Revisiting tradeoffs between Rubisco kinetic parameters

Avi I. Flamholz[a], Noam Prywes[b], Uri Moran[c], Dan Davidi[c], Yinon M. Bar-On[c], Luke M. Oltrogge[a], Rui Alves[d,e], David Savage[a] and Ron Milo*[,c]

[a]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720;
[b]Innovative Genomics Institute, University of California, Berkeley, CA 94704;
[c]Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel;
[d]Institute of Biomedical Research of Lleida IRBLleida, 25198, Lleida, Catalunya, Spain;
[e]Departament de Ciències Mèdiques Bàsiques, University of Lleida, 25198, Lleida, Catalunya, Spain;

*Corresponding Author: ron.milo@weizmann.ac.il.

Abstract

Rubisco is the primary carboxylase of the Calvin cycle, the most abundant enzyme in the biosphere, and one of the best-characterized enzymes. Based on correlations between Rubisco kinetic parameters, it is widely posited that constraints embedded in the catalytic mechanism enforce tradeoffs between CO_2-specificity, $S_{C/O}$, and maximum carboxylation rate, $k_{cat,C}$. However, the reasoning that established this view was based on data from ≈20 organisms. Here, we re-examine models of tradeoffs in Rubisco catalysis using a dataset from ≈300 organisms. Correlations between kinetic parameters are substantially attenuated in this larger dataset, with the inverse relationship between $k_{cat,C}$ and $S_{C/O}$ being a key example. Nonetheless, measured kinetic parameters display extremely limited variation, consistent with a view of Rubisco as a highly-constrained enzyme. More than 95% of $k_{cat,C}$ values are between 1 and 10 s^{-1} and no measured $k_{cat,C}$ exceeds 15 s^{-1}. Similarly, $S_{C/O}$ varies by only 30% among Form I Rubiscos and < 10% among C_3 plant enzymes. Limited variation in $S_{C/O}$ forces a strong positive correlation between the catalytic efficiencies (k_{cat}/K_M) for carboxylation and oxygenation, consistent with a model of Rubisco catalysis in which increasing the rate of CO_2 addition to the enzyme-substrate complex requires an equal increase to the O_2 addition rate. Altogether, these data suggest that Rubisco evolution is tightly constrained by the physicochemical limits of CO_2/O_2 discrimination.

Introduction

Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) is the primary carboxylase of the Calvin-Benson-Bassham (CBB) cycle - the carbon fixation cycle responsible for growth throughout the green lineage and many other autotrophic taxa - and the ultimate source of

nearly all carbon atoms entering the biosphere.[1] Typically, 20-30% of total soluble protein in C_3 plant leaves is Rubisco.[2] As Rubisco is so highly expressed and plants are the dominant constituents of planetary biomass,[3] it is often said that Rubisco is the most abundant enzyme on Earth.[1,4] Since Rubisco is ancient (> 2.5 billion years old), abundant, and remains central to biology, one might expect it to be exceptionally fast. But Rubisco is not fast.[5–8] Typical central metabolic enzymes have a turnover number $k_{cat} \approx 80\ s^{-1}$,[7] but > 95% of Rubisco carboxylation $k_{cat,C}$ values are between 1-10 s^{-1} and no measured $k_{cat,C}$ values exceed 15 s^{-1}.

In addition to relatively low $k_{cat,C}$ values, Rubisco reacts with O_2 in a process called oxygenation (Figure 1A). Although both carboxylation and oxygenation of the five-carbon substrate ribulose 1,5-bisphosphate (RuBP) are energetically favorable,[9] carboxylation is the productive reaction for incorporating carbon from CO_2 into precursors that generate biomass (Figure 1B). While it may play a role in sulfur, nitrogen, and energy metabolism,[10,11] oxygenation is often considered counterproductive as it occupies Rubisco active sites and yields a product, 2-phosphoglycolate (2PG), that is not part of the CBB cycle and must be recycled through metabolically-expensive photorespiration at a partial loss of carbon.[12] As such, oxygenation can substantially reduce the net rate of carboxylation by Rubisco, depending on CO_2 and O_2 concentrations and the kinetic parameters of the particular enzyme. There are at least four distinct Rubisco isoforms in nature,[13] but all isoforms catalyze carboxylation and oxygenation of RuBP through the multistep mechanism described in Figures 1A and 1C.[14,15] Even though many autotrophs depend on Rubisco carboxylation for growth, all known Rubiscos are relatively slow carboxylases and fail to exclude oxygenation.

The fastest-carboxylating Rubisco observed (at 25 ^{o}C) is from the cyanobacterium *S. elongatus* PCC 7942.[16] This enzyme has a maximum per-active site carboxylation rate ($k_{cat,C}$) of 14 s^{-1}. However, because present-day atmosphere contains abundant O_2 and relatively little CO_2 (≈21% O_2, ≈0.04% CO_2), PCC 7942 Rubisco carboxylates at a rate 20-fold below maximum in ambient conditions ($R_C \approx 0.7\ s^{-1}$ per active site, rate law in Figure 1A). Due to its relatively low CO_2-specificity, PCC 7942 Rubisco will also oxygenate RuBP appreciably in ambient conditions ($R_O \approx 0.3\ s^{-1}$), necessitating substantial photorespiratory flux to recycle 2PG. As downstream processing of 2PG by the C_2 photorespiratory pathway leads to the loss of one carbon for every two 2PG,[11,12] every two oxygenations "undoes" a carboxylation. In ambient air, therefore, the net rate of carboxylation by PCC 7942 Rubisco would be $f = R_C - R_O/2 \approx 0.6\ s^{-1}$, or ≈4% of $k_{cat,C}$. Given the kinetics of PCC 7942 Rubisco, it is not surprising that all known cyanobacteria use a CO_2-concentrating mechanism to ensure Rubisco functions in a CO_2-rich environment. Elevated CO_2 ensures that oxygenation is competitively inhibited and that carboxylation proceeds at near-maximum rate.[17] Thirty fold enrichment of CO_2 above ambient increases the carboxylation rate of PCC 7942 Rubisco to 8.9 s^{-1} and suppresses oxygenation to ≈0.14 s^{-1}, giving a net carboxylation rate $f = 8.8\ s^{-1}$ per active site (≈60% of $k_{cat,C}$).

For comparison, the Rubisco from spinach leaves (*S. oleracea*) is characteristic of plant Rubiscos in having lower $k_{cat,C} \approx 3\ s^{-1}$ and much greater CO_2-affinity than the *S. elongatus* enzyme (spinach half-maximum CO_2 concentration $K_C \approx 12$ μM, PCC 7942 $K_C \approx 170$ μM). As a

result, the spinach enzyme outperforms the cyanobacterial one in ambient air, with $R_C \approx 1.2\ s^{-1}$, $R_O \approx 0.4\ s^{-1}$, and $f \approx 1\ s^{-1}$. Spinach is a C_3 plant, meaning it does not have a CO_2-concentrating mechanism, which may explain why it employs a slow-but-specific Rubisco. Still, most central metabolic enzymes catalyze far more than 1 reaction per second,[7] leading many to wonder if Rubisco catalysis could be improved. Improved Rubisco carboxylation might increase C_3 crop yields,[18,19] but a substantially improved enzyme has evaded bioengineers for decades.[20] The repeated evolution of diverse CO_2 concentrating mechanisms, which modulate the catalytic environment rather than Rubisco itself, raises further doubts about whether Rubisco catalysis can be strictly improved.[21]

Various nomenclature has been used to describe the kinetics of Rubisco carboxylation and oxygenation since its discovery in the 1950s.[5,6,22,23] Here we use $k_{cat,C}$ and $k_{cat,O}$ to denote turnover numbers (maximum rates per active site , s^{-1} units) for carboxylation and oxygenation respectively. K_C and K_O denote the Michaelis constants (half-saturation concentrations in μM units) for carboxylation and oxygenation. The specificity factor $S_{C/O} = (k_{cat,C}/K_C) / (k_{cat,O}/K_O)$ is a unitless measure of the relative preference for CO_2 over O_2 (Figure 1D). Since $S_{C/O}$ relates only to the ratio of kinetic parameters, higher $S_{C/O}$ does not necessarily imply higher carboxylation rates. Rather, absolute carboxylation and oxygenation rates depend on CO_2 and O_2 concentrations (Figure 1B) which can vary between organisms and environments (SI).

As data on bacterial, archaeal and plant carboxylases has accumulated over the decades, many researchers have noted that fast-carboxylating Rubiscos are typically less CO_2-specific.[24–26] In other words, Rubiscos with high $k_{cat,C}$ were observed to have lower $S_{C/O}$ due either to lower CO_2-affinity (high K_C) or more efficient oxygenation (higher $k_{cat,O}/K_O$). Negative correlation between $k_{cat,C}$ and $S_{C/O}$ is often cited to motivate the idea that a tradeoff between carboxylation rate and specificity constrains Rubisco evolution.[5,6,26,27]

It is worth pausing to clarify the concepts of “tradeoff,” “constraint,” and “correlation” (Figure 2). Correlation indicates an apparent linear (or log-log, etc.) relationship between two kinetic parameters. Correlations between enzyme kinetic parameters can result from a “tradeoff” due to two distinct kinds of underlying constraints (Figure 2).[6] In the “mechanistic coupling” scenario, the enzymatic mechanism forces a strict quantitative relationship between two kinetic parameters such that varying one forces the other to vary in a defined manner. This results in a situation where the value of one parameter strictly determines the other and vice versa (i.e. an “equality constraint,” Figure 2A). This could arise for Rubisco, for example, if a single catalytic step (e.g. enolization of RuBP, Figure 1A) determines the rates of CO_2 and O_2 entry both.

In the “selection within limits” model (Figure 2B), in contrast, the catalytic mechanism imposes an upper limit on kinetic parameters, i.e. an inequality constraint. A clear correlation between parameters will emerge only if there is sufficient selection to reach the boundary. To highlight the difference between these models, consider the kinetics of ancestral enzymes. In the “mechanistic coupling” model, kinetic parameters of ancestors should lie along the same curve as present-day enzymes because the grey regions off the curve are disallowed. Selection could act by moving enzymes along the line of mechanistic coupling - e.g. from a region of high

selectivity and low rate towards higher rate and lower selectivity (Figure 2A). According to the "selection within limits" model, in contrast, ancestral enzymes can lie beneath the upper limit determined by the catalytic mechanism (Figure 2B). This second model requires selection to produce a situation where the kinetics of present-day Rubiscos extracted from various organisms trace out a curve determined by the upper limit enforced by the mechanism.[28]

Previous research advanced two distinct families of mechanistic models to explain correlations between Rubisco kinetic parameters.[5,6] The first model, which we term "$k_{cat,C}$-K_C coupling," hypothesizes a tradeoff between the rate and affinity of carboxylation that leads to a negative correlation between $k_{cat,C}$ and $S_{C/O}$ (Figure S2).[5] A second model, which was advanced in a study which the last author of this work participated in,[6] hypothesizes that multiple tradeoffs constrain Rubisco such that kinetic parameters can only vary along a one dimensional curve. In addition to $k_{cat,C}$-K_C coupling, this work hypothesized a tradeoff between catalytic efficiencies for carboxylation and oxygenation (coupling $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$) wherein improving carboxylation efficiency also improves oxygenation efficiency.

These mechanistic models are substantively different. Though both models imply limitations on the concurrent improvement of $k_{cat,C}$ and $S_{C/O}$, "$k_{cat,C}$-K_C coupling" relates only to carboxylation kinetics, leaving the possibility that oxygenation kinetics are unconstrained. Coupling between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$, in contrast, relates to both reaction pathways. While these models appeal to physical and chemical intuition, they are based on data from only ≈20 organisms. Moreover, "mechanistic coupling" and "selection within limits" could plausibly underly either model (Figure 2).[6]

Here we take advantage of the accumulation of new data to revisit correlations and tradeoffs between Rubisco kinetic parameters. We collected and curated literature measurements of ≈300 Rubiscos. Though diverse organisms are represented, the Form I Rubiscos of C_3 plants make up the bulk of the data (>80%, Figure 3A). Most previously-reported correlations between Rubisco kinetic parameters are substantially attenuated in this dataset, with the negative correlation between $k_{cat,C}$ and specificity $S_{C/O}$ being a key example. Weakened $k_{cat,C}$-$S_{C/O}$ and $k_{cat,C}$-K_C correlations imply that these parameters are not straightforwardly mechanistically coupled, suggesting that models of $k_{cat,C}$-K_C coupling should be revisited in future experiments. Overall, weakened correlations call into question previous claims that (i) Rubisco kinetics are constrained to evolve on a one-dimensional line and (ii) natural Rubiscos are optimized to suit environmental CO_2 and O_2 concentrations.[5,6]

Despite weakened correlations, Rubisco kinetic parameters display extremely limited variation. $k_{cat,C}$ varies by only 50% among Form I Rubiscos, and $S_{C/O}$ varies even less than that (≈30%, Figure 3C). Limited variation in $S_{C/O}$ forces a strong positive power-law correlation between the catalytic efficiencies for carboxylation ($k_{cat,C}/K_C$) and oxygenation ($k_{cat,O}/K_O$).[6] We propose a simple model of mechanistic coupling that explains how constraints on the Rubisco mechanism could restrict variation in $S_{C/O}$. In this model, variation in catalytic efficiency ($k_{cat,C}/K_C$ and $k_{cat,O}/K_O$) derives solely from gating substrate access to the active site complex, which could

help explain why Rubisco has been so recalcitrant to improvement by mutagenesis and rational engineering.

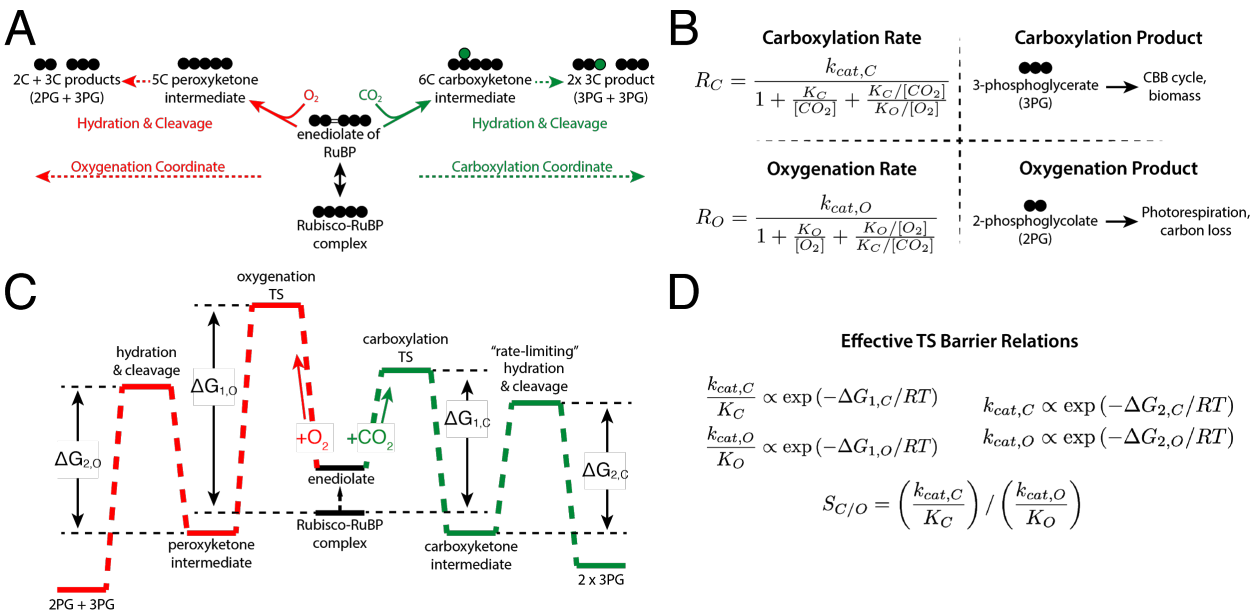

Figure 1: *Description of the catalytic mechanism of Rubisco. The "middle-out" diagram in (A) shows the ordered mechanisms of carboxylation and oxygenation. Circles represent carbon atoms. RuBP is isomerized to an enediolate before carboxylation or oxygenation. Addition of CO_2 or O_2 to the enediolate of RuBP are considered irreversible as are the subsequent hydration and cleavage steps of carboxylation and oxygenation arms. (B) Carboxylation displays effective Michaelis-Menten kinetics (maximum catalytic rate $k_{cat,C}$, half-maximum CO_2 concentration $K_M = K_C$) with competitive inhibition by O_2 (assuming half-maximum inhibitory O_2 concentration $K_i = K_O$). Carboxylation results in net addition of one carbon to the five-carbon RuBP, producing two 3PG molecules. 3PG is part of the CBB cycle and can therefore be used to continue the cycle and produce biomass. Oxygenation also displays effective Michaelis-Menten kinetics ($k_{cat,O}$, $K_M = K_O$, half-max inhibitory CO_2 concentration $K_I = K_C$). Oxygenation of RuBP produces one 3PG and one 2PG. Rates of carboxylation (R_C) and oxygenation (R_O) are calculated from kinetic parameters and the CO_2 and O_2 concentrations. The reaction coordinate diagram in (C) describes carboxylation and oxygenation as a function of two "effective" barriers.[6] The first effective barrier includes enolization and gas addition while the second includes hydration and cleavage. (D) Given standard assumptions (SI), catalytic efficiencies (k_{cat}/K_M) are related to the height of the first effective barrier while k_{cat}s are related to the second. The first barrier to oxygenation is drawn higher than for carboxylation because oxygenation is typically slower than carboxylation. Net reactions of RuBP carboxylation and oxygenation are both quite thermodynamically favorable.*[9]

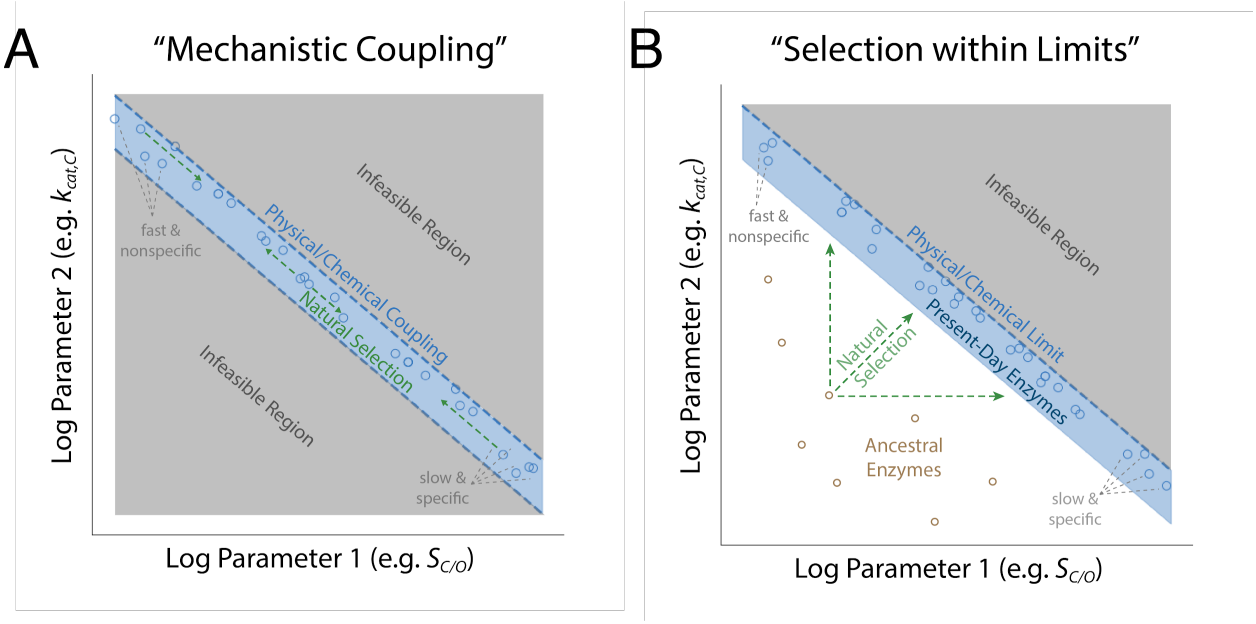

Figure 2: *Scenarios that produce strong correlations between enzyme kinetic parameters. As the log of the kinetic parameters are linearly related to energy barriers, linear energetic tradeoffs should manifest as log-log correlations between kinetic parameters (power laws). Panel (A) describes a situation in which two kinetic parameters are inextricably linked by the enzyme mechanism, diagrammed here as negative coupling between $k_{cat,C}$ and $S_{C/O}$ as an example. These couplings take the form of "equality constraints" where one parameter determines the other within measurement error. Correlation is expected so long as diverse enzymes are measured. In (A) selection moves enzymes along the blue curve, but cannot produce enzymes off the curve (grey) because they are infeasible. Panel (B) diagrams an alternative scenario in which the enzyme mechanism imposes an upper limit on two parameters (an inequality constraint). In the "selection within limits" scenario, effective selection is required for correlation to emerge because sub-optimal enzymes (e.g. ancestral sequences) are feasible. In the examples plotted, different environmental CO_2 and O_2 concentrations should select for different combinations of rate ($k_{cat,C}$) and affinity ($S_{C/O}$), resulting in present-day enzymes occupying distinct regions of the plots in (A) and (B).*

Materials and Methods

Data collection and curation. We reviewed the literature to find Rubisco kinetic data measured at 25 °C and near pH 8. Ultimately 61 primary literature studies were included, yielding 335 $S_{C/O}$, 284 $k_{cat,C}$, 316 K_C, and 254 K_o values for Rubiscos from 304 distinct organisms (Datasets S1 and S2). We also recorded 58 measurements of the Michaelis constant for RuBP (K_{RuBP}). Experimental error was recorded for all of these values (when reported) along with the pH, temperature and other metadata. Data was filtered as described in SI. $k_{cat,O}$ is usually not measured directly,[29] but is rather inferred as $k_{cat,O} = (k_{cat,C}/K_C) / (S_{C/O}/K_O)$. We assumed that experimental error is normally distributed and used 10^4-fold bootstrapping to estimate 199 $k_{cat,O}$ values and 95% confidence intervals thereof. We used an identical procedure to estimate $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ and confidence intervals thereof (SI). Altogether, we were able to calculate 274 $k_{cat,C}/K_C$ and 199 $k_{cat,O}/K_O$ values. Datasets S1 and S2 provide all source and inferred data respectively.

Fitting power laws. Certain model Rubiscos are measured frequently. For example, we found 12 independent measurements of the spinach Rubisco. In such cases, the median measured value was used to avoid bias in correlation and regression analyses. In contrast to textbook examples with one independent and one dependent variable, there is experimental error associated with both variables in all scatter plots shown here (e.g. plotting $k_{cat,C}$ against K_C in Figure 5B). As such we used total least squares linear regression in log scale to fit relationships between Rubisco parameters. Because R^2 values of total least squares fits do not convey the explained fraction of Y axis variance, they are challenging to interpret. We instead report the degree of correlation as Pearson R values of log-transformed values. Bootstrapping was used to determine 95% confidence intervals for the Pearson correlation coefficient, power-law exponents and prefactors (i.e. the slopes and intercepts of linear fits in log-log scale). In each iteration of the bootstrap, data were subsampled to 90% with replacement. Total least squares regression was applied to each subsample and the procedure was repeated 10^4 times to determine 95% confidence intervals. Python source code is available at github.com/flamholz/rubisco.

Results

An extended dataset of Rubisco kinetic parameters

To augment existing data, we collected literature data on ≈300 Rubiscos including representatives of clades and physiologies that had been poorly represented in earlier datasets e.g. diatoms, ferns, CAM plants and anaerobic bacteria (Figure 3A). We collected kinetic parameters associated with carboxylation and oxygenation - S, K_C, $k_{cat,C}$, K_O and $k_{cat,O}$ - as well as measurements of the RuBP Michaelis constant (half-maximum RuBP concentration, K_{RuBP}) and experimental uncertainty for all values where available. All data considered below were measured at 25 ^oC and near pH 8 to ensure that measured values are comparable (SI). Notably, Rubisco assays are challenging to perform and variation in measurements across labs is expected. Some of the spread in the data may come from systematic differences between labs and assay methods. Rubisco activation state, for example, may differ between methods and preparations.[15] Though we cannot resolve this issue here, we were careful to review each study's methods, document a small number of problematic measurements, and record experimental error when reported (Dataset S1).

The resulting dataset contains Rubisco measurements from a total of 304 distinct species, including 335 $S_{C/O}$ values, 284 $k_{cat,C}$ values, 316 K_C values, 254 K_O values, and 199 $k_{cat,O}$ values (Figure 3B). $k_{cat,O}$ values are rarely measured directly (SI) and are typically inferred as $k_{cat,O} = (k_{cat,C}/K_C) / (S_{C/O}/K_O)$.[29] The Michaelis constant for RuBP (K_{RuBP}) is measured infrequently and only 58 values were extracted. We were able to estimate catalytic efficiencies for carboxylation ($k_{cat,C}/K_C$) in 274 cases and for oxygenation ($k_{cat,O}/K_O$) in 199 cases (Methods). Though the data include measurements of some Form II, III and II/III Rubiscos, they remain highly focused on the Form I Rubiscos found in cyanobacteria, diatoms, algae and higher plants, which make up > 95% of the dataset (Figure 3B). As such, we focus here on the kinetic parameters of Form I Rubiscos, (abbreviated FI Rubisco).

Rubisco kinetic parameters display very narrow dynamic range (Figure 3C). The geometric standard deviation (σ^*) expresses multiplicative variability in the dataset and is well-below one order-of-magnitude ($\sigma^* \ll 10$) for all parameters. Rubisco displays particularly low variation in $k_{cat,C}$ ($\sigma^* = 1.5$) as compared to other enzymes for which 20 or more k_{cat} measurements are available (median $\sigma^* \approx 7$, Figure S5). Specificity $S_{C/O}$ displays the least variation of all parameters ($\sigma^* = 1.3$). This is due in part to overrepresentation of C_3 plants in the dataset, which occupy a narrow range of $S_{C/O} \approx 80\text{-}120$. Nonetheless, measurements of $S_{C/O}$ for FI and FII enzymes are clearly distinct, with values ranging from and 7-15 for FII measurements and ≈50-200 for FI (Figure 3C).

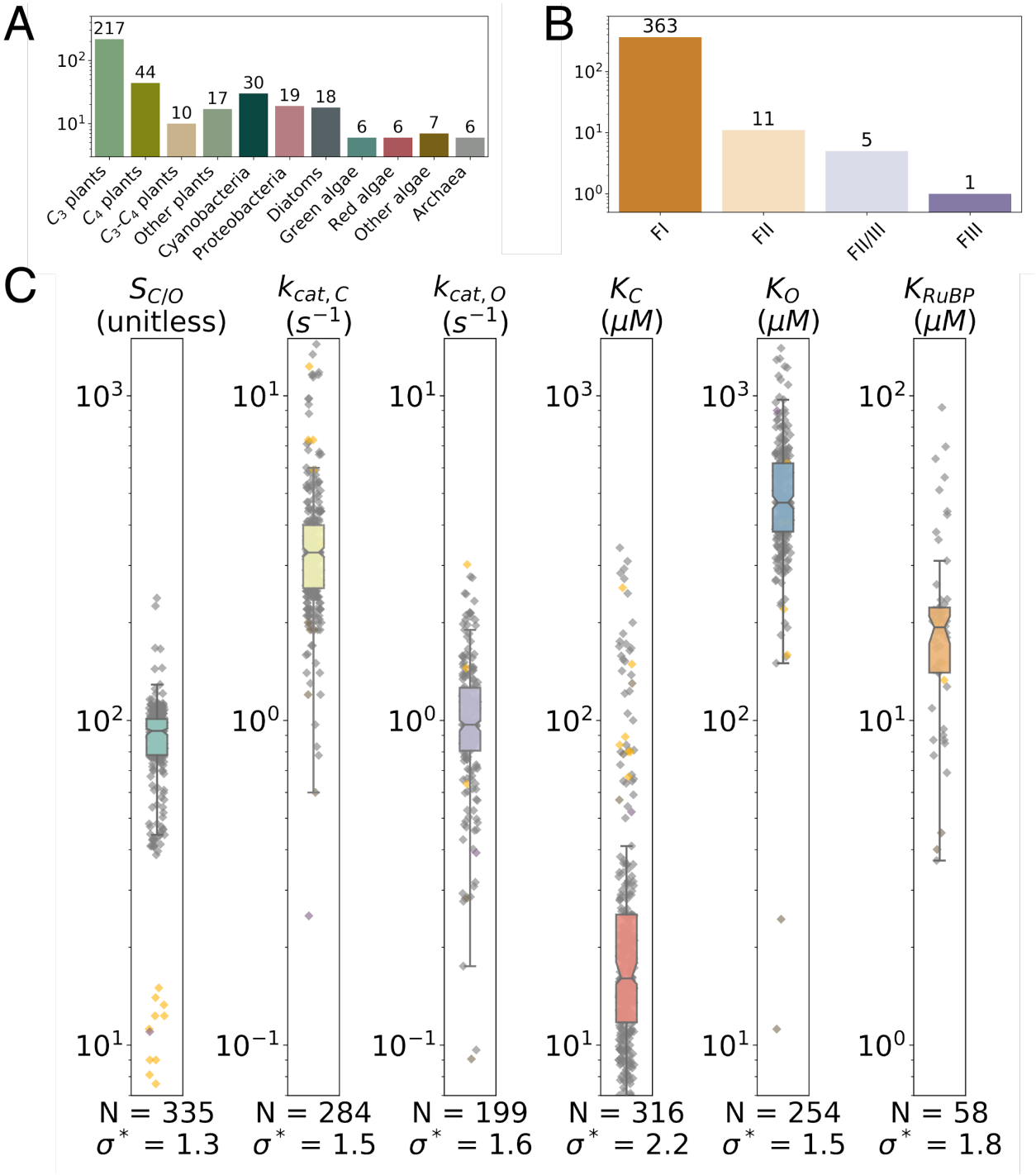

Figure 3: *Summary of the full extended dataset. We collected measurements of Rubisco kinetic parameters from a variety of organisms (A) representing four classes of Rubisco isoforms (B). Form I enzymes from plants, cyanobacteria and algae make up the bulk of the data (A-B). (C) Rubisco kinetic parameters display narrow dynamic range. The box-plot and grey points describe the distribution of Form I Rubiscos while Form II Rubiscos are in yellow. Colored boxes give the range of the central 50% of FI values and the notch indicates the median. N is the number values and σ^* gives the geometric standard deviation of Form I data. $\sigma^* < 3$ for all parameters, meaning a single standard deviation varies over less*

than threefold. All data are from wild-type Rubiscos measured at 25 °C near pH 8. More detailed histograms are given in Figure S4.

Energetic tradeoffs tend to produce power law correlations

All kinetic parameters ($S_{C/O}$, $k_{cat,C}$, K_C, $k_{cat,O}$ and K_O) are mathematically related to the microscopic rate constants of the Rubisco mechanism (SI). Given common assumptions about irreversible and rate limiting steps (SI), this multi-step mechanism can be simplified so that logarithms of measured kinetic parameters are proportional to effective transition state barriers (Figure 1C-D, SI). As such, correlations between kinetic parameters will emerge if effective transition state barriers vary together (Figure 2). If, for example, lowering the effective transition state barrier to CO_2 addition ($\Delta G_{1,C}$) requires an increase to the effective barrier to the subsequent hydration and cleavage steps of carboxylation ($\Delta G_{2,C}$), then we should observe a negative linear correlation $\Delta G_{1,C} \propto -\Delta G_{2,C}$. Since $k_{cat,C}/K_C$ is related to the first effective carboxylation barrier ($k_{cat,C}/K_C \propto exp\,(-\Delta G_{1,C}/RT)$) and $k_{cat,C}$ to the second ($k_{cat,C} \propto exp\,(-\Delta G_{2,C}/RT)$), linear correlation between transition state barrier heights translates to log-scale correlation between kinetic parameters such that $ln\,(k_{cat,C}/K_C) \propto -\,ln\,(k_{cat,C})$. These relationships are known as power laws and motivate us and others to investigate the kinetic parameters on a log-log scale.

We expect to observe strong power-law correlations between pairs of kinetic parameters in two cases. Either (i) the associated energy barriers co-vary because they are linked by the enzymatic mechanism ("mechanistic coupling," Figure 2A) or (ii) the mechanism imposes an upper bound on the sum (or difference) of two barrier heights. In case (ii), strong selection favors the emergence of enzymes at-or-near the imposed limit ("selection within limits," Figure 2B). As Rubisco is the central enzyme of photoautotrophic growth, it likely evolved under selection pressure towards maximizing the net rate of carboxylation in each host, and so either of these scenarios is plausible *a priori*. Notably, host physiology and growth environments can affect the catalytic environment: Rubiscos from different organisms will experience different temperature, pH and prevailing CO_2 and O_2 concentrations due, for example, to an anaerobic host or a CO_2 concentrating mechanism enriching CO_2.[6] Different conditions should favor different combinations of kinetic parameters (Figure 2).

Correlations between kinetic parameters of Form I Rubiscos

We performed a correlation analysis to investigate relationships between kinetic parameters of Form I Rubiscos. Figure 4 gives log-scale Pearson correlations between parameters that are measured directly: $k_{cat,C}$, K_C, K_O, $S_{C/O}$ and K_{RuBP}. Linear scale correlations are reported in Figure S7.

Overall, correlations are weaker in the extended dataset than documented in previous studies of smaller datasets.[5,6] Nonetheless, we observed modestly strong, statistically significant correlations between $k_{cat,C}$ and $S_{C/O}$ ($R = -0.56$, $p < 10^{-10}$), $k_{cat,C}$ and K_C ($R = 0.48$, $p < 10^{-10}$), K_C and $S_{C/O}$ ($R = -0.66$, $p < 10^{-10}$), and K_C and K_O ($R = 0.56$, $p < 10^{-10}$). Since Rubisco kinetic parameters are mathematically interrelated through the microscopic mechanism as it is

commonly understood, some level of correlation is expected. For example, when we derive expressions for $k_{cat,C}$ and K_C from the Rubisco mechanism, they share common factors that should produce some correlation even in the absence of underlying coupling (SI). Similarly, $S_{C/O}$ is defined as $(k_{cat,C}/K_C)$ / $(k_{cat,O}/K_O)$ and could correlate negatively with K_C for this reason. Because modest correlation is expected irrespective of underlying tradeoffs, the correlations in Figure 4 do not necessarily support any particular tradeoff model.

Correlations between $k_{cat,C}$ and $S_{C/O}$ as well as $k_{cat,C}$ and K_C were previously highlighted to support particular mechanistic tradeoff models.[5,6] However, these correlations are substantially attenuated by the addition of new data (Figures 4 and 5). Plotting $k_{cat,C}$ against $S_{C/O}$ (Figure 5A) shows that these parameters are modestly correlated, with $R \approx 0.6$ as compared to $R \approx 0.9$ in previous analyses.[5,6] Figure 5A also highlights the extremely limited and stereotyped variation in $S_{C/O}$ where Rubiscos from organisms sharing a particular physiology (e.g. C_3 plants or cyanobacteria) occupy a very narrow range of $S_{C/O}$ values. Multiplicative standard deviations (σ^*) are less than 1.25 in all cases. Plotting $k_{cat,C}$ against K_C (Figure 5B) shows that this correlation is also weakened, with $R \approx 0.5$ as compared to ≈ 0.9 previously.[6] We interpret weakened correlations as evidence that previously-proposed tradeoff models should be revisited. We therefore proceed to evaluate the correlations predicted by specific tradeoff models, with an eye towards understanding the restricted variation in $S_{C/O}$ shown in Figure 5A.

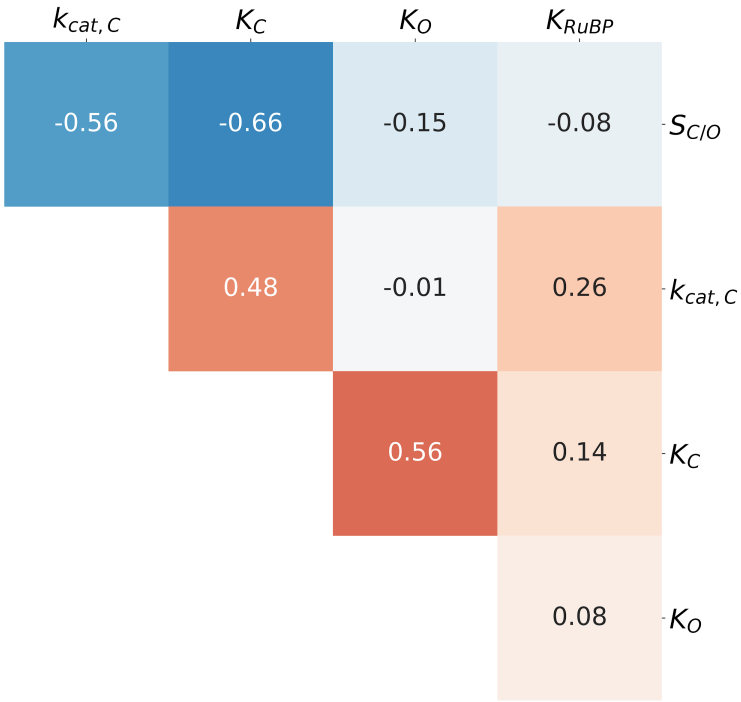

Figure 4: *Correlations between measured kinetic parameters are attenuated by the addition of new data. The figure gives Pearson correlations (R) between pairs of log-transformed kinetic parameters of Form I Rubiscos. When multiple measurements of the same enzyme were available, the median value was used (Methods).* $S_{C/O}$*-*K_C*,* $S_{C/O}$*-*$k_{cat,C}$*, and* K_C*-*$k_{cat,C}$ *correlations are of particular interest because they were highlighted in previous works, which found R = 0.8-0.95. None of these pairs have R values exceeding 0.7 in the extended dataset.*

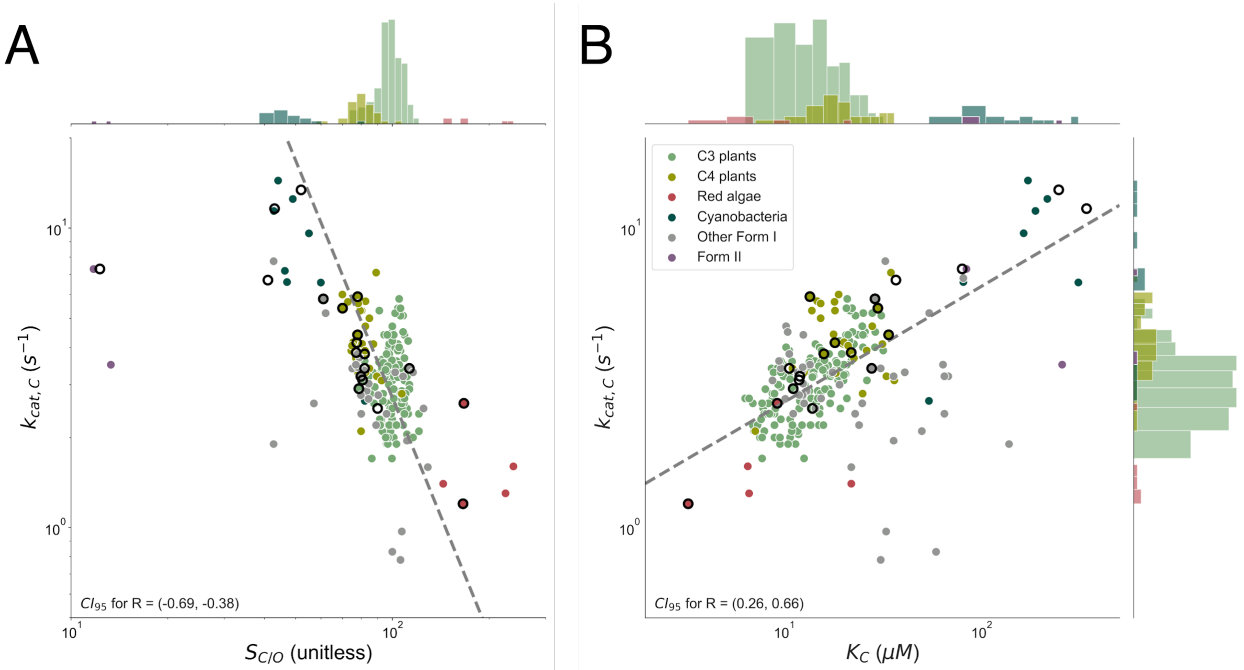

Figure 5: *Focal correlations of previous analyses are not robust to new data. Points with black outlines are from Savir et al. 2010 and dashed grey lines represent the best fit to FI Rubisco data. Histograms for $k_{cat,C}$, $S_{C/O}$ and K_C are plotted on parallel axes. Panel (A) plots $k_{cat,C}$ against $S_{C/O}$. $k_{cat,C}$ and $S_{C/O}$ correlate with R ≈ -0.6 among FI Rubiscos as compared to ≈ 0.9 previously.[5,6] 95% confidence intervals are (-4.0, -2.0) for the fit exponent and ($3x10^4$, $2x10^8$) for the prefactor (slope and intercept in log-log scale respectively) indicating that the form of $k_{cat,C}$-$S_{C/O}$ correlation is very uncertain. Notably, $S_{C/O}$ displays very limited variation overall and especially within physiological groupings with sufficient data. Median $S_{C/O}$ values are 177 for red algae (σ^* = 1.2, N=6), 98 for C_3 plants (σ^* = 1.1, N=162), 80 for C_4 plants (σ^* = 1.1, N=35), and 48 for cyanobacteria (σ^* = 1.1, N=16). Panel (B) plots $k_{cat,C}$ against K_C. Here, R ≈ 0.5 as compared to ≈ 0.9 previously.[6] This fit is more robust, with 95% confidence intervals of (0.3, 0.5) and (0.8, 1.5) for the fit exponent and prefactor respectively.*

Re-evaluation of Proposed Tradeoff Models

Two distinct mechanistic tradeoff models have been advanced.[5,6] The first model, which we term $k_{cat,C}$-K_C coupling, posits that increased specificity towards CO_2 necessitates a slower maximum carboxylation rate, $k_{cat,C}$.[5,6] It was proposed that this tradeoff is due to stabilization of the first carboxylation transition state (TS).[5] Under this model, a stable Rubisco-TS complex produces high CO_2-specificity but slows the subsequent carboxylation steps and limits $k_{cat,C}$ (Figure S2). This proposal can be cast in energetic terms by relating the measured catalytic parameters to effective transition state barrier heights (Figure 1D, SI). This model can be construed in energetic terms as follows: lowering the effective barrier to CO_2 addition ($\Delta G_{1,C}$ in Figure 6A) will make Rubisco more CO_2-specific even if oxygenation kinetics remain unchanged.[6] $k_{cat,C}$-K_C coupling posits a negative coupling between CO_2 addition and the subsequent carboxylation steps of hydration and bond cleavage (effective barrier height $\Delta G_{2,C}$ diagrammed in Figure 6A). Therefore, the energetic interpretation of this model predicts a

negative correlation between $\Delta G_{1,C}$ and $\Delta G_{2,C}$ and, as a result, a negative power-law correlation between $k_{cat,C}$ and $k_{cat,C}/K_C$.[6]

In previous work, $k_{cat,C}$ and $k_{cat,C}/K_C$ were found to correlate strongly on a log-log scale.[6] The reported correlation, however, is not strongly supported by our dataset (Figure 6B). The true barrier height to CO_2 addition depends on the CO_2 concentration, which could partially explain the apparent lack of correlation. However, correlation is not improved by restricting focus to C_3 plants for which data is abundant and for which measured leaf CO_2 concentrations vary by only 20-30% due to variation in CO_2 conductance and Rubisco activity.[30,31]

Absence of correlation does not necessarily imply the absence of an underlying mechanistic limitation. Rather, if the Rubisco mechanism limits the joint improvement of $k_{cat,C}$ and $k_{cat,C}/K_C$, much decreased correlation over the extended dataset ($R < 0.4$) could result from several factors including measurement error, undersampling of Rubiscos with high $k_{cat,C}$ (e.g. from cyanobacteria) or, alternatively, insufficient selection pressure. Diminished correlation, with many points observed below the previous correlation line, suggests that the "mechanistic coupling" model is less likely than "selection within limits" in this case (Figure 2).

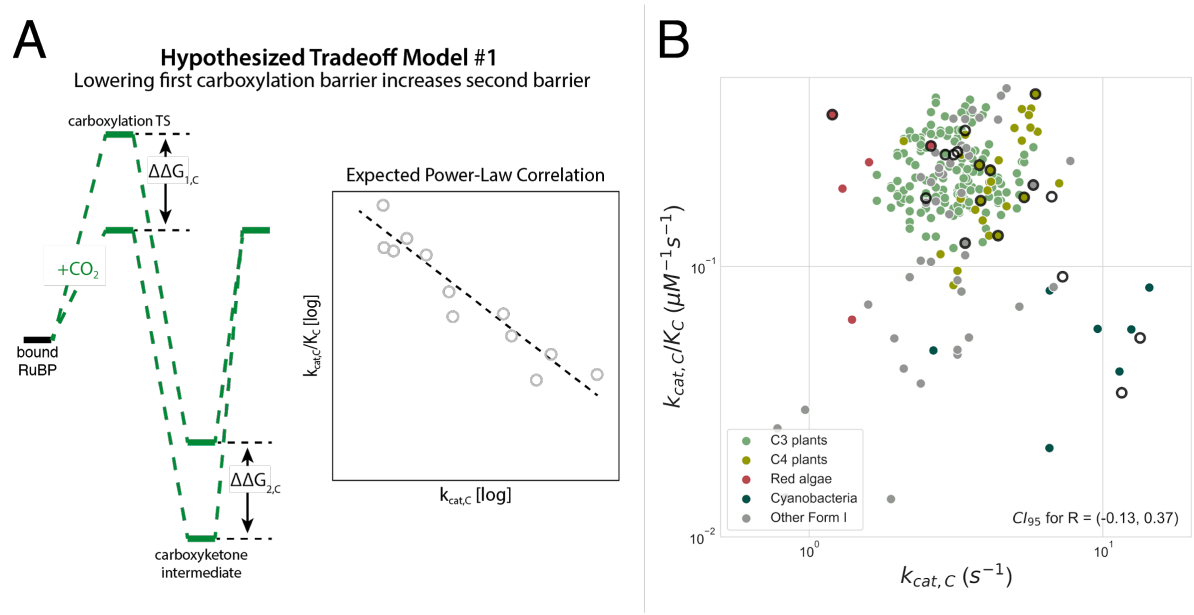

***Figure 6**: Negative power-law correlation between $k_{cat,C}$ and $k_{cat,C}/K_C$ is not supported by the extended dataset. In the model diagrammed in panel (A), CO_2-specific Rubiscos have low barriers to enolization and CO_2 addition (first effective carboxylation barrier $\Delta G_{1,C}$), but lowering the first effective barrier necessarily increases the second effective barrier ($\Delta G_{2,C}$), reducing $k_{cat,C}$. In this view, stabilizing the first carboxylation TS also enhances selectivity but also slows carboxylation (Figure S2). $\Delta G_{1,C}$ and $\Delta G_{2,C}$ should be negatively correlated, which would manifest as negative power-law correlation between $k_{cat,C}$ and $k_{cat,C}/K_C$ under certain assumptions (SI). (B) The extended dataset does not evidence the expected correlation ($R = 0.02$, $p = 0.8$ for Form I enzymes). While previous analyses gave $R \approx -0.9$,[6] the 95% confidence interval for R now includes 0.0. Restricting focus to particular physiologies like C_3 plants does not recover the expected correlation.*

The second mechanistic model, wherein faster CO_2 addition entails faster O_2 addition,[6] is well-supported by the extended dataset (Figure 7). This model was previously supported by a power-law correlation between catalytic efficiencies for carboxylation and oxygenation ($k_{cat,C}/K_C \sim (k_{cat,O}/K_O)^2$). As $k_{cat,C}/K_C$ is exponentially related to the first effective carboxylation barrier ($k_{cat,C}/K_C \sim \exp(-\Delta G_{1,C})$) and $k_{cat,O}/K_O$ to the first effective oxygenation barrier ($k_{cat,O}/K_O \sim \exp(-\Delta G_{1,O})$), correlation was taken to imply that decreasing the barrier to CO_2 addition also decreases the barrier to O_2 addition (Figure 7A). Our dataset supports a similar power-law, albeit with an exponent of ≈1.0 instead of ≈2.0.

Again, we found that $S_{C/O}$ varies little among Form I Rubiscos (Figure 3C) and even less within C_3 plants, cyanobacteria and other physiological groupings (Figures 5A and 7B). $S_{C/O} = (k_{cat,C}/K_C) / (k_{cat,O}/K_O)$ by definition, so the fact that $S_{C/O}$ is approximately constant forces a positive power-law relationship of $\log(k_{cat,C}/K_C) = \log(k_{cat,O}/K_O) + \log(S_{C/O})$. Indeed, Form I enzymes display a remarkably high-confidence power-law relationship between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ ($R = 0.94$, $p < 10^{-10}$). Since $S_{C/O}$ is the only free parameter in this equation and it is nearly-constant, the geometric mean of $S_{C/O}$ measurements (≈ 90 for Form I Rubiscos) can be used to predict $k_{cat,O}/K_O$ as $(S_{C/O})^{-1} \times (k_{cat,C}/K_C)$. This simple approach, which uses a power-law exponent of 1.0 and prefactor of $(S_{C/O})^{-1}$, predicts Form I $k_{cat,O}/K_O$ values with $R^2 = 0.80$, nearly as accurate as fitting both the prefactor and exponent as free parameters ($R^2 = 0.81$). As shown in Figure 7C, predictions of $k_{cat,O}/K_O = (S_{C/O})^{-1} \times (k_{cat,C}/K_C)$ generally improve when considering specific physiological groupings like C_3 and C_4 plants because $S_{C/O}$ varies so little within these groups. Assuming roughly constant $S_{C/O}$ forces a 1:1 relationship $\Delta G_{1,C} = \Delta G_{1,O} + C$, meaning that decreasing the CO_2 addition barrier $\Delta G_{1,C}$ is associated with an *equal* decrease to the O_2 addition barrier $\Delta G_{1,O}$.

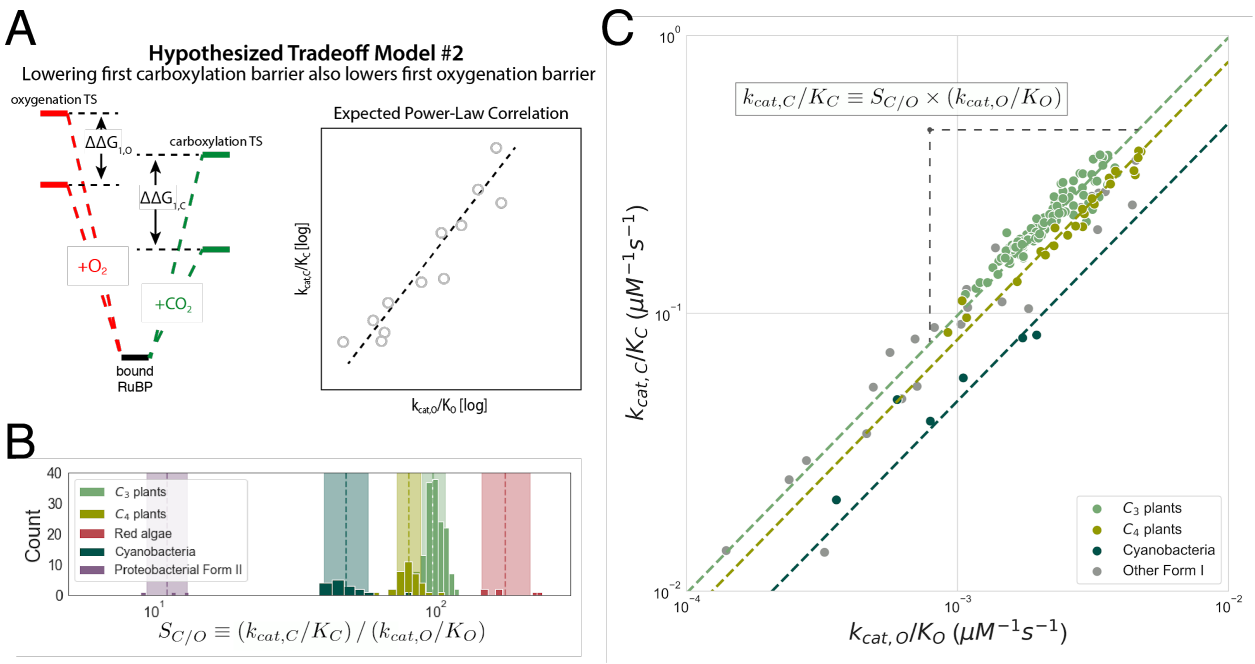

Figure 7: *The second mechanistic proposal is remarkably well-supported by the extended dataset. (A) In this proposal, mutations increasing the rate of CO_2 addition to the Rubisco-RuBP complex also increase the rate of O_2 addition. In energetic terms, lowering the effective barrier to enolization and CO_2 addition ($\Delta G_{1,C}$) lowers the first effective barrier to O_2 addition ($\Delta G_{1,O}$) as well. Given this model, barrier heights should be positively correlated, which would manifest as a positive linear correlation on a log-log plot of $k_{cat,C}/K_C$ against $k_{cat,O}/K_O$. (B) $S_{C/O}$ displays limited variation within physiological groups such as C_3 and C_4 plants for which we have substantial data. Dashed lines give the geometric mean of $S_{C/O}$ values. The multiplicative standard deviation, σ^*, sets the width of the shaded region. (C) $S_{C/O} = (k_{cat,C}/K_C) / (k_{cat,O}/K_O)$, so restricted $S_{C/O}$ variation implies a power-law relationship $(k_{cat,C}/K_C) = S_{C/O} \times (k_{cat,O}/K_O)$. $k_{cat,C}/K_C$ is strongly correlated with $k_{cat,O}/K_O$ on a log-log scale ($R = 0.94$, $p < 10^{-10}$). Fitting FI measurements gives $k_{cat,C}/K_C = 119\ (k_{cat,O}/K_O)^{1.04}$. A 95% confidence interval exponent is (0.94, 1.13), which includes 1.0. The geometric mean of measured $S_{C/O}$ values predicts $k_{cat,O}/K_O = (k_{cat,C}/K_C)/S_{C/O}$ and vice versa. This simple approach accurately predicts of $k_{cat,O}/K_O$ for FI Rubiscos (prediction $R^2 = 0.80$), C_3 plants ($R^2 = 0.84$), C_4 plants ($R^2 = 0.96$) and cyanobacteria ($R^2 = 0.79$). Other groups, e.g. red algae, are omitted because of insufficient data.*

Implications for the mechanism of CO_2/O_2 discrimination by Rubisco

A 1:1 relationship between effective barriers to CO_2 and O_2 addition suggests that a single factor controls both barriers. We offer a simple model based on the mechanism of Rubisco that can produce a 1:1 correlation between barrier heights and constant $S_{C/O}$. In this model, the RuBP-bound active site fluctuates between reactive and unreactive states (Figures 8A). The fraction of enzyme in the reactive state is denoted ϕ. In the unreactive state neither oxygenation or carboxylation can proceed. In the reactive state, either gas can react at its intrinsic rate, which does not vary across Rubiscos of the same class (Figure 8B). Since RuBP must undergo enolization in order for carboxylation or oxygenation to occur, ϕ may be determined by the degree of enolization of RuBP (SI).

This model can be phrased quantitatively as $\frac{k_{cat,C}}{K_C} \propto \phi\ exp\ (-\Delta G^*{}_{1,C}/RT)$ and $\frac{k_{cat,O}}{K_O} \propto \phi\ exp\ (-\Delta G^*{}_{1,O}/RT)$ where $\Delta G^*{}_{1,C}$ and $\Delta G^*{}_{1,O}$ are the intrinsic reactivities of the RuBP enediolate to CO_2 and O_2 respectively. Under this model, $S_{C/O}$ should be roughly constant, which forces a power-law relationship between $\frac{k_{cat,C}}{K_C}$ and $\frac{k_{cat,O}}{K_O}$ with an exponent 1.0 (Figure 7C, SI). Variation in $\frac{k_{cat,C}}{K_C}$ and $\frac{k_{cat,O}}{K_O}$ implies that ϕ can vary between related Rubiscos, perhaps by evolutionary tuning of the equilibrium constant for RuBP enolization. $S_{C/O}$ is independent of the equilibrium fraction of on-enzyme RuBP enolization, so variation in enolization should affect $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ without altering $S_{C/O}$ (SI). Rather, $S_{C/O}$ is determined by the difference $\Delta G^*{}_{1,O} - \Delta G^*{}_{1,C}$ and so changes to the conformation of the RuBP enediolate might explain characteristic differences between $S_{C/O}$ of C_3 plant and cyanobacterial Rubiscos.[5,8] See SI for a derivation of this model and further discussion of its implications.

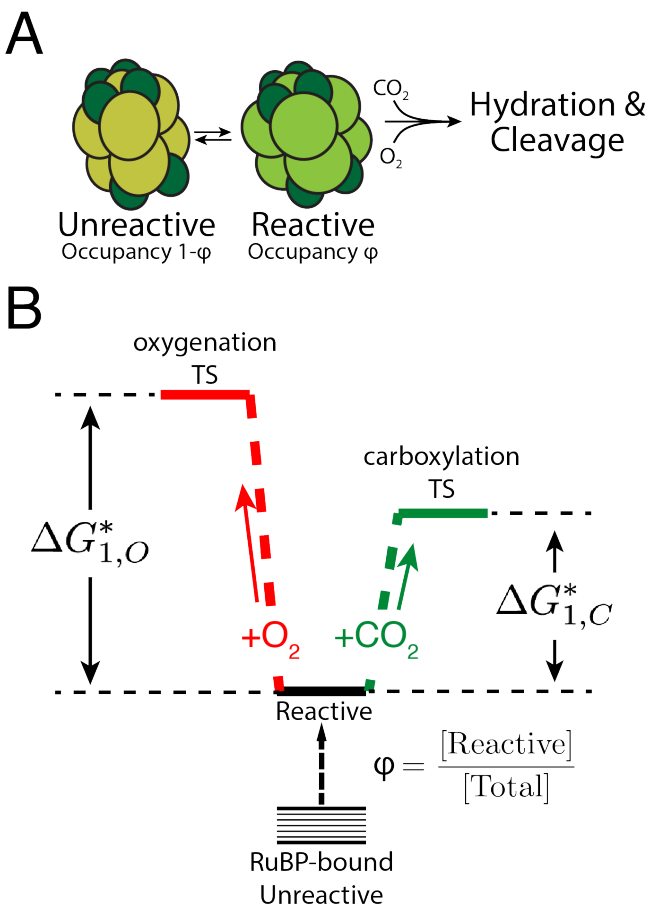

Figure 8: *A power-law relationship between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ can be explained by an active site that fluctuates between "reactive" and "unreactive" states. (A) In this model, CO_2 and O_2 react with bound RuBP only when enzyme is in the reactive state, which has occupancy φ. (B) φ can vary between related enzymes. In the reactive state, CO_2 and O_2 react with the bound RuBP with intrinsic reactivities $\Delta G^*_{1,C}$ and $\Delta G^*_{1,O}$ that do not vary between related Rubiscos. If the difference in intrinsic reactivities ($\Delta G^*_{1,O} - \Delta G^*_{1,C}$) is constant, we derive a power-law relationship between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ with an exponent of 1.0. This relationship requires constant $S_{C/O}$ (SI).*

Discussion

We collected and analyzed literature measurements of ≈300 Rubiscos (Figure 3A). The literature is very phylogenetically-biased, with the readily-purified plant Rubiscos making up >80% of the data (Figure 3B). Despite incomplete coverage, some trends are clear. Rubisco kinetic parameters display extremely limited dynamic range, with multiplicative standard deviations being less than threefold in all cases (Figure 3C). $k_{cat,C}$ and $S_{C/O}$ appear particularly constrained. Rubisco displays much less k_{cat} variability than any other enzyme for which sufficient data is available (Figure S5). 97% of $k_{cat,C}$ values are between 1 and 10 s^{-1} and the highest $k_{cat,C}$ measured at 25 °C (14 s^{-1}, *S. elongatus* PCC 7942[16]) is only ≈20 times the lowest reported Form I value (0.8 s^{-1} from the diatom *Cylindrotheca* N1[32]). Altogether, these data suggest that there is some limitation on the maximum rate of carboxylation by Rubisco in the presence of O_2.

Focusing on O_2, we find that measured Rubiscos oxygenate slowly: more than half of $k_{cat,O}$ measurements are less than 1 s^{-1} and $k_{cat,C}$ is 4 times greater than $k_{cat,O}$ on average (Figures 3C

and S4A). Similarly, O_2 affinity is quite low in general: the median K_O is ≈ 470 μM, nearly double the Henry's law equilibrium of water with a 21% O_2 atmosphere (≈ 270 μM at 25 ^oC).

With a multiplicative standard deviation of 1.3, $S_{C/O}$ displays the least variation all Rubisco kinetic parameters (Figure 3C and S4A). Figures 5A and 7B highlight the stereotyped variation in $S_{C/O}$, where C_3 plant, C_4 plant, cyanobacterial and red algal enzymes display very limited variation around characteristic $S_{C/O}$ values. All groups have multiplicative standard deviations σ^* < 1.25. Nonetheless, Form I Rubiscos are about an order-of-magnitude more CO_2-specific than the few characterized Form II, III and II/III enzymes (Figure 7B, SI). This might be explained by the prevalence of FII, FIII and FII/III enzymes in bacteria and archaea that fix CO_2 in anaerobic conditions, where it is doubtful that oxygenation affects organismal fitness. We note, however, that there is substantial variation among measurements of the model FII Rubisco from *R. rubrum* (Figure S4B). This and the paucity of data on non-Form I Rubiscos (Figure 3B) indicates that more measurements are required to evaluate FII, FII/III and FIII enzymes. As such, we focused here on FI Rubiscos, for which data is abundant.

Rubisco kinetics were previously argued to vary in a one-dimensional landscape[6] and hypothesized to be "nearly perfectly optimized."[5] Overall, FI Rubiscos appear less constrained than previously supposed. Figure 4 documents an overall reduction in correlation between FI Rubisco kinetic parameters and the dataset is no longer well-approximated as one-dimensional (Figure S8). Many natural Rubiscos appear suboptimal in plots of $k_{cat,C}$ against $S_{C/O}$ since other enzymes have roughly equal $S_{C/O}$ but higher $k_{cat,C}$ (Figures 5A and S6). Weakened correlations could be due to measurement error and systematic biases, though we find this explanation unlikely because (i) measurements of Form I Rubiscos from similar organisms are broadly consistent (Figure S6); (ii) some correlations remain strong and statistically significant across the entire dataset; (iii) systematic bias towards C_3 plants would tend to increase correlations; and (iv) standardization of Rubisco assays using stoichiometric inhibitors to quantify active sites should improve data quality over time (SI). Reduced correlations therefore lead us to reject the notion that Rubisco kinetics vary in a strictly one-dimensional landscape and to revisit previous models of mechanistic tradeoff.

The mechanistic models described in Figures 6 and 7 are based on a simple chemical intuition: the intrinsic difficulty of discriminating CO_2 and O_2 requires the enzyme to differentiate between carboxylation and oxygenation transition states. The requirement of transition state discrimination is a direct consequence of two assumptions supported by experimental evidence.[22] Briefly, it is assumed that addition of either gas is irreversible and that there is no binding site for CO_2 or O_2 and, thus, no "Michaelis complex" for either gas.[5,6,22,33,34] If CO_2 bound a specific site on Rubisco before reacting, it might be possible to modulate K_C by mutation without substantially affecting the kinetics of subsequent reaction steps. In the unlikely case that gas addition is substantially reversible,[34,35] we might expect to find Rubiscos that evolved enhanced selectivity by energy-coupled kinetic proofreading. Energy coupling can enable amplification of selectivity due to differential CO_2 and O_2 off-rates.[36] The fact that no such Rubiscos have been found suggests that gas addition is irreversible or that CO_2 and O_2 off-rates are incompatible with kinetic proofreading in some other way.[6,37]

As Rubisco likely does not bind CO_2 directly, it was hypothesized that high CO_2-specificity (large $S_{C/O}$) is realized by discriminating between the first carboxylation and oxygenation transition states, i.e. between the developing carboxyketone and peroxyketone (Figures S1 and S2).[5] A late carboxylation transition state would be maximally discriminable because the developing carboxylic acid is distinguishable from the peroxyl group of the oxygenation intermediate. The extraordinarily tight binding of the carboxyketone analog CABP to plant Rubisco provides strong support for a late carboxylation transition state.[5] Since a late transition state resembles the carboxyketone intermediate, it was argued that CO_2-specific Rubiscos must tightly bind the intermediate, slowing the subsequent reaction steps and restricting $k_{cat,C}$ (Figure S2).[5]

As $k_{cat,C}/K_C$ is related to the effective barrier to enolization and CO_2 addition ($\Delta G_{1,C}$) and $k_{cat,C}$ is related to the effective barrier to hydration and cleavage ($\Delta G_{2,C}$, Figure 1D), an energetic framing of this model argues that lowering $\Delta G_{1,C}$ (increasing $k_{cat,C}/K_C$) entails increasing $\Delta G_{2,C}$ (lowering $k_{cat,C}$, Figure 6A).[6] Despite nuanced differences, we collectively term these models $k_{cat,C}$-K_C coupling due to the hypothesized coupling of carboxylation kinetics. Though these models are motivated by the need to discriminate between CO_2 and O_2, they invoke a tradeoff between carboxylation steps only. That is: specificity requires tighter binding of the carboxylation intermediate, which slows downstream processing of that same intermediate, irrespective of O_2.

Three correlations previously supported $k_{cat,C}$-K_C coupling: correlation between $k_{cat,C}$ and $S_{C/O}$, between $k_{cat,C}$ and K_C, and between $k_{cat,C}$ and $k_{cat,C}/K_C$. $k_{cat,C}$ and $S_{C/O}$ remain negatively correlated in our larger dataset, but more weakly than previously observed (Figure 5A). The same is true for $k_{cat,C}$ and K_C (Figure 5B) and for $k_{cat,C}$-$k_{cat,C}/K_C$ (Figure 6B). Rather than arguing for strong coupling of carboxylation kinetics, Figure 5 highlights the stereotyped variation in $S_{C/O}$ described above. We interpret weakened correlations as implying that carboxylation kinetics are not strictly coupled. Considering residuals of the $k_{cat,C}$-K_C fit (Figure S9) shows that outliers include recent measurements of cyanobacterial[38] and diatom[39] Rubiscos, which fall well-below the fit line. This is consistent with a “selection within limits” view of $k_{cat,C}$-K_C coupling (Figure 2B).

The second mechanistic tradeoff model posits that faster CO_2 addition to the Rubisco-RuBP complex necessarily allows faster O_2 addition. This model was previously supported by a positive power-law correlation between the catalytic efficiencies for carboxylation and oxygenation ($k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ respectively), which can be understood as a positive coupling of the effective barriers to enolization and gas addition for CO_2 and O_2 ($\Delta G_{1,C}$ and $\Delta G_{1,O}$, Figure 7A). We showed that extremely limited and stereotyped variation in $S_{C/O} = (k_{cat,C}/K_C) / (k_{cat,O}/K_O)$ necessitates a power-law correlation with exponent 1.0 between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ (Figure 7B-C). An exponent of 1.0 implies that decreasing $\Delta G_{1,C}$ (enabling faster carboxylation) requires a *roughly equal* reduction to $\Delta G_{1,O}$ (enabling faster oxygenation as well). Although several research groups have attempted to isolate improved Rubisco mutants, none of the mutants examined so far exceed wild-type enzymes on these axes (Figure S11).

A power law relation with an exponent of 1.0 can be seen as resulting from an active site that fluctuates between a reactive and unreactive state (Figure 8A). This coarse-grained model is motivated by the Rubisco mechanism in two ways. Since Rubisco likely does not bind CO_2 or O_2 directly, active site concentrations are determined by solution concentrations (e.g. in the chloroplast stroma). Rubisco could close the active site to diffusion to limit O_2 entry,[40] but this would slow carboxylation as well. Similarly, RuBP must enolize for oxygenation or carboxylation to proceed (Figure 1A) so modulating the degree of enolization would affect both reaction pathways equally.[14,15] In either case, the average occupancy of the reactive state mechanistically couples the rates of CO_2 and O_2 addition (Figure 2A) and throttles the subsequent steps of carboxylation and oxygenation equally (Figure 8).

In previous work, where Rubisco kinetics were thought to vary in a one-dimensional landscape, setting $k_{cat,C}$ determined all other kinetic parameters.[6] In this setting, it was argued that Rubisco kinetic parameters are determined by the prevailing CO_2 and O_2 concentrations since a unique choice of parameters on the one-dimensional curve maximizes the net rate of carboxylation.[6] Since the data is no longer clearly one-dimensional, we cannot argue that Rubisco is "perfectly optimized" to match prevailing concentrations. Moreover, the model presented in Figure 8 sets no upper limit on $k_{cat,C}$, suggesting that selection for increased carboxylation in the absence of O_2 could produce Rubisco mutants with superlative $k_{cat,C}$ values (i.e. $k_{cat,C} >> 15\ s^{-1}$). Such enzymes might be found in anaerobic bacteria and would be of interest in probing the limits of Rubisco catalysis.

The prospect of engineering an improved Rubisco is tantalizing, not only because it could plausibly improve crop yields,[18] but also because the task tests our understanding of enzymes on a very basic level. It is clear from the data presented here that there is some evolutionary constraint on Rubisco catalysis. Surely a superlative Rubisco would have arisen if it was mutationally accessible from existing enzymes. More detailed biochemical investigation of naturally-occuring Rubiscos will help delineate the evolutionary constraints imposed on Rubisco kinetics. Still, the Rubisco large subunit displays extremely limited sequence variation.[41] Perhaps exploring a wider swath of sequence space via protein engineering techniques[42–44] would enable strict improvements to Rubisco kinetics? We argue that biochemical and bioengineering techniques should be used in concert to probe the limits of Rubisco catalysis and propose several avenues of future research to evaluate the prospects of Rubisco engineering.

First, the kinetics of non-plant Rubiscos should be characterized more thoroughly. These should include the Form II, III and II/III enzymes of bacteria and archaea as well as FI enzymes of bacteria and diverse eukaryotic autotrophs.[13,39] Ideally these enzymes would be chosen in a manner that maximizes sequence and phylogenetic diversity[45] and characterized for their binding (e.g. of RuBP and CABP) and catalytic activity (measuring $k_{cat,C}$, K_C, $k_{cat,O}$, K_O and $S_{C/O}$) as a function of temperature and pH.[29,46,47] A facile assay for direct measurement of oxygenation would also reduce the number of assumptions made in measuring and analyzing Rubisco kinetics.[29] These data would help resolve whether Rubisco isoforms display characteristic differences in catalytic potential. It is possible, for example, that non-Form I enzymes are

subject to different constraints than FI Rubiscos and might serve as useful chassis for engineering.

It is also important to revisit the classic experiments undergirding our understanding of the Rubisco catalytic mechanism, especially those supporting the central assumptions that (i) there is no Michaelis complex for CO_2 or O_2 and (ii) that gas addition is irreversible.[22,34,35] These assumptions substantially constrain CO_2 specificity. If we were to find Rubiscos for which these assumptions are relaxed, they might serve as a basis for engineering a fast-and-selective carboxylase. On the other hand, all Rubiscos may share the same limitations. Since these limitations are likely described as couplings between transition state barriers (as in Figures 7 and 8) measurements of barriers heights for a wide variety of Rubiscos would enable more direct testing of tradeoff models. One avenue for drawing inference about barrier heights is by measuring the binding energies of intermediate and transition state analogs.[5,48] Kinetic isotope effects for CO_2 and O_2 report indirectly on the relevant barriers[49] and can be measured by mass spectrometry.[50] Investigating the relationship between transition state barriers and kinetic parameters will help delineate which reaction steps limit carboxylation and oxygenation in different Rubisco lineages.[5]

There remains some disagreement about the precise ordering of carboxylation steps[5,14,15] and the mechanism of oxygenation is not well understood.[48] Chemical reasoning about the mechanisms of Rubisco carboxylation and oxygenation would benefit from progress in structural biology - intermediate and transition state analogs should be used to capture the active site at various points along the reaction trajectory.[14,40,48,51] If experiments and structural analyses confirm that the above assumptions hold for all Rubiscos, it would greatly limit our capacity to engineer Rubisco and strongly suggest that alternative strategies for improving carbon fixation should be pursued.[19,52–54] If, however, these assumptions are invalidated, many enzyme engineering strategies would become viable. Such data and analyses will be instrumental in guiding the engineering of carbon fixation for the next decade.

Acknowledgements

We would like to thank Uri Alon, Kapil Amarnath, Doug Banda, Arren Bar-Even, Cecilia Blikstad, Jack Desmarais, Woodward Fischer, Vahe Galstyan, Laura Helen Gunn, Itai Halevy, Robert Nichols, Elad Noor, Jeremy Roop, Yonatan Savir, Patrick Shih, Daniel Stolper, Dan Tawfik, Guillaume Tcherkez, Tsvi Tlusty and Renee Wang for helpful conversations and comments on the manuscript. This research was supported by the US National Science Foundation (grant MCB-1818377); the European Research Council (project NOVCARBFIX 646827); the Israel Science Foundation (grant No. 740/16); the ISF-NRF Singapore joint research program (grant No. 7662712); Beck-Canadian Center for Alternative Energy Research; Dana and Yossie Hollander; Ullmann Family Foundation; Helmsley Charitable Foundation; The Larson Charitable Foundation; Wolfson Family Charitable Trust; Charles Rothschild; Selmo Nussenbaum. R.M. is the Charles and Louise Gartner professional chair. A.I.F. was supported by a National Science Foundation Graduate Research Fellowship. Y.M.B. is an Azrieli fellow.

Competing Interests

The authors declare no competing interests.

References

(1) Raven, J. A. (2013) Rubisco: still the most abundant protein of Earth? *New Phytol. 198*, 1–3.
(2) Galmés, J., Kapralov, M. V., Andralojc, P. J., Conesa, M. À., Keys, A. J., Parry, M. A. J., and Flexas, J. (2014) Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ. 37*, 1989–2001.
(3) Bar-On, Y. M., Phillips, R., and Milo, R. (2018) The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U. S. A. 115*, 6506–6511.
(4) Bar-On, Y. M., and Milo, R. (2019) The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U. S. A. 116*, 4738–4743.
(5) Tcherkez, G. G. B., Farquhar, G. D., and Andrews, T. J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U. S. A. 103*, 7246–7251.
(6) Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010) Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci. U. S. A. 107*, 3475–3480.
(7) Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry 50*, 4402–4410.
(8) Bathellier, C., Tcherkez, G., Lorimer, G. H., and Farquhar, G. D. (2018) Rubisco is not really so bad. *Plant Cell Environ. 41*, 705–716.
(9) Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res. 40*, D770–5.
(10) Bloom, A. J., and Lancaster, K. M. (2018) Manganese binding to Rubisco could drive a photorespiratory pathway that increases the energy efficiency of photosynthesis. *Nat Plants 4*, 414–422.
(11) Busch, F. A., Sage, R. F., and Farquhar, G. D. (2018) Plants increase CO2 uptake by assimilating nitrogen via the photorespiratory pathway. *Nat Plants 4*, 46–54.
(12) Bauwe, H., Hagemann, M., and Fernie, A. R. (2010) Photorespiration: players, partners and origin. *Trends Plant Sci. 15*, 330–336.
(13) Liu, D., Ramya, R. C. S., and Mueller-Cajar, O. (2017) Surveying the expanding prokaryotic Rubisco multiverse. *FEMS Microbiol. Lett. 364*.
(14) Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C., and Lorimer, G. H. (1998) Mechanism of Rubisco: The Carbamate as General Base. *Chem. Rev. 98*, 549–562.
(15) Andersson, I. (2008) Catalysis and regulation in Rubisco. *J. Exp. Bot. 59*, 1555–1568.
(16) Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R., and Parry, M. A. J. (2016) Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2. *Plant J. 85*, 148–160.
(17) Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A. 113*, E5354–62.
(18) Zhu, X.-G., Long, S. P., and Ort, D. R. (2010) Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol. 61*, 235–261.
(19) Wu, A., Hammer, G. L., Doherty, A., von Caemmerer, S., and Farquhar, G. D. (2019) Quantifying impacts of enhancing photosynthesis on crop yield. *Nat Plants 5*, 380–388.

(20) Spreitzer, R. J., and Salvucci, M. E. (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol. 53*, 449–475.
(21) Raven, J. A., Beardall, J., and Sánchez-Baracaldo, P. (2017) The possible evolution and future of CO2-concentrating mechanisms. *J. Exp. Bot. 68*, 3701–3716.
(22) Pierce, J., Lorimer, G. H., and Reddy, G. S. (1986) Kinetic mechanism of ribulosebisphosphate carboxylase: evidence for an ordered, sequential reaction. *Biochemistry 25*, 1636–1644.
(23) Wildman, S. G. (2002) Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate carboxylase-oxygenase). *Photosynth. Res. 73*, 243–250.
(24) Jordan, D. B., and Ogren, W. L. (1981) Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature 291*, 513–515.
(25) Parry, M. A. J., Keys, A. J., and Gutteridge, S. (1989) Variation in the Specificity Factor of C3 Higher Plant Rubiscos Determined by the Total Consumption of Ribulose-P2. *J. Exp. Bot. 40*, 317–320.
(26) Bainbridge, G., Madgwick, P., Parmar, S., Mitchell, R., Paul, M., Pitts, J., Keys, A. J., and Parry, M. A. J. (1995) Engineering Rubisco to change its catalytic properties. *J. Exp. Bot. 46*, 1269–1276.
(27) Shih, P. M., Occhialini, A., Cameron, J. C., Andralojc, P. J., Parry, M. A. J., and Kerfeld, C. A. (2016) Biochemical characterization of predicted Precambrian RuBisCO. *Nat. Commun. 7*, 10382.
(28) Shoval, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., Dekel, E., Kavanagh, K., and Alon, U. (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science 336*, 1157–1160.
(29) Boyd, R. A., Cavanagh, A. P., Kubien, D. S., and Cousins, A. B. (2019) Temperature response of Rubisco kinetics in Arabidopsis thaliana: thermal breakpoints and implications for reaction mechanisms. *J. Exp. Bot. 70*, 231–242.
(30) Evans, J. R., Sharkey, T. D., Berry, J. A., and Farquhar, G. D. (1986) Carbon Isotope Discrimination measured Concurrently with Gas Exchange to Investigate CO2 Diffusion in Leaves of Higher Plants. *Funct. Plant Biol. 13*, 281–292.
(31) Caemmerer, S. V., and Evans, J. R. (1991) Determination of the Average Partial Pressure of CO2 in Chloroplasts From Leaves of Several C3 Plants. *Funct. Plant Biol. 18*, 287–305.
(32) Read, B. A., and Tabita, F. R. (1994) High substrate specificity factor ribulose bisphosphate carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant cyanobacterial RubiSCO containing “algal” residue modifications. *Arch. Biochem. Biophys. 312*, 210–218.
(33) Andrews, T. J., and Lorimer, G. H. (1987) 3 - Rubisco: Structure, Mechanisms, and Prospects for Improvement, in *Photosynthesis* (Hatch, and Boardman, N. K., Eds.), pp 131–218. Academic Press.
(34) Tcherkez, G. G., Bathellier, C., Farquhar, G. D., and Lorimer, G. H. (2018) Commentary: Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO’s Efficiency May Not Be So Constrained After All. *Front. Plant Sci. 9*, 929.
(35) Cummins, P. L., Kannappan, B., and Gready, J. E. (2018) Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO’s Efficiency May Not Be So Constrained After All. *Front. Plant Sci. 9*, 183.
(36) Hopfield, J. J. (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U. S. A. 71*, 4135–4139.
(37) Savir, Y., and Tlusty, T. (2007) Conformational proofreading: the impact of conformational changes on the specificity of molecular recognition. *PLoS One 2*, e468.
(38) Gubernator, B., Bartoszewski, R., Kroliczewski, J., Wildner, G., and Szczepaniak, A. (2008) Ribulose-1,5-bisphosphate carboxylase/oxygenase from thermophilic cyanobacterium Thermosynechococcus elongatus. *Photosynth. Res. 95*, 101–109.

(39) Young, J. N., Heureux, A. M. C., Sharwood, R. E., Rickaby, R. E. M., Morel, F. M. M., and Whitney, S. M. (2016) Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *J. Exp. Bot. 67*, 3445–3456.
(40) Andersson, I., and Backlund, A. (2008) Structure and function of Rubisco. *Plant Physiol. Biochem. 46*, 275–291.
(41) Kapralov, M. V., and Filatov, D. A. (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol. 7*, 73.
(42) Silberg, J. J., Endelman, J. B., and Arnold, F. H. (2004) SCHEMA-guided protein recombination. *Methods Enzymol. 388*, 35–42.
(43) Fowler, D. M., and Fields, S. (2014) Deep mutational scanning: a new style of protein science. *Nat. Methods 11*, 801–807.
(44) Chin, J. W. (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem. 83*, 379–408.
(45) Akiva, E., Copp, J. N., Tokuriki, N., and Babbitt, P. C. (2017) Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc. Natl. Acad. Sci. U. S. A. 114*, E9549–E9558.
(46) Orr, D. J., Alcântara, A., Kapralov, M. V., Andralojc, P. J., Carmo-Silva, E., and Parry, M. A. J. (2016) Surveying Rubisco Diversity and Temperature Response to Improve Crop Photosynthetic Efficiency. *Plant Physiol. 172*, 707–717.
(47) Sharwood, R. E., Ghannoum, O., Kapralov, M. V., Gunn, L. H., and Whitney, S. M. (2016) Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C3 photosynthesis. *Nat Plants 2*, 16186.
(48) Tcherkez, G. (2016) The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ. 39*, 983–997.
(49) Hayes, J. M. (2001) Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes. *Rev. Mineral. Geochem. 43*, 225–277.
(50) McNevin, D. B., Badger, M. R., Whitney, S. M., von Caemmerer, S., Tcherkez, G. G. B., and Farquhar, G. D. (2007) Differences in Carbon Isotope Discrimination of Three Variants of D-Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Reflect Differences in Their Catalytic Mechanisms. *J. Biol. Chem. 282*, 36068–36076.
(51) Schneider, G., Lindqvist, Y., and Brändén, C. I. (1992) RUBISCO: structure and mechanism. *Annu. Rev. Biophys. Biomol. Struct. 21*, 119–143.
(52) McGrath, J. M., and Long, S. P. (2014) Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol. 164*, 2247–2261.
(53) Bar-Even, A. (2017) Daring metabolic designs for enhanced plant carbon fixation. *Plant Sci.*
(54) South, P. F., Cavanagh, A. P., Liu, H. W., and Ort, D. R. (2019) Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science 363*, eaat9077.