frontiers in Plant Science

HYPOTHESIS AND THEORY
published: 01 March 2018
doi: 10.3389/fpls.2018.00183

Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All

*Peter L. Cummins, Babu Kannappan and Jill E. Gready**

Gready Group, Department of Genome Science, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia

The ubiquitous enzyme Ribulose 1,5-bisphosphate carboxylase-oxygenase (RuBisCO) fixes atmospheric carbon dioxide within the Calvin-Benson cycle that is utilized by most photosynthetic organisms. Despite this central role, RuBisCO's efficiency surprisingly struggles, with both a very slow turnover rate to products and also impaired substrate specificity, features that have long been an enigma as it would be assumed that its efficiency was under strong evolutionary pressure. RuBisCO's substrate specificity is compromised as it catalyzes a side-fixation reaction with atmospheric oxygen; empirical kinetic results show a trend to tradeoff between relative specificity and low catalytic turnover rate. Although the dominant hypothesis has been that the active-site chemistry constrains the enzyme's evolution, a more recent study on RuBisCO stability and adaptability has implicated competing selection pressures. Elucidating these constraints is crucial for directing future research on improving photosynthesis, as the current literature casts doubt on the potential effectiveness of site-directed mutagenesis to improve RuBisCO's efficiency. Here we use regression analysis to quantify the relationships between kinetic parameters obtained from empirical data sets spanning a wide evolutionary range of RuBisCOs. Most significantly we found that the rate constant for dissociation of CO_2 from the enzyme complex was much higher than previous estimates and comparable with the corresponding catalytic rate constant. Observed trends between relative specificity and turnover rate can be expressed as the product of negative and positive correlation factors. This provides an explanation in simple kinetic terms of both the natural variation of relative specificity as well as that obtained by reported site-directed mutagenesis results. We demonstrate that the kinetic behaviour shows a lesser rather than more constrained RuBisCO, consistent with growing empirical evidence of higher variability in relative specificity. In summary our analysis supports an explanation for the origin of the tradeoff between specificity and turnover as due to competition between protein stability and activity, rather than constraints between rate constants imposed by the underlying chemistry. Our analysis suggests that simultaneous improvement in both specificity and turnover rate of RuBisCO is possible.

Keywords: RuBisCO, carbon fixation, photosynthesis, enzyme kinetics and specificity, protein evolution, evolutionary constraints, enzyme-complex stability, gas-substrate binding

OPEN ACCESS

Edited by:
Hartmut Stützel, Leibniz University of Hanover, Germany

Reviewed by:
Grant Pearce, University of Canterbury, New Zealand
Qiang Wang, Institute of Hydrobiology (CAS), China

****Correspondence:***
Jill E. Gready
jill.gready@anu.edu.au

Specialty section:
This article was submitted to Plant Biophysics and Modeling, a section of the journal Frontiers in Plant Science

Received: *30 August 2017*
Accepted: *31 January 2018*
Published: *01 March 2018*

Citation:
Cummins PL, Kannappan B and Gready JE (2018) Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All. Front. Plant Sci. 9:183. doi: 10.3389/fpls.2018.00183

INTRODUCTION

Ribulose 1,5-bisphosphate carboxylase-oxygenase (RuBisCO) is the enzyme responsible for the fixation of carbon derived from atmospheric CO_2 as part of the Calvin-Benson cycle that leads to production of the glucose essential for growth in most photosynthetic organisms. However, RuBisCO has a low turnover rate in higher plants ($\sim$3 s^{-1}) and the efficiency of carbon fixation by the enzyme is compromised by a competing reaction with atmospheric O_2 that leads to photorespiration at high cost to the organism in terms of both energy and loss of carbon. A recent analysis of k_{cat} and K_M values of several thousand enzymes (Bar-Even et al., 2011) has shown that RuBisCO's catalytic rate, k_{cat}, and efficiency (k_{cat}/K_M) are not unusually low compared with values of the "average" enzyme (see their **Figure 1**), even though much lower than fast enzymes at the diffusion-controlled limit, for a variety of reasons including absence of strong evolutionary selection pressure and substrate properties, especially low molecular mass and hydrophobicity, limiting K_M optimization. A later analysis (Bar-Even et al., 2015) showed that enzyme-substate encounters for the "average" enzyme are not productive("futile"), again for various reasons. The insights from these analyses are useful in placing RuBisCO's catalytic rate and efficiency in the context of all enzymes, especially the significant dissociation rate for CO_2 we find in this work, but nonetheless puzzles remain as RuBisCO has been subject to very strong evolutionary pressure.

To mitigate this apparent torpidity of the enzyme, organisms have co-evolved other strategies for maintaining levels of photosynthesis. The observed large variations in RuBisCO kinetic parameters from photosynthetic organisms in different kingdoms down to different species (Jordan and Ogren, 1981, 1983) is a consequence of co-evolution with resource allocation into other strategies that lead to enhanced photosynthesis (largely by way of more efficient CO_2 and nitrogen utilization) and suppressed photorespiration (Badger and Andrews, 1987; Badger et al., 1998).

Cyanobacterial RuBisCOs are characterized by lower values of activity with CO_2 relative to that of O_2 (the relative specificity, $S_{C/O}$) and higher catalytic turnover rates (k^C_{cat}). These organisms utilize a carbon-concentrating mechanism (CCM) which compensates for the lower $S_{C/O}$ and limits photorespiration by increasing the CO_2/O_2 ratio at the site of fixation, while taking advantage of the higher k^C_{cat} by reducing RuBisCO concentration and hence the requirement for nitrogen. Some non-green algae with higher $S_{C/O}$ do not express a CCM but instead the lower k^C_{cat} is mitigated by increasing RuBisCO and, hence, higher investment of nitrogen in RuBisCO protein. In higher plants, the kinetic balances and photosynthetic pathways lie somewhere in the middle of these two extremes. In C_3 plants $S_{C/O}$ is generally greater and k^C_{cat} less than in C_4 plants expressing CCMs (Yeoh et al., 1980; Seemann et al., 1984; Ghannoum et al., 2005), while others are characterized as C_3-C_4 intermediate or C_4-like (Kubien et al., 2008).

Understanding the nature of constraints imposed on RuBisCO's intrinsic efficiency is important for directing future research on photosynthesis. Study of RuBisCO activity has become a focus for improving photosynthesis (Bainbridge et al., 1995; Peterhansel et al., 2008; Gready and Kannappan, 2009; Whitney et al., 2011; Parry et al., 2013; Carmo-Silva et al., 2015) with a major aim of improving crop yields. However, some doubt has been cast on whether it can be significantly improved via mutation because of a hypothesis of "underlying constraints" in the chemistry of the reaction (Tcherkez et al., 2006; Savir et al., 2010; Tcherkez, 2013).

In the present study, we argue that this conclusion may have resulted from unsupported assumptions of the kinetic models and limited data sets used in the analyses. Resolving the precise nature of the constraints imposed on RuBisCO kinetics is clearly pivotal to providing direction of future research into improving photosynthesis. The rate constants (**Figure 1**) determine, and therefore ultimately limit, the physical binding of substrates, the breaking and formation of chemical bonds, and finally the release of products (Lorimer, 1981; Cleland et al., 1998; Andersson, 2008; Kannappan and Gready, 2008).

Although methods for computing individual rate constants from kinetic data have not been widely implemented for RuBisCO (McNevin et al., 2006), the more commonly measured kinetic parameters (k^C_{cat}, k^O_{cat}, K_C, K_O, and $S_{C/O} = \frac{k^C_{cat} K_O}{k^O_{cat} K_C}$), *in vitro*, are generally functions of these. Here we derive the equations for the kinetic mechanism (**Figure 1**) and estimate the mean (or expected) values for rate constants using regression analysis. Utilizing the compilation in **Table 1**, which includes the data used by Savir et al. in their analysis (Savir et al., 2010), we performed our own linear regression analysis on a wider range of data sets. This analysis was extended to other plant data (Galmés et al., 2014; Prins et al., 2016) to assist in validating the results. We found that the rate constants for dissociation of the CO_2 and O_2 substrates (k_6 and k_{12} in **Figure 1**) are much larger relative to the corresponding catalytic rate than previously assumed and consequently have a significant effect on the kinetics. We also suggest the constraints on RuBisCO may be better explained by competing selection pressures, rather than by positive selection within hypothetical constraints (Tcherkez et al., 2006; Tcherkez, 2013) imposed by the chemical mechanism.

Our results and conclusions are indicative of a less constrained RuBisCO and are consistent with observed variations in the kinetics of a wider range of wild type and mutant RuBisCO that are now available, although such kinetic data is regrettably still sparse.

METHODS

We consider the rate constants k_i for the kinetic mechanism (**Figure 1**) to be a set of general random variables (Koralov and Sinai, 2007). The expected value, $E(k_i) \equiv \langle k_i \rangle$, is the mean value of k_i, i.e., averaged over a number of sequences. In principle these averages can be extracted using both linear and non-linear regression methods to establish functional relationships between the RuBisCO kinetic parameters. As K_C and K_O depend explicitly on k^C_{cat} and k^O_{cat}, respectively, we restrict the independent variables (predictors) to k^C_{cat} and k^O_{cat}. The dependent (response) variables whose expected values, conditional on k^C_{cat} or k^O_{cat},

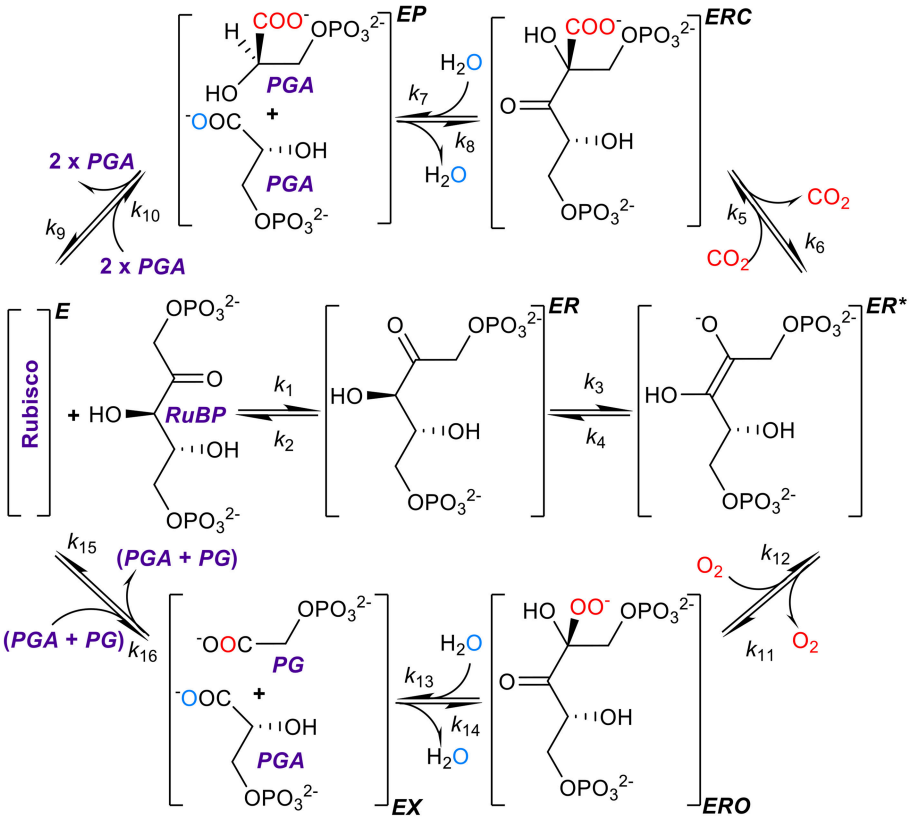

FIGURE 1 | The kinetic mechanism of RuBisCO. RuBisCO must first be activated by carbamylation and binding of Mg^{2+} before it processes three substrates, ribulose bisphosphate (RuBP), and carbon dioxide or oxygen, the complete reactions taking place over several stages (Lorimer, 1981; Cleland et al., 1998; Andersson, 2008; Kannappan and Gready, 2008). RuBP binds first forming a complex (*ER*) with the activated form of the enzyme (*E*), followed by enolization of RuBP (*ER**) which facilitates binding with the carbon dioxide or oxygen molecule to form the *ERC* or *ERO* enzyme-substrate complexes. After hydrolysis, the six-carbon compound formed by the addition of carbon dioxide to RuBP breaks at a C-C bond forming a product complex (*EP*) which dissociates into two three-carbon compounds, 3-phosphoglyceric acid (*PGA*), with the addition of two protons. Oxygenation proceeds through analogous steps except that the dissociation products are one *PGA* molecule and one of 2-phospho-glycolate (*PG*). Atoms originating from free CO_2 and O_2 are shown in red, and oxygen atom originating from the water molecule used for hydration is shown in aqua blue.

are determined by regression are then K_C, K_O and $S_{C/O}$, e.g., $E\left(K_C \middle| k_{cat}^C\right) \equiv \langle K_C \rangle$.

The Generalized Extreme Studentized Deviate (ESD) test (Rosner, 1983) was used with *P*-value of 0.05 to eliminate multiple outliers in the data prior to regression analysis. The regression parameters were then used to estimate the expected values of various terms in the kinetic equations. We can illustrate the procedure by considering a more simplistic single-intermediate kinetic mechanism where the Michaelis constant is given by $K_M = \frac{(k_{cat}+k_{off})}{k_{on}}$ (e.g, Roberts, 1977; Farquhar, 1979). Enzyme assays typically provide K_M and k_{cat} but insufficient data to determine k_{on} and k_{off} which are, respectively, the rate constants for the binding and dissociation of substrate (e.g., CO_2 or O_2). However, if we consider that the rate constants k_{cat}, k_{on} and k_{off} randomly fluctuate over a number of sequences, a linear correlation, $\langle K_M \rangle$, may be obtained between K_M and k_{cat} from which the gradient and intercept give the expected values $\left\langle \frac{1}{k_{on}} \right\rangle$ and $\left\langle \frac{k_{off}}{k_{on}} \right\rangle$, respectively, and using the approximation $\langle xy \rangle \approx \langle x \rangle \langle y \rangle$ for a finite number of random variables x and y, we can hence determine the expected values of the rate constants $\langle k_{on} \rangle$ and $\langle k_{off} \rangle$. Although K_M is linearly dependent on k_{cat}, we should not necessarily expect to observe any correlation, as high variances may be associated with the other two terms, k_{on} and k_{off}. Where a linear correlation exists, we may infer that the rate constants k_{on} and k_{off} are fairly constant (low variance), while a non-linear correlation would be consistent with an additional correlation between k_{cat} and at least one of these other two terms. Statistical (regression) methods are here used to show how these different scenarios are represented in the available kinetic data.

RESULTS

Kinetic Equations

In deriving the following kinetic equations for this mechanism (**Figure 1**) we assumed only that both k_{10} and k_{16} are very much smaller than any of the remaining rate constants (effectively, $k_{10} = k_{16} = 0$). We emphasize that no such approximations ($k_i = 0$) were made anywhere else in the derivation. The Michaelis constants (K_M) for carboxylation and oxygenation are

TABLE 1 | RuBisCO kinetic parameters.

Species	Ref.	k_{cat}^{C} (s^{-1})	k_{cat}^{O} (s^{-1})	$S_{C/O}$(mol/mol)	K_O (μM)	K_C (μM)
Higher plant C_3 (*Triticum aestivum*)	a	2.5	1.45	90	730	14
Higher plant C_4-like (*Flareria brownie*)	b	2.58	0.91	83.8	378	12.8
Higher plant C_3-C_4 (*Flaveria sonorensis*)	b	2.69	2.46	84.3	785	10.2
Higher plant C_3-C_4 (*Flaveria ramosissima*)	b	2.77	2.09	79.8	722	12.0
Higher plant C_3-C_4 (*Flaveria angustifolia*)	b	2.86		83.2		13.1
Higher plant C_3 (*Chenopodium alba*)	a	2.91	1.37	78.7	415	11.2
Higher plant C_3 (*Flaveria pringlei*)	a	3.1	2.14	80.8	666	12.0
Higher plant C_4 (*Paspalum dilatatum*)	c	3.11	0.74	88	415	19.9
Higher plant C_3 (*Flaveria cronquistii*)	b	3.13	2.34	81	653	10.8
Higher plant C_3-C_4 (*Flaveria floridana*)	b	3.19	1.96	84.5	686	13.2
Higher plant C_3 (*Spinacia oleracea*)	b	3.20	1.90	79.8	574	12.1
Higher plant C_3-C_4 (*Flaveria chloraefolia*)	b	3.35	2.45	81.6	740	12.4
Higher plant C_3 (*Nicotiana tabacum*)	a	3.4	1.11	82	295	10.7
Higher plant C_4 (*Cynodon dactylon*)	c	3.41	0.73	89	402	21
Higher plant C_3-C_4 (*Flaveria linearis*)	b	3.43	1.46	78.1	415	12.5
Higher plant C_4-like (*Flaveria palmeri*)	b	3.54	0.60	83.8	193	13.5
Higher plant C_4 (*Flaveria kochiana*)	b	3.68	0.32	77	150	22.7
Higher plant C_3 (*Spinacia oleracea*)	a	3.7	1.59	80	480	14
Higher plant C_4-like (*Flaveria vaginata*)	b	3.78	1.98	78.7	880	21.4
Higher plant C_4 (*Zoysia japonica*)	c	3.78	0.98	84.1	403	18.5
Higher plant C_4 (*Amaranthus hybridus*)	a	3.8	1.85	82	640	16
Higher plant C_4 (*Flaveria australasica*)	a	3.84	0.70	77.2	309	22.0
Higher plant C_4 (*Zea mays*)	d	4.05	0.32	74.9	157	26.2
Higher plant C_4 (*Amaranthus edulis*)	a	4.14	0.85	77.5	289	18.2
Higher plant C_4 (*Flaveria bidentis*)	b	4.16	1.74	75.5	639	20.2
Higher plant C_4 (*Zea mays*)	a	4.4	1.34	78	810	34
Higher plant C_4 (*Flaveria trinervia*)	b	4.42	2.15	77	671	17.9
Higher plant C_4 (*Sorghum bicolor*)	a	5.4		70		30
Higher plant C_4 (*Zea mays*)	d	5.5	1.31	88	397	19
Higher plant C_4 (*Potulaca oleraca*)	a	5.9		78		13.6
Green algae (*Chlamydomonas reinhardtii*)	a	5.8	1.57	61	480	29
Cyanobacteria (*Synechococcus* 6301)	a	11.6	0.77	43	972	340
Cyanobacteria (*Synechococcus* 7002)	a	13.4	1.36	52	1300	246
Nongreen algae (*Cylindrotheca* sp. N1)	e	0.78		106	1292	31
Nongreen algae (*Olisthodiscus luteus*	e	0.83		101	692	59
Nongreen algae (*Galdieria sulfuraria*)	a	1.2	0.82	166	374	3.3
Nongreen algae (*Cyanidium caldarium*	e	1.3		224		6.7
Nongreen algae *Porphyridium cruentum*	e	1.6		129	1574	22
Nongreen algae *Cyanidium partita*	e	1.6		238		6.6
Nongreen algae *Cylindrotheca fusiformis*	e	1.95		110	568	36
Nongreen algae (*Griffithsia monilis*)	a	2.6		167		9.3
Nongreen algae (*Phaeodactylum tricornutum*)	a	3.4	0.50	113	467	28
Diatom (*Bellerochea cf. horologicalis*)	d	2.1			764	50
Diatom (*Thalassiosira oceania*)	d	2.4	0.44	80	954	65
Diatom (*Chaetoceros muelleri*)	d	2.4	0.46	96	425	23
Diatom (*Chaetoceros calcitrans*)	d	2.6	0.75	57	413	25
Diatom (*Phaeodactylum tricornutum*)	d	3.2	0.49	108	592	36
Diatom (*Skeletonema marinoi*)	d	3.2			883	68
Diatom (*Thalassiosira weissflogii*)	d	3.2	1.27	79	2032	65
Diatom (*Phaeodactylum tricornutum*)	d	3.3	0.46	116	664	41
Diatom (*Chaetoceros calcitrans*)	d	3.4	0.72	75	490	31
Diatom (*Fragilariopsis cylindrus*)	d	3.5	0.47	77	667	64
Diatom (*Cylindrotheca fusiformis*)	d	3.7		79		
Bacteria (*Chromatium vinosum*)	a	6.7	1.28	41	290	37
Bacteria (*Rhodospirillum rubrum*)	a	7.3	3.01	12.3	406	80

Data compiled by Savir et al. (2010) are highlighted in green.

a Savir et al. (2010); b Kubien et al. (2008); c Carmo-Silva et al. (2010); d Young et al. (2016); e Badger et al. (1998).

then given, respectively, by equations of the form (Equations A23, A24; see Appendix in Supplementary Materials for details of derivations)

$$K_C = \frac{\left(k_{cat}^C + \gamma_C k_6\right)}{K_R k_5} \quad (1)$$

$$K_O = \frac{\left(k_{cat}^O + \gamma_O k_{12}\right)}{K_R k_{11}} \quad (2)$$

The general equation for the specificity of carboxylation relative to that of oxygenation (relative specificity) is then (Equation A25).

$$S_{C/O} = \frac{S_C}{S_O} = \frac{k_{cat}^C K_O}{K_C k_{cat}^O} = \frac{k_5 k_{cat}^C (k_{cat}^O + \gamma_O k_{12})}{k_{11} k_{cat}^O (k_{cat}^C + \gamma_C k_6)} \quad (3)$$

In Equation (3), the relative specificity ($S_{C/O}$) is formally a function of 10 rate constants $\left(k_5..k_9, k_{11}..k_{15}\right)$, five for each of the carboxylation and oxygenation reactions. $K_R = \frac{k_3}{(k_3+k_4)}$ is a function only of rate constants for the enolization step (Equation A22), i.e., independent of carboxylation or oxygenation, and $0 < \gamma < 1$. Both k_{cat}^C and γ_C are formally functions of k_3, k_7, k_8 and k_9 (Equation A26). It is evident (Equation A26) that if k_7 is the slow step that determines the maximum catalytic rate ($k_{cat}^C = k_7$), then $\gamma_C = 1$. Similarly (Equation A27), if $k_{cat}^O = k_{13}$, then $\gamma_O = 1$. However, we need not make these types of assumptions here, and simply regard $\gamma_C k_6$ and $\gamma_O k_{12}$ as effective dissociation rate constants.

Michaelis Constants

The results of linear regression analysis performed on a number of data sets are summarized in **Table 2**. The green algae, bacteria and cyanobacteria data in **Table 1** and other plant species (Galmés et al., 2014) could not be considered individually for analysis due to the small numbers of observations ($N < 3$). The log-scale plots (**Figures 2A,B**) of K_C over the full range of k_{cat}^C values in **Table 1** suggest a linear correlation and hence regression analysis of $\ln(K_C)$ on k_{cat}^C ("All data" sets in **Table 2**). $P < 0.05$ for both coefficients were obtained only for carboxylation using the "All data" sets (**Figures 2A,B**), carboxylation using a subset of the C_3 plants (Galmés et al., 2014), oxygenation using Triticeae data (Prins et al., 2016) and oxygenation using only the higher plant data (**Figure 2C**). The residuals were found to be near-normally distributed (**Figure 3**). Reliable expected values for effective CO_2 and O_2 dissociation rate constants can be derived from the coefficients in regressions (**Table 2**) that yield $P < 0.05$ for both coefficients (i.e., both the gradient and intercept). The results are given in **Table 3**. For the regression of $\ln(K_C)$ on k_{cat}^C, equating the first terms ($a_1 + a_1 b_1 k_{cat}^C$) in the expansion of the exponential form ($a_1 e^{b_1 k_{cat}^C}$) with Equation (1) we find that the value of K_C at $k_{cat}^C = 0$ is given by $\left\langle \frac{\gamma_C k_6}{K_R k_5} \right\rangle = a_1$. From the regression analysis carried out using the full data set in **Table 1** (**Figure 2A**) and the subset utilized by Savir et al. (2010) (**Figure 2B**), we obtain values of $a_1 = 9.7$ μM and $a_1 = 4.5$ μM, respectively. From the expansion of the exponential we also find that $\left\langle \frac{1}{K_R k_5} \right\rangle \approx a_1 b_1$ at $k_{cat}^C = 0$, where the two estimates are $a_1 b_1 = 2.2$ μM.s and $a_1 b_1 = 1.5$ μM.s, respectively. In **Figures 2A,B**, Equation (1), which will obviously deviate from the trend line as k_{cat}^C increases, has been graphed using these values. Combining these results obtained for $\left\langle \frac{\gamma_C k_6}{K_R k_5} \right\rangle$ and $\left\langle \frac{1}{K_R k_5} \right\rangle$ we estimate (at $k_{cat}^C = 0$) expected effective rate constants for CO_2 dissociation ($\langle \gamma_C k_6 \rangle$) of 4.3 s^{-1} and 3.0 s^{-1}, respectively. Assuming the scheme (**Figure 1**) correctly describes the kinetic mechanism, the deviation from linear behavior suggests there exists at least one type of correlation between rate constants. From Equation (1), the expected value of $K_R k_5$ conditional on k_{cat}^C in terms of regression parameters a_1 and b_1 is then given by (**Figure 4A**).

$$\langle K_R k_5 \rangle = \frac{\left(k_{cat}^C + \langle \gamma_C k_6 \rangle\right)}{a_1 e^{b_1 k_{cat}^C}}. \quad (4)$$

Therefore, we may also use Equation (4) to define the expected effective dissociation constant conditional on k_{cat}^C as (**Figure 4B**).

$$\langle K_D^C \rangle = \frac{\langle \gamma_C k_6 \rangle}{\langle K_R k_5 \rangle}. \quad (5)$$

TABLE 2 | Linear regressions of K_M or $\ln(K_M)$ on k_{cat} for various data sets of sample size N: Coefficients of y-intercept, K_M or $\ln(K_M)$, and x-variable (gradient), k_{cat}, with standard errors (SE), P-values and 95% ($P = 0.05$) confidence intervals.

Regression	Coefficients		SE	*P*-value	Lower 95%	Upper 95%
Other than C_3 Plants[a]	K_C	15.3	3.4	0.001	7.8	22.7
($N = 14$, *P-value* = 0.39)	k_{cat}^C	−0.9	1.0	0.39	−3.0	1.3
C_3 Plants[a]	K_C	4.5	1.4	0.007	1.5	7.5
($N = 14$, *P-value* = 0.008)	k_{cat}^C	1.4	0.4	0.008	0.4	2.4
C_3 Plants[a,c]	K_C	5.2	2.5	0.05	−0.04	10.4
($N = 21$, *P-value* = 0.072)	k_{cat}^C	1.6	0.8	0.07	−0.2	3.3
Higher Plants[b]	K_C	3.2	8.9	0.73	−17.0	23.5
($N = 11$, *P-value* = 0.13)	k_{cat}^C	3.7	2.2	0.13	−1.3	8.7
Higher Plants[c]	K_C	2.9	4.2	0.51	−5.8	11.5
($N = 30$, *P-value* = 0.002)	k_{cat}^C	3.8	1.1	0.002	1.5	6.1
Non-green algae[c]	K_C	28.6	15.0	0.10	−6.8	64.0
($N = 9$, *P-value* = 0.66)	k_{cat}^C	−3.7	8.0	0.66	−22.5	15.2
Diatoms[c]	K_C	26.8	36.6	0.49	−57.7	111
($N = 10$, *P-value* = 0.60)	k_{cat}^C	6.8	12.3	0.60	−21.6	35.3
Triticeae[d]	K_C	9.8	3.2	0.03	1.6	17.9
($N = 7$, *P-value* = 0.15)	k_{cat}^C	1.8	1.1	0.15	−0.9	4.6
Triticeae[d]	K_O	315	35.8	0.0003	223	408
($N = 7$, *P-value* = 0.023)	k_{cat}^O	138	42.6	0.02	28.5	247
Higher Plants[c]	K_O	115	52.1	0.04	7.5	222
(**Figure 2C**)						
($N = 27$, *P-value* $< 10^{-5}$)	k_{cat}^O	278	33.1	$<10^{-5}$	210	346
All Data[c] (**Figure 2A**)	$\ln(K_C)$	2.3	0.2	$<10^{-5}$	1.9	2.6
($N = 54$, *P-value* $< 10^{-5}$)	k_{cat}^C	0.23	0.04	$<10^{-5}$	0.15	0.31
All Data[b] (**Figure 2B**)	$\ln(K_C)$	1.5	0.2	$<10^{-5}$	1.1	1.9
($N = 19$, *P-value* $< 10^{-5}$)	k_{cat}^C	0.34	0.03	$<10^{-5}$	0.27	0.40

[a]*Data from Table 1 in Galmés et al. (2014)* [b]***Table 1*** *Savir et al. (2010)* [c] ***Table 1*** [d]*25°C data from Table 2 in Prins et al. (2016)*

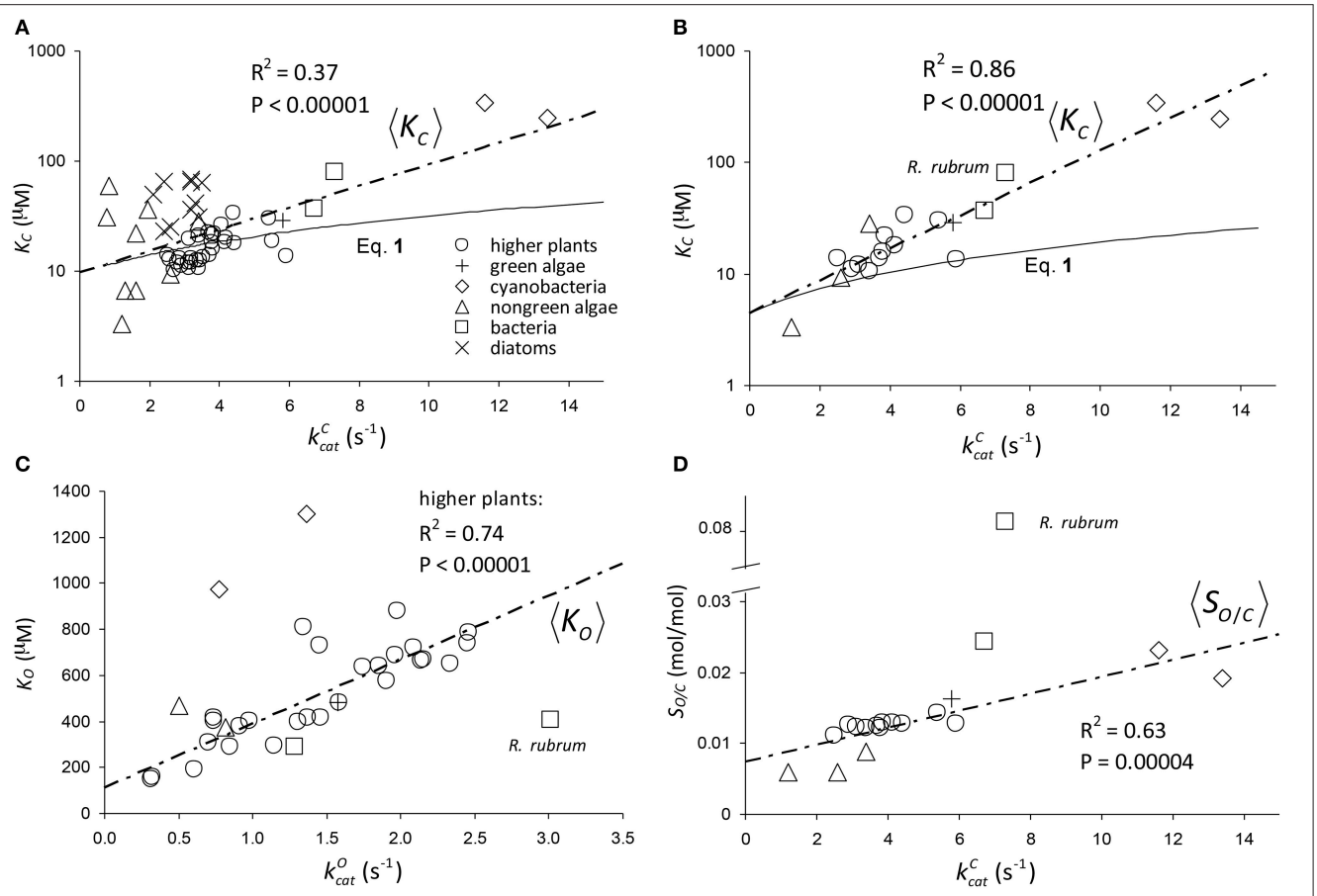

FIGURE 2 | Regression of: **(A)** K_C on k^C_{cat} using all data (**Table 1**) in the regression. The parameters of the exponential, $a_1 e^{b_1 k^C_{cat}}$, are $a_1 = 9.7$ μM and $b_1 = 0.23$ s. **(B)** K_C on k^C_{cat} using only the data compiled by Savir et al. (2010) in the regression. The parameters of the exponential are $a_1 = 4.5$ μ M and $b_1 = 0.34$ s. The Form II RuBisCO, *R. rubrum*, is not a significant outlier. **(C)** K_O on k^O_{cat} using all higher plant data (**Table 1**) only. The gradient and intercept of the regression line are 278 μM.s and 114 μM, respectively. Form II RuBisCO, *R. rubrum*, and the cyanobacteria are the significant outliers by the ESD test with $P = 0.05$ (Rosner, 1983). **(D)** Reciprocal relative specificity ($S_{O/C} = \frac{1}{S_{C/O}}$) on k^C_{cat} using the data compiled by Savir et al. (2010). The Form II RuBisCO, *R. rubrum* is the only significant outlier by the ESD test ($P = 0.05$), due mainly to its relatively higher value for $S_O = \frac{k^O_{cat}}{K_O}$ (**Figure 2C**), and was not included in the regression. The gradient and intercept are 1.2×10^{-3} s and 7.4×10^{-3} mol/mol, respectively. Note that in **(A,B)** K_C is graphed in logarithmic scale and Equation (1) has been graphed using the parameters at $k^C_{cat} = 0$ as derived from the regression analysis (see text).

In **Figure 4** it is assumed (Tcherkez et al., 2006) that the exponential increase in $\langle K_C \rangle$ conditional on k^C_{cat} arises from $\langle K_R k_5 \rangle$ (one correlation effect, i.e., due to CO_2 binding) while $\langle \gamma_C k_6 \rangle$ is a constant in Equation (4). Alternatively, in **Figure 5** we have assumed that variation arises from $\langle \gamma_C k_6 \rangle$ (another correlation effect i.e., due to CO_2 dissociation) while $\langle K_R k_5 \rangle$ is now the constant. Here the respective constants are the values of $\langle \gamma_C k_6 \rangle$ and $\langle K_R k_5 \rangle$ at $k^C_{cat} = 0$ as determined from the regression (**Figure 2B**). There is, of course, also the possibility that variability in both $\langle K_R k_5 \rangle$ and $\langle \gamma_C k_6 \rangle$ contribute to the non-linear behavior of $\langle K_C \rangle$, i.e., both $\langle K_R k_5 \rangle$ and $\langle \gamma_C k_6 \rangle$ are conditional on k^C_{cat}. In general, therefore, we could ascribe any functional dependence for either $\langle K_R k_5 \rangle$ or $\langle \gamma_C k_6 \rangle$ to this non-linear behavior.

For the regression of K_O on k^O_{cat} (**Figure 2C**), we have included only the data for all higher plants (**Table 1**). Unlike the above regressions of K_C on k^C_{cat} there are no indications of any deviations from non-linear behavior. The graph of K_O on k^O_{cat} for the higher plants in particular clearly conforms to a linear function, and the residuals of regressed K_O data are near normally distributed (**Figure 3**). From the intercept we find the expected value of the dissociation constant

$$\langle K^O_D \rangle = \frac{\langle \gamma_O k_{12} \rangle}{\langle K_R k_{11} \rangle} \approx 110\ \mu\text{M} \tag{6}$$

and from the gradient we obtain the constant

$$\left\langle \frac{1}{K_R k_{11}} \right\rangle \approx 280\ \mu\text{M.s}. \tag{7}$$

From Equations (6, 7) we estimate the expected value of the effective O_2 dissociation rate constant, $\langle \gamma_O k_{12} \rangle \approx 0.3\ \text{s}^{-1}$. Finally,

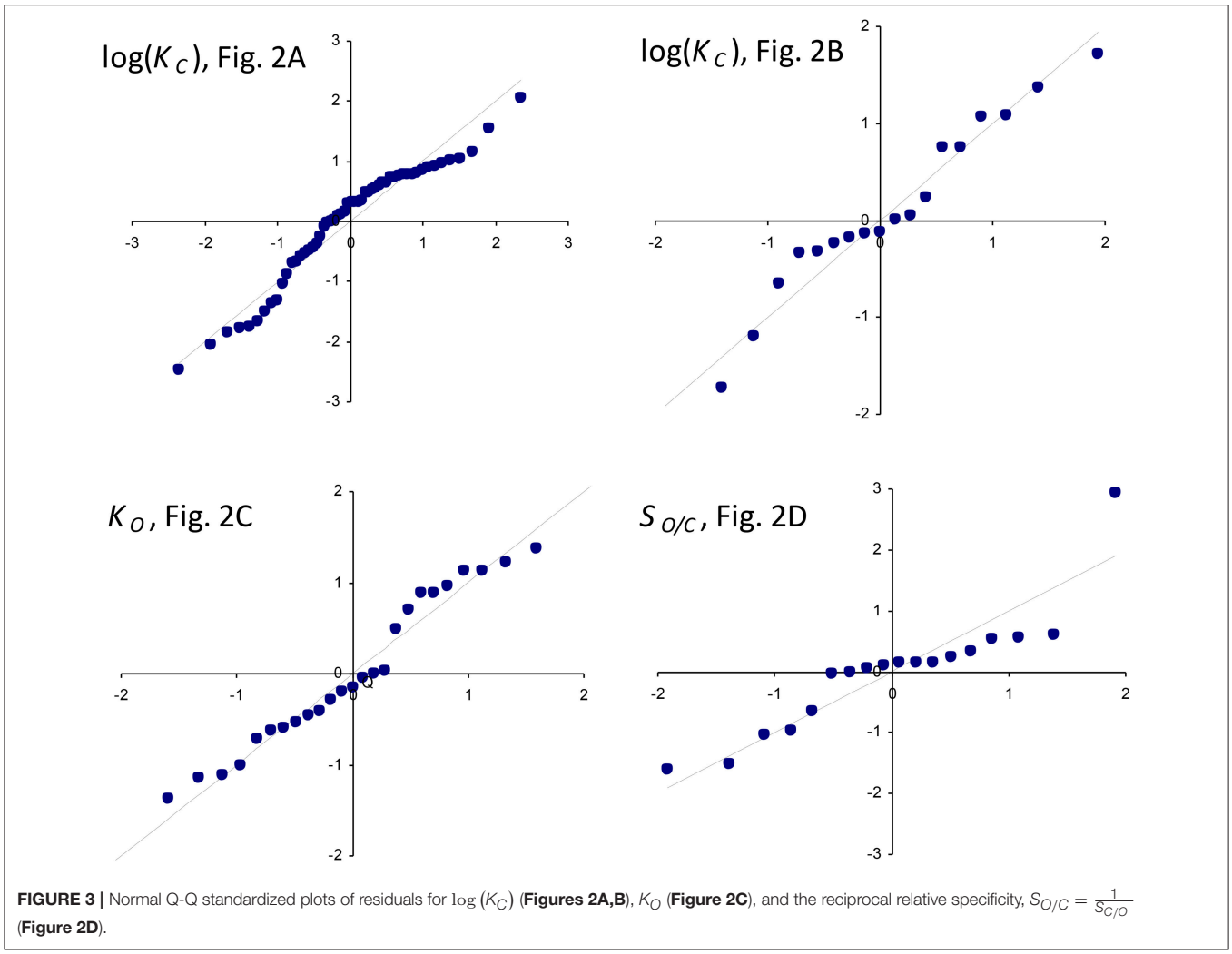

FIGURE 3 | Normal Q-Q standardized plots of residuals for $\log(K_C)$ (**Figures 2A,B**), K_O (**Figure 2C**), and the reciprocal relative specificity, $S_{O/C} = \frac{1}{S_{C/O}}$ (**Figure 2D**).

TABLE 3 | Expected values of dissociation rate constants (s^{-1}) for carboxylation ($\gamma_C K_6$) and oxygenation ($\gamma_C K_{12}$) with standard errors and corresponding 95% confidence intervals calculated from coefficients (gradient and intercept) with $P < 0.05$ in **Table 2**.

Rate constant	$\gamma_C K_6$ [a]	$\gamma_C K_6$ [b]	$\gamma_C K_6$ [c]	$\gamma_C K_{12}$ [d]	$\gamma_C K_{12}$ [e]
Expected value	4.4	3.0	3.2	0.4	2.3
Standard Error	±0.8	±0.3	±1.4	±0.2	±0.8
95% Confidence Interval	±1.6	±0.6	±3.0	±0.4	±1.9

[a] ***Table 1*** [b] *Savir et al. (2010)* (***Table 1***) [c] *C_3 plant data from Table 1 in Galmés et al. (2014)* [d] *higher plants* (***Table 1***) [e] *25°C Triticeae data from Table 2 in Prins et al. (2016).*

from the above determinations of $\left\langle \frac{1}{K_R k_5} \right\rangle$ (from **Figure 2B**) and $\left\langle \frac{1}{K_R k_{11}} \right\rangle$ we can estimate the expected CO_2 to O_2 ratio of the rate constants for binding at $k^C_{cat} = 0$ as $\left\langle \frac{k_5}{k_{11}} \right\rangle \approx 190$.

Relative Specificity

The graph of reciprocal relative specificity, $S_{O/C} = \frac{1}{S_{C/O}}$, against k^C_{cat} (**Figure 2D**) suggests a linear dependence. The residuals of regressed $S_{O/C}$ data are near normally distributed (**Figure 3**). We first consider the expected value of $S_{C/O}$ conditional on k^C_{cat} as the reciprocal of the equation for the straight line that describes $\langle S_{O/C} \rangle$, i.e.,

$$\langle S_{C/O} \rangle = \frac{1}{(a_2 + b_2 k^C_{cat})} \tag{8}$$

where $a_2 = 7.4 \times 10^{-3}$ mol/mol and $b_2 = 1.2 \times 10^{-3}$ s are the regression parameters (**Figure 2D**). Although Equation (8) generally provides a good fit to the data (**Figure 6**), it clearly does not display the correct limiting behavior as k^C_{cat} approaches zero Equation (3). However, defining the expected value as the ratio $\langle S_{C/O} \rangle = \frac{\langle S_C \rangle}{\langle S_O \rangle}$ and substituting $\langle S_C \rangle = \frac{k^C_{cat}}{a_1 e^{b_1 k^C_{cat}}}$ (**Figure 2A**), the expected value of $S_{C/O}$ conditional on k^C_{cat} can be written as

$$\langle S_{C/O} \rangle = \frac{\langle S_O \rangle^{-1} k^C_{cat}}{a_1 e^{b_1 k^C_{cat}}}. \tag{9}$$

As there are no correlations between k^C_{cat} and k^O_{cat} (**Figure 7A**) or K_O (**Figure 7B**), S_O is also not correlated (**Figure 7C**), and so

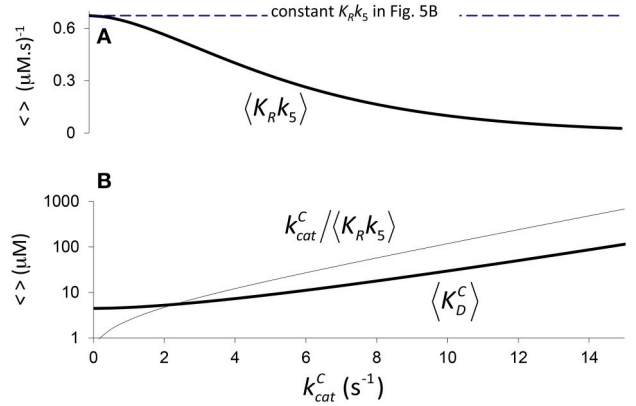

FIGURE 4 | Components of K_C if correlation is due to CO_2 binding. **(A)** $\langle K_R k_5 \rangle$ Equation (4) calculated assuming a constant value of $\langle \gamma_C k_6 \rangle = 3\,s^{-1}$ (**Figure 2B**, $a_1 = 4.5\,\mu M$, $b_1 = 0.34\,s$), and **(B)** the corresponding components of $\langle K_C \rangle$, i.e. $\frac{k_{cat}^C}{\langle K_R k_5 \rangle}$ and $\langle K_D^C \rangle$ Equation (5), derived from Equation (1) ($\langle K_C \rangle = \frac{k_{cat}^C}{\langle K_R k_5 \rangle} + \langle K_D^C \rangle$). **(B)** only is graphed in logarithmic scale.

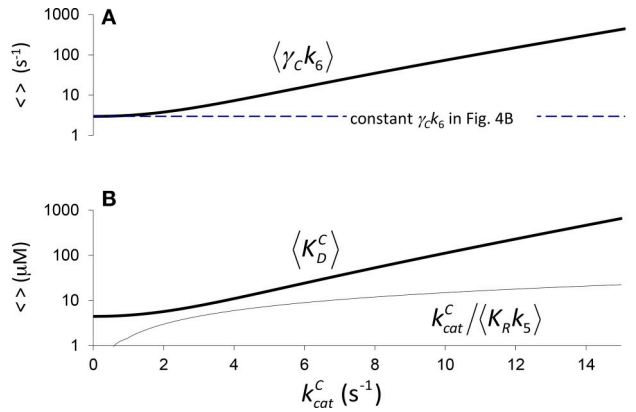

FIGURE 5 | Components of K_C if correlation is due to CO_2 dissociation. **(A)** $\langle \gamma_C k_6 \rangle = \langle K_R k_5 \rangle a_1 e^{b_1 k_{cat}^C} - k_{cat}^C$, (from rearranging Equation 4) calculated assuming a constant value of $\langle K_R k_5 \rangle = 0.7\,(\mu M.s)^{-1}$ (**Figure 2B**, $a_1 = 4.5\,\mu M$, $b_1 = 0.34\,s$), and **(B)** the corresponding components of $\langle K_C \rangle$, i.e., $\frac{k_{cat}^C}{\langle K_R k_5 \rangle}$ and $\langle K_D^C \rangle$. Both **(A,B)** are graphed in logarithmic scale.

the best possible approximation for Equation (9) takes the form $S_{C/O} \propto S_C$. The constant $\langle S_O \rangle^{-1}$ in Equation (9) can therefore be estimated by a linear regression of $S_{C/O}$ (excluding the outlier, *R. rubrum*, **Figure 2D**) on $\frac{k_{cat}^C}{a_1 e^{b_1 k_{cat}^C}}$ subject to the constraint $S_{C/O} = 0$ at $k_{cat}^C = 0$ to obtain the correct general equation for the expected value of $S_{C/O}$ conditional on k_{cat}^C as (**Figure 6**).

$$\langle S_{C/O} \rangle \approx \frac{490 k_{cat}^C}{a_1 e^{b_1 k_{cat}^C}}. \tag{10}$$

Assuming correlation (**Figure 2B**) arises from CO_2 binding, the factor implicit in Equation (10) corresponding to $\left\langle \frac{k_5}{k_{11}} \right\rangle$ (**Figure 6**)

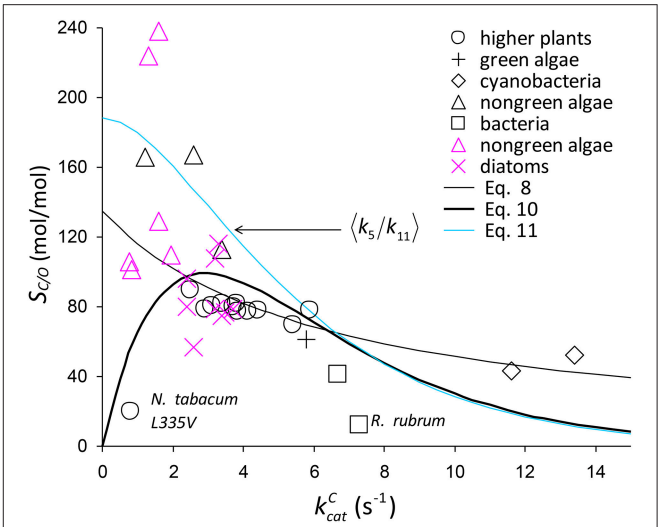

FIGURE 6 | Selection of $S_{C/O}$ data from **Table 1.** The symbols in black are from the compilation of Savir et al. (2010). The $S_{C/O}$ value for *N. tabacum* L335V mutant is shown. Also on the graph is $\langle S_{C/O} \rangle$ given by Equation (8) and Equation (10), including a possible factor of Equation (10), $\left\langle \frac{k_5}{k_{11}} \right\rangle$ (Equation 11).

that is also conditional on k_{cat}^C is estimated by (Equations 4, 7, **Figure 4A**).

$$\left\langle \frac{k_5}{k_{11}} \right\rangle \approx 280 \langle K_R k_5 \rangle. \tag{11}$$

Mutant Example

We use Equation (3) to rationalize the *in vitro* kinetic data for the Leu to Val mutation at position 335 (L335V) in tobacco (Whitney et al., 1999). The decrease in k_{cat}^C from 3.43 s^{-1} in the wild type to 0.81 s^{-1} in the mutant is accompanied by a large decrease also in $S_{C/O}$ from 81 to 20 mol/mol. In **Figure 8**, $S_{C/O}$ is plotted against k_{cat}^C assuming that in Equation (3) the term $\frac{k_5(k_{cat}^O + \gamma_O k_{12})}{k_{11} k_{cat}^O}$ is constant on the curve, i.e.,

$$S_{C/O} \propto \frac{k_{cat}^C}{(k_{cat}^C + \gamma_C k_6)} \tag{12}$$

We determine the constant factor such that $S_{C/O} = 81$ mol/mol for the wild-type tobacco at the two limits ($k_{cat}^C \gg \gamma_C k_6$ and $k_{cat}^C \ll \gamma_C k_6$) for specific values of $\gamma_C k_6 = 1, 2, 3$ and $4\,s^{-1}$. Note that in the limit $k_{cat}^C \gg \gamma_C k_6$ we obtain $S_{C/O} = \frac{k_5}{k_{11}} = 81$ mol/mol, while the lower limit for $k_{cat}^C \ll \gamma_C k_6$ gives $S_{C/O} = 0$. Noting that $\left\langle \frac{k_5}{k_{11}} \right\rangle = \frac{\left\langle \frac{K_R}{k_5} \right\rangle}{\left\langle \frac{K_R}{k_{11}} \right\rangle}$, the remaining kinetic parameters [$K_C = 10.7\,\mu M$, $k_{cat}^O = 1.17\,s^{-1}$, $K_O = 295\,\mu M$ for wild type, and $K_C = 5.1\,\mu M$, $k_{cat}^O = 0.39\,s^{-1}$, $K_O = 48.9\,\mu M$ for the mutant] (Whitney et al., 1999) can be used to simply determine the expected value of the ratio $\frac{k_5}{k_{11}}$ as.

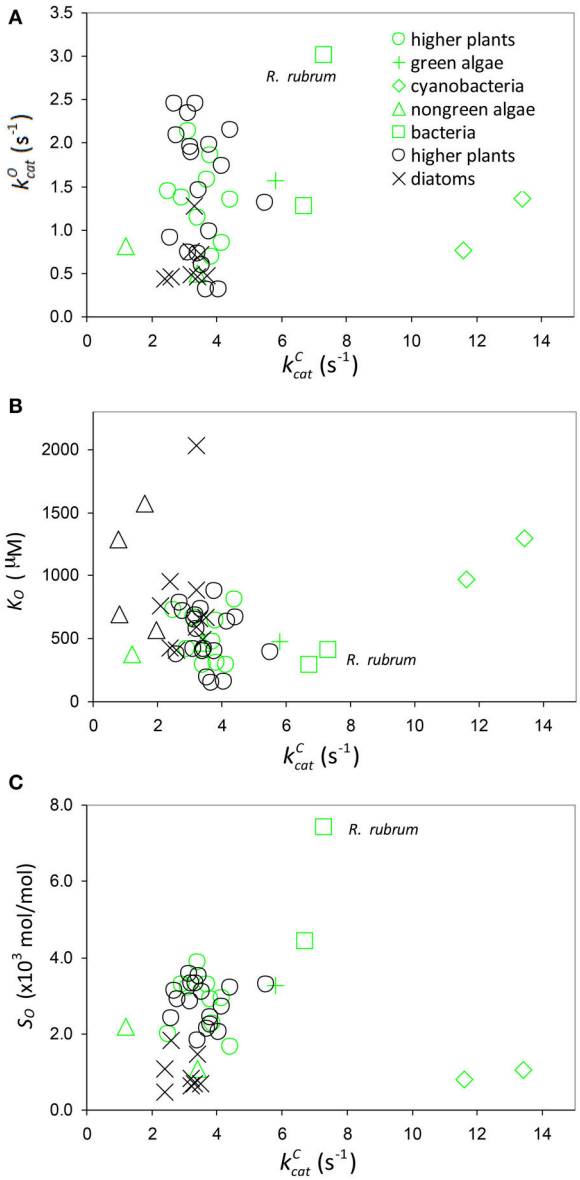

FIGURE 7 | Oxygenation parameters. Scatter plots of **(A)** k^O_{cat} against k^C_{cat} **(B)** K_O against k^C_{cat} and **(C)** specificity, $S_O = \frac{k^O_{cat}}{K_O}$, against k^C_{cat} including all data in **Table 1**. Data points highlighted in green are those compiled by Savir et al. (2010).

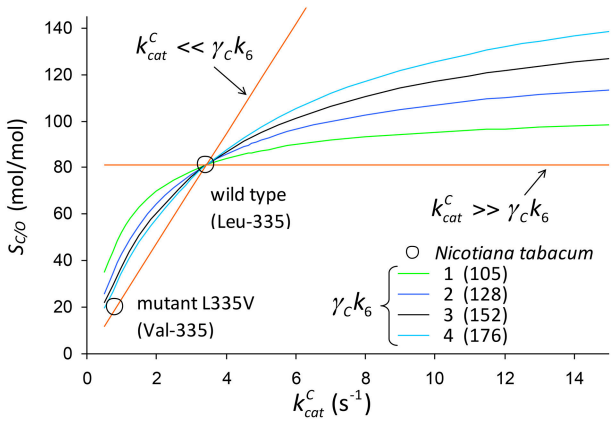

FIGURE 8 | $S_{C/O}$ (Equation 3) plotted against k^C_{cat} assuming that $\frac{k_5\left(k^O_{cat}+\gamma_O k_{12}\right)}{k_{11}k^O_{cat}}$ is constant on the curve (Equation 12). The numbers in parentheses are the values of $\frac{k_5\left(k^O_{cat}+\gamma_O k_{12}\right)}{k_{11}k^O_{cat}}$ in Equation (3) that give $S_{C/O} = 81.1$ mol/mol for wild-type tobacco given CO_2 dissociation rate constants of $\gamma_C k_6 = 1, 2, 3$ and $4\,s^{-1}$. For the wild type $k^C_{cat} = 3.43\,s^{-1}$, and for the mutant (Val-335) $k^C_{cat} = 0.81\,s^{-1}$ and $S_{C/O} = 20.1$ mol/mol (Whitney et al., 1999).

$$\left\langle \frac{k_5}{k_{11}} \right\rangle = \frac{\Delta k^C_{cat} K_O}{\Delta k^O_{cat} \Delta K_C} \approx 150 \qquad (13)$$

where Δ is the difference between wild type and mutant.

DISCUSSION

Significant Dissociation of CO_2 and O_2 Substrates

The trend lines (**Figure 2**) clearly intercept the vertical axes well above zero, indicating significant expected values for the dissociation constants $\gamma_C k_6$ and $\gamma_O k_{12}$. However, the rate constant for CO_2 dissociation has been previously estimated as not more than about 5% of k^C_{cat} (Pierce et al., 1986; McNevin et al., 2007), so that it has generally been assumed that $\frac{k^C_{cat}}{(k^C_{cat}+\gamma_C k_6)} \approx 1$. Our estimates (**Figures 2A,B**, **Table 3**) of the expected value (at least for low k^C_{cat}) are much higher, and find support in the kinetics modeling study of RuBisCO from spinach. We find that the expected values of dissociation rate constants ($\gamma_C k_6$) for the binding of the substrate CO_2 are 4.3 s^{-1} (**Figure 2A**), 3.0 s^{-1} (**Figure 2B**), and 3.1 s^{-1} for a subset of C_3 plants (Galmés et al., 2014; Prins et al., 2016), noting that the differences are not statistically significant (**Table 3**). These values can be compared with 1.6 ±1.1 s^{-1} estimated for the CO_2 dissociation rate constant in spinach (McNevin et al., 2006, 2007), and the 5 – 10 μM range of $\langle K^C_D \rangle$ for lower values of k^C_{cat} (**Figures 4B**, **5B**) is also consistent with a $K^C_D = \frac{k_6}{k_5}$ of 3 μM for spinach RuBisCO (McNevin et al., 2006). The effective CO_2 dissociation rate constant, $\gamma_C k_6$, impacts the k^C_{cat} dependence of $S_{C/O}$ (**Figure 8**). As k^C_{cat} approaches $\gamma_C k_6$ Equation (12) describes the rapid decline in $S_{C/O}$ due to increasing probability that the CO_2 will dissociate from RuBP before catalysis takes place. The observed values of k^C_{cat} and $S_{C/O}$ for the L335V mutant (Whitney et al., 1999) are entirely consistent with a $\gamma_C k_6$ greater than k^C_{cat}. The expected value of $\frac{k_5}{k_{11}}$ as given by Equation (13) is also consistent with the value obtained when averaged over a larger number of RuBisCOs with lower k^C_{cat} (**Figure 6**). Thus, changes in the gas-substrate binding in the mutant RuBisCO appear to be minimal, the bulk of the effect being described by Equation (12). The dissociation rate constant of O_2 is generally considered effectively zero (Tcherkez, 2013, 2016). However, although the expected value of 0.4 ± 0.4 s^{-1} for $\gamma_O k_{12}$ in higher plants obtained here (**Table 3**) is

significantly lower than the mean k_{cat}^{O} of 1.3 ± 0.2 s^{-1} (from data in **Table 1**) it is still sufficient to have an impact on K_O (Equation 2). Additionally, the expected value of 2.3 ± 1.9 s^{-1} for $\gamma_O k_{12}$ in Triticeae (**Table 3**) and the corresponding mean k_{cat}^{O} of 0.83 ± 0.16 s^{-1} (Prins et al., 2016) are not significantly different. Statistical analysis of the available data therefore suggests the expected (or average) value of the dissociation rate is not significantly lower than that of the catalytic rate. Moreover, a knowledge of rate differences in any particular RuBisCO requires more kinetic data than is currently available. Consequently, there is no justification for generally neglecting either of the dissociation rate constants, $\gamma_C k_6$ or $\gamma_O k_{12}$, i.e., assuming they are an order of magnitude or more lower than the corresponding catalytic rates, as has been done previously (Tcherkez, 2013, 2016).

The Tight-Binding Hypothesis

Assuming $\left\langle \frac{k_5}{k_{11}} \right\rangle$ decreases with increasing k_{cat}^{C} (Equation 11, **Figure 6**), it could be regarded as a proxy for $S_{C/O}$ (Tcherkez, 2013). Also as the specificity of oxygenation, S_O, is not correlated with k_{cat}^{C} (**Figure 7C**), the variation in $\left\langle \frac{k_5}{k_{11}} \right\rangle$ would be largely constrained to the dependence of $\langle k_5 \rangle$ on k_{cat}^{C} (**Figure 4A**). It has been hypothesized (Tcherkez et al., 2006; Tcherkez, 2013) that such a constraint is to be expected from the predicted energetics of the reaction as tighter binding of CO_2 to ribulose bisphosphate (increasing k_5) would necessarily raise the activation free energy (decreasing k_{cat}^{C}) required for the subsequent steps leading to turnover of product. However, the generality of this tight-binding (TB) hypothesis has come under question (Hanson, 2016) for its inability to explain the variations in $S_{C/O}$ that have been observed in some RuBisCOs (Young et al., 2016). It would seem that the TB hypothesis suffers from a more fundamental problem in that it is based on an incomplete and unrepresentative data distribution. In the present analysis, Equation (8) provides the better fit $R^2 = 0.63$ to the selected data (**Figure 2D**), although it is not the more general equation for $\langle S_{C/O} \rangle$ (Equation 10, **Figure 6**). Similar types of relationships that provide an even tighter fit to the data have been reported elsewhere: $K_C \propto \left(k_{cat}^{C}\right)^2$ [$R^2 = 0.90$] and $S_{C/O} \propto \left(k_{cat}^{C}\right)^{-0.51}$ [$R^2 = 0.79$] (Savir et al., 2010). The TB hypothesis is posited on $\frac{k_5}{k_{11}}$ determining the dependence of $S_{C/O}$ on k_{cat}^{C}. Significantly, all of these analyses are in fact conditional on $k_{cat}^{C} \gg \gamma_C k_6$, i.e, neglect of the CO_2 dissociation rate constant, k_6. However, the high level of variance in K_C and $S_{C/O}$ (**Figures 2A**, **6**, respectively) argues for a more cautious data interpretation in the regression analysis. Statistically, the quadratic (Savir et al., 2010) and exponential (**Figure 2B**) forms both describe the dependence of K_C on k_{cat}^{C} equally well, but only the latter, more general case (Equations 3, 10), allows nonzero values for $\gamma_C k_6$ (**Figure 5A**).

Rate Constants May Not Be Highly Correlated

The deviation of any given data point (**Figure 6**) from the expected value (Equation 10) can be attributed to variations in the parameters of Equation (3). We expect that S_O will generally produce random variations in $S_{C/O}$ (**Figure 7C**), although, possibly lower k_{11} (higher K_O, **Figure 2C**) for the cyanobacteria may in part account for a systematic reduction in $S_{C/O}$. The CO_2 dissociation term, $\gamma_C k_6$, will certainly become apparent at low enough k_{cat}^{C} values (**Figures 4**, **5**). In particular, variations in $\gamma_C k_6$ may contribute significantly to the large variance seen in the non-green algae (**Figures 2A**, **6**). If the catalytic rate correlates with k_5, regression analysis defines only the first moment, $\langle k_5 \rangle$, of the distribution (**Figure 4A** and Equation 11, **Figure 6**), and provides no information on the variance. In the absence of any coupling, mutations produce random changes in the underlying rate constants, k_i. Irrespective of whether rate constants are correlated, the expected value of k_i is given by $\frac{\sum_s^n k_i^s}{n}$ where k_i^s is the value of a rate constant for a given sequence (s). In reality, the composition of the sequence space, Ω (i.e., any number of known sequences), will be determined in varying degrees by genetic drift and natural selection, as these determine the probability that a mutation becomes fixed. If the variations in k_i^s themselves are entirely random (zero correlation), we might expect both $S_{C/O}$ and k_{cat}^{C} at the high end of their observed values, as there is nothing to constrain them and the combined effect should have become fixed in some species by positive selection. The TB hypothesis attempts to explain this absence of both high $S_{C/O}$ and high k_{cat}^{C} by positive selection processes occurring within particular constraints (**Figures 4A**, **6**) imposed on the chemical reaction steps (Tcherkez et al., 2006), but it may also be explained by competing selection pressures. The essential difference is that the origin of the evolutionary constraints is shifted from k_i^s to Ω.

Competing Selection Pressures May Constrain RuBisCo

From a biophysical perspective, thermodynamic stability is recognized as the most important constraint on the evolution of proteins and their ability to acquire new function (Tokuriki and Tawfik, 2009; Sikosek and Chan, 2014). The necessity of a protein to maintain the integrity of its folded structure despite the destabilizing effects of accumulated mutations results in only a small percentage being fixed by positive selection. Consequently, in the evolution of C_3 to C_4 plants, destabilizing mutations that are selected on the basis of improved activity are followed by mutations that restore stability with little impact on activity (Studer et al., 2014). This leads to an apparent tradeoff between activity and stability that may well limit the ability of RuBisCO to fix the number of mutations required to increase both $S_{C/O}$ and k_{cat}^{C}. Depending on the sub-cellular CO_2/O_2 ratio, the fixed mutations increase specificity (for low ratio) or catalytic rate (for high ratio), or a varying combination of both, whichever best optimizes photosynthesis.

Potential for Optimizing Carbon Fixation

The origin of the constraint(s) has significant implications for the optimization of RuBisCO activity. If the constraint is on Ω (i.e., from competing selection pressures) rather than k_i^s, greater variability may be exhibited. To what extent the functional limits of RuBisCO are reflected in the minimum and maximum values of kinetic parameters is not yet clear for RuBisCOs with

higher k_{cat}^{C} because of the absence of empirical data. Much effort has been directed toward research on higher plants with particular emphasis on the evolution of C_3 to C_4 plants with their associated CCMs, although the recent work on diatoms may now help stimulate investigations into a more diverse range of photosynthetic organisms (Hanson, 2016; Young et al., 2016). Diatoms and C_3 plants share very similar k_{cat}^{C}, although the variance, var($S_{C/O}$), for diatoms is relatively large (with corresponding variations in CCM expression), whereas for C_3 plants var($S_{C/O}$) is barely significant (**Figure 6**, **Table 1**). This could raise the possibility of improving specificity, if not k_{cat}^{C}, in higher plants. It is perhaps not surprising that the non-green (red) algae, from which diatoms have evolved with somewhat lower k_{cat}^{C} values, also exhibit high var($S_{C/O}$) (**Figure 6**). The data distributions are incomplete (**Figures 2A**, **6**, **Table 1**); there is a scarcity of data for green algae, photosynthetic bacteria and cyanobacteria, with k_{cat}^{C} values between 6 s^{-1} and 14 s^{-1}. Discoveries of significant variance among these also may provide important clues on how to achieve increases in both k_{cat}^{C} and $S_{C/O}$ in higher plants.

CONCLUSION

The results of our analysis using regression analysis on updated RuBisCO-kinetic data sets suggest that CO_2 dissociation from the RuBisCO gas-addition complex is generally more important in rationalizing the observed variations in the kinetics of RuBisCO than hitherto assumed (Tcherkez et al., 2006; Tcherkez, 2013). Moreover, we have identified significant variations in the statistical correlations between K_M and k_{cat} in higher plants, i.e., the non-linear correlation for carboxylation as opposed to the linear correlation for oxygenation. These findings cast doubt on the hypothesis (Tcherkez et al., 2006; Savir et al., 2010; Tcherkez, 2013) that RuBisCO is so tightly constrained by the active-site chemistry that its activity is effectively optimized. Rather, the current body of kinetic parameters exhibits far more plasticity than this hypothesis predicts. We suggest that the possibility that the apparent tradeoff observed between k_{cat}^{C} and $S_{C/O}$ could arise from competing selection pressures on RuBisCO activity and stability (Studer et al., 2014) be given more attention. The relative strengths of these selection pressures would determine the strength of the constraints and, thus, the possibilities of improving the kinetics of RuBisCO by site-directed mutagenesis. Indeed, although published comments (Griffiths, 2006; Gutteridge and Pierce, 2006) on the paper of Tcherkez et al. (2006) noted the vastness of sequence space that would need to be sampled, neither showed any positivity that a rational method to increase the efficiency of such a search was possible merely noting (Griffiths, 2006) directed evolution as a possibility. However, a method to reduce the sequence-search space for RuBisCO has since been reported in a patent (Gready and Kannappan, 2009).

In summary, there is still wide conjecture in the literature regarding the mechanisms by which plants ultimately regulate photosynthesis (Igamberdiev, 2015), and the absolute limitations of RuBisCO functionality have only been partly explored, as recent studies (Hanson, 2016; Young et al., 2016) suggest. Consequently, the potential for increasing both the catalytic turnover and relative specificity in higher plants with the view to improving photosynthesis remains to be fully tested. As argued (Hanson, 2016), kinetic data for a wider diversity of RuBisCOs are much needed and will likely prove useful in guiding the reengineering of higher-plant RuBisCOs with both significantly higher turnover rate and specificity. Our analysis suggests that such simultaneous improvement in both specificity and turnover rate is possible, and that competing selection pressures of activity and stability better explain the nature of constraints. Improved understanding of these competing selection pressures is much needed.

AUTHOR CONTRIBUTIONS

PC, BK, and JG designed and performed the research, wrote the paper and approved it for submission.

ACKNOWLEDGMENTS

We sincerely thank Dr. Andrey Bliznyuk for checking the manuscript and helpful comments. We also acknowledge the Australian NCI (National Computational Infrastructure) for computing support. We thank the reviewers for helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00183/full#supplementary-material

REFERENCES

Andersson, I. (2008). Catalysis and regulation in Rubisco. *J. Exp. Bot.* 59, 1555–1568. doi: 10.1093/jxb/ern091

Badger, M. R., and Andrews, T. J. (1987). "Co-evolution of rubisco and CO_2 concentrating mechanisms," in *Progress in Photosynthesis, Research: Vol. 3 Proceedings of the VIIth International Congress on Photosynthesis Providence, Rhode Island, U. S. A.,* ed J. Biggins. (Dordrecht: Springer Netherlands), 601–609.

Badger, M. R., Andrews, T. J., Whitney, S. M., Ludwig, M., Yellowlees, D. C., Leggat, W., et al. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. *Can. J, Bot.* 76, 1052–1071. doi: 10.1139/b98-074

Bainbridge, G., Madgwick, P., Parmar, S., Mitchell, R., Paul, M., Pitts, J., et al. (1995). Engineering Rubisco to change its catalytic properties. *J. Exp. Bot.* 46, 1269–1276. doi: 10.1093/jxb/46.special_issue.1269

Bar-Even, A., Milo, R., Noor, E., and Tawfik, D. S. (2015). The moderately efficient enzyme: futile encounters and enzyme floppiness. *Biochemistry* 54, 4969–4977. doi: 10.1021/acs.biochem.5b00621

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., et al. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50, 4402–4410. doi: 10.1021/bi2002289

Carmo-Silva, A. E., Keys, A. J., Andralojc, P. J., Powers, S. J., Arrabaça, M. C., and Parry, M. A. J. (2010). Rubisco activities, properties, and regulation in three different C(4) grasses under drought. *J. Exp. Bot.* 61, 2355–2366. doi: 10.1093/jxb/erq071

Carmo-Silva, E., Scales, J. C., Madgwick, P. J., and Parry, M. A. (2015). Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant Cell Environ.* 38, 1817–1832. doi: 10.1111/pce.12425

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. doi: 10.1021/cr970010r

Farquhar, G. D. (1979). Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Arch. Biochem. Biophys.* 193, 456–468. doi: 10.1016/0003-9861(79)90052-3

Galmés, J., Kapralov, M. V., Andralojc, P. J., Conesa, M. À., Keys, A. J., Parry, M. A. J., et al. (2014). Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ.* 37, 1989–2001. doi: 10.1111/pce.12335

Ghannoum, O., Evans, J. R., Chow, W. S., Andrews, T. J., Conroy, J. P., and von Caemmerer, S. (2005). Faster rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C(4) grasses. *Plant Physiol.* 137, 638–650. doi: 10.1104/pp.104.054759

Gready, J. E., and Kannappan, B. (2009). *Process for Generation of Protein and Uses Thereof.* US patent No 12/422,190. Washington, DC: U.S. Patent and Trademark Office.

Griffiths, H. (2006). Plant biology: designs on Rubisco. *Nature* 441, 940–941. doi: 10.1038/441940a

Gutteridge, S., and Pierce, J. (2006). A unified theory for the basis of the limitations of the primary reaction of photosynthetic CO_2 fixation: was Dr. Pangloss right? *Proc. Natl. Acad. Sci. U.S.A.* 103, 7203–7204. doi: 10.1073/pnas.0602075103

Hanson, D. T. (2016). Breaking the rules of Rubisco catalysis. *J. Exp. Bot.* 67, 3180–3182. doi: 10.1093/jxb/erw197

Igamberdiev, A. U. (2015). Control of Rubisco function via homeostatic equilibration of CO_2 supply. *Front. Plant Sci.* 6:106. doi: 10.3389/fpls.2015.00106

Jordan, D. B., and Ogren, W. L. (1981). Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* 291, 513–515. doi: 10.1038/291513a0

Jordan, D. B., and Ogren, W. L. (1983). Species variation in kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* 227, 425–433. doi: 10.1016/0003-9861(83)90472-1

Kannappan, B., and Gready, J. E. (2008). Redefinition of Rubisco carboxylase reaction reveals origin of water for hydration and new roles for active-site residues. *J. Am. Chem. Soc.* 130, 15063–15080. doi: 10.1021/ja803464a

Koralov, L. B., and Sinai, Y. G. (2007). *Theory of Probability and Random Processes*. Berlin; Heidelberg: Springer-Verlag.

Kubien, D. S., Whitney, S. M., Moore, P. V., and Jesson, L. K. (2008). The biochemistry of Rubisco in Flaveria. *J. Exp. Bot.* 59, 1767–1777. doi: 10.1093/jxb/erm283

Lorimer, G. H. (1981). The carboxylation and oxygenation of ribulose 1,5-bisphosphate: the primary events in photosynthesis and photorespiration. *Annu. Rev. Plant Physiol.* 32, 349–382. doi: 10.1146/annurev.pp.32.060181.002025

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. doi: 10.1074/jbc.M706274200

McNevin, D., von Caemmerer, S., and Farquhar, G. (2006). Determining RuBisCO activation kinetics and other rate and equilibrium constants by simultaneous multiple non-linear regression of a kinetic model. *J. Exp. Bot.* 57, 3883–3900. doi: 10.1093/jxb/erl156

Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., et al. (2013). Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* 64, 717–730. doi: 10.1093/jxb/ers336

Peterhansel, C., Niessen, M., and Kebeish, R. M. (2008). Metabolic engineering towards the enhancement of photosynthesis. *Photochem. Photobiol.* 84, 1317–1323. doi: 10.1111/j.1751-1097.2008.00427.x

Pierce, J., Andrews, T. J., and Lorimer, G. H. (1986). Reaction intermediate partitioning by ribulose-bisphosphate carboxylases with differing substrate specificities. *J. Biol. Chem.* 261, 10248–10256.

Prins, A., Orr, D. J., Andralojc, P. J., Reynolds, M. P., Carmo-Silva, E., and Parry, M. A. J. (2016). Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. *J. Exp. Bot.* 67, 1827–1838. doi: 10.1093/jxb/erv574

Roberts, D. V. (1977). *Enzyme Kinetics*. Cambridge: Cambridge University Press.

Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure. *Technometrics* 25, 165–172. doi: 10.1080/00401706.1983.10487848

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. doi: 10.1073/pnas.0911663107

Seemann, J. R., Badger, M. R., and Berry, J. A. (1984). Variations in the specific activity of ribulose-1,5-bisphosphate carboxylase between species utilizing differing photosynthetic pathways. *Plant Physiol.* 74, 791–794. doi: 10.1104/pp.74.4.791

Sikosek, T., and Chan, H. S. (2014). Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* 11:20140419. doi: 10.1098/rsif.2014.0419

Studer, R. A., Christin, P. A., Williams, M. A., and Orengo, C. A. (2014). Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2223–2228. doi: 10.1073/pnas.1310811111

Tcherkez, G. (2013). Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters. *Plant Cell Environ.* 36, 1586–1596. doi: 10.1111/pce.12066

Tcherkez, G. (2016). The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ.* 39, 983–997. doi: 10.1111/pce.12629

Tcherkez, G. G., 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. doi: 10.1073/pnas.0600605103

Tokuriki, N., and Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19, 596–604. doi: 10.1016/j.sbi.2009.08.003

Whitney, S. M., Houtz, R. L., and Alonso, H. (2011). Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol.* 155, 27–35. doi: 10.1104/pp.110.164814

Whitney, S. M., von Caemmerer, S., Hudson, G. S., and Andrews, T. J. (1999). Directed mutation of the Rubisco large subunit of tobacco influences photorespiration and growth. *Plant Physiol.* 121, 579–588. doi: 10.1104/pp.121.2.579

Yeoh, H.-H., Badger, M. R., and Watson, L. (1980). Variations in K_mCO_2 of Ribulose-1,5-bisphosphate carboxylase among grasses. *Plant Physiol.* 66, 1110–1112. doi: 10.1104/pp.66.6.1110

Young, J. N., Heureux, A. M., Sharwood, R. E., Rickaby, R. E., Morel, F. 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. doi: 10.1093/jxb/erw163

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.