Supporting Information

for

The mechanism of Rubisco catalyzed carboxylation reaction: chemical aspects involving acid-base chemistry and functioning of the molecular machine.

Immacolata C. Tommasi*

Dipartimento di Chimica, Università di Bari "A. Moro", v. Orabona, 4, 70126 Bari, Italy.

(15 pages)

* Corresponding author. Tel.: +39-080-544-2484; e-mail: immacolata.tommasi@uniba.it

Section S1. The Global Carbon Budget.[1]

The Global Carbon Budget report published within the Global Carbon Project is an accurate assessment of anthropogenic carbon dioxide emissions and CO_2-fluxes within environmental compartments contributing to the steadily increase of the average amount of atmospheric CO_2 (referred as "CO_2 growth rate" , G_{ATM}). The CO_2-budget accounts for CO_2-emissions (including emissions form fuels and industrial activities, E_{FOS} and land use change, E_{LUC}. E_{FOS} takes into account emission form fuels and industrial activities excluding the cement carbonation sink) and CO_2-sinks (including terrestrial, S_{LAND} and oceans sinks, S_{OCEAN}) resulting in the mass balance shown in Eq. 1 (budget imbalance, B_{IM}, compensates for overestimated emissions or underestimated sinks and accounts approximately for 0.37 $GtCO_2y^{-1}$.

$$E_{FOS} + E_{LUC} = G_{ATM} + S_{OCEAN} + S_{LAND} + B_{IM} \quad \text{Eq.1}$$

Through the text of the Reports CO_2 fluxes are expressed in $GtCy^{-1}$ unit. To convert $GtCy^{-1}$ into the $GtCO_2$/year) the following equation can be used: 1 GtC = 3.664 $GtCO_2$.

According to the Global Carbon Budget Report published in 2020, S_{LAND} and S_{OCEAN} are estimated at 12.4 $GtCO_2$/y and 9.16 $GtCO_2$/y respectively (data are averaged globally for the decade 2010-2019) while E_{FOS} and E_{LUC} fluxes are estimated at 35.1 and 5.8 $GtCO_2$/y.

Table S1. [1]

Cumulative CO_2 for different time of periods in GtC. All uncertainties are reported as ±1□. The budget imbalance provides a measure of the discrepancies among the nearly independent estimates. Its uncertainty exceeds ± 60 GtC. The method used here does not capture the loss of additional sink capacity from reduced forest cover, which is about 20 GtC for the years 1850-2018 and would exacerbate the budget imbalance. All values are rounded to the nearest 5 GtC, and therefore columns do not necessarily add to zero. (Table from ref. 1, used without changes from original source distributed under the creative common attribution 4.0 licence, https://essd.copernicus.org/articles/12/3269/2020/).

Units of GtC	1750-2019	1850-2014	1959-2019	1850-2018	1850-2020[a]
Emissions					
Fossil CO_2 emissions (E_{FOS})	445 ± 20	395 ± 20	365 ± 20	445 ± 20	455 ± 20
Land use change CO_2 emissions (E_{LUC})	255 ± 70[b]	200 ± 60[c]	85 ± 40[d]	210 ± 60[b]	210 ± 60
Total emissions	700 ± 75	595 ± 65	450 ± 30	650 ± 65	665 ± 65
Partitioning					
Growth rate in atmospheric CO_2 concentration (G_{ATM})	285 ± 5	235 ± 5	205 ± 5	265 ± 5	270 ± 5
Ocean sink (S_{OCEAN})[e]	170 ± 20	145 ± 20	105 ± 20	160 ± 20	165 ± 20
Terrestrial sink (S_{LAND})	230 ± 50	195 ± 50	145 ± 35	210 ± 55	215 ± 55
Budget imbalance					
BIM = E_{FOS} + E_{LUC} – (G_{ATM} + S_{OCEAN} + S_{LAND})	20	20	0	20	20

[a]Using projections for the year 2020. Uncertainties are the same as for the 1850–2019 period. [b]Cumulative E_{LUC} 1750–1849 of 30 GtC based on multi-model mean of Pongratz et al. (2009),[2] Shevliakova et al. (2009),[3] Zaehle et al. (2011),[4] and Van Minnen et al. (2009).[5] The 1850–2019 period from mean of H&N (Houghton and Nassikas, 2017)[6] and BLUE (Hansis et al., 2015).[7] The 1750–2018 uncertainty is estimated from standard deviation of DGVMs over 1850–2018 scaled by 1750–2018 emissions.
[c] Cumulative E_{LUC} based on H&N and BLUE. Uncertainty is estimated from the standard deviation of DGVM estimates.
[d] Cumulative E_{LUC} based on H&N and BLUE. Uncertainty is formed from the uncertainty in annual E_{LUC} over 1959–2019, which is 0.7 GtC yr^{-1} multiplied by the length of the time series. [e]Ocean sink uncertainty from IPCC (Denman et al., 2007).[8]

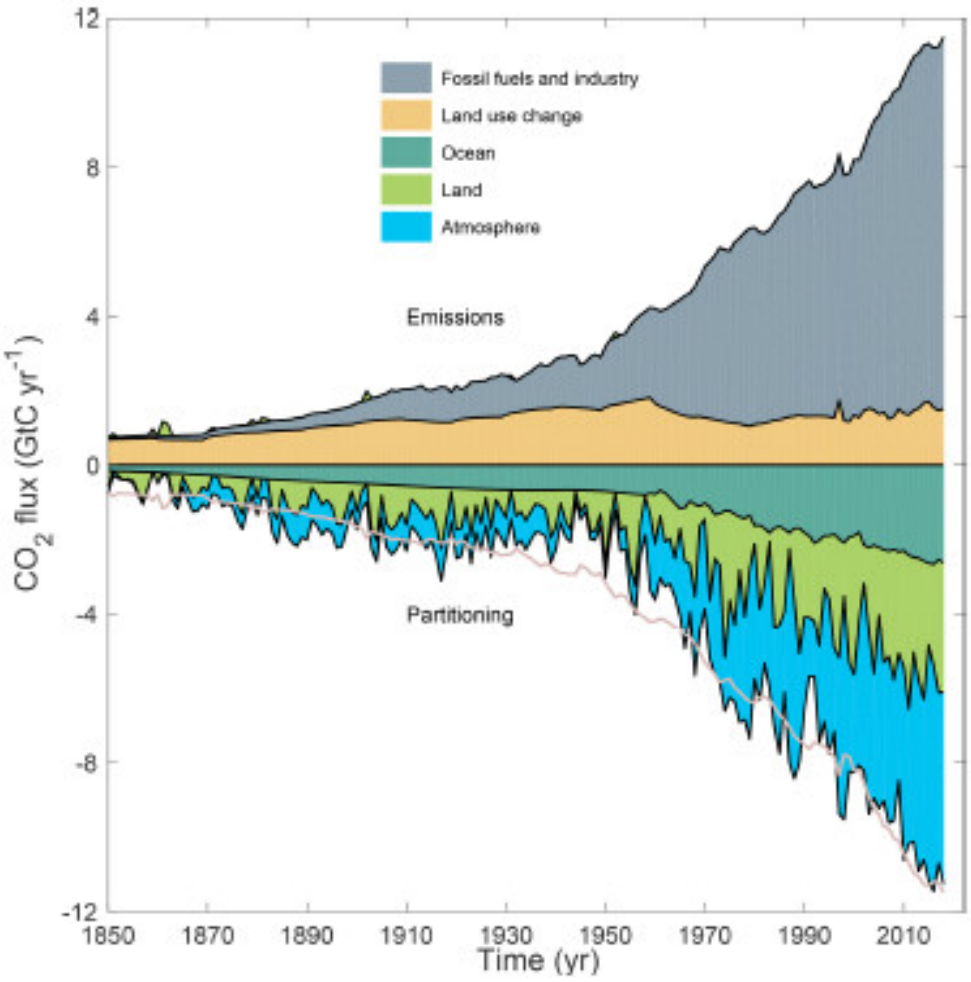


Figure S1 Combined components of the global carbon budget illustrated as a function of time, for fossil CO_2 emissions (E_{FOS}, grey) and emissions from land use change (E_{LUC}, brown), as well as their partitioning among the atmosphere (G_{ATM}, blue), ocean (S_{OCEAN}, turquoise), and land (S_{LAND}, green). The partitioning is based on nearly independent estimates from observations

(for G_{ATM}) and from process model ensembles constrained by data (for S_{OCEAN} and S_{LAND}), and it does not exactly add up to the sum of the emissions, resulting in a budget imbalance, which is represented by the difference between the bottom pink line (reflecting total emissions) and the sum of the ocean, land, and atmosphere. (Figure from ref. 1, used without changes from original source distributed under the creative common attribution 4.0 licence, https://essd.copernicus.org/articles/12/3269/2020/).

Table S2.[1] Decadal mean in the five components of the anthropogenic CO_2 budget for different periods and the last year available. All values are in giga tonnes of carbon per year, and uncertainties are reported as ±1□. The table also shows the budget imbalance (BIM), which provides a measure of the discrepancies among the nearly independent estimates and has an uncertainty exceeding ±1 GtC yr^{-1}. A positive imbalance means the emissions are overestimated and/or the sinks are too small. All values are rounded to the nearest 0.1 GtC and therefore columns do not necessarily add to zero (Table from ref. 1, used without changes from original source distributed under the creative common attribution 4.0 licence, https://essd.copernicus.org/articles/12/3269/2020/).

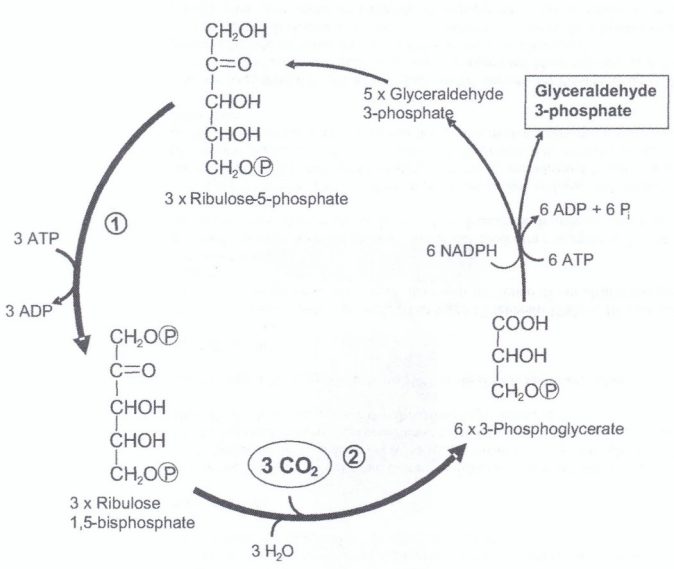
Mean GtCyr^{-1}	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2009-2019	2019
Total emissions ($E_{FOS} + E_{LUC}$)							
Fossil CO_2 emissions (E_{FF})	3.0 ± 0.2	4.7 ± 0.2	5.4 ± 0.3	6.3 ± 0.3	7.7 ± 0.4	9.4 ± 0.5	9.7 ± 0.5
Land use change CO_2 emissions (E_{LUC})	1.5 ± 0.7	1.3 ± 0.7	1.3 ± 0.7	1.4 ± 0.7	1.4 ± 0.7	1.6 ± 0.7	1.8 ± 0.7
Total emissions	4.5 ± 0.7	5.9 ± 0.7	6.7 ± 0.8	7.6 ± 0.8	9.1 ± 0.8	10.9 ± 0.9	11.5 ± 0.9
Partitioning							
Growth rate in atmospheric CO_2 concentration (G_{ATM})	1.8 ± 0.07	2.8 ± 0.07	3.4 ± 0.02	3.2 ± 0.02	4.1 ± 0.02	5.1 ± 0.02	5.1 ± 0.2
Ocean sink (S_{OCEAN})	1.0 ± 0.6	1.3 ± 0.4	1.7 ± 0.4	2.0 ± 0.5	2.1 ± 0.6	2.5 ± 0.6	2.6 ± 0.6
Terrestrial sink (S_{LAND})	1.3 ± 0.4	2.1 ± 0.4	2.0 ± 0.7	2.6 ± 0.7	2.9 ± 0.6	3.4 ± 0.9	3.1 ± 1.2
Budget imbalance							
BIM = $E_{FOC} + E_{LUC} – (G_{ATM} + S_{OCEAN} + S_{LAND})$	0.5	-0.2	-0.4	0.1	0	-0.1	0.3

Section S2. The Calvin Cycle.[9]

The catalytic cycle is initiated by Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) enzyme that fixes CO_2 into ribulose-1,5-disphosphate (the CO_2-acceptor) with formation of two mol of 3-phosphoglycerate. The latter is further converted into glyceraldeide-3-phosphate and, subsequently, in ribulose-5-phosphate and ribulose-1,5-disphosphate, with regeneration of the CO_2-acceptor.

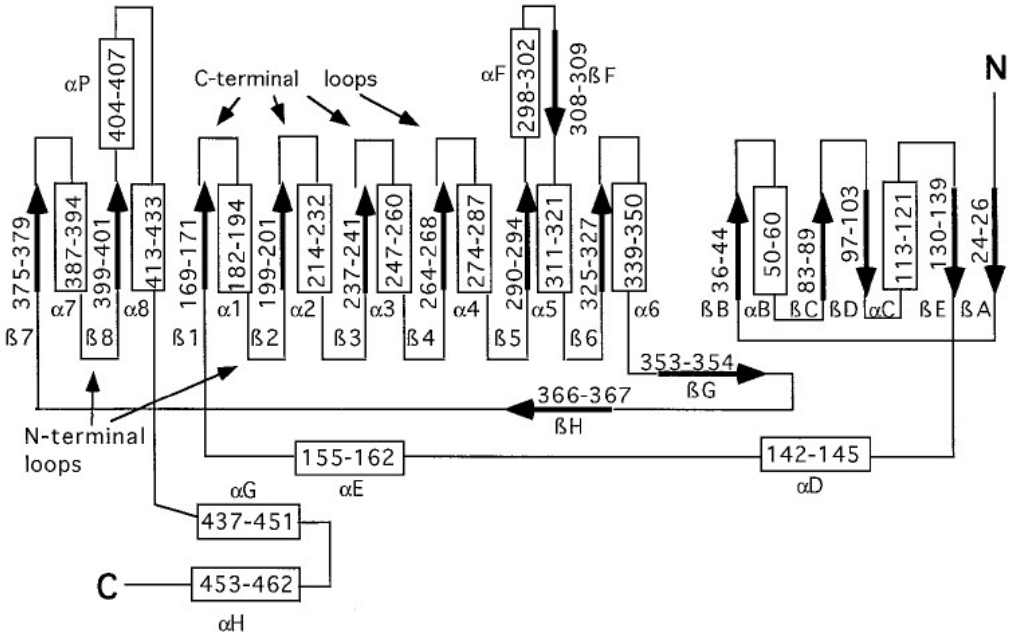
The neat overall reaction is the well-known Eq. 2 with production of carbohydrates.

$$CO_2 + 4\ H + n\ ATP = (CH_2O) + H_2O + nADP + nPi \qquad \text{Eq. 2}$$



Scheme S1. A schematic representation of the Calvin-Benson-Bassham (CBB) cycle active in photosynthetic organisms. (Scheme S1 from ref. 9, used by permission of John Wiley and Sons (copyright 2010).

Section S3. Diagram showing the secondary structure elements of the large subunit of Rubisco from spinach.[10]



Scheme S2. Connectivity diagram showing the secondary structure of the large subunit of Rubisco. Rectangles indicate α-helices, arrows indicate β-strands; numbering of helices and strands follows Knight et al. (1990). Helices, strands and amino acid residues are numbered according to Knight and co-workers (Knight 1990). Numbers indicate amino acids included in helices or strands. Only some of the C-terminal and N-terminal loops are labeled. (Scheme S2 from ref. 10, used by permission of John Wiley and Sons (copyright 1997).

Section S4. Primary and secondary structure elements of the S subunit from several organisms.[11]

The secondary structure elements of the S subunit are the following:

αA – βA – βAβB loop –βB – αB – βC – βD – βE – βF

Figure S2 Small subunit sequences aligned according to Rubisco x-ray crystal structures. The sequences are from *C. reinhardtii* (*CR*), *Spinacea oleracea* (*SO*), *Nicotiana tabacum* (*NT*), *Pisum sativum* (*PS*), *Synechococcus* (*SY*), *Anabaena* (*AB*), *Alcaligenes eutrophus* (*AE*), and *Galdieria partita* (*GP*). Residues that comprise α-helixes A and B are colored *dark gray*, and those that comprise β-strands A through F are colored *light gray*. (Figure S2 from ref. 11, used without changes from original source distributed under the CC-BY licence, https://www.sciencedirect.com/science/article/pii/S0021925820380650?via%3Dihub).

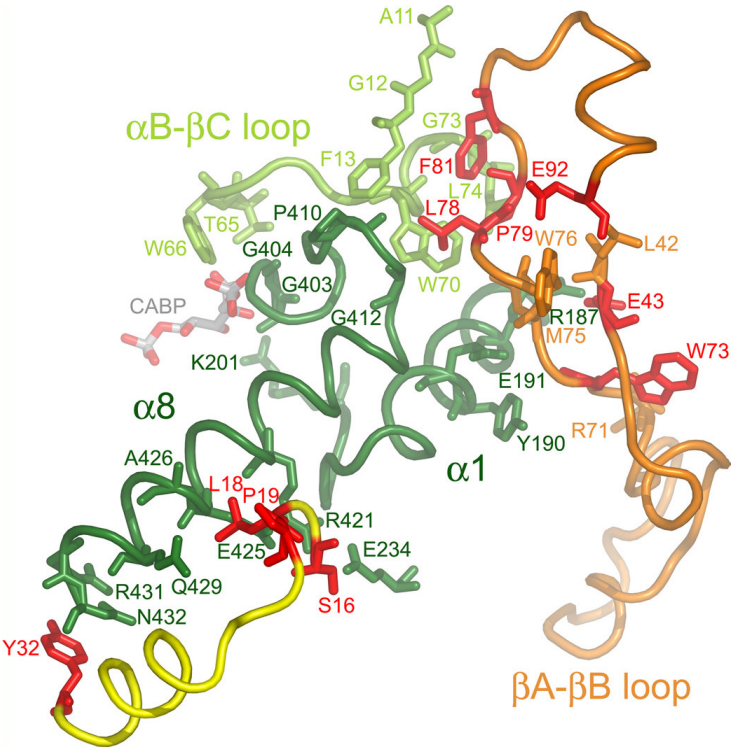


Figure S3[11] Ribbon representation of the secondary structure elements of S subunit of *Chlamydomonas* Rubisco (Protein Data Bank code 1GK8, evidenced in orange color) as plotted from the x-ray crystal structure deposited at PDB. Secondary structure elements of the L subunit functionally related to the S subunit are evidenced in green and yellow color. Residues conserved in greater than 95% of all known small subunit sequences are colored *red.* ((Figure S3 from ref. 11, used without changes from original source distributed under the CC-BY licence, https://www.sciencedirect.com/science/article/pii/S0021925820380650?via%3Dihub).

Section 5. Cumulative CO_2 and O_2 density around the active site of Rubisco as modelled by M. Van Lun et al, 2014.[12]

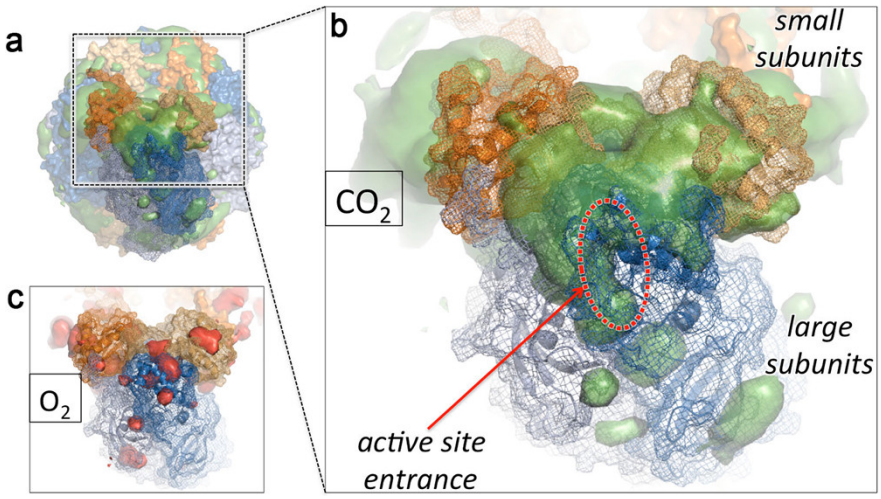


Figure S4. Cumulative CO_2 and O_2 density around the active site of Rubisco. (a) Hexadecameric enzyme (*C. reinhardtii* Rubisco closed complex, PDB code 1gk8) with large subunits in blue and small subunits in orange. The highest density of CO_2 after 20 ns simulation is shown as a green smooth surface. (b) Close-up around the active site with residues from large (blue) and small (orange) subunits
represented as a mesh. CO_2 density in green has a direct connection between the large and small subunits that may aid in attracting and guiding CO_2 toward the active site. The active site area (circled in red) is covered by flexible loops. (c) Cumulative O_2 density (red) after 20 ns simulation. The orientation is the same as in (b), and the simulation was run at the same concentration as for CO_2. (Figure S4 from ref. 12, used by permission of America Chemical Society (copyright 2014).

Section S6. Multistep mechanism for the Rubisco carboxylation and oxygenation reactions including microscopic rate constants associated to each elementary step as proposed by Cummins and Gready in 2018.[13]

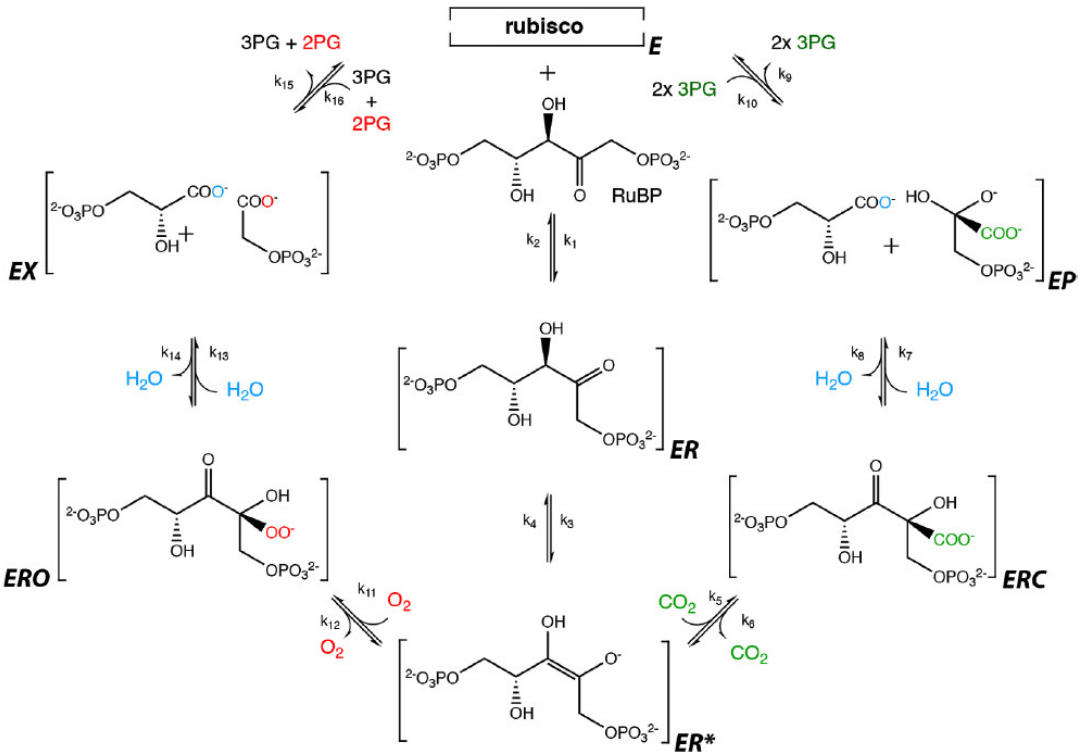


Figure S5. The mechanism of RuBP enolization (central steps), with description of the RuBP complete oxygenation (lefts side) and complete carboxylation (right side) reactions following the nomenclature of Cummins et al. 2018. (Figure taken from Ref. 19)
E denotes Rubisco in the active state (mentioned as ECM complex in Fig. 4)
ER denotes the Lys-$COO^{-}Mg^{2+}$-RuBP complex (mentioned as ECMR in Fig. 4)
ER* denotes the Lys-$COO^{-}Mg^{2+}$-RuBP* complex after RuBP enolization.

Carboxylation reaction (right side).
ERC the complex formed after CO_2 bind the ER* complex.
EP state: hydration and cleavage of the ERC complex leads to the formation of two enzyme-bound 3-phosphoglycerate molecules (3PG) in the EP state, each of which have 3 carbon atoms.
Release of two 3PG from the **EP** state.

Oxygenation reaction (left side).
ERO the complex formed after O_2 binding to the ER* complex.
EX state: hydration and cleavage of the ERO complex produces enzyme-bound 3PG and one 2-phosphoglycolate (2PG). 2PG has two carbon atoms and is not part of the CBB cycle.
Release of 3PG and 2PG from the **EX** state.

(Figure S5 from ref. 13, used without changes from original source distributed under the CC-BY licence, https://www.frontiersin.org/articles/10.3389/fpls.2018.00183/full)

Table S3. Definition of Rubisco kinetic parameters (according to Milo, 2019).[14]

The carboxylation reaction displays effective Michaelis−Menten kinetics (maximum catalytic rate $k_{cat,C}$, half-maximum CO_2 concentration $K_M = K_C$).[20]

The oxygenation reaction also displays effective Michaelis−Menten kinetics ($k_{cat,O}$, $K_M = K_O$, half-maximum inhibitory CO_2 concentration $K_I = K_C$).

$k_{cat,C}$	Is the maximum rate per-active site carboxylation rate (also named turnover number) defined as: mol of CO_2 fixed into products/mol of active sites s (measured in units of s^{-1}) or number of CO_2 molecules fixed into products/active site s.
$k_{cat,O}$	Is the maximum rate per-active site oxygenation rate (also named turnover number) defined as: mol of O_2 fixed into products /mol of active sites s (measured in units of s^{-1}). or number of O_2 molecules fixed into products/active site s.
K_C	Effective Michaelis constant (half-saturation concentrations in μM units) for CO_2
K_O	Effective Michaelis constant (half-saturation concentrations in μM units) for O_2
$R_C = \frac{k_{cat,C}}{1 + \frac{K_C}{[CO_2]} + \frac{K_C\,[CO_2]}{K_O / [O_2]}}$	R_C is the per-active site rate of carboxylation at saturating RuBP with competitive inhibition by O_2 (assuming half-maximum inhibitory O_2 concentration Ki = K_O). R_C is strongly influenced by $[CO_2]$ and $[O_2]$ at the Rubisco active site. Rates of carboxylation (R_C) are calculated from kinetic parameters and the CO_2 and O_2 concentrations.

$R_O = \frac{k_{cat,O}}{1 + \frac{K_O}{[O_2]} + \frac{K_O\,[O_2]}{K_C / [CO_2]}}$	R_O is the per-active site rate of oxygenation.at saturating RuBP with competitive inhibition by CO_2 (asuming half-maximum inhibitory CO_2 concentration $K_I = K_C$). R_O is strongly influenced by $[CO_2]$ and $[O_2]$ at the Rubisco active site. Rates of oxygenation (R_O) are calculated from kinetic parameters and the CO_2 and O_2 concentrations.
$k_{cat,C}/K_C$	Catalytic efficiency for carboxylation. Can be viewed as the rate of productive capture of the substrate (i.e. leading to complete catalysis) (non so bad)
$k_{cat,O}/K_O$	Catalytic efficiency for oxygenation. Can be viewed as the rate of productive capture of the substrate (i.e. leading to complete catalysis). (non so bad)
V_C	When RuBP is saturating, per-active site enzyme rates can be multiplied by the concentration of active Rubisco ([**E**]) to calculate the total rates of carboxylation V_C. $V_C = k_{cat,C}\,[\mathbf{E}]$
V_O	When RuBP is saturating, per-active site enzyme rates can be multiplied by the concentration of active Rubisco ([**E**]) to calculate the total rates of carboxylation V_O. $V_O = k_{cat,O}\,[\mathbf{E}]$
By definition: $S_{C/O} = \left(\frac{k_{cat,C}}{K_C}\right) / \left(\frac{k_{cat,O}}{K_O}\right)$	Sc/o is the specificity factor Unitless measure of the Rubisco relative preference for CO_2 over O_2
$\frac{R_C}{R_O} = S_{C/O}\,\frac{[CO_2]}{[O_2]}$	To measure $S_{C/O}$, the rate of O_2 and CO_2 incorporation are measured simultaneously in the same sealed assay vessel. Plotting R_C / R_O against $[CO_2]/[O_2]$ gives $S_{C/O}$ as the slope of linear fit passing through $R_C/R_O = 0$. According to the previous equation after measuring $S_{C/O}$ and $k_{cat,C}$ it can be calculated $k_{cat,O.}$

Reference list.

1) P. Friedlingstein et al.: Global Carbon Budget 2020, *Earth Syst. Sci. Data*, **2020**, *12*, 3269–3340.

2) J. Pongratz, C. H. Reick, T. Raddatz and M. Claussen Effects of anthropogenic land cover change on the carbon cycle of the last millennium, Global Biogeochem. Cy., **2009**, 23, GB4001, https://doi.org/10.1029/2009GB003488.

3) E. Shevliakova, S.W. Pacala, S. Malyshev, G.C. Hurtt, P.C.D. Milly, J. P. Caspersen, L.T. Sentman, J.P. Fisk, C. Wirth and C. Crevoisier, Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, Global Biogeochem. Cy**., 2009**, 23, GB2022, https://doi.org/10.1029/2007GB003176.

4) S. Zaehle, P. Ciais, A.D. Friend and V. Prieur, Carbon benefits of anthropogenic reactive nitrogen offset by nitrous oxide emissions, Nat. Geosci., **2011**, 4, 601–605, https://doi.org/10.1038/ngeo1207.

5) J. G. Van Minnen, K. Klein Goldewijk, E. Stehfest, B. Eickhout, G. van Drecht and R. Leemans, The importance of three centuries of land-use change for the global and regional terrestrial carbon cycle, Climatic Change, **2009**, 97, 123–144, https://doi.org/10.1007/s10584-009-9596-0.

6) R.A. Houghton and A. A. Nassikas, Global and regional fluxes of carbon from land use and land cover change 1850–2015, Global Biogeochem. Cy., **2017**, 31, 456–472, https://doi.org/10.1002/2016gb005546.

7) E. Hansis, S. J. Davis and J. Pongratz, Relevance of methodological choices for accounting of land use change carbon fluxes, Global Biogeochem. Cy., **2015**, 29, 1230–1246, https://doi.org/10.1002/2014GB004997.

8) K. L. Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., Leite da Silva Dias, P., Wofsy, S. C., and Zhang, X.: Couplings Between Changes in the Climate System and Biogeochemistry, in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Marquis, M., Averyt, K., Tignor, M. M. B., Miller, H. L., and Chen, Z. L., Cambridge University Press, Cambridge, UK and New York, USA, 2007, 499–587.

9) Ivan A. Berg, Daniel Kockelkorn, W. Hugo Ramos-Vera, Rafael Say, Jan Zarzycki, in "Carbon Dioxide as Chemical Feedstock", Ed. Prof. Michele Aresta, 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Chapter 3, 33-53.

10) E. A. Kellogg and N. D. Juliano, American Journal of Botany, **1997**, 84, 413–428.

11) T. Genkov and R. J. Spreitzer, THE JOURNAL OF BIOLOGICAL CHEMISTRY, **2009**, 284, 44, 30105–30112.

12) M. van Lun, J. S. Hub, D. van der Spoel and I. Andersson, J. Am. Chem. Soc. **2014**, 136, 3165−3171.

13) P. L. Cummins, B. Kannappan and J. E. Gready, Front. Plant Sci. 2018, 9, 183.

14) A. I. Flamholz, N. Prywes, U. Moran, D. Davidi, Y. M. Bar-On, L. M. Oltrogge, R. Alves, D. Savage and R. Milo, *Biochemistry,* **2019**, 58, 3365–3376