ANNUAL REVIEWS

Annual Review of Biochemistry

Rubisco Function, Evolution, and Engineering

Noam Prywes,[1] Naiya R. Phillips,[2] Owen T. Tuck,[3] Luis E. Valentin-Alvarado,[4] and David F. Savage[1,2,4,5]

[1]Innovative Genomics Institute, University of California, Berkeley, California, USA; email: savage@berkeley.edu

[2]Department of Molecular and Cell Biology, University of California, Berkeley, California, USA

[3]Department of Chemistry, University of California, Berkeley, California, USA

[4]Graduate Group in Microbiology, University of California, Berkeley, California, USA

[5]Howard Hughes Medical Institute, University of California, Berkeley, California, USA

Annu. Rev. Biochem. 2023. 92:17.1–17.26

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-040320-101244

Keywords

rubisco, carbon fixation, enzyme engineering, Calvin cycle, carboxylases

Abstract

Carbon fixation is the process by which CO_2 is converted from a gas into biomass. The Calvin–Benson–Bassham cycle (CBB) is the dominant carbon-consuming pathway on Earth, driving >99.5% of the ~120 billion tons of carbon that are converted to sugar by plants, algae, and cyanobacteria. The carboxylase enzyme in the CBB, ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco), fixes one CO_2 molecule per turn of the cycle into bioavailable sugars. Despite being critical to the assimilation of carbon, rubisco's kinetic rate is not very fast, limiting flux through the pathway. This bottleneck presents a paradox: Why has rubisco not evolved to be a better catalyst? Many hypothesize that the catalytic mechanism of rubisco is subject to one or more trade-offs and that rubisco variants have been optimized for their native physiological environment. Here, we review the evolution and biochemistry of rubisco through the lens of structure and mechanism in order to understand what trade-offs limit its improvement. We also review the many attempts to improve rubisco itself and thereby promote plant growth.

Contents

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) is the carboxylase of the Calvin–Benson–Bassham cycle (CBB), where it fixes CO_2 onto a ribulose bisphosphate (RuBP) sugar, producing two molecules of 3-phosphoglycerate (3PG) (**Figure 1**). This reaction is central to the carbon cycle and converts ~100 gigatons of carbon from CO_2 into biomass annually—approximately 10 times more than annual human emissions (1, 2).

The balance of CO_2 uptake and release by the biosphere has supported a stable atmospheric composition over the past few millions years of ~78% N_2, ~21% O_2, and ≈0.02–0.04% CO_2,

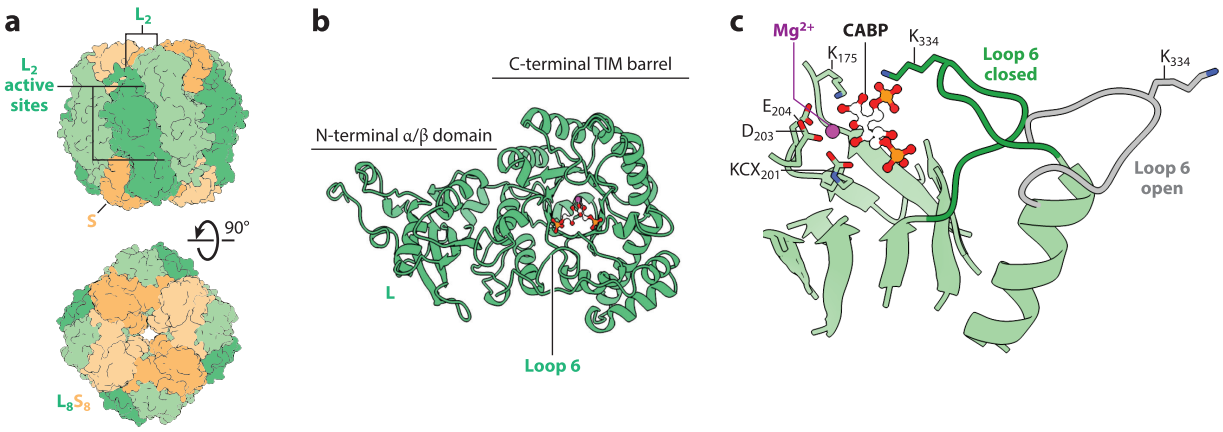


Figure 1

A triptych digest of structural features of Form I rubisco from spinach bound to inhibitor CABP (PDB IDs: 8RUC, 9RUB) (3, 4). (*a*) The L_8S_8 architecture contains the conserved L_2 homodimeric functional unit present in all rubiscos. (*b*) Substrate binds at the face of the C-terminal TIM barrel after capture of divalent cation and carbamylation of an active-site lysine. CABP mimics the structure of RuBP after CO_2 addition. (*c*) Loop 6 extends over the center of the TIM barrel during binding and participates in catalysis (*green*, +CABP; *gray*, apo conformation). Abbreviations: CABP, 2-carboxyarabinitol-1,5-bisphosphate; L, large subunit; PDB ID, Protein Data Bank identifier; RuBP, ribulose-1,5-bisphosphate; S, small subunit; TIM, triosephosphate isomerase.

Table 1 Carboxylases found in nature

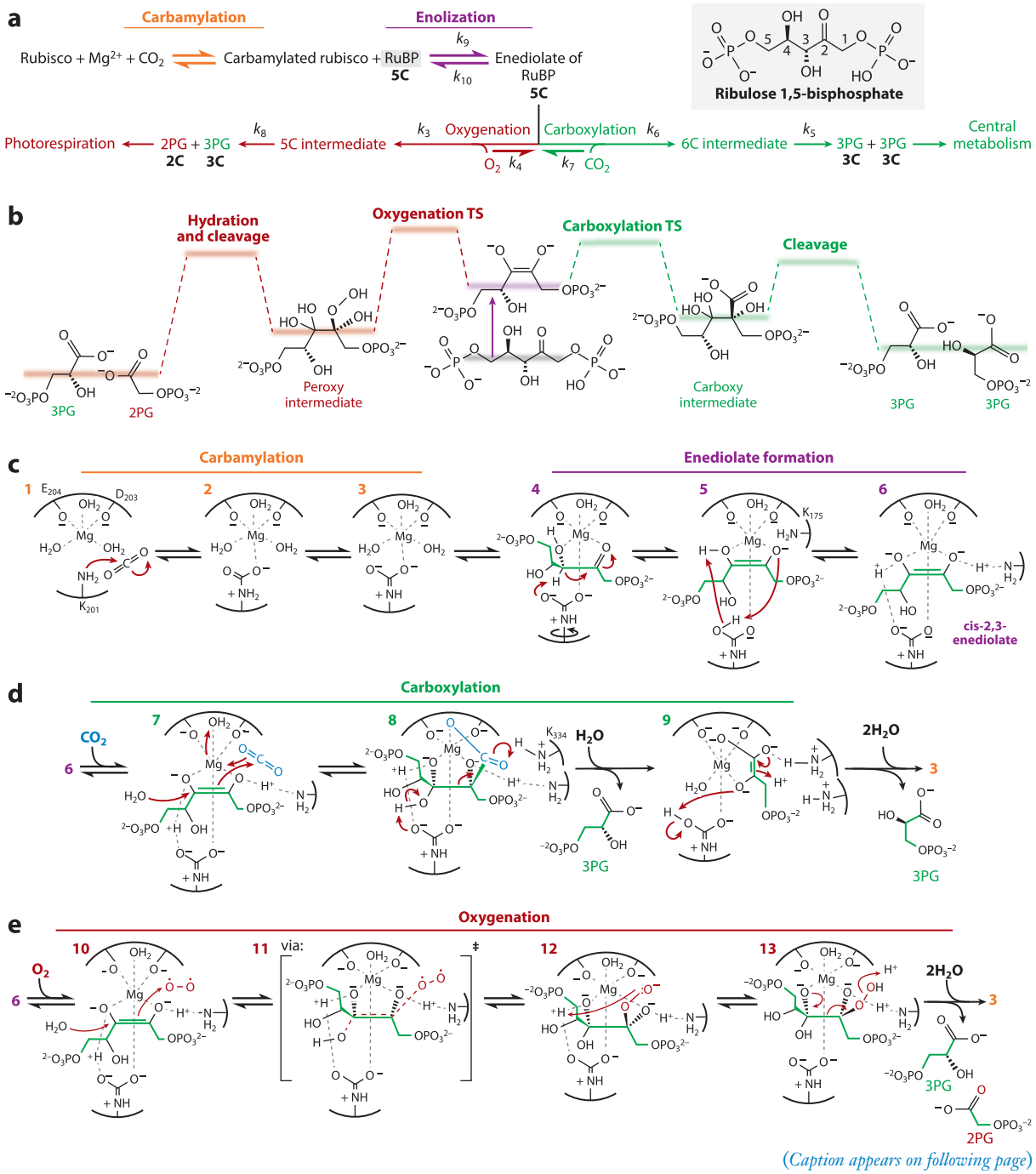
Carboxylase Class	Difficulty with oxygen?	Energetic coupling?	Carbon-fixation pathway	Substrate
Rubisco	Confuses O_2 and CO_2	Substrate (C–C cleavage)	Natural and artificial	CO_2
Reducing carboxylases (e.g., crotonyl-CoA carboxylase)	None	NAD(P)H	Artificial	CO_2
ATP-dependent carboxylases (e.g., Ac-CoA carboxylase)	None	ATP	Natural and artificial	HCO_3
Vitamin K–dependent carboxylase	None	O_2	None	CO_2
Ferredoxin oxidoreductases (e.g., KGOR)	Usually O_2 intolerant	Ferredoxin	Natural and artificial	CO_2
Amine carboxylases (e.g., carbamoyl phosphate synthase)	None	ATP	None	HCO_3
Glycine dehydrogenase	None	Reduced disulfide	None	CO_2
Aromatic carboxylase (PurE or PurK)	None	Substrate or ATP	None	CO_2/HCO_3
Formate dehydrogenase[a]	Usually O_2 intolerant	NAD(P)H, Ferredoxin or Ferrocytochrome b1	Natural	CO_2

Abbreviations: Ac-CoA, acetylcoenzyme A; ATP, adenosine triphosphate; KGOR, 2-ketoglutarate ferredoxin oxidoreductase; NAD(P)H, reduced nicotinamide adenine dinucleotide (phosphate); rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase.
[a]While not technically carboxylases, formate (and carbon monoxide) dehydrogenases reduce CO_2 for use in metabolism.

but this was not always the case. Some ~3 billion years ago, when rubisco first evolved, CO_2 levels were likely quite high and O_2 certainly quite low (5). The historical atmospheric composition may help explain one of the seeming paradoxes of rubisco biochemistry: Rubisco is notable among carboxylases for also reacting promiscuously with O_2 (**Table 1**) (6, 7). The off-target oxygenation of RuBP produces 3PG and a molecule of 2-phosphoglycolate (2PG), the latter of which must be recycled in a carbon salvage pathway, termed photorespiration. Ironically, the oxygenic photosynthetic metabolism enabled by rubisco caused the rise in atmospheric O_2 levels and the attendant inhibition of the enzyme.

The apparent inefficiency of rubisco has led some to malign it as slow or confused (8), average (9), or not really so bad (10). In terms of maximum rate (k_{cat}) or catalytic efficiency (k_{cat}/K_M), rubisco ranks close to the median among characterized enzymes (9). It is far from a diffusion-limited perfect enzyme, but perhaps rubisco deserves to be graded on a curve: Perfect enzymes often catalyze easier reactions—reactions that proceed at appreciable uncatalyzed rates. The uncatalyzed carboxylation of RuBP is far too slow to measure, but its rate constants have been estimated computationally (11). By this measure, rubisco is highly effective, conferring a 10^{15}- to 10^{18}-fold rate enhancement (10). This is a reflection of the difficulty of rubisco's chemical mechanism, in particular the challenge associated with having CO_2—a small, hydrophobic, uncharged, relatively inert molecule—as a substrate (**Figure 2**). As a result of its kinetic parameters, rubisco is highly expressed in plants to increase total carboxylation and is therefore the most abundant protein on the planet (2).

Rubisco has received significant attention as a target for protein engineering, but attempts to improve it face a steeper challenge than is typical (15, 16). Successful protein engineering campaigns often optimize a property orthogonal to the natural function of an enzyme; for example, one starts with a small amount of promiscuous activity in an enzyme that catalyzes a different reaction and, over successive rounds of screening or selection, cultivates the promiscuous activity (17). In contrast to these empirical best practices, the desired improvements to rubisco (in carboxylation



(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Mechanisms of rubisco catalysis. (*a*) Middle-out scheme. Rate constants are numbered as described previously (6, 12). (*b*) Simplified landscape of chemical reactions at the rubisco active site. Relative intermediate and transition state energies are not drawn to scale. (*c*) Carbamylation, RuBP binding, and preparation of the activated cis-2,3-enediolate (State 6) (13). (*d*) Carboxylation of enediolate intermediate, State 6. Gas addition and hydration are shown as a concerted mechanistic step. (*e*) Oxygenation of State 6, giving one molecule of 3PG and one molecule of 2PG. Drawn to depict plausible steps in SET. Proton assignments informed by isotope-labeling experiments (14). All active-site residues are numbered based on spinach rubisco. Energy levels not to scale. Abbreviations: 2PG, 2-phosphoglycolate; 3PG, 3-phosphoglycerate; RuBP, ribulose bisphosphate; SET, single electron transfer; TS, transition state.

rate, CO_2 affinity, and specificity for CO_2 over O_2) are all axes along which evolution has already acted for billions of years with apparently limited results (18). Improvements in one rubisco biochemical parameter (e.g., kinetic rate of carboxylation, $k_{cat,C}$) may come at the expense of another (e.g., specificity, $S_{C/O}$). A variety of trade-offs, discussed in the section titled Kinetics, Trade-offs, and Optimization, have been proposed based on biochemical data and mechanistic chemical logic (6, 7). Relevant parameters include rate constants $k_{cat,C}$ and $k_{cat,O}$, Michaelis constants K_C and K_O, and the ratio of carbon- and oxygen-specific catalytic efficiencies $S_{C/O}$.

The earliest rubisco mutagenesis studies were motivated by a desire to improve rubisco's rate and specificity (19, 20). Efforts have focused on regions near the active site and a mobile region termed loop 6 that is known to be involved in catalysis (21). Rational mutagenesis has not been successful in generating superlative rubiscos, i.e., enhancing carboxylation while limiting oxygenation. Mutant libraries in bacteria have been used to evolve rubisco in high throughput since the early 1990s (22). *Escherichia coli* has been engineered in a variety of ways to serve as a chassis for rubisco library selections (23). These selections have generally resulted in rubiscos with improved expression and stability, though some have yielded enzymes with faster $k_{cat,C}$s (24, 25) and lower K_Cs (26). An alternative means of understanding the sequence–function landscape is through natural diversity. Although initial studies focused on plant rubisco, the bulk of sequence and functional diversity is found in microbes (27, 28), and a wide swath of this diversity remains unexplored.

Even if rubisco could be improved, it is debatable whether this would translate to faster-growing autotrophs. At low light or high CO_2 concentrations, the maximum rate of electron transport (J_{max}) limits growth (29, 30). A similar argument has been made regarding improvements to specificity. Although photorespiration may seem wasteful, it may be necessary for nitrogen uptake (31, 32), and nitrogen is often the limiting factor for growth (33). However, increased rubisco carboxylation flux may improve growth and yield in agricultural settings (34–36) when resources other than carbon are not limiting (37) or when the downstream carbon-sink capacity is artificially increased (38). Thus far, however, all attempts to improve plant growth by rubisco replacement have failed.

Here, we first review recent and historical insights into one of the most pivotal enzymes to life on earth. We then review how the structure and catalytic mechanism influence the catalytic rate of rubisco and then explore the diverse superfamily of rubisco homologs. Finally, we summarize the history of attempts to improve rubisco function both in vitro and in vivo and finish by highlighting where new insights and practical advances may be achieved.

RUBISCO FUNCTION

A Brief History of Rubisco

Rubisco was discovered twice, first as an unidentified enzyme of unusually high abundance in leaves and later as the first step of CO_2 assimilation in plants. Although the initial discovery, that as much as 50% of soluble leaf protein is a single species in gel electrophoresis, was viewed skeptically, repeated crystallization (39) would prove that there was only one protein in the suspect

gel band. Originally called fraction 1 protein based on ammonium sulfate fractionation (40), it was eventually realized this protein was the same enzyme identified as the first step of the CBB pathway—ribulose diphosphate carboxylase (41, 42). C_3 photosynthesis is named after phosphoglycerate, the triose product of the carboxylase reaction. Calvin and coworkers (43) had already correctly speculated that there was an enediolate intermediate (**Figure 2*b***, center) bound to magnesium that would attack CO_2 and subsequently hydrolyze. The relatively slow activity of rubisco was also noted with curiosity in these early studies (44). Different forms of rubisco were first recognized in *Rhodospirillum rubrum* and *Rhodobacter* in 1968 (45).

It took until the early 1970s to learn that rubisco engaged in promiscuous oxygenation activity (46, 47), which led to its current name—ribulose bisphosphate carboxylase/oxygenase—provided by David Eisenberg (40). Careful analysis of the order of addition of reactants further revealed that rubisco requires an activation step wherein a specific active-site lysine (48, 49) is carbamylated with CO_2 (50, 51)—there are thus two CO_2 molecules at the active site during catalysis, one as a cofactor and the other as a substrate. Carbonic anhydrase analysis showed that CO_2 and not bicarbonate is the substrate for carboxylation (52); tritium exchange experiments proved Calvin was right about reversible enediol formation (53, 54); and nuclear magnetic resonance studies confirmed the order of cofactor, metal, and substrate binding (55). Mendelian inheritance of the small subunit (SSU) (56) and maternal inheritance of the large subunit (LSU) (57) demonstrated the respective nuclear and chloroplastic localization of the genes. Just a few years after the first DNA sequence of maize rubisco was determined (58), mutagenesis studies were pinpointing the key active-site residues conserved between bacterial and plant rubiscos (59, 60). At the same time, X-ray crystal structures were solved for both of these forms (61, 62) (for more history, see 40, 63).

Structure

All rubiscos have the same basic functional unit of a homodimer of LSUs (L_2), with each subunit consisting of an N-terminal α/β domain and a C-terminal $(\beta/\alpha)_8$ triosephosphate isomerase (TIM) barrel (**Figure 2*a*,*b***). The structure of the rubisco large-subunit dimer is well characterized and highly conserved, even between homologs with only ∼30% sequence identity (27, 64). The subunits are antiparallel with the two active sites situated at the interfaces between the N and C termini. A polar binding pocket is formed across the opening of one C-terminal barrel with positively charged regions on opposite sides of the barrel to anchor the two phosphate groups on RuBP, while the coordinated Mg^{2+} orients C2 of the substrate for its reaction with CO_2 (**Figures 1*b*,*c*** and **2*a*–c**). This pocket is formed primarily by highly conserved residues in the loops connecting the β strands and α helices of the barrel, while two loop regions from the neighboring N-terminal domain close part of the top of the binding pocket (65). Upon binding to RuBP, the flexible loop 6, connecting β-strand 6 to α-helix 6 in the C-terminal domain, undergoes a ≈12-Å shift that encloses the active site (66, 67) and is thought to stabilize the reaction intermediates. All bona fide rubiscos have a high degree of conservation of the active-site residues (68), and mutations to active-site residues have been reviewed in detail (69, 70). Mutation of any of the roughly 20 conserved residues surrounding the active site, discussed more below in the section titled Engineering, leads to a decrease in catalytic activity or specificity.

The rubisco family of proteins contains a number of forms, separated by sequence, taxonomic distribution, or biochemical characteristics (see the section titled Evolution). The L_2 dimer is the core feature of rubisco structure, but it often assembles into high-order oligomers across the phylogeny. For example, the first rubisco that was isolated (71) was found to assemble into a heterohexadecameric structure (L_8S_8) composed of four L_2 dimers assembled in a ring, with the dimer–dimer interfaces capped at the top and bottom by additional SSUs (72, 73) (**Figure 2*a***).

This class of rubisco is called Form I and has been further classified into five main subtypes: IA, found mostly in proteobacteria, cyanobacteria with α-carboxysomes, and the alga *Paulinella*; IB in β-cyanobacteria, green algae, and plants; IC mostly in proteobacteria; ID in red algae, diatoms, haptophytes, and other algae; and IE, generally from actinobacteria; plus a variety of newly characterized and unknown forms largely found in Chloroflexi bacteria (**Figure 4*c***) (27, 74). Form II, II/III, and III rubiscos lack the SSU but often form higher-order oligomers of the L_2 dimer. They are primarily found in prokaryotes.

The SSU is not strictly necessary for carboxylation, but its absence results in a dramatic loss of complex stability, substrate-binding affinity, and catalytic activity in model plant rubiscos (75), leading to some debate over its function (reviewed in 76, 77). Form I is the only class of rubiscos with an SSU and also demonstrates a higher CO_2 specificity than other forms, suggesting the SSU plays some role in selectivity. However, there is currently no confirmed mechanism by which the SSU provides this increased specificity. One theory is that the SSU stabilizes the LSU, improving its robustness to mutation and allowing it to explore a productive region of the fitness landscape with higher CO_2 specificity (78). Removing the SSU greatly reduces the thermal stability of the *Synechococcus elongatus* 6301 heterohexadecamer, but it is not absolutely required for the formation of stable oligomers, as Form I′ rubiscos form octamers and Forms II and III form $(L_2)_n$ assemblies without SSUs (79, 80). The SSU has more recently been shown to have a possible role in regulating rubisco activity, as differentially expressed isoforms of the tobacco rubisco SSUs yield holoenzymes with moderate differences in $k_{cat,C}$ and K_C values (81). It has also been suggested that the formation of the holoenzyme creates unique binding sites at the subunit interfaces wherein rubisco can serve as a protein–protein interaction hub to recruit additional factors (82–85). It is also worth noting that fusing the LSUs and SSUs into one protein produces functional rubisco in tobacco (86), which is useful for further research efforts to determine the role of individual SSU isoforms *in planta*.

While Form I rubiscos tend to have higher specificity for CO_2 and more opportunity for regulation, they also require a suite of folding chaperones in order to avoid misfolding and aggregation (reviewed in 87). *Arabidopsis thaliana* rubisco has been shown to require five chloroplast-specific chaperones, as well as the GroEL/ES chaperonins, to fold in *E. coli*, and these chaperones were species specific. That is, the *E. coli* strain expressing *A. thaliana* rubisco chaperones could produce only *A. thaliana* rubisco and not *Nicotiana tabacum* rubisco (88). This presents a major challenge to studying Form I rubiscos: Heterologous expression of a given rubisco requires first identification and expression of its accompanying chaperones, which is itself a tedious process in slow-growing plants or unculturable bacteria and algae.

Mechanisms

Rubisco catalysis occurs in a series of composite steps: preparation of the active site, RuBP binding, enediol formation, gas addition (CO_2 or O_2), and cleavage to produce two 3PG or one 3PG and one 2PG from CO_2 and O_2, respectively (**Figure 2*a***). Computational modeling of rubisco led by Gready and colleagues and Tcherkez (89, 90) permits speculation about the heights of the energy barriers (**Figure 2*b***). Here, we summarize proposed mechanistic steps in the reaction (**Figure 2*c–e***). Efforts to improve rubisco activity will benefit from considering the nuances of these individual steps along with the chemical and structural constraints imposed by the active site.

Carbamylation of a conserved active-site lysine is the crucial activation step required for both carboxylation and oxygenation (**Figure 2*c***). The carbamyl moiety is derived from the attack of a lysyl amine on a CO_2 molecule (a different CO_2 molecule than the one that is added to RuBP), generating a zwitterionic intermediate. The coordinated divalent Mg^{2+} stabilizes the *aci* form of

the carbamate (State 3, **Figure 2*c***), prolonging the modification and priming the carbamyl group for its function as a general base in later stages of catalysis (13, 48, 50).

After Mg^{2+} binding and carbamylation, RuBP enters the active site, binds to Mg^{2+} (State 4, **Figure 2*c***), and is converted to the *cis*-2,3-enediolate, with the carbamyl oxygen that is not bound to Mg^{2+} acting as a general base (reviewed extensively in 13). Evidence from crystallography and isotopic labeling suggest that proton transfer activates the enediolate such that the enediolate oxygens are nearly bidentate with respect to the metal (State 6, **Figure 2*c***). This results in a strained pseudo-five-membered ring poised to act as a nucleophile. In carboxylation, CO_2 approaches from the solvent-exposed face and adds, along with H_2O, across the enediolate to give the 6-carbon intermediate. Rubisco does not bind CO_2 noncovalently with a dedicated binding pocket, in a classical Michaelis complex. Rather, covalent bond formation happens immediately upon CO_2 entering the active site (55). There is debate over the mechanism of hydration and CO_2 addition: Lorimer and colleagues (91) showed that the mechanism was most likely concerted, with His294 acting as the base (**Figure 2*d***, State 7, and **Figure 3*c***). Additional evidence from chemical simulations (92) and isotope experiments (93) complicate these assertions, though no conclusive mechanism can be drawn to date. The structure of rubisco bound to the competitive inhibitor carboxyarabinitol-1,5-bisphosphate (CABP), a mimic of the reaction intermediate found in State 8 in **Figure 2*d***, illustrates the location of the nascent carboxyl group and serves as a useful tool for dissecting the mechanism. Once hydrated (State 8, **Figure 2*d***), cleavage of the long C2–C3 bond occurs rapidly via carbamate-mediated deprotonation (13) and stereospecific protonation, furnishing two molecules of 3PG bound for central metabolism (**Figure 2*a***).

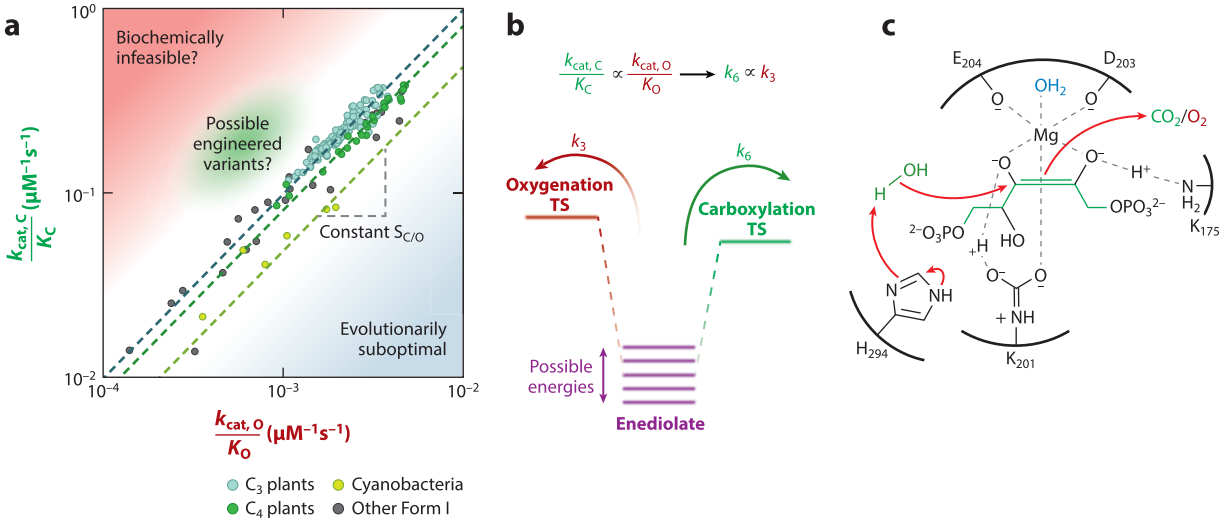


Figure 3

Example of a possible mechanistic trade-off: the empirical link between the carboxylation and oxygenation efficiencies. (*a*) Catalytic efficiencies for carboxylation and oxygenation for Form I rubiscos (data from 18). The values are fit to a line in log-space that has a slope very close to 1. (*b*) The trade-off represented by the line between measured kinetic parameters can be interpreted as a correlation between the rate constants for gas addition of CO_2 and O_2. The correlation could arise from a change in the energy of the enediolate bound in the active site with the TS energies held constant. (*c*) More detail for State 6 in **Figure 2*c***. Some of the residues that may be involved in determining the energy of bound enediolate and the activation energies for gas addition are depicted. This step of the mechanism precedes both oxygenation and carboxylation. Abbreviations: K_C, Michaelis constant for carboxylation; $k_{cat,C}$, kinetic rate of carboxylation; $k_{cat,O}$, kinetic rate of oxygenation; K_O, Michaelis constant for oxygenation; $S_{C/O}$, specificity (ratio of carbon- and oxygen-specific catalytic efficiencies), TS, transition state.

Rubisco oxygenation, which produces one 3PG and one photorespiration-bound 2PG, is notable because the reaction of the singlet endiolate with the biradical triplet oxygen is spin forbidden and, in nearly all other oxygenases besides rubisco, is achieved with cofactors. Indeed, no other carboxylase has promiscuous oxygenase activity (**Table 1**). Experimentalists have only recently begun to elucidate the mechanism of rubisco oxygenation. Work by Tcherkez and colleagues (14) demonstrated that rubisco catalyzes oxygenation by single electron transfer (SET), wherein a superoxide radical transfers an electron to RuBP, generating two radicals that can recombine freely. They base their argument on both the invariance of the reaction to the ^{18}O kinetic isotope effect in rubiscos with different CO_2/O_2 specificities and the computed reduction potentials of RuBP and superoxide. These data suggest that the enediolate is energetically poised for SET oxygenation and that diversity in $S_{C/O}$ is the result of minute differences in the active-site electric environments of various rubiscos (94). Another plausible mechanism for spin-forbidden oxygenation is intersystem crossing (ISC), where one reactant changes its electronic state to match the other. For example, torsionally strained RuBP in the active site can, in principle, be excited from a singlet to a triplet state. Kannappan, Cummins, and Gready (89) performed large-scale quantum chemical calculations that suggest that, in disagreement with the SET model, the enediolate forms a caged biradical complex with O_2. Both the ISC and SET camps agree that preparation of the enediolate for facile carboxylation invites oxygenation and that there is no obvious way to rationally design a rubisco whose active site configuration strongly disfavors oxygen addition without compromising carboxylation. The discovery of an oxygenase-only rubisco isoform (see the section titled Evolution) suggests that the relationship between oxygenation and carboxylation may not be symmetric. Carboxylation may necessitate oxygenation, but the reverse is not the case: It is possible to arrange an active site with an apparent $S_{C/O}$ of zero (95). This result is less surprising when we consider that rubisco carboxylation, while favorable energetically, with a $\Delta G° \approx -30$ kJ/mol, pales in comparison to the oxygenation $\Delta G° \approx -500$ kJ/mol. Similarly, it was shown that carboxylation can also be suppressed by replacing Mg^{2+} with Co^{2+} in the active site (96).

Several posttranslational modifications apart from lysine carbamylation can interface with the rubisco mechanism. Nitrosylation of an active-site cysteine occludes RuBP from binding before carbamylation is complete (97, 98). RuBP binds extremely strongly, even in uncarbamylated active sites, impeding activity and requiring a chaperone, rubisco activase, to remove RuBP and regenerate the active site (99) (reviewed in 87). Rubisco S-nitrosylation via nitric oxide (NO) is increasingly appreciated as an important modulator in plant metabolism and may represent a potential avenue for control of rubisco activity (100). Some autotrophs exert additional control over the active site by using binders that mimic catalytic intermediates. For example, the dephosphorylated analog of the 2-carboxy-3-ketoarabinitol 1,5-bisphosphate (CKABP) intermediate (**Figure 2*b*,*d***), carboxyarabinitol 1-phosphate (CA1P), binds the carbamylated active site in the absence of sulfate ions and has been implicated in protection of rubisco from proteolysis during nighttime inactivity (101, 102).

Kinetics, Trade-offs, and Optimization

It has long been observed that specific rubiscos (i.e., those with high $S_{C/O}$) tend to be slower carboxylases (i.e., have low $k_{cat,C}$) (103). The characterization of natural and engineered rubisco variants led to the emergence of several trade-off models to explain such correlations in the biochemical data. Kinetic models for rubisco activity have been developed in vitro (6, 12, 104) and in vivo (105), though only a few of the many possible parameters are routinely measured. As kinetic measurements of rubiscos from various species trickled in over the decades, a picture emerged of an enzyme caught between two or more conflicting priorities, forced to compromise (6, 7, 103).

While there are still very few available measurements from many rubisco forms (**Supplemental Figure 1**), it is possible to compare the kinetics of rubiscos from individual classes, especially the Form IB variety found in many cyanobacteria, green algae, and all plant chloroplasts, which are the most extensively studied.

One clear trend in kinetic parameters across species is the correlation between catalytic efficiencies for carboxylation and oxygenation (i.e., $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$) (**Figure 3**). This result has been interpreted to imply that while rubisco can easily evolve to increase or decrease both the carboxylation and oxygenation efficiencies, it is very difficult to improve the ratio between them (7, 18). Since the selectivity measure is simply the ratio between the carboxylation and oxygenation efficiencies [$S_{C/O} = (k_{cat,C}/K_C)/(k_{cat,O}/K_O)$], it follows that $S_{C/O}$ should not vary greatly (**Figure 3*a***). Within the Form IB rubiscos, $S_{C/O}$ varies only 30% (18, 106). A related trade-off, between selectivity and rate, is common among enzymes, and many enzymes, like rubisco, have a hard time distinguishing between substrates through different binding energies. These enzymes are forced to use transition state stabilization differences (15). Mechanistically, the observed trade-off may be related to initial enediolate formation, which can then react with O_2 or CO_2 with little ability to discriminate (**Figure 3*b***). The correlations supporting other proposed trade-offs, like that between $k_{cat,C}$ and K_C (6, 7), have attenuated as more data have been collected (18, 106, 107).

k_{cat} and K_M are composites of the rate constants for individual steps in the reaction mechanism, so correlations between these phenomenological parameters must reflect dependencies between the energy barriers associated with individual reaction steps. For instance, it has been proposed that the observed correlation between carboxylation and oxygenation efficiencies can be rewritten as a correlation between the on rate of CO_2 and O_2; i.e., k_3 is proportional to k_6 (**Figures 2*a*** and **3*b***). This correlation in turn can be explained by the proposal that the activation energies for carboxylation and oxygenation can grow or shrink but only in proportion to one another. A simple way for this to happen would be if the energy of the enediolate were to change while the transition state energies remain fixed (**Figure 3*b***) (7, 18, 90, 108). As the energy of the bound enediolate rises, the Hammond postulate dictates that the transition states become more reactant like, which would result in worse selectivity because the reactant in carboxylation and oxygenation is the same (6). In addition, the basis for a trade-off can ultimately be found in the structure and electronics of the enzyme active site. One proposal for the modest negative correlation between $k_{cat,C}$ and K_C is that a key active-site residue, Lys175, may act as a base in enolization and as an acid in hydration and cleavage, which could result in a trade-off where improvement to one of its functions comes at a cost to the other. Some possible trade-off sources are shown in **Figure 3*c***, including Lys175 (109), His294 (110), and Lys201 (111). Crucially, each of these residues is involved in two or more mechanistic steps, creating the potential for improvements in one function coming at the cost of another. Advances in molecular modeling (92), more measurements of individual rate constants from different species (104), more data on rubiscos outside of Form I (28), and molecular evolution studies will improve our understanding of the mechanistic basis, or lack thereof, of apparent trade-offs.

A final trend relates to host physiology and strengthens the case for trade-offs. Many organisms possess CO_2-concentrating mechanisms (CCMs) that increase the concentration of CO_2 near rubisco (112, 113). Elevated CO_2 promotes carboxylation and competitively inhibits oxygenation (114). Thus, the evolutionary pressure in CCM-bearing organisms is shifted toward increasing $k_{cat,C}$, and the influence of the other kinetic parameters is diminished. CCMs are found in plants, algae, and bacteria and have evolved convergently, resulting in a variety of different strategies. In bacteria, CCMs sequester rubisco inside a protein compartment with locally high CO_2 produced via active carbon uptake. Alternatively, so-called C_4 plants use specialized anatomy to separate initial CO_2 fixation with a nonrubisco carboxylase (e.g., phosphoenolpyruvate carboxylase) in order

to facilitate a second rubisco carboxylation, the latter carried out in specialized cells with high CO_2 (115). Rubiscos from CCM-bearing organisms generally have higher carboxylation rate constants ($k_{cat,C}$), lower selectivities ($S_{C/O}$) and weaker affinities for CO_2 (K_C) (7, 18, 106), implying that evolution cannot isolate a superlative variant and instead optimizes the system around rubisco to achieve fast, on-target catalysis.

Rubisco Among Carboxylases

When considering the complex mechanism of rubisco catalysis, and the steps peculiar to the reaction(s) it catalyzes, a paradox begins to materialize: Why is >99% (1) of carbon still fixed through the CBB cycle despite rubisco's well-known limitations? This question follows from an intuitive premise that evolution could surely do better than relying on a slow and nonspecific catalyst to produce nearly all organic carbon in the biosphere. One place to look for answers is among other carboxylases—there are dozens of carboxylases and roughly ten different carbon-fixation pathways (116, 117). These comparisons offer three primary answers, none of them completely satisfying: oxygen sensitivity, energetic cost, and metabolic inflexibility. Some pathways rely on enzymes that are nonfunctional, or nearly so, in the presence of oxygen [e.g., reverse tricarboxylic acid (TCA) cycle (see 118)]. Oxygen insensitive pathways generally require more energy per fixed carbon. In addition, the Calvin cycle is well integrated into the pentose phosphate pathway and the remainder of central carbon metabolism (**Supplemental Figure 2**), while the two alternative oxygen-tolerant pathways found in prokaryotes (116, 117) use metabolites that are less integrated into central metabolism.

Various carboxylases, and decarboxylases run in reverse, have been engineered to produce organic molecules using CO_2 as a substrate (119). Decarboxylases often display worse thermodynamics (**Supplemental Figure 3**) and kinetics than carboxylases but in many cases can be forced to catalyze carboxylation especially in the presence of elevated CO_2 (120, 121). Among carboxylases, rubisco is unique in several respects:

1. It is the only one to carboxylate RuBP and thus uniquely compatible with the CBB cycle.
2. Among carboxylases, it is the only one that reacts promiscuously with O_2.
3. It is in a small minority of carboxylases that do not directly couple activity to an energetic cofactor (other CBB reactions supply the required energy and reducing potential).

Some clear advantages exist among alternative carboxylases:

1. Many carboxylases use the more accessible substrate bicarbonate, which is easier to bind due to its charge and high concentration (typically ~100 times the concentration of CO_2 in water at intracellular pH). These enzymes gain the ability to separate substrate binding from reactivity by using an ATP to convert bicarbonate to carboxyphosphate in a classic umpolung mechanism wherein a nucleophile becomes an electrophile (122).
2. Some enzymes reduce their substrates concomitant with fixation, which typically reduces the number of ATPs hydrolyzed in their pathways overall, improving energetic efficiency (123).

Despite rubisco's drawbacks, it is unlikely that a scalable, engineered pathway will dethrone the CBB in the near future. Many carboxylases would be difficult to integrate into carbon-fixation pathways because their substrates and products are not part of central metabolism [e.g., the aromatic carboxylases (**Table 1**; for more detail, see **Supplemental File 2**)]. Alternative synthetic carbon-fixation cycles using other carboxylases have also been proposed (124–127), and engineering these cycles in vivo is an ongoing grand challenge for synthetic biology. A variant of the CBB,

on the other hand, has been successfully engineered into *E. coli* strains, converting them from heterotrophs to autotrophs (128, 129). This was possible, in part, because *E. coli* lacks only two CBB enzymes—*prk* and rubisco—the remainder being part of the ubiquitous pentose phosphate pathway. As synthetic biology approaches to metabolic engineering become more widespread, extensive pathway engineering and host integration may open the door to replacing the CBB in microbes and eventually in plants.

RUBISCO EVOLUTION

After the initial discovery of a distinction between Form I L_8S_8 plant rubiscos and Form II L_2 microbial rubiscos (45), additional forms have been discovered. Form III rubiscos were discovered in 1999 (130, 131); they are distinguished from other forms by their presence in archaea and their association with non-CBB pathways. They account for a large proportion of the diversity of rubiscos, and many do not fall neatly into well-defined clades (**Figure 4*c***; for details, see the **Supplemental Text**; for sequences used to produce the trees, see **Supplemental File 3**; and for tree files, see **Supplemental Files 4–7**). There is no formal threshold for the establishment of a new form, and the phylogeny and nomenclature of these proteins remains in flux.

Rubisco enzymes are not always associated with the CBB cycle (132). Some plants use rubisco outside of the CBB to improve the efficiency of seed oil biosynthesis (133). In some heterotrophic bacteria the CBB can serve as a secondary electron sink (132). Form II/III and III rubiscos typically serve roles in nucleoside salvage (134, 135) or the reductive hexulose-phosphate pathway (136) and not in any version of the CBB. They are found in genomes apparently lacking phosphoribulokinase (*prk*) genes encoding the enzyme immediately upstream of rubisco in the CBB cycle. However, exceptions have been found in some Gottesmanbacteria, Deltaproteobacteria, and Chloroflexi species; they have *prk* in their genomes and are proposed to operate some modified form of the CBB (137, 138). In our phylogenetic analysis, we find that the latter group of rubiscos branch adjacent to the most divergent Form I rubiscos (compare Form III-transaldolase, indicated by the asterisk, to Form Iα in **Figure 4*c***), inviting evolutionary speculation that would benefit from additional structural studies and metagenomic sequencing.

Form IV rubiscos were discovered in 2000 (139) and have alternately been called rubisco-like proteins (RLP)—hereafter we use rubisco as a term that excludes RLPs—because they catalyze reactions other than RuBP carboxylation. Due to their clear homology to rubiscos, RLPs were immediately recognized as either evolutionary precursors or derivatives (139). A number of different chemistries have been identified among various RLP clades, and one clade is known to harbor RLPs that catalyze at least three chemical reactions (**Figure 4*c***). Despite recent discoveries of RLP chemistries (95, 140), the majority of named clades have no known function and some clades may be diverse enough to harbor additional chemistries (e.g., DeepYkr). Some RLPs can complement knockouts of others (141), as can bona fide rubiscos (142). Thus, despite extreme divergence in sequence, the underlying mechanisms of the various RLPs and rubisco are similar enough to catalyze each others' reactions (though no carboxylating RLP has yet been discovered). It is unknown if this plasticity is a contributor to trade-offs in rubisco catalysis.

Sequences from environmental samples have repeatedly shed light on the evolutionary path of rubisco. The observation that the rubisco tree has many deeply branching archaeal Form III rubiscos led to the hypothesis that rubisco evolution began here and, through many instances of horizontal gene transfer (HGT) and structural innovations, led to the phylogenetic distribution observed today (**Figure 4*a***) (27). By this hypothesis, the ancestor was a bona fide rubisco (**Figure 4*b***). An alternative hypothesis, based on mechanistic considerations and structural parsimony, is that bona fide rubiscos evolved from an ancestral RLP (143). Rubiscos and all RLPs

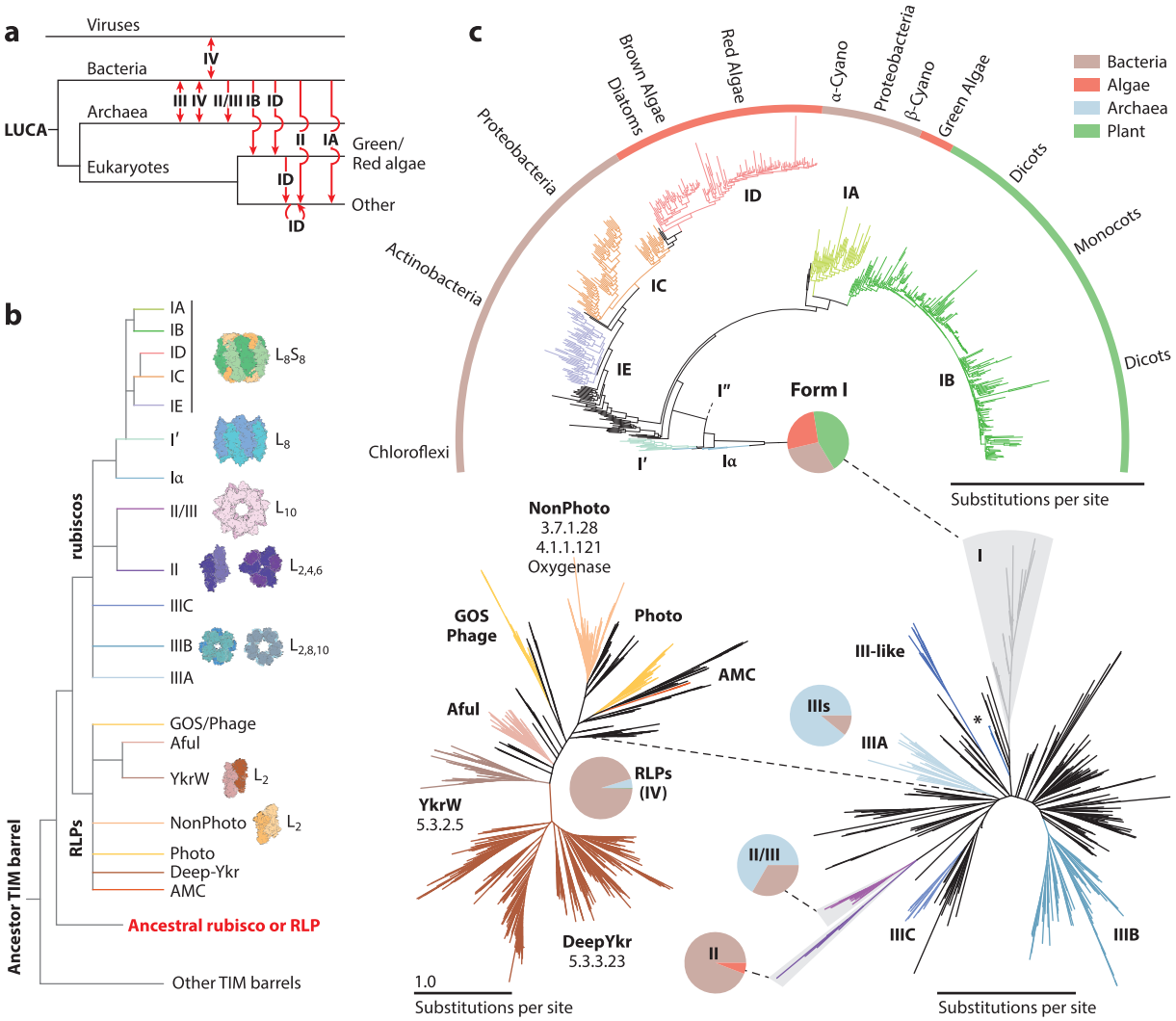


Figure 4

(*a*) A schematic showing horizontal gene transfer over the history of rubisco evolution. The event labeled IB is the initial endosymbiosis of cyanobacteria to form green algae. The ID event is the transfer of proteobacterial rubisco into the red algal chloroplast (large and small subunits together). Further ID transfers represent secondary or higher level endosymbioses. The Form II transfer is into the dinoflagellate nucleus. The last Form IA represents transfer from a cyanobacteria into *Paulinella*. (*b*) Simplified diagram of the evolution of the rubisco protein family. The outgroup labeled in red represents the last common ancestor of RLPs and rubisco. It remains an open question what reaction(s) that ancestor catalyzed. Branch lengths in this tree are meaningless. Representative crystal structures of oligomeric complexes are displayed for each clade where they are available (for PDB IDs, see **Supplemental File 1**). (*c*) A tree of RLPs (*bottom left*), all rubiscos (*center*), and Form I rubiscos (*top right*). Host organisms are indicated over the Form I tree but are not exclusive. Pie charts and the color bar over the Form I tree represent proportions of clades from a 90% identity dereplication clustering and are colored according to the legend at the top right. The asterisk indicates the location of the transaldolase Form III rubisco variants (137). Scale bars represent the number of substitutions per site. Abbreviations: LUCA, last universal common ancestor of cellular life; AMC, acid mine consortia; GOS, global ocean sequencing program; PDB ID, Protein Data Bank identifier; photo, found in phototrophic organisms; RLP, rubisco-like protein; TIM, triosephosphate isomerase.

use mechanisms that involve the formation of an enolate (**Supplemental File 1**), but only the mechanism of rubisco contains a second step where the enolate attacks CO_2—RLP enolates usually attack protons (143). However, one RLP does have a mechanism with a hydrolysis step, like rubisco (140), and another exclusively catalyzes oxygenation of the same intermediate as rubisco, bending the definition of RLPs (95). The discovery of a suitable TIM-barrel protein to act as an outgroup to the rubisco superfamily could settle this debate (**Figure 4*b***).

More recent evolutionary developments have been more amenable to phylogenetic analysis. HGT of rubiscos is widespread (68, 138). Plants acquired their Form IB rubiscos from cyanobacteria during endosymbiosis, with subsequent partitioning of the LSU and SSU between the chloroplast and nuclear genomes, respectively. Plant SSUs evolve faster than the LSU in part as a result of this partition (144). Proteobacteria often have multiple Form I and II rubiscos in their genomes, with different biochemical characteristics, perhaps as a response to their mixotrophic lifestyles (74). Although the chloroplasts of green and nongreen algae derive from the same endosymbiotic event, Form ID rubiscos in nongreen algae are derived from an evolutionary transfer of a Form IC rubisco to their chloroplast genome (LSU and SSU), possibly from an alphaproteobacteria (145). One exception is in dinoflagellates, which have a Form II rubisco acquired from other alphaproteobacteria (146) in their nuclear genomes. A group of rubiscos were also recently found in the genomes of Myoviridae phage and their hosts, Beckwithbacteria (138). This raises the possibility that viruses act as a vehicle for HGT across domains of life. The study of Form I rubisco evolution has benefitted from a clear root in the tree (**Figure 4*c***); evolutionary intermediates have been inferred through metagenomic sampling (79, 147). One such rubisco from *Promineofilum breve* has been termed a Form I′ rubisco because it lacks an SSU, though it still has an L_8 oligomeric form and catalyzes the standard rubisco reaction, albeit with low specificity for CO_2 (79). Further analysis of another L_8 clade, Form I″, has traced the evolution of the SSU through ancestral reconstructions (148). These reconstructions suggest that the SSU plays a proteostatic role in maintaining active enzyme complexes in the face of the destabilization incurred by evolutionary selection for higher selectivity against oxygenation.

There have been a number of attempts to capture the natural biochemical diversity of rubiscos in order to better constrain the range of possible kinetic parameters available to the enzyme family. In order to find more divergent rubisco sequences, metagenomic libraries have been assembled and then cloned into bacterial strains that have rubisco knocked out. Metagenomic libraries can come from sources like soil, rivers (149), hydrothermal vents (150), ocean water (151), or deep groundwater stimulated with acetate to promote bacterial growth (149), and the bacterial strains can be photoheterotrophs like *Rhodobacter capsulatus* (149, 151) or simply *E. coli* (150); metagenomic rubiscos can be purified from these sources for characterization. Rubiscos from these samples typically perform worse in in vivo complementation or in vitro than known rubisco genes, but vast new regions of sequence space remain unexplored, and several new rubisco clades have recently been discovered [e.g., Form III-like, Form Iα (138, 147)]. Perhaps the most compelling instance of this approach was a recent survey of Form II rubisco diversity, which uncovered the fastest rubisco yet from a betaproteobacterial *Gallionella* species (28), i.e., the rubisco with the highest $k_{cat,C}$ measured thus far. Cyanobacterial rubiscos have generally been assumed to be the fastest variants, so this result suggests there may be additional untapped diversity in natural sequences.

Rubisco is a slowly evolving enzyme (144), so it is possible that the observed biochemical limitations are a function of phylogenetic constraint—rubiscos may be hampered by their ancestry and not underlying biochemical constraints (107). This interpretation has been called into question (108), since it is clear that in the case of rubiscos with CCMs, evolution is at least fast enough to adjust rubisco kinetics to match local CO_2 concentrations (7). Further sampling and characterization promises to constrain the strength of this phylogenetic signal (107). This may help determine

to what degree plant rubiscos are evolving too slowly in a changing environment. Metagenomic sampling also promises to reveal additional evolutionary steps and to expand the known functional repertoire of the rubisco superfamily (many Form III rubiscos and RLPs are not found in defined groups).

RUBISCO ENGINEERING

The motivation for mutagenizing rubisco has been to both test mechanistic hypotheses and improve rubisco's function in order to enhance photosynthesis. Point mutations are the smallest steps that can be taken in sequence space and, for many years, were the only mutations accessible. Point mutations near the active site were instrumental in uncovering mechanistic steps (13) but have not been successful in generating enhanced enzymes. Mutations have been installed in many key locations on the enzyme, including the active site and loop 6, without much improvement (21). This is not surprising as it is exceedingly difficult to predict the effect of a mutation on an enzyme's efficiency, and it is unlikely that a point mutation would have a strong positive effect on catalysis, since evolution would be very likely to find such a modification. Some improvements have been reported, but they are generally weak or else come with a trade-off, for instance, improved $S_{C/O}$ at the cost of a reduction in $k_{cat,C}$ (152).

Laboratory Evolution, Trade-Offs, and Optimization

Directed evolution is the intentional improvement of biological fitness through iterative selection (**Figure 5**). It is inherently a test of trade-offs and also a tool to explore the biochemical mechanism. There are two components to any experiment in molecular evolution: a library of variants (**Figure 5*a***) and a selection or assay system (**Figure 5*b***). The construction of a library must integrate the goal of the experiment (e.g., is there a particular part of the protein to target?) with the

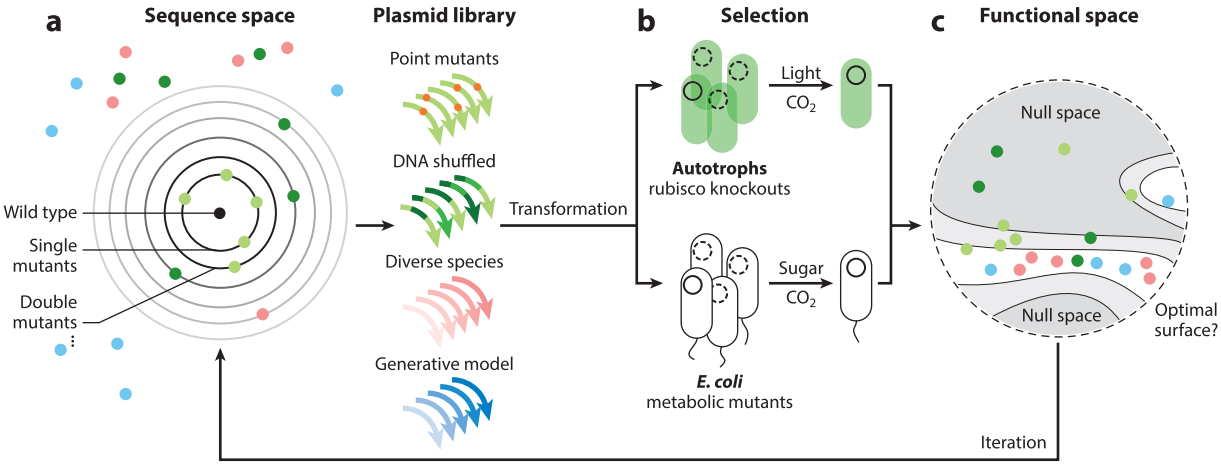


Figure 5

(*a*) Mutant libraries are made in a variety of ways, resulting in different distributions throughout sequence space. Libraries are made through environmental sampling, random mutagenesis, or synthesized oligo pools. Distance in sequence space is measured in number of mutations. Sequence space is discrete but high dimensional. (*b*) Genetic screening for functional rubisco can be accomplished using either rubisco knockout strains of autotrophs or rubisco-dependent mutants of *Escherichia coli*. (*c*) Mutants that survive selection may have clustered properties. Analysis of these clusters may reveal optimal surfaces in functional space. Functional space is also high dimensional. Dimensions include but are not limited to $k_{cat,C}$, $k_{cat,O}$, K_C, K_O, T_m, $K_{M,RuBP}$, k_3, k_6, k_5, k_8, K_E (the enolization constant), turnover number, maximum expression level, average activation level, $k_{on,RuBP}$, $k_{off,3PG}$, $k_{off,2PG}$, etc.

constraints of the assay (e.g., how many variants can be tested?) (153). Libraries can be made in both random and nonrandom fashions, such as through error-prone PCR or cloning with designed oligonucleotides, respectively. At ∼1,500 bases, the gene for the LSU of rubisco presents many challenges relative to model proteins in that the inherent combinatorics leads to large libraries, complicating cloning, and more difficult assays when using modern tools such as short-read DNA sequencing. Recent advances in machine learning, drawing on both sequence and structure, have improved the design of libraries for better exploration of the sequence–function landscape (154, 155), but these are just beginning to be applied to rubisco (37, 156).

Selection systems must be developed individually to match the desired protein function. For rubisco, the desired outcome of protein evolution is generally faster $k_{cat,C}$, lower K_C, or higher $S_{C/O}$ (ideally all three). Two main modes of selection systems have been developed: genetically tractable autotrophs with replaceable rubisco genes and rubisco-dependent *E. coli* mutants (**Figure 5*b***). In the case of autotrophs, their growth is rubisco limited under conditions where they have sufficient energetic resources. Species that have been used to test rubisco replacements include oxygenic photoautotrophs like the model cyanobacteria *Synechocystis* sp. PCC6803 (22), as well as facultative autotrophs like *R. capsulatus* (26, 157), *Rhodopseudomonas palustris* (158, 159), or *Cupriavidus necator* (160).

The first selection system achieved in *E. coli* relied on rubisco alleviating toxicity induced by *prk* expression (161). RuBP toxicity is a result of it being a metabolic dead-end. No enzyme in *E. coli* makes use of this metabolite, so it shunts away useful carbon and builds up in the cell. Rubisco takes RuBP as a substrate and can alleviate this toxicity. This strategy is inherently prone to cheaters that disable *prk* function through mutations. An improved strategy was developed where the *prk* gene is fused to an antibiotic resistance gene, preventing most cheater mutations (162). Implementation in an *E. coli* strain that is especially susceptible to *prk* toxicity enhanced the system further (163).

Other selection systems in *E. coli* are built on the central placement of 3PG, the product of rubisco catalysis; 3PG sits at the nexus of four central metabolic pathways (**Supplemental Figure 2**). The first strategy proposed involved a deletion of the GAPDH gene (8, 23). This metabolic lesion allows *E. coli* to grow only when provided with two carbon sources: one for sugar synthesis and upper glycolysis and another for lipid and energetic pathways including the TCA cycle. GAPDH is a critical link between these two parts of *E. coli* metabolism that has no redundant replacement in the genome. It can, however, be replaced by the combined actions of *prk* and rubisco, which form a link between upper and lower metabolism through the pentose phosphate pathway. *E. coli* with GAPDH removed and *prk*/rubisco added can, in principle, survive on one carbon source alone that would trickle down to the lower metabolism through rubisco, permitting growth. In practice, *E. coli* could not be grown in this mode and required a small amount of added amino acids (164); the resulting selection worked in part because of the same principle of RuBP detoxification. An algorithm was developed to automate the generation of selective strains (129). One such strain was a phosphoglycerate mutase knockout that was used to evolve an *E. coli* strain capable of growth using rubisco, first to supply all of the upper metabolites in the cell (129) and eventually all carbon biomass (128). Likewise, a ribose-5-phosphate isomerase knockout was used to investigate the function of CCMs (114). Neither of these latter strains have yet been used to assay rubisco libraries, though additional selection systems have already been proposed (165).

Many different starting points have been chosen for these selections including form I, II, II/III, and III rubiscos (**Figure 5*a***). Because it is difficult to distinguish between improved kinetics and increased expression, the majority of hits from these selections have been rubisco mutants with higher expression levels in the host, often as a result of improved folding and stability. Improvements in kinetic parameters have been modest with a few exceptions. A twofold increase in k_{cat} was achieved in the Form II/III rubisco from the extremophile *Methanococcoides burtonii*, though this

enzyme had a very low starting rate, implying that it was not optimized for catalysis at room temperature (24). In two cases, a twofold improvement in k_{cat} was achieved by changing the protein chaperone complement in *E. coli* (21, 25). Due to inconsistencies in the reported rates of rubiscos between different laboratories, it can be difficult to assess or explain these improvements. The selected mutations were all far from the active site so the mechanistic basis for these biochemical improvements remains to be explained. Additional kinetic measurements and structural studies are necessary.

Plant Synthetic Biology

One major goal in rubisco engineering is to improve plant growth. As atmospheric CO_2 levels climb, it may be possible to improve plant growth by simply replacing the rubisco genes of crop plants with enzymes that are adapted to higher CO_2 levels, like those of cyanobacteria. It has been possible to edit the chloroplast genome of the model plant *N. tabacum* since 1993 (166), and soon thereafter, a transgenic line with the rubisco LSU moved to the nucleus was created (167). Since then, a variety of rubiscos of different forms have been expressed in the tobacco chloroplast, in some cases with specific chaperones (**Supplemental Text** and **Supplemental Table 1**). These mutant tobacco plants grow uniformly more slowly than wild type, even in chambers with CO_2 elevated to 1%. They also have higher CO_2 compensation points (i.e., the CO_2 concentration where there is net assimilation of carbon), in many cases because the rubiscos used have lower CO_2 affinity than the endogenous tobacco gene. In nearly all cases, the cause of poor growth is clear, as rubisco expression levels are reduced by an order of magnitude compared to wild type (168). In one case, rubisco levels were reduced by only a factor of two, and growth at high CO_2 was comparable to wild type (169). Likewise, extremely similar rubiscos have been swapped with minimal effect (170). Although the Whitney lab (171) has developed a tobacco line with *R. rubrum* rubisco that is more readily transformable, chloroplast transformation remains an enormous bottleneck, and all such studies suffer from the ability to test few rubisco variants *in planta*. Because the SSU is found in the nuclear genome of most plants, modification of the SSU or its genetic regulatory elements present a possible alternative that avoids this bottleneck (reviewed in depth in 77), although further research is required to determine the compatibility of different SSUs and their effects on rubisco catalysis *in planta*.

PERSPECTIVES AND FUTURE WORK

Despite decades of research, the promises of rubisco engineering remain unrealized at every level. There are still deep unanswered questions about the chemical mechanism of catalysis and what implications it has on the prospects for improving rubisco carboxylation rate and specificity. The evolutionary constraints on the enzyme are unknown and poorly sampled. Massively improved rubisco variants have not emerged, and improvements in plant photosynthesis have not been achieved through rubisco replacement. On the other hand, much progress has been made in revealing the chemical mechanism of oxygenation (14, 89). Metagenomic mining has provided an enormous wealth of rubisco isoforms to study mechanism and evolution (138). High-throughput biochemistry has recently uncovered a record-breaking rubisco variant (28), and improvements in plant growth have been achieved through increasing rubisco levels (172, 173).

One source of future advances may be from additional detailed biochemical characterization of phylogenetically distant rubisco isoforms that can better constrain the full breadth of rubisco catalytic possibilities. This knowledge will come in the form of a sharper account of the etiology of trade-offs. An added benefit of characterizations of previously overlooked rubisco isoforms is a new source of starting points for directed evolution. The specificity of plant chaperones for their

cognate rubiscos also remains a large barrier to heterologously assaying and engineering variants, and deciphering this code will surely accelerate advances on Form I.

Continued improvements in selection systems (**Supplemental Figure 2**) are essential in order to test larger libraries that would unlock the benefits of computational design approaches. Advances in machine learning models, coupled with mounting empirical sequences and structural data, will also help explore the functional regions of the sequence landscape, which is vast due to the large size of rubisco. For example, many bacterial rubiscos apparently do not have the same chaperone requirements as relatively similar plant variants. Machine learning methods (154, 155) may be able to recombine sequences in a productive fashion in order to better understand folding landscapes and to isolate variants combining the most desirable features of Form I enzymes. Alternatively, even if future experimentation shows that rubisco is constrained by trade-offs, laboratory evolution promises to reveal the precise shape of those constraints (the optimal surface in **Figure 5*c***) and to test hypotheses in a proof-by-synthesis fashion. For example, current trade-off models suggest that rubiscos with higher k_{cat}s and lower specificities should be evolvable from slower and more specific variants.

It is likely that the complexity of expressing rubisco *in planta* has limited experimental successes, and so improvements in plant genome editing are key to translating rubisco improvements into the field. Successful heterologous expression of rubisco *in planta* depends on the ability to test a large number of alternative gene cassettes in order to balance gene expression of the LSUs and SSUs along with all of their associated chaperones. This is especially true for Form IB and ID rubiscos that have a heavy dependence on the proteostatic machinery for folding and activation (87, 174, 175). Of particular importance is further improving chloroplast genome editing technology, which currently suffers from low throughput and long time horizons. Even if biochemical improvements fail to materialize, heterologous replacement of rubisco from red algae is predicted to be advantageous (175, 176), due to the slow evolution of rubisco in plants (144).

If direct improvements to rubisco fail to materialize, alternative strategies for photosynthetic engineering that work around the constraints of rubisco have been proposed and are being pursued (177). There are several successes in the engineering of pathways closely related to extant plant physiology, such as in the case of the photorespiratory bypass (178) or increased expression of CBB components (179), which have led to double-digit percent yield increases. Alternatively, there are more ambitious proposals for reengineering plant physiology with CCM components that could offer greater gains but that come with the downside of significant complexity (113, 114, 180). Looking further into the future, it may eventually be possible to circumvent the downsides of rubisco entirely by switching to alternative carbon-fixation cycles, either natural or artificial. Until then, rubisco engineering holds a key position in efforts to improve photosynthesis.

DISCLOSURE STATEMENT

D.F.S. is a cofounder and scientific advisory board member of Scribe Therapeutics. This company is not involved in this work in any way. The remaining authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Jacob West-Roberts, Alex Jaffe, Jill Banfield, and Nat Thompson for helpful conversations on building rubisco trees. We thank Jack Desmarais and Elad Noor for invaluable advice related to thermodynamic interpretations and help with data analysis. We thank Abhishek Bhatt for help in assembling the carboxylase table. We thank Avi Flamholz, as well as Luke Oltrogge

and Muntathar Al-Shimary, for their insightful comments when preparing the manuscript. The work was supported by the Howard Hughes Medical Institute, National Institutes of Health K99 award 5K99GM141455-02 to N.P., National Science Foundation grant MCB-1818377 to D.F.S., and US Department of Energy grant DE-SC00016240 to D.F.S.

LITERATURE CITED

1. Raven JA. 2009. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* 56:177–92
2. Bar-On YM, Milo R. 2019. The global mass and average rate of rubisco. *PNAS* 116(10):4738–43
3. Andersson I. 1996. Large structures at high resolution: the 1.6 Å crystal structure of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate. *J. Mol. Biol.* 259:160–74
4. Lundqvist T, Schneider G. 1991. Crystal structure of the ternary complex of ribulose-1,5-bisphosphate carboxylase, Mg(II) and activator CO_2 at 2.3-Å resolution. *Biochemistry* 30:904–8
5. Fischer WW, Hemp J, Johnson JE. 2016. Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* 44:647–83
6. Tcherkez GGB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *PNAS* 103:7246–51
7. Savir Y, Noor E, Milo R, Tlusty T. 2010. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *PNAS* 107:3475–80
8. Morell MK, Paul K, Kane HJ, Andrews TJ. 1992. Rubisco: maladapted or misunderstood. *Aust. J. Bot.* 40:431–41
9. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, et al. 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–10
10. Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. 2018. Rubisco is not really so bad. *Plant Cell Environ*. 41:705–16
11. Zhan C-G, Niu S, Ornstein RL. 2001. Theoretical studies of nonenzymatic reaction pathways for the three reaction stages of the carboxylation of ribulose-1,5-bisphosphate. *J. Chem. Soc. Perkin Trans. 2* 2001:23–29
12. Farquhar GD. 1979. Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Arch Biochem. Biophys.* 193:456–68
13. Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, Lorimer GH. 1998. Mechanism of rubisco: the carbamate as general base. *Chem. Rev.* 98:549–62
14. Bathellier C, Yu L-J, Farquhar GD, Coote ML, Lorimer GH, Tcherkez G. 2020. Ribulose 1,5-bisphosphate carboxylase/oxygenase activates O_2 by electron transfer. *PNAS* 117:24234–42
15. Tawfik DS. 2014. Accuracy-rate tradeoffs: How do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* 21:73–80
16. Bar-Even A, Salah Tawfik D. 2013. Engineering specialized metabolic pathways—is there a room for enzyme improvements? *Curr. Opin. Biotechnol.* 24:310–19
17. Arnold FH. 2018. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed. Engl.* 57:4143–48
18. Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, et al. 2019. Revisiting trade-offs between rubisco kinetic parameters. *Biochemistry* 58:3365–76
19. Gutteridge S, Lorimer G, Pierce J. 1988. Details of the reactions catalysed by mutant forms of rubisco. *Plant Physiol. Biochem.* 26:675–82
20. Gutteridge S, Sigal I, Thomas B, Arentzen R, Cordova A, Lorimer G. 1984. A site-specific mutation within the active site of ribulose-1,5-bisphosphate carboxylase of *Rhodospirillum rubrum*. *EMBO J.* 3:2737–43
21. Cai Z, Liu G, Zhang J, Li Y. 2014. Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. *Protein Cell* 5:552–62

22. Amichay D, Levitz R, Gurevitz M. 1993. Construction of a *Synechocystic* PCC6803 mutant suitable for the study of variant hexadecameric ribulose bisphosphate carboxylase/oxygenase enzymes. *Plant Mol. Biol.* 23:465–76
23. Mueller-Cajar O, Whitney SM. 2008. Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynth Res*. 98:667–75
24. Wilson RH, Alonso H, Whitney SM. 2016. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci. Rep.* 6:22284
25. Durão P, Aigner H, Nagy P, Mueller-Cajar O, Hartl FU, Hayer-Hartl M. 2015. Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* 11:148–55
26. Satagopan S, Huening KA, Tabita FR. 2019. Selection of cyanobacterial (*Synechococcus* sp. strain PCC 6301) RubisCO variants with improved functional properties that confer enhanced CO_2-dependent growth of *Rhodobacter capsulatus*, a photosynthetic bacterium. *MBio* 10(4):e01537-19
27. Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* 59:1515–24
28. Davidi D, Shamshoum M, Guo Z, Bar-On YM, Prywes N, et al. 2020. Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.* 39:e104081
29. Farquhar GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* 149:78–90
30. Wu A, Brider J, Busch FA, Chen M, Chenu K, et al. 2022. A cross-scale analysis to understand and quantify effects of photosynthetic enhancement on crop growth and yield. bioRxiv 2022.07.06.498957. **http://doi.org/10.1101/2022.07.06.498957**
31. Busch FA, Sage RF, Farquhar GD. 2018. Plants increase CO_2 uptake by assimilating nitrogen via the photorespiratory pathway. *Nat. Plants* 4:46–54
32. Bloom AJ, Lancaster KM. 2018. Manganese binding to Rubisco could drive a photorespiratory pathway that increases the energy efficiency of photosynthesis. *Nat. Plants* 4:414–22
33. Sinclair TR, Rufty TW, Lewis RS. 2019. Increasing photosynthesis: unlikely solution for world food problem. *Trends Plant Sci*. 24(11):1032–39
34. Raines CA. 2003. The Calvin cycle revisited. *Photosynth. Res.* 75:1–10
35. Iñiguez C, Aguiló-Nicolau P, Galmés J. 2021. Improving photosynthesis through the enhancement of Rubisco carboxylation capacity. *Biochem. Soc. Trans.* 49:2007–19
36. Zhu X-G, Portis AR Jr., Long SP. 2004. Would transformation of C_3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant Cell Environ*. 27:155–65
37. Iqbal WA, Miller IG, Moore RL, Hope IJ, Cowan-Turner D, Kapralov MV. 2021. Rubisco substitutions predicted to enhance crop performance through carbon uptake modelling. *J. Exp. Bot.* 72:6066–75
38. Singh AK, Santos-Merino M, Sakkos JK, Walker BJ, Ducat DC. 2022. Rubisco regulation in response to altered carbon status in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Physiol*. 189(2):874–88
39. Kawashima N, Wildman SG. 1971. Studies on fraction-I protein. I. Effect of crystallization of fraction-I protein from tobacco leaves on ribulose diphosphate carboxylase activity. *Biochim. Biophys. Acta Protein Struct.* 229:240–49
40. Wildman SG. 2002. Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate carboxylase-oxygenase). *Photosynth. Res.* 73:243–50
41. Dorner RW, Kahn A, Wildman SG. 1957. The proteins of green leaves. VII. Synthesis and decay of the cytoplasmic proteins during the life of the tobacco leaf. *J. Biol. Chem.* 229:945–52
42. Quayle JR, Fuller RC, Benson AA, Calvin M. 1954. Enzymatic carboxylation of ribulose diphosphate. *J. Am. Chem. Soc.* 76:3610–11
43. Bassham JA, Benson AA, Kay LD, Harris AZ, Wilson AT, Calvin M. 1954. The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. *J. Am. Chem. Soc.* 76:1760–70
44. Weissbach A, Horecker BL, Hurwitz J. 1956. The enzymatic formation of phosphoglyceric acid from ribulose diphosphate and carbon dioxide. *J. Biol. Chem.* 218:795–810

45. Anderson LE, Price GB, Fuller RC. 1968. Molecular diversity of the ribulose-1,5-diphosphate carboxylase from photosynthetic microorganisms. *Science* 161:482–84
46. Bowes G, Ogren WL, Hageman RH. 1971. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophys. Res. Commun.* 45:716–22
47. Lorimer GH, Andrews TJ, Tolbert NE. 1973. Ribulose diphosphate oxygenase. II. Further proof of reaction products and mechanism of action. *Biochemistry* 12:18–23
48. Lorimer GH, Miziorko HM. 1980. Carbamate formation on the ε-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by CO_2 and Mg^{2+}. *Biochemistry* 19:5321–28
49. Lorimer GH. 1981. Ribulosebisphosphate carboxylase: amino acid sequence of a peptide bearing the activator carbon dioxide. *Biochemistry* 20:1236–40
50. Lorimer GH, Badger MR, Andrews TJ. 1976. The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry* 15:529–36
51. Miziorko HM. 1979. Ribulose-1,5-biphosphate carboxylase. Evidence in support of the existence of distinct CO_2 activator and CO_2 substrate sites. *J. Biol. Chem.* 254:270–72
52. Cooper TG, Filmer D, Wishnick M, Lane MD. 1969. The active species of "CO_2" utilized by ribulose diphosphate carboxylase. *J. Biol. Chem.* 244:1081–83
53. Saver BG, Knowles JR. 1982. Ribulose 1,5-bisphosphate carboxylase: enzyme-catalyzed appearance of solvent tritium at carbon 3 of ribulose 1,5-bisphosphate reisolated after partial reaction. *Biochemistry* 21:5398–403
54. Sue JM, Knowles JR. 1982. Ribulose-1,5-bisphosphate carboxylase: fate of the tritium label in [3-^{3}H]ribulose 1,5-bisphosphate during the enzyme-catalyzed reaction. *Biochemistry* 21:5404–10
55. Gutteridge S, Parry MAJ, Schmidt CNG, Feeney J. 1984. An investigation of ribulosebisphosphate carboxylase activity by high resolution ^{1}H NMR. *FEBS Lett.* 170:355–59
56. Kawashima N, Wildman SG. 1972. Studies on fraction I protein. IV. Mode of inheritance of primary structure in relation to whether chloroplast or nuclear DNA contains the code for a chloroplast protein. *Biochim. Biophys. Acta Nucleic Acids Protein Synth.* 262:42–49
57. Chan PH, Wildman SG. 1972. Chloroplast DNA codes for the primary structure of the large subunit of fraction I protein. *Biochim. Biophys. Acta Nucleic Acids Protein Synth.* 277:677–80
58. McIntosh L, Poulsen C, Bogorad L. 1980. Chloroplast gene sequence for the large subunit of ribulose bisphosphatecarboxylase of maize. *Nature* 288:556–60
59. Niyogi SK, Foote RS, Mural RJ, Larimer FW, Mitra S, et al. 1986. Nonessentiality of histidine 291 of *Rhodospirillum rubrum* ribulose-bisphosphate carboxylase/oxygenase as determined by site-directed mutagenesis. *J. Biol. Chem.* 261:10087–92
60. Hartman FC, Soper TS, Niyogi SK, Mural RJ, Foote RS, et al. 1987. Function of Lys-166 of *Rhodospirillum rubrum* ribulosebisphosphate carboxylase/oxygenase as examined by site-directed mutagenesis. *J. Biol. Chem.* 262:3496–501
61. Schneider G, Lindqvist Y, Brändén CI, Lorimer G. 1986. Three-dimensional structure of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* at 2.9 Å resolution. *EMBO J.* 5:3409–15
62. Chapman MS, Suh SW, Cascio D, Smith WW, Eisenberg D. 1987. Sliding-layer conformational change limited by the quaternary structure of plant RuBisCO. *Nature* 329:354–56
63. Sharkey TD. 2022. The discovery of rubisco. *J. Exp. Bot.* 74:510–19
64. Li H, Sawaya MR, Tabita FR, Eisenberg D. 2005. Crystal structure of a RuBisCO-like protein from the green sulfur bacterium *Chlorobium tepidum*. *Structure* 13:779–89
65. Andersson I, Knight S, Schneider G, Lindqvist Y, Lundqvist T, et al. 1989. Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature* 337:229–34
66. Schreuder HA, Knight S, Curmi PM, Andersson I, Cascio D, et al. 1993. Crystal structure of activated tobacco rubisco complexed with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate. *Protein Sci.* 2:1136–46
67. Taylor TC, Andersson I. 1996. Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the activated unliganded spinach enzyme. *Nat. Struct. Biol.* 3:95–101

68. Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. 2007. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol. Mol. Biol. Rev.* 71:576–99
69. Spreitzer RJ, Salvucci ME. 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* 53:449–75
70. Andersson I, Backlund A. 2008. Structure and function of Rubisco. *Plant Physiol. Biochem.* 46:275–91
71. Wildman SG, Bonner J. 1947. The proteins of green leaves; isolation, enzymatic properties and auxin content of spinach cytoplasmic proteins. *Arch. Biochem.* 14:381–413
72. Nishimura, Takabe, Sugiyama. Structure and function of chloroplast proteins: XIX. Dissociation of spinach leaf ribulose-1,5-diphosphate carboxylase by mercuribenzoate. *J. Biochem.* 74:945–54
73. Baker TS, Suh SW, Eisenberg D. 1977. Structure of ribulose-1,5-bisphosphate carboxylase-oxygenase: form III crystals. *PNAS* 74:1037–41
74. Badger MR, Bek EJ. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *J. Exp. Bot.* 59:1525–41
75. Andrews TJ. 1988. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* 263:12213–19
76. Spreitzer RJ. 2003. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* 414:141–49
77. Mao Y, Catherall E, Díaz-Ramos A, Greiff GRL, Azinas S, et al. 2022. The small subunit of Rubisco and its potential as an engineering target. *J. Exp. Bot.* 74:543–61
78. Mallik S, Goloubinoff P, Tawfik DS. 2021. On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. *PNAS* 118:e2020885118
79. Banda DM, Pereira JH, Liu AK, Orr DJ, Hammel M, et al. 2020. Novel bacterial clade reveals origin of form I Rubisco. *Nat. Plants* 6:1158–66
80. Satagopan S, Chan S, Jeanne Perry L, Robert Tabita F. 2014. Structure-function studies with the unique hexameric form II ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) from *Rhodopseudomonas palustris*. *J. Biol. Chem.* 289:21433–50
81. Lin MT, Stone WD, Chaudhari V, Hanson MR. 2020. Small subunits can determine enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli*. *Nat. Plants* 6:1289–99
82. He S, Chou H-T, Matthies D, Wunder T, Meyer MT, et al. 2020. The structural basis of Rubisco phase separation in the pyrenoid. *Nat. Plants* 6:1480–90
83. Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF. 2020. Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* 27:281–87
84. Wang H, Yan X, Aigner H, Bracher A, Nguyen ND, et al. 2019. Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* 566:131–35
85. Blikstad C, Dugan EJ, Laughlin TG, Liu MD. 2021. Discovery of a carbonic anhydrase-Rubisco supercomplex within the alpha-carboxysome. bioRxiv 2021.11.05.467472. **https://doi.org/10.1101/2021.11.05.467472**
86. Whitney SM, Kane HJ, Houtz RL, Sharwood RE. 2009. Rubisco oligomers composed of linked small and large subunits assemble in tobacco plastids and have higher affinities for CO_2 and O_2. *Plant Physiol.* 149:1887–95
87. Bracher A, Whitney SM, Hartl FU, Hayer-Hartl M. 2017. Biogenesis and metabolic maintenance of rubisco. *Annu. Rev. Plant Biol.* 68:29–60
88. Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, et al. 2017. Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* 358:1272–78
89. Kannappan B, Cummins PL, Gready JE. 2019. Mechanism of oxygenase-pathway reactions catalyzed by rubisco from large-scale Kohn–Sham density functional calculations. *J. Phys. Chem. B* 123:2833–43
90. Tcherkez G. 2013. Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters. *Plant Cell Environ.* 36:1586–96
91. Lorimer GH, Andrews TJ, Pierce J, Schloss JV. 1986. 2′-carboxy-3-keto-D-arabinitol 1,5-bisphosphate, the six-carbon intermediate of the ribulose bisphosphate carboxylase reaction. *Philos. Trans. R. Soc. B* 313:397–407

92. Cummins PL, Kannappan B, Gready JE. 2018. Revised mechanism of carboxylation of ribulose-1,5-biphosphate by rubisco from large scale quantum chemical calculations. *J. Comput. Chem.* 39:1656–65
93. Tcherkez GGB, Bathellier C, Stuart-Williams H, Whitney S, Gout E, et al. 2013. D_2O solvent isotope effects suggest uniform energy barriers in ribulose-1,5-bisphosphate carboxylase/oxygenase catalysis. *Biochemistry* 52:869–77
94. Fried SD, Boxer SG. 2017. Electric fields and enzyme catalysis. *Annu. Rev. Biochem.* 86:387–415
95. Kim SM, Lim HS, Lee SB. 2018. Discovery of a RuBisCO-like protein that functions as an oxygenase in the novel D-hamamelose pathway. *Biotechnol. Bioprocess Eng.* 23:490–99
96. Robison PD, Martin MN, Tabita FR. 1979. Differential effects of metal ions on *Rhodospirillum rubrum* ribulosebisphosphate carboxylase/oxygenase and stoichiometric incorporation of HCO_3^- into a cobalt(III)–enzyme complex. *Biochemistry* 18:4453–58
97. Stec B. 2012. Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *PNAS* 109:18785–90
98. Valegård K, Andralojc PJ, Haslam RP, Pearce FG, Eriksen GK, et al. 2018. Structural and functional analyses of Rubisco from arctic diatom species reveal unusual posttranslational modifications. *J. Biol. Chem.* 293:13033–43
99. Bhat JY, Miličić G, Thieulin-Pardo G, Bracher A, Maxwell A, et al. 2017. Mechanism of enzyme repair by the AAA^+ chaperone rubisco activase. *Mol. Cell* 67:744–56.e6
100. Astier J, Rasul S, Koen E, Manzoor H, Besson-Bard A, et al. 2011. S-nitrosylation: an emerging post-translational protein modification in plants. *Plant Sci.* 181:527–33
101. Gutteridge S, Parry MAJ, Burton S, Keys AJ, Mudd A, et al. 1986. A nocturnal inhibitor of carboxylation in leaves. *Nature* 324:274–76
102. Khan S, Andralojc PJ, Lea PJ, Parry MAJ. 1999. 2′-Carboxy-D-arabitinol 1-phosphate protects ribulose 1,5-bisphosphate carboxylase/oxygenase against proteolytic breakdown. *Eur. J. Biochem.* 266:840–47
103. Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* 291:513–15
104. McNevin D, von Caemmerer S, Farquhar G. 2006. Determining RuBisCO activation kinetics and other rate and equilibrium constants by simultaneous multiple non-linear regression of a kinetic model. *J. Exp. Bot.* 57:3883–900
105. von Caemmerer S, Evans JR, Hudson GS, Andrews TJ. 1994. The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. *Planta* 195:88–97
106. Iñiguez C, Capó-Bauçà S, Niinemets Ü, Stoll H, Aguiló-Nicolau P, Galmés J. 2020. Evolutionary trends in RuBisCO kinetics and their co-evolution with CO_2 concentrating mechanisms. *Plant J.* 101:897–918
107. Bouvier JW, Emms DM, Rhodes T, Bolton JS, Brasnett A, et al. 2021. Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off. *Mol. Biol. Evol.* 38:2880–96
108. Tcherkez G, Farquhar GD. 2021. Rubisco catalytic adaptation is mostly driven by photosynthetic conditions – not by phylogenetic constraints. *J. Plant Physiol.* 267:153554
109. Tcherkez GG, Bathellier C, Farquhar GD, Lorimer GH. 2018. Commentary: Directions for optimization of photosynthetic carbon fixation: RuBisCO's efficiency may not be so constrained after all. *Front. Plant Sci.* 9:929
110. Cummins PL, Kannappan B, Gready JE. 2019. Response: Commentary: Directions for optimization of photosynthetic carbon fixation: RuBisCO's efficiency may not be so constrained after all. *Front. Plant Sci.* 10:1426
111. Badger MR, Sharwood RE. 2022. Rubisco, the imperfect winner: "It's all about the base." *J. Exp. Bot.* 74:562–80
112. Raven JA, Cockell CS, De La Rocha CL. 2008. The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philos. Trans. R. Soc. B* 363:2641–50
113. Hennacy JH, Jonikas MC. 2020. Prospects for engineering biophysical CO_2 concentrating mechanisms into land plants to enhance yields. *Annu. Rev. Plant Biol.* 71:461–85
114. Flamholz AI, Dugan E, Blikstad C, Gleizer S, Ben-Nissan R, et al. 2020. Functional reconstitution of a bacterial CO_2 concentrating mechanism in *Escherichia coli*. *eLife* 9:e59882

115. Sage RF. 2004. The evolution of C_4 photosynthesis. *New Phytol.* 161:341–70
116. Bar-Even A, Noor E, Milo R. 2012. A survey of carbon fixation pathways through a quantitative lens. *J. Exp. Bot.* 63:2325–42
117. Fuchs G. 2011. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* 65:631–58
118. Bar-Even A, Flamholz A, Noor E, Milo R. 2012. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta Bioener.* 1817:1646–59
119. Aleku GA, Roberts GW, Titchiner GR, Leys D. 2021. Synthetic enzyme-catalyzed CO_2 fixation reactions. *ChemSusChem* 14:1781–804
120. Martin J, Eisoldt L, Skerra A. 2018. Fixation of gaseous CO_2 by reversing a decarboxylase for the biocatalytic synthesis of the essential amino acid l-methionine. *Nat. Catal.* 1:555–61
121. Satanowski A, Dronsella B, Noor E, Vögeli B, He H, et al. 2020. Awakening a latent carbon fixation cycle in *Escherichia coli*. *Nat. Commun.* 11:5812
122. Knowles JR. 1989. The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* 58:195–221
123. Erb TJ. 2011. Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* 77:8466–77
124. Schwander T, Schada von Borzyskowski L, Burgener S, Cortina NS, Erb TJ. 2016. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 354:900–4
125. Bar-Even A, Noor E, Lewis NE, Milo R. 2010. Design and analysis of synthetic carbon fixation pathways. *PNAS* 107:8889–94
126. Zelcbuch L. 2015. *Implementing synthetic carbon fixation pathway in the model organism* Escherichia Coli. PhD Thesis, Weizmann Inst. Sci., Rehovot, Isr. **https://doi.org/10.34933/WIS.000033**
127. Trudeau DL, Edlich-Muth C, Zarzycki J, Scheffen M, Goldsmith M, et al. 2018. Design and in vitro realization of carbon-conserving photorespiration. *PNAS* 115:E11455–64
128. Gleizer S, Ben-Nissan R, Bar-On YM, Antonovsky N, Noor E, et al. 2019. Conversion of *Escherichia coli* to generate all biomass carbon from CO_2. *Cell* 179:1255–63.e12
129. Antonovsky N, Gleizer S, Noor E, Zohar Y, Herz E, et al. 2016. Sugar synthesis from CO_2 in *Escherichia coli*. *Cell* 166:115–25
130. Watson GM, Yu JP, Tabita FR. 1999. Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic *Archaea*. *J. Bacteriol.* 181:1569–75
131. Ezaki S, Maeda N, Kishimoto T, Atomi H, Imanaka T. 1999. Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1. *J. Biol. Chem.* 274:5078–82
132. Liu D, Ramya RCS, Mueller-Cajar O. 2017. Surveying the expanding prokaryotic Rubisco multiverse. *FEMS Microbiol. Lett.* 364:fnx156
133. Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432:779–82
134. Wrighton KC, Castelle CJ, Varaljay VA, Satagopan S, Brown CT, et al. 2016. RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J.* 10:2702–14
135. Sato T, Atomi H, Imanaka T. 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315:1003–6
136. Kono T, Mehrotra S, Endo C, Kizu N, Matusda M, et al. 2017. A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat. Commun.* 8:14007
137. Frolov EN, Kublanov IV, Toshchakov SV, Lunev EA, Pimenov NV, et al. 2019. Form III RubisCO-mediated transaldolase variant of the Calvin cycle in a chemolithoautotrophic bacterium. *PNAS* 116:18638–46
138. Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. 2019. Lateral gene transfer shapes the distribution of RuBisCO among candidate phyla radiation bacteria and DPANN archaea. *Mol. Biol. Evol.* 36:435–46
139. Hanson TE, Tabita FR. 2001. A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *PNAS* 98:4397–402
140. Carter MS, Zhang X, Huang H, Bouvier JT, Francisco BS, et al. 2018. Functional assignment of multiple catabolic pathways for D-apiose. *Nat. Chem. Biol.* 14:696–705

141. Kreel NE, Tabita FR. 2007. Substitutions at methionine 295 of *Archaeoglobus fulgidus* ribulose-1,5-bisphosphate carboxylase/oxygenase affect oxygen binding and CO_2/O_2 specificity. *J. Biol. Chem.* 282:1341–51
142. Dey S, North JA, Sriram J, Evans BS, Tabita FR. 2015. In vivo studies in *Rhodospirillum rubrum* indicate that ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) catalyzes two obligatorily required and physiologically significant reactions for distinct carbon and sulfur metabolic pathways. *J. Biol. Chem.* 290:30658–68
143. Erb TJ, Zarzycki J. 2018. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr. Opin. Biotechnol.* 49:100–7
144. Bouvier JW, Emms DM, Kelly S. 2022. Slow molecular evolution of rubisco limits adaptive improvement of CO2 assimilation. bioRxiv 2022.07.06.498985. **http://doi.org/10.1101/2022.07.06.498985**
145. Delwiche CF, Palmer JD. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol. Biol. Evol.* 13:873–82
146. Whitney SM, Shaw DC, Yellowlees D. 1995. Evidence that some dinoflagellates contain a ribulose-1,5-bisphosphate carboxylase/oxygenase related to that of the α-proteobacteria. *Proc. Biol. Sci.* 259:271–75
147. West-Roberts JA, Matheus-Carnevali PB, Schoelmerich MC, Al-Shayeb B, Thomas AD, et al. 2021. The *Chloroflexi* supergroup is metabolically diverse and representatives have novel genes for non-photosynthesis based CO_2 fixation. bioRxiv 2021.08.23.457424. **http://doi.org/10.1101/2021.08.23.457424**
148. Schulz L, Guo Z, Zarzycki J, Steinchen W, Schuller JM, et al. 2022. Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* 378:155–60
149. Varaljay VA, Satagopan S, North JA, Witte B, Dourado MN, et al. 2016. Functional metagenomic selection of ribulose 1,5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environ. Microbiol.* 18:1187–99
150. Böhnke S, Perner M. 2015. A function-based screen for seeking RubisCO active clones from metagenomes: novel enzymes influencing RubisCO activity. *ISME J.* 9:735–45
151. Witte B, John D, Wawrik B, Paul JH, Dayan D, Tabita FR. 2010. Functional prokaryotic RubisCO from an oceanic metagenomic library. *Appl. Environ. Microbiol.* 76:2997–3003
152. Spreitzer RJ, Peddi SR, Satagopan S. 2005. Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *PNAS* 102:17225–30
153. Higgins SA, Savage DF. 2018. Protein science by DNA sequencing: how advances in molecular biology are accelerating biochemistry. *Biochemistry* 57:38–46
154. Lu H, Diaz DJ, Czarnecki NJ, Zhu C, Kim W, et al. 2022. Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* 604:662–67
155. Bryant DH, Bashir A, Sinai S, Jain NK, Ogden PJ, et al. 2021. Deep diversification of an AAV capsid protein by machine learning. *Nat. Biotechnol.* 39:691–96
156. Iqbal WA, Lisitsa A, Kapralov MV. 2022. Predicting plant Rubisco kinetics from RbcL sequence data using machine learning. *J. Exp. Bot.* 74:638–50
157. Paoli GC, Vichivanives P, Tabita FR. 1998. Physiological control and regulation of the *Rhodobacter capsulatus cbb* operons. *J. Bacteriol.* 180:4258–69
158. Satagopan S, North JA, Arbing MA, Varaljay VA, Haines SN, et al. 2019. Structural perturbations of *Rhodopseudomonas palustris* form II RuBisCO mutant enzymes that affect CO_2 fixation. *Biochemistry* 58:3880–92
159. Yoshida S, Atomi H, Imanaka T. 2007. Engineering of a type III rubisco from a hyperthermophilic archaeon in order to enhance catalytic performance in mesophilic host cells. *Appl. Environ. Microbiol.* 73:6254–61
160. Satagopan S, Tabita FR. 2016. RubisCO selection using the vigorously aerobic and metabolically versatile bacterium *Ralstonia eutropha*. *FEBS J.* 283:2869–80
161. Parikh MR, Greene DN, Woods KK, Matsumura I. 2006. Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E. coli*. *Protein Eng. Des. Sel.* 19:113–19
162. Wilson RH, Martin-Avila E, Conlan C, Whitney SM. 2018. An improved *Escherichia coli* screen for Rubisco identifies a protein–protein interface that can enhance CO_2-fixation kinetics. *J. Biol. Chem.* 293:18–27

163. Zhou Y, Whitney S. 2019. Directed evolution of an improved rubisco; in vitro analyses to decipher fact from fiction. *Int. J. Mol. Sci.* 20:5019
164. Mueller-Cajar O, Morell M, Whitney SM. 2007. Directed evolution of rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochemistry* 46:14067–74
165. Aslan S, Noor E, Benito Vaquerizo S, Lindner SN, Bar-Even A. 2020. Design and engineering of *E. coli* metabolic sensor strains with a wide sensitivity range for glycerate. *Metab. Eng.* 57:96–109
166. Svab Z, Maliga P. 1993. High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *PNAS* 90:913–17
167. Kanevski I, Maliga P. 1994. Relocation of the plastid rbcL gene to the nucleus yields functional ribulose-1,5-bisphosphate carboxylase in tobacco chloroplasts. *PNAS* 91:1969–73
168. Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, Hanson MR. 2014. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* 513:547–50
169. Occhialini A, Lin MT, Andralojc PJ, Hanson MR, Parry MAJ. 2016. Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO_2. *Plant J.* 85:148–160
170. Martin-Avila E, Lim Y-L, Birch R, Dirk LMA, Buck S, et al. 2020. Modifying plant photosynthesis and growth via simultaneous chloroplast transformation of rubisco large and small subunits. *Plant Cell* 32:2898–916
171. Whitney SM, Andrews TJ. 2003. Photosynthesis and growth of tobacco with a substituted bacterial Rubisco mirror the properties of the introduced enzyme. *Plant Physiol.* 133:287–94
172. Yoon D-K, Ishiyama K, Suganami M, Tazoe Y, Watanabe M, et al. 2020. Transgenic rice overproducing Rubisco exhibits increased yields with improved nitrogen-use efficiency in an experimental paddy field. *Nat. Food* 1:134–39
173. Salesse-Smith CE, Sharwood RE, Busch FA, Kromdijk J, Bardal V, Stern DB. 2018. Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. *Nat. Plants* 4:802–10
174. Lin MT, Hanson MR. 2018. Red algal Rubisco fails to accumulate in transplastomic tobacco expressing *Griffithsia monilis RbcL* and *RbcS* genes. *Plant Direct.* 2:e00045
175. Gunn LH, Martin Avila E, Birch R, Whitney SM. 2020. The dependency of red Rubisco on its cognate activase for enhancing plant photosynthesis and growth. *PNAS* 117:25890–96
176. Whitney SM, Baldet P, Hudson GS, Andrews TJ. 2001. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* 26:535–47
177. Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, et al. 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *PNAS* 112:8529–36
178. South PF, Cavanagh AP, Liu HW, Ort DR. 2019. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* 363:eaat9077
179. Driever SM, Simkin AJ, Alotaibi S, Fisk SJ, Madgwick PJ, et al. 2017. Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philos. Trans. R. Soc. B* 372:20160384
180. Long BM, Hee WY, Sharwood RE, Rae BD, Kaines S, et al. 2018. Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* 9:3570