Modeling With Uncertainty Quantification Identifies Essential Features of a Non-Canonical Algal Carbon-Concentrating Mechanism

Anne K. Steensma [1,2 a], Joshua A.M. Kaste [3 a], Junoh Heo [4], Douglas J. Orr [5], Chih-Li Sung [4], Yair Shachar-Hill [1], Berkley J. Walker [1,2 *]

1. Department of Plant Biology, Michigan State University, East Lansing, MI, 48824
2. Michigan State University – Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, MI, 48824
3. Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48824
4. Department of Statistics and Probability, Michigan State University, East Lansing, MI, 48824
5. Lancaster Environment Center, Lancaster University, Lancaster, United Kingdom, LA1 4YQ

Corresponding author

* denotes corresponding author in author line
Berkley J. Walker
Room 106 Plant Biology Laboratories, 612 Wilson Road, East Lansing, MI, 48824
(517) 355-3928
berkley@msu.edu

---

[a] Equal contributions: AKS and JAMK

Abstract

The thermoacidophilic red alga *Cyanidioschyzon merolae* survives its challenging environment likely in part by operating a carbon-concentrating mechanism (CCM). Here, we demonstrated that *C. merolae*'s cellular affinity for CO_2 is stronger than its rubisco affinity for CO_2. This provided further evidence that *C. merolae* operates a CCM while lacking structures and functions characteristic of CCMs in other organisms. To test how such a CCM could function, we created a mathematical compartmental model of a simple CCM distinct from those previously described in detail. The results supported the feasibility of this proposed minimal and non-canonical CCM in *C. merolae.* To facilitate robust modeling of this process, we incorporated new physiological and enzymatic data into the model, and we additionally trained a surrogate machine-learning model to emulate the mechanistic model and characterized the effects of model parameters on key outputs. This parameter exploration enabled us to identify model features that influenced whether the model met experimentally-derived criteria for functional carbon-concentration and efficient energy usage. Such parameters included cytosolic pH, bicarbonate pumping cost and kinetics, cell radius, carboxylation velocity, number of thylakoid membranes, and CO_2 membrane permeability. Our exploration thus suggested that a novel CCM could exist in *C. merolae* and illuminated essential features necessary for CCMs to function.

Significance

Carbon-concentrating mechanisms (CCMs) are processes which boost photosynthetic efficiency. By developing modeling approaches to robustly describe CCMs in organisms where biochemical data is limited, such as extremophile algae, we can better understand how organisms survive environmental challenges. We demonstrate an interdisciplinary modeling approach which efficiently sampled from large parameter spaces and identified features (e.g., compartment permeability, pH, enzyme characteristics) which determine the function and energy cost of a simple CCM. This approach is new to compartmental photosynthetic modeling, and could facilitate effective use of models to inform experiments and rational engineering. For example, engineering CCMs into crops may improve agricultural productivity, and could benefit from models defining the structural and biochemical features necessary for CCM function.

Introduction

Cyanidioschyzon merolae is a red microalga found in moist environments surrounding geothermal sulfur springs. This species is extremophilic, with optimal laboratory growth conditions including low pH (~ 2) and high temperatures (~ 42 °C) (1, 2). *C. merolae* and other thermo-acidophilic red algae draw interest for their unique biology and simple characteristics, which position them as useful model organisms and as candidates for biotechnology applications (3–6). For example, *C. merolae* is of interest because it is one of few organisms which relies on photosynthesis in geothermal spring environments, where hot and acidic conditions restrict the availability of inorganic carbon and challenge biological carbon fixation (1, 7). Notably, organisms of acid waters can only access approximately 10 micromolar inorganic carbon, as the inorganic carbon pool at acid pH is primarily the volatile species CO_2. In comparison, organisms of near-neutral and alkaline waters may have access to several millimolar of inorganic carbon, due to accumulation of the involatile bicarbonate (8).

C. merolae is thought to survive in its challenging environment in part by operating a carbon-concentrating mechanism (CCM) (9–11). CCMs boost carbon-fixation efficiency by concentrating CO_2 around rubisco, providing ample substrate for carbon-fixation and inhibiting a competing oxygen-fixation reaction of rubisco. Evidence supporting a CCM in *C. merolae* includes measured accumulation of carbon in the cell, $\delta^{13}C$ consistent with a CCM, similar growth rates under ambient and elevated CO_2, transcriptional response of potential CCM genes to CO_2 fluctuations, and substantial CO_2 assimilation at low environmental CO_2 concentrations (9–12). However, many of these indications of the CCM are not definitive: in particular, it is not known how much of *C. merolae*'s ability to assimilate CO_2 efficiently could be explained by the affinity of *C. merolae* rubisco for CO_2. Thus, we here provide further evidence for the CCM in *C. merolae* by demonstrating that the affinity of *C. merolae* cells for CO_2 is better than could be explained by the affinity of *C. merolae* rubisco for CO_2.

C. merolae's CCM may be described as a "novel" or "non-canonical" CCM, as the *C. merolae* CCM must operate differently from the few CCMs which are well-characterized. Unlike algae and cyanobacteria with well-characterized CCMs, *C. merolae* is not able to take up external bicarbonate, and *C. merolae* lacks anatomy associated with the pyrenoid CCM organelle (10, 11, 13, 14). The absence of these CCM features in *C. merolae* challenges our understanding of how algal CCMs work, and presents the opportunity to define essential CCM components. We thus used mathematical modeling, informed by new experimental measurements, to explore how the *C. merolae* CCM may function.

Research on CCMs has long employed mathematical models to understand the components of functional CCMs in model cyanobacteria and algae. A particular area of interest in CCM modeling is the possibility of boosting crop productivity by engineering CCMs into crops which lack CCMs (15–18). We sought to add to the inspiration for these engineering efforts by modeling a heat-tolerant CCM with minimal components which offers unique possibilities for plant engineering (19). To draw robust conclusions about cellular characteristics which can support a CCM, we used state-of-the-art statistical methods to define the effects of model parameters on the predicted photosynthetic phenotype while limiting unwarranted *a priori* assumptions.

Some sets of model input parameters produced model outputs which met empirically-based criteria for functional carbon concentration and efficient energy usage, and we identified input parameters which have substantial impacts on the model outputs. Overall, our model of a hypothetical biophysical CCM which requires minimal enzymes and anatomical features (**Figure**

1) appears to represent a feasible CCM structure in *C. merolae*, which invites further research into the sources of environmental resilience in extremophile algae.

Methods

Experimental data collection

Extraction, purification, and kinetic assays of *C. merolae* rubisco; and measurement of gas-exchange parameters by open-path infra-red gas analysis, were as detailed in **Supplemental Methods.**

Model details

The hypothetical CCM described in this study (**Figure 1**) was modeled as a set of well-mixed compartments and represented as a system of ordinary differential equations (ODEs). In this minimal biophysical CCM, carbon diffuses into the cell as CO_2, is trapped in the cytosol as bicarbonate by action of carbonic anhydrase, and is pumped into the chloroplast, where a second carbonic anhydrase provides CO_2 around rubisco. No pyrenoid diffusion barrier is present, though we accounted for potential effects of the concentric thylakoids which are present in *C. merolae* and many other aquatic photosynthetic organisms (22).

The model geometry is based on the cellular structure of *C. merolae* as apparent in published micrographs of this alga (22–31). The modeled cell and its boundary layer form a series of concentric spherical well-mixed compartments. The cell is enclosed by a lipid bilayer of radius $Radius_{cell}$. The cell is surrounded by a medium boundary layer of radius $2 * Radius_{cell}$, beyond which lies an infinite external medium. The cell contains a cytosol of radius $Radius_{cell}$ and a chloroplast stroma space of radius $0.25 * Radius_{cell}$.

Molecules cross the boundary of the stroma space according to diffusion or transport equations. For flux calculations, the boundary consists of 1 to 7 lipid bilayers of negligible thickness that are evenly spaced from $0.5 * Radius_{cell}$ to $0.25 * Radius_{cell}$. This boundary structure represents the fact that the *C. merolae* chloroplast is surrounded by a chloroplast envelope and by approximately 4 to 6 thylakoids which appear as concentric circles or spirals in thin-section microscopy (22). A range of possible transport scenarios (how many membranes molecules must cross when crossing between the cytosol and stroma, and how much energy this crossing costs) are captured by varying parameters *Membranes* and $Pump_{cost}$.

Diffusion through lipid membranes was described using estimates of conductivity of lipid membranes to the chemical species in question:

$$J_{membrane\ diffusion} = Conductivity_X * ([X]_A - [X]_B) \qquad \textbf{(E1)}$$

Where $Conductivity_X$ is the conductivity – in units of $\mu m^3/s$ – of chemical species X through a lipid bilayer, and $[X]_A$ and $[X]_B$ are the concentrations of that species on the two sides of that lipid bilayer. Diffusion into or out of the medium boundary layer was described as an analogous simple diffusion flux, with conductivity determined according to diffusion coefficients through water at the boundary layer thickness. Lipid permeability coefficients for CO_2 and HCO_3^- and the water diffusion coefficient for O_2 were sourced from the literature (**Table S1**), and other necessary gas permeability and diffusion coefficients were determined from the literature values by Graham's law of diffusion:

$$\frac{r_1}{r_2} = \sqrt{\frac{M_1}{M_2}} \qquad \textbf{(}\boldsymbol{E2}\textbf{)}$$

Where the rates of diffusion r_1 and r_2 for two different ideal gases, here CO_2 and O_2, are related according to their two molar masses M_1 and M_2.

To describe diffusion of CO_2, HCO_3^-, and O_2 through variable numbers of stacked thylakoid membranes, an overall conductivity through all of the layers was calculated as:

$$Overall\ Conductivity = \left(\sum_{i=1}^{n}(4\pi r_n^2 * Conductivity_X)^{-1}\right)^{-1} \qquad \textbf{(}\boldsymbol{E3}\textbf{)}$$

Where r_n is the radius of the sphere formed by the nth thylakoid membrane. This overall conductivity value is then used in **(E1)** to describe the movement of a chemical species from the outer stroma into the inner stroma space, as shown in **Figure 1**. We assume that small gas molecules diffuse easily around membrane proteins, so that the diffusion of CO_2 and O_2 through any modeled membrane is potentially impeded by increased path length, but is not impeded by CO_2 and O_2 passing through high-resistance protein material.

Spontaneous interconversion of CO_2 and HCO_3^- **(E4-5)** was described using simple first-order kinetics, according to the rate constant of the dehydration (slower) step of the interconversion:

$$J_{CO_2\ hydration} = k_2[CO_2] \qquad \textbf{(}\boldsymbol{E4}\textbf{)}$$

$$J_{HCO_3^-\ dehydration} = k_{-2}[HCO_3^-][H^+] \qquad \textbf{(}\boldsymbol{E5}\textbf{)}$$

Note that CO_2 must first be hydrated to H_2CO_3, which is then deprotonated to yield the HCO_3^- ion. However, because the interconversion of HCO_3^- and H_2CO_3 is essentially instantaneous relative to the hydration-dehydration reaction, here we ignore the H_2CO_3 species and approximate the spontaneous interconversion as the hydration-dehydration reaction.

The interconversion of CO_2 and HCO_3^- by carbonic anhydrase was described as in (16):

$$J_{CA} = \frac{[CA] * CA_{kcat} * \left([CO_2] - \frac{[HCO_3^-][H^+]}{K_a}\right)}{K_m^{CO_2} + [HCO_3]\left(\frac{K_m^{CO_2}}{K_m^{HCO_3^-}}\right) + [CO_2]} \qquad \textbf{(}\boldsymbol{E6}\textbf{)}$$

Where the K_a value is the overall K_a for the CO_2/HCO_3^- system. This value is temperature-sensitive and was calculated using the R package *seacarb* package (32). Other potentially temperature-sensitive parameters receive temperature adjustments according to Q_{10} or Q_{15} factors.

Carboxylation by rubisco was described as with the assumption that CO_2 is limiting, as in (33):

$$v_c = \frac{Vmax_{carboxylation}[CO_2]}{\left([CO_2] + K_m^{CO_2}\left(1 + \frac{[O_2]}{K_m^{O_2}}\right)\right)} \qquad (\boldsymbol{E7})$$

To estimate oxygenation, we estimate v_c/v_o (carboxylation flux over oxygenation flux) from the CO_2/O_2 specificity ($S_{c/o}$) of rubisco and chloroplast CO_2 and O_2 concentrations **(E8)**, and then use this to arrive at v_o.

$$\frac{v_c}{v_o} = S_{co}\left(\frac{[CO_2]}{[O_2]}\right) \qquad (\boldsymbol{E8})$$

The pumping of HCO_3^- across the stack of thylakoid membranes by a bicarbonate pump was described by simple Michaelis-Menten kinetics:

$$J_{HCO_3^-\,pump} = \frac{V_{max}[HCO_3^-]}{K_m + [HCO_3^-]} \qquad (\boldsymbol{E9})$$

Respiration in the light (R_L) was estimated from experimental data according to a modified Kok method, by measuring under sub-saturating light intensities and extrapolating CO_2 release in the absence of light (**Figure 2**). The mean measured value of R_L was normalized to cell size for use in the model: we assume that the empirical measurement of R_L we obtained was, on a per cell basis, characteristic of a *C. merolae* cell of a radius of 1 µm. Under the assumption that R_L should vary proportionally with cell volume, we normalized R_L as follows:

$$R_{L_{normalized}} = R_{L_{measured}} \frac{Volume}{Volume_{1um}} \qquad (\boldsymbol{E10})$$

ATP costs for the cell were estimated as:

$$ATP_{total} = 3v_c + 3.5v_o + \left(J_{HCO_3^-\,pump} * Membranes * Pump_{cost}\right) \qquad (\boldsymbol{E11})$$

Where *Membranes* is the number of thylakoid stacks and $Pump_{cost}$ is the assumed cost, in ATP, of pumping a single HCO_3^- ion across a lipid bilayer by the hypothesized pump.

A full list of all flux equations and the system of ODEs used to describe the system can be found in **Supplemental Materials**.

Definition of reasonable model output values

To ensure the model reproduced experimental results, we used new and published experimental data to set acceptable bounds for the following model outputs: CO_2 compensation point (Γ_{CO2}), the ratio of ATP consumption flux to net CO_2 assimilation flux (ATP per CO_2), the steady-state CO_2 concentration in the chloroplast stroma (stromal CO_2), and the ratio of oxygen-fixation flux to carbon-fixation flux (v_o/v_c). Selection and justification of these bounds are detailed in **Supplemental Methods**.

Model optimization and estimation of simulated compensation point

Steady-state fluxes and metabolite concentrations were solved using *odeint()* from Python's SciPy library (34). Latin hypercube parameter sampling (35) and curve-fitting to generate compensation point estimates were as detailed in **Supplemental Methods**.

Parameter exploration and surrogate model selection

In order to thoroughly explore the 19-dimensional parameter space in a computationally-feasible way, we trained a surrogate machine-learning model on the mechanistic CCM model. By emulating the intricacies of the mechanistic model, surrogate modeling faithfully captures dynamics of complex systems while alleviating the substantial computational costs associated with obtaining results. Surrogate modeling additionally gave us access to powerful statistical tools for machine-learning model analysis, including SHapley Additive exPlanations (SHAP) (36) and partial dependence (PD) plots (37).

To identify the optimal surrogate model for parameter exploration, we compared four popular machine-learning models: eXtreme Gradient Boosting (XGBoost) (38), Local approximate Gaussian Process (laGP) (39), single-layer Neural Network (NN) (40), and Deep Neural Network (DNN) (38). We collected a 240,000-sized dataset, where the outputs were simulated from the CCM model at space-filling input locations. 90% of the data was used for training the surrogate, and the remaining 10% was used as the test dataset to validate the model performance. The evaluation of prediction performance was based on the root-mean-square error (RMSE):

$$RMSE = \sqrt{\sum_{i=1}^{n_{test}} \frac{(y_i - \hat{y}_i)^2}{n_{test}}},$$

where y_i is the i-th test output and $\hat{y}_i$ is the i-th predicted model output.

Model outputs had varying scales and degrees of skew, so to effectively compare prediction performance on different model outputs, a normalized RMSE (NRMSE) was calculated. The NRMSE was calculated as the RMSE divided by $y_{max} - y_{min}$, where y_{max} is the highest test output and y_{min} is the lowest test output.

From the model evaluation (**Table S2**), it appears that XGBoost outperformed other models for v_o/v_c and ATP per CO_2, and remained comparable for Γ_{CO2} and stromal CO_2. As such, XGBoost was used as the surrogate model for further analyses.

Results and Discussion

Rubisco kinetics demonstrated that *C. merolae* operates a CCM

In previous work, we determine that if *C. merolae* has rubisco kinetics similar to other red algae, then this alga must operate a CCM to maintain its measured photosynthetic efficiency. Alternatively, its measured photosynthetic efficiency could be explained by unprecedented rubisco kinetics, meaning enzyme properties favoring carbon-fixation over oxygen-fixation to an unprecedented degree (10). Here we confirmed that *C. merolae* rubisco kinetics are similar to those of other red-type (Form 1D) rubiscos (41–43). *C. merolae* rubisco had a strong affinity for CO_2

(low K_C), a poor affinity for O_2 (high K_O), and a slow carboxylation rate (low $kcat_C$) (**Figure 2**). Consistent with other studies, $kcat_C$ and K_C were higher when measured at increased temperature, while K_O was lower. Although K_O is a component of rubisco specificity ($S_{c/o}$) and $S_{c/o}$ decreases with increased temperature, *in vitro* K_O is observed to decrease with increased assay temperature in some species (42, 44, 45).

These kinetics findings indicated *C. merolae* does operate a CCM, as *C. merolae* cells had higher affinity for CO_2 than *C. merolae* rubisco (8.71 ± 1.7 µM cell K_C vs. 24.9 ± 3.2 µM rubisco K_C at 45 °C, $p = 0.008$ by two-sample t-test) (**Figure 2**). This result adds to the indications of the CCM in *C. merolae* (9–11).

Quantitative modeling showed that a hypothesized CCM can explain *C. merolae*'s carbon-concentrating behavior

To explore how the *C. merolae* CCM may operate, we constructed a functional model of a CCM (**Figure 1**). This model demonstrated that there were parameter sets consistent with the empirical literature that result in a functional CCM, despite the minimal model structure (**Figure 3**). Our results provided quantitative support for a CCM taking inorganic carbon from the environment solely through CO_2 diffusion into the cell, which we term a "non-canonical" or "novel" CCM due to its differences in structure and function from CCMs that have been characterized in detail. Though there is speculation that extremophilic red algae may use a C_4-like CCM, it has been previously proposed that acidophile algae may accumulate carbon by a "bicarbonate-trap" or "acid-loading" mechanism similar to our modeled CCM (7, 12, 15, 46, 47). Briefly, bicarbonate would be concentrated for enzymatic action by bringing inorganic carbon speciation near equilibrium in near-neutral cellular compartments, since the predominant inorganic carbon species from pH ~6 to ~10 is the poorly-membrane-permeable bicarbonate.

We used two strategies to deeply explore the model parameter space and ensure that our conclusions were robust. First, the model included new experimental data on gas-exchange and rubisco parameters central to photosynthetic efficiency (**Figure 2**). Second, we developed a method for thoroughly assessing the model's sensitivity to the value of model parameters of interest. Specifically, we were interested in 19 of the 43 model parameters which were biologically interesting in relation to the function of a novel CCM and which were not well-characterized physical constants (**Table S1**). We thus sampled input parameter sets through a Latin hypercube design (35). Latin hypercube sampling enhanced analysis accuracy by mitigating sampling bias, as it produced parameter sets distributed throughout the 19-dimensional parameter space of interest. Then, each input parameter set was used to parameterize the model and to generate a set of outputs for analysis.

Some of the input parameter sets produced outputs consistent with a functional CCM with reasonable energy cost. Of particular interest were the parameter sets which met all the empirically-based criteria for a realistic and functional CCM (criteria selection described in **Supplemental Methods**). 13,998 of 240,000 (6%) of parameter sets fulfilled the two competing objectives of functional carbon concentration (corresponding to outputs of low Γ_{CO2}, high stromal CO_2, and low v_o/v_c) and efficient energy usage (corresponding to output of low ATP per CO_2) (**Figure 2**).

The generated parameter sets allowed us to explore the trade-offs associated with various features related to the CCM. For example, adding additional concentric thylakoids slightly improved carbon concentration by presenting barriers to CO_2 leakage out of the chloroplast, but incurred additional energy costs (**Figure 4, Figures S1 – S2**). This is consistent with other modeling studies indicating that thylakoid membranes could affect inorganic carbon diffusion (15, 48).

Machine-learning-based surrogate models identified the parameters that most influence CCM efficiency

Like most mathematical models of photosynthetic systems, this model faced the challenge of drawing robust conclusions while using parameters which, although bounded by their relationship to physical processes, have substantial uncertainty (**Table S1**). To model a system with limited biochemical data while not constraining input parameters to a greater degree than was supported by the literature, it was important to assess uncertainties which seemed likely to have substantial and interdependent effects on the model. For example, the input parameter describing permeability of a lipid bilayer to CO_2 ($Plip_{CO2}$) has reported values ranging over several orders of magnitude (**Table S1**). Furthermore, the effect of $Plip_{CO2}$ in the model depended on the value of other parameters, such as the number of lipid bilayers which pose a barrier to carbon moving between the stroma and cytosol (*Membranes*). $Plip_{CO2}$ and similar parameters were unlikely to be satisfactorily explored by classical local sensitivity analyses, which involve tracking model outputs when individual parameters are varied by a set fraction of the parameter's original value. Therefore, to reveal which model conditions were necessary for the modeled CCM and to identify interesting directions for future investigation, we used statistical methods to identify impactful parameters and to identify which input spaces corresponded to target output ranges. These statistical methods involved training a surrogate machine-learning model on our CCM model inputs and outputs. Interpretations of this surrogate model identified which zones in the input parameter space contained the most combinations fulfilling output criteria **(Figure 5 lower left)**, quantified how much each input parameter affected the prediction of outputs by the surrogate model (**Figure 5 upper right**), and visualized the response of model outputs to inputs (**Figures S4 – S7).**

Some input parameters had little impact on model outputs. For these parameters, values from across the input range were evenly represented in the parameter sets meeting all output criteria, which is reassuring for future modeling and engineering efforts that involve these features. The parameters with relatively little impact on outputs included values related to carbonic anhydrase concentration and kinetics (*[CA]*, *CAkcat*, Km_{CO2} and Km_{HCO3-} for carbonic anhydrases), chloroplast pH, and values related to bicarbonate membrane permeability ($Plip_{HCO3-}$, $Q10_{PlipHCO3-}$, **Figure 5, Figures S4 – S8**).

Other parameters were more constraining in the model, indicating their importance in producing a functional CCM. For example, six parameters appeared to impact all four of the target model outputs in the mean absolute SHAP plots: V_c, $Vmax_{pump}$, Km_{pump}, pH in the cytosol, $PlipCO_2$, and *Membranes*. As might be expected in a model relying on a cytosolic bicarbonate trap followed by bicarbonate pumping, parameter sets that successfully and efficiently concentrated carbon tended to have cytosolic pH at or above the pH where bicarbonate predominates (cytosol pH above 6), and tended to have a lower ATP cost of pumping bicarbonate (low $Pump_{cost}$), as well as faster and higher-affinity bicarbonate pumps (high $Vmax_{pump}$, low Km_{pump}).

Other features enriched in parameter sets meeting output criteria were a cell radius in the middle of the input range (moderate $Radius_{cell}$), and a lower CO_2 membrane permeability (low $Plip_{CO2}$, **Figure 5, Figures S4 – S9**). This suggested an important relationship between the volumes where metabolism occurs and the surface areas which present diffusion barriers between compartments. As the radius of the cell increases, CO_2 loss from R_L may overcome the ability of the cell to acquire carbon through passive diffusion into the cell. Conversely, as the radius of the cell decreases, less absolute bicarbonate pumping would be necessary to achieve high rubisco saturation, especially when rubisco is slow (low V_c). In low-radius scenarios, "over-pumping" bicarbonate could reduce energy efficiency.

In silico knockouts identified experimental targets for further characterization of the *C. merolae* CCM

The modeling also suggested interesting directions for investigating enzymatic components of the CCM. Alternative models with CCM enzymes removed (carbonic anhydrases or bicarbonate pumping not functional) were less likely to meet the criterion of a Γ_{CO2} indicative of functional carbon concentration, but tended to have lower ATP per CO_2 cost than the model with all enzymes present (**Figure 4, Figures S1 – S2**).

The modeled CCM functioned without fine details of cellular structure that support photosynthesis in other organisms, such as rubisco aggregation into an area smaller than the stroma, recapture of mitochondrially-respired CO_2, and perforations or interconnections in concentric thylakoids (9, 49, 50). It may still be of interest to explore whether similar structures exist in *C. merolae,* and to investigate the biochemical and molecular basis for this novel CCM.

Further applications of surrogate modeling and uncertainty quantification

More broadly, the statistical approach adopted in this paper represents an advance in metabolic and biochemical modeling. By training a surrogate model on the parameter space of mechanistic biological models, we can understand and account for high-dimensional uncertainty in model parameters. Metabolic modeling in general has been highlighted as a particularly promising application of surrogate modeling, since metabolic modeling has biotechnological potential but is challenged by the complexity of metabolism and by the "trial and error" process which is often required to produce a working metabolic model (21). Surrogate modeling has found uses in dynamic flux balance analysis and process modeling for bioprocesses (51, 52). Our work expands on these investigations by demonstrating what is to our knowledge the first application of surrogate modeling to ODE-based compartmental modeling of biological systems. Our methods may be particularly valuable for models that have poorly-defined parameters or are extremely computationally expensive. For example, the implementation of surrogate modeling described here could alleviate current limitations in interpreting reaction-diffusion models and genome-scale metabolic models (21).

Effective parameter exploration and analysis may generally be useful in confronting global challenges. Here, we used statistical sampling, surrogate modeling, and uncertainty quantification methods to investigate how aquatic organisms achieve the high photosynthetic efficiency that enables them to be responsible for approximately half of global photosynthetic CO_2 consumption (53). Similar modeling techniques may be applied effectively to any system: for example, as part of engineering efforts for bioproduction, crop resilience, and other goals, it may be useful to determine which features of a system are essential or inflexible *in silico* before devoting resources to *in vivo* experimentation.

In conclusion, the extremophilic red microalga *C. merolae* operates a CCM, as evidenced by this alga having gas-exchange behavior which was not explained by its rubisco properties. Mathematical modeling suggested that this CCM could consist of a minimal mechanism which includes thylakoid membranes as diffusion barriers. Robust parameter exploration and statistical analysis, aided by the use of a surrogate model, allowed us to quantify the sensitivity of our model to parameter uncertainties, identify important parameter interactions, and identify key determinants of CCM efficiency. Therefore, in addition to supporting the presence of a novel CCM in *C. merolae*, our results shed light on what conditions must be met for this CCM to function and the essential elements of biophysical CCMs in general.

Code availability

Model code used in this study can be accessed via GitHub: https://github.com/anne-steensma/Cmerolae_CCM_model.

Acknowledgments and funding sources

Work in the laboratory of BJW is supported by Grant Number DE-FG02-91ER20021 from the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences of the United States Department of Energy. Work in the laboratory of YSH was supported by Grant Number DE-SC0018269 from the United States Department of Energy. AKS and JAMK were additionally supported by a predoctoral training award from Grant Number T32-

GM110523 from the National Institute of General Medical Sciences of the National Institutes of Health. JAMK received additional support from the National Science Foundation Research Traineeship Program, grant number DGE-1828149. JH and C-LS are supported by Grant Number DMS-2113407 from the National Science Foundation. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

We thank Dr. Mark Seger and Dr. Peter Lammers (Arizona Center for Algae Technology and Innovation) for kindly providing inocula and biomass of *C. merolae*. We additionally thank the following individuals for valuable supplies and technical insight: Dr. Sigal Lechno-Yossef and Damien Sheppard (Fast Protein Liquid Chromatography technical consultation), Ludmila Roze (protein gel technical consultation), Dr. Josh Vermaas and Dr. Daipayan Sarkar (insightful discussion of lipid membrane permeability).

Figures

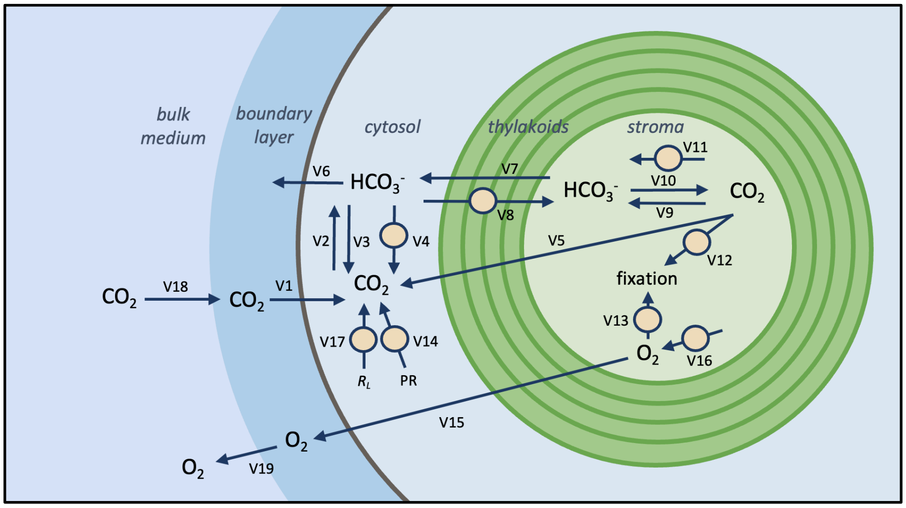
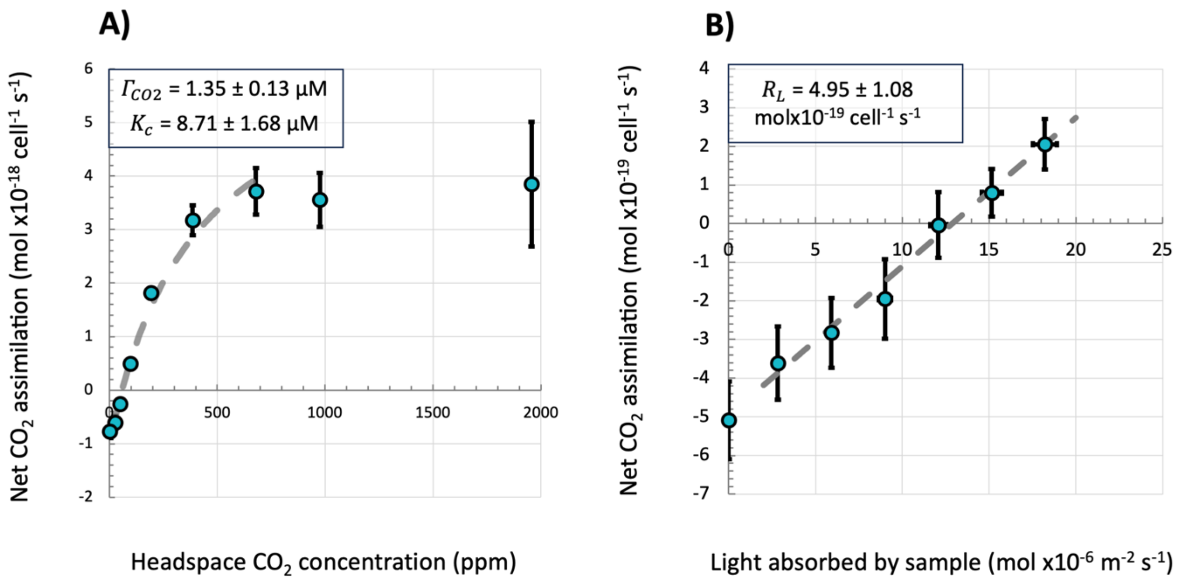


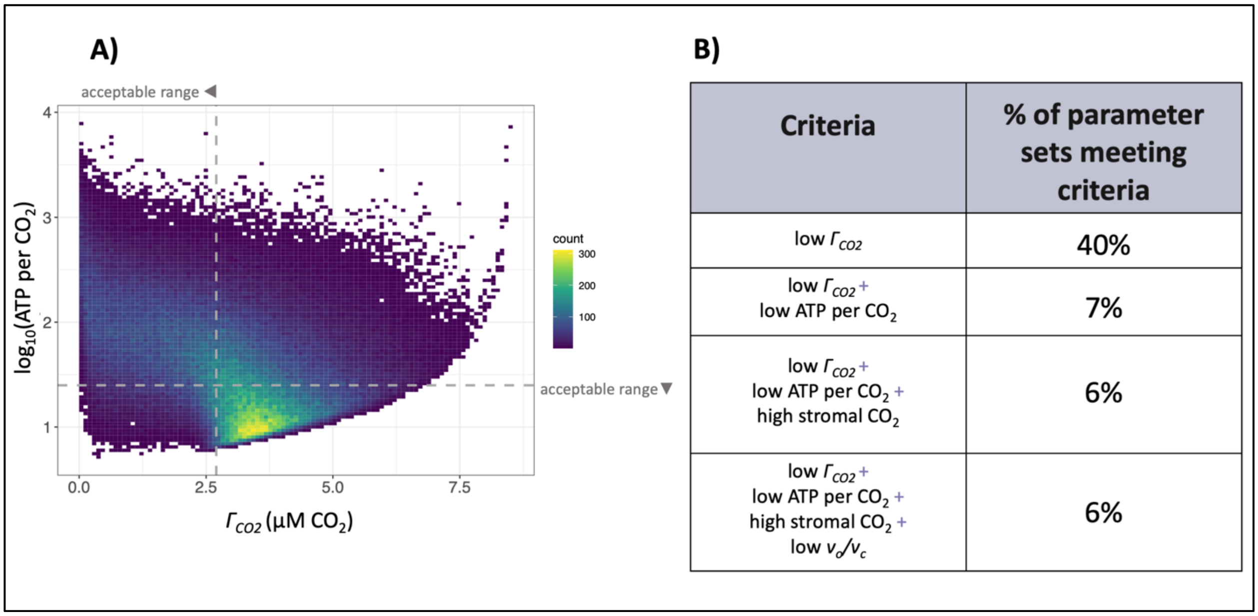
Figure 1. Cross-section of model structure. This model describes fluxes (indicated by arrows) and pools (indicated by molecular formulas) of a simplified dissolved inorganic carbon system (CO_2, HCO_3^-) and of oxygen (O_2). Molecule pools can be present in several well-mixed compartments: the bulk external medium surrounding the cell, an unstirred boundary layer of medium around the cell, the cytosol, or a central stromal space of the chloroplast. Circles mark enzymatically-catalyzed fluxes. Compartments are not drawn to scale. *PR* = photorespiratory CO_2 release, R_L = respiration in the light. All fluxes are reversible and are assigned an arbitrary direction, except those fluxes which represent producing or consuming material.



C)

	Measured at 25 °C	Measured at 45 °C
kcat$_C$ (s^{-1})	1.6 ± 0.1	7.2 ± 0.2
K$_C$ (µM)	11.7 ± 1.4	24.9 ± 3.2
K$_O$ (µM)	1237.8 ± 413.7	479.2 ± 43.5

Figure 2. Experimental data incorporated into the model. **(A,B).** Response of net assimilation in *C. merolae* to **(A)** CO$_2$ availability and **(B)** light availability. Points are mean ± SE (n = 3), and parameters calculated from the data are indicated in the upper left corner of each plot as mean ± SE. Dashed lines indicate trend fits used to determine K_C and R_L. The linear fit used to determine Γ_{CO2} is not pictured. **(C)** Kinetic properties of *C. merolae* rubisco. Rubisco turnover rate for CO$_2$ fixation ($kcat_C$), Michaelis-Menten constant of CO$_2$ fixation (K_C), and Michaelis-Menten constant of O$_2$ fixation (K_O) were measured at 25 and 45 °C. Data is mean ± SE, n = 4.



Criteria	% of parameter sets meeting criteria
low Γ_{CO2}	40%
low Γ_{CO2} + low ATP per CO_2	7%
low Γ_{CO2} + low ATP per CO_2 + high stromal CO_2	6%
low Γ_{CO2} + low ATP per CO_2 + high stromal CO_2 + low v_o/v_c	6%

Figure 3. (A) Values of key model outputs. Parameter sets are organized into a 2-dimensional histogram according to their output values of Γ_{CO2} and ATP per CO_2, with dashed lines indicating bounds for acceptable values of these outputs. 80 parameter sets (0.03% of total) are not pictured on the figure, as they produced negative ATP per CO_2 values and could not be log-transformed. **(B)** Percentages of parameter sets meeting various combinations of output criteria.

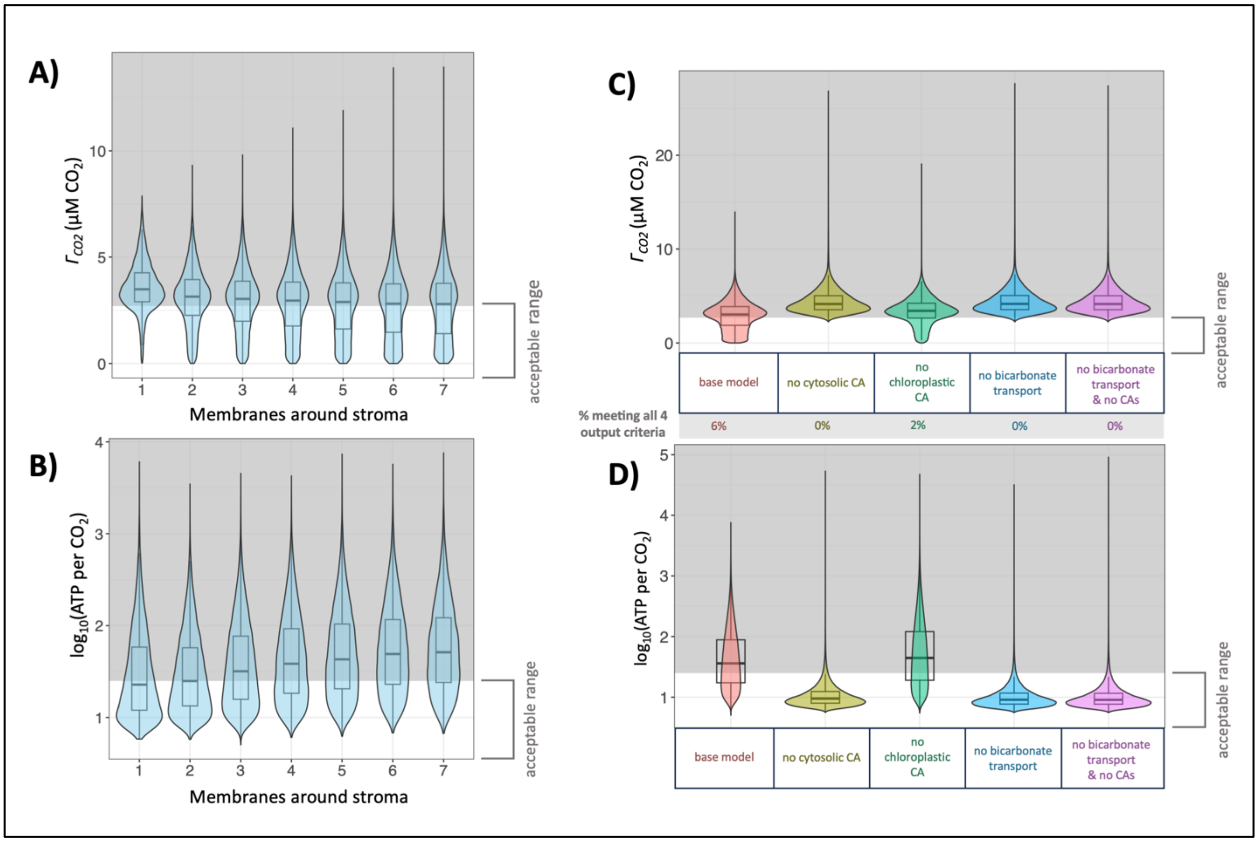


Figure 4. (A, B) Effect of model input parameter *Membranes* (x-axis) on key model outputs. Distribution of parameter set outputs for each value of *Membranes* is represented by a box plot overlaid on a violin plot. Shaded areas represent unacceptable values of outputs. **(A)** Effect of *Membranes* on model output Γ_{CO2}. **(B)** Effect of *Membranes* on model output ATP per CO$_2$. 80 parameter sets (0.03% of total) are not pictured in this panel, as they produced negative ATP per CO$_2$ values and could not be log-transformed. **(C, D)** Effect on key model outputs when bicarbonate transport or carbonic anhydrases (CAs) are removed from the model. Distribution of parameter set outputs for each scenario is represented by a box plot overlaid on a violin plot. Shaded areas represent out-of-bounds values of outputs. The same sampling of input parameter sets was run through models representing each scenario. **(C)** Γ_{CO2} in model scenarios where various model features removed, with indication of how many parameter sets met output criteria in each scenario. **(D)** ATP per CO$_2$ in model scenarios where bicarbonate transport activity at the chloroplast boundary is removed. 6,991 parameter sets producing negative ATP per CO$_2$ values (0.6% of total) are not pictured in this panel.

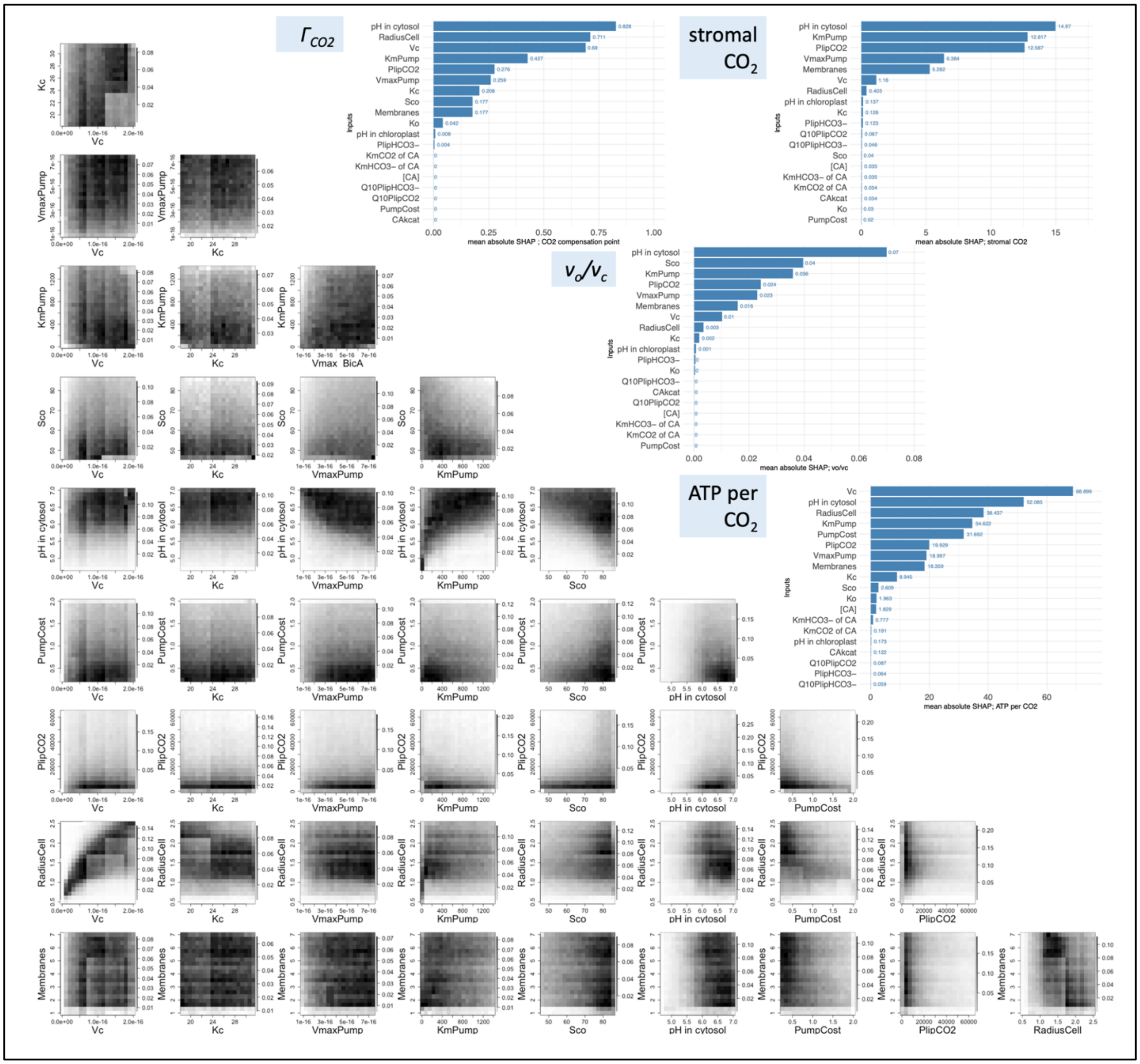


Figure 5. (upper right bar plots) Mean absolute SHapley Additive exPlanations (SHAP) plots for each output criterion. **(lower left density plots)** Density plots of parameter sets meeting all output criteria, organized by selected pairwise input parameter (input parameters pictured are those input parameters with high SHAP values for all output criteria). Darker areas indicate areas where more parameter sets meeting criteria occur. Scales of color vary for each plot).

References

---

1. S. Miyagishima, J. L. Wei, H. Nozaki, S. Hirooka, Cyanidiales: Evolution and Habitats. *ICyanidioschyzon Merolaei New Model Eukaryote Cell Organelle Biol.*, 3–16, (2017).
2. S. Miyagishima, J. L. Wei, Procedures for Cultivation, Observation, and Conventional Experiments in *Cyanidioschyzon merolae*. *ICyanidioschyzon Merolaei New Model Eukaryote Cell Organelle Biol.*, 31–42, (2017).
3. S. Y. Miyagishima, K. Tanaka, The Unicellular Red Alga *Cyanidioschyzon merolae* - The Simplest Model of a Photosynthetic Eukaryote. *Plant Cell Physiol* **62**, 926–941, (2021).
4. D. Y. Rahman, F. D. Sarian, A. van Wijk, M. Martinez-Garcia, M. J. E. C. van der Maarel, Thermostable phycocyanin from the red microalga *Cyanidioschyzon merolae*, a new natural blue food colorant. *J Appl Phycol* **29**, 1233–1239, (2017).
5. M. Seger, *et al.*, Engineered ketocarotenoid biosynthesis in the polyextremophilic red microalga *Cyanidioschyzon merolae* 10D. *Metab. Eng. Commun.* **17**, e00226, (2023).
6. M. Villegas-Valencia, *et al.*, Cultivation of the polyextremophile *Cyanidioschyzon merolae* 10D during summer conditions on the coast of the Red Sea and its adaptation to hypersaline sea water. *Front Microbiol* **14**, (2023).
7. W. Gross, Ecophysiology of algae living in highly acidic environments. *Hydrobiologia* **433**, 31–37, (2000).
8. C. Oesterhelt, E. Schmälzlin, J. M. Schmitt, H. Lokstein, Regulation of photosynthesis in the unicellular acidophilic red alga *Galdieria sulphuraria*. *Plant J* **51**, 500–511, (2007).
9. N. Rademacher, *et al.*, Transcriptional response of the extremophile red alga *Cyanidioschyzon merolae* to changes in CO_2 concentrations. *J Plant Physiol* **217**, 49–56, (2017).
10. A. K. Steensma, Y. Shachar-Hill, B. J. Walker, The carbon-concentrating mechanism of the extremophilic red microalga *Cyanidioschyzon merolae*. *Photosynth Res* **156**, 247–264, (2023).
11. D. Zenvirth, M. Volokita, A. Kaplan, Photosynthesis and Inorganic Carbon Accumulation in the Acidophilic Alga *Cyanidioschyzon merolae*. *Plant Physiol* **77**, 237–239, (1985).
12. N. Rademacher, *et al.*, Photorespiratory glycolate oxidase is essential for the survival of the red alga *Cyanidioschyzon merolae* under ambient CO_2 conditions. *J Exp Bot* **67**, 3165–3175, (2016).
13. M. R. Badger, *et al.*, The diversity and coevolution of Rubisco, plastids, pyrenoids and chloroplast-based CO_2-concentrating mechanisms in the algae. *Can J Bot* **76**, 1052–1071, (1998).

14. O. Misumi, *et al.*, *Cyanidioschyzon merolae* Genome. A Tool for Facilitating Comparable Studies on Organelle Biogenesis in Photosynthetic Eukaryotes. *Plant Physiol* **137**, 567–585, (2005).

15. C. Fei, A. T. Wilson, N. M. Mangan, N. S. Wingreen, M. C. Jonikas, Modelling the pyrenoid-based CO_2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nat Plants* **8**, 583–595, (2022).

16. J. M. McGrath, S. P. Long, Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol* **164**, 2247–2261, (2014).

17. G. D. Price, *et al.*, The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. *J. Exp. Bot.* **64**, 753–768, (2013).

18. Joshua A.M. Kaste, Berkley J. Walker, Yair Shachar-Hill, Biophysical carbon concentrating mechanisms in land plants: insights from reaction-diffusion modeling. *bioRxiv*, 2024.01.04.574220, (2024), https://doi.org/10.1101/2024.01.04.574220.

19. O. Misumi, T. Kuroiwa, S. Hirooka, "Application of the Tolerance to Extreme Environment to Land Plants" in *Cyanidioschyzon Merolae: A New Model Eukaryote for Cell and Organelle Biology*, T. Kuroiwa, *et al.*, Eds. (Springer Singapore, 2017), , pp. 325–341.

20. R. B. Gramacy, *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences* (Taylor & Francis Group, 2020).

21. I. M. Gherman, *et al.*, Bridging the gap between mechanistic biological models and machine learning surrogates. *PLoS Comput. Biol.* **19**, e1010988, (2023).

22. T. M. Ichinose, A. H. Iwane, Cyotological Analyses by Advanced Electron Microscopy. *ICyanidioschyzon Merolaei New Model Eukaryote Cell Organelle Biol.*, 129–152, (2017).

23. R. Itoh, *et al.*, Two ftsH-family genes encoded in the nuclear and chloroplast genomes of the primitive red alga *Cyanidioschyzon merolae*. *Plant Mol. Biol.* **41**, 321–337, (1999).

24. T. Kuroiwa, The primitive red algae *Cyanidium caldarium* and *Cyanidioschyzon merolae* as model system for investigating the dividing apparatus of mitochondria and plastids. *BioEssays* **20**, 344–354, (1998).

25. M. Miyagishima, *et al.*, Visualization of the microbody division in *Cyanidioschyzon merolae* with the fluorochrome brilliant sulfoflavin. *Protoplasma* **201**, 115–119, (1998).

26. T. Moriyama, N. Mori, N. Nagata, N. Sato, Selective loss of photosystem I and formation of tubular thylakoids in heterotrophically grown red alga *Cyanidioschyzon merolae*. *Photosynth Res* **140**, 275–287, (2018).

27. K. A. Reimer, *et al.*, The sole LSm complex in Cyanidioschyzon merolae associates with pre-mRNA splicing and mRNA degradation factors. *RNA* **23**, 952–967, (2017).

28. N. Sato, T. Moriyama, N. Mori, M. Toyoshima, Lipid metabolism and potentials of biofuel and high added-value oil production in red algae. *World J Microbiol Biotechnol* **33**, (2017).

29. K. Toda, H. Takano, S. Miyagishima, H. Kuroiwa, T. Kuroiwa, Characterization of a chloroplast isoform of serine acetyltransferase from the thermo-acidiphilic red alga *Cyanidioschyzon merolae*. *Biochim Biophys Acta - Mol Cell Res* **1403**, 72–84, (1998).

30. F. Yagisawa, *et al.*, Mitotic inheritance of endoplasmic reticulum in the primitive red alga *Cyanidioschyzon merolae*. *Protoplasma* **249**, 1129–1135, (2012).

31. F. Yagisawa, H. Kuroiwa, T. Fujiwara, T. Kuroiwa, Intracellular Structure of the Unicellular Red Alga *Cyanidioschyzon merolae* in Response to Phosphate Depletion and Resupplementation. *Cytologia (Tokyo)* **81**, 341–347, (2016).

32. H. Lavigne, A. Proye, J.-P. Gattuso, seacarb: Calculates parameters of the seawater carbonate system, (2019), . Deposited 2019.

33. G. D. Farquhar, S. von Caemmerer, J. A. Berry, A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* **149**, 78–90, (1980).

34. P. Virtanen, *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272, (2020).

35. M. D. McKay, R. J. Beckman, W. J. Conover, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21**, 239–245, (1979).

36. S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (Curran Associates Inc., 2017), , pp. 4768–4777.

37. J. H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **29**, 1189–1232, (2001).

38. T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 785–794, (2016).

39. R. B. Gramacy, D. W. Apley, Local Gaussian Process Approximation for Large Computer Experiments. *J. Comput. Graph. Stat.* **24**, 561–578, (2015).

40. G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning With Applications in R* (Springer Science + Business Media, 2013).

41. B. A. Read, F. R. Tabita, High substrate specificity factor ribulose bisphosphate carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant cyanobacterial RubiSCO containing "algal" residue modifications. *Arch. Biochem. Biophys.* **312**, 210–218, (1994).

42. K. Uemura, Anwaruzzaman, S. Miyachi, A. Yokota, Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase from Thermophilic Red Algae with a Strong Specificity for CO_2 Fixation. *Biochem Biophys Res Commun* **233**, 568–571, (1997).

43. S. M. Whitney, P. Baldet, G. S. Hudson, T. J. Andrews, Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J* **26**, 535–547, (2001).

44. D. B. Jordan, W. L. Ogren, The CO_2/O_2 specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase : Dependence on ribulose bisphosphate concentration, pH and temperature. *Planta* **161**, 308–313, (1984).

45. A. Prins, *et al.*, Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. *J. Exp. Bot.* **67**, 1827–1838, (2016).

46. G. Curien, *et al.*, Mixotrophic growth of the extremophile *Galdieria sulphuraria* reveals the flexibility of its carbon assimilation metabolism. *New Phytol* **231**, 326–338, (2021).

47. L. E. Fridlyand, Models of CO2 concentrating mechanisms in microalgae taking into account cell and chloroplast structure. *Biosystems* **44**, 41–57, (1997).

48. S. Thoms, M. Pahlow, D. A. Wolf-Gladrow, Model of the carbon concentrating mechanism in chloroplasts of eukaryotic algae. *J Theor Biol* **208**, 295–313, (2001).

49. J. Barrett, P. Girr, L. C. M. Mackinder, Pyrenoids: CO_2-fixing phase separated liquid organelles. *Biochim Biophys Acta - Mol Cell Res* **1868**, (2021).

50. R. Nevo, *et al.*, Thylakoid membrane perforations and connectivity enable intracellular traffic in cyanobacteria. *EMBO J.* **26**, 1467–1473, (2007).

51. A. D. Mountraki, B. Benjelloun-Mlayah, A. C. Kokossis, A Surrogate Modeling Approach for the Development of Biorefineries. *Front. Chem. Eng.* **2**, (2020).

52. R. D. de Oliveira, M. N. Guedes, J. Matias, G. A. C. Le Roux, Nonlinear Predictive Control of a Bioreactor by Surrogate Model Approximation of Flux Balance Analysis. *Ind. Eng. Chem. Res.* **60**, 14464–14475, (2021).

53. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**, 237–240, (1998).

54. B. J. Walker, D. M. Kramer, N. Fisher, X. Fu, Flexibility in the Energy Balancing Network of Photosynthesis Enables Safe Operation under Changing Environmental Conditions. *Plants* **9**, (2020).

55. C. Bellasio, S. J. Burgess, H. Griffiths, J. M. Hibberd, A high throughput gas exchange screen for determining rates of photorespiration or regulation of C_4 activity. *J. Exp. Bot.* **65**, 3769–3779, (2014).

56. T. Fujiwara, M. Ohnuma, Procedures for Transformation and Their Applications in *Cyanidioschyzon merolae*. *Cyanidioschyzon Merolae New Model Eukaryote Cell Organelle Biol.*, 87–103, (2017).

57. D. J. Orr, E. Carmo-Silva, Extraction of RuBisCO to Determine Catalytic Constants. *Methods Mol Biol* **1770**, 229–238, (2018).

58. J. Carter, B. P. Petersen, S. A. Printz, T. L. Sorey, T. T. Kroll, Quantitative Application for SDS − PAGE in a Biochemistry Lab. *J Chem Educ* **90**, 1255–1256, (2013).

59. D. S. Kubien, C. M. Brown, H. J. Kane, Quantifying the Amount and Activity of Rubisco in Leaves. *Photosynth. Res. Protoc.*, 349–362, (2010).

60. R. E. Sharwood, O. Ghannoum, S. M. Whitney, Prospects for improving CO_2 fixation in C_3-crops through understanding C_4-Rubisco biogenesis and catalytic diversity. *Curr. Opin. Plant Biol.* **31**, 135–142, (2016).

61. N. Loganathan, Y. C. Tsai, O. Mueller-Cajar, Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proc Natl Acad Sci U A* **113**, 14019–14024, (2016).

62. D. Tholen, X. G. Zhu, The mechanistic basis of internal conductance: a theoretical analysis of mesophyll cell photosynthesis and CO_2 diffusion. *Plant Physiol* **156**, 90–105, (2011).

63. S. Larsson, H. Björkbacka, C. Forsman, G. Samuelsson, Molecular cloning and biochemical characterization of carbonic anhydrase from *Populus tremula* × *tremuloides*. *Plant Mol. Biol.* **34**, 583–592, (1997).

64. B. H. Gibbons, J. T. Edsall, Rate of Hydration of Carbon Dioxide and Dehydration of Carbonic Acid at 25°. *J. Biol. Chem.* **238**, 3502–3507, (1963).

65. G. M. Bond, *et al.*, Development of Integrated System for Biomimetic CO_2 Sequestration Using the Enzyme Carbonic Anhydrase. *Energy Fuels* **15**, 309–316, (2001).

66. Y. Pocker, J. S. Y. Ng, Plant carbonic anhydrase. Properties and carbon dioxide hydration kinetics. *Biochemistry* **12**, 5127–5134, (1973).

67. Y. Pocker, R. R. Miksch, Plant carbonic anhydrase. Properties and bicarbonate dehydration kinetics. *Biochemistry* **17**, 1119–1125, (1978).

68. M. R. Badger, G. D. Price, B. M. Long, F. J. Woodger, The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. *J. Exp. Bot.* **57**, 249–265, (2006).

69. J. Du, B. Förster, L. Rourke, S. M. Howitt, G. D. Price, Characterisation of Cyanobacterial Bicarbonate Transporters in *E. coli* Shows that SbtA Homologs Are Functional in This Heterologous Expression System. *PLoS One* **9**, (2014).

70. J. Gutknecht, M. A. Bisson, F. C. Tosteson, Diffusion of carbon dioxide through lipid bilayer membranes: effects of carbonic anhydrase, bicarbonate, and unstirred layers. *J. Gen. Physiol.* **69**, 779–794, (1977).

71. A. Missner, *et al.*, Carbon dioxide transport through membranes. *J. Biol. Chem.* **283**, 25340–25347, (2008).

72. G. D. Price, M. R. Badger, S. von Caemmerer, The Prospect of Using Cyanobacterial Bicarbonate Transporters to Improve Leaf Photosynthesis in C_3 Crop Plants. *Plant Physiol* **155**, 20–26, (2011).

73. P. Han, D. M. Bartels, Temperature Dependence of Oxygen Diffusion in H_2O and D_2O. *J. Phys. Chem.* **100**, 5597–5602, (1996).

74. T. M. Ichinose, A. H. Iwane, Long-term live cell cycle imaging of single *Cyanidioschyzon merolae* cells. *Protoplasma* **258**, 651–660, (2021).

75. Y. Imoto, Y. Yoshida, Cellular Structure of *Cyanidioschyzon merolae*: A Minimum Set of Organelles. *ICyanidioschyzon Merolaei New Model Eukaryote Cell Organelle Biol.*, 17–30, (2017).

76. M. Matsuzaki, *et al.*, Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657, (2004).

77. A. Merola, *et al.*, Revision of *Cyanidium caldarium*. Three species of acidophilic algae. *Plant Biosyst.* **115**, 189–195, (1981).

78. Y. Sekiguchi, *et al.*, Coherent X-Ray Diffraction Imaging of *Cyanidioschyzon merolae*. *ICyanidioschyzon Merolaei New Model Eukaryote Cell Organelle Biol.*, 153–173, (2017).

79. A. I. Flamholz, *et al.*, Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochem.* **58**, 3365–3376, (2019).

80. H. Nozaki, *et al.*, A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biol* **5**, 28, (2007).

81. C. Yoshida, M. Murakami, A. Niwa, M. Takeya, T. Osanai, Efficient extraction and preservation of thermotolerant phycocyanins from red alga Cyanidioschyzon merolae. *J. Biosci. Bioeng.* **131**, 161–167, (2021).

SUPPLEMENTAL MATERIAL

Supplemental Methods

Model optimization and estimation of simulated compensation point

In order to characterize the response of key outputs and robustness of conclusions to a wide range of possible parameterizations of the model, we used Latin Hypercube Sampling to explore 240,000 parameter combinations according to the bounds specified in (**Table S1**). These simulations were run on Michigan State University's High Performance Computing Cluster. Compensation point estimates were generated for every parameter set by running the model at external CO_2 concentrations ranging from 0.0001 to 1000 μM, constructing a cubic spline from the resulting curve of net CO_2 assimilation vs. external CO_2 concentration, and identifying the root of this spline to find the compensation point. Each simulation was verified to reach steady-state (metabolite concentration solutions changing 0.01% or less from previous value).

Definition of reasonable output values

CO_2 compensation point (Γ_{CO2})

We accepted Γ_{CO2} values less than or equal to 2.70 μM, corresponding to no more than twice the mean measured value (**Figure 2**).

Ratio of ATP consumption flux to net CO_2 assimilation flux (ATP per CO_2)

We accepted ATP per CO_2 values which were less than or equal to 25 and greater than 0. Measured light response curves indicated how much additional light absorption drives additional CO_2 assimilation (**Figure 2**) We used this data to estimate how much additional ATP production drives an additional CO_2 assimilation, using the photon per ATP values for various light-reaction pathways (53), the cylindrical geometry of the gas-exchange sample chamber, and the measured density of cells in the sample. The resulting estimated values were: 13.8 ± 2.19 ATP produced/CO_2 assimilated (mean ± SE, assuming cyclic and linear electron flow operating equally) or 17.4 ± 2.76 ATP produced/CO_2 assimilated (mean ± SE, assuming linear electron flow only operating). This suggests that ATP per CO_2 values of up to roughly 25 are supported by photosynthetic electron flow. The lower bound of the acceptable range excludes a few parameter sets outputting negative ATP per CO_2, since these parameter sets represent particularly non-functional CCM scenarios with negative net assimilation values under ambient CO_2 conditions.

Steady-state CO_2 concentration in the chloroplast stroma (stromal CO_2)

We accepted chloroplast CO_2 concentration values of greater than or equal to the CO_2 concentration in the medium under 400 ppm CO_2 atmosphere, by the logic that a functional CCM should result in rubisco accessing a greater CO_2 concentration than is available from ambient medium.

Ratio of oxygen fixation flux to carbon fixation flux (v_o/v_c)

We accepted v_o/v_c values less than or equal to 0.3, based on data and models indicating that plants without CCMs are unlikely to achieve v_o/v_c less than approximately 0.3 (54).

Experimental data collection: gas-exchange measurements

Cyanidioschyzon merolae 10D was grown as cultures in Erlenmeyer flasks in 50 mL of medium containing 40 mM $(NH_4)_2SO_4$, 4 mM $MgSO_4 \cdot 7H_2O$, 8 mM KH_2PO_4, 0.75 mM $CaCl_2 \cdot 2H_2O$, 1 mL L^{-1} Hutner's Trace Elements solution, and H_2SO_4 to pH 2.7 (recipe modified from MA2 medium recipe of (55)). Cultures were maintained at 40 °C under 100 µmol m^{-2} s^{-1} white light, with aeration by shaking at 100 rpm. For gas-exchange measurements, cultures of OD_{750} 1.0 – 1.2 were resuspended in growth medium to OD_{750} 0.6 (1.60×10^7 – 3.68×10^7 cells/mL). Gas-exchange parameters were measured in a LI-6800-18 Aquatic Chamber (LI-COR Biosciences) at 45 °C, following the procedures of (10).

Experimental data collection: rubisco kinetics measurements

We purified rubisco from *C. merolae* biomass with a protocol adapted from (2, 56). Approximately 60 grams of biomass were lysed by freeze-thawing followed by mechanical homogenization. Crude rubisco was polyethylene-glycol-precipitated from clarified homogenate and purified by FPLC. FPLC fractions eluting under the major UV trace peak were assayed by SDS-PAGE and by spectrophotometric rubisco activity assay (procedures adapted from (57, 58)) (**Figure S3**). Fractions containing active semi-pure rubisco were pooled, concentrated with a 100 kDa centrifugal concentration filter, and snap-frozen for use in rubisco assays.

Purified rubisco was used to determine catalytic properties as described previously (43), with some alterations to protein desalting and activation: concentrated protein aliquots were first diluted with activation mix containing 100 mM Bicine-NaOH pH 8.0, 20 mM $MgCl_2$, 10 mM $NaHCO_3$, and 1 % (v/v) Plant Protease Inhibitor cocktail (Sigma-Aldrich, UK). Rubisco was then activated at 45 °C for 15 min before being used in $^{14}CO_2$ consumption assays at either 25 °C or 45 °C with CO_2 concentrations of 8, 16, 24, 36, 68, and 100 µM. To determine K_O, these CO_2 concentrations were combined with concentrations of either 0, 21, 40, or 70 % (v/v) O_2. $kcat_C$ was determined using measurements with 0% O_2. An aliquot of the activated protein was used for determination of Rubisco active sites via ^{14}C-CABP binding using the method of (59). For ^{14}C-CABP binding, protein aliquots were incubated at 45°C for 15 mins with ^{14}C-CABP to maximize binding, prior to application to Sephadex columns as previously described (60). Aliquots were also analyzed via SDS-PAGE alongside known concentrations of plant type Rubisco to strengthen estimates of Rubisco content.

ODE System

$$\frac{d[CO_2]_{cytosol}}{dt} = \frac{V_1 + V_3 + V_{14} + V_{17} - V_2 - V_4 - V_5}{Volume_{cytosol}}$$

$$\frac{d[CO_2]_{boundary}}{dt} = \frac{V_{18} - V_1}{Volume_{boundary}}$$

$$\frac{d[HCO_3^-]_{cytosol}}{dt} = \frac{V_2 + V_4 + V_6 - V_3 - V_7 - V_8}{Volume_{cytosol}}$$

$$\frac{d[CO_2]_{chloroplast}}{dt} = \frac{V_5 + V_{10} - V_{11} - V_{12} - V_9}{Volume_{chloroplast}}$$

$$\frac{d[HCO_3^-]_{chloroplast}}{dt} = \frac{V_7 + V_8 + V_9 + V_{11} - V_{10}}{Volume_{chloroplast}}$$

$$\frac{d[O_2]_{chloroplast}}{dt} = \frac{V_{16} - V_{13} - V_{15}}{Volume_{chloroplast}}$$

$$\frac{d[O_2]_{boundary}}{dt} = \frac{V_{19} + V_{15}}{Volume_{boundary}}$$

Model fluxes

See **Table S1** and main text for parameter sources, values, and definitions.

Diffusion of inorganic carbon through membranes or boundary layer (V1, V5, V6, V7, V15, V18, V19)

Implemented as described in Methods.

$$V_1 = CO_2\ Conductivity_{boundary\ to\ cytosol}([CO_2]_{boundary} - [CO_2]_{cytosol})$$

$$V_5 = CO_2\ Conductivity_{cytosol\ to\ chloroplast}([CO_2]_{cytosol} - [CO_2]_{chloroplast})$$

$$V_6 = HCO_3^-\ Conductivity_{boundary\ to\ cytosol}([HCO_3^-]_{boundary} - [HCO_3^-]_{cytosol})$$

$$V_7 = HCO_3^-\ Conductivity_{cytosol\ to\ chloroplast}([HCO_3^-]_{cytosol} - [HCO_3^-]_{chloroplast})$$

$$V_{15} = Conductivity_{boundary\ to\ chloroplast}([O_2]_{chloroplast} - [O_2]_{boundary})$$

$$V_{18} = CO_2\ Conductivity_{exterior\ to\ boundary}([CO_2]_{exterior} - [CO_2]_{boundary})$$

$$V_{19} = O_2\ Conductivity_{exterior\ to\ boundary}([O_2]_{exterior} - [O_2]_{boundary})$$

Spontaneous interconversion of dissolved inorganic carbon species (V2, V3, V9, V10)

Implemented as described in Methods.

$$V_2 = CO_{2_{cytosol}} * k_2 * Volume_{cytosol}$$

$$V_3 = [HCO_3^-]_{cytosol} * [H^+]_{cytosol} * k_{-2} * Volume_{cytosol}$$

$$V_9 = CO_{2_{chloroplast}} * k_2 * Volume_{chloroplast}$$

$$V_{10} = [HCO_3^-]_{chloroplast} * [H^+]_{chloroplast} * k_{-2} * Volume_{chloroplast}$$

Carbonic-anhydrase-mediated interconversion of inorganic carbon (V4, V11)

Implemented as described in Methods.

$$V_4 = \frac{[CA]_{cytosol} * CA_{kcat} * \left([CO_2]_{cytosol} - \frac{[HCO_3^-]_{cytosol}[H^+]_{cytosol}}{K_a}\right)}{K_{m_2}^{CO_2} + [HCO_3]_{cytosol}\left(\frac{K_m^{CO_2}}{K_m^{HCO_3^-}}\right) + [CO_2]_{cytosol}}$$

$$V_{11} = \frac{[CA]_{chloroplast} * CA_{kcat} * \left([CO_2]_{chloroplast} - \frac{[HCO_3^-]_{chloroplast}[H^+]_{chloroplast}}{K_a}\right)}{K_{m_2}^{CO_2} + [HCO_3]_{chloroplast}\left(\frac{K_m^{CO_2}}{K_m^{HCO_3^-}}\right) + [CO_2]_{chloroplast}}$$

Active transport (pumping) of bicarbonate from cytosol to stroma (V8)

Implemented as described in Methods.

$$V_8 = \frac{Vmax_{pump}[HCO_3^+]_{cytosol}}{[HCO_3^+]_{cytosol} + Km_{pump}}$$

Carbon dioxide fixation by rubisco (V12)

Implemented as described in Methods.

$$V_{12} = \frac{Vmax_{carboxylation}[CO_2]_{chloroplast}}{\left([CO_2]_{chloroplast} + K_m^{CO_2}\left(1 + \frac{[O_2]_{chloroplast}}{K_m^{O_2}}\right)\right)}$$

Oxygen fixation by rubisco (V13)

Implemented as described in Methods.

$$V_{13} = \frac{V_{12}}{\left(\frac{v_c}{v_o}\right)}$$

Evolution of carbon dioxide in cytosol as a result of photorespiration (V14)

This flux is determined based on the stoichiometry of photorespiration.

$$V_{14} = \frac{1}{2}V_{13}$$

Evolution of oxygen into stroma from thylakoid action (V16)

This flux is determined based on the stoichiometry of photosynthesis.

$$V_{16} = V_{12}$$

Evolution of carbon dioxide in cytosol as a result of respiration in the light (V17)

Implemented as described in Methods.

$$V_{17} = R_L\left(\frac{Volume_{cell}}{Volume_{cell\ with\ 1\mu m\ radius}}\right)$$

Supplemental Figures

Table S1. Parameter values or ranges used in the model. Values are known or assumed to be at 25 °C, unless otherwise specified.

Parameter name	**Parameter definition**	**Parameter value or range of values**	**Source or rationale**
[CA]	carbonic anhydrase concentration, used for both chloroplast and cytosolic	0.04 – 0.69 mol m^{-3}	range of stromal CA concentrations calculated by (61)

	carbonic anhydrases		
$[H^+]_{chloroplast}$	proton concentration in the chloroplast stroma, from pH in chloroplast	$10^{-6.35} – 10^{-8}$ M	range based on generic stroma in light pH to average *C. merolae* intracellular pH reported by (11)
$[H^+]_{cytosol}$	proton concentration in the cytosol, from pH in cytosol	$10^{-4.7} – 10^{-7}$ M	range based on neutral pH down to lowest *C. merolae* cytosolic pH implied by (11)
$[HCO_3^-]_{boundary}$	HCO_3^- concentration in medium boundary layer	0 μM	assumption that HCO_3^- concentrations are negligible at typical growth pH of *C. merolae*
CA_{kcat}	carbonic anhydrase hydration rate constant	$0.16 \times 10^6 – 0.3 \times 10^6$ s^{-1}	range of values reported and cited by (62)
$CO2_{ppm}$	air concentration of CO_2, used for calculation of ambient medium CO_2 concentration	400 ppm	generic air CO_2 concentration
$Henry_{CO2}$	standard-temperature Henry's law constant for CO_2, used for calculation of ambient medium CO_2 concentration	0.035 mol kg^{-1} bar^{-1}	NIST
$Henry_{O2}$	standard-temperature	0.0013 mol kg^{-1} bar^{-1}	NIST

	Henry's law constant for O_2, used for calculation of ambient medium O_2 concentration		
$Henry_{tempCO2}$	Henry's law temperature dependence constant for CO_2, used for calculation of ambient medium CO_2 concentration	2400 K	NIST
$Henry_{tempO2}$	Henry's law temperature dependence constant for O_2, used for calculation of ambient medium O_2 concentration	1700 K	NIST
k_{-2}	rate constant for uncatalyzed CO_2 hydration	13.7 s^{-1}	(63)
k_2	rate constant for uncatalyzed CO_2 dehydration	0.062 s^{-1}	(64)
K_C	affinity of rubisco for CO_2	18.5 – 31.3 µM	this study (measured value at 45 °C, range is mean ± 2 SEs)
K_m^{CO2} in carbonic anhydrase equations	carbonic anhydrase affinity for CO_2	1.0 – 17.9 mol m^{-3}	range of values from (65)
K_m^{HCO3-}	carbonic anhydrase	26.6 – 98.0 mol m^{-3}	range of values from (66)

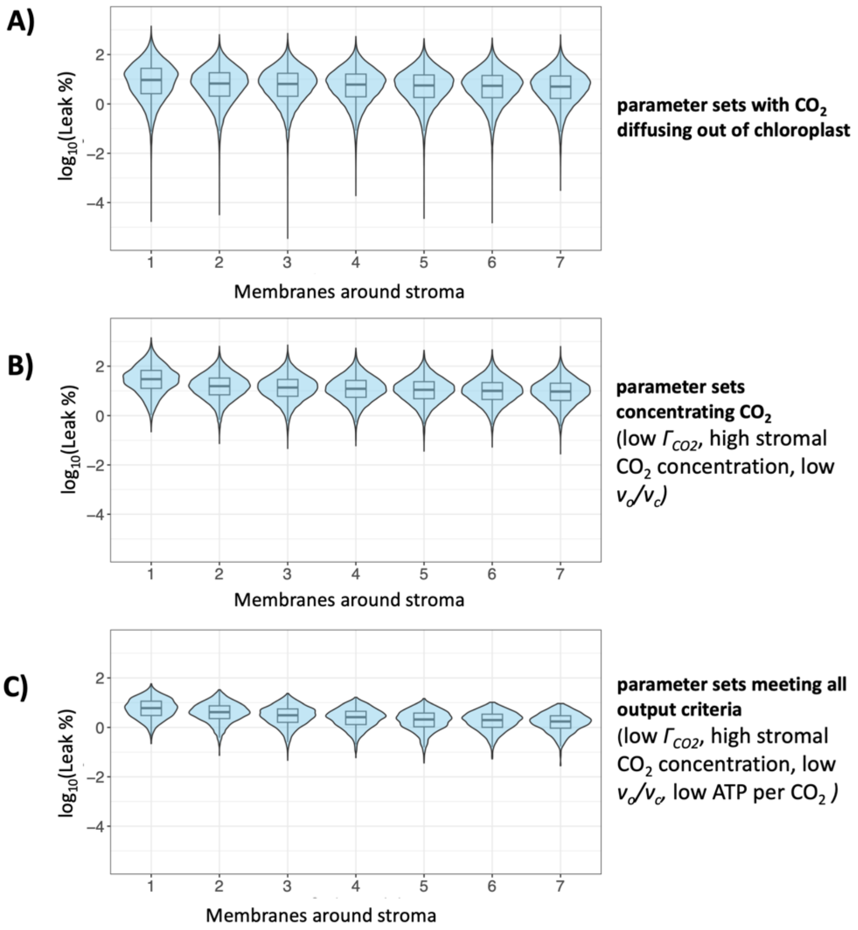
	affinity for HCO_3^-		
Km_{pump}	bicarbonate transporter affinity for bicarbonate	2 – 350 µM	range based on values for high- and low-affinity bicarbonate transporters from cyanobacteria (summarized in (67, 68))
K_O	affinity of rubisco for CO_2	394.0 – 564.6 µM	this study (measured value at 45 °C, range is mean ± 2 SEs) **(Figure 2)**
Membranes	number of lipid bilayers around the chloroplast, including outer envelope	1 – 7 bilayers (integer values only)	see “Methods: Model geometry”
$O2_{percent}$	air concentration of O_2, used for calculation of medium O_2 concentration	21%	generic air O_2 concentration
pKa	CO_2 and HCO_3^- system overall pKa	5.97	(30)
$Plip_{CO2}$	CO_2 permeability coefficient of a double lipid layer membrane, used in conductivity calculations	$3.5x10^2$ – $1.6x10^4$ µm / s	range from 1/10th of value of (69) to value of (70)
$Plip_{HCO3-}$	HCO_3^- permeability coefficient of a double lipid layer membrane, used in conductivity calculations	$2x10^{-9}$ – $1.2x10^{-6}$ m/s	range of values summarized in (61)
$pressure_{atm}$	atmospheric pressure	1.01325 bar	standard atmospheric pressure

$Pump_{cost}$	the assumed cost of pumping a single HCO_3^- ion across a lipid bilayer	0.25 – 2 ATP	range from lowest estimated value of (71) to twice a generic value 1
$P_{waterO2}$	O_2 diffusion coefficient of water, used in conductivity calculations	3050 $m^2\ s^{-1}$	O_2 diffusion coefficient at 45.1 °C (72)
$Q_{10}\ CA_{kcat}$	Q_{10} for parameter CA_{kcat}	2	generic value
$Q_{10\ k\text{-}2}$	Q_{10} for parameter k_{-2}	2	generic value
$Q_{10\ k2}$	Q_{10} for parameter k_2	2	generic value
$Q_{10}\ K_m^{CO2}$	Q_{10} for parameter K_m^{CO2}	2	generic value
$Q_{10}\ K_m^{HCO3-}$	Q_{10} for parameter K_m^{HCO3-}	2	generic value
$Q_{10}\ Km_{pump}$	Q_{10} for parameter Km_{pump}	2	generic value
$Q_{10\ PlipCO2}$	Q_{10} for parameter $Plip_{CO2}$	1 - 2	range from generic value 2 to assuming that organisms fully maintain membrane permeability across temperature
$Q_{10\ PlipHCO3-}$	Q_{10} for parameter $Plip_{HCO3-}$	1 - 2	range from generic value 2 to assuming that organisms fully maintain membrane permeability across temperature
$Q_{10}\ S_{co}$	Q_{10} for parameter S_{co}	0.6	measured value for a thermophile red algae rubisco (40)
$Q_{10}\ V_c$	Q_{10} for parameter V_c	4.5	this study (ratio between $kcat_C$ measured at 45 °C and $kcat_C$ measured at 25 °C) (**Figure 2**)
$Q_{10}\ Vmax_{pump}$	Q_{10} for parameter $Vmax_{pump}$	2	generic value
$Radius_{cell}$	radius of a *C. merolae* cell	0.5 – 2.5 μM	(1, 73–77)

R_L	rate of respiration in the light	$4.95\text{x}10^{-19}$ mol s^{-1} $cell^{-1}$	this study (estimated and normalized from gas-exchange data as described in Methods) (**Figure 2**)
$S_{c/o}$	specificity of rubisco for CO_2 over O_2	129 – 238 (molar ratio)	range of mean specificities $S_{c/o}$ for "taxonomy=Red algae" or "taxonomy=Rhodophyte" in rubisco kinetics meta-analysis (78)
temperature	temperature at which the model is run	45 °C	a typical growth temperature of *C. merolae* (2)
V_C	maximum rate of CO_2 fixation by rubisco	15 – 44 mM/s	range of ± 50% of the value used by (15)
$Vmax_{pump}$	maximum rate of bicarbonate pumping	$185\text{x}10^{-6}$ – $22.48\text{x}10^{-6}$ mol μm^{-2} s^{-1}	range between maximum rates of two different cyanobacterial bicarb transporters, calculated by (16)

Table S2. The test root-mean-square errors (RMSEs) and normalized RMSEs (NRMSEs) of four machine-learning surrogate models: eXtreme Gradient Boosting (XGBoost), Local approximate Gaussian Process (laGP), single-layer Neural Network (NN), and Deep Neural Network (DNN).

RMSE	**XGBoost**	**laGP**	**NN**	**DNN**
Γ_{CO2}	0.1982	1.5054	0.7899	0.1458
stromal CO_2	15.7500	73.2884	66.0004	15.6920
v_o/v_c	0.0132	0.1248	0.0547	0.0613
ATP per CO_2	49.8090	173.7715	161.0012	72.2325
NRMSE	**XGBoost**	**laGP**	**NN**	**DNN**
Γ_{CO2}	0.0142	0.1080	0.0567	0.0105
stromal CO_2	0.0043	0.0201	0.0181	0.0043
v_o/v_c	0.0086	0.0813	0.0356	0.0400
ATP per CO_2	0.0084	0.0293	0.0271	0.0121



D)

Membranes	1	2	3	4	5	6	7
% meeting all 4 output criteria	3.3%	8.3%	7.2%	6.0%	5.2%	4.5%	4.3%

Figure S1. Effect of model input parameter *Membranes* (x-axis) on CO_2 leakage from the chloroplast (represented as $\log_{10}$(Leak %): the log-transformed percentage relationship between the CO_2 flux from the chloroplast to cytosol and the CO_2 assimilation flux). **(A)** Results for parameter sets with CO_2 diffusing out of the chloroplast (V5 steady-state flux towards cytosol, rather than towards chloroplast) (n = 191,345). **(B)** Results for parameter sets concentrating CO_2 (low Γ_{CO2}, high stromal CO_2 concentration, low v_o/v_c) (n = 92,764). **(C)** Results for parameter sets meeting all output criteria (low Γ_{CO2}, high stromal CO_2 concentration, low v_o/v_c, low ATP per CO_2) (n = 13,998).

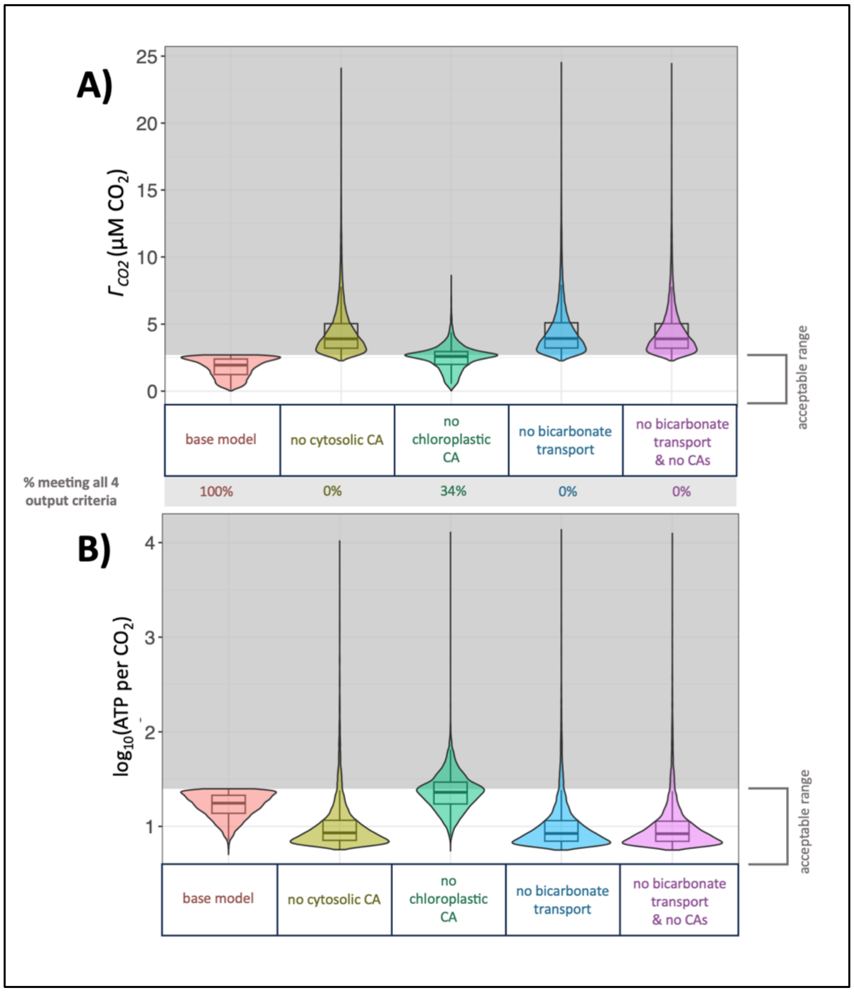


Figure S2. Effect on key model outputs when bicarbonate transport or carbonic anhydrases (CAs) are removed from the model, looking only at the 13,998 parameter sets that met all four output criteria in the base model. Distribution of parameter set outputs for each scenario is represented by a box plot overlaid on a violin plot. Shaded areas represent unacceptable values of outputs. The same sampling of input parameter sets was run through models representing each scenario. **(A)** Γ_{CO2} in model scenarios where various model features removed, with indication of how many parameter sets met output criteria in each scenario. **(B)** ATP per CO$_2$ in model scenarios where bicarbonate transport activity at the chloroplast boundary is removed. 2,083 parameter sets producing negative ATP per CO$_2$ values (3% of total) are not pictured in this panel due to log-transformation.

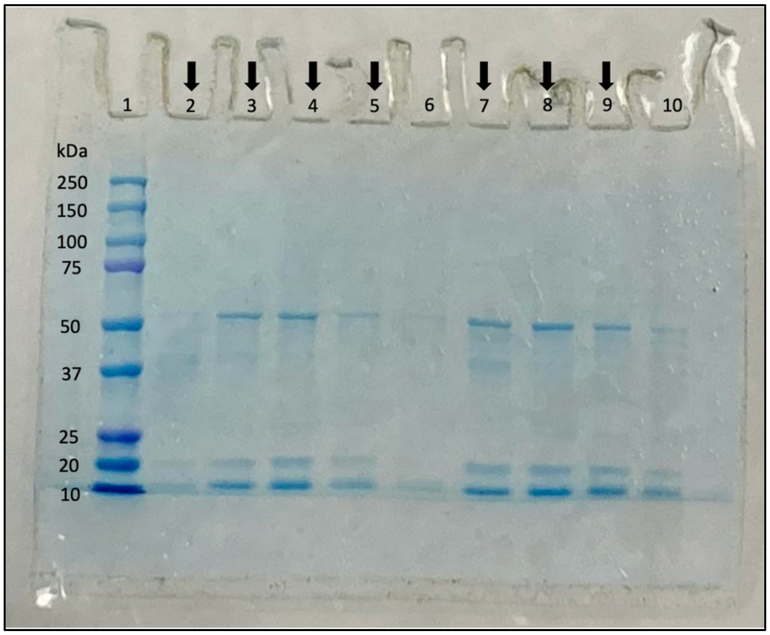


Figure S3. SDS-PAGE analysis of rubisco preparation. Lane 1: protein standards (BioRad Precision Plus Protein Dual Color Standards). Lanes 2 - 10 : fractions from fast protein liquid chromatography (FPLC) rubisco preparation. Three major bands are observable on the gel, which are believed to correspond to the rubisco large subunit (expected molecular weight 54 kDa), the rubisco small subunit (expected molecular weight 16 kDa), and the accessory pigment phycocyanin (composed of two subunits with reported molecular weights of about 15 – 18 kDa each). Arrows indicate fractions with rubisco activity detectable by a spectrophotometric assay, which were pooled for further preparation and analysis (see Methods for details). Molecular weights of rubisco were predicted from amino acid sequences CMV013C and CMV014C (*Cyanidioschyzon merolae* Genome Project v3, http://czon.jp/) (75, 79) using the Protein Molecular Weight Tool from the bioinformatics.org Sequence Manipulation Suite. Molecular weights of phycocyanin were reported by (4, 80).

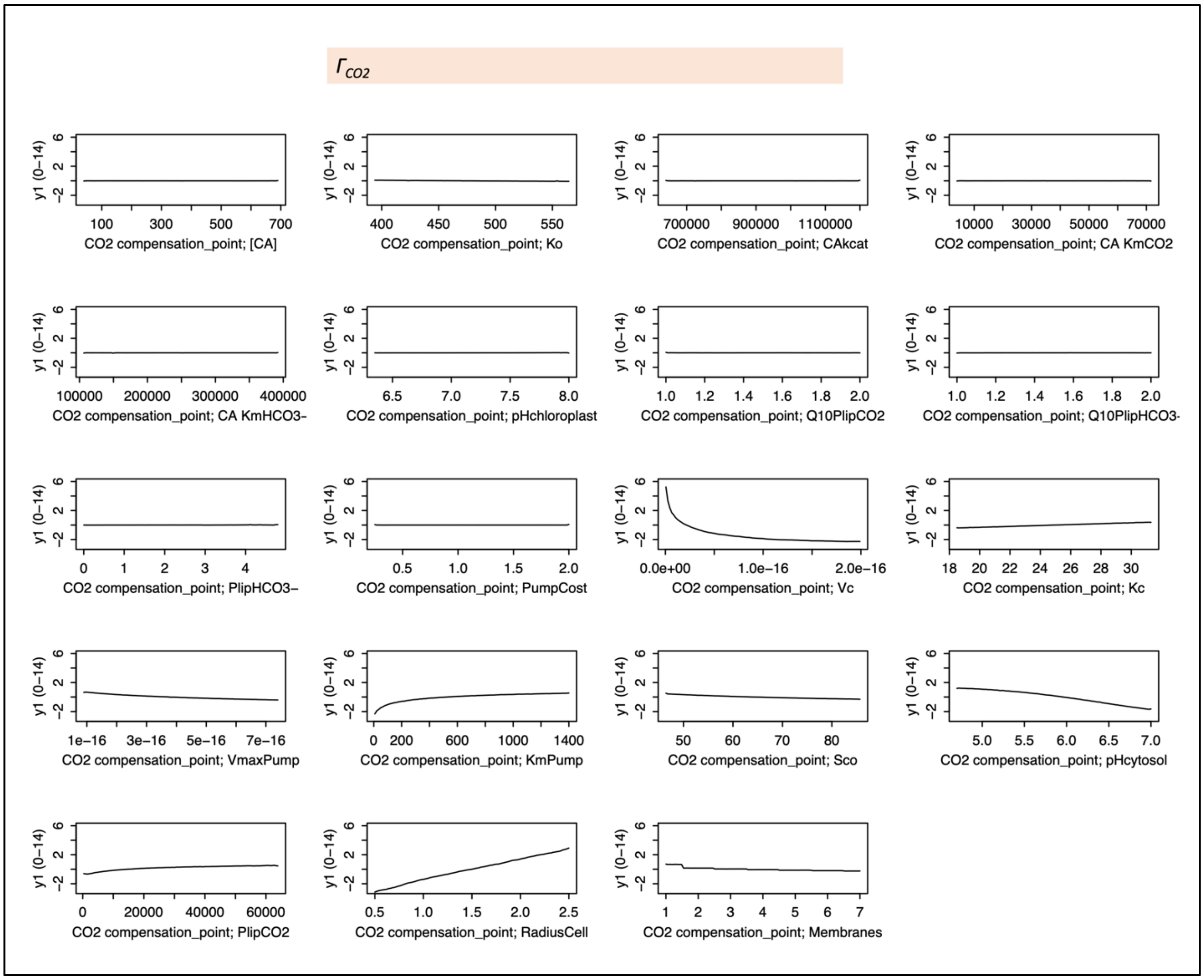
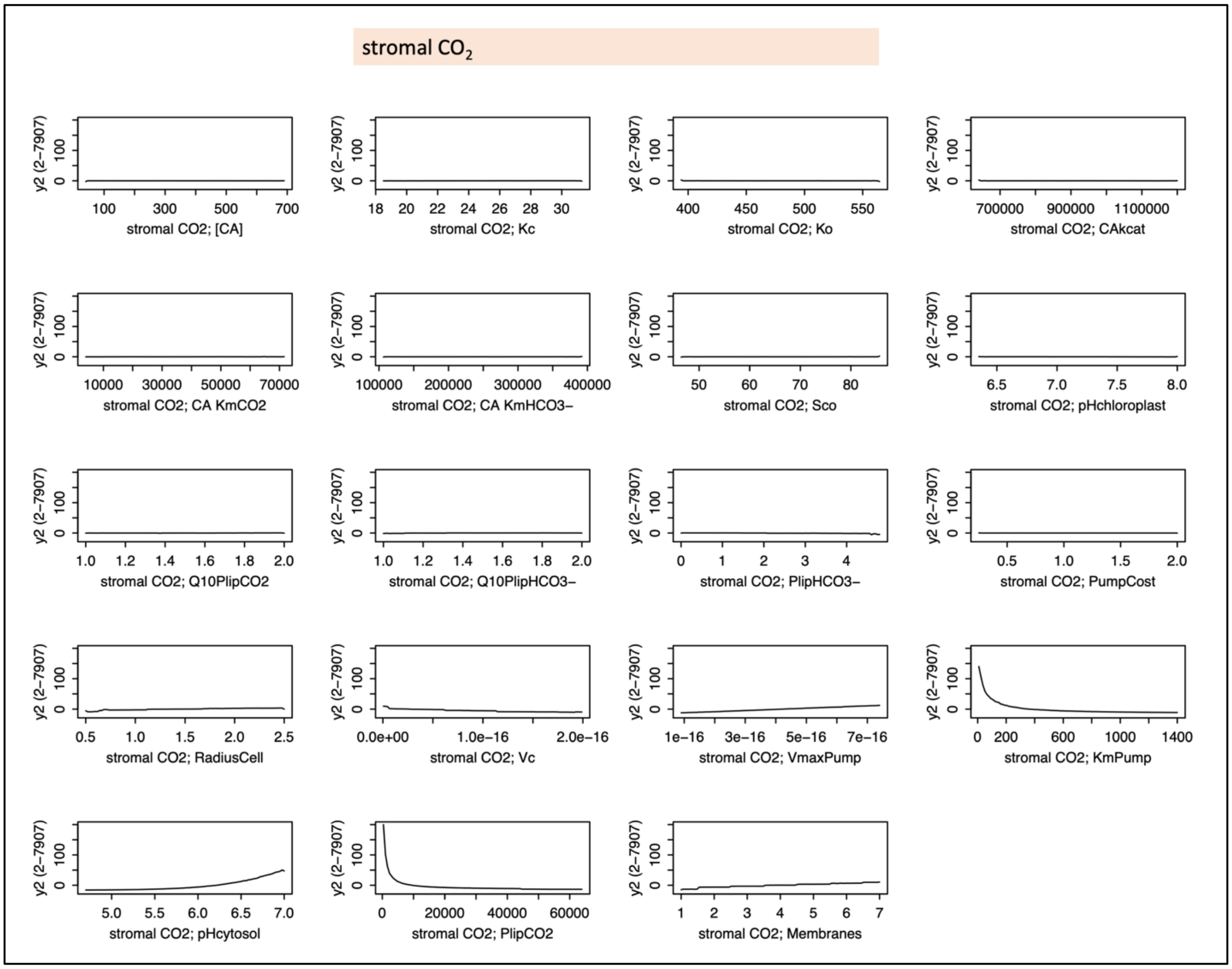


Figure S4. Partial dependence (PD) plots of first-order effects for Γ_{CO2}.



733
734 **Figure S5.** Partial dependence (PD) plots of first-order effects for stromal CO$_2$.
735

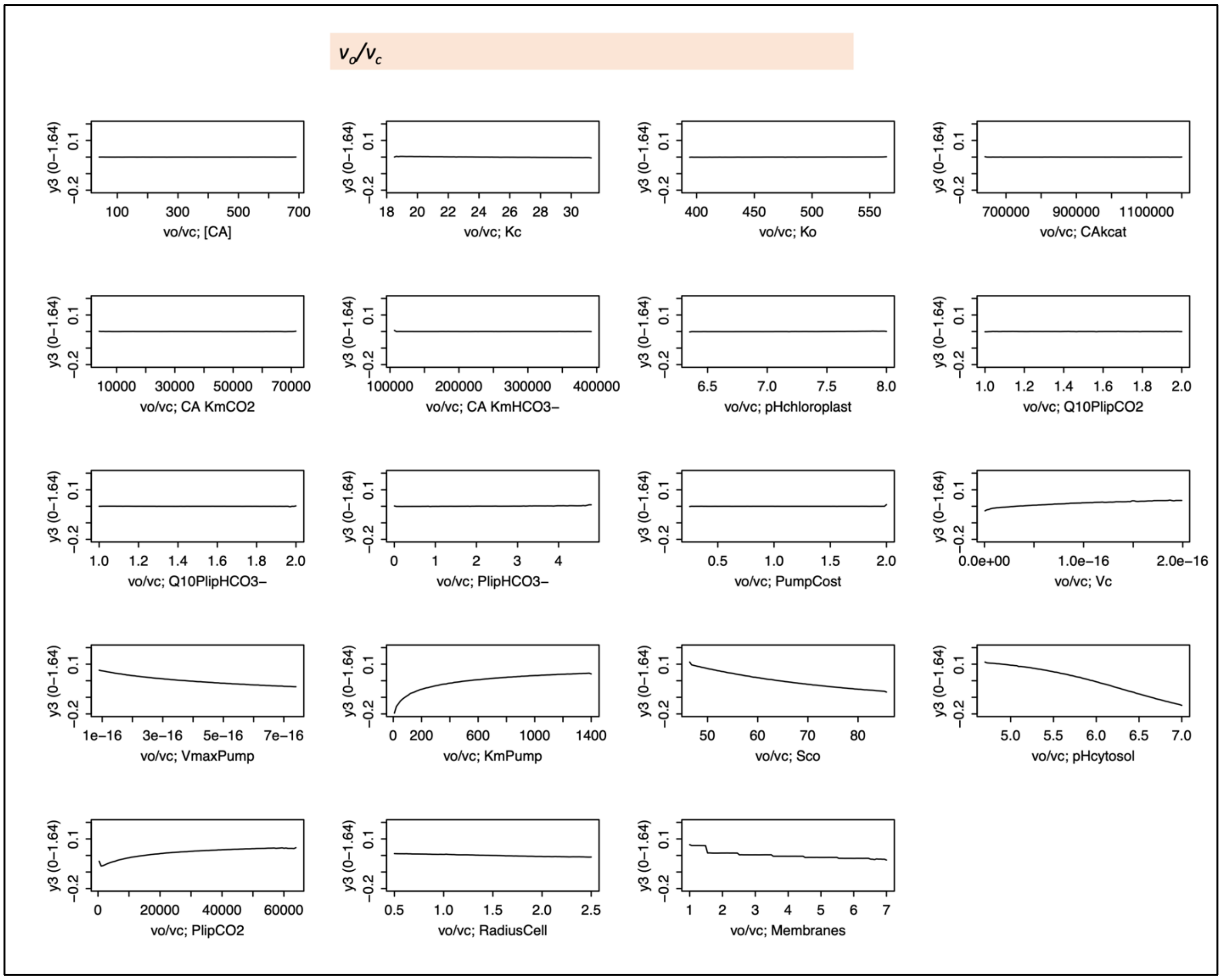


Figure S6. Partial dependence (PD) plots of first-order effects for v_o/v_c.

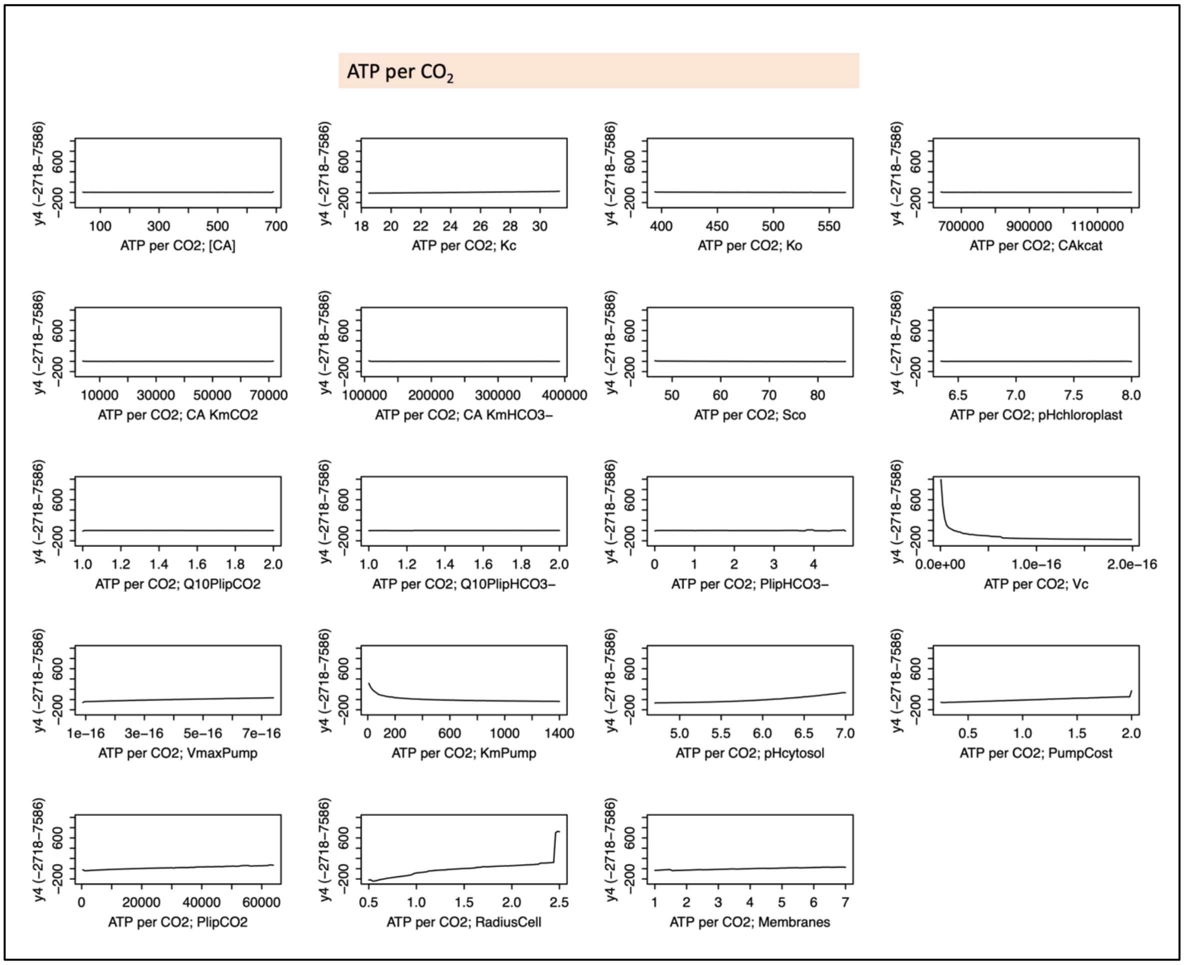


Figure S7. Partial dependence (PD) plots of first-order effects for ATP per CO_2.

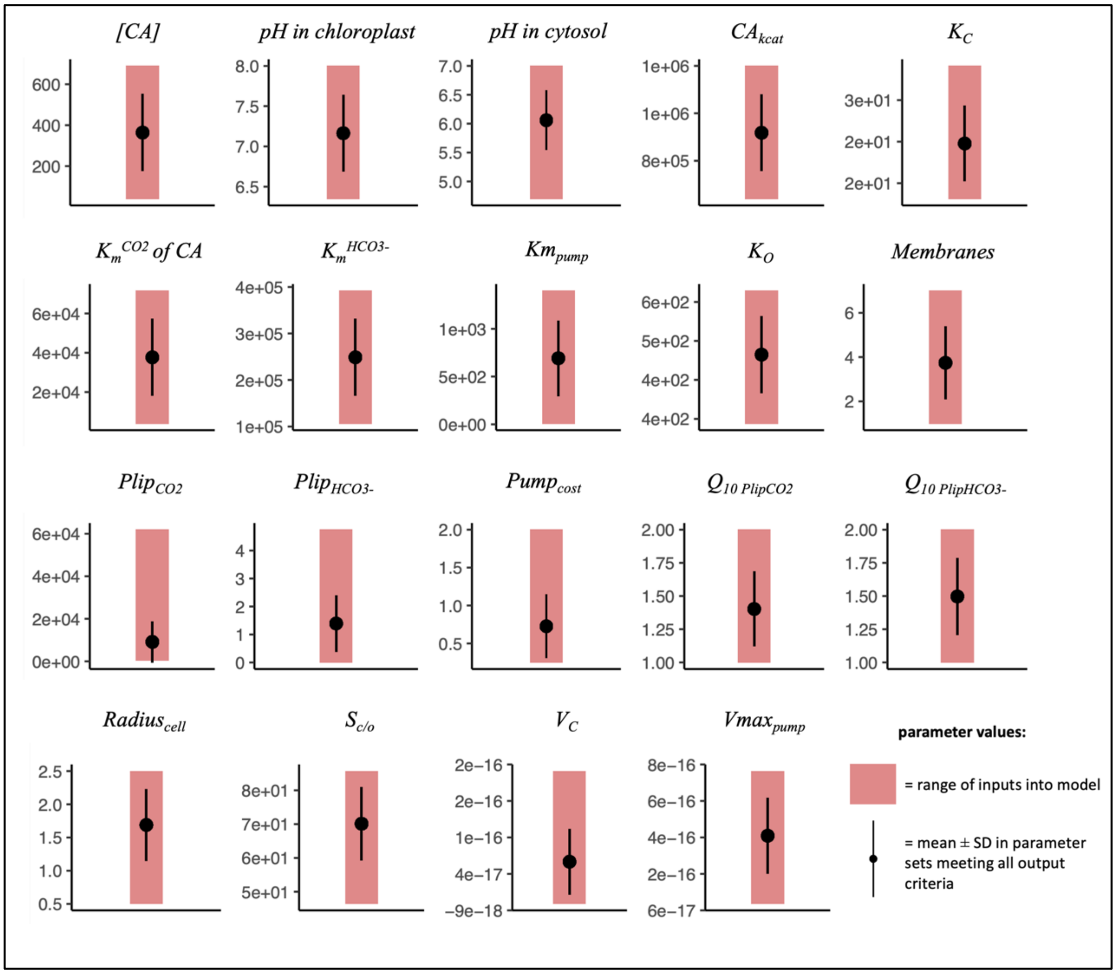


Figure S8. Ranges of parameters in all parameter sets (pink shaded areas) versus in parameter sets meeting all output criteria (black points with error bars indicating ± one standard deviation.