	1.86 ± 0.74‰	4.87 ± 0.74‰	10.11 ± 0.54‰
$\delta^{18}O$	-16.58 ± 0.25‰	22.72 ± 0.25‰	22.72 ± 0.25‰

Table 4.11. Example of scrambling correction.
Values are reported as mean ± s.e. for one measurement of iFhp on the Production Ultra using γ = 0.045. The raw measurement ("Vs. Caltech Ref Gas") is first corrected to the international standard of AIR or VSMOW ("Vs. Intl (No scrambling corr."), and then the scrambling correction is applied ("Final Reported Values," Eqn. 4.19).

We checked our scrambling-corrected values against previously reported *in vitro* values for NOR. *ΔnosZΔfhp*, which only has NOR, was corrected using γ = 0.110 and iNOR was corrected using γ = 0.045. All corrected values overlap with previous *in vitro* measurements of a NOR enzyme purified from *Paracoccus denitrificans* ATCC 35512 (Yamazaki et al. 2014) (Fig. 4.15). In addition, as an internal check, the two samples measured on both the Production and Prototype Ultras (0225 and 0230, Fig. 4.14) gave similar values, implying that the scrambling factors are similar on both instruments. Indeed, γ = 0.045 ± 0.136 for the Production Ultra and γ = 0.110 ± 0.002 are similar within uncertainty.
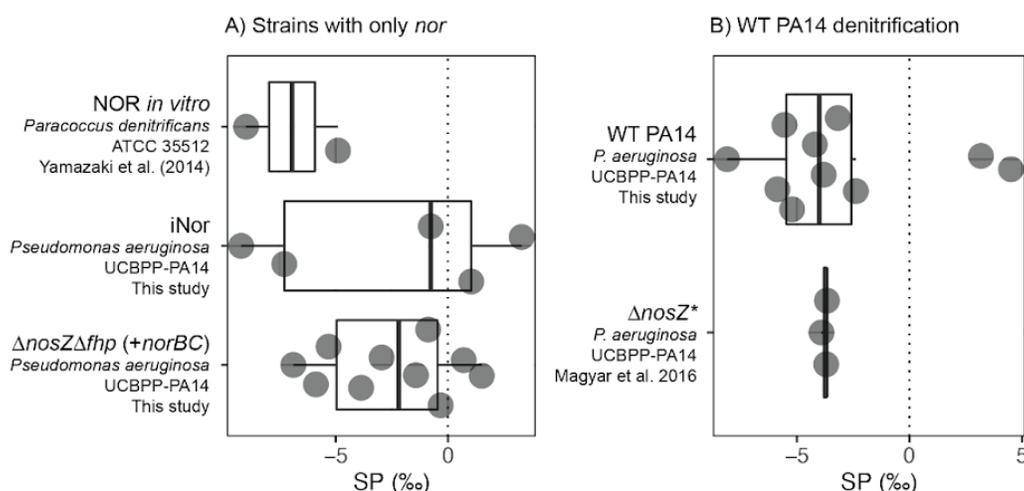
Fig. 4.15. Comparing scrambling corrected values with prior studies.
(A) Scrambling-corrected SP values for strains with only NOR (iNor, Δ*nosZ*Δ*fhp*) compared to previously published values by (Yamazaki et al. 2014) of an *in vitro* NOR purified from *Paracoccus denitrificans* ATCC 35512. To accurately compare data across studies, each data point shows one biological replicate; data from (Yamazaki et al. 2014) is presented as the average of Experiments A and the single data point from Experiment C from their study. Δ*nosZ*Δ*fhp*, measured on the Prototype Ultra, was corrected using $\gamma = 0.110$ and iNOR, measured on the Production Ultra, was corrected using $\gamma = 0.045$. **(B)** Comparison of scrambling-corrected WT PA14 SP values from this study and (Magyar et al. 2016), both grown in similar batch culture, denitrifying conditions. The strain used in (Magyar et al. 2016) was reported as "*Pseudomonas aeruginosa* strain PA14 Δ*nosZ*," but we found through PCR amplification of the *nosZ* gene location that the deletion was not successful – therefore, this strain is the same as our study and is the WT strain. This is indicated as "Δ*nosZ**" in the figure. Values from this study and (Magyar et al. 2016) were both measured on the Prototype Ultra and corrected using $\gamma = 0.110$.

4.7.6 Isotopic composition of DETA NONOate

We calculated the average $\delta^{15}N$ of the initial NO reactant used in the suspension assays by measuring the difference in $\delta^{15}N$ between the full and decomposed NO-donor, DETA NONOate (#82120, Cayman Chemical Company). DETA NONOate ($C_4H_{13}N_5O_2$) is a pH-dependent NO-donor that decays following first order kinetics to release two moles of NO per mole of DETA NONOate. At pH 7.4, it has a half life of 20 hours at 37°C and a half life of 56 hours at 22-25°C. The manufacturer states that DETA NONOate is very stable in alkaline solutions, so 500 mM stock solutions at pH 11.7 (adjusted with NaOH) were kept at -80°C until use.

To measure the $\delta^{15}N$ of the full DETA NONOate molecule (5 Nitrogens), 8 μL of 500 mM DETA NONOate stock was added to 12 μL of 10 mM NaOH to create a final concentration of 200 mM DETA NONOate at pH 11.8. Three 5 μL aliquots of this solution were pipetted into individual 4x6 mm pressed tin capsules (Costech Analytical Technologies) and left to dry overnight. Capsules were then sealed closed for analysis on an EA-IRMS (Elemental Analyzer Isotope Ratio Mass Spectrometer).

To measure the $\delta^{15}N$ of the decayed DETA NONOate molecule (3 Nitrogens), 10 μL of 500 mM DETA NONOate stock was added to 10 μL of 10 mM NaOH to create a final concentration of 250 mM DETA NONOate at pH 11.7. Then, 0.5 μL of 12M HCl was added to acidify the solution to pH 0.54. Small bubbles could be seen forming immediately after adding the acid. This solution was allowed to sit on the bench top at room temperature (≈22°C) overnight. Then, three 5 μL aliquots of this solution were pipetted into individual 4x6 mm pressed tin capsules, left to dry overnight, and then sealed for analysis on an EA-IRMS.

We also prepared two NaOH and HCl blanks. Two 5 μL aliquots of 10 mM NaOH and two 1 μL aliquots of 12M HCl were pipetted into individual 4x6 mm tin capsules, left to dry overnight, and then sealed for further analysis on an EA-IRMS.

Samples were then analyzed for $\delta^{15}N$ on a Delta-V Advantage with Gas Bench and Costech elemental analyzer, where N-containing phases of the sample are fully combusted and then subsequently reduced to N_2 for isotopic measurement. Before measuring samples, two tests were performed to ensure the instrument was functioning normally: i) An 'on/off' test where an internal N_2 standard was opened and closed to ensure instrument sensitivity and to establish a baseline intensity at a 'zero' N_2 concentration, and ii) A linearity test where the concentration of the internal N_2 standard was increased linearly within the designated sensitivity range of the instrument to ensure that a linear increase in N_2 concentration corresponds to a linear increase in electrical signal on the collector cups. We measured at three masses (28-30 amu). The instrument was also tuned to ensure that each mass was measured at the center of its mass peak.

Three analytical replicates each of the full and decayed DETA NONOate molecule were run, while two analytical replicates each of the NaOH and HCl blanks were run. All samples were bracketed at the beginning and end of the run by a suite of external isotope standards (Urea $\delta^{15}N$ = 0.0‰; Acetanilide $\delta^{15}N$ = 19.56 ± 0.03‰) and tin capsule blanks. Each sample measurement itself was also internally bracketed by four total repetitions (two before, two after) of the internal N_2 standard. After sample analysis, each sample peak was visually inspected to ensure peaks were not saturating the detector cups (i.e. within instrument's operating range) and that the N_2 peak detected was eluting at the correct time (235 seconds in the run method used). Sample peaks that did not meet this criteria were discarded, though in this run no peaks were discarded.

Data was then processed to correct for blanks and to correct measured $\delta^{15}N$ to finalized $\delta^{15}N$ values. In this run, all blanks (tin capsule, NaOH, HCl) were so small that they were negligible and were not detected by the instrument's auto peak-detect software. Because the volumes were so small, an accurate $\delta^{15}N$ and peak area could not be determined so the blank's contribution to $\delta^{15}N$ and total N is neglected. $\delta^{15}N$ values were then corrected using the external Urea and Acetanilide standards run; a linear regression was made using these standards and the correction applied to all samples. On average, the correction decreased the measured $\delta^{15}N$ by 0.2‰.

We then calculated the average $\delta^{15}N$ of the released NO molecules by mass balance using the equation:

$$2 * \delta^{15}N_{2N} = 5 * \delta^{15}N_{5N} - 3 * \delta^{15}N_{3N}$$

<div align="right">Equation 4.20</div>

Where $\delta^{15}N_{2N}$, $\delta^{15}N_{5N}$, and $\delta^{15}N_{3N}$ refer to the average $\delta^{15}N$ of the released NO molecules, the full DETA NONOate molecule, and the decayed DETA NONOate molecule respectively. All results are reported in Table S11.

Sample ID (# of N)	δ^{15}N (‰)
Full Donor (5 N)	-22.95 ± 0.15‰
Decomposed Donor (3 N)	-23.54 ± 0.24‰
Released N (2 N)	-22.08 ± 0.29‰

Table 4.12. Isotopic composition of DETA NONOate
Measured δ^{15}N values of the full and decomposed NO-donor, DETA NONOate. The δ^{15}N of the released nitrogens was calculated by mass balance. Values represent mean ± s.d. of three replicates.

4.7.7 Nitrate and batch culture aliquots

At each sampling point for batch culture assays where an N_2O measurement was performed (end-exponential and end-stationary), a ≈5 mL aliquot of the liquid culture was taken and immediately flash frozen in liquid nitrogen. Aliquots were then kept frozen at -80°C until isotopic analysis on an EA-IRMS. In addition, for each sampling batch an aliquot of the added 1 M KNO_3 stock, the SCFM Amended media, DI water, and 1 M sodium succinate stock were flash frozen and stored as well. When samples were ready for analysis on the EA-IRMS, all flash frozen aliquots were thawed at room temperature and aliquots were pipetted in triplicate into individual 5x9 mm pressed tin capsules (Costech Analytical Technologies) and left to dry overnight.

Samples were analyzed for δ^{15}N in a similar manner as DETA NONOate above on a Delta-V Advantage with Gas Bench and Costech elemental analyzer. Raw measurement outputs were also blank corrected in the same way. Some tin capsule blanks, especially those early in the run, were so small that an accurate δ^{15}N and peak area could not be determined; therefore, their contribution to δ^{15}N and total N is neglected. However, some tin capsules contributed small amounts to δ^{15}N the total N and this contribution was corrected out using the formula:

$$\delta_s = \frac{n_t \delta_t - n_b \delta_b}{n_t - n_b}$$

<div align="right">Equation 4.21</div>

Where s indicates the corrected sample value, t indicates total (i.e. the measured value before correction) and b indicates blank. δ indicates δ^{15}N value and

n indicates amount of sample. Here, we used the total peak area of mass 28 and 29 for *n*.

Peak area all was also corrected for tin blanks:

$$n_s = n_t - n_b$$

<div align="right">Equation 4.22</div>

$$\sigma_{n_s} = \sqrt{\sigma_{n_t}^2 + \sigma_{n_b}^2}$$

<div align="right">Equation 4.23</div>

After correcting all samples and standards for $\delta^{15}N$ and Peak Area All (Mass 28 and 29) using the tin capsule blanks, standards were used to correct sample $\delta^{15}N$ values to reportable values and total moles of N. The acetanilide standard was dissolved in acetone and the urea standard was dissolved in water for pipetting; tin capsules were then dried before analysis. Acetone and water blanks were prepared and measured, where only acetone or water was pipetted into a tin capsule and let dry. These blanks looked just like the tin capsule blanks (similar peak area, similar $\delta^{15}N$) so we concluded that the main blank contribution was by the capsules themselves and not the acetone or water.

$\delta^{15}N$ values were then corrected using the external Urea and Acetanilide standards run; a linear regression was made using these standards and the correction applied to all samples. On average, the correction decreased the measured $\delta^{15}N$ by 0.2‰. A similar linear regression was made using the Peak Area All measurement to calculate total N amounts.

The $\delta^{15}N$ of nitrate and of the total N in SCFM Amended of each batch are shown in Table 4.5. SCFM Amended has multiple sources of nitrogen, primarily from amino acids (Palmer et al. 2007). Sodium succinate was also measured but did not have any detectable amounts of N. The fraction of total N remaining was calculated by dividing the total N measured by the total N initially added.

4.7.8 Rayleigh curves

Prior studies modeling SP in bacteria have adapted Rayleigh curves (Mariotti et al. 1981) to calculate the isotopic fractionation of SP, termed ε^{SP}, or ε^{456} and ε^{546}. ("456" and "546" denote $^{14}N^{15}N^{16}O$ and $^{15}N^{14}N^{16}O$ respectively; (Kantnerová et al. 2022; Sutka et al. 2006)) so that these ε values can then be used in a Hayes-style model of isotope flux modeling (Hayes 2001). Prior studies have either plotted SP ((Sutka et al. 2006)) or delta values of specific isotopocules (i.e. δ^{456} denoting $\delta^{15}N$ of $^{14}N^{15}N^{16}O$ vs. $^{14}N_2^{16}O$ (Kantnerová et al. 2022)) vs. -(f*lnf)/(1-f), where *f* is the fraction of the remaining substrate.

We constructed Rayleigh curves for batch culture assays of NOR-only strains, *ΔnosZΔfhp* and iNOR, using *f* values calculated from degree of nitrate consumption derived from $\delta^{15}N$ measurements of media aliquots as described in

Section 5.2 ($f_{nitrate}$) or amount of N$_2$O production compared to amount of nitrate initially added (f_{N2O}). $f_{nitrate}$ was calculated based on the total amount of N measured through EA-IRMS. However, $f_{nitrate}$ reflects the fraction of total N remaining in solution because N from nitrate vs. SCFM Amended could not be distinguished in this analysis, though most of the initial nitrogen comes from nitrate (Table 4.5). f_{N2O} is defined as:

$$f_{N2O} = 1 - \frac{2 * n_{N2O}}{n_{NO3-}}$$

Equation 4.24

Where n_{N2O} are the moles of N$_2$O produced and n_{NO3-} are the moles of nitrate initially added. This equation assumes that every mole of nitrate that is taken up by the denitrification pathway results in two moles of NO; this is the same assumption used by (Kantnerová et al. 2022).

Fitted values for *m* and *b* gave very large uncertainties and were of low confidence (Fig. 4.16). In particular, fitted values using f_{N2O} had extremely large uncertainties due to the narrow range of the x-axis – i.e. nitrate was given at saturating conditions so the nitrate pool was not very depleted, resulting in f_{N2O} values around 0.99. $f_{nitrate}$ varied over a larger range, likely because not all the nitrate consumed ended up as N$_2$O (i.e. due to assimilatory nitrate processes, or from loss along the denitrification pathway), so the approach for calculating *f* in Eqn. 4.24 likely gives an overestimate. However, overall no variation in SP is seen in the NOR-only strains, *Δ*nosZ*Δ*fhp and iNOR.
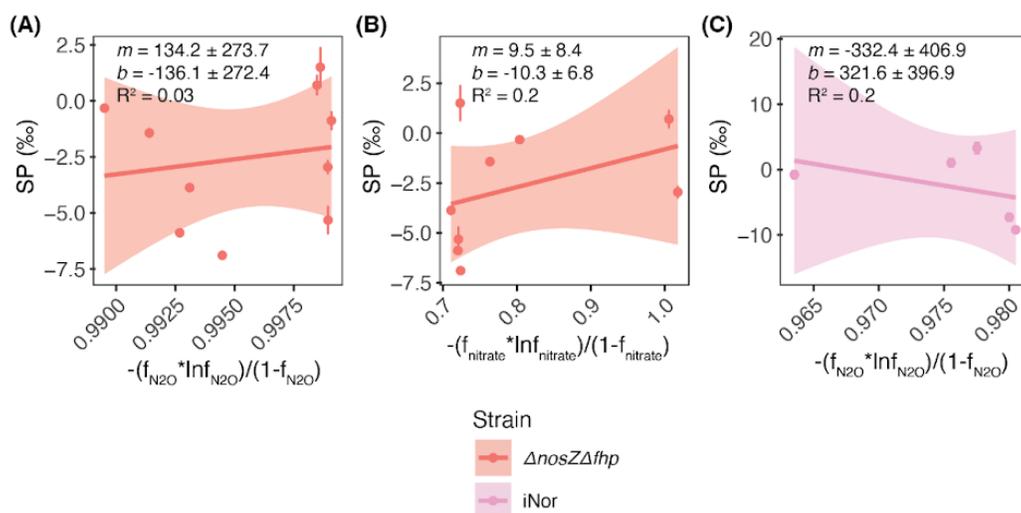
Fig. 4.16. Rayleigh plots of NOR-only strains.
Rayleigh plots for strains with only NOR – $\Delta nosZ\Delta fhp$ **(A,B)** and iNor **(C)**. f (fraction of substrate remaining) was calculated either from the amount of nitrate remaining in batch culture ($f_{nitrate}$) or based on moles of N_2O produced compared to the amount of nitrate initially added (f_{N2O}). Results of linear regression are shown in the upper right corner of each plot; all analyses and data visualization were performed using R Statistical Software (v4.1.0; (R Core Team 2021)) and the ggplot2 package (v3.3.6; (Wickham et al. 2016)).

4.7.9 AnnoTree Search Parameters

A phylogram of species with annotated Fhp/Hmp sequences was first made from the NCBI database (Fig. 4.17). Phylogram was made to include representative strains from a range of known pathogens. The amino acid sequence of Fhp from WT PA14 was used (PA14_29640). Two strains with a high and low sequence similarity were selected for further N_2O screening and SP measurement. Fhp from *S. aureus* shows 31.6% sequence similarity to Fhp from *P. aeruginosa*, while Fhp from *A. baumannii* shows 98.5% similarity. Next, Fhp and NorBC were queried from AnnoTree, a functionally annotated database of >27,000 bacterial and >1,500 archaeal genomes (Mendler et al. 2019). Since Fhp from *S. aureus* shows 31.6% sequence similarity to Fhp from *P. aeruginosa*, the default search parameters were used: % identity: 30; E value: 0.00001; % subject alignment: 70; % query alignment: 70. Results are shown in Tables 4.12-4.14 at the phylum level.
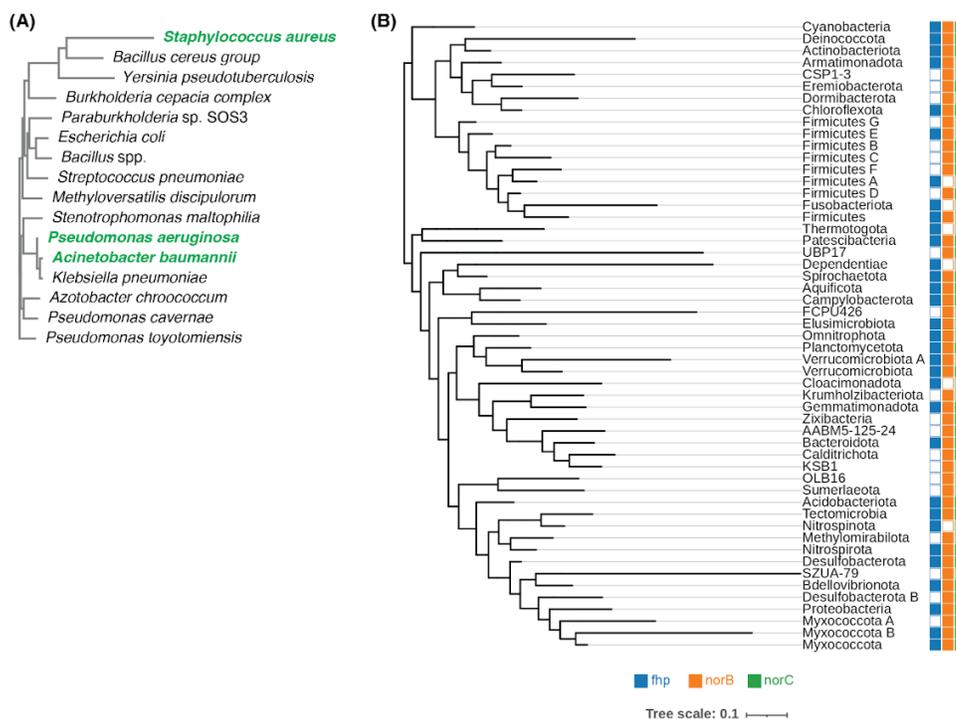
Fig. 4.17. Phylogeny of Fhp in Bacteria.
(A) Phylogram of annotated Fhp/Hmp amino acid sequences in the NCBI database; phylogram was curated to show a representative group of pathogenic bacteria. Strains in green were measured for SP in this study. **(B)** Tree showing abundance of Fhp, NorB and NorC across Bacteria at the Phylum level, annotated in AnnoTree (Mendler et al. 2019) and visualized using the interactive tree of life (iTOL). Search parameters for AnnoTree were: % identity: 30; E value: 0.00001; % subject alignment: 70; % query alignment: 70.

Bacterial Phylum	Number of genome hits	Proportion of all hits	Number of genomes in clade
Proteobacteria	3761	52.90%	9474
Myxococcota	28	0.39%	168
Myxococcota_B	1	0.01%	2
Bdellovibrionota	4	0.06%	110
Desulfobacterota	4	0.06%	560
Nitrospirota	3	0.04%	138
Nitrospinota	1	0.01%	22
Tectomicrobia	1	0.01%	4
Acidobacteriota	14	0.20%	380
Bacteroidota	318	4.47%	3781
Gemmatimonadota	4	0.06%	101
Cloacimonadota	1	0.01%	27
Verrucomicrobiota	36	0.51%	478
Verrucomicrobiota_A	2	0.03%	52
Planctomycetota	52	0.73%	376
Omnitrophota	2	0.03%	83
Elusimicrobiota	3	0.04%	66
Campylobacterota	60	0.84%	323
Aquificota	4	0.06%	39
Spirochaetota	17	0.24%	310
Dependentiae	1	0.01%	26
Patescibacteria	13	0.18%	1131
Thermotogota	1	0.01%	63
Firmicutes	1086	15.28%	2737
Fusobacteriota	3	0.04%	70
Firmicutes_A	62	0.87%	2636
Firmicutes_E	2	0.03%	39
Actinobacteriota	1492	20.99%	4261
Deinococcota	10	0.14%	92
Chloroflexota	31	0.44%	520
Armatimonadota	1	0.01%	36
Cyanobacteria	91	1.28%	727
Total:	7109	1	28832

Table 4.13. Fhp AnnoTree query results in Bacteria.
Fhp (KEGG ID K05916) query results in AnnoTree (Mendler et al. 2019) at the phylum
level for Bacteria. Default search parameters were used: % identity: 30; E value: 0.00001;
% subject alignment: 70; % query alignment: 70.

Bacterial Phylum	Number of genome hits	Proportion of all hits	Number of genomes in clade
Proteobacteria	1381	74.49%	9474
Myxococcota	7	0.38%	168
Myxococcota_B	1	0.05%	2
Myxococcota_A	3	0.16%	16
Desulfobacterota_B	1	0.05%	28
CG2-30-70-394	1	0.05%	
Bdellovibrionota	11	0.59%	110
Desulfobacterota	22	1.19%	560
Nitrospirota	20	1.08%	138
Tectomicrobia	1	0.05%	4
CG2-30-53-67	1	0.05%	
Acidobacteriota	6	0.32%	380
Bacteroidota	228	12.30%	3781
AABM5-125-24	1	0.05%	5
Zixibacteria	2	0.11%	16
Verrucomicrobiota	8	0.43%	478
Planctomycetota	4	0.22%	376
Campylobacterota	53	2.86%	323
Aquificota	4	0.22%	39
Spirochaetota	44	2.37%	310
UBP17	1	0.05%	2
Patescibacteria	1	0.05%	1131
Firmicutes_D	1	0.05%	67
Firmicutes_F	2	0.11%	38
Firmicutes_C	11	0.59%	225
Firmicutes_B	24	1.29%	183
Actinobacteriota	1	0.05%	4261
Deinococcota	8	0.43%	92
Eremiobacterota	3	0.16%	34
Cyanobacteria	3	0.16%	727
Total:	1854	1	22968

Table 4.14. NorB and NorC AnnoTree query results in Bacteria.
NorB (KEGG ID K04561) and NorC (KEGG ID K02305) query results in AnnoTree (Mendler et al. 2019) at the phylum level for Bacteria. Default search parameters were used: % identity: 30; E value: 0.00001; % subject alignment: 70; % query alignment: 70.

Bacterial Phylum	Number of genome hits	Proportion of all hits	Number of genomes in clade
Proteobacteria	524	92.74%	9474
Myxococcota	1	0.18%	168
Myxococcota_B	1	0.18%	2
Bacteroidota	25	4.42%	3781
Verrucomicrobiota	3	0.53%	478
Planctomycetota	1	0.18%	376
Campylobacterota	5	0.88%	323
Aquificota	1	0.18%	39
Spirochaetota	3	0.53%	310
Cyanobacteria	1	0.18%	727
Total:	565	1	15678

Table 4.15. Fhp, NorB and NorC AnnoTree query results in Bacteria.
Fhp (KEGG ID K05916), NorB (KEGG ID K04561) and NorC (KEGG ID K02305) query results in AnnoTree (Mendler et al. 2019) at the phylum level for Bacteria. Default search parameters were used: % identity: 30; E value: 0.00001; % subject alignment: 70; % query alignment: 70.
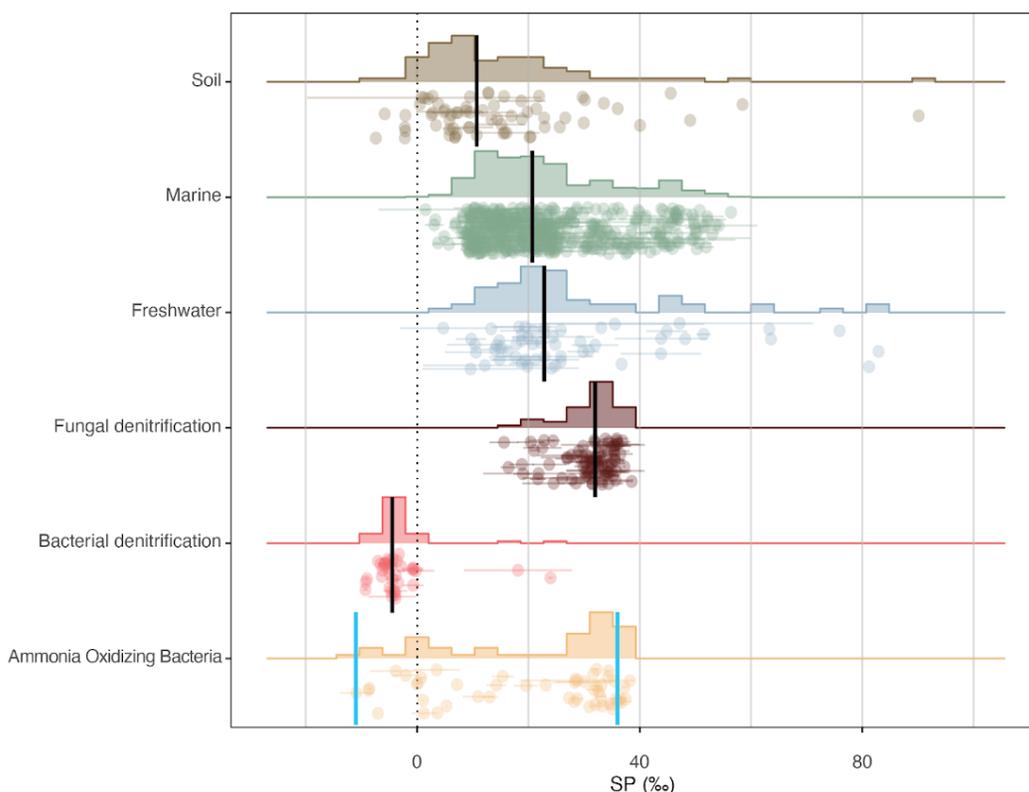
Fig. 4.18. Full literature compilation of environmental and end-member SP values.
Literature compilation of environmental (31 references; *n* = 622 (Kelly et al. 2023; Sasaki et al. 2011; Toyoda et al. 2009; Well et al. 2005; Koba et al. 2009; Well et al. 2012; Wong et al. 2021; Mander et al. 2014; Li et al. 2022; Westley et al. 2006; Toyoda 2002; Fujii et al. 2013; Popp et al. 2002; Yamagishi et al. 2007; Farías et al. 2009; Charpentier et al. 2007; Gluschankoff et al. 2023; Kelly et al. 2021; Monreal et al. 2022; Casciotti et al. 2018; Park et al. 2011; Pérez et al. 2001; Yamulki et al. 2001; Bol et al. 2003; Zou et al. 2014; Toyoda et al. 2011; Yano et al. 2014; Opdyke et al. 2009; Ostrom et al. 2010; Kato et al. 2013; Koehler et al. 2012)) and pure culture (15 references; *n* = 172 (Yamazaki et al. 2014; Toyoda et al. 2005; Sutka et al. 2003; Sutka et al. 2004; Kantnerová et al. 2022; Haslun et al. 2018; Magyar 2017; Sutka et al. 2008; Maeda et al. 2015; Rohe et al. 2014; Yang et al. 2014; Sutka et al. 2006; Frame and Casciotti 2010; Jung et al. 2014; Ostrom et al. 2007) SP data. Environmental data (Soil, Marine, Freshwater) are *in situ* measurements; therefore soil incubation studies were not included. Soil includes forest, cropland, grassland and wetlands. Freshwater includes lakes, rivers and groundwater. To the best of our ability, each data point in the Environmental data represents an individual measurement; for studies that did not report the full dataset, the mean ± s.d. is used instead (shown as circle with error bar). Biogenic endmembers (Fungal denitrification, Bacterial denitrification and Ammonia Oxidizing Bacteria (AOB)) are from *in vitro* studies of pure strains or enzymes. SP values of ammonia oxidizing archaea (AOA) are not included because measurements were performed on enrichment cultures rather than purified strains (Santoro et al. 2011; Jung et al. 2014), though they have SP values that lie within the positive spread of AOB studies (≈10-30‰). Each data point represents a unique biological replicate; for studies that did not report a full data set, the mean ± s.d. is used instead. Vertical black bar shows median; blue vertical bars of AOB indicate end-member values (roughly -11‰ for nitrifier-denitrification and 36‰ for NH_2OH decomposition) that the SP of AOB has been found to vary between based on growth conditions (Frame and Casciotti 2010) due to multiple pathways of N_2O formation (Stein 2019). All analyses and data visualization were performed using R Statistical Software (v4.1.0; (R Core Team 2021)) and the ggplot2 package (v3.3.6; (Wickham et al. 2016)).

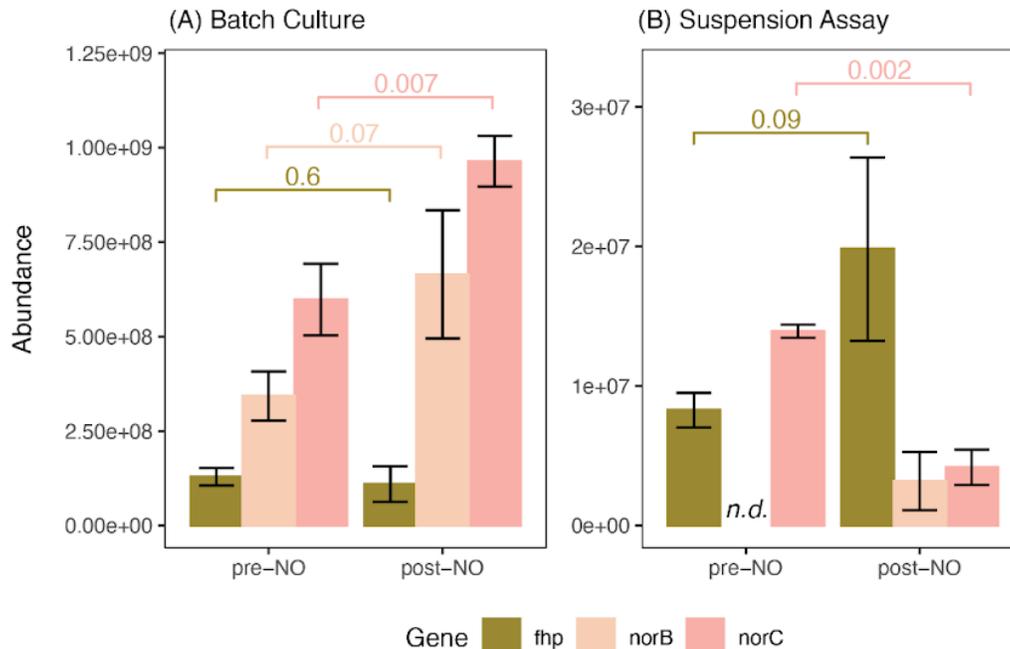**Fig. 4.19: Fhp, NorB and NorC protein abundances.**
(A) Protein abundances for *fhp*, *norB* and *norC* before and after NO-addition for WT PA14 grown in batch culture. The ratio of *fhp* to *norB* and *norC* is presented in the main text. *P* values were calculated using Welch's t-test for two independent groups and variance was not assumed to be the same across groups. See main text Fig. 2B for experimental set-up. **(B)** Results for WT PA14 grown in suspension assays; see main text Fig. 2C for experimental set-up. *P* value could not be calculated for *norB* in the suspension assay because it was not detected; detection limit was XX. For both panels, values represent the mean ± s.d. of three biological replicates. All analyses and data visualization were performed using R Statistical Software (v4.1.0; (R Core Team 2021)) and the ggplot2 package (v3.3.6; (Wickham et al. 2016)).
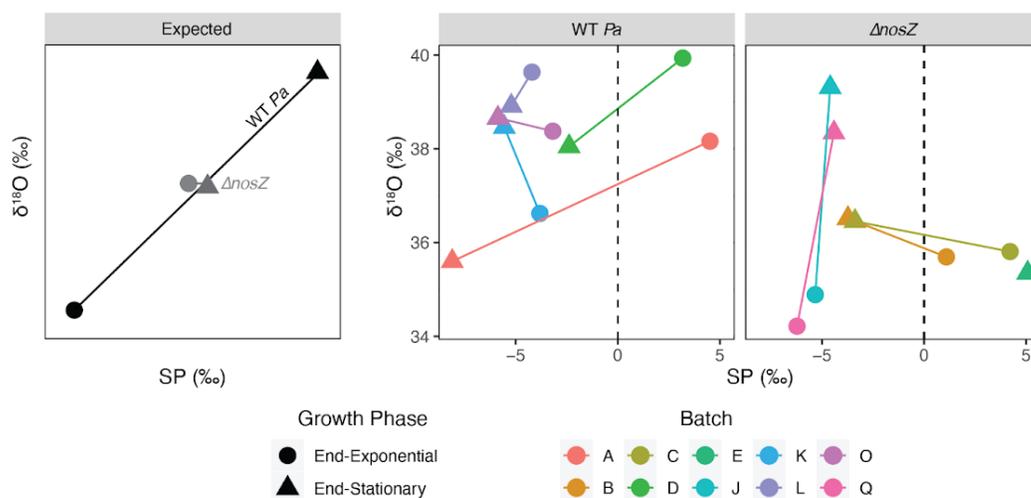
Fig. 4.20: SP and δ¹⁸O data for *ΔnosZ* and *ΔnosZΔfhp*.

Prior studies have shown that NOS can increase the SP and $\delta^{18}O$ of the residual N_2O pool through preferential cleavage of the ^{14}N-^{16}O vs. ^{15}N-^{18}O bond in N_2O (Casciotti et al. 2018; Ostrom et al. 2007). Therefore, it is expected that in batch culture conditions, end-stationary growth phase (triangles) data for WT *Pa* would be more positive in SP and $\delta^{18}O$ than end-exponential (circles), and that this trend should go away with deletion of the *nosZ* gene. However, WT *Pa* and *ΔnosZ* do not show consistent trends for both measurements.
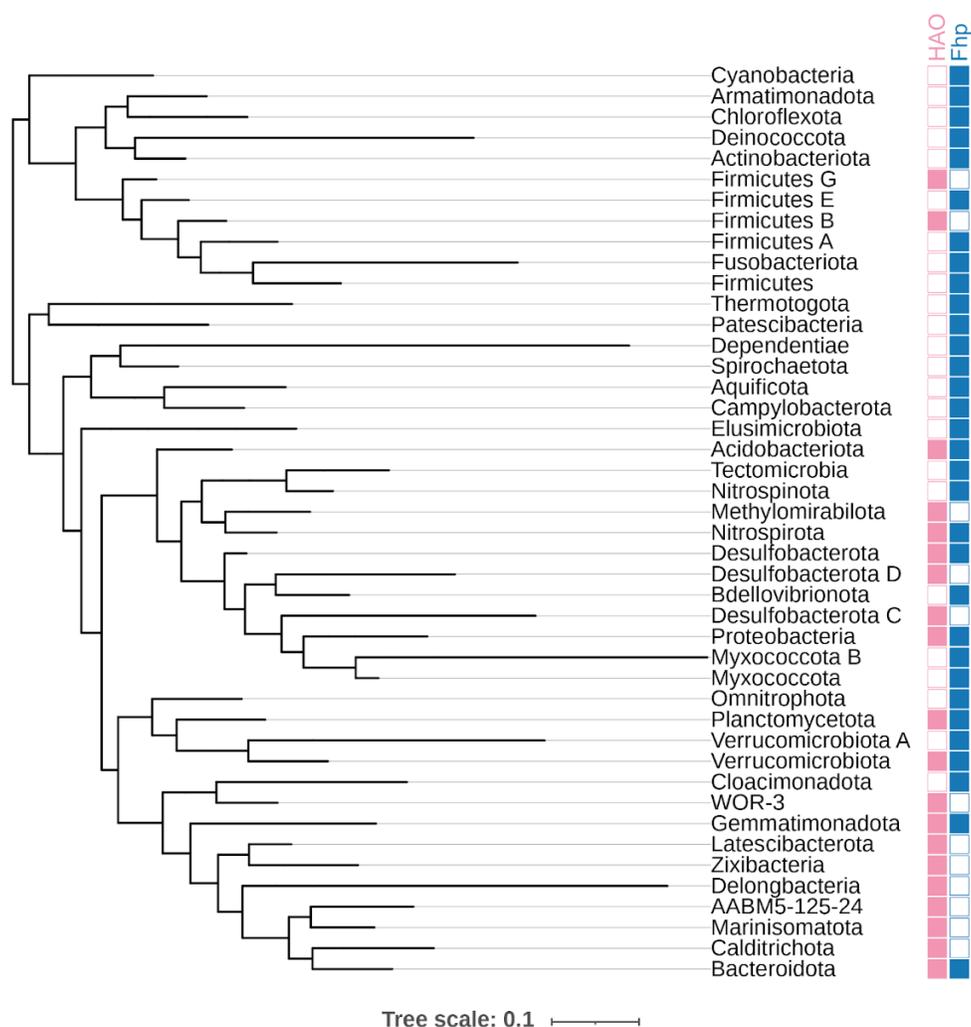
Fig. 4.21: Phylogeny of HAO and Fhp in Bacteria.
Genome hits for hydroxylamine oxidoreductase (HAO; K10535) and Fhp (K05916) in bacteria at the phylum level from AnnoTree (Mendler et al. 2019). Search parameters were used: % identity: 30; E value: 0.00001; % subject alignment: 70; % query alignment: 70. Results were visualized using the Interactive Tree of Life (iTOL); leaves without hits for HAO or Fhp were trimmed for clarity. HAO catalyzes the oxidation of NH_2OH to NO (Caranto and Lancaster 2017) and is used as a proxy for ammonia oxidizing bacteria.
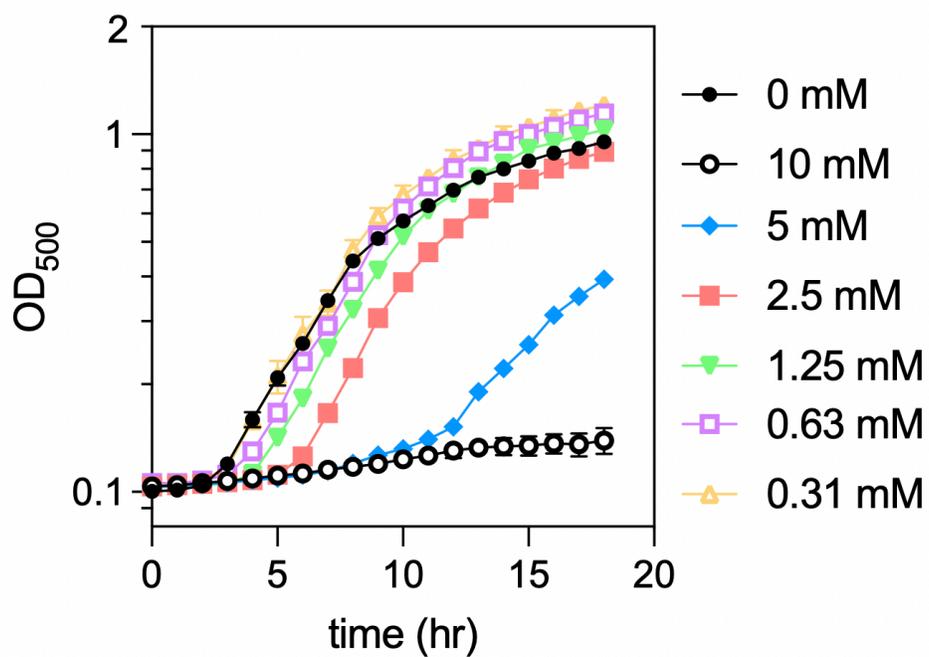
Fig. 4.22: Growth of WT *Pa* with DETA-NONOate titration.
WT *Pa* was grown in LB in the presence of DETA-NONOate with 2-fold dilutions ranging from ~300 μM to 10 mM and OD500 monitored over time. A concentration of less than 1 mM DETA-NONOate did not appreciably affect growth.

Species	SP References	Fhp accession number	NOR accession number
Pseudomonas aeruginosa	This study; Magyar et al. (2016); Magyar et al. (2017)	A0A0H2ZC95	A0A0H2ZLE2 (NorB); A0A0H2ZKE8 (NorC)
Pseudomonas fluorescens	Toyoda et al. (2005)	A0A448BJZ8 or A0A8H2RPK4	A0A0D0T5F4 (NorB); A0A0D0S4Z1 (NorC)
Paracoccus denitrificans	Toyoda et al. (2005); Ostrom et al. 2007	A1B2P2	Q51663 (NorB); Q51662 (NorC)
Pseudomonas chlororaphis; *Pseudomonas aureofaciens* subsp. nov., comb. nov.	Magyar et al 2017; Kantnerova et al 2022; Sutka et al. 2006; Haslun et al. 2018	A0A5M7CAB6	Q9F0W6 (NorB); Q9F0W7 (NorC)
Pseudomonas stutzeri (*Stutzerimonas stutzeri*)	Ostrom et al 2007	Q5W5T4	P98008 (NorB); Q52527 (NorC)

Table 4.16. Fhp and NOR accession numbers for previously measured bacterial denitrifiers. The accession number for Fhp or NorB and NorC of denitrifying strains used in prior SP studies. A close strain relative, whose genome has been sequenced, was used. Fhp is also annotated as Hmp or NOD (nitric oxide dioxygenase). (Ostrom et al. 2007) used "*Pseudomonas stutzeri* (provided by J. M. Tiedje)" and *Pseudomonas denitrificans* ATCC 13867; *P. stutzeri* is also known as *Stutzerimonas stutzeri*. (Magyar et al. 2016; Magyar 2017) used *Pseudomonas aeruginosa* UCBPP-PA14 and *Pseudomonas aureofaciens* ATCC 13985. (Toyoda et al. 2005) used *Pseudomonas fluorescens* ATCC 13525 and *Paracoccus denitrificans* ATCC 17741 (also known as 19376). (Kantnerová et al. 2022) used *Pseudomonas aureofaciens* ATCC 13985. (Sutka et al. 2006) used *Pseudomonas aureofaciens* ATCC 13985 and *Pseudomonas chlororaphis* ATCC 43928. However, DNA-DNA hybridization experiments has led to the reclassification of *P. aureofaciens* into *P. chlororaphis* (Peix et al. 2007) – therefore the strain "*Pseudomonas aureofaciens* ATCC 13985" used by (Sutka et al. 2006; Kantnerová et al. 2022; Magyar 2017) is now a subspecies of *P. chlororaphis* with the proposed taxonomic name "*P. chlororaphis* subsp. *aureofaciens* subsp. nov., comb. nov. [with the type strain DSM 6698^T (=ATCC 13985^T=NCIMB 9030^T)] (Peix et al. 2007)." Therefore, *P. aureofaciens* and *P. chlororaphis* are grouped together in the graph above. (Haslun et al. 2018) specified that they use the strains "*Pseudomonas chlororaphis* subsp. chlororaphis (ATCC 43928; *P. chlororaphis*) and *Pseudomonas chlororaphis* subsp. *aureofaciens* (ATCC 13985; *P. aureofaciens*)."

Bacterial Phylum	Number of genome hits	Proportion of all hits	Number of genomes in clade
Proteobacteria	7	0.4375	9474
Verrucomicrobiota	3	0.1875	478
Planctomycetota	6	0.375	376
Total:	16	1	10328

Table 4.17: Fhp and HAO AnnoTree query results in Bacteria.
Fhp (K05916) and HAO (K10535) query results in AnnoTree (Mendler et al. 2019) at the phylum level for Bacteria. HAO was used as a proxy for ammonia oxidizing bacteria (AOB). Default search parameters were used: % identity: 30; E value: 0.00001; % subject alignment: 70; % query alignment: 70.

4.8 References

Begun, G M, and L Landau. 1961. "Mass Spectra and Metastable Transitions in Isotopic Nitrous Oxides." The Journal of Chemical Physics 35 (2): 547–51. https://doi.org/10.1063/1.1731966.

Bergkessel, Megan, David W Basta, and Dianne K Newman. 2016. "The Physiology of Growth Arrest: Uniting Molecular and Environmental Microbiology." Nature Reviews. Microbiology 14 (9): 549–62. https://doi.org/10.1038/nrmicro.2016.107.

Bodor, Attila, Naila Bounedjoum, György Erik Vincze, Ágnes Erdeiné Kis, Krisztián Laczi, Gábor Bende, Árpád Szilágyi, Tamás Kovács, Katalin Perei, and Gábor Rákhely. 2020. "Challenges of Unculturable Bacteria: Environmental Perspectives." Reviews in Environmental Science and Bio/Technology 19 (1): 1–22. https://doi.org/10.1007/s11157-020-09522-4.

Bol, R, S Toyoda, S Yamulki, J M B Hawkins, L M Cardenas, and N Yoshida. 2003. "Dual Isotope and Isotopomer Ratios of N2O Emitted from a Temperate Grassland Soil after Fertiliser Application." Rapid Communications in Mass Spectrometry 17 (22): 2550–56. https://doi.org/10.1002/rcm.1223.

Bonamore, Alessandra, and Alberto Boffi. 2008. "Flavohemoglobin: Structure and Reactivity." IUBMB Life 60 (1): 19–28. https://doi.org/10.1002/iub.9.

Caranto, Jonathan D, and Kyle M Lancaster. 2017. "Nitric Oxide Is an Obligate Bacterial Nitrification Intermediate Produced by Hydroxylamine Oxidoreductase." Proceedings of the National Academy of Sciences of the United States of America 114 (31): 8217–22. https://doi.org/10.1073/pnas.1704504114.

Casciotti, K L, M Forbes, J Vedamati, B D Peters, T S Martin, and C W Mordy. 2018. "Nitrous Oxide Cycling in the Eastern Tropical South Pacific as Inferred from Isotopic and Isotopomeric Data." Deep Sea Research Part II: Topical Studies in Oceanography 156 (October): 155–67. https://doi.org/10.1016/j.dsr2.2018.07.014.

Casciotti, Karen L, John Karl Böhlke, Matthew R McIlvin, Stanley J Mroczkowski, and Janet E Hannon. 2007. "Oxygen Isotopes in Nitrite: Analysis, Calibration, and Equilibration." Analytical Chemistry 79 (6): 2427–36. https://doi.org/10.1021/ac061598h.

Charpentier, J, L Farias, N Yoshida, N Boontanon, and P Raimbault. 2007. "Nitrous Oxide Distribution and Its Origin in the Central and Eastern South Pacific Subtropical Gyre." Biogeosciences 4 (5): 729–41. https://doi.org/10.5194/bg-4-729-2007.

Choi, Kyoung-Hee, and Herbert P Schweizer. 2006. "Mini-Tn7 Insertion in Bacteria with Single AttTn7 Sites: Example Pseudomonas Aeruginosa." Nature Protocols 1 (1): 153–61. https://doi.org/10.1038/nprot.2006.24.

Cook, Gregory M, Chris Greening, Kiel Hards, and Michael Berney. 2014. "Energetics of Pathogenic Bacteria and Opportunities for Drug Development." Advances in Microbial Physiology 65 (November): 1–62. https://doi.org/10.1016/bs.ampbs.2014.08.001.

Denk, Tobias R.A., Joachim Mohn, Charlotte Decock, Dominika Lewicka-Szczebak, Eliza Harris, Klaus Butterbach-Bahl, Ralf Kiese, and Benjamin Wolf. 2017. "The Nitrogen Cycle: A Review of Isotope Effects and Isotope Modeling Approaches." Soil Biology and Biochemistry 105 (February): 121–37. https://doi.org/10.1016/j.soilbio.2016.11.015.

Eiler, John M., Matthieu Clog, Paul Magyar, Alison Piasecki, Alex Sessions, Daniel Stolper, Michael Deerberg, Hans-Juergen Schlueter, and Johannes Schwieters. 2013. "A High-Resolution Gas-Source Isotope Ratio Mass Spectrometer." International Journal of Mass Spectrometry 335 (February): 45–56. https://doi.org/10.1016/j.ijms.2012.10.014.

Farías, Laura, Maribeb Castro-González, Marcela Cornejo, José Charpentier, Juan Faúndez, Narin Boontanon, and Naohiro Yoshida. 2009. "Denitrification and Nitrous Oxide Cycling within the Upper Oxycline of the Eastern Tropical South Pacific Oxygen Minimum Zone." Limnology and Oceanography Letters 54 (1): 132–44. https://doi.org/10.4319/lo.2009.54.1.0132.

Ferousi, Christina, Sean H Majer, Ida M DiMucci, and Kyle M Lancaster. 2020. "Biological and Bioinspired Inorganic N-N Bond-Forming Reactions." Chemical Reviews 120 (12): 5252–5307. https://doi.org/10.1021/acs.chemrev.9b00629.

Frame, C H, and K L Casciotti. 2010. "Biogeochemical Controls and Isotopic Signatures of Nitrous Oxide Production by a Marine Ammonia-Oxidizing Bacterium." Biogeosciences 7 (9): 2695–2709. https://doi.org/10.5194/bg-7-2695-2010.

Fujii, Ayako, Sakae Toyoda, Osamu Yoshida, Shuichi Watanabe, Ken'ichi Sasaki, and Naohiro Yoshida. 2013. "Distribution of Nitrous Oxide Dissolved in Water Masses in the Eastern Subtropical North Pacific and Its Origin Inferred from Isotopomer Analysis." Journal of Oceanography 69 (2): 147–57. https://doi.org/10.1007/s10872-012-0162-4.

Gibson, Daniel G, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." Nature Methods 6 (5): 343–45. https://doi.org/10.1038/nmeth.1318.

Gluschankoff, Noah, Alyson E Santoro, Carolyn Buchwald, and Karen L Casciotti. 2023. "Shifts in the Isotopic Composition of Nitrous Oxide between El Niño and La Niña in the Eastern Tropical South Pacific." Authorea, Inc., February. https://doi.org/10.22541/essoar.167751609.92910201/v1.

Haslun, Joshua A., Nathaniel E. Ostrom, Eric L. Hegg, and Peggy H. Ostrom. 2018. "Estimation of Isotope Variation of N2O during Denitrification by Pseudomonas Aureofaciens and Pseudomonas Chlororaphis: Implications for N2O Source Apportionment." Biogeosciences 15 (12): 3873–82. https://doi.org/10.5194/bg-15-3873-2018.

Hayes, J M. 1983. "Practice and Principles of Isotopic Measurements in Organic Geochemistry." Organic Geochemistry of Contemporaneous and Ancient Sediments 5.

———. 2001. "Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes." Reviews in Mineralogy and Geochemistry 43 (1): 225–77. https://doi.org/10.2138/gsrmg.43.1.225.

Homyak, Peter M, Joseph C Blankinship, Kenneth Marchus, Delores M Lucero, James O Sickman, and Joshua P Schimel. 2016. "Aridity and Plant Uptake Interact to Make Dryland Soils Hotspots for Nitric Oxide (NO) Emissions."

Proceedings of the National Academy of Sciences of the United States of America 113 (19): E2608-16. https://doi.org/10.1073/pnas.1520496113.

Johnson, J B. 1928. "Thermal Agitation of Electricity in Conductors." Physical Review 32 (1): 97–109. https://doi.org/10.1103/PhysRev.32.97.

Jung, Man-Young, Reinhard Well, Deullae Min, Anette Giesemann, Soo-Je Park, Jong-Geol Kim, So-Jeong Kim, and Sung-Keun Rhee. 2014. "Isotopic Signatures of N2O Produced by Ammonia-Oxidizing Archaea from Soils." The ISME Journal 8 (5): 1115–25. https://doi.org/10.1038/ismej.2013.205.

Kaiser, Jan, Sunyoung Park, Kristie A Boering, Carl A M Brenninkmeijer, Andreas Hilkert, and Thomas Röckmann. 2004. "Mass Spectrometric Method for the Absolute Calibration of the Intramolecular Nitrogen Isotope Distribution in Nitrous Oxide." Analytical and Bioanalytical Chemistry 378 (2): 256–69. https://doi.org/10.1007/s00216-003-2233-2.

Kantnerová, Kristýna, Shohei Hattori, Sakae Toyoda, Naohiro Yoshida, Lukas Emmenegger, Stefano M. Bernasconi, and Joachim Mohn. 2022. "Clumped Isotope Signatures of Nitrous Oxide Formed by Bacterial Denitrification." Geochimica et Cosmochimica Acta, May. https://doi.org/10.1016/j.gca.2022.05.006.

Kato, Tomomichi, Sakae Toyoda, Naohiro Yoshida, Yanhong Tang, and Eitaro Wada. 2013. "Isotopomer and Isotopologue Signatures of N2O Produced in Alpine Ecosystems on the Qinghai-Tibetan Plateau." Rapid Communications in Mass Spectrometry 27 (13): 1517–26. https://doi.org/10.1002/rcm.6595.

Kelly, C, C Manning, C Frey, J Kaiser, N Gluschankoff, and K Casciotti. 2023. "Pyisotopomer: A Python Package for Obtaining Intramolecular Isotope Ratio Differences from Mass Spectrometric Analysis of Nitrous Oxide Isotopocules."

Kelly, Colette L, Cara Manning, Claudia Frey, Jan Kaiser, Noah Gluschankoff, and Karen L Casciotti. 2023. "Pyisotopomer: A Python Package for Obtaining Intramolecular Isotope Ratio Differences from Mass Spectrometric Analysis of Nitrous Oxide Isotopocules." Rapid Communications in Mass Spectrometry, March, e9513. https://doi.org/10.1002/rcm.9513.

Kelly, Colette L., Nicole M. Travis, Pascale A. Baya, and Karen L. Casciotti. 2021. "Quantifying Nitrous Oxide Cycling Regimes in the Eastern Tropical North Pacific Ocean with Isotopomer Analysis." Global Biogeochemical Cycles 35 (2). https://doi.org/10.1029/2020GB006637.

Koba, K, K Osaka, Y Tobari, S Toyoda, N Ohte, M Katsuyama, N Suzuki, et al. 2009. "Biogeochemistry of Nitrous Oxide in Groundwater in a Forested Ecosystem Elucidated by Nitrous Oxide Isotopomer Measurements." Geochimica et Cosmochimica Acta 73 (11): 3115–33. https://doi.org/10.1016/j.gca.2009.03.022.

Koehler, Birgit, Marife D. Corre, Kristin Steger, Reinhard Well, Erwin Zehe, Juvia P. Sueta, and Edzo Veldkamp. 2012. "An In-Depth Look into a Tropical Lowland Forest Soil: Nitrogen-Addition Effects on the Contents of N2O, CO2 and CH4 and N2O Isotopic Signatures down to 2-m Depth." Biogeochemistry 111 (1–3): 695–713. https://doi.org/10.1007/s10533-012-9711-6.

Kolpen, M, T Bjarnsholt, C Moser, C R Hansen, L F Rickelt, M Kühl, C Hempel, T Pressler, N Høiby, and P Ø Jensen. 2014. "Nitric Oxide Production by Polymorphonuclear Leucocytes in Infected Cystic Fibrosis Sputum Consumes Oxygen." Clinical and Experimental Immunology 177 (1): 310–19. https://doi.org/10.1111/cei.12318.

Kolpen, Mette, Michael Kühl, Thomas Bjarnsholt, Claus Moser, Christine Rønne Hansen, Lars Liengaard, Arsalan Kharazmi, Tanja Pressler, Niels Høiby, and Peter Østrup Jensen. 2014. "Nitrous Oxide Production in Sputum from Cystic Fibrosis Patients with Chronic Pseudomonas Aeruginosa Lung Infection." Plos One 9 (1): e84353. https://doi.org/10.1371/journal.pone.0084353.

Krichels, Alexander H., Peter M. Homyak, Emma L. Aronson, James O. Sickman, Jon Botthoff, Hannah Shulman, Stephanie Piper, Holly M. Andrews, and G. Darrel Jenerette. 2022. "Rapid Nitrate Reduction Produces Pulsed NO and N2O Emissions Following Wetting of Dryland Soils." Biogeochemistry 158 (2): 233–50. https://doi.org/10.1007/s10533-022-00896-x.

Kuypers, Marcel M M, Hannah K Marchant, and Boran Kartal. 2018. "The Microbial Nitrogen-Cycling Network." Nature Reviews. Microbiology 16 (5): 263–76. https://doi.org/10.1038/nrmicro.2018.9.

Li, Xiaofei, Mengting Qi, Dengzhou Gao, Min Liu, Jordi Sardans, Josep Peñuelas, and Lijun Hou. 2022. "Nitrous Oxide Emissions from Subtropical Estuaries: Insights for Environmental Controls and Implications." Water Research 212 (April): 118110. https://doi.org/10.1016/j.watres.2022.118110.

Maeda, Koki, Aymé Spor, Véronique Edel-Hermann, Cécile Heraud, Marie-Christine Breuil, Florian Bizouard, Sakae Toyoda, Naohiro Yoshida, Christian Steinberg, and Laurent Philippot. 2015. "N2O Production, a Widespread Trait in Fungi." Scientific Reports 5 (April): 9697. https://doi.org/10.1038/srep09697.

Magyar, Paul M, Victoria J Orphan, and John M Eiler. 2016. "Measurement of Rare Isotopologues of Nitrous Oxide by High-Resolution Multi-Collector Mass Spectrometry." Rapid Communications in Mass Spectrometry 30 (17): 1923–40. https://doi.org/10.1002/rcm.7671.

Magyar, Paul Macdonoald. 2017. "Insights into Pathways of Nitrous Oxide Generation from Novel Isotopologue Measurements." Doctoral dissertation, California Institute of Technology.

Mander, Ulo, Reinhard Well, Daniel Weymann, Kaido Soosaar, Martin Maddison, Arno Kanal, Krista Lõhmus, Jaak Truu, Jürgen Augustin, and Julien Tournebize. 2014. "Isotopologue Ratios of N2O and N2 Measurements Underpin the Importance of Denitrification in Differently N-Loaded Riparian Alder Forests." Environmental Science & Technology 48 (20): 11910–18. https://doi.org/10.1021/es501727h.

Mariotti, A, J C Germon, P Hubert, P Kaiser, R Letolle, A Tardieux, and P Tardieux. 1981. "Experimental Determination of Nitrogen Kinetic Isotope Fractionation: Some Principles; Illustration for the Denitrification and Nitrification Processes." Plant and Soil 62 (3): 413–30. https://doi.org/10.1007/BF02374138.

Mendler, Kerrin, Han Chen, Donovan H Parks, Briallen Lobb, Laura A Hug, and Andrew C Doxey. 2019. "AnnoTree: Visualization and Exploration of a Functionally Annotated Microbial Tree of Life." Nucleic Acids Research 47 (9): 4442–48. https://doi.org/10.1093/nar/gkz246.

Mohn, Joachim, Christina Biasi, Samuel Bodé, Pascal Boeckx, Paul J Brewer, Sarah Eggleston, Heike Geilmann, et al. 2022. "Isotopically Characterised N2 O Reference Materials for Use as Community Standards." Rapid Communications in Mass Spectrometry 36 (13): e9296. https://doi.org/10.1002/rcm.9296.

Mohn, Joachim, Benjamin Wolf, Sakae Toyoda, Cheng-Ting Lin, Mao-Chang Liang, Nicolas Brüggemann, Holger Wissel, et al. 2014. "Interlaboratory Assessment of Nitrous Oxide Isotopomer Analysis by Isotope Ratio Mass Spectrometry and Laser Spectroscopy: Current Status and Perspectives." Rapid Communications in Mass Spectrometry 28 (18): 1995–2007. https://doi.org/10.1002/rcm.6982.

Monreal, Patrick J., Colette L. Kelly, Nicole M. Travis, and Karen L. Casciotti. 2022. "Identifying the Sources and Drivers of Nitrous Oxide Accumulation in the Eddy-influenced Eastern Tropical North Pacific Oxygen-deficient Zone." Global Biogeochemical Cycles 36 (6). https://doi.org/10.1029/2022GB007310.

Nyquist, H. 1928. "Thermal Agitation of Electric Charge in Conductors." Physical Review 32 (1): 110–13. https://doi.org/10.1103/PhysRev.32.110.

Opdyke, Matthew R., Nathaniel E. Ostrom, and Peggy H. Ostrom. 2009. "Evidence for the Predominance of Denitrification as a Source of N2 O in Temperate Agricultural Soils Based on Isotopologue Measurements." Global Biogeochemical Cycles 23 (4): n/a-n/a. https://doi.org/10.1029/2009GB003523.

Ostrom, Nathaniel E, Hasand Gandhi, Tyler B Coplen, Sakae Toyoda, J K Böhlke, Willi A Brand, Karen L Casciotti, et al. 2018. "Preliminary Assessment of Stable Nitrogen and Oxygen Isotopic Composition of USGS51 and USGS52 Nitrous Oxide Reference Gases and Perspectives on Calibration Needs." Rapid Communications in Mass Spectrometry 32 (15): 1207–14. https://doi.org/10.1002/rcm.8157.

Ostrom, Nathaniel E., Adam Pitt, Robin Sutka, Peggy H. Ostrom, A. Stuart Grandy, Kristin M. Huizinga, and G. Philip Robertson. 2007. "Isotopologue Effects during N2 O Reduction in Soils and in Pure Cultures of Denitrifiers." Journal of Geophysical Research 112 (G2). https://doi.org/10.1029/2006JG000287.

Ostrom, Nathaniel E., Robin Sutka, Peggy H. Ostrom, A. Stuart Grandy, Kristin M. Huizinga, Hasand Gandhi, Joseph C. von Fischer, and G. Philip Robertson. 2010. "Isotopologue Data Reveal Bacterial Denitrification as the Primary Source of N2O during a High Flux Event Following Cultivation of a Native Temperate Grassland." Soil Biology and Biochemistry 42 (3): 499–506. https://doi.org/10.1016/j.soilbio.2009.12.003.

Palmer, Kelli L, Lindsay M Aye, and Marvin Whiteley. 2007. "Nutritional Cues Control Pseudomonas Aeruginosa Multicellular Behavior in Cystic Fibrosis Sputum." Journal of Bacteriology 189 (22): 8079–87. https://doi.org/10.1128/JB.01138-07.

Park, S, T Pérez, K A Boering, S E Trumbore, J Gil, S Marquina, and S C Tyler. 2011. "Can N2 O Stable Isotopes and Isotopomers Be Useful Tools to Characterize Sources and Microbial Pathways of N2 O Production and Consumption in Tropical Soils?" Global Biogeochemical Cycles 25 (1): n/a-n/a. https://doi.org/10.1029/2009GB003615.

Peix, Alvaro, Angel Valverde, Raúl Rivas, José M Igual, Martha-Helena Ramírez-Bahena, Pedro F Mateos, Ignacio Santa-Regina, Claudino Rodríguez-Barrueco, Eustoquio Martínez-Molina, and Encarna Velázquez. 2007. "Reclassification of Pseudomonas Aurantiaca as a Synonym of Pseudomonas Chlororaphis and Proposal of Three Subspecies, P. Chlororaphis Subsp. Chlororaphis Subsp.

Nov., P. Chlororaphis Subsp. Aureofaciens Subsp. Nov., Comb. Nov. and P. Chlororaphis Subsp. Aurantiaca Subsp. Nov., Comb. Nov." International Journal of Systematic and Evolutionary Microbiology 57 (Pt 6): 1286–90. https://doi.org/10.1099/ijs.0.64621-0.

Pérez, T, S E Trumbore, S C Tyler, P A Matson, I Ortiz-Monasterio, T Rahn, and D W T Griffith. 2001. "Identifying the Agricultural Imprint on the Global N2 O Budget Using Stable Isotopes." Journal of Geophysical Research 106 (D9): 9869–78. https://doi.org/10.1029/2000JD900809.

Poole, R K, and M N Hughes. 2000. "New Functions for the Ancient Globin Family: Bacterial Responses to Nitric Oxide and Nitrosative Stress." Molecular Microbiology 36 (4): 775–83.

Popp, Brian N., Marian B. Westley, Sakae Toyoda, Tatsuya Miwa, John E. Dore, Naohiro Yoshida, Terri M. Rust, et al. 2002. "Nitrogen and Oxygen Isotopomeric Constraints on the Origins and Sea-to-Air Flux of N2 O in the Oligotrophic Subtropical North Pacific Gyre." Global Biogeochemical Cycles 16 (4): 12-1-12–10. https://doi.org/10.1029/2001GB001806.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (version 4.1.0). Computer software.

Rohe, Lena, Traute-Heidi Anderson, Gesche Braker, Heinz Flessa, Anette Giesemann, Dominika Lewicka-Szczebak, Nicole Wrage-Mönnig, and Reinhard Well. 2014. "Dual Isotope and Isotopomer Signatures of Nitrous Oxide from Fungal Denitrification--a Pure Culture Study." Rapid Communications in Mass Spectrometry 28 (17): 1893–1903. https://doi.org/10.1002/rcm.6975.

Santoro, Alyson E, Carolyn Buchwald, Matthew R McIlvin, and Karen L Casciotti. 2011. "Isotopic Signature of N(2)O Produced by Marine Ammonia-Oxidizing Archaea." Science 333 (6047): 1282–85. https://doi.org/10.1126/science.1208239.

Sasaki, Y, K Koba, M Yamamoto, A Makabe, Y Ueno, M Nakagawa, S Toyoda, N Yoshida, and M Yoh. 2011. "Biogeochemistry of Nitrous Oxide in Lake Kizaki, Japan, Elucidated by Nitrous Oxide Isotopomer Analysis." Journal of Geophysical Research 116 (G4). https://doi.org/10.1029/2010JG001589.

Schmidt, Ingo, Rob J M van Spanning, and Mike S M Jetten. 2004. "Denitrification and Ammonia Oxidation by Nitrosomonas Europaea Wild-Type, and NirK- and

NorB-Deficient Mutants." *Microbiology* 150 (Pt 12): 4107–14. https://doi.org/10.1099/mic.0.27382-0.

Schottky, W. 1918. "Über Spontane Stromschwankungen in Verschiedenen Elektrizitätsleitern." *Annalen Der Physik* 362 (23): 541–67. https://doi.org/10.1002/andp.19183622304.

Sigman, D M, K L Casciotti, M Andreani, C Barford, M Galanter, and J K Böhlke. 2001. "A Bacterial Method for the Nitrogen Isotopic Analysis of Nitrate in Seawater and Freshwater." *Analytical Chemistry* 73 (17): 4145–53. https://doi.org/10.1021/ac010088e.

Spero, Melanie A, and Dianne K Newman. 2018. "Chlorate Specifically Targets Oxidant-Starved, Antibiotic-Tolerant Populations of Pseudomonas Aeruginosa Biofilms." *MBio* 9 (5). https://doi.org/10.1128/mBio.01400-18.

Stanton, Chloe L, Christopher T Reinhard, James F Kasting, Nathaniel E Ostrom, Joshua A Haslun, Timothy W Lyons, and Jennifer B Glass. 2018. "Nitrous Oxide from Chemodenitrification: A Possible Missing Link in the Proterozoic Greenhouse and the Evolution of Aerobic Respiration." *Geobiology* 16 (6): 597–609. https://doi.org/10.1111/gbi.12311.

Stein, Lisa Y. 2019. "Insights into the Physiology of Ammonia-Oxidizing Microorganisms." *Current Opinion in Chemical Biology* 49 (April): 9–15. https://doi.org/10.1016/j.cbpa.2018.09.003.

Sutka, Robin L, Gerard C Adams, Nathaniel E Ostrom, and Peggy H Ostrom. 2008. "Isotopologue Fractionation during N(2)O Production by Fungal Denitrification." *Rapid Communications in Mass Spectrometry* 22 (24): 3989–96. https://doi.org/10.1002/rcm.3820.

Sutka, R L, N E Ostrom, P H Ostrom, J A Breznak, H Gandhi, A J Pitt, and F Li. 2006. "Distinguishing Nitrous Oxide Production from Nitrification and Denitrification on the Basis of Isotopomer Abundances." *Applied and Environmental Microbiology* 72 (1): 638–44. https://doi.org/10.1128/AEM.72.1.638-644.2006.

Sutka, R L, N E Ostrom, P H Ostrom, H Gandhi, and J A Breznak. 2003. "Nitrogen Isotopomer Site Preference of N2O Produced by Nitrosomonas Europaea and Methylococcus Capsulatus Bath." *Rapid Communications in Mass Spectrometry* 17 (7): 738–45. https://doi.org/10.1002/rcm.968.

———. 2004. "Nitrogen Isotopomer Site Preference of N2O Produced ByNitrosomonas Europaea AndMethylococcus Capsulatus Bath." *Rapid*

Communications in Mass Spectrometry 18 (12): 1411–12. https://doi.org/10.1002/rcm.1482.

Tian, Hanqin, Rongting Xu, Josep G Canadell, Rona L Thompson, Wilfried Winiwarter, Parvadha Suntharalingam, Eric A Davidson, et al. 2020. "A Comprehensive Quantification of Global Nitrous Oxide Sources and Sinks." Nature 586 (7828): 248–56. https://doi.org/10.1038/s41586-020-2780-0.

Toyoda, Sakae, Hiroyuki Iwai, Keisuke Koba, and Naohiro Yoshida. 2009. "Isotopomeric Analysis of N2O Dissolved in a River in the Tokyo Metropolitan Area." Rapid Communications in Mass Spectrometry 23 (6): 809–21. https://doi.org/10.1002/rcm.3945.

Toyoda, Sakae, Hidenori Mutobe, Hiroaki Yamagishi, Naohiro Yoshida, and Yasunori Tanji. 2005. "Fractionation of N2O Isotopomers during Production by Denitrifier." Soil Biology and Biochemistry 37 (8): 1535–45. https://doi.org/10.1016/j.soilbio.2005.01.009.

Toyoda, Sakae, Yuuri Suzuki, Shohei Hattori, Keita Yamada, Ayako Fujii, Naohiro Yoshida, Rina Kouno, Kouki Murayama, and Hiroshi Shiomi. 2011. "Isotopomer Analysis of Production and Consumption Mechanisms of N2O and CH4 in an Advanced Wastewater Treatment System." Environmental Science & Technology 45 (3): 917–22. https://doi.org/10.1021/es102985u.

Toyoda, Sakae, and Naohiro Yoshida. 1999. "Determination of Nitrogen Isotopomers of Nitrous Oxide on a Modified Isotope Ratio Mass Spectrometer." Analytical Chemistry 71 (20): 4711–18. https://doi.org/10.1021/ac9904563.

Toyoda, Sakae. 2002. "Production Mechanism and Global Budget of N2 O Inferred from Its Isotopomers in the Western North Pacific." Geophysical Research Letters 29 (3): 1037. https://doi.org/10.1029/2001GL014311.

Wang, Zhengrong, Edwin A. Schauble, and John M. Eiler. 2004. "Equilibrium Thermodynamics of Multiply Substituted Isotopologues of Molecular Gases." Geochimica et Cosmochimica Acta 68 (23): 4779–97. https://doi.org/10.1016/j.gca.2004.05.039.

Well, R, H Flessa, F Jaradat, S Toyoda, and N Yoshida. 2005. "Measurement of Isotopomer Signatures of N2 O in Groundwater." Journal of Geophysical Research 110 (G2): n/a-n/a. https://doi.org/10.1029/2005JG000044.

Well, Reinhard, Wolfram Eschenbach, Heinz Flessa, Carolin von der Heide, and Daniel Weymann. 2012. "Are Dual Isotope and Isotopomer Ratios of N2O

Useful Indicators for N2O Turnover during Denitrification in Nitrate-Contaminated Aquifers?" Geochimica et Cosmochimica Acta 90 (August): 265–82. https://doi.org/10.1016/j.gca.2012.04.045.

Westley, Marian B, Brian N Popp, and Terri M Rust. 2007. "The Calibration of the Intramolecular Nitrogen Isotope Distribution in Nitrous Oxide Measured by Isotope Ratio Mass Spectrometry." Rapid Communications in Mass Spectrometry 21 (3): 391–405. https://doi.org/10.1002/rcm.2828.

Westley, Marian B., Hiroaki Yamagishi, Brian N. Popp, and Naohiro Yoshida. 2006. "Nitrous Oxide Cycling in the Black Sea Inferred from Stable Isotope and Isotopomer Distributions." Deep Sea Research Part II: Topical Studies in Oceanography 53 (17–19): 1802–16. https://doi.org/10.1016/j.dsr2.2006.03.012.

Wickham, H, W Chang, and M H Wickham. 2016. "Package 'Ggplot2'." Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2 (1): 1–189.

Wilbert, Steven A, and Dianne K Newman. 2022. "The Contrasting Roles of Nitric Oxide Drive Microbial Community Organization as a Function of Oxygen Presence." Current Biology 32 (24): 5221-5234.e4. https://doi.org/10.1016/j.cub.2022.10.008.

Wong, Wei Wen, Moritz F. Lehmann, Thomas Kuhn, Caitlin Frame, Seng Chee Poh, Ian Cartwright, and Perran L.M. Cook. 2021. "Nitrogen and Oxygen Isotopomeric Constraints on the Sources of Nitrous Oxide and the Role of Submarine Groundwater Discharge in a Temperate Eutrophic Salt-wedge Estuary." Limnology and Oceanography 66 (4): 1068–82. https://doi.org/10.1002/lno.11664.

Yamagishi, Hiroaki, Marian B. Westley, Brian N. Popp, Sakae Toyoda, Naohiro Yoshida, Shuichi Watanabe, Keisuke Koba, and Yasuhiro Yamanaka. 2007. "Role of Nitrification and Denitrification on the Nitrous Oxide Cycle in the Eastern Tropical North Pacific and Gulf of California." Journal of Geophysical Research 112 (G2). https://doi.org/10.1029/2006JG000227.

Yamazaki, T, T Hozuki, K Arai, S Toyoda, K Koba, T Fujiwara, and N Yoshida. 2014. "Isotopomeric Characterization of Nitrous Oxide Produced by Reaction of Enzymes Extracted from Nitrifying and Denitrifying Bacteria." Biogeosciences (Online) 11 (10): 2679–89. https://doi.org/10.5194/bg-11-2679-2014.

Yamulki, S, S Toyoda, N Yoshida, E Veldkamp, B Grant, and R Bol. 2001. "Diurnal Fluxes and the Isotopomer Ratios of N(2)O in a Temperate Grassland Following Urine Amendment." Rapid Communications in Mass Spectrometry 15 (15): 1263–69. https://doi.org/10.1002/rcm.352.

Yang, Hui, Hasand Gandhi, Nathaniel E Ostrom, and Eric L Hegg. 2014. "Isotopic Fractionation by a Fungal P450 Nitric Oxide Reductase during the Production of N2O." Environmental Science & Technology 48 (18): 10707–15. https://doi.org/10.1021/es501912d.

Yano, Midori, Sakae Toyoda, Takeshi Tokida, Kentaro Hayashi, Toshihiro Hasegawa, Akiko Makabe, Keisuke Koba, and Naohiro Yoshida. 2014. "Isotopomer Analysis of Production, Consumption and Soil-to-Atmosphere Emission Processes of N2O at the Beginning of Paddy Field Irrigation." Soil Biology and Biochemistry 70 (March): 66–78. https://doi.org/10.1016/j.soilbio.2013.11.026.

Yeung, Laurence Y. 2016. "Combinatorial Effects on Clumped Isotopes and Their Significance in Biogeochemistry." Geochimica et Cosmochimica Acta 172 (January): 22–38. https://doi.org/10.1016/j.gca.2015.09.020.

Yoon, Sang Sun, Robert F Hennigan, George M Hilliard, Urs A Ochsner, Kislay Parvatiyar, Moneesha C Kamani, Holly L Allen, et al. 2002. "Pseudomonas Aeruginosa Anaerobic Respiration in Biofilms: Relationships to Cystic Fibrosis Pathogenesis." Developmental Cell 3 (4): 593–603. https://doi.org/10.1016/s1534-5807(02)00295-2.

Yoshida, N, and S Toyoda. 2000. "Constraining the Atmospheric N2O Budget from Intramolecular Site Preference in N2O Isotopomers." Nature 405 (6784): 330–34. https://doi.org/10.1038/35012558.

Zou, Yun, Yuhei Hirono, Yosuke Yanai, Shohei Hattori, Sakae Toyoda, and Naohiro Yoshida. 2014. "Isotopomer Analysis of Nitrous Oxide Accumulated in Soil Cultivated with Tea (Camellia Sinensis) in Shizuoka, Central Japan." Soil Biology and Biochemistry 77 (October): 276–91. https://doi.org/10.1016/j.soilbio.2014.06.016.

5. Conclusion

Our knowledge of the microbial and biochemical world has rapidly expanded over the past few decades. Metagenomics has revealed many surprises, including novel clades of enzymes like rubisco (Banda et al. 2020; Liu et al. 2022; Schulz et al. 2022) and novel metabolic pathways like comammox (complete ammonia oxidation; (Daims et al. 2015; van Kessel et al. 2015)). This, in turn, has challenged our conventional frameworks for rationalizing Earth's elemental 'cycles,' one where specialized classes of organisms neatly and completely transform one elemental species into another. The terms we use to describe nitrogen transformations carry the legacy of these assumptions – nitrifiers oxidize ammonia to nitrate ('nitrify'), and then denitrifiers turn nitrate back into dinitrogen ('de-nitrify'). In reality, nitrogen transformations spanning eight redox states (-3 to +5) can be carried out by a wide range of diverse organisms (see (Kuypers et al. 2018) for review), and pathways thought to be carried out by specific organisms, like denitrification, are instead carried out by complex communities where each member only catalyzes a portion of the full pathway (Gowda et al. 2022). Even the word 'cycle' implies a completeness of knowledge that does not exist – in reality, most elemental cycles, like the nitrogen cycle, are not closed, and thermodynamics predict that there are yet many novel biochemical reactions waiting to be discovered (see (O'Malley and Walsh 2021) for review).

How do we meaningfully respond to these challenges as the geochemistry community, especially our isotopic tools? First, a mechanistic understanding of KIEs could enable predictions based on non-isotopic enzymatic characteristics (i.e. turnover rates, substrate availability, Michaelis constant). This way, KIE measurements from a representative subset of enzymes could be used to predict the KIEs of that class of enzyme overall. This would also enable us to predict the KIEs of novel enzymes gleaned from metagenomic data (i.e. enzymes that have not yet been purified and measured *in vitro*). A mechanistic understanding of KIEs could also enable us to address variations in KIE seen in the literature – i.e. are 1‰ variations meaningful or do they result from methodological differences? What about 10‰ variations? Such knowledge would allow us, for example, to answer why the SP of bacterial NORs, measured by multiple labs, seems to vary on the order of 10‰ (Chapter 4), or why a Form I' rubisco has a smaller KIE than a Form I even though the active sites of both enzymes are nominally the same (Chapter 3; (Wang et al. 2023)). SP measurements may be a good model system – the N_2O molecule is just complicated enough to offer additional isotopic constraints (i.e. site-specific isotope enrichments); there are multiple classes of enzymes that catalyze the NO reduction reaction (i.e. NORs, Fhps, p450nor, and flavo-diiron proteins); there is good abiotic data and predictions to compare our enzymatic results to. This data may allow us to systematically test certain hypotheses, like if the electronegativity of residues around the active site affect SP values – for example, though Fhp and fungal NORs have a heme Fe active site, Fhp gives SP values ~10‰ while fungal NORs give values of ~30‰. Comparison of their active

sites may help shed light on this problem. Overall, systematic measurement of diverse native enzymes, paired with kinetic modeling and their abiotic reaction counterparts, may help us better understand the reaction mechanisms that lead to enzymatic isotopic fractionation.

Concurrently, building a faster pipeline for KIE measurements is key to increasing the amount of high-quality KIE data available. The fact that the iNOR strain (rhamnose-induced expression of NOR; Chapter 4) gave similar values to previously published *in vitro* NOR measurements (Yamazaki et al. 2014) show that, at least in well-studied organisms like *P. aeruginosa*, alternatives to *in vitro* KIE measurements exist. In addition, Chapters 2 and 3 leveraged collaborations with biochemistry labs that specialize in rubisco purification – additional collaborations between isotope geochemistry and biochemistry labs are sure to bear fruit. Alternatively, a lab that specifically works at the intersection of these problems could make significant progress towards these goals.

However, though such knowledge of KIEs *in vitro* is critical, biochemical reactions are not catalyzed by naked enzymes in the environment – therefore, an understanding of microbial physiology and KIEs *in vivo* is essential for applying such knowledge to natural environments. As shown in Chapter 2 (Wang et al. 2023), the overall expression of rubisco and other enzymes' KIE at the biomass level is highly dependent on external environmental conditions (i.e. pCO$_2$ and light levels). And as shown in Chapter 4, knowing extra- and intracellular concentrations of NO are necessary for predicting protein abundances of two similarly regulated enzymes, NOR and Fhp. Therefore, by measuring KIEs both *in vitro* and *in vivo* in environmentally relevant experimental conditions, significant progress can be made.

Finally, we must keep an open mind when studying biologically mediated reactions. Our work on N$_2$O (Chapter 4) showed that a focus on dissimilatory pathways caused the broader class of NO reductases to be neglected. Our work on rubisco (Chapters 2 and 3) questions how strictly uniformitarianism can be applied to biology – though uniformitarian approaches may be readily applied to abiotic processes, it's unclear how it can be applied to biological processes which undergo substantial evolution over geologic timescales. Overall, 'biogeochemistry' has, at times, become a catch-all term for all biologically mediated chemical reactions on Earth environments. Critically re-focusing 'biogeochemistry' to 'biochemistry of the Earth' could facilitate new ways of understanding and investigating our natural world. Future work at the interface of isotope geochemistry, biochemistry and microbial physiology is likely to offer a path forward.

5.1 References

Banda, Douglas M, Jose H Pereira, Albert K Liu, Douglas J Orr, Michal Hammel, Christine He, Martin A J Parry, et al. 2020. "Novel Bacterial Clade Reveals Origin of Form I Rubisco." Nature Plants 6 (9): 1158–66. https://doi.org/10.1038/s41477-020-00762-4.

Daims, Holger, Elena V Lebedeva, Petra Pjevac, Ping Han, Craig Herbold, Mads Albertsen, Nico Jehmlich, et al. 2015. "Complete Nitrification by Nitrospira Bacteria." Nature 528 (7583): 504–9. https://doi.org/10.1038/nature16461.

Gowda, Karna, Derek Ping, Madhav Mani, and Seppe Kuehn. 2022. "Genomic Structure Predicts Metabolite Dynamics in Microbial Communities." Cell 185 (3): 530-546.e25. https://doi.org/10.1016/j.cell.2021.12.036.

Kessel, Maartje A H J van, Daan R Speth, Mads Albertsen, Per H Nielsen, Huub J M Op den Camp, Boran Kartal, Mike S M Jetten, and Sebastian Lücker. 2015. "Complete Nitrification by a Single Microorganism." Nature 528 (7583): 555–59. https://doi.org/10.1038/nature16459.

Kuypers, Marcel M M, Hannah K Marchant, and Boran Kartal. 2018. "The Microbial Nitrogen-Cycling Network." Nature Reviews. Microbiology 16 (5): 263–76. https://doi.org/10.1038/nrmicro.2018.9.

Liu, Albert K., Jose H. Pereira, Alexander J. Kehl, Daniel J. Rosenberg, Douglas J. Orr, Simon K. S. Chu, Douglas M. Banda, et al. 2022. "Structural Plasticity Enables Evolution and Innovation of RuBisCO Assemblies." Sci. Adv. 8 (34). https://doi.org/10.1126/sciadv.adc9440.

O'Malley, Maureen A, and David A Walsh. 2021. "Rethinking Microbial Infallibility in the Metagenomics Era." FEMS Microbiology Ecology 97 (8). https://doi.org/10.1093/femsec/fiab092.

Schulz, Luca, Zhijun Guo, Jan Zarzycki, Wieland Steinchen, Jan M Schuller, Thomas Heimerl, Simone Prinz, Oliver Mueller-Cajar, Tobias J Erb, and Georg K A Hochberg. 2022. "Evolution of Increased Complexity and Specificity at the Dawn of Form I Rubiscos." Science 378 (6616): 155–60. https://doi.org/10.1126/science.abq1416.

Wang, Renée Z, Robert J Nichols, Albert K Liu, Avi I Flamholz, Juliana Artier, Doug M Banda, David F Savage, John M Eiler, Patrick M Shih, and Woodward W Fischer. 2023. "Carbon Isotope Fractionation by an Ancestral Rubisco Suggests That Biological Proxies for CO2 through Geologic Time Should Be Reevaluated." Proceedings of the National Academy of Sciences of the United States of America 120 (20): e2300466120. https://doi.org/10.1073/pnas.2300466120.

Wang, Renée Z., Albert K. Liu, Douglas M. Banda, Woodward W. Fischer, and Patrick M. Shih. 2023. "A Bacterial Form I' Rubisco Has a Smaller Carbon Isotope Fractionation than Its Form I Counterpart." Biomolecules, April.

Yamazaki, T, T Hozuki, K Arai, S Toyoda, K Koba, T Fujiwara, and N Yoshida. 2014. "Isotopomeric Characterization of Nitrous Oxide Produced by Reaction of Enzymes Extracted from Nitrifying and Denitrifying Bacteria." Biogeosciences (Online) 11 (10): 2679–89. https://doi.org/10.5194/bg-11-2679-2014