Revisiting tradeoffs in Rubisco kinetic parameters

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

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

Abstract

Rubisco is 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 tradeoffs embedded in the catalytic mechanism constrain its specificity and maximum catalytic rate. However, the reasoning that established this view was based on data from ≈20 organisms. We re-examine these tradeoff models using a dataset from ≈300 organisms. Most correlations are substantially attenuated, with the inverse relationship between carboxylation k_{cat} and specificity $S_{C/O}$ being a key example. Only one tradeoff model survives in our dataset. In this model, increasing catalytic efficiency (k_{cat}/K_M) for carboxylation requires increased catalytic efficiency for the competing oxygenation reaction, evidenced by strong power-law correlation between catalytic efficiencies. Our results imply that Rubisco evolution is constrained primarily by the physicochemical limits of O_2/CO_2 discrimination, which should reframe efforts to engineer this very central enzyme.

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 (Raven, 2013). Typically, 20-30% of total soluble protein in C3 plant leaves is Rubisco (Galmés et al., 2014). As Rubisco is so highly expressed and plants are the dominant constituents of planetary biomass (Bar-On et al., 2018), it is often said that Rubisco is the most abundant enzyme on Earth (Raven, 2013). 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 (Bar-Even et al., 2011; Bathellier et al., 2018; Savir et al., 2010; Shih et al., 2016; Tcherkez et al., 2006). Typical central metabolic enzymes have a

maximum catalytic rate k_{cat} ≈ 80 s^{-1} (Bar-Even et al., 2011), but more than 95% of Rubisco carboxylation k_{cat} values are between 1-10 s^{-1}.

In addition to relatively low carboxylation k_{cat} values, all known Rubiscos are capable of reacting "wastefully" with O_2 in a process called oxygenation (Figure 1A-B). Both carboxylation and oxygenation of the Rubisco substrate ribulose 1,5-bisphosphate (RuBP) are energetically favorable, but only carboxylation is considered productive because it incorporates carbon from CO_2 into precursors that can generate biomass (Figure 1AB). Oxygenation is 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 (Bauwe et al., 2010; Busch et al., 2017). As such, oxygenation by Rubisco can substantially reduce the net rate of carboxylation. There are at least four distinct Rubisco isoforms in nature (Jaffe et al., 2018; Liu et al., 2017), but all known Rubiscos catalyze both carboxylation and oxygenation of RuBP through the multistep mechanism described in Figure 1A (Andersson, 2008; Cleland et al., 1998). Despite the fact that many autotrophs depend on Rubisco carboxylation for growth, all known Rubiscos are relatively slow carboxylases and fail to exclude oxygenation (Figure 1A-B).

As a concrete example, the fastest-carboxylating Rubisco ever observed (at 25 ^{o}C) is from the freshwater cyanobacterium *S. elongatus* PCC 7942 (Occhialini et al., 2016). The PCC 7942 Rubisco has a maximum per-active site carboxylation rate ($k_{cat,C}$) of 14.4 s^{-1}. However, because present day atmosphere contains abundant O_2 and relatively little CO_2 (≈21% O_2, ≈0.04% CO_2), this Rubisco carboxylates at a rate 20-fold below maximum in ambient conditions (≈0.7 s^{-1} per active site, rate law in Figure 1A). Due to its relatively low specificity towards CO_2, PCC 7942 Rubisco will also oxygenate RuBP in ambient conditions at a rate (≈0.3 s^{-1}) that is about half the carboxylation rate, which would necessitate substantial photorespiratory flux to recycle 2PG. As downstream processing of 2PG by the canonical C_2 photorespiratory pathway leads to the loss of one carbon for every two 2PG (Bauwe et al., 2010; Busch et al., 2017), every two oxygenations "undoes" a carboxylation. As such, the net rate of carboxylation by PCC 7942 Rubisco is roughly $R_C - R_O/2 \approx 0.7 - 0.15 = 0.55$ carboxylations per second, more than 25 times less than $k_{cat,C}$. Indeed, all known cyanobacteria use a CO_2-concentrating mechanism (CCM) to ensure that Rubisco functions in a CO_2-rich environment. High CO_2 ensures that oxygenation is inhibited and that carboxylation proceeds at near-maximum rate (Mangan et al., 2016; Reinhold et al., 1991). Just tenfold enrichment of CO_2 above ambient would increase the carboxylation rate of PCC 7942 Rubisco to ≈5 s^{-1} and suppress oxygenation to ≈0.2 s^{-1} per active site, giving a net carboxylation rate of ≈ 4.6 s^{-1} per active site.

For comparison, the well-studied Rubisco from spinach leaves (*S. oleracea*) carboxylates much more slowly ($k_{cat,C}$ ≈ 3 s^{-1}) but also has much greater affinity towards CO_2 than the *S. elongatus* enzyme (half-maximum CO_2 concentration K_C ≈ 12 μM for spinach as compared to K_C ≈ 170 μM

for *S. elongatus* PCC 7942). As a result, the spinach enzyme outperforms the cyanobacterial enzyme in present day atmosphere, achieving a carboxylation rate of ≈1.2 s^{-1} and an oxygenation rate of ≈0.4 s^{-1}. This represents a net carboxylation rate of ≈ 1 s^{-1}, nearly double that of the cyanobacterial example above. Spinach is a C3 plant, meaning it does not have a CO_2 concentrating mechanism, which likely explains why it evolved to use a slow-but-specific enzyme for catalysis in ambient conditions. Still, many other enzymes catalyze far more than 1.2 reaction per second (Bar-Even et al., 2011), which leads many to wonder if Rubisco catalysis could be improved by engineering. Improved Rubisco carboxylation is expected to increase C3 crop yields (Zhu et al., 2010), but a substantially improved enzyme has evaded bioengineers for decades (Spreitzer and Salvucci, 2002). The repeated evolution of diverse CCMs, which modulate the catalytic environment rather than the enzyme itself, raises further doubts about whether Rubisco catalysis can be strictly improved (Raven et al., 2017).

Various nomenclature has been used to describe the kinetics of Rubisco carboxylation and oxygenation (Pierce et al., 1986; Savir et al., 2010; Tcherkez et al., 2006) since its discovery in the 1950s (Wildman, 2002). Here we use $k_{cat,C}$ and $k_{cat,O}$ to denote turnover numbers (maximum per active site catalytic rates in units of s^{-1}) 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 1A-C). Since $S_{C/O}$ relates only to the ratio of kinetic parameters, it should be noted that higher $S_{C/O}$ does not necessarily imply higher carboxylation rates. Rather, absolute carboxylation and oxygenation rates depend on the CO_2 and O_2 concentrations which can vary between organisms.

As data on bacterial, archaeal and plant rubiscos has accumulated over the decades, many researchers have noted that fast-carboxylating Rubiscos are typically less CO_2-specific (Bainbridge et al., 1995; Jordan and Ogren, 1981; Parry et al., 1989). 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 higher catalytic efficiency towards O_2 ($k_{cat,O}/K_O$). This negative correlation between $k_{cat,C}$ and $S_{C/O}$ is often cited to motivate the idea that the Rubisco mechanism imposes a tradeoff between carboxylation rate and specificity that constrains the evolution of this enzyme (Bainbridge et al., 1995; Savir et al., 2010; Shih et al., 2016; Tcherkez et al., 2006). Indeed, if the Rubisco mechanism imposes a tradeoff between $k_{cat,C}$ and $S_{C/O}$ we would expect strong correlation between those parameters because Rubisco is so central to autotrophic life and has, therefore, likely experienced strong selection pressure. As diagrammed in Figure 1C, strong selection for Rubisco carboxylation could in theory push the enzyme towards a point where its kinetics can be improved no further. Since different kinetics are preferable in different CO_2 and O_2 concentrations (as described above) strong selection is claimed to produce a situation in which the kinetics of natural enzymes trace out a curve determined by the underlying tradeoff (Savir et al., 2010; Shoval et al., 2012; Tcherkez et al., 2006).

Two distinct tradeoff models have been proposed to explain the observed correlations between Rubisco kinetic parameters (Savir et al., 2010; Tcherkez et al., 2006). Although the proposed

models are substantively different, both models imply limitations on the concurrent improvement of the maximum carboxylation rate ($k_{cat,C}$) and specificity ($S_{C/O}$) of natural Rubiscos. While these hypotheses appeal to physical and chemical intuition, they are based on data from only ≈20 organisms. Here we take advantage of the accumulation of new data - more than 200 Rubisco variants have been characterized since 2010 - to examine whether new data evidence the same correlations. We find that most previously-reported correlations between Rubisco kinetic parameters are substantially attenuated by the addition of new data, with the negative correlation between $k_{cat,C}$ and specificity $S_{C/O}$ being a key example.

Only one previously-reported correlation remains both strong and statistically significant in the extended dataset - a power-law correlation between the catalytic efficiency for carboxylation ($k_{cat,C}/K_C$) and the catalytic efficiency for oxygenation ($k_{cat,O}/K_O$) first reported in (Savir et al., 2010). We propose a simple physico-chemical model based on the Rubisco mechanism that can explain this very strong correlation. 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.

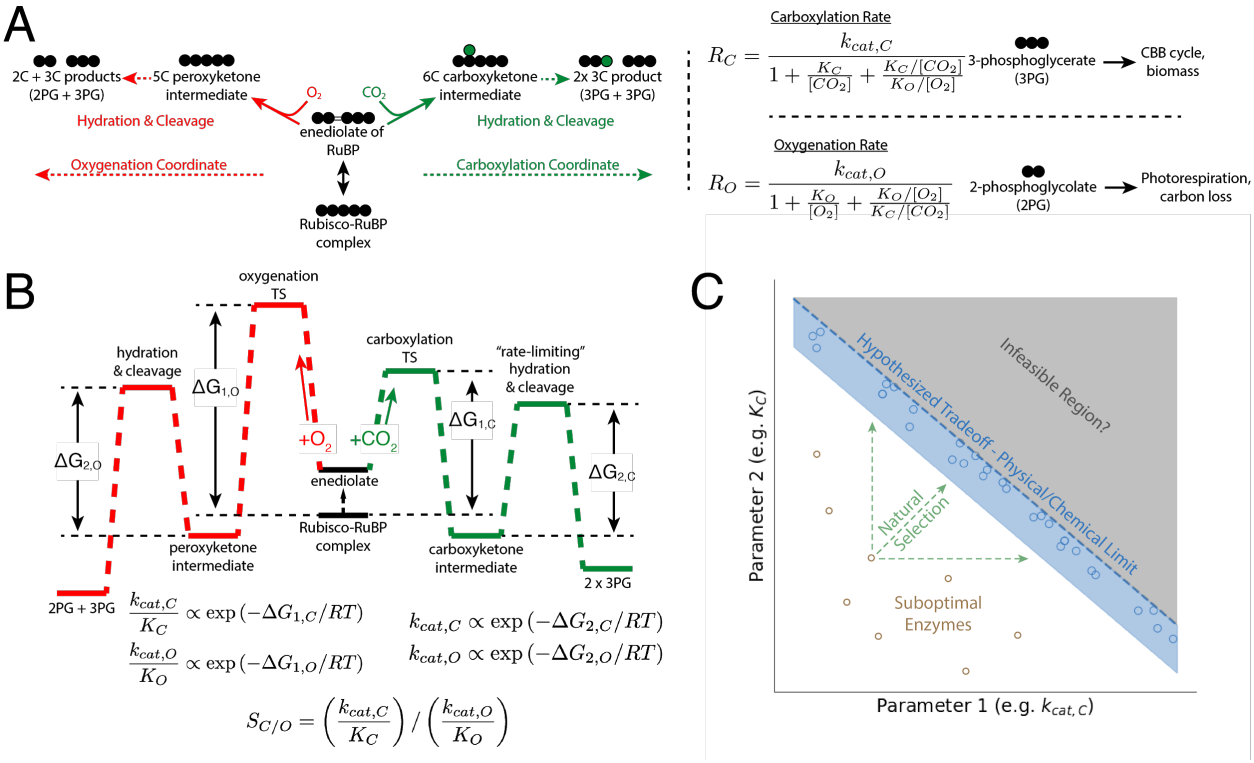

Figure 1: *Description of the catalytic mechanism of Rubisco. The "middle-out" diagram in Panel A shows the ordered mechanisms of carboxylation and oxygenation. Circles represent carbon atoms. RuBP is first isomerized to an enediolate before carboxylation or oxygenation can occur. Addition of CO_2 or O_2 to the enediolate of RuBP are considered irreversible as are the subsequent hydration and cleavage steps of carboxylation and oxygenation arms. 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, which produces 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. 2PG is not part of the CBB cycle and must be recycled through photorespiration to avoid the loss of both carbons in 2PG to central metabolism. Per-active site rates of carboxylation (R_C) and oxygenation (R_O) can be calculated from kinetic parameters and the CO_2 and O_2 concentrations. The reaction coordinate diagram in panel (B) mirrors panel A and describes Rubisco carboxylation and oxygenation as a function of two "effective" barriers as in (Savir et al., 2010). The first effective barrier describes enolization and gas addition while the second describes hydration and bond cleavage. Given standard assumptions (described in SI), the respective catalytic efficiencies (k_{cat}/K_M) are related to the height of the first effective barrier while the k_{cat}s are related to the second. The first barrier to oxygenation is drawn higher than for carboxylation because Rubisco oxygenation is typically much slower than carboxylation. The net reactions of RuBP carboxylation and oxygenation are both quite thermodynamically favorable ($\Delta_r G'^m \approx$ -28 kJ/mol and -520 kJ/mol respectively (Flamholz et al., 2012)). As kinetic parameters are linearly related to the log of effective energy barriers, energetic tradeoffs should manifest as linear correlations in a log-log plot of kinetic parameters (C). As Rubisco is central to photoautotrophic growth, we expect that natural selection has pushed the enzyme towards the upper limits of its catalytic capacity - i.e. towards the blue shaded region.

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 2A). 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 were measured at 25 °C and near pH 8 to ensure that measured values are comparable (Methods).

The resulting dataset contains Rubisco kinetic parameters from a total of 286 distinct species including 319 $S_{C/O}$ values, 275 $k_{cat,C}$ values, 310 K_C values, 198 $k_{cat,O}$ values and 256 K_O values (Figure 2B). The Michaelis constant for RuBP is not measured frequently and so only 45 values were extracted. In 198 cases there was sufficient data to calculate catalytic efficiencies for carboxylation ($k_{cat,C}/K_C$) and oxygenation ($k_{cat,O}/K_O$, 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 2B). As such, we focus here on the kinetic parameters of Form I Rubiscos.

Rubisco kinetic parameters display very narrow dynamic range, with geometric standard deviations being well-below one order-of-magnitude for all parameters (Figure 2C). The geometric standard deviation, denoted σ^*, expresses multiplicative variability in the dataset.

Rubisco displays extremely low variation in $k_{cat,C}$ (σ^* = 1.5), especially in comparison to other enzymes for which > 20 k_{cat} measurements are available (Figure S4). The median σ^* for these other enzymes is 6.9, more than fourfold higher than for Rubisco. Specificity $S_{C/O}$ displays the least variation (σ^* = 1.3) of all parameters, though this may be due in part to overrepresentation of C3 plants in the dataset, which occupy a narrow range of $S_{C/O}$ ≈ 80-120. Nonetheless, measurements of $S_{C/O}$ for Form I and Form II enzymes are clearly distinct in this dataset, with values ranging from and 7-15 for Form IIs and roughly 50-200 for Form Is (Figure 2C, SI).

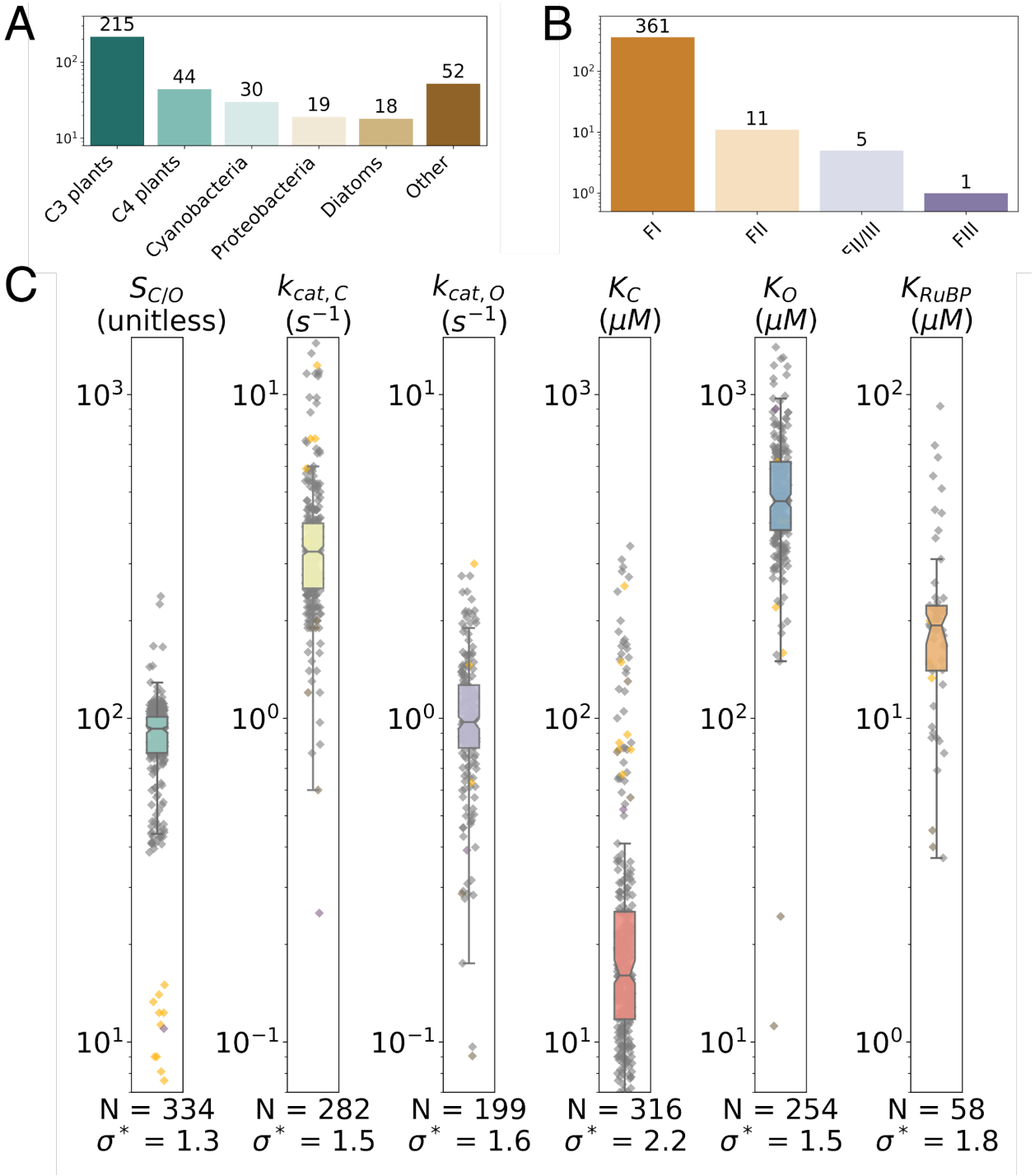

Figure 2: *Summary of the full extended dataset. We collected measurements of Rubisco kinetic parameters from a variety of organisms (A) representing four classes of Rubisco isoforms (B). The bulk of data represent Form I enzymes from green lineage organisms (A-B). As shown in panel C, the assembled 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. Whiskers show the range of parameter values among Form 1 enzymes, the colored box gives the range of the central 50% of Form I values and the notch indicates the median. N is the number values and the geometric standard deviation of Form I data is reported as σ^*. $\sigma^* < 3$ for all parameters, meaning that a single standard deviation varies over less than*

threefold. All data presented in this figure are from wild-type Rubiscos measured at 25 °C near pH 8. More detailed histograms are given in Figure S3.

Energetic Tradeoffs Tend to Produce Power-Law Correlations

All measured kinetic parameters ($S_{C/O}$, $k_{cat,C}$, K_C, $k_{cat,O}$ and K_O) are by definition mathematically related to the microscopic rate constants of the Rubisco mechanism (SI). Given common assumptions about irreversible and rate limiting steps (elaborated below and in the SI), this multi-step mechanism can be simplified so that logarithms of measured kinetic parameters are proportional to effective transition state (TS) barriers (Figure 1B, SI). As such, tradeoffs between kinetic parameters are expected to emerge if effective TS barriers are constrained to vary together (Figure 1C). If, for example, lowering the effective TS 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 between these barrier heights such that $\Delta G_{1,C} \propto -\Delta G_{1,C}$. Since, as shown in Figure 1B, $k_{cat,C}/K_C$ is related to the first effective carboxylation barrier ($\ln(k_{cat,C}/K_C) \propto -\Delta G_{1,C}$) and $k_{cat,C}$ to the second ($\ln(k_{cat,C}) \propto -\Delta G_{2,C}$), linear correlation between transition state barrier energies should translate into log-scale correlation between kinetic parameters such that $ln\,(k_{cat,C}/K_C) \propto -ln\,(k_{cat,C})$. These log-linear 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 when three conditions are met: (I) the associated energy barriers are subject to a tradeoff that forces them to vary together; (II) these constraints affect the net rate of carboxylation by Rubisco; and (III) the selection pressure imposed during Rubisco evolution was sufficient to reach the limits imposed by the tradeoff (as diagrammed in Figure 1C). As Rubisco is the central enzyme of photoautotrophic growth, we assume here that it evolved under selection pressure towards maximizing the net rate of carboxylation in each host (Savir et al., 2010). Notably, different host physiologies and growth environments can affect the catalytic environment Rubisco experiences - Rubiscos in different organisms experience different temperature, pH and prevailing CO_2 and O_2 concentrations (e.g. due to an anaerobic host or a CCM enriching CO_2), which we expect to select for different combinations of kinetic parameters (Figure 1C).

Correlations between kinetic parameters of Form I Rubiscos

As in (Savir et al., 2010; Tcherkez et al., 2006) we performed a correlation analysis to investigate relationships between Rubisco kinetic parameters. Pairwise correlations between log-transformed Form I Rubisco kinetic parameters are given in Figure 3 (linear scale correlations are reported in Figure S8).

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 (Savir et al., 2010; Tcherkez et al., 2006). In previous analyses these pairs correlated very strongly, with Pearson correlation coefficients R ≈

0.9 for both. However, both correlations are substantially attenuated by the addition of new data ($R \approx 0.6$, Figure 3). Figure 4 inspects these two correlations in greater detail. Figure 4A plots $k_{cat,C}$ against $S_{C/O}$ and shows that these parameters are only modestly correlated in the extended dataset, with $R \approx 0.6$ (and extremely sensitive to outliers) as compared to $R \approx 0.9$ in previous analyses (Savir et al., 2010; Tcherkez et al., 2006). Similarly, Figure 4B plots $k_{cat,C}$ against K_C and shows that this correlation is also weakened, with $R \approx 0.5$ as compared to $R \approx 0.9$ in previous work (Savir et al., 2010). We interpret the weakened correlations detailed in Figures 3-4 as evidence that previously proposed tradeoffs may need to be experimentally revisited. Examining Figure 3 shows that only one pair of parameters, $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$, correlate with $R > 0.7$ on a log scale. This correlation between catalytic efficiencies for carboxylation and oxygenation is the strongest observed by far ($R = 0.93$, $P < 10^{-10}$, Figure 3). We discuss possible explanations for this very strong correlation in detail below.

One might wonder why so many pairs of Rubisco kinetic parameters correlate with appreciable R values (e.g. $R > 0.3$ for 11 of 28 pairs in Figure 3). We note that some level of correlation is expected because the measured parameters are mathematically interrelated through the microscopic mechanism of Rubisco as it is commonly understood. 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 level of correlation even in the absence of any tradeoff (SI). Similarly, $S_{C/O}$ is defined as $(k_{cat,C}/K_C) / (k_{cat,O}/K_O)$ and might correlate positively with $k_{cat,C}$ for this reason. Because modest correlation is expected, we focus here only on very strong correlations since these may yield insight into mechanistic constraints on Rubisco evolution.

Principal components analysis (PCA) of Rubisco kinetic parameters was previously used to interrogate constraints on Rubisco evolution. It was argued that Rubisco adaptation is constrained to a one-dimensional landscape because the first principal component (PC1) explained > 90% of the variance in Rubisco kinetics. In a one-dimensional landscape model all kinetic parameters are tightly interrelated so that changing one (e.g. $k_{cat,C}$) forces all others to assume predetermined values (Savir et al., 2010). However, our extended dataset is not well-approximated as one-dimensional. While the orientation of PC1 is not substantially altered by the addition of tenfold more measurements, it now explains ≈70% instead of >90% of the variance in Rubisco kinetics (Savir et al., 2010). Three principal components are required to explain >90% of the variation in our extended dataset (SI), consistent with the overall reduction in pairwise correlation documented in Figure 3. We therefore proceed to ask whether the correlations predicted by specific tradeoff models advanced in (Savir et al., 2010; Tcherkez et al., 2006) are supported by the extended dataset.

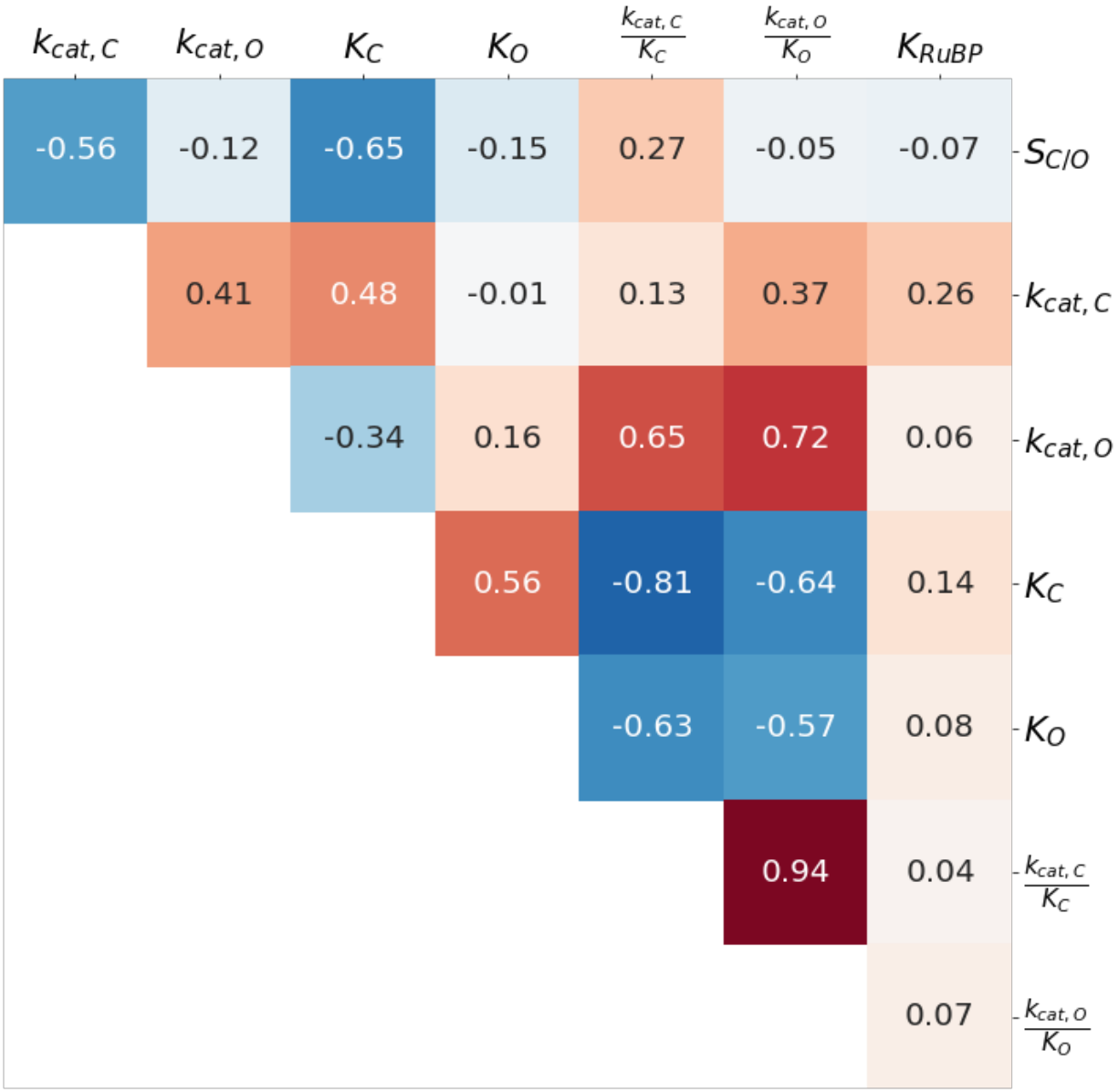

Figure 3: *Correlations between kinetic parameters are mostly attenuated by addition of new data. The figure summarizes Pearson correlations (R) between pairs of log-transformed Form I Rubisco kinetic parameters. When multiple measurements of the same Rubisco were available, the median value was used (Methods). The $S_{C/O}$-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. None of these pairs give R > 0.7. The strongest observed correlation is between the catalytic efficiencies for carboxylation and oxygenation, $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ (R = 0.94).*

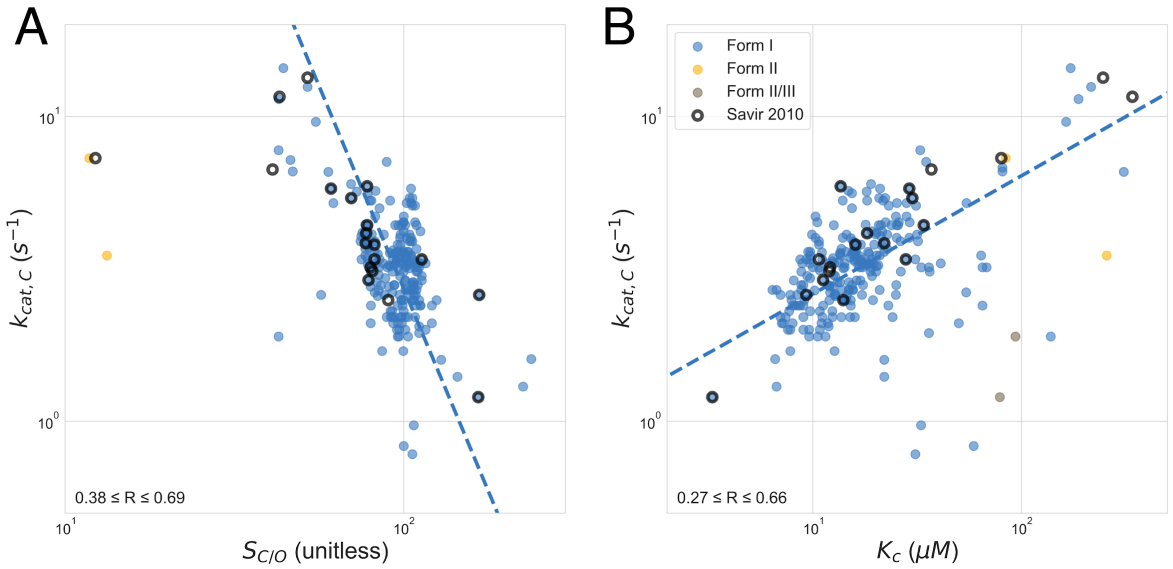

Figure 4: *Focal correlations of previous analyses are not robust to new data. Points with black outlines are those in Savir et al. 2010 and dashed blue lines represent the best fit to all Form I Rubiscos in the extended dataset. Panel A plots the maximum carboxylation rate $k_{cat,C}$ against specificity $S_{C/O}$ as in (Tcherkez et al., 2006). Considering only Form I Rubiscos, $k_{cat,C}$ and $S_{C/O}$ correlate with $R \approx -0.6$. Bootstrapping gives very wide 95% confidence intervals (CIs) of (-4.0, -2.0) for the fit exponent and ($3x10^4$, $3x10^8$) for the exponential prefactor (the slope and intercept in log-log scale respectively) indicating that the form of $k_{cat,C}$-$S_{C/O}$ correlation is very uncertain. Panel B plots $k_{cat,C}$ against the Michaelis constant for CO_2 (K_C) as in (Savir et al., 2010; Tcherkez et al., 2006). $R \approx 0.5$ as compared to the previously reported value of 0.92. This fit is substantially more robust to outliers with bootstrapping giving 95% CIs of (0.3, 0.5) and (0.8, 1.5) for the fit exponent and prefactor respectively. More detailed plots are given in Figure S5.*

Re-evaluation of Proposed Tradeoff Models

Two distinct mechanistic tradeoff models were advanced in (Savir et al., 2010; Tcherkez et al., 2006). Savir et al. 2010 cast these proposals in energetic terms by relating the measured catalytic parameters to effective transition state barrier heights (Figure 1B, SI). The first tradeoff model posits that that increased specificity towards CO_2 necessitates a slower maximum carboxylation rate, $k_{cat,C}$ (Savir et al., 2010; Tcherkez et al., 2006). Tcherkez et al. 2006 propose that this tradeoff is caused by stabilization of the first carboxylation transition state. In 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 model can be construed in energetic terms as follows: lowering the effective barrier to CO_2 addition ($\Delta G_{1,C}$ in Figure 5A) will make Rubisco more CO_2-specific even if none of the oxygenation parameters change. The tradeoff model 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 5A).

Therefore, the energetic interpretation of the first 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 $k_{cat,C}/K_C$.

The energetic interpretation of this model was previously supported by an inverse power-law relationship between $k_{cat,C}$ and $k_{cat,C}/K_C$ (Savir et al., 2010). Our extended dataset does not, however, conform to the reported power-law (Figure 5B). It should be noted that 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 C3 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 (Caemmerer and Evans, 1991; Evans et al., 1986).

Absence of correlation does not necessarily imply the absence of a tradeoff. Rather, if the Rubisco mechanism couples $k_{cat,C}$ and $k_{cat,C}/K_C$, much decreased correlation over the extended dataset ($R < 0.4$) could result from several factors including bias in data collection leading to undersampling of faster Rubiscos (e.g. those from cyanobacteria) or, alternatively, insufficient selection pressure.

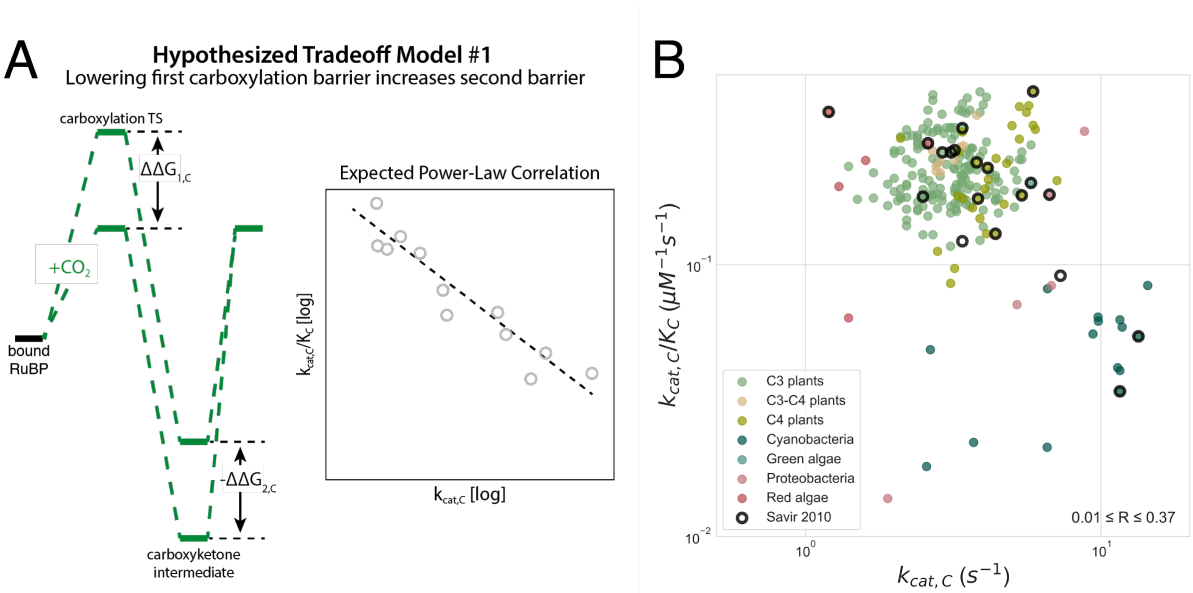

***Figure 5**: Negative power-law correlation between $k_{cat,C}$ and $k_{cat,C}/K_C$ is not supported by the extended dataset. Under the tradeoff model in panel (A), CO_2-specific Rubiscos have low barriers to enolization and CO_2 addition (first effective carboxylation barrier $\Delta G_{1,C}$), but lowering the first effective barrier necessarily increases the height of the effective barrier to subsequent carboxylation steps ($\Delta G_{2,C}$). This tradeoff might be due to coupling between the carboxylation transition state and the carboxyketone carboxylation intermediate (Tcherkez et al., 2006), where stabilizing the TS also stabilizes the intermediate diagrammed in panel (A) and described in Figure S2. In this case we would expect $\Delta G_{1,C}$ and $\Delta G_{2,C}$ to be negatively correlated, which would manifest as negative linear correlation on a log-log plot of $k_{cat,C}$ vs. $k_{cat,C}/K_C$. (B) The extended dataset does not evidence the expected power-law correlation ($R = 0.02$, $P = 0.8$ for Form I enzymes). Fitting the entire dataset gives $R = 0.13$ as compared to the previously-reported $R = -0.95$ in (Savir et al., 2010), where two outliers were omitted. Restricting focus to particular physiologies like C3 plants does not recover the expected correlation.*

The second mechanistic tradeoff model (Savir et al., 2010) - wherein faster CO_2 addition entails faster O_2 addition as well - is extremely well-supported by the addition of new data (Figure 6). This model was previously supported by a power-law relationship with an exponent of 0.5 between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ ($k_{cat,O}/K_O \propto (k_{cat,C}/K_C)^{0.5}$). As $k_{cat,C}/K_C$ is exponentially related to the first effective carboxylation barrier ($\ln(k_{cat,C}/K_C) \propto -\Delta G_{1,C}$) and $k_{cat,O}/K_O$ to the first effective oxygenation barrier ($\ln(k_{cat,O}/K_O) \propto -\Delta G_{1,O}$), the power-law relationship was taken to imply that decreasing the barrier to CO_2 addition will also decrease the barrier to O_2 addition ($0.5\ \Delta G_{1,C} - \Delta G_{1,O} = C$, Figure 6A).

The extended dataset evidences clear power-law correlation between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ (Figure 6B). While some Form II enzymes appear to be strictly inferior to the Form I enzymes on these axes, there is a clear "front" in the $k_{cat,C}/K_C$ vs. $k_{cat,O}/K_O$ plot. Most measurements lie along a robust line of positive correlation in a log-log plot. Fitting the Form I enzymes gives a remarkably high-confidence ($R = 0.93$, $P < 10^{-10}$) power-law relationship with and exponent of roughly 1.0: $k_{cat,C}/K_C \propto (k_{cat,O}/K_O)^{1.06}$ (Figure 6B).

$S_{C/O}$ is defined as the ratio of $k_{cat,C}/K_C$ to $k_{cat,O}/K_O$. A power law exponent of ≈1 implies a roughly constant slope relationship between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$, which in turn implies that $S_{C/O}$ is constant. However $S_{C/O}$ is clearly not constant - it varies about tenfold between Form I and Form II Rubiscos and about threefold among Form Is (Figure 2C). Subdividing the Form I enzymes by host physiology (e.g. C3 plants, C4 plants, cyanobacteria, etc.) reveals that all groups with sufficient data display a strong and statistically-significant power-law relationship between $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ (Figure 6C, SI (Tcherkez, 2016)). The power-law exponent differs consistently from the value of 0.5 given by (Savir et al., 2010). We now find a roughly 1:1 relationship of $\Delta G_{1,C} - \Delta G_{1,O} = C$, meaning that a decrease in the CO_2 addition barrier is associated with an *equal* decrease in the barrier to O_2 addition. We estimate a 95% confidence interval (CI) of 0.98-1.24 for the exponent of this power law relationship for Form I enzymes, or about double the previously-reported value.

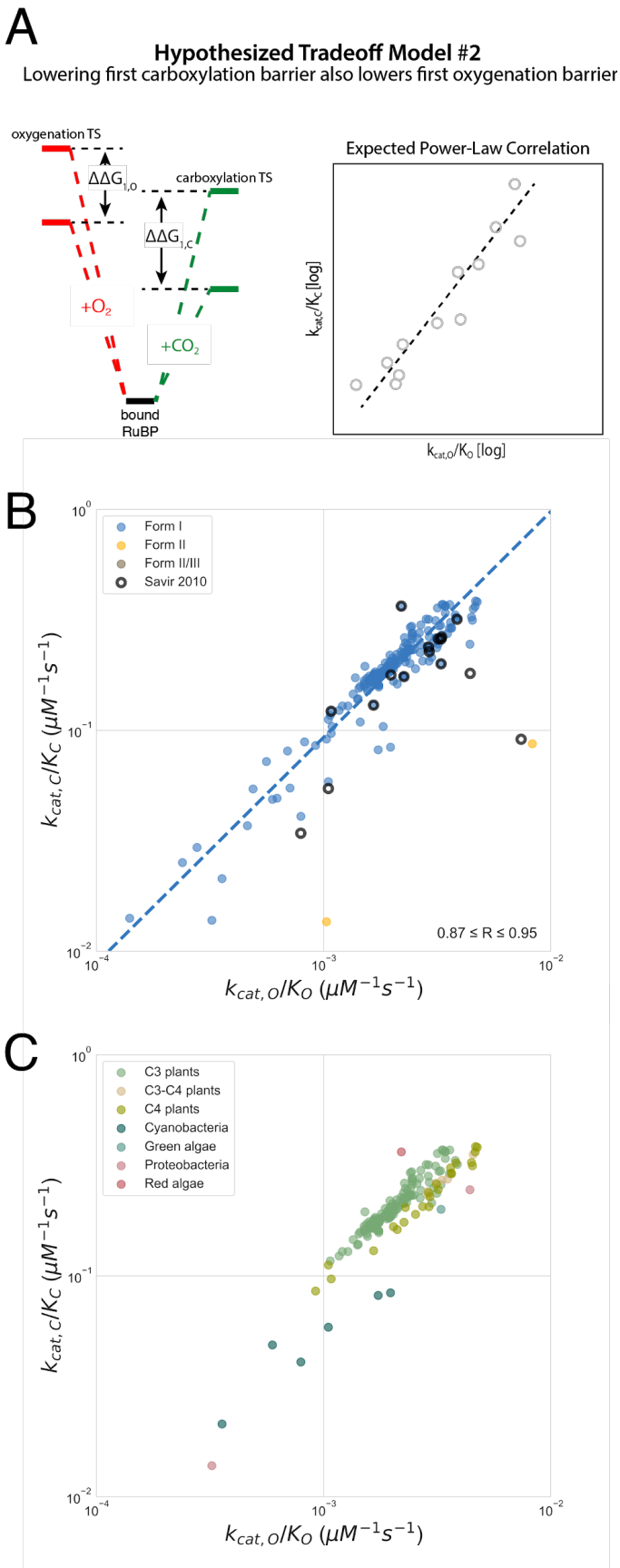

Figure 6: *The second mechanistic proposal of (Savir et al., 2010) is remarkably well-supported by the extended dataset, but with a different power-law exponent. (A) In this proposal, CO_2 and O_2 addition rates are coupled, with faster CO_2 addition necessitating faster O_2 addition. This can be framed in energetic terms, where 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, we would expect the barrier heights to be positively correlated, which would manifest as a positive linear correlation on a log-log plot of $k_{cat,C}/K_C$ vs $k_{cat,O}/K_O$. (B) most measurements cluster along a power-law (linear in log-log) front in the $k_{cat,C}/K_C$ vs $k_{cat,O}/K_O$ plot (dashed blue line, R = 0.94). While some Form I Rubiscos appear to lie beneath this front, Form II and Form II/III enzymes deviate most profoundly. A total least squares fit to the Form I enzymes produces a very strong power-law correlation ($P < 10^{-10}$, blue dashed line). 95% CIs for the exponent and prefactor are (0.93, 1.1) and (63, 199), respectively. The best fit power law is $k_{on,C} \sim (k_{on,O})^{1.04}$, but forcing $k_{on,C} \sim (k_{on,O})^{1.0}$ gives a fit of nearly identical quality. (C) Restricting focus to particular physiologies - e.g. C3 and C4 plants, cyanobacteria - reveals that each grouping obeys a distinct power law. These power laws differ primarily in the exponential prefactor, which causes variation in the Y-intercept but not the slope on a log-log plot. 95% CIs on the power-law exponent are (0.87, 1.01) for C3 plants, (0.82, 1.01) for C4 plants, (0.38, 1.31) for C3-C4 plants and (0.38, 1.06) for cyanobacteria.*

Implications for the mechanism of CO_2/O_2 discrimination by Rubisco

Figure 6 shows that effective barriers to CO_2 and O_2 addition and oxygenation appear to vary in proportion with each other ($\Delta G_{1,C}$ - $\Delta G_{1,O}$ ≈ constant). A roughly 1:1 correlation between effective barriers to CO_2 and O_2 addition suggests that a single factor controls both. We offer a model based on the known catalytic mechanism of Rubisco that could produce a 1:1 relationship between barriers. In this model, the RuBP-bound Rubisco active site fluctuates between reactive and unreactive states (Figures 7A and S10). The fraction of enzyme in the reactive state is denoted ϕ. In the unreactive state neither oxygenation or carboxylation is possible. In the reactive state, either gas can react at an intrinsic rate that does not vary across Rubiscos of the same class. Since RuBP must undergo enolization in order for either 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 enediolate of RuBP to CO_2 and O_2 respectively (SI, Figure S10). Under this model, we expect to observe a power-law relationship with exponent 1.0 between $\frac{k_{cat,C}}{K_C}$ and $\frac{k_{cat,O}}{K_O}$(SI). Since $S_{C/O}$ = ($k_{cat,C}/K_C$) / ($k_{cat,O}/K_O$), it should be roughly constant under this model (SI). Though $S_{C/O}$ varies the least of all measured Rubisco kinetic parameters (Figure 2C), it is not constant. However, Rubiscos isolated from hosts belonging to the same physiological grouping - e.g. C3 or C4 plants - do display a characteristic and roughly constant $S_{C/O}$ value independent of $k_{cat,C}/K_C$ (Figure 7B).

This model implies that ϕ can vary between related Rubiscos, perhaps by evolutionary tuning of the equilibrium constant for RuBP enolization. Since $S_{C/O}$ is independent of the equilibrium fraction of on-enzyme RuBP enolization (K_E), variation in K_E would affect $k_{cat,C}/K_C$ and $k_{cat,O}/K_O$ without changing $S_{C/O}$ (SI). Though individual groups of Rubiscos have roughly constant $S_{C/O}$, specificity clearly varies between C3 plants and cyanobacteria, for example. Variation in specificity could be achieved by adjusting the difference between intrinsic reactivities $\Delta G^*_{1,O} - \Delta G^*_{1,C}$through changes to the conformation of the enediolate of RuBP. This would produce roughly constant $S_{C/O}$ among C3 plants while permitting variation in $S_{C/O}$ between C3 plants, cyanobacteria and proteobacterial Form I Rubiscos. A full derivation of this model and discussion of its potential implications is given in the SI.

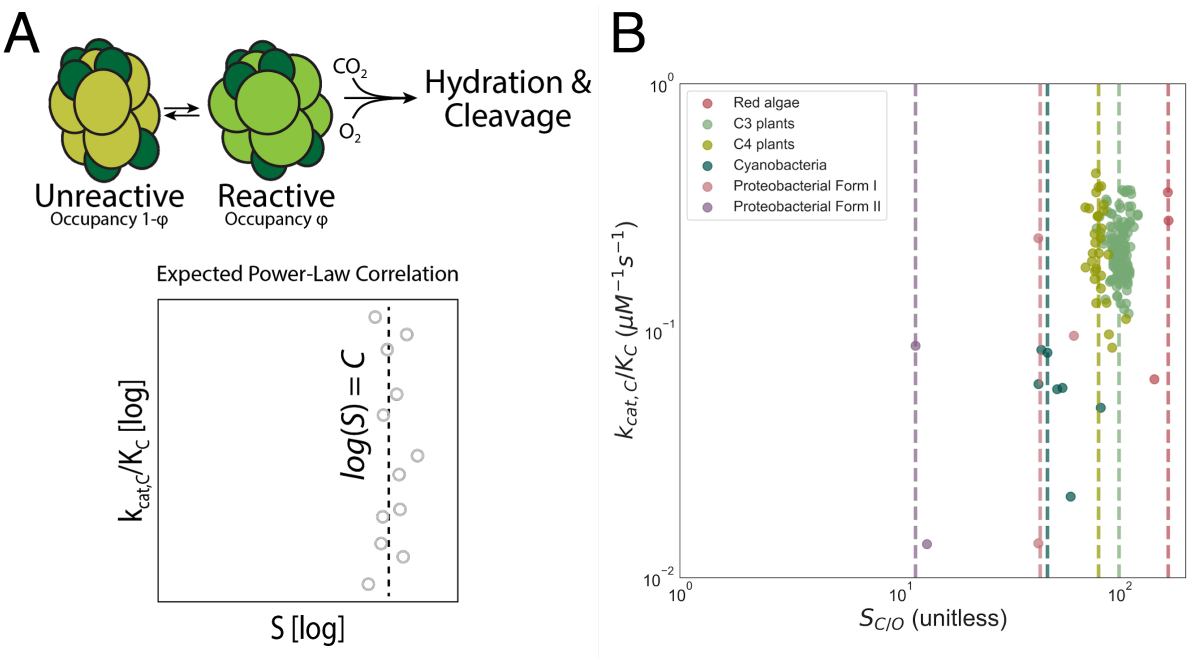

Figure 7: *A power-law relationship between $k_{cat,C}/K_C$ vs $k_{cat,O}/K_O$ with an exponent of roughly 1.0 can be explained by an active site that fluctuates between "reactive" and "unreactive" states. (A) In this model CO_2 and O_2 can react with the bound RuBP only if the enzyme is in the reactive state. If the difference in intrinsic reactivities of the active site complex ($\Delta G^*_{1,O} - \Delta G^*_{1,C}$) is organism-independent, we derive a power-law relationship between $k_{cat,C}/K_C$ vs $k_{cat,O}/K_O$ that has an exponent of 1.0 (SI, Figure S8). This model predicts that $S_{C/O}$ is constant. However, $S_{C/O}$ varies 3-4 fold across Form I Rubiscos. (B) Rubiscos within the same physiological grouping - e.g. C3 or C4 plants - have roughly constant $S_{C/O}$ independent of $k_{cat,C}/K_C$ (log scale standard deviations $\sigma^* \leq 0.2$ in all cases). The dashed vertical line is drawn at the median $S_{C/O}$ value for each group, with red algal Rubiscos having the highest measured $S_{C/O}$ values ($S_{C/O} \approx 150$) at 25 °C.*

Discussion

We collected and analyzed roughly 300 literature measurements of Rubisco kinetic parameters (Figure 2A). The collection is quite biased, with the readily-purified Rubiscos of land plants making up ≈80% of the data (Figure 2B). Better sampling of Rubisco diversity including more algal, bacterial and archaeal Rubiscos would greatly improve our understanding of the evolution and capacity of this enzyme (Liu et al., 2017). Despite incomplete coverage, some trends are clear. All Rubisco kinetic parameters display limited dynamic range, with standard deviations in log-scale being less than one order-of-magnitude in all cases (Figure 2C). Rubisco appears particularly constrained in that it displays much less k_{cat} variability than any other enzyme for which sufficient data is available (Figure S4). For many other enzymes, k_{cat} varies over 100-fold or more. In contrast, 97% of Rubisco $k_{cat,C}$ data are between 1 and 10 s^{-1}. The highest Rubisco $k_{cat,C}$ measured at 25 °C is 14.4 s^{-1} (*S. elongatus* PCC 7942, (Occhialini et al., 2016)), only 18-fold greater than the lowest measured Form I $k_{cat,C}$ (0.83 s^{-1} from the diatom *Cylindrotheca* N1,

(Read and Tabita, 1994)). Altogether, these data suggest that there is some limitation on the maximum carboxylation rate by Rubisco in the presence of O_2.

Other extremal Rubiscos of note include an enzyme from the thermophillic red alga *G. sulphuraria*, which has the lowest $K_C \approx 3$ μM. Red algal Rubiscos are generally the most CO_2-specific ($S_{C/O} \approx 160$-200) measured at 25 °C (Uemura et al., 1997; Whitney et al., 2001). Many Rubiscos are quite slow oxygenators with more than half of measurements having $k_{cat,O} < 1$ s^{-1} (Figure 2A). Similarly, many Rubiscos have quite low O_2 affinity: the median K_O is ≈ 470 μM, nearly double the Henry's law equilibrium of water with a 21% O_2 atmosphere at 25 °C. Rubisco from the diatom *Thalassiosira weissflogii*, for example, has a $K_O \approx 2$ mM (Young et al., 2016), corresponding to roughly ten times the ambient O_2 concentration (Sander, 2015).

Specificity $S_{C/O}$ varies the least of all Rubisco parameters (Figure 2C). Nonetheless, Form I Rubiscos are much more CO_2-specific than their Form II, III and II/III counterparts (Figure 7B, SI). This might be explained by the prevalence of Form II, III and II/III enzymes in bacteria and archaea that fix CO_2 in anaerobic conditions where oxygenation should be negligible. We note, however, that there is substantial variation among measurements of the model Form II Rubisco from *R. rubrum*. This and the general paucity of data on non-Form I Rubiscos indicates that more measurements are required to evaluate stereotyped differences within and between Form II, III and II/III Rubiscos. As such, we chose to focus here on Form I Rubiscos for which data is abundant (Figure 2).

Overall, we observed that Rubisco appears less constrained than previously supposed. Rubisco kinetics were previously argued to vary in a one-dimensional landscape (Savir et al., 2010) and hypothesized be "nearly perfectly optimized" (Tcherkez et al., 2006). If these arguments were true, we would expect very limited variation in Rubisco kinetics because all enzymes should attain near-optimal parameters (as diagrammed in Figure 1C). However, as shown in Figure S6, the extended dataset is not strictly one-dimensional. Consistent with this analysis, Figures 3 and S3 document an an overall reduction in correlation between Form I Rubisco kinetic parameters. There are some stereotyped differences between Rubiscos from different kinds of organisms - cyanobacterial Rubiscos are among the fastest (highest $k_{cat,C}$) and red algal Rubiscos are the most CO_2-specific (highest $S_{C/O}$). However, when we plot the assembled data in Figure 4A, there is only modest correlation between $k_{cat,C}$ and $S_{C/O}$. Similarly, Figure 4B shows relatively modest correlation between $k_{cat,C}$ and K_C. Overall weakened correlations led us to reject the notion that Rubisco kinetics vary in a one-dimensional landscape and to investigate whether previously-suggested tradeoff models truly constrain the evolution of Rubisco.

The mechanistic tradeoff models described in Figures 5-6 are based on a simple chemical intuition: that the intrinsic difficulty of binding and discriminating between CO_2 and O_2 requires the enzyme to differentiate between carboxylation and oxygenation transition states. The requirement of TS discrimination is a direct consequence of two common assumptions that are supported by experimental evidence (Pierce et al., 1986). 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 so-called "Michaelis complex" (Andrews and Lorimer, 1987; Pierce et al., 1986; Savir et al., 2010; Tcherkez et al., 2018, 2006). If CO_2 bound a specific site on Rubisco before reacting, K_C could be modulated by mutation without substantially affecting the kinetics of subsequent reaction steps. In the unlikely case that gas addition is substantially reversible, (Cummins et al., 2018; Tcherkez et al., 2018) we would expect to find Rubiscos that evolved enhanced selectivity by energy-coupled kinetic proofreading. Energy coupling would enable amplification of selectivity determined by differential CO_2 and O_2 off-rates (Hopfield, 1974). The fact that no such Rubiscos have been found suggests that gas addition is irreversible or that the off-rates of CO_2 and O_2 are incompatible with kinetic proofreading in some other way (Savir et al., 2010; Savir and Tlusty, 2007).

As Rubisco likely does not bind CO_2 directly, Tcherkez et al. 2006 hypothesize that high specificity towards CO_2 (i.e. large $S_{C/O}$) is realized by discriminating between the first carboxylation and oxygenation transition states (i.e. between the carboxyketone and the peroxyketone, Figure S2). A late carboxylation TS would be maximally discriminable from the oxygenation TS because the developing carboxylic acid is distinguishable from the peroxyl group of the oxygenation intermediate (Tcherkez et al., 2006). Since a late TS resembles the carboxyketone carboxylation intermediate, Tcherkez et al. further argue that CO_2-specific Rubiscos must tightly bind the carboxyketone, which could slow the subsequent hydration and cleavage steps and restrict $k_{cat,C}$ (Figure S2). Though this model is motivated by the need for discrimination between CO_2 and O_2, it implies coupling between the kinetics of carboxylation steps alone. That is: specificity requires tighter binding of the carboxylation intermediate, which slows downstream processing of that same intermediate irrespective of the oxygenation steps.The extraordinarily tight binding of the carboxyketone analog CABP to plant Rubisco provides strong support for the idea of TS discrimination. Negative correlation between $k_{cat,C}$ and $S_{C/O}$ was taken to support the idea of tighter TS binding slowing carboxylation (Tcherkez et al., 2006).

Savir et al. 2010 articulate a related model, noting that $k_{cat,C}$ and $k_{cat,C}/K_C$ are inversely correlated in their dataset (Savir et al., 2010). Since $k_{cat,C}/K_C$ is related to the effective barrier to enolization and CO_2 addition and $k_{cat,C}$ is related to the effective barrier to hydration and cleavage (Figure 1B), Savir et al. argue that lowering the effective barrier to CO_2 addition entails a higher barrier for the subsequent steps (i.e. a lower $k_{cat,C}$, Figure 5A). In both of these descriptions, the initial steps of carboxylation are negatively coupled to the subsequent steps in a manner that produces the observed correlations. However, those correlations - between $S_{C/O}$ and $k_{cat,C}$, K_C and $k_{cat,C}$ and $k_{cat,C}/K_C$ and $k_{cat,C}$ - are attenuated by the addition of new measurements (Figures 3-4) which calls these proposals into question. Importantly, we do not argue that the chemical logic advanced by (Tcherkez et al., 2006) is incorrect, but rather that the assembled data do not support such a tradeoff being optimized over the evolution of Form I Rubiscos.

The second tradeoff model posited by (Savir et al., 2010) is that faster CO_2 addition to the Rubisco-RuBP complex necessarily allows faster O_2 addition. This model can be motivated by the catalytic mechanism of Rubisco as well. Since Rubisco likely does not bind CO_2 or O_2

directly, the concentrations of CO_2 and O_2 in the Rubisco active site should be determined by their solution concentrations alone (e.g. in the chloroplast stroma). Rubisco might limit the active site concentration of O_2 by closing the active site to diffusion, but this would slow CO_2 entry and carboxylation as well. 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 both CO_2 and O_2 (Figure 6A, SI). We showed that the extended dataset strongly supports this power-law relation and suggests lowering the effective CO_2 addition barrier (enabling faster carboxylation) requires a *roughly equal* reduction in the effective barrier to O_2 addition (i.e. enabling faster oxygenation as well). Though several research groups have attempted to isolate improved Rubisco mutants, none of the mutants examined so far exceed the 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 7A). In this model, the average occupancy of the reactive state dictates the rate of CO_2 and O_2 addition and throttles the subsequent steps of carboxylation and oxygenation equally (Figure 7). This model can be mapped onto the Rubisco mechanism by noting that RuBP must be enolized before CO_2 or O_2 can react, suggesting that the occupancy of the reactive state (φ) is related to the degree of enolization of RuBP (SI, Figure S10). One implication of this model is that $S_{C/O}$ is roughly constant (SI). While $S_{C/O}$ does vary over roughly tenfold across the entire dataset and 3-4 fold across Form I enzymes (Figure 1B), Rubiscos from the same physiological groupings display roughly constant $S_{C/O}$ values independent of $k_{cat,C}/K_C$ (Figure 7B). More measurements of bacterial Form I, II and III Rubiscos as well as the notably high-specificity Form ID enzymes of red algae will be crucial to evaluate the generality of this observation.

In previous work, where Rubisco kinetics were thought to vary in a one-dimensional landscape, setting $k_{cat,C}$ determined all other kinetic parameters (Savir et al., 2010). In this setting it was argued that Rubisco kinetic parameters were wholly determined by the prevailing CO_2 and O_2 concentrations since there was a unique choice of parameters on the one-dimensional curve that maximize the net rate of carboxylation (Savir et al., 2010). Since the data is no longer clearly one-dimensional, we cannot argue that Rubisco is “perfectly optimized” to match prevailing concentrations. Moreover, the single surviving tradeoff model does not, on its own, explain why we have not found faster-carboxylating Rubiscos. The model presented in Figures 6 and 7 describes a tradeoff between CO_2 and O_2 addition, but sets no upper limit on $k_{cat,C}$, suggesting that selection for increased carboxylation in the absence of O_2 could produce Rubiscos with superlative $k_{cat,C}$ values (i.e. $k_{cat,C} >> 15\ s^{-1}$). Such an enzyme might be used as a basis for engineering fast-and-selective Rubiscos, something that might indeed be possible if $k_{cat,C}$, K_C and $S_{C/O}$ are not tightly linked to each other (Figure 4).

The prospect of engineering an improved Rubiscos is tantalizing not only because it could plausibly increase crop yields substantially (Zhu et al., 2010), but also because the task tests our understanding of proteins and enzymes on a very basic level. It is clear from the data presented here that there is some evolutionary constraint on Rubisco catalysis. Indeed, no

known Rubisco has a $k_{cat,C}$ greater than 15 s^{-1} and no measured $S_{C/O}$ exceeds 250. Surely a superlative Rubisco would have arisen if it was mutationally accessible from existing Rubiscos. However, the large subunit of Rubisco displays extremely limited sequence variation (Kapralov and Filatov, 2007). Perhaps exploring a wider swath of sequence space via protein engineering techniques (Chin, 2014; Fowler and Fields, 2014; Silberg et al., 2004) would enable strict improvements to Rubisco kinetics? In order to better-resolve the evolutionary constraints imposed on Rubisco kinetics and evaluate the prospects of advanced Rubisco engineering, we suggest several avenues for future research.

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 Form I enzymes of cyanobacteria and diverse Eukaryotic autotrophs (Liu et al., 2017). Ideally these enzymes would be sampled from accumulated genomic data in a manner that maximizes sequence and phylogenetic diversity (Akiva et al., 2017) 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 (Orr et al., 2016; Sharwood et al., 2016). These data would likely resolve whether Rubisco isoforms display characteristic differences in catalytic potential. It is possible, for example, that Form II, III or II/III enzymes are subject to different constraints than Form I Rubiscos and might serve as useful chassis for bioengineering.

Furthermore, it is important to revisit the classic experiments undergirding our understanding of the Rubisco catalytic mechanism, especially those supporting the central assumptions that (a) there is no Michaelis complex for CO_2 or O_2 and (b) that gas addition is irreversible (Cummins et al., 2018; Pierce et al., 1986; Tcherkez et al., 2018). As mentioned above, these assumptions imply substantial limitation on CO_2 specificity by, for example, disallowing a kinetic proofreading based mechanism for the amplification of specificity. If we were to find Rubiscos for which these assumptions are relaxed, they might be used as a basis for future engineering of a fast-and-selective carboxylase. On the other hand, it may be the case that all Rubiscos share these same limitations and are constrained by the same tradeoffs. Since tradeoffs in Rubisco catalysis are likely described by couplings between transition state barriers (e.g. as in Figure 6) it would be very useful to measure TS barrier heights for many variants. One avenue for further investigation would be measurement of carbon and oxygen kinetic isotope effects (KIEs) for a wide variety of Rubiscos. Kinetic isotope effects report indirectly on TS barrier heights (Hayes, 2001; McNevin et al., 2007) and KIEs could plausibly be measured in relatively high throughput via mass spectrometry. Investigating the relationship between kinetic isotope effects and kinetic parameters will hopefully refine our understanding of the Rubisco mechanism and help clarify whether different families of Rubisco enzymes are subject to the same constraints (Tcherkez et al., 2006).

There remains some disagreement about the precise ordering of Rubisco carboxylation steps (Andersson, 2008; Cleland et al., 1998; Tcherkez et al., 2006) and the mechanism of oxygenation is not well understood (Tcherkez, 2016). Chemical reasoning about the mechanisms of Rubisco carboxylation and oxygenation would benefit from progress in structural biology - intermediates and transition state analogs should be used to capture the active site at

various points along the reaction trajectory (Andersson and Backlund, 2008; Cleland et al., 1998; Schneider et al., 1992; Stec, 2012; Tcherkez, 2016). 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 (Bar-Even, 2017; McGrath and Long, 2014; South et al., 2019). If, however, these assumptions are invalidated, many enzyme engineering strategies would be viable. Such data and analyses will be instrumental in guiding the engineering of carbon fixation for the next decade.

Methods

Data collection and curation. We reviewed the literature to find Rubisco kinetic data measured at 25 ^{o}C and near pH 8. Ultimately 61 primary literature studies were included, yielding 334 $S_{C/O}$, 282 $k_{cat,C}$, 316 K_C, and 254 K_o values for Rubiscos from 304 distinct organisms (Datasets S1 and S2). We also recorded 52 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, but is rather inferred as $k_{cat,O} = (k_{cat,C}/K_C) / (S_{C/O}/K_O)$. When an uncertainty is reported, we assumed that the underlying experimental noise is normally distributed and used 10^4-fold bootstrapping to estimate 198 $k_{cat,O}$ 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. As certain model Rubiscos are measured frequently (e.g. we found 12 independent measurements of the Rubisco from spinach), multiple measurements are available for some Rubiscos. In these cases we used the median measured value in correlation and regression analyses to avoid bias. 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. 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. 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 to determine a point estimate of R, the power-law exponent and prefactor. This procedure was repeated 10^4 times to determine a 95% confidence intervals on the above parameters. Python source code is available at github.com/flamholz/rubisco.

Acknowledgements

We would like to thank Uri Alon, Kapil Amarnath, Doug Banda, Arren Bar-Even, Cissi 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 Department of Energy Grant (DE-SC00016240); 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

Akiva E, Copp JN, Tokuriki N, Babbitt PC. 2017. Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc Natl Acad Sci U S A* **114**:E9549–E9558.

Andersson I. 2008. Catalysis and regulation in Rubisco. *J Exp Bot* **59**:1555–1568.

Andersson I, Backlund A. 2008. Structure and function of Rubisco. *Plant Physiol Biochem* **46**:275–291.

Andrews TJ, Lorimer GH. 1987. 3 - Rubisco: Structure, Mechanisms, and Prospects for Improvement In: Hatch, Boardman NK, editors. Photosynthesis. Academic Press. pp. 131–218.

Bainbridge G, Madgwick P, Parmar S, Mitchell R, Paul M, Pitts J, Keys AJ, Parry MAJ. 1995. Engineering Rubisco to change its catalytic properties. *J Exp Bot* **46**:1269–1276.

Bar-Even A. 2017. Daring metabolic designs for enhanced plant carbon fixation. *Plant Sci*. doi:10.1016/j.plantsci.2017.12.007

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*. doi:10.1021/bi2002289

Bar-On YM, Phillips R, Milo R. 2018. The biomass distribution on Earth. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1711842115

Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. 2018. Rubisco is not really so bad. *Plant Cell Environ* **41**:705–716.

Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: players, partners and origin. *Trends Plant Sci* **15**:330–336.

Busch FA, Sage RF, Farquhar GD. 2017. Plants increase CO2 uptake by assimilating nitrogen via the photorespiratory pathway. *Nat Plants*. doi:10.1038/s41477-017-0065-x

Caemmerer SV, Evans JR. 1991. Determination of the Average Partial Pressure of CO2 in Chloroplasts From Leaves of Several C3 Plants. *Funct Plant Biol* **18**:287–305.

Chin JW. 2014. Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem* **83**:379–408.

Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, Lorimer GH. 1998. Mechanism of Rubisco: The Carbamate as General Base. *Chem Rev* **98**:549–562.

Cummins PL, Kannappan B, Gready JE. 2018. Directions for Optimization of Photosynthetic

728 Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All. *Front Plant*
729 *Sci* **9**:183.
730 Evans JR, Sharkey TD, Berry JA, Farquhar GD. 1986. Carbon Isotope Discrimination measured
731 Concurrently with Gas Exchange to Investigate CO2 Diffusion in Leaves of Higher Plants.
732 *Funct Plant Biol* **13**:281–292.
733 Flamholz A, Noor E, Bar-Even A, Milo R. 2012. eQuilibrator--the biochemical thermodynamics
734 calculator. *Nucleic Acids Res* **40**:D770–5.
735 Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nat*
736 *Methods* **11**:801–807.
737 Galmés J, Kapralov MV, Andralojc PJ, Conesa MÀ, Keys AJ, Parry MAJ, Flexas J. 2014.
738 Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and
739 evolutionary trends. *Plant Cell Environ* **37**:1989–2001.
740 Hayes JM. 2001. Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes.
741 *Rev Mineral Geochem* **43**:225–277.
742 Hopfield JJ. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic
743 processes requiring high specificity. *Proc Natl Acad Sci U S A* **71**:4135–4139.
744 Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. 2018. Lateral gene transfer shapes the
745 distribution of RuBisCO among Candidate Phyla Radiation bacteria and DPANN archaea.
746 *Mol Biol Evol*. doi:10.1093/molbev/msy234
747 Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose biphosphate
748 carboxylase/oxygenase. *Nature* **291**:513.
749 Kapralov MV, Filatov DA. 2007. Widespread positive selection in the photosynthetic Rubisco
750 enzyme. *BMC Evol Biol* **7**:73.
751 Liu D, Ramya RCS, Mueller-Cajar O. 2017. Surveying the expanding prokaryotic Rubisco
752 multiverse. *FEMS Microbiol Lett* **364**. doi:10.1093/femsle/fnx156
753 Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF. 2016. pH determines the energetic
754 efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc Natl Acad Sci U S A*
755 **113**:E5354–62.
756 McGrath JM, Long SP. 2014. Can the cyanobacterial carbon-concentrating mechanism increase
757 photosynthesis in crop species? A theoretical analysis. *Plant Physiol* **164**:2247–2261.
758 McNevin DB, Badger MR, Whitney SM, von Caemmerer S, Tcherkez GGB, Farquhar GD. 2007.
759 Differences in Carbon Isotope Discrimination of Three Variants of D-Ribulose-1,5-
760 bisphosphate Carboxylase/Oxygenase Reflect Differences in Their Catalytic Mechanisms. *J*
761 *Biol Chem* **282**:36068–36076.
762 Occhialini A, Lin MT, Andralojc PJ, Hanson MR, Parry MAJ. 2016. Transgenic tobacco plants
763 with improved cyanobacterial Rubisco expression but no extra assembly factors grow at
764 near wild-type rates if provided with elevated CO2. *Plant J* **85**:148–160.
765 Orr DJ, Alcântara A, Kapralov MV, Andralojc PJ, Carmo-Silva E, Parry MAJ. 2016. Surveying
766 Rubisco Diversity and Temperature Response to Improve Crop Photosynthetic Efficiency.
767 *Plant Physiol* **172**:707–717.
768 Parry MAJ, Keys AJ, Gutteridge S. 1989. Variation in the Specificity Factor of C3 Higher Plant
769 Rubiscos Determined by the Total Consumption of Ribulose-P2. *J Exp Bot* **40**:317–320.
770 Pierce J, Lorimer GH, Reddy GS. 1986. Kinetic mechanism of ribulosebisphosphate
771 carboxylase: evidence for an ordered, sequential reaction. *Biochemistry* **25**:1636–1644.
772 Raven JA. 2013. Rubisco: still the most abundant protein of Earth? *New Phytol* **198**:1–3.
773 Raven JA, Beardall J, Sánchez-Baracaldo P. 2017. The possible evolution, and future, of CO2-
774 concentrating mechanisms. *J Exp Bot*. doi:10.1093/jxb/erx110
775 Read BA, Tabita FR. 1994. High substrate specificity factor ribulose bisphosphate
776 carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant
777 cyanobacterial RubiSCO containing "algal" residue modifications. *Arch Biochem Biophys*
778 **312**:210–218.

779 Reinhold L, Kosloff R, Kaplan A. 1991. A model for inorganic carbon fluxes and photosynthesis
780 in cyanobacterial carboxysomes. *Can J Bot* **69**:984–988.
781 Sander R. 2015. Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmos*
782 *Chem Phys* **15**:4399–4981.
783 Savir Y, Noor E, Milo R, Tlusty T. 2010. Cross-species analysis traces adaptation of Rubisco
784 toward optimality in a low-dimensional landscape. *Proc Natl Acad Sci U S A* **107**:3475–
785 3480.
786 Savir Y, Tlusty T. 2007. Conformational proofreading: the impact of conformational changes on
787 the specificity of molecular recognition. *PLoS One* **2**:e468.
788 Schneider G, Lindqvist Y, Brändén CI. 1992. RUBISCO: structure and mechanism. *Annu Rev*
789 *Biophys Biomol Struct* **21**:119–143.
790 Sharwood RE, Ghannoum O, Kapralov MV, Gunn LH, Whitney SM. 2016. Temperature
791 responses of Rubisco from Paniceae grasses provide opportunities for improving C3
792 photosynthesis. *Nat Plants* **2**:16186.
793 Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA. 2016. Biochemical
794 characterization of predicted Precambrian RuBisCO. *Nat Commun* **7**:10382.
795 Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U. 2012.
796 Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science*
797 **336**:1157–1160.
798 Silberg JJ, Endelman JB, Arnold FH. 2004. SCHEMA-guided protein recombination. *Methods*
799 *Enzymol* **388**:35–42.
800 South PF, Cavanagh AP, Liu HW, Ort DR. 2019. Synthetic glycolate metabolism pathways
801 stimulate crop growth and productivity in the field. *Science* **363**:eaat9077.
802 Spreitzer RJ, Salvucci ME. 2002. Rubisco: structure, regulatory interactions, and possibilities for
803 a better enzyme. *Annu Rev Plant Biol* **53**:449–475.
804 Stec B. 2012. Structural mechanism of RuBisCO activation by carbamylation of the active site
805 lysine. *Proceedings of the National Academy of Sciences* **109**.
806 doi:10.1073/pnas.1210754109
807 Tcherkez G. 2016. The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ*
808 **39**:983–997.
809 Tcherkez GG, Bathellier C, Farquhar GD, Lorimer GH. 2018. Commentary: Directions for
810 Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So
811 Constrained After All. *Front Plant Sci* **9**:929.
812 Tcherkez GGB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused
813 substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly
814 optimized. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.0600605103
815 Uemura K, Anwaruzzaman, Miyachi S, Yokota A. 1997. Ribulose-1,5-bisphosphate
816 carboxylase/oxygenase from thermophilic red algae with a strong specificity for CO2
817 fixation. *Biochem Biophys Res Commun* **233**:568–571.
818 Whitney SM, Baldet P, Hudson GS, Andrews TJ. 2001. Form I Rubiscos from non-green algae
819 are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J* **26**:535–547.
820 Wildman SG. 2002. Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate
821 carboxylase-oxygenase). *Photosynth Res* **73**:243–250.
822 Young JN, Heureux AMC, Sharwood RE, Rickaby REM, Morel FMM, Whitney SM. 2016. Large
823 variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-
824 concentrating mechanisms. *J Exp Bot* **67**:3445–3456.
825 Zhu X-G, Long SP, Ort DR. 2010. Improving photosynthetic efficiency for greater yield. *Annu*
826 *Rev Plant Biol* **61**:235–261.

827