Primer

Evolution and origins of rubisco

Leah J. Taylor-Kearney[1], Renée Z. Wang[1,2], and Patrick M. Shih[1,3,4,5,*]

Rubisco (D-ribulose 1,5-bisphosphate carboxylase/oxygenase) is the most abundant enzyme in the world, constituting up to half of the soluble protein content in plant leaves. Such is its ubiquity that its chemical fingerprint can be detected in the geological record spanning billions of years. Rubisco catalyses the conversion of inorganic CO_2 into organic sugars, which underpin almost all of the biosphere, including our entire food chain. Due to its central role in the global carbon cycle, rubisco has been the subject of intense research for over 50 years. Rubisco is often considered inefficient due to its slow rate of carboxylation compared with other central metabolism enzymes, and its promiscuous oxygenase activity, which competes with the productive carboxylation reaction. It is hoped that engineering improved CO_2 fixation will have significant advantages in agriculture and climate change mitigation. However, rubisco has proven difficult to engineer, with decades of efforts yielding limited results. Recent research has focused on reconstructing the evolutionary trajectory of rubisco to help elucidate its cryptic origins. Such evolutionary studies have led to a better understanding of both the origins of more complex rubisco forms and the broader relationship between rubisco's structure and function.

While there are numerous distinct forms of rubisco found across the tree of life, most global carbon fixation is driven by the form I rubisco found in plants, algae, and some bacteria. Rubisco is the central enzyme in the Calvin-Benson-Bassham (CBB) cycle. It catalyses the fixation of a single CO_2 molecule to the 5-carbon sugar, ribulose 1,5 bisphosphate (RuBP), producing two molecules of 3-phosphoglycerate (3PG). Rubisco's oxygenation reaction yields one molecule of 3PG and one molecule of 2-phosphoglycolate (2PG). While 3PG provides organic carbon to various downstream metabolic pathways, 2PG inhibits carbon metabolism and results in net carbon loss during its catabolism.

Our planet's atmosphere has dramatically changed over the course of Earth's 4.5 billion year old history. It is thought that the rise of atmospheric O_2 and concurrent fall of CO_2 necessitated certain adaptations to ensure efficient carbon fixation. Many metabolic innovations are thought to have evolved around the Great Oxygenation Event (~2.5 billion years ago; Ga), which coincides with the emergence of oxygenic photosynthesis. However, questions still remain regarding rubisco evolution — the answers to which are key to both bioengineering efforts and larger, conceptual ideas about protein evolution in response to environmental change.

Origin of form I rubisco

The different rubisco forms (i.e., form I, II, II/III and III rubiscos) are defined by sequence, structure, function, and phylogenetic relationships. Forms I and II are CBB-associated, found in plants, algae, and select bacteria. The form III and II/III rubiscos are found primarily in archaea, which lack a CBB, and catalyse non-photosynthetically driven CO_2 fixation using sugars synthesised from nucleotide metabolism or as part of the reductive hexulose-phosphate pathway. Rubisco catalysis includes a series of steps: active site lysine carbamylation, Mg(II) binding, RuBP binding, enediol formation, addition of CO_2 or O_2, and cleavage to produce products. It has been suggested that the oxygenation reaction is a vestige of the environment in which rubisco evolved. Rubisco is thought to have emerged early in Earth's history, potentially as early as 3.8 billion years ago, when atmospheric CO_2 concentrations were much higher and O_2 levels were much lower. Therefore, rubisco's early evolution was not initially influenced by a need to differentiate between CO_2 and O_2.

Certain innovations in rubisco structure and host physiology have evolved to limit the oxygenation reaction. Such adaptations support the idea that rubisco evolution is tightly constrained by CO_2/O_2 discrimination. For example, carbon-concentrating mechanisms have convergently evolved in a wide range of organisms to create CO_2 enriched environments around rubisco, competitively inhibiting the oxygenation reaction. While not all oxygenic phototrophs use carbon-concentrating mechanisms to inhibit oxygenation, all have evolved rubiscos with improved abilities to distinguish between CO_2 and O_2. Rubsico's ability to distinguish between gases is described by a specificity factor ($S_{C/O}$), and is defined by the rates of carboxylation and oxygenation (k_{cat}) and the Michaelis-Menten (K_M) constants for CO_2 and O_2 [$S_{C/O} = (k_{cat,C}/K_C)/(k_{cat,O}/K_O)$]. There exists a well-documented kinetic trade-off between rubisco's k_{cat} and its specificity. The form I rubiscos, for example, have the highest $S_{C/O}$ values of all forms but tend to have lower $k_{cat,C}$ values. The inverse is also true, in that rubiscos with higher $k_{cat,C}$ values often have lower $S_{C/O}$ values. The form I rubiscos in particular have greatly altered their quaternary structure by incorporating a small subunit. The role the small subunit plays in form I catalysis has long been debated, with theories ranging from modulating specificity to stabilising rubisco during folding and catalysis.

Structurally, a rubisco must consist of at least two large subunits, which assemble head to tail, forming the active sites. The form II, II/III and IIIs can vary their oligomeric state by binding multiple large subunit pairs (L_{2+2n}), resulting in homo-oligomers. By contrast, the form I has a unique hetero-oligomeric assembly (L_8S_8); with an octameric large subunit core and eight additional non-catalytic small subunits. The small subunits reside in between the large subunit dimers, capping both ends of the protein to form the L_8S_8 structure. As the small subunit was the most obvious structural difference between the form I and the other, less specific forms, it has been assumed that the SSU is responsible for the form I's higher CO_2 specificity. However, due to extant form I rubisco's reliance on the small subunit for stability and catalysis, it has been difficult to test this theory directly.

Recently, it was demonstrated that the carboxylation efficiency and specificity of the form I rubisco likely improved when the L_8S_8 assembly first evolved. An ancestrally reconstructed form I rubisco, which was active without a small subunit, had its specificity increased with the addition of a small subunit binding partner. While some modest gains in specificity could be attributed to the allosteric interaction between the large and the small subunit

solely, the specificity could be doubled with certain amino acid substitutions. These findings suggest that the biggest, indirect, contribution the small subunit likely made to specificity was to increase the available sequence space, allowing for variants that would have otherwise been impossible. This aligns with the fact that many extant form I rubiscos from anaerobes, which do not have to deal with O_2, have reverted to lower specificities while retaining their small subunit. However, the increased sequence space also enabled the accumulation of substitutions that made the large subunit reliant on the small subunit for solubility, thus trapping the L_8S_8 assembly. Until recently, the steps taken from simpler, small subunit-deficient rubisco forms to the complex form I had been poorly resolved. However, novel extant lineages have helped contextualise the evolution of the form I, furthering our understanding of important molecular transitions that preceded the L_8S_8.

The recent discovery and characterization of three sister clades to form I rubisco — dubbed form I', I" and Iα — have helped retrace the origin and evolution of form I rubisco (Figure 1). Of the three, the form Iα clade is the most distantly related to the canonical form I clade and assembles as a basic L_2 dimer. Moving up the phylogeny, the form I' and I" clades transitioned to assemble as a L_8 octamer. Ultimately, the octameric assembly enabled the incorporation of the small subunit, which gave rise to the form I clade that dominates our carbon cycle today. Interestingly, form I' enzyme kinetics suggest that increases in specificity may not necessarily require the small subunit. The form I' was found in the genome of '*Candidatus* Promineofilum breve', from the order Anaerolineales, which typically contains obligate anaerobes. The form I' has a typical kinetic trade-off in that it has a low $k_{cat,C}$ (~2 s^{-1}) and a higher specificity factor ($S_{C/O} \approx 36$). While the form I' specificity factor is lower than the median specificity of most plant form I rubiscos ($S_{C/O} \approx 98$), the form I' specificity is comparable to the median specificity for cyanobacterial form Is ($S_{C/O} \approx 48$). By contrast, the form II and III rubiscos, which also lack a small subunit, have much lower specificities ($S_{C/O} \approx 4$–18). This suggests that while the small subunit can play a role in CO_2 selectivity, it may not necessarily be deterministic, but rather one avenue of biochemical innovation.

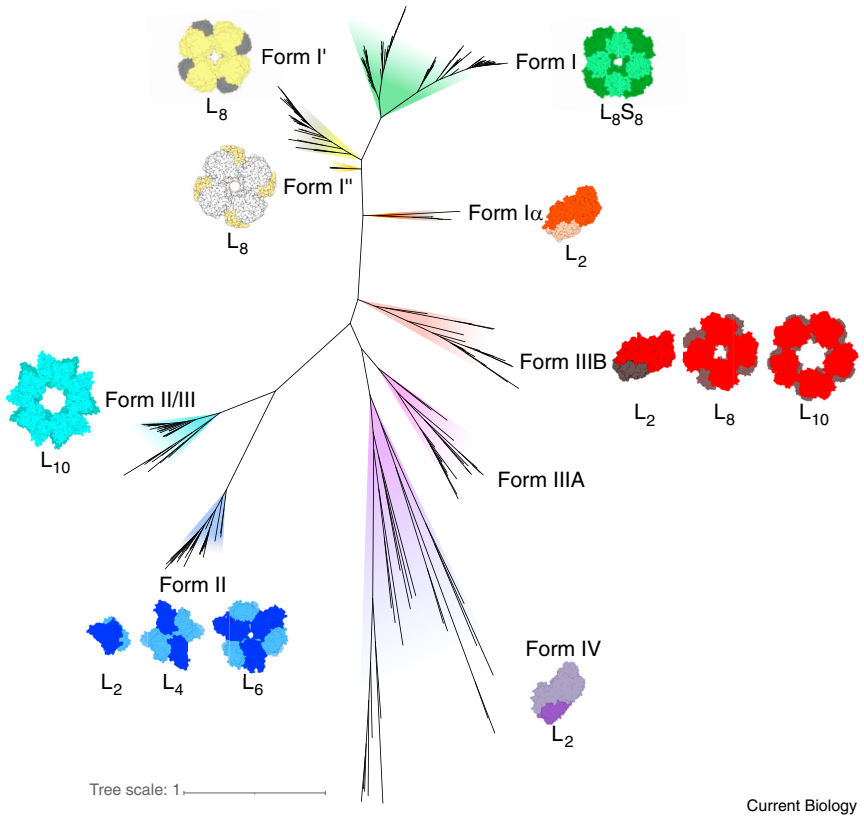

Figure 1. An unrooted maximum likelihood phylogenetic tree of rubisco sequences. Contrasting colours highlight the dimeric interface between the large subunits, or between the large subunit and small subunits of the form I. Many rubisco forms vary their higher order oligomeric states by binding multiple identical dimeric large subunits (L_{2+2n}). The form I is unique, forming a hexadecameric L_8S_8 structure, with a core octameric large subunit and eight non-catalytic small subunits. Form I PDB: 1RBL; form I' PDB: 6URA; form I" PDB: 8U66; form Iα: alpha fold model; form IV PDB: 2QYG; form II PDB: 5RUB, PDB: 7T1C, PDB: 5C2C; form II/III PDB: 5MAC; form IIIB PDB: 8DHT, PDB: 2CWX, PDB: 1GEH.

Structural innovation beyond the form I rubisco

The form I rubisco has likely been entrenched in the L_8S_8 assembly due to small subunit incorporation. Although the form I rubisco is still the most O_2-tolerant and arguably successful lineage today, it is not clear how entrenchment may impact the long-term evolutionary trajectory of the clade. By contrast, other rubisco forms explore many large subunit configurations beyond the L_8. For example, the form II/III rubiscos form L_{10} assemblies. The form II and III rubiscos exhibit great structural plasticity, adopting multiple homo-oligomeric states (form II as L_2, L_4, L_6 and form III as L_2, L_8, L_{10}).

The form II clade in particular demonstrates remarkable structural diversity and plasticity. Recent work by Liu *et al.* reconstructed the complex evolution of form II, revealing multiple interconversion events between dimers and hexamers. The discovery of a tetramer with a novel inter-dimer interface further highlighted the structural flexibility of the form II clade and illustrated how readily novel protein–protein interactions can evolve. These evolutionary studies were further complemented by protein engineering efforts that showed how even a few mutations could break hexamers into dimers, or turn dimers into hexamers.

How quaternary structure may constrain both the biochemistry and evolutionary trajectory of rubisco remains poorly understood. However, diversity-driven form II rubisco studies

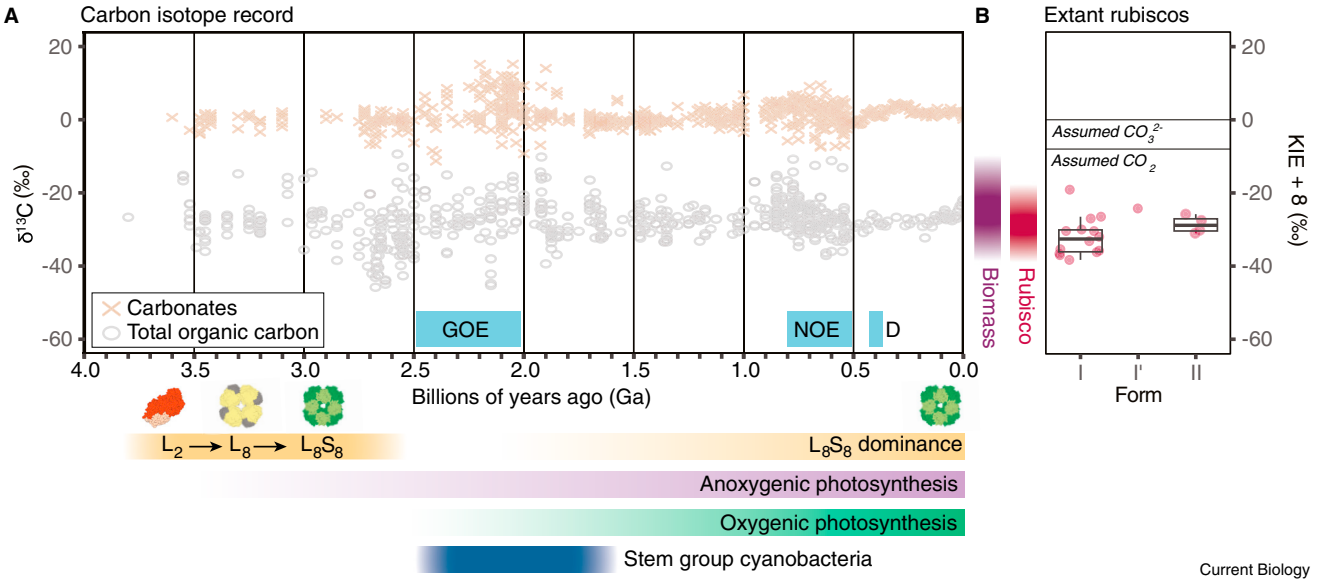

Figure 2. Carbon isotope record and extant rubisco kinetic isotope effects.
(A) The carbon isotope record, which are the ^{13}C vs. ^{12}C concentrations ($\delta^{13}C$) of inorganic carbon pools (preserved largely as carbonates; brown crosses) and organic carbon pools (preserved as total organic carbon in sedimentary rocks; grey circles) through time. Data are from Krissansen-totton *et al.* (2015). Oxygen is thought to have increased in a stepwise fashion through Earth's history (GOE = Great Oxygenation Event; ~2.5–2.0 Ga; NOE = Neoproterozoic Oxygenation Event; 0.8–0.5 Ga; D = oxygenation during the Devonian; ~0.4 Ga). The evolution of oxygenic photosynthesis, from anoxygenic photosynthesis, is thought to have caused the GOE. Stem group cyanobacteria, which have rubisco and the CBB cycle, are thought to have evolved either prior to or after the GOE. Rubisco is thought to have evolved ~3.8 Ga, but the timing of L_8S_8 evolution is uncertain, and it is not clear when the L_8S_8 and CBB cycle rose to become the most ecologically abundant carbon fixation pathway today. (B) The KIEs of extant rubiscos with the $CO_2 \rightleftarrows HCO_3^-$ equilibrium isotope effect added (KIE + 8) so that it can be compared with the rock record; the equilibrium isotope effect is assumed for 25°C. The spread in $\delta^{13}C$ of all rubiscos and biomass (red, purple bars) are shown to the left.

have helped us better understand how rubisco's structure can affect biochemical function. Liu *et al.* engineered two L_2 variants derived from an L_6 using two separate single point mutations. While just a single point mutation at the inter-dimer interface was sufficient to form the L_2, kinetic characterisation revealed both L_2 variants had higher specificities compared to wild type. The change in specificity was dictated by higher K_O values for both variants, which can be interpreted as a decrease in affinity for O_2, even though these mutations were distal to the active site. Exactly how the enzyme's affinity for O_2 was reduced by the L_6–L_2 conversion is unknown. It does, however, further support the idea that changes in higher order oligomeric assemblies may play an important role in modulating rubisco kinetics. Future studies will help clarify the role distal residues play in controlling rubisco oligomerization and function. In turn, this may help reveal the biophysical limitations of the small subunit-dependent form I rubisco clade and innovations therein.

Evolutionary origin and timing of rubisco

Evolutionarily, questions remain regarding when rubisco first appeared, and when the L_8S_8 became so ecologically dominant. Although phylogenetics enables relative comparisons of extant proteins, it does not provide evolutionary timing. For that, one must turn towards the geologic record. Isotopic data from the geologic record has largely been interpreted as evidence for rubisco evolving in the Archaean prior to the Great Oxygenation Event about 2.5 Ga and prior to the rise of stem group cyanobacteria (Figure 2A). However, the exact timing of key transitions like the evolution of the L_8S_8 is still largely unknown (Figure 2A). In addition, interpretation of the record itself is debated due to uncertainties posed by abiotic processes that can alter the record, and uncertainties regarding rubisco's intrinsic kinetic isotope effect (KIE).

Archean evidence of rubisco evolution is largely reliant on the carbon isotope record; a globally compiled record of stable carbon isotope ratios (^{13}C vs.^{12}C) in sedimentary rocks up to ~3.8 Ga (Figure 2A). Carbon isotope ratios are typically reported in delta (δ) notation where negative $\delta^{13}C$ values are more ^{12}C-enriched. Modern rubiscos display a KIE where $^{12}CO_2$ is preferentially fixed over $^{13}CO_2$, and the magnitude of this preference is reported as the KIE where larger KIE values indicate a greater relative preference for $^{12}CO_2$ vs. $^{13}CO_2$. Since CBB-utilizers are the dominant primary producers in modern ecosystems, all organic carbon on Earth today is relatively ^{13}C-depleted compared with inorganic carbon, reflecting rubisco's preference for ^{12}C. This isotopic signature is then preserved in rocks. Thus, if rubisco was active throughout geologic time in a manner similar to today, carbon isotope measurements are effectively chemical fossils that can be used to extend our observations of rubisco into the Archaean.

The current interpretation of the carbon isotope record as evidence for rubisco's emergence in the Archaean is based on two key assumptions; the first is that the carbon isotope record

has accurately recorded the ancient global carbon cycle. To interpret the carbon isotope record as a history of rubisco evolution, we first need to know what the $\delta^{13}C$ of 'true' CO_2 and organic carbon was in the past. However, certain biotic and abiotic processes can alter the $\delta^{13}C$ values of inorganic carbon and organic carbon in the rock record, casting doubt on whether the measured $\delta^{13}C$ values can be attributed to rubisco solely. Another difficulty is that sediments can preserve regional rather than global signals, and that the carbon cycle records both biotic and abiotic processes like volcanism. In addition, though inorganic carbon is preserved as carbonates, we are primarily interested in CO_2 since this is the inorganic carbon species that rubisco fixes. However, there is an equilibrium carbon isotopic fractionation between CO_2 and HCO_3^- where ^{13}C prefers to be in HCO_3^- by ~8‰ at 25°C (Figure 2B); therefore, one must infer temperature and other parameters to reconstruct the $\delta^{13}C$ of CO_2 through deep time. If these challenges can be surmounted, and the effect of abiotic process controlled for, then the carbon isotope record may offer a record of early rubisco evolution.

The second assumption is that the ε_{Rub} values of extant rubiscos can be applied readily to the past, despite the fact that rubisco and the CBB cycle have evolved substantially over geologic timescales. Though the enzyme rubisco is responsible for the primary isotopic composition of biomass, physiology causes the KIE to decrease in bulk biomass (Figure 2B). Since bulk biomass from diverse organisms and physiologies is preserved in the rock record, this increases the range of uncertainty that must be taken into account. Currently, the spread in measured KIE values of modern rubiscos and biomass (Figure 2B) loosely matches the variation in $\delta^{13}C$ of sedimentary organic carbon throughout Earth history (Figure 2A). Assuming that the $\delta^{13}C$ of rocks measured today captures what 'true' $\delta^{13}C$ values were in the past, one can conclude that rubisco evolved roughly 3.8 Ga, and certain isotopic data may even extend this date to roughly 4.1 Ga. However, this interpretation is debated, given the large range of uncertainty on rubisco KIEs, and the many potential sources of noise in the carbon isotope record as noted above.

Ultimately, much of the validity of these interpretations and assumptions hinge on a mechanistic understanding of rubisco KIEs. Intriguingly, there are OC $\delta^{13}C$ values that are more ^{12}C-enriched than one would expect based on extant KIE values (Figure 2), and these values are often interpreted as evidence for other early metabolisms. This, however, is based on an assumption that the KIE values of the few sampled extant rubiscos represent the KIE values of all rubiscos, extinct or extant. Notably, of the tens of thousands of rubisco sequences that have been deposited into databases, fewer than fifteen natural rubisco KIE values have been measured, and most are of form I. Therefore, it is not clear how applicable KIE values of extant rubiscos are to interpreting the past. For example, the form I' rubisco — a potential analogue for rubisco prior to the Great Oxidation Event — has a smaller KIE than a comparable form I (16 vs. 22‰, respectively). Similarly, a reconstructed, inferred ancestral form IB rubisco dating to roughly 1 Ga also has a smaller KIE than the extant form IB (17 vs. 25‰, respectively). These data suggest that KIEs of ancient rubiscos may have been quite different, suggesting that the KIEs of modern rubiscos may not be directly applied to interpreting the past. KIE measurements at key transitions in rubisco evolution may therefore help us answer some outstanding questions regarding the timing of rubisco evolution and the ecological dominance of the CBB cycle before and after the Great Oxidation Event since we are heavily reliant on the carbon isotope record for rubisco's Archaean history.

Outlook and future directions

Recent advances have highlighted how diversity-driven studies that better sample and characterise extant rubiscos can provide the requisite first-order knowledge needed to drive better hypotheses and build a more comprehensive understanding of rubisco evolution. Second-order questions elucidating how rubisco has evolved can be addressed using various techniques, such as ancestral sequence reconstruction and isotopic measurements. Finally, the application of this knowledge is critical to better resolving how our planet and its carbon cycle have changed as a function of rubisco.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FURTHER READING

Banda, D.M., Pereira, J.H., Liu, A.K., Orr, D.J., Hammel, M., He, C., Parry, M.A.J., Carmo-Silva, E., Adams, P.D., Banfield, J.F., *et al*. (2020). Novel bacterial clade reveals origin of form I Rubisco. Nat. Plants *6*, 1158–1166.

Des Marais, D.J. (2001). Isotopic evolution of the biogeochemical carbon cycle during the Precambrian. Rev. Mineral. Geochem. *43*, 555–578.

Flamholz, A.I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y.M., Oltrogge, L.M., Alves, R., Savage, D., and Milo, R. (2019). Revisiting trade-offs between rubisco kinetic parameters. Biochemistry *58*, 3365–3376.

Kono, T., Mehrotra, S., Endo, C., Kizu, N., Matusda, M., Kimura, H., Mizohata, E., Inoue, T., Hasunuma, T., Yokota, A., *et al*. (2017). A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. Nat. Commun. *8*, 14007.

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

Liu, A.K., Pereira, J.H., Kehl, A.J., Rosenberg, D.J., Orr, D.J., Chu, S.K.S., Banda, D.M., Hammel, M., Adams, P.D., Siegel, J.B., *et al*. (2022). Structural plasticity enables evolution and innovation of RuBisCO assemblies. Sci. Adv. *8*, eadc9440.

Liu, A.K., Kaeser, B., Chen, L., West-Roberts, J., Taylor-Kearney, L.J., Lavy, A., Günzing, D., Li, W.-J., Hammel, M., Nogales, E., *et al*. (2023). Deep-branching evolutionary intermediates reveal structural origins of form I rubisco. Curr. Biol. *33*, 5316–5325.e3.

Prywes, N., Phillips, N.R., Tuck, O.T., Valentin-Alvarado, L.E., and Savage, D.F. (2023). Rubisco function, evolution, and engineering. Annu. Rev. Biochem. *92*, 385–410.

Raven, J.A., Cockell, C.S., and De La Rocha, C.L. (2008). The evolution of inorganic carbon concentrating mechanisms in photosynthesis. Philos. Trans. R. Soc. B Biol. Sci. *363*, 2641–2650.

Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science *315*, 1003–1006.

Schidlowski, M. (2001). Carbon isotopes as biogeochemical recorders of life over 3.8 Ga of Earth history: evolution of a concept. Precambrian Res. *106*, 117–134.

Schulz, L., Guo, Z., Zarzycki, J., Steinchen, W., Schuller, J.M., Heimerl, T., Prinz, S., Mueller-Cajar, O., Erb, T.J., and Hochberg, G.K.A. (2022). Evolution of increased complexity and specificity at the dawn of form I Rubiscos. Science *378*, 155–160.

[1]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720, USA. [3]Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA 94608, USA. [4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Innovative Genomics Institute, Berkeley, CA, USA.
*E-mail: pmshih@berkeley.edu