1 **Full title**
2
3 Essentials of a non-canonical algal carbon-concentrating mechanism: modeling with uncertainty
4 quantification
5
6 **Short title**
7
8 Exploring models of carbon-concentration
9
10 **Author line**
11
12 Anne K. Steensma [1,2], Joshua A.M. Kaste [3], Junoh Heo [4], Douglas J. Orr [5], Chih-Li Sung [4],
13 Yair Shachar-Hill [1], Berkley J. Walker [1,2] [*]
14 **Equal contributions (co-first authors):** AKS and JAMK
15 **ORCIDs:** AKS 0000-0003-2146-9417, JAMK 0000-0003-1942-0315, DJO 0000-0003-1217-
16 537X, CLS 0000-0003-4622-5195, YSH 0000-0001-8793-5084, BJW 0000-0001-5932-6468
17
18 **Author affiliations**
19
20   1. Department of Plant Biology, Michigan State University, East Lansing, MI, 48824
21   2. Michigan State University – Department of Energy Plant Research Laboratory, Michigan
22   State University, East Lansing, MI, 48824
23   3. Department of Biochemistry and Molecular Biology, Michigan State University,
24   East Lansing, MI, 48824
25   4. Department of Statistics and Probability, Michigan State University, East Lansing, MI,
26   48824
27   5. Lancaster Environment Center, Lancaster University, Lancaster, United Kingdom, LA1
28   4YQ
29   The current address of author JAMK is: Department of Civil and Environmental
30   Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801.
31
32 **Corresponding author information**
33
34 * denotes corresponding author in author line
35 Berkley J. Walker
36 Room 106 Plant Biology Laboratories, 612 Wilson Road, East Lansing, MI, 48824
37 (517) 355-3928
38 berkley@msu.edu
39
40 **Preprint servers**
41
44
45 **Keywords**
46

carbon-concentrating mechanisms, metabolic modeling, surrogate modeling

Glossary

ATP per CO$_2$: the ratio of adenosine triphosphate consumption flux to net carbon dioxide assimilation flux.
CA: carbonic anhydrase enzyme.
CCM: carbon-concentrating mechanism.
DNN: Deep Neural Network.
FPLC: Fast protein liquid chromatography.
GP: Gaussian Process
laGP: Local approximate Gaussian Process.
NN: single-layer Neural Network.
NRMSE: normalized root-mean-square error.
ODE: ordinary differential equation.
PD plot: partial dependence plot.
Q$_{10}$ (or Q$_{15}$): temperature response factor, representing the response of a parameter when temperature rises by 10 (or 15) degrees.
R_L: respiration in the light, a non-photorespiratory release of CO$_2$ during photosynthesis.
RMSE: root-mean-square error.
SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis.
SHAP: SHapley Additive exPlanations.
Stromal CO$_2$: the steady-state carbon dioxide concentration in the chloroplast stroma.
v_o/v_c: the ratio of oxygen-fixation flux to carbon-fixation flux.
XGBoost: eXtreme Gradient Boosting (a machine-learning model).
Γ_{CO2}: carbon dioxide compensation point, the carbon dioxide concentration at which net carbon assimilation is zero.
Other model parameter definitions are listed in Table S1.

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 we have seen 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,

93   and CO_2 membrane permeability. Our exploration thus suggested that a non-canonical CCM
94   could exist in *C. merolae* and illuminated essential features necessary for CCMs to function
95   generally.
96
97   **Introduction**
98
99   *Cyanidioschyzon merolae* is a red microalga found in moist environments surrounding
100   geothermal sulfur springs. This species is extremophilic, with optimal laboratory growth
101   conditions including low pH (~ 2) and high temperatures (~ 42 °C) (Miyagishima and Wei,
102   2017; Miyagishima *et al.*, 2017). *C. merolae* and other thermo-acidophilic red algae draw
103   interest for their unique biology and simple characteristics, which position them as useful model
104   organisms and as candidates for biotechnology applications (Rahman *et al.*, 2017; Miyagishima
105   and Tanaka, 2021; Seger *et al.*, 2023; Villegas-Valencia *et al.*, 2023). For example, *C. merolae* is
106   of interest because it is one of few organisms which relies on photosynthesis in geothermal
107   spring environments, where hot and acidic conditions restrict the availability of inorganic carbon
108   and challenge biological carbon fixation (Gross, 2000; Miyagishima *et al.*, 2017). Notably,
109   organisms of acid waters can only access approximately 10 micromolar inorganic carbon, as the
110   inorganic carbon pool at acid pH is primarily the volatile species CO_2. In comparison, organisms
111   of near-neutral and alkaline waters may have access to several millimolar of inorganic carbon,
112   due to accumulation of the involatile bicarbonate (Oesterhelt *et al.*, 2007).
113   *C. merolae* is thought to survive in its challenging environment in part by operating a
114   carbon-concentrating mechanism (CCM) (Zenvirth, Volokita and Kaplan, 1985; Rademacher *et*
115   *al.*, 2017; Steensma, Shachar-Hill and Walker, 2023). CCMs boost carbon-fixation efficiency by
116   concentrating CO_2 around rubisco, providing ample substrate for carbon-fixation and inhibiting a
117   competing oxygen-fixation reaction of rubisco. Evidence supporting a CCM in *C. merolae*
118   includes measured accumulation of radiolabeled carbon in the cell, $\delta^{13}C$ consistent with a CCM,
119   transcriptional response of potential CCM genes to CO_2 fluctuations, and substantial CO_2
120   assimilation at low environmental CO_2 concentrations (Zenvirth, Volokita and Kaplan, 1985;
121   Rademacher *et al.*, 2017; Steensma, Shachar-Hill and Walker, 2023). However, many of these
122   indications of the CCM are not definitive: in particular, it is not known how much of *C.*
123   *merolae*'s ability to assimilate CO_2 efficiently could be explained by the affinity of *C. merolae*
124   rubisco for CO_2. Thus, we here provide further evidence for the CCM in *C. merolae* by
125   demonstrating that the affinity of *C. merolae* cells for CO_2 is better than could be explained by
126   the affinity of *C. merolae* rubisco for CO_2.
127   *C. merolae*'s CCM may be described as a "non-canonical" CCM, since the *C. merolae*
128   CCM must operate differently from the few CCM types which are well-characterized.
129   For example, unlike algae and cyanobacteria with well-characterized CCMs, *C. merolae* is not
130   able to take up external bicarbonate, and *C. merolae* lacks anatomy associated with the pyrenoid
131   CCM organelle (Zenvirth, Volokita and Kaplan, 1985; Badger *et al.*, 1998; Misumi *et al.*, 2005;
132   Steensma, Shachar-Hill and Walker, 2023). The absence of these CCM features in *C. merolae*
133   challenges our understanding of what components are required for a functional CCM, and
134   presents the opportunity to define essential CCM components. While previous work has
135   discussed CO_2 as a source of carbon for the CCM (Fridlyand, Kaplan and Reinhold, 1996; Price,
136   2011), there has been little quantitative exploration of whether a CCM could function while
137   lacking both facilitated carbon uptake and specialized compartments such as the pyrenoid or
138   carboxysome. We thus used mathematical modeling, informed by new experimental
139   measurements, to explore how the *C. merolae* CCM may function.

140   Research on CCMs has long employed mathematical models to understand the
141 components of functional CCMs in model cyanobacteria and algae, with a particular area of
142 interest in CCM modeling being the possibility of boosting crop productivity by engineering
143 CCMs into crops which lack CCMs (Price *et al.*, 2013; McGrath and Long, 2014; Fei *et al.*,
144 2022; Kaste, Walker and Shachar-Hill, 2024). By developing modeling approaches to robustly
145 describe CCMs in organisms where biochemical data is limited, such as extremophile algae, we
146 can better understand how organisms survive environmental challenges. Here we add to these
147 engineering efforts by modeling a heat-tolerant CCM with minimal components which offers
148 unique possibilities for plant synthetic biology (Misumi, Kuroiwa and Hirooka, 2017). To draw
149 robust conclusions about cellular characteristics which can support a CCM, we used state-of-the-
150 art statistical methods to define the effects of model parameters on the predicted photosynthetic
151 phenotype while limiting unwarranted *a priori* assumptions. We demonstrate an interdisciplinary
152 modeling approach which efficiently sampled from large parameter spaces and identified
153 features (e.g., compartment permeability, pH, enzyme characteristics) that determine the function
154 and energy cost of a simple CCM. This approach is to our knowledge new to compartmental
155 photosynthetic modeling, and could facilitate effective use of models to inform experiments and
156 rational engineering.
157   Some sets of model input parameters produced model outputs which met empirically-
158 based criteria for functional carbon concentration and efficient energy usage, and we identified
159 input parameters which have substantial impacts on the model outputs. Overall, our model of a
160 hypothetical biophysical CCM which requires minimal enzymes and anatomical features (**Figure
161 1**) appears to represent a feasible CCM structure in *C. merolae*, which invites further research
162 into the sources of environmental resilience in extremophile algae.
163
164 **Methods**
165
166 **Experimental data collection: gas-exchange measurements**
167 *Cyanidioschyzon merolae* 10D was grown as cultures in Erlenmeyer flasks in 50 mL of medium
168 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
169 mL L^{-1} Hutner's Trace Elements solution, and H_2SO_4 to pH 2.7 (recipe modified from MA2
170 medium recipe of (Fujiwara and Ohnuma, 2017)). Cultures were maintained at 40 °C under 100
171 µmol m^{-2} s^{-1} white light, with aeration by shaking at 100 rpm. For gas-exchange measurements,
172 cultures of OD_{750} 1.0 – 1.2 were resuspended in growth medium to OD_{750} 0.6 ($1.60x10^7$ –
173 $3.68x10^7$ cells/mL). Gas-exchange parameters were measured in a LI-6800-18 Aquatic Chamber
174 (LI-COR Biosciences) at 45 °C and with normalization to cell count data from a hemocytometer
175 slide, following the procedures of (Steensma, Shachar-Hill and Walker, 2023) and with a
176 protocol similar to (Davey and Lawson, 2024).
177
178 **Experimental data collection: rubisco kinetics measurements**
179 We purified rubisco from *C. merolae* biomass with a protocol adapted from (Miyagishima and
180 Wei, 2017; Orr and Carmo-Silva, 2018). Approximately 60 grams of biomass were lysed by
181 freeze-thawing followed by mechanical homogenization. Crude rubisco was polyethylene-
182 glycol-precipitated from clarified homogenate and purified by fast protein liquid chromatography
183 (FPLC). FPLC fractions eluting under the major UV trace peak were assayed by SDS-PAGE and
184 by spectrophotometric rubisco activity assay (procedures adapted from (Kubien, Brown and
185 Kane, 2010; Carter *et al.*, 2013)) (**Figure S3**). Fractions containing active semi-pure rubisco

186 were pooled, concentrated with a 100 kDa centrifugal concentration filter, and snap-frozen for
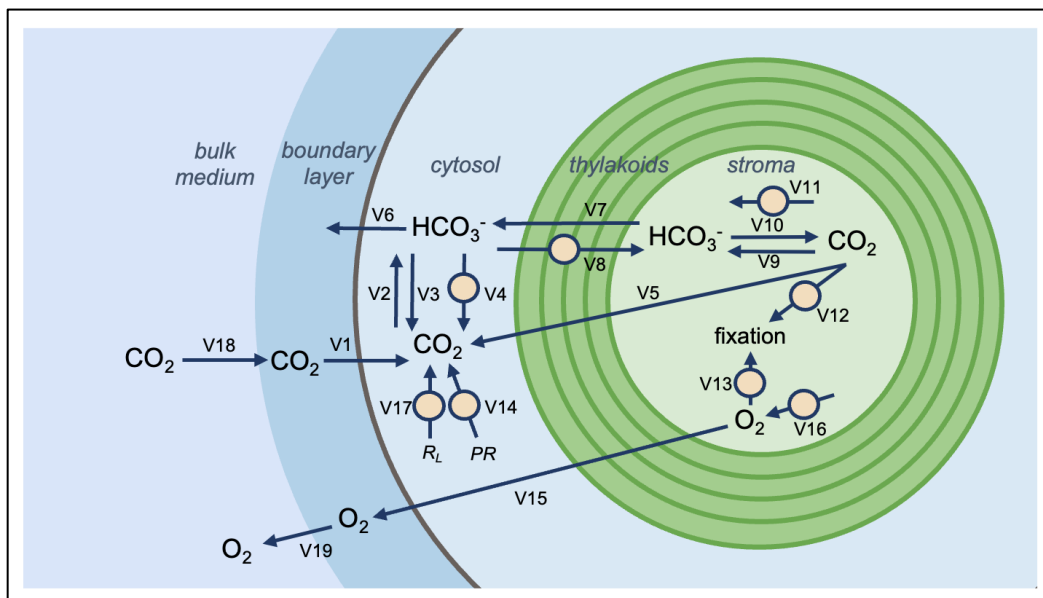187 use in rubisco assays.
188   Purified rubisco was used to determine catalytic properties as described previously in
189 detail (Prins *et al.*, 2016), with some alterations to protein desalting and activation: concentrated
190 protein aliquots were first diluted with activation mix containing 100 mM Bicine-NaOH pH 8.0,
191 20 mM $MgCl_2$, 10 mM $NaHCO_3$, and 1 % (v/v) Plant Protease Inhibitor cocktail (Sigma-Aldrich,
192 UK). Rubisco was then activated at 45 °C for 15 min before being used in $^{14}CO_2$ consumption
193 assays at either 25 °C or 45 °C with CO_2 concentrations of 8, 16, 24, 36, 68, and 100 μM. To
194 determine K_O, these CO_2 concentrations were combined with concentrations of either 0, 21, 40,
195 or 70 % (v/v) O_2. $kcat_C$ was determined using measurements with 0% O_2. An aliquot of the
196 activated protein was used for determination of Rubisco active sites via ^{14}C-CABP binding using
197 the method of (Sharwood, Ghannoum and Whitney, 2016). For ^{14}C-CABP binding, protein
198 aliquots were incubated at 45°C for 15 mins with ^{14}C-CABP to maximize binding, prior to
199 application to Sephadex columns as previously described (Loganathan, Tsai and Mueller-Cajar,
200 2016). Aliquots were also analyzed via SDS-PAGE alongside known concentrations of plant
201 type Rubisco to strengthen estimates of Rubisco content.
202
203 **Model details**
204 The hypothetical CCM described in this study (**Figure 1**) was modeled as a set of well-mixed
205 compartments and represented as a system of ordinary differential equations (ODEs). In this
206 minimal biophysical CCM, carbon diffuses into the cell as CO_2, is trapped in the cytosol as
207 bicarbonate by action of carbonic anhydrase, and is pumped into the chloroplast, where a second
208 carbonic anhydrase provides CO_2 around rubisco. No pyrenoid diffusion barrier is present, as
209 neither a starch sheath nor a clear organized subcompartment for rubisco have been described in
210 *C. merolae*. However, we accounted for potential effects of the concentric thylakoids which are
211 present in *C. merolae* and many other aquatic photosynthetic organisms (Ichinose and Iwane,
212 2017). Carbonic anhydrases (CAs) and bicarbonate transporters are essential components of
213 known biophysical CCMs and thus essential components of a CCM model (Beardall and Raven,
214 2020). These components (V4, V11, V8) are discussed in more detail below.
215

216

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.*

The model geometry is based on the cellular structure of *C. merolae* as apparent in published micrographs of this alga (Kuroiwa, 1998; Miyagishima *et al.*, 1998; Toda *et al.*, 1998; Itoh *et al.*, 1999; Yagisawa *et al.*, 2012, 2016; Ichinose and Iwane, 2017; Reimer *et al.*, 2017; Sato *et al.*, 2017; Moriyama *et al.*, 2018). 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 contains a cytosol of radius $Radius_{cell}$ and a chloroplast stroma space of radius $0.25 * Radius_{cell}$. The cell is surrounded by a medium boundary layer of radius $2 * Radius_{cell}$, beyond which lies an infinite external medium. Though varying fluid dynamic conditions strongly impact the size of boundary layers such as gas surface films or phycospheres, these layers are reported to be on the order of magnitude of 1 cell radius (Guterman and Ben-Yaakov, 1987; Seymour *et al.*, 2017).

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 microscopy examinations (Ichinose and Iwane, 2017). 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 (V1, V6, V5, V7, V15) was described using estimates of conductivity of lipid membranes to the chemical species in question:

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

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 between the medium boundary layer and bulk medium (V18, V19) 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:

$$\frac{r_1}{r_2} = \sqrt{\frac{M_1}{M_2}}\ \#(E2)$$

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 (V5), HCO_3^- (V7), and O_2 (V15) through variable numbers of stacked thylakoid membranes, an overall conductivity through all of the layers was calculated as:

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

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^- , as in V2, V3, V9, and V10 (**E4-5**), was described using simple first-order kinetics, accordring to the rate constant of the dehydration (slower) step of the interconversion:

$$J_{CO_2\ hydration} = k_2[CO_2] \#(\textbf{E4})$$
$$J_{HCO_3^-\ dehydration} = k_{-2}[HCO_3^-][H^+] \#(\textbf{E5})$$

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. It was observed in (Mangan *et al.*, 2016) that the significantly higher permeability of H_2CO_3 relative to $HCO3^-$, coupled with the rapid interconversion of these species, results in a greater permeability through lipid membranes of this joint H_2CO_3/HCO_3^- pool than would be expected from HCO_3^- permeability alone. To account for this while accommodating the simplification of not including the H_2CO_3 species, we explored a range of possible lipid permeabilities to HCO_3^- and CO_2 that substantially overlaps with the range of inorganic carbon permeability values from (Mangan *et al.*, 2016).

The interconversion of CO_2 and HCO_3^- by carbonic anhydrase (V4, V11) was described as in (McGrath and Long, 2014):

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

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 (Lavigne, Proye and Gattuso, 2019). Other potentially temperature-sensitive parameters receive temperature adjustments according to Q_{10} or Q_{15} factors as in (von Caemmerer, 2000). In *C. merolae*, CA inhibitors have not been shown to affect oxygen evolution, but it remains plausible that CAs are involved in photosynthesis, since genes homologous to CCM CAs show transcript increases in response to lowered CO_2 availability (Rademacher *et al.*, 2017; Parys *et al.*, 2021). Of the two putative CAs with the most dramatic transcriptional response to CO_2, one protein has a computationally-predicted chloroplast targeting sequence and has been fluorescence-localized between the mitochondrion and chloroplast, while the other protein has no predicted targeting sequence and has been fluorescence-localized in the cytosol (Rademacher *et al.*, 2017; Steensma, Shachar-Hill and Walker, 2023).

295   Carboxylation by rubisco (V12) was described as with the assumption that CO_2 is
296   limiting, as in (Farquhar, von Caemmerer and Berry, 1980):

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

297   To estimate oxygenation (V13), we estimate v_c/v_o (carboxylation flux over oxygenation flux)
298   from the CO_2/O_2 specificity ($S_{c/o}$) of rubisco and chloroplast CO_2 and O_2 concentrations (**E8**),
299   and then use this to arrive at v_o.

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

300   The pumping of HCO_3^- across the stack of thylakoid membranes by a bicarbonate pump (V8)
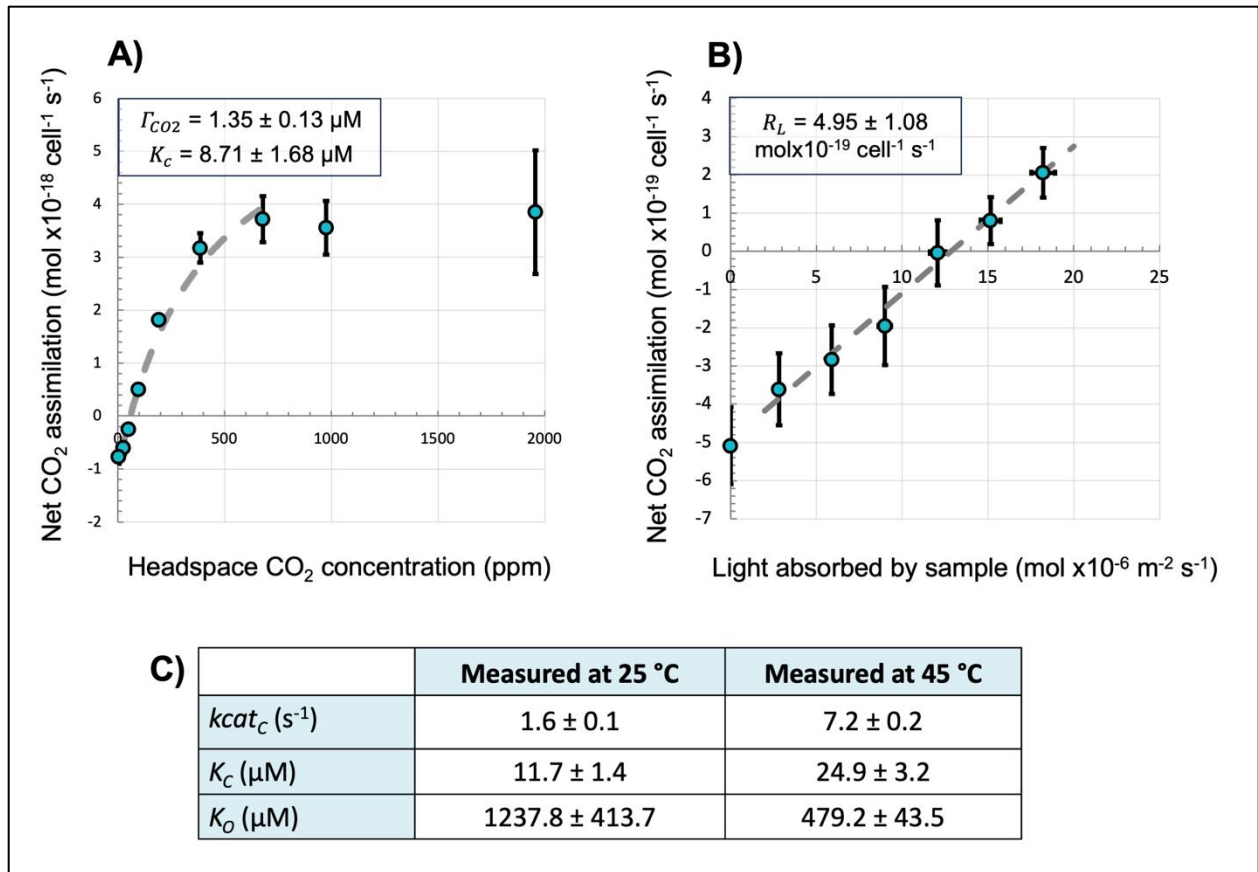301   was described by simple Michaelis-Menten kinetics:

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

302   Concerning what is known about bicarbonate transport in *C. merolae*, it is difficult to identify
303   bicarbonate transporters by homology (Price and Howitt, 2011; Steensma, Shachar-Hill and
304   Walker, 2023). *C. merolae* would have minimal access to extracellular bicarbonate even if
305   bicarbonate were substantially available in its acidic environment, as is evident from
306   radiolabelling and from gas-exchange conducted at varying pH (Zenvirth, Volokita and Kaplan,
307   1985; Steensma, Shachar-Hill and Walker, 2023). Bicarbonate transport at the chloroplast or
308   thylakoids is an key feature of biophysical CCMs (Price *et al.*, 2008; Spalding, 2008).
309   Photorespiratory CO_2 release (V14) and photosynthetic oxygen evolution (V16) were
310   determined by the stoichiometry described in **S1 Supporting Information.** Non-
311   photorespiratory CO_2 release occurring during photosynthesis, known as respiration in the light
312   (R_L) (Xu *et al.*, 2021), was estimated from gas-exchange data according to a modified Kok
313   method (V17). Assimilation was measured under sub-saturating light intensities and extrapolated
314   to estimate CO_2 release in the absence of light (**Figure 2B**). The resulting mean measured value
315   of R_L was normalized to cell size for use in the model: we assume that the empirical
316   measurement of R_L we obtained was, on a per cell basis, characteristic of a *C. merolae* cell of a
317   radius of 1 μm. Under the assumption that R_L should vary proportionally with cell volume, we
318   normalized R_L as follows:
319

320             $$R_{L normalized} = R_{L measured}\frac{Volume}{Volume_{1um}}\#(\boldsymbol{E10})$$

321

A)

$\Gamma_{CO2} = 1.35 \pm 0.13$ μM
$K_c = 8.71 \pm 1.68$ μM

Net CO$_2$ assimilation (mol x10^{-18} cell^{-1} s^{-1})

Headspace CO$_2$ concentration (ppm)

B)

$R_L = 4.95 \pm 1.08$ molx10^{-19} cell^{-1} s^{-1}

Net CO$_2$ assimilation (mol x10^{-19} cell^{-1} s^{-1})

Light absorbed by sample (mol x10^{-6} m^{-2} s^{-1})

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

322

323 **Figure 2.** *Experimental data incorporated into the model. (**A, B**). Response of net assimilation in*
324 C. merolae *to (**A**) CO$_2$ availability and (**B**) light availability. Points are mean ± SE (n = 3), and*
325 *parameters calculated from the data are indicated in the upper left corner of each plot as mean*
326 *± SE. Dashed lines indicate trend fits used to determine Michaelis-Menten constant of CO$_2$*
327 *fixation (K$_C$) and respiration in the light (R$_L$). The linear fit used to determine CO$_2$*
328 *compensation point (Γ$_{CO2}$) is not pictured but is described in **Methods**. (**C**) Kinetic properties of*
329 C. merolae *rubisco. Rubisco turnover rate for CO$_2$ fixation (kcat$_C$), Michaelis-Menten constant*
330 *of CO$_2$ fixation (K$_C$), and Michaelis-Menten constant of O$_2$ fixation (K$_O$) were measured at 25*
331 *and 45 °C. Data is mean ± SE, n = 4.*

332

333 ATP costs for the cell were estimated as:

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

334 Where *Membranes* is the number of thylakoid stacks and *Pump$_{cost}$* is the assumed cost, in ATP,
335 of pumping a single HCO$_3^-$ ion across a lipid bilayer by the hypothesized pump.
336   A full list of all flux equations and the system of ODEs used to describe the system can
337 be found in **S1 Supporting Information**.

338

Definition of reasonable model output values

340 To ensure the model reproduced experimental results, we used newly measured and published
341 experimental data to set acceptable bounds for the following model outputs: CO$_2$ compensation
342 point (Γ$_{CO2}$), the ratio of ATP consumption flux to net CO$_2$ assimilation flux (ATP per CO$_2$), the

343   steady-state CO_2 concentration in the chloroplast stroma (stromal CO_2), and the ratio of oxygen-
344   fixation flux to carbon-fixation flux (v_o/v_c).
345
346   *CO_2 compensation point (Γ_{CO2})*
347   We accepted *Γ_{CO2}* values less than or equal to 2.70 μM, corresponding to no more than twice the
348   mean measured value (**Figure 2**).
349
350   *Ratio of ATP consumption flux to net CO_2 assimilation flux (ATP per CO_2)*
351   We accepted ATP per CO_2 values which were less than or equal to 25 and greater than 0. These
352   bounds are supported by measured light response curves which indicated how much additional
353   light absorption drives a certain amount of additional CO_2 assimilation (**Figure 2**). We used this
354   data to estimate how much additional ATP production drives an additional CO_2 assimilation,
355   using the photon per ATP values for various light-reaction pathways (Walker *et al.*, 2020), the
356   cylindrical geometry of the gas-exchange sample chamber, and the measured density of cells in
357   the sample. The resulting estimated values were: 13.8 ± 2.19 ATP produced/CO_2 assimilated
358   (mean ± SE, assuming cyclic and linear electron flow operating equally) or 17.4 ± 2.76 ATP
359   produced/CO_2 assimilated (mean ± SE, assuming linear electron flow only operating). This
360   suggests that ATP per CO_2 values of up to ~25 are supported by photosynthetic electron flow.
361   The lower bound of the acceptable range excludes a few parameter sets outputting negative ATP
362   per CO_2, since these parameter sets represented particularly non-functional CCM scenarios with
363   negative net assimilation values under ambient CO_2 conditions.
364
365   *Steady-state CO_2 concentration in the chloroplast stroma (stromal CO_2)*
366   We accepted chloroplast CO_2 concentration values of greater than or equal to the CO_2
367   concentration in the medium under 400 ppm CO_2 atmosphere, by the logic that a functional CCM
368   should result in rubisco accessing a greater CO_2 concentration than is available from ambient
369   medium.
370
371   *Ratio of oxygen fixation flux to carbon fixation flux (v_o/v_c)*
372   We accepted *v_o/v_c* values less than or equal to 0.3, based on data and models indicating that
373   plants without CCMs are unlikely to achieve *v_o/v_c* less than approximately 0.3 (Bellasio *et al.*,
374   2014).
375
376   **Model optimization and estimation of simulated compensation point**
377   Steady-state fluxes and metabolite concentrations were solved using *odeint()* from Python's
378   SciPy library (Virtanen *et al.*, 2020) with error control handled by maintaining the following
379   inequality:

$$\max\left(\frac{errors(y)}{error_{weights}(y)}\right) \le 1$$

380   Where *errors* is a vector of local errors against computed outputs *y* and *$error_{weights}$* is a vector of
381   weights:

$$error_{weights} = tolerance_{relative} * |y| + tolerance_{absolute}$$

382   Where *$tolerance_{relative}$* and *$tolerance_{absolute}$* are the relative and absolute tolerance values set in the
383   *odeint()* solver. We use the default value of for these tolerances from SciPy version 1.10.0. All
384   simulations were verified to reach steady-state (metabolite concentration solutions changing
385   0.01% or less from previous value). An end time of sufficient length was chosen to ensure that

simulations successfully reached steady-state. The maximum number of step sizes allowed for each time point was manually set to 5,000 as this was found to allow our simulations to reach steady-state without optimization difficulties. Other optimization parameters, such as the maximum and minimum step sizes, were left at their default settings as well and controlled by the optimizer. Using these settings, 100% (240,000/240,000) of all simulations successfully reached a steady-state solution in all model architectures.

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 (McKay, Beckman and Conover, 1979) 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. CO_2 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.

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 additional results from a mechanistic model. Surrogate modeling additionally gave us access to powerful statistical tools for machine-learning model analysis, including SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) and partial dependence (PD) plots (Friedman, 2001).

To identify the optimal surrogate model for parameter exploration, we compared four popular machine-learning models: eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), Local approximate Gaussian Process (laGP) (Gramacy and Apley, 2015), single-layer Neural Network (NN) (James *et al.*, 2013), and Deep Neural Network (DNN) (Chen and Guestrin, 2016). We collected a 240,000-sized dataset, where the outputs were simulated from the mechanistic 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 dataset was divided into training and test sets using a random sampling approach. Specifically, we used the *sample()* function in R with a fixed seed. 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.

428 The XGBoost model was trained using a max number of boosting iterations of 1000 with
429 the evaluation metric of the root-mean-square error. The laGP model used the nearest neighbor
430 method for prediction. The NN model is a simple feedforward neural network with a logistic
431 activation function $\frac{1}{1+e^{-x}}$ for regression tasks. The error function used for the calculation of the
432 error was the sum of squared errors. The threshold parameter for the partial derivatives of the
433 error function as stopping criteria for the NN model was set to half the range of the target
434 variable.

435 The DNN model consists of two hidden layers containing 64 and 32 units respectively,
436 both using rectified linear unit (ReLU) activation functions max(x, 0). The DNN model was
437 trained using the adaptive moment estimation (Adam) optimizer and mean squared error (MSE)
438 as the loss function. The model was trained for 40 epochs, with the learning algorithm processing
439 the entire training dataset 40 times. A batch size of 240 was used, indicating the number of
440 samples processed before updating the model's internal parameters. Moreover, 20% of the
441 training data was set aside for validation purposes during the training process.

442 **Results and Discussion**

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

446 In previous work, we determine that if *C. merolae* has rubisco kinetics similar to other red algae,
447 then this alga must operate a CCM to maintain its measured photosynthetic efficiency.
448 Alternatively, its measured photosynthetic efficiency could be explained by unprecedented
449 rubisco kinetics, meaning enzyme properties favoring carbon-fixation over oxygen-fixation to an
450 unprecedented degree (Steensma, Shachar-Hill and Walker, 2023). Here we confirmed that *C.
451 merolae* rubisco kinetics are similar to those of other red-type (Form 1D) rubiscos (Read and
452 Tabita, 1994; Uemura *et al.*, 1997; Whitney *et al.*, 2001). *C. merolae* rubisco had a strong
453 affinity for CO_2 (low K_C), a poor affinity for O_2 (high K_O), and a slow carboxylation rate (low
454 $kcat_C$) (**Figure 2**). Consistent with other studies, $kcat_C$ and K_C were higher when measured at
455 increased temperature, while K_O was lower. Although K_O is in the denominator of rubisco
456 specificity ($S_{c/o}$) and $S_{c/o}$ decreases with increased temperature, *in vitro* K_O is observed to
457 decrease with increased assay temperature in some species (Jordan and Ogren, 1984; Uemura *et
458 al.*, 1997; Prins *et al.*, 2016).
459 These kinetics findings indicated *C. merolae* does operate a CCM, as *C. merolae* cells
460 had higher affinity for CO_2 than *C. merolae* rubisco (8.71 ± 1.7 μM cell K_C vs. 24.9 ± 3.2 μM
461 rubisco K_C at 45 °C, p = 0.008 by two-sample *t*-test) (**Figure 2**). This result adds to the evidence
462 of a CCM in *C. merolae* (Zenvirth, Volokita and Kaplan, 1985; Rademacher *et al.*, 2017;
463 Steensma, Shachar-Hill and Walker, 2023).

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

468 To explore how the *C. merolae* CCM may operate, we constructed a functional model of a CCM
469 (**Figure 1**). This model demonstrated that there were parameter sets consistent with the empirical
470 literature that result in a functional CCM, despite the minimal model structure lacking structures
471 like a pyrenoid or carboxysome (**Figure 3**). Cyanobacterial CCM models have also supported

472 reduction to a simple model with only two compartments from the cell membrane inwards
473 (Mangan and Brenner, 2014).
474   Our results provided quantitative support for a CCM taking inorganic carbon from the
475 environment solely through CO_2 diffusion into the cell without specialized compartments, which
476 we term a "non-canonical" CCM due to its differences in structure and function from CCMs that
477 have been characterized in detail. *C. merolae* has a different structure and environment than the
478 "canonical" CCMs of *Chlamydomonas reinhardtii* and of model cyanobacteria, which allowed
479 us to explore a biology and a parameter space which are different from those in previous CCM
480 models.
481   Though there is speculation that extremophilic red algae may use a C_4-like CCM, it has
482 been previously proposed that acidophile algae may accumulate carbon by a "bicarbonate-trap"
483 or "acid-loading" mechanism similar to our modeled CCM (Gehl and Colman, 1985; Fridlyand,
484 1997; Gross, 2000; Rademacher *et al.*, 2016; Curien *et al.*, 2021; Fei *et al.*, 2022). Briefly, this
485 mechanism would involve bicarbonate being concentrated for enzymatic action by bringing
486 inorganic carbon speciation near equilibrium in near-neutral cellular compartments, since the
487 predominant inorganic carbon species from pH ~6 to ~10 is the poorly-membrane-permeable
488 bicarbonate.
489   Various facilitated CO_2 uptake mechanisms exist in CCM-containing organisms, such as
490 the NDH-I complexes in cyanobacteria and the periplasmic CA system in algae (Fridlyand,
491 Kaplan and Reinhold, 1996; Moroney *et al.*, 2011; Price, 2011). We here test a different model
492 where inorganic carbon enters the cell solely by passive CO_2 diffusion into the cytosol, followed
493 by the action of non-vectorial cytosolic carbonic anhydrase. In contrast to the well-studied
494 cyanobacterial and algal systems, where growth under limiting CO_2 is supported by active
495 bicarbonate uptake and the accumulation of cytosolic bicarbonate above equilibrium levels
496 (Price and Badger, 1989; Price *et al.*, 2004; Duanmu *et al.*, 2009), our model functions as a CCM
497 without taking any bicarbonate from the environment.
498   Another unique feature of our model is the nature of the diffusion barrier surrounding
499 rubisco. Cyanobacteria encapsulate rubisco in a proteinaceous shell called the carboxysome,
500 which is thought to provide a diffusion barrier to CO_2 (Price *et al.*, 2008). The model alga *C.*
501 *reinhardtii* aggregates rubisco into an organelle called the pyrenoid, which in wild-type cells is
502 surrounded by a starch sheath that may serve as a diffusion barrier. In contrast to the well-studied
503 system of *C. reinhardtii*, there has been comparatively less investigation into algae which lack
504 starch sheaths or lack pyrenoids entirely (Morita *et al.*, 1999; Barrett, Girr and Mackinder, 2021).
505 Thus, to broaden our knowledge of CCM anatomy, we modeled an arrangement where rubisco is
506 diffuse within a series of concentric thylakoid membranes. This allowed us to further investigate
507 whether membranes, which are thought to be highly permeable to CO_2 (Gutknecht, Bisson and
508 Tosteson, 1977; Missner *et al.*, 2008), could impact carbon-concentration, and how carbon-
509 concentration could function without a carboxysome or pyrenoid.
510
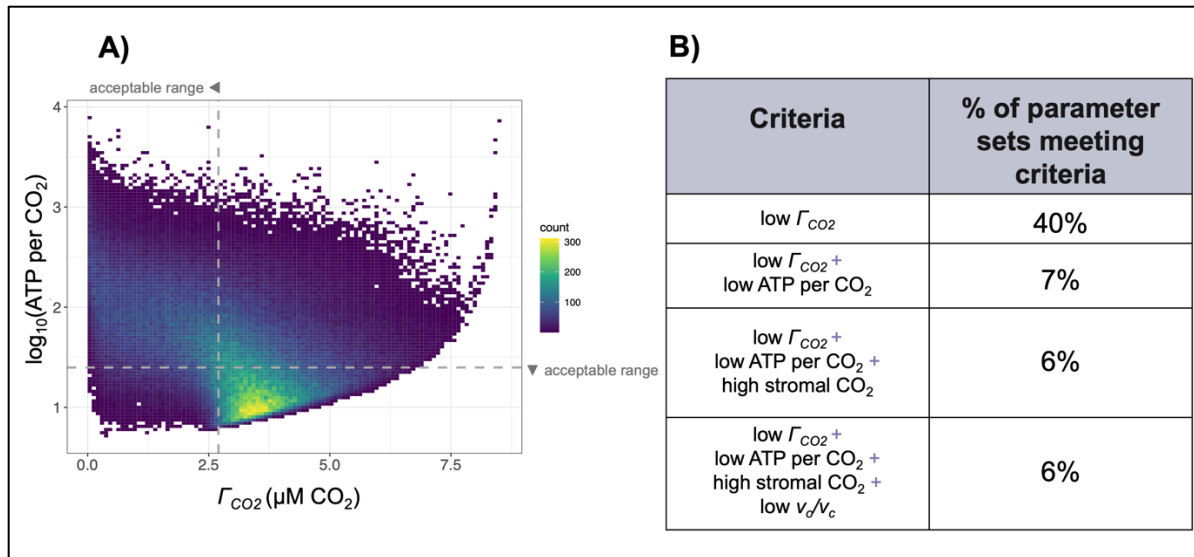
Figure 3. Values of key model outputs. *(A) Parameter sets are organized into a 2-dimensional histogram according to their output values of Γ_{CO2} and ATP per CO$_2$, with dashed lines indicating bounds for acceptable values of these outputs. 80 parameter sets (0.03% of total) are not pictured on the 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.*

To investigate these and other features of interest, 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 hypothetical *C. merolae* CCM and which were not well-characterized physical constants (**Table S1**). We thus sampled input parameter sets with varying numbers for these parameters of interest. We sampled parameter sets through a Latin hypercube design (McKay, Beckman and Conover, 1979) which 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 **S1 Supporting Information**). 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 of carbon transport (**Figures 4, S1 – S2**). This is consistent with other modeling studies indicating that thylakoid membranes could affect inorganic carbon

543 diffusion, and with observations of pyrenoids surrounded by layers of thylakoids in hornworts
544 (Thoms, Pahlow and Wolf-Gladrow, 2001; Fei *et al.*, 2022; Robison *et al.*, 2024).

545
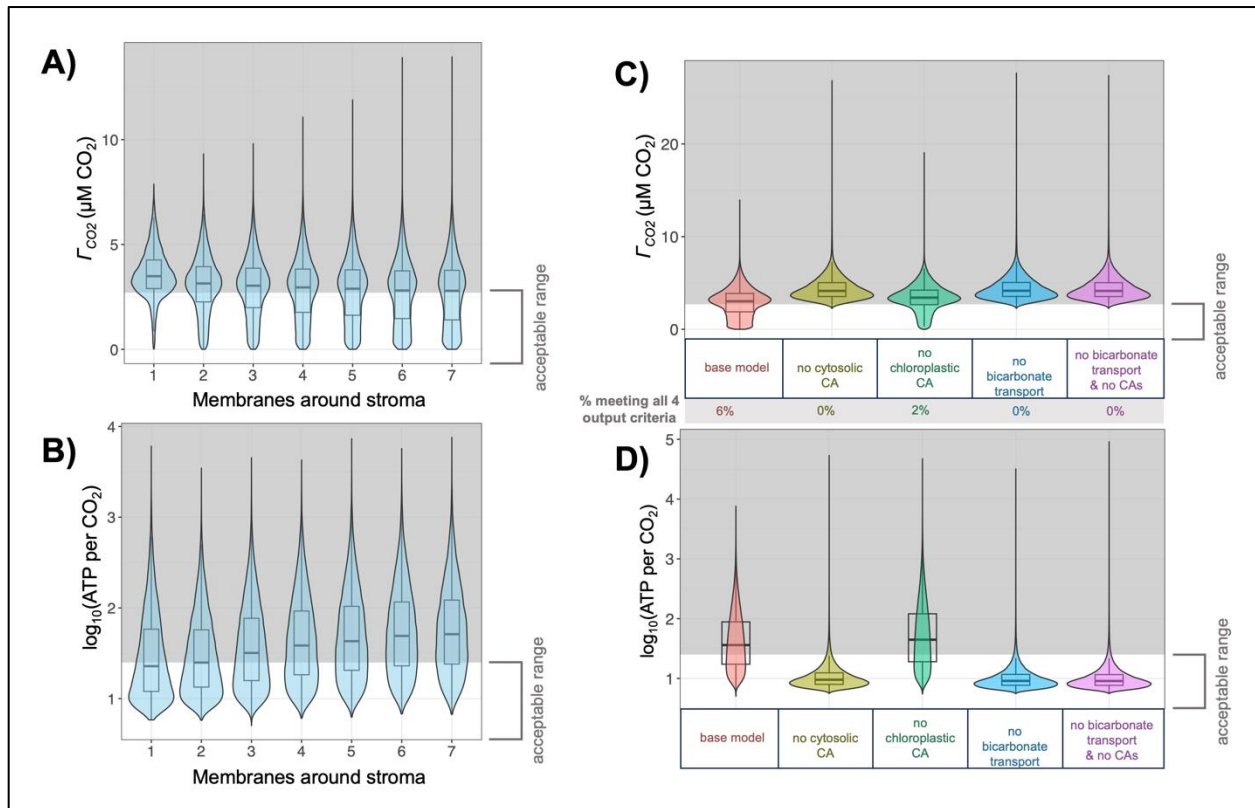
546
547 *Figure 4. Effect of select input parameters on key model outputs. (A, B) Effect of model input*
548 *parameter* Membranes *(x-axis) on key model outputs. Distribution of parameter set outputs for*
549 *each value of* Membranes *is represented by a box plot overlaid on a violin plot. Shaded areas*
550 *represent unacceptable values of outputs. (A) Effect of* Membranes *on model output Γ_{CO_2}. (B)*
551 *Effect of* Membranes *on model output ATP per CO_2. 80 parameter sets (0.03% of total) are not*
552 *pictured in this panel, as they produced negative ATP per CO_2 values and could not be log-*
553 *transformed. (C, D) Effect on key model outputs when bicarbonate transport or carbonic*
554 *anhydrases (CAs) are removed from the model. Distribution of parameter set outputs for each*
555 *scenario is represented by a box plot overlaid on a violin plot. Shaded areas represent out-of-*
556 *bounds values of outputs. The same sampling of input parameter sets was run through models*
557 *representing each scenario. (C) Γ_{CO_2} in model scenarios where various model features removed,*
558 *with indication of how many parameter sets met output criteria in each scenario. (D) ATP per*
559 *CO_2 in model scenarios where bicarbonate transport activity at the chloroplast boundary is*
560 *removed. 6,991 parameter sets producing negative ATP per CO_2 values (0.6% of total) are not*
561 *pictured in this panel.*

562
563 **Machine-learning-based surrogate models identified the parameters that most influence**
564 **CCM efficiency**

565
566 Like most mathematical models of photosynthetic systems, this model faced the challenge of
567 drawing robust conclusions while using parameters which, although bounded by their
568 relationship to physical processes, have substantial uncertainty (**Table S1**). To model a system

15

569 with limited biochemical data while not constraining input parameters to a greater degree than
570 was supported by the literature, it was important to assess uncertainties which seemed likely to
571 have substantial and interdependent effects on the model. For example, the input parameter
572 describing permeability of a lipid bilayer to CO_2 ($Plip_{CO2}$) has reported values ranging over
573 several orders of magnitude (**Table S1**). Furthermore, the effect of $Plip_{CO2}$ in the model
574 depended on the value of other parameters, such as the number of lipid bilayers which pose a
575 barrier to carbon moving between the stroma and cytosol (*Membranes*). Various sensitivity
576 analyses are available for ODE models, but $Plip_{CO2}$ and similar parameters were unlikely to be
577 satisfactorily explored by classical local sensitivity analyses, which involve tracking model
578 outputs when individual parameters are varied by a set fraction of the parameter's original value.
579 Therefore, to reveal which model conditions were necessary for the modeled CCM to function
580 biologically, and to identify interesting directions for future investigation, we used statistical
581 methods to identify impactful parameters and to identify which input spaces corresponded to
582 target output ranges. These statistical methods involved training a surrogate machine-learning
583 model on our CCM model inputs and outputs. Interpretations of this surrogate model identified
584 which zones in the input parameter space contained the most combinations fulfilling output
585 criteria (**Figure 5 lower left**), quantified how much each input parameter affected the prediction
586 of outputs by the surrogate model (**Figure 5 upper right**), and visualized the response of model
587 outputs to inputs (**Figures S4 – S7**).
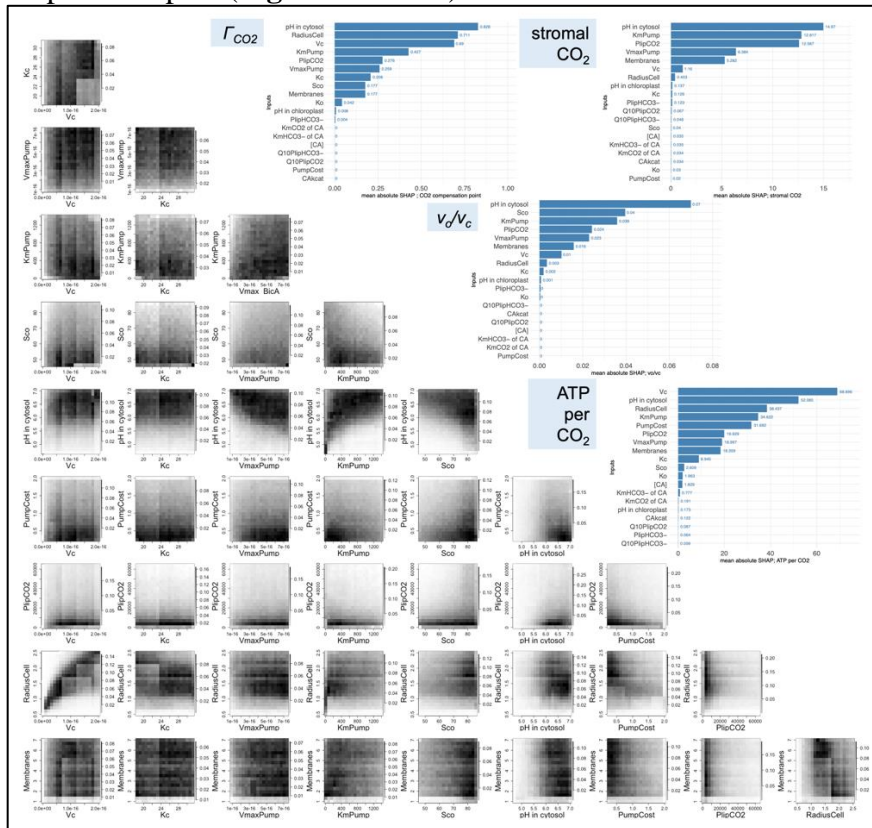
588
589 ***Figure 5. Statistical investigation of parameters affecting model output.*** *(upper right bar plots)*
590 *Mean absolute SHapley Additive exPlanations (SHAP) plots for each output criterion. (lower left*
591 *density plots) Density plots of parameter sets meeting all output criteria, organized by selected*
592 *pairwise input parameter (input parameters pictured are those input parameters with high SHAP*

593 *values for all output criteria). Darker areas indicate areas where more parameter sets meeting*
594 *criteria occur. Scales of color vary for each plot).*
595

596   Some input parameters had little impact on model outputs with the tested input ranges.
597 For these parameters, values from across the input range were evenly represented in the
598 parameter sets meeting all output criteria. The parameters with relatively little impact on outputs
599 included values related to carbonic anhydrase concentration and kinetics (*[CA]*, *CAkcat*, Km_{CO2}
600 and Km_{HCO3-} for carbonic anhydrases), chloroplast pH, and values related to bicarbonate
601 membrane permeability (*$Plip_{HCO3-}$*, *$Q10_{PlipHCO3-}$*, **Figures 5, S4 – S8**). While it is possible that
602 these aspects of the CCM may become impactful if varied beyond the tested range (e.g., if
603 engineering efforts produce carbonic anhydrase concentrations falling outside the range of
604 literature values we used), these parameters did not emerge as particularly impactful in our
605 exploration. Due to how fast the interconversion of inorganic carbon species by carbonic
606 anhydrase is, the enzyme is likely capable of keeping inorganic carbon species close to their
607 equilibrium concentrations across the range of values we explored for its kinetics. Given this, it
608 is unsurprising that model outputs varied little with respect to carbonic-anhydrase-related
609 parameters, even though the complete absence of these enzymes was deleterious (**Figure 4**).
610   Other parameters were more constraining in the model, indicating their importance in
611 producing a functional CCM. For example, six parameters appeared to impact all four of the
612 target model outputs in the mean absolute SHAP plots: *V_c*, *$Vmax_{pump}$*, *Km_{pump}*, pH in the cytosol,
613 *$Plip_{CO2}$*, and *Membranes* (**Figure 5**). Sobol′ analysis (Sobol′, 2001) of the surrogate model
614 produced similar results (**Figure S9**). As might be expected in a model relying on a cytosolic
615 bicarbonate trap followed by bicarbonate pumping, parameter sets that successfully and
616 efficiently concentrated carbon tended to have cytosolic pH at or above the pH where
617 bicarbonate predominates (cytosol pH above 6), and tended to have a lower ATP cost of
618 pumping bicarbonate (low *$Pump_{cost}$*), as well as faster and higher-affinity bicarbonate pumps
619 (high *$Vmax_{pump}$*, low *Km_{pump}*) (**Figure 5**).
620   Other features enriched in parameter sets meeting output criteria were a cell radius in the
621 middle of the input range (moderate *$Radius_{cell}$*), and a lower CO_2 membrane permeability (low
622 *$Plip_{CO2}$*, **Figure 5, Figure S4 – S9**). This suggested an important relationship between the
623 volumes where metabolism occurs and the surface areas which present diffusion barriers
624 between compartments. As the radius of the cell increases, CO_2 loss from R_L may overcome the
625 ability of the cell to acquire carbon through passive diffusion into the cell. Conversely, as the
626 radius of the cell decreases, less absolute bicarbonate pumping would be necessary to achieve
627 high rubisco saturation, especially when rubisco is slow (low *V_c*). In low-radius scenarios, "over-
628 pumping" bicarbonate could reduce energy efficiency.
629

630 ***In silico* knockouts identified experimental targets for further characterization of the *C.**
631 ***merolae* CCM**
632

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

638 The modeled CCM functioned without fine details of cellular structure that support
639 photosynthesis in other organisms, such as rubisco aggregation into an area smaller than the
640 stroma, carbonic anhydrases with restricted distributions and directions (i.e., lumenal and
641 vectorial carbonic anhydrases), recapture of mitochondrially-respired CO_2, and perforations or
642 interconnections in concentric thylakoids (Nevo *et al.*, 2007; Rademacher *et al.*, 2017; Barrett,
643 Girr and Mackinder, 2021). Our work thus expands on previous models with detailed chloroplast
644 geometry (Fei *et al.*, 2022) by demonstrating that efficient carbon capture may occur in a simple
645 case when rubisco and carbonic anhydrase are diffuse within a series of concentric thylakoid
646 spheres. It may still be of interest to explore what chloroplast ultrastructures structures support
647 photosynthesis in *C. merolae,* and to investigate the biochemical and molecular basis for this
648 non-canonical CCM.
649
650 **<u>Further applications of surrogate modeling and uncertainty quantification</u>**
651
652 More broadly, the statistical approach adopted in this paper represents an advance in metabolic
653 and biochemical modeling. By training a surrogate model on the parameter space of mechanistic
654 biological models, we can understand and account for high-dimensional uncertainty in model
655 parameters. Metabolic modeling in general, especially complex metabolic modelling, has been
656 highlighted as a particularly promising application of surrogate modeling, as metabolic modeling
657 has biotechnological potential but is challenged by the complexity of metabolism and by the
658 "trial and error" process which is often required to produce a working metabolic model
659 (Gherman *et al.*, 2023). Surrogate modeling has found uses in dynamic flux balance analysis and
660 process modeling for bioprocesses (Mountraki, Benjelloun-Mlayah and Kokossis, 2020; de
661 Oliveira *et al.*, 2021). Our work expands on these investigations by demonstrating what is to our
662 knowledge the first application of surrogate modeling to ODE-based compartmental modeling of
663 biological systems. Our methods may be particularly valuable for models that have poorly-
664 defined parameters or are extremely computationally expensive. For example, the
665 implementation of surrogate modeling described here could alleviate current limitations in
666 interpreting reaction-diffusion models and genome-scale metabolic models (Gherman *et al.*,
667 2023). Even for our relatively-simple model, the run time for 240,000 simulations was several
668 hours and required use of a computing cluster. In contrast, surrogate modelling could be run
669 locally on a laptop computer, and was able to generate 240,000 predictions for all four outputs of
670 interest in less than 10 seconds, easily creating a large dataset for analysis and allowing for
671 precise sensitivity estimation. We compared this with a Sobol′ sensitivity analysis (Sobol′, 2001)
672 performed with the original model with a sample size of n = 163,840, comparable to the number
673 of parameter sets and outputs used to train the surrogate model. Despite the generation of these
674 samples taking several hours of computation time, this approach yielded extremely imprecise
675 and uninterpretable results, suggesting that substantially more computational investment would
676 be necessary to achieve acceptably precise sensitivity estimates **(Figure S10)**. With NRMSE
677 below 1.5% in our validation (**Table S2**), the computational gains associated with the surrogate
678 modeling approach outweighed the near-negligible potential error introduced by an inexact
679 surrogate.
680 Important considerations in any surrogate modelling application include the sample size
681 required to train the model, and limitations of surrogate models for out-of-sample predictions.
682 Surrogates should be used cautiously for out-of-sample predictions, particularly in high-
683 dimensional settings where training data is limited (Forrester, Sóbester and Keane, 2008).

Regarding the sample size, early studies (Chapman *et al.*, 1994; Jones, Schonlau and Welch, 1998; Loeppky, Sacks and Welch, 2009) suggested using around $10d$ samples, where d is the input dimension, for building an accurate Gaussian Process (GP) surrogate model. GP surrogates are particularly effective for small datasets and provide uncertainty quantification, which is valuable for assessing the confidence of out-of-sample predictions (Gramacy, 2020). If the desired accuracy is not achieved, one can improve the model by increasing the sample size through adaptive strategies such as active learning (MacKay, 1992), which allows for more efficient use of additional data to further enhance accuracy. Recent studies have also provided guidance on determining the run size required for a GP surrogate to achieve a pre-specified level of out-of-sample prediction accuracy (Harari *et al.*, 2017). In scenarios where high extrapolation performance is critical, one may consider using physics-informed surrogates, which tend to be more reliable in out-of-sample contexts. These surrogate models incorporate physical laws into their training process and offer improved performance for out-of-sample predictions, especially when physical dynamics play a significant role. Examples of physics-informed surrogates include a manifold-constrained GP surrogate that adheres to an underlying ODE system (Yang, Wong and Kou, 2021), or Physics-Informed Neural Networks (PINNs) (Raissi, Perdikaris and Karniadakis, 2019).

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 a particular aquatic organism achieve the high photosynthetic efficiency that enables them collectively to be responsible for approximately half of global photosynthetic CO_2 consumption (Field *et al.*, 1998). 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 *in silico* determine which features of a system are essential or inflexible throughout ranges of interest before devoting resources to *in vivo* experimentation.

Conclusions

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. 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 non-canonical 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.

Acknowledgments

Funding information

Data availability

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

Author contributions (CRediT Taxonomy definitions)

Conceptualization: AKS, JAMK, CLS, YSH, BJW. **Data Curation:** AKS, JAMK. **Formal Analysis:** AKS, JAMK, JH, DJO. **Funding Acquisition:** AKS, JAMK, CLS, YSH, BJW. **Investigation:** AKS, JAMK, DJO. **Methodology:** all authors. **Project Administration:** AKS, JAMK, CLS, YSH, BJW. **Resources:** all authors. **Software:** AKS, JAMK, JH. **Supervision:** AKS, JAMK, CLS, YSH, BJW. **Validation:** AKS, JAMK. **Visualization:** AKS, JH. **Writing – Original Draft Preparation:** AKS. **Writing – Review & Editing:** all authors.

References

Badger, M.R. *et al.* (1998) 'The diversity and coevolution of Rubisco, plastids, pyrenoids and chloroplast-based CO_2-concentrating mechanisms in the algae', *Can. J. Bot*, 76(6), pp. 1052–1071. Available at: https://doi.org/10.1139/b98-074.

Barrett, J., Girr, P. and Mackinder, L.C.M. (2021) 'Pyrenoids: CO_2-fixing phase separated liquid organelles', *Biochim. Biophys. Acta - Mol. Cell Res.*, 1868(5). Available at: https://doi.org/10.1016/j.bbamcr.2021.118949.

Beardall, J. and Raven, J.A. (2020) 'Structural and Biochemical Features of Carbon Acquisition in Algae', in A.W.D. Larkum, A.R. Grossman, and J.A. Raven (eds) *Photosynthesis in Algae: Biochemical and Physiological Mechanisms*. Cham: Springer International Publishing, pp. 141–160. Available at: https://doi.org/10.1007/978-3-030-33397-3_7.

Bellasio, C. *et al.* (2014) 'A high throughput gas exchange screen for determining rates of photorespiration or regulation of C_4 activity', *Journal of Experimental Botany*, 65(13), pp. 3769–3779. Available at: https://doi.org/10.1093/jxb/eru238.

775   von Caemmerer, S. (2000) *Biochemical models of leaf photosynthesis*. Collingwood, Australia:
776   CSIRO Publishing.

777   Carter, J. *et al.* (2013) 'Quantitative Application for SDS − PAGE in a Biochemistry Lab', *J.*
778   *Chem. Educ.*, 90(9), pp. 1255–1256. Available at: https://doi.org/10.1021/ed300390j.

779   Chapman, W.L. *et al.* (1994) 'Arctic sea ice variability: Model sensitivities and a multidecadal
780   simulation', *Journal of Geophysical Research: Oceans*, 99(C1), pp. 919–935. Available at:
781   https://doi.org/10.1029/93JC02564.

782   Chen, T. and Guestrin, C. (2016) 'XGBoost: A Scalable Tree Boosting System', *Proceedings of*
783   *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
784   San Francisco, California, USA: Association for Computing Machinery. Available at:
785   https://doi.org/10.1145/2939672.2939785.

786   Curien, G. *et al.* (2021) 'Mixotrophic growth of the extremophile *Galdieria sulphuraria* reveals
787   the flexibility of its carbon assimilation metabolism', *New Phytol.* Available at:
788   https://doi.org/10.1111/nph.17359.

789   Davey, P. and Lawson, T. (2024) 'Measurements of Carbon Assimilation in Aquatic Systems.',
790   *Methods in molecular biology (Clifton, N.J.)*, 2790, pp. 95–120. Available at:
791   https://doi.org/10.1007/978-1-0716-3790-6_6.

792   Duanmu, D. *et al.* (2009) 'Knockdown of limiting-CO2-induced gene HLA3 decreases HCO3-
793   transport and photosynthetic Ci affinity in Chlamydomonas reinhardtii.', *Proceedings of the*
794   *National Academy of Sciences of the United States of America*, 106(14), pp. 5990–5995.
795   Available at: https://doi.org/10.1073/pnas.0812885106.

796   Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980) 'A biochemical model of
797   photosynthetic CO_2 assimilation in leaves of C_3 species', *Planta*, 149(1), pp. 78–90. Available at:
798   https://doi.org/10.1007/BF00386231.

799   Fei, C. *et al.* (2022) 'Modelling the pyrenoid-based CO_2-concentrating mechanism provides
800   insights into its operating principles and a roadmap for its engineering into crops', *Nat. Plants*.
801   Available at: https://doi.org/10.1038/s41477-022-01153-7.

802   Field, C.B. *et al.* (1998) 'Primary Production of the Biosphere: Integrating Terrestrial and
803   Oceanic Components', *Science*, 281(5374), pp. 237–240. Available at:
804   https://doi.org/10.1126/science.281.5374.237.

805   Forrester, A.I.J., Sóbester, A. and Keane, A.J. (2008) *Engineering Design via Surrogate*
806   *Modelling: A Practical Guide*. John Wiley & Sons, Ltd.

807   Fridlyand, L., Kaplan, A. and Reinhold, L. (1996) 'Quantitative evaluation of the role of a
808   putative CO2-scavenging entity in the cyanobacterial CO2-concentrating mechanism.', *Bio*
809   *Systems*, 37(3), pp. 229–238. Available at: https://doi.org/10.1016/0303-2647(95)01561-2.

810 Fridlyand, L.E. (1997) 'Models of CO2 concentrating mechanisms in microalgae taking into
811 account cell and chloroplast structure.', *Bio Systems*, 44(1), pp. 41–57. Available at:
812 https://doi.org/10.1016/s0303-2647(97)00042-7.

813 Friedman, J.H. (2001) 'Greedy Function Approximation: A Gradient Boosting Machine', *The*
814 *Annals of Statistics*, 29(5), pp. 1189–1232.

815 Fujiwara, T. and Ohnuma, M. (2017) 'Procedures for Transformation and Their Applications in
816 *Cyanidioschyzon merolae*', *Cyanidioschyzon merolae: A New Model Eukaryote for Cell and*
817 *Organelle Biology*. Edited by T. Kuroiwa et al. Singapore, Singapore: Springer Nature.

818 Gehl, K.A. and Colman, B. (1985) 'Effect of External pH on the Internal pH of *Chlorella*
819 *saccharophila*', *Plant Physiol.* Available at: https://doi.org/10.1104/pp.77.4.917.

820 Gherman, I.M. *et al.* (2023) 'Bridging the gap between mechanistic biological models and
821 machine learning surrogates.', *PLoS computational biology*, 19(4), p. e1010988. Available at:
822 https://doi.org/10.1371/journal.pcbi.1010988.

823 Gramacy, R.B. (2020) *Surrogates: Gaussian Process Modeling, Design, and Optimization for*
824 *the Applied Sciences*. Taylor & Francis Group (Texts in Statistical Science).

825 Gramacy, R.B. and Apley, D.W. (2015) 'Local Gaussian Process Approximation for Large
826 Computer Experiments', *Journal of Computational and Graphical Statistics*, 24(2), pp. 561–578.

827 Gross, W. (2000) 'Ecophysiology of algae living in highly acidic environments', *Hydrobiologia*.
828 Available at: https://doi.org/10.1023/A:1004054317446.

829 Guterman, H. and Ben-Yaakov, S. (1987) 'Exchange rates of O2 and CO2 between an algal
830 culture and atmosphere', *Water Research*, 21(1), pp. 25–34. Available at:
831 https://doi.org/10.1016/0043-1354(87)90095-9.

832 Gutknecht, J., Bisson, M.A. and Tosteson, F.C. (1977) 'Diffusion of carbon dioxide through lipid
833 bilayer membranes: effects of carbonic anhydrase, bicarbonate, and unstirred layers.', *The*
834 *Journal of general physiology*, 69(6), pp. 779–794. Available at:
835 https://doi.org/10.1085/jgp.69.6.779.

836 Harari, O. *et al.* (2017) 'Computer Experiments: Prediction Accuracy, Sample Size and Model
837 Complexity Revisited', *Statistica Sinica* [Preprint]. Available at:
838 https://doi.org/10.5705/ss.202016.0217.

839 Ichinose, T.M. and Iwane, A.H. (2017) 'Cyotological Analyses by Advanced Electron
840 Microscopy', *Cyanidioschyzon merolae: A New Model Eukaryote for Cell and Organelle*
841 *Biology*. Edited by T. Kuroiwa et al. Singapore, Singapore: Springer Nature.

842 Itoh, R. *et al.* (1999) 'Two ftsH-family genes encoded in the nuclear and chloroplast genomes of
843 the primitive red alga *Cyanidioschyzon merolae*', *Plant molecular biology*, 41(3), pp. 321–337.
844 Available at: https://doi.org/10.1023/a:1006369104530.

845   James, G. *et al.* (2013) *An Introduction to Statistical Learning With Applications in R*. Springer
846   Science + Business Media (Springer Texts in Statistics).

847   Jones, D.R., Schonlau, M. and Welch, W.J. (1998) 'Efficient Global Optimization of Expensive
848   Black-Box Functions', *Journal of Global Optimization*, 13(4), pp. 455–492. Available at:
849   https://doi.org/10.1023/A:1008306431147.

850   Jordan, D.B. and Ogren, W.L. (1984) 'The CO_2/O_2 specificity of ribulose 1,5-bisphosphate
851   carboxylase/oxygenase : Dependence on ribulose bisphosphate concentration, pH and
852   temperature', *Planta*. Available at: https://doi.org/10.1007/BF00398720.

853   Kaste, J.A.M., Walker, B.J. and Shachar-Hill, Y. (2024) 'Reaction-diffusion modeling provides
854   insights into biophysical carbon concentrating mechanisms in land plants', *Plant Physiology*, p.
855   kiae324. Available at: https://doi.org/10.1093/plphys/kiae324.

856   Kubien, D.S., Brown, C.M. and Kane, H.J. (2010) 'Quantifying the Amount and Activity of
857   Rubisco in Leaves', *Photosynthesis Research Protocols*. Edited by R. Carpentier. Totowa, NJ:
858   Humana Press (Methods in Molecular Biology).

859   Kuroiwa, T. (1998) 'The primitive red algae *Cyanidium caldarium* and *Cyanidioschyzon*
860   *merolae* as model system for investigating the dividing apparatus of mitochondria and plastids',
861   *BioEssays*. Available at: https://doi.org/10.1002/(SICI)1521-1878(199804)20:4<344::AID-
862   BIES11>3.0.CO;2-2.

863   Lavigne, H., Proye, A. and Gattuso, J.-P. (2019) 'seacarb: Calculates parameters of the seawater
864   carbonate system'. Available at: https://rdrr.io/rforge/seacarb/.

865   Loeppky, J.L., Sacks, J. and Welch, W.J. (2009) 'Choosing the Sample Size of a Computer
866   Experiment: A Practical Guide', *Technometrics*, 51(4), pp. 366–376. Available at:
867   https://doi.org/10.1198/TECH.2009.08040.

868   Loganathan, N., Tsai, Y.C. and Mueller-Cajar, O. (2016) 'Characterization of the
869   heterooligomeric red-type rubisco activase from red algae', *Proc. Natl. Acad. Sci. U S A*.
870   20161121st edn. Available at: https://doi.org/10.1073/pnas.1610758113.

871   Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', in
872   *Proceedings of the 31st International Conference on Neural Information Processing Systems*.
873   Long Beach, California, USA: Curran Associates Inc., pp. 4768–4777.

874   MacKay, D.J.C. (1992) 'Information-Based Objective Functions for Active Data Selection',
875   *Neural Computation*, 4, pp. 590–604.

876   Mangan, N.M. *et al.* (2016) 'pH determines the energetic efficiency of the cyanobacterial CO_2
877   concentrating mechanism', *Proc. Natl. Acad. Sci. U S A*. 20160822nd edn. Available at:
878   https://doi.org/10.1073/pnas.1525145113.

879   Mangan, N.M. and Brenner, M.P. (2014) 'Systems analysis of the CO_2 concentrating mechanism
880   in cyanobacteria', *eLife*. Available at: https://doi.org/10.7554/eLife.02043.

881 McGrath, J.M. and Long, S.P. (2014) 'Can the cyanobacterial carbon-concentrating mechanism
882 increase photosynthesis in crop species? A theoretical analysis', *Plant Physiol.* 20140218th edn.
883 Available at: https://doi.org/10.1104/pp.113.232611.

884 McKay, M.D., Beckman, R.J. and Conover, W.J. (1979) 'A Comparison of Three Methods for
885 Selecting Values of Input Variables in the Analysis of Output from a Computer Code',
886 *Technometrics*, 21(2), pp. 239–245. Available at: https://doi.org/10.2307/1268522.

887 Missner, A. *et al.* (2008) 'Carbon dioxide transport through membranes', *The Journal of*
888 *biological chemistry*, 283(37), pp. 25340–25347. Available at:
889 https://doi.org/10.1074/jbc.M800096200.

890 Misumi, O. *et al.* (2005) '*Cyanidioschyzon merolae* Genome. A Tool for Facilitating
891 Comparable Studies on Organelle Biogenesis in Photosynthetic Eukaryotes', *Plant Physiol.*
892 Available at: https://doi.org/10.1104/pp.104.053991.

893 Misumi, O., Kuroiwa, T. and Hirooka, S. (2017) 'Application of the Tolerance to Extreme
894 Environment to Land Plants', in T. Kuroiwa et al. (eds) *Cyanidioschyzon merolae: A New Model*
895 *Eukaryote for Cell and Organelle Biology*. Singapore: Springer Singapore, pp. 325–341.
896 Available at: https://doi.org/10.1007/978-981-10-6101-1_20.

897 Miyagishima, M. *et al.* (1998) 'Visualization of the microbody division in *Cyanidioschyzon*
898 *merolae* with the fluorochrome brilliant sulfoflavin', *Protoplasma*, 201(1), pp. 115–119.
899 Available at: https://doi.org/10.1007/BF01280718.

900 Miyagishima, S. *et al.* (2017) 'Cyanidiales: Evolution and Habitats', *Cyanidioschyzon merolae:*
901 *A New Model Eukaryote for Cell and Organelle Biology*. Edited by T. Kuroiwa et al. Singapore,
902 Singapore: Springer Nature.

903 Miyagishima, Shin-ya and Wei, J.L. (2017) 'Procedures for Cultivation, Observation, and
904 Conventional Experiments in *Cyanidioschyzon merolae*', *Cyanidioschyzon merolae: A New*
905 *Model Eukaryote for Cell and Organelle Biology*. Edited by T. Kuroiwa et al. Singapore,
906 Singapore: Springer Nature.

907 Miyagishima, S.Y. and Tanaka, K. (2021) 'The Unicellular Red Alga *Cyanidioschyzon merolae* -
908 The Simplest Model of a Photosynthetic Eukaryote', *Plant Cell Physiol.* Available at:
909 https://doi.org/10.1093/pcp/pcab052.

910 Morita, E. *et al.* (1999) 'Role of pyrenoids in the CO2-concentrating mechanism: comparative
911 morphology, physiology and molecular phylogenetic analysis of closely related strains of
912 Chlamydomonas and Chloromonas (Volvocales)', *Planta*, 208(3), pp. 365–372. Available at:
913 https://doi.org/10.1007/s004250050571.

914 Moriyama, T. *et al.* (2018) 'Selective loss of photosystem I and formation of tubular thylakoids
915 in heterotrophically grown red alga *Cyanidioschyzon merolae*', *Photosynth. Res.* Available at:
916 https://doi.org/10.1007/s11120-018-0603-z.

917 Moroney, J.V. *et al.* (2011) 'The carbonic anhydrase isoforms of *Chlamydomonas reinhardtii*:
918 intracellular location, expression, and physiological roles', *Photosynth. Res.* Available at:
919 https://doi.org/10.1007/s11120-011-9635-3.

920 Mountraki, A.D., Benjelloun-Mlayah, B. and Kokossis, A.C. (2020) 'A Surrogate Modeling
921 Approach for the Development of Biorefineries', *Frontiers in Chemical Engineering*, 2.
922 Available at: https://doi.org/10.3389/fceng.2020.568196.

923 Nevo, R. *et al.* (2007) 'Thylakoid membrane perforations and connectivity enable intracellular
924 traffic in cyanobacteria', *The EMBO Journal*, 26(5), pp. 1467–1473. Available at:
925 https://doi.org/10.1038/sj.emboj.7601594.

926 Oesterhelt, C. *et al.* (2007) 'Regulation of photosynthesis in the unicellular acidophilic red alga
927 *Galdieria sulphuraria*', *Plant J.* Available at: https://doi.org/10.1111/j.1365-313X.2007.03159.x.

928 de Oliveira, R.D. *et al.* (2021) 'Nonlinear Predictive Control of a Bioreactor by Surrogate Model
929 Approximation of Flux Balance Analysis', *Industrial & Engineering Chemistry Research*,
930 60(40), pp. 14464–14475. Available at: https://doi.org/10.1021/acs.iecr.1c01242.

931 Orr, D.J. and Carmo-Silva, E. (2018) 'Extraction of RuBisCO to Determine Catalytic Constants',
932 *Methods Mol. Biol.* Available at: https://doi.org/10.1007/978-1-4939-7786-4_13.

933 Parys, E. *et al.* (2021) 'Photosynthesis of the *Cyanidioschyzon merolae* cells in blue, red, and
934 white light', *Photosynth. Res.* Available at: https://doi.org/10.1007/s11120-020-00796-x.

935 Price, G.D. *et al.* (2004) 'Identification of a SulP-type bicarbonate transporter in marine
936 cyanobacteria', *Proc. Natl. Acad. Sci. U S A.* Available at:
937 https://doi.org/10.1073/pnas.0405211101.

938 Price, G.D. *et al.* (2008) 'Advances in understanding the cyanobacterial CO_2-concentrating-
939 mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and
940 prospects for engineering into plants', *J. Exp. Bot.* Available at:
941 https://doi.org/10.1093/jxb/erm112.

942 Price, G.D. (2011) 'Inorganic carbon transporters of the cyanobacterial CO2 concentrating
943 mechanism.', *Photosynthesis research*, 109(1–3), pp. 47–57. Available at:
944 https://doi.org/10.1007/s11120-010-9608-y.

945 Price, G.D. *et al.* (2013) 'The cyanobacterial CCM as a source of genes for improving
946 photosynthetic CO2 fixation in crop species.', *Journal of experimental botany*, 64(3), pp. 753–
947 768. Available at: https://doi.org/10.1093/jxb/ers257.

948 Price, G.D. and Badger, M.R. (1989) 'Expression of Human Carbonic Anhydrase in the
949 Cyanobacterium *Synechococcus* PCC7942 Creates a High CO_2-Requiring Phenotype: Evidence
950 for a Central Role for Carboxysomes in the CO_2 Concentrating Mechanism', *Plant Physiol.*
951 Available at: https://doi.org/10.1104/pp.91.2.505.

952   Price, G.D. and Howitt, S.M. (2011) 'The cyanobacterial bicarbonate transporter BicA: its
953   physiological role and the implications of structural similarities with human SLC26 transporters',
954   *Biochemistry and Cell Biology*, 89(2), pp. 178–188. Available at: https://doi.org/10.1139/O10-
955   136.

956   Prins, A. *et al.* (2016) 'Rubisco catalytic properties of wild and domesticated relatives provide
957   scope for improving wheat photosynthesis', *Journal of Experimental Botany*, 67(6), pp. 1827–
958   1838. Available at: https://doi.org/10.1093/jxb/erv574.

959   Rademacher, N. *et al.* (2016) 'Photorespiratory glycolate oxidase is essential for the survival of
960   the red alga *Cyanidioschyzon merolae* under ambient CO_2 conditions', *J. Exp. Bot.* Available at:
961   https://doi.org/https:/doi.org/10.1093/jxb/erw118.

962   Rademacher, N. *et al.* (2017) 'Transcriptional response of the extremophile red alga
963   *Cyanidioschyzon merolae* to changes in CO_2 concentrations', *J. Plant Physiol.* Available at:
964   https://doi.org/10.1016/j.jplph.2017.06.014.

965   Rahman, D.Y. *et al.* (2017) 'Thermostable phycocyanin from the red microalga *Cyanidioschyzon*
966   *merolae*, a new natural blue food colorant', *J. Appl. Phycol.* 20161121st edn. Available at:
967   https://doi.org/10.1007/s10811-016-1007-0.

968   Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) 'Physics-informed neural networks: A
969   deep learning framework for solving forward and inverse problems involving nonlinear partial
970   differential equations', *Journal of Computational Physics*, 378, pp. 686–707. Available at:
971   https://doi.org/10.1016/j.jcp.2018.10.045.

972   Read, B.A. and Tabita, F.R. (1994) 'High substrate specificity factor ribulose bisphosphate
973   carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant
974   cyanobacterial RubiSCO containing "algal" residue modifications.', *Archives of biochemistry*
975   *and biophysics*, 312(1), pp. 210–218. Available at: https://doi.org/10.1006/abbi.1994.1301.

976   Reimer, K.A. *et al.* (2017) 'The sole LSm complex in Cyanidioschyzon merolae associates with
977   pre-mRNA splicing and mRNA degradation factors', *RNA*. 20170321st edn. Available at:
978   https://doi.org/10.1261/rna.058487.116.

979   Robison, T.A. *et al.* (2024) 'Hornworts reveal a spatial model for pyrenoid-based CO_2-
980   concentrating mechanisms in land plants', *bioRxiv*, p. 2024.06.26.600872. Available at:
981   https://doi.org/10.1101/2024.06.26.600872.

982   Sato, N. *et al.* (2017) 'Lipid metabolism and potentials of biofuel and high added-value oil
983   production in red algae', *World J. Microbiol. Biotechnol.* Available at:
984   https://doi.org/10.1007/s11274-017-2236-3.

985   Seger, M. *et al.* (2023) 'Engineered ketocarotenoid biosynthesis in the polyextremophilic red
986   microalga *Cyanidioschyzon merolae* 10D', *Metabolic Engineering Communications*, 17, p.
987   e00226. Available at: https://doi.org/10.1016/j.mec.2023.e00226.

Seymour, J.R. *et al.* (2017) 'Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships', *Nature Microbiology*, 2(7), p. 17065. Available at: https://doi.org/10.1038/nmicrobiol.2017.65.

Sharwood, R.E., Ghannoum, O. and Whitney, S.M. (2016) 'Prospects for improving CO_2 fixation in C_3-crops through understanding C_4-Rubisco biogenesis and catalytic diversity.', *Current opinion in plant biology*, 31, pp. 135–142. Available at: https://doi.org/10.1016/j.pbi.2016.04.002.

Sobol′, I.M. (2001) 'Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates', *Mathematics and Computers in Simulation*, 55(1), pp. 271–280. Available at: https://doi.org/10.1016/S0378-4754(00)00270-6.

Spalding, M.H. (2008) 'Microalgal carbon-dioxide-concentrating mechanisms: Chlamydomonas inorganic carbon transporters', *Journal of Experimental Botany*, 59(7), pp. 1463–1473. Available at: https://doi.org/10.1093/jxb/erm128.

Steensma, A.K., Shachar-Hill, Y. and Walker, B.J. (2023) 'The carbon-concentrating mechanism of the extremophilic red microalga *Cyanidioschyzon merolae*', *Photosynth. Res.* 20230213th edn. Available at: https://doi.org/10.1007/s11120-023-01000-6.

Thoms, S., Pahlow, M. and Wolf-Gladrow, D.A. (2001) 'Model of the carbon concentrating mechanism in chloroplasts of eukaryotic algae', *J. Theor. Biol.* Available at: https://doi.org/10.1006/jtbi.2000.2219.

Toda, K. *et al.* (1998) 'Characterization of a chloroplast isoform of serine acetyltransferase from the thermo-acidiphilic red alga *Cyanidioschyzon merolae*', *Biochim. Biophys. Acta - Mol. Cell Res.* Available at: https://doi.org/10.1016/S0167-4889(98)00031-7.

Uemura, K. *et al.* (1997) 'Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase from Thermophilic Red Algae with a Strong Specificity for CO_2 Fixation', *Biochem. Biophys. Res. Commun.* Available at: https://doi.org/10.1006/bbrc.1997.6497.

Villegas-Valencia, M. *et al.* (2023) '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.* 20230420th edn. Available at: https://doi.org/10.3389/fmicb.2023.1157151.

Virtanen, P. *et al.* (2020) 'SciPy 1.0: fundamental algorithms for scientific computing in Python', *Nature Methods*, 17(3), pp. 261–272. Available at: https://doi.org/10.1038/s41592-019-0686-2.

Walker, B.J. *et al.* (2020) 'Flexibility in the Energy Balancing Network of Photosynthesis Enables Safe Operation under Changing Environmental Conditions', *Plants*, 9(3). Available at: https://doi.org/10.3390/plants9030301.

Whitney, S.M. *et al.* (2001) 'Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts', *Plant J.* Available at: https://doi.org/10.1046/j.1365-313x.2001.01056.x.

1025   Xu, Y. *et al.* (2021) 'The metabolic origins of non-photorespiratory CO2 release during
1026   photosynthesis: a metabolic flux analysis', *Plant Physiology*, 186(1), pp. 297–314. Available at:
1027   https://doi.org/10.1093/plphys/kiab076.

1028   Yagisawa, F. *et al.* (2012) 'Mitotic inheritance of endoplasmic reticulum in the primitive red alga
1029   *Cyanidioschyzon merolae*', *Protoplasma*, 249(4), pp. 1129–1135. Available at:
1030   https://doi.org/10.1007/s00709-011-0359-1.

1031   Yagisawa, F. *et al.* (2016) 'Intracellular Structure of the Unicellular Red Alga *Cyanidioschyzon
1032   merolae* in Response to Phosphate Depletion and Resupplementation', *Cytologia*, 81(3), pp.
1033   341–347. Available at: https://doi.org/10.1508/cytologia.81.341.

1034   Yang, S., Wong, S.W.K. and Kou, S.C. (2021) 'Inference of dynamic systems from noisy and
1035   sparse data via manifold-constrained Gaussian processes.', *Proceedings of the National Academy
1036   of Sciences of the United States of America*, 118(15). Available at:
1037   https://doi.org/10.1073/pnas.2020397118.

1038   Zenvirth, D., Volokita, M. and Kaplan, A. (1985) 'Photosynthesis and Inorganic Carbon
1039   Accumulation in the Acidophilic Alga *Cyanidioschyzon merolae*', *Plant Physiol.* Available at:
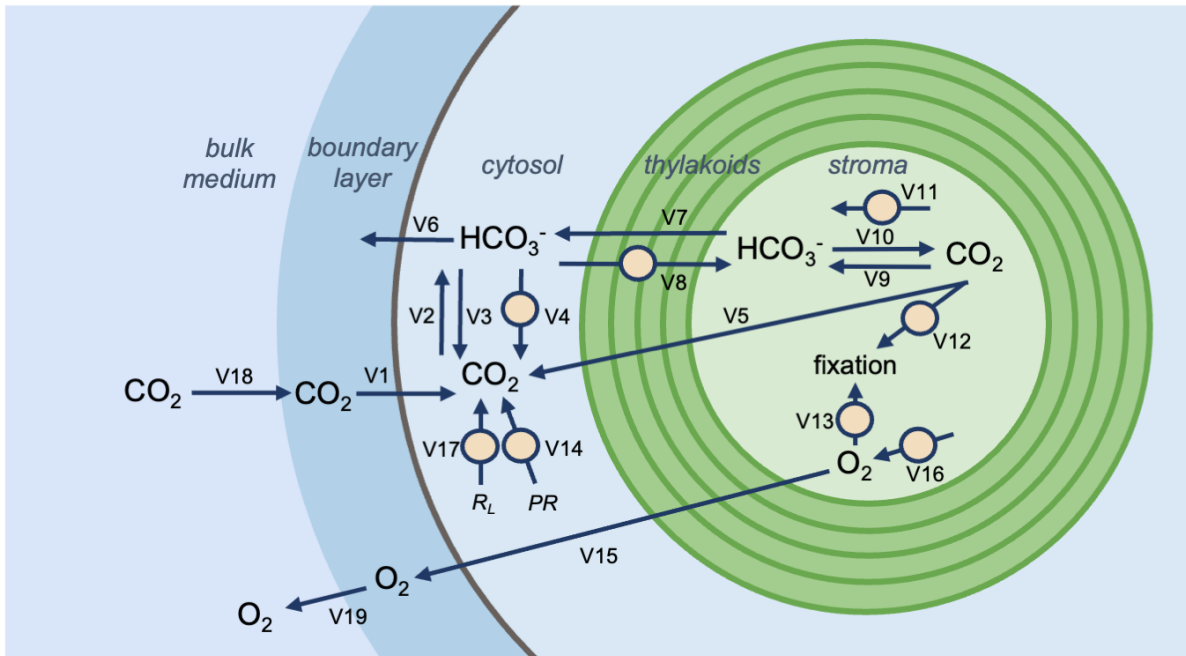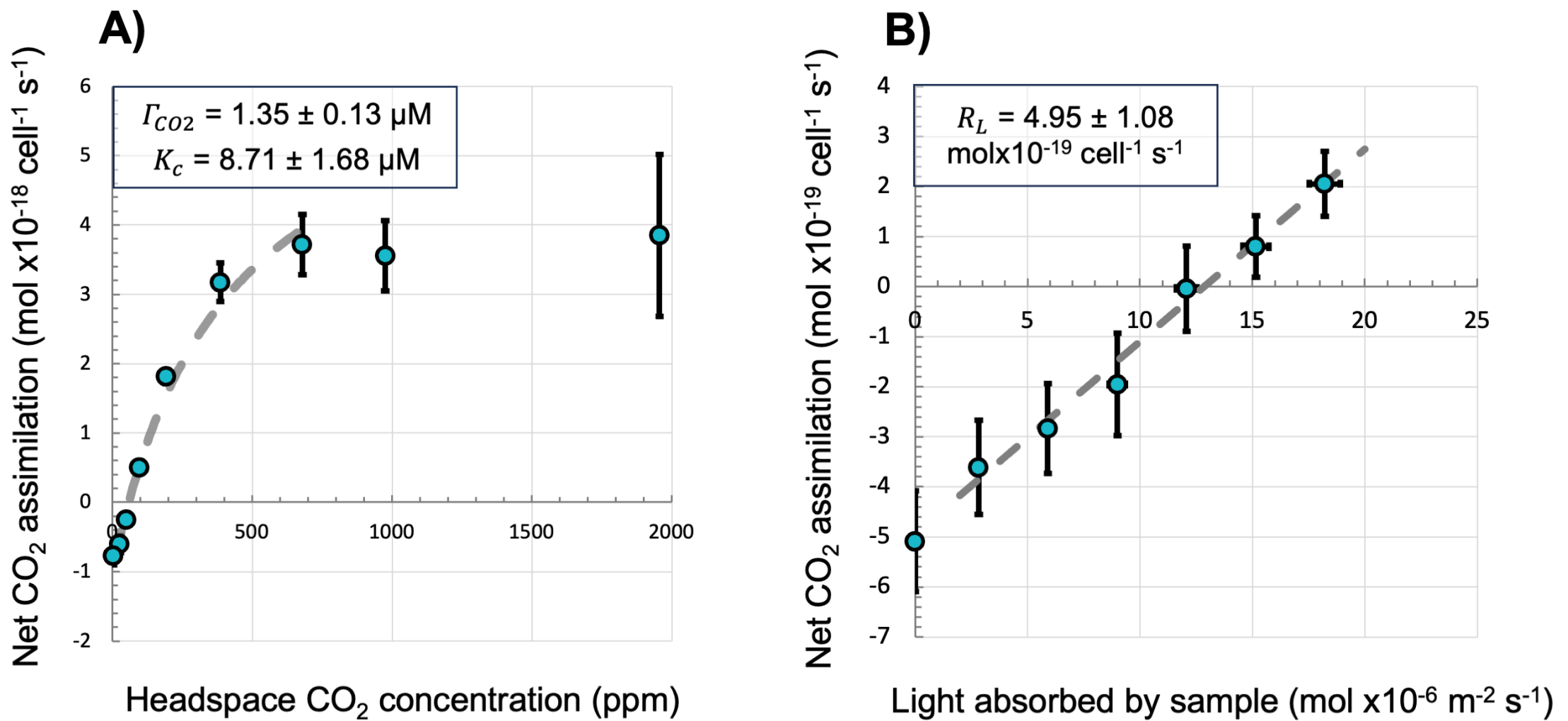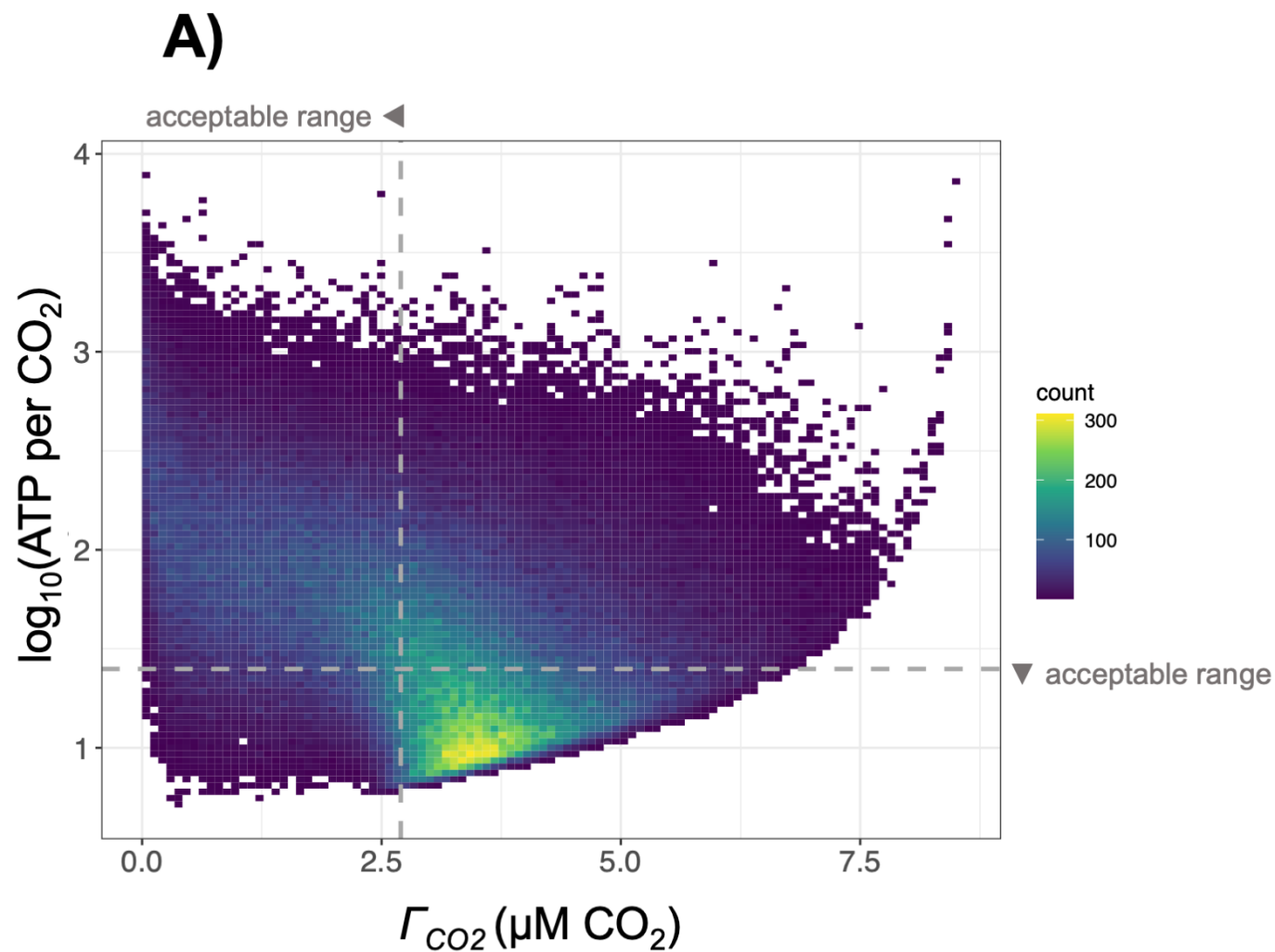1040   https://doi.org/10.1104/pp.77.1.237.

1041
1042

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.*

A)

Γ_{CO2} = 1.35 ± 0.13 µM
K_c = 8.71 ± 1.68 µM

Net CO$_2$ assimilation (mol x10^{-18} cell^{-1} s^{-1})

Headspace CO$_2$ concentration (ppm)

B)

R_L = 4.95 ± 1.08 molx10^{-19} cell^{-1} s^{-1}

Net CO$_2$ assimilation (mol x10^{-19} cell^{-1} s^{-1})

Light absorbed by sample (mol x10^{-6} m^{-2} s^{-1})

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 Michaelis-Menten constant of CO$_2$ fixation (K$_C$) and respiration in the light (R$_L$). The linear fit used to determine CO$_2$ compensation point (Γ_{CO2}) is not pictured but is described in* **Methods**. **(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.*

Figure 3. Values of key model outputs. (A) *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.*
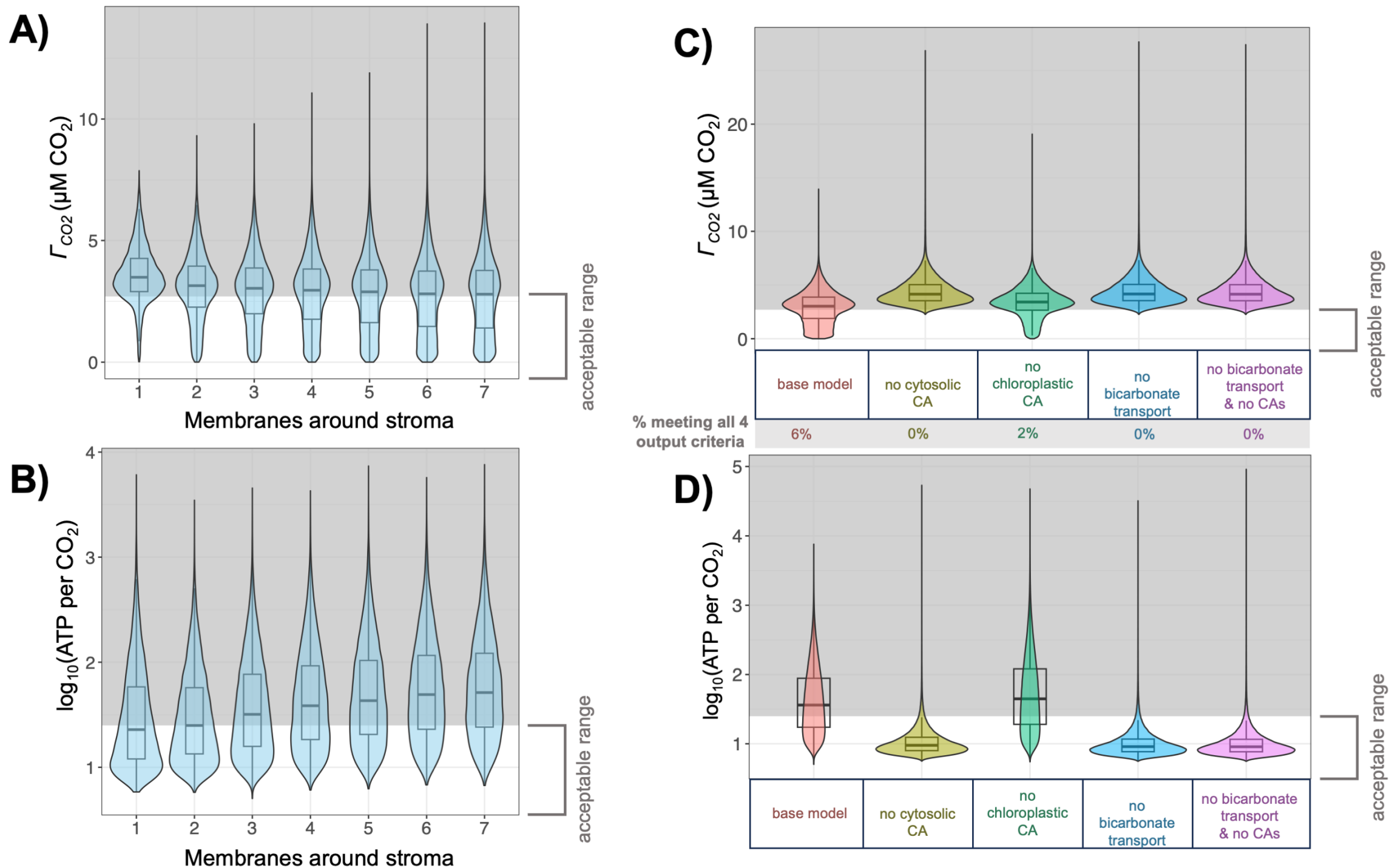
Figure 4. Effect of select input parameters on key model outputs. *(A, B) Effect of model input parameter* Membranes *(x-axis) on key model outputs. Distribution of parameter set outputs for each value of* Membranes *is represented by a box plot overlaid on a violin plot. Shaded areas represent unacceptable 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.*
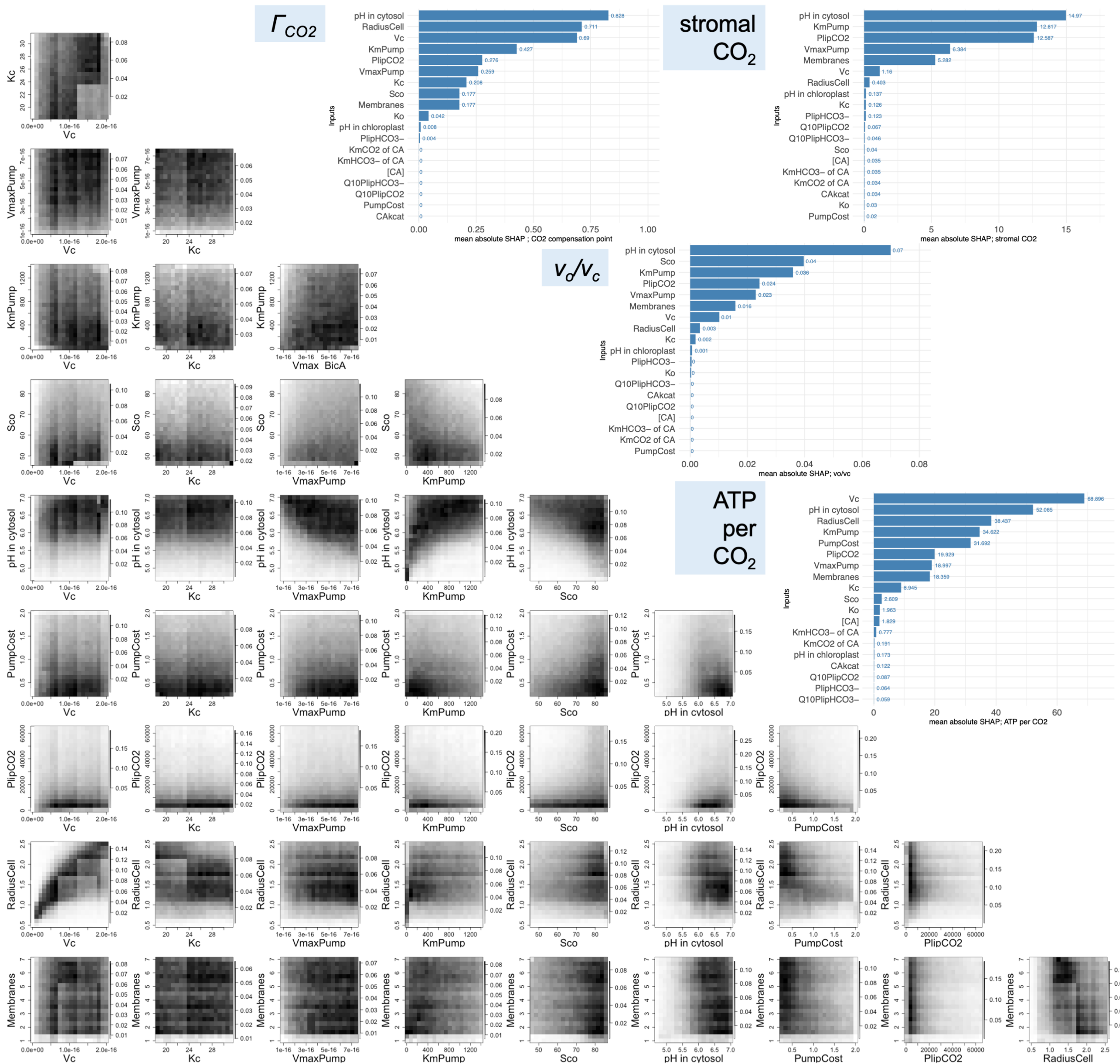
Figure 5. Statistical investigation of parameters affecting model output. *(upper right bar plots) Mean absolute SHapley Additive exPlanations (SHAP) plots for each output criterion. (lower left density plots) Density plots of parameter sets meeting all output criteria, organized by selected pairwise 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).*

Parsed Citations

1+ □ −□ for regression tasks. The error function used for the calculation of the error was the sum of squared errors. The threshold parameter for the partial derivatives of the error function as stopping criteria for the NN model was set to half the range of the target variable. The DNN model consists of two hidden layers containing 64 and 32 units respectively, both using rectified linear unit (ReLU) activation functions max(x, 0). The DNN model was trained using the adaptive moment estimation (Adam) optimizer and mean squared error (MSE) as the loss function. The model was trained for 40 epochs, with the learning algorithm processing the entire training dataset 40 times. A batch size of 240 was used, indicating the number of samples processed before updating the model's internal parameters. Moreover, 20% of the training data was set aside for validation purposes during the training process. 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 (Steensma, Shachar-Hill and Walker, 2023). Here we confirmed that C. merolae rubisco kinetics are similar to those of other red-type (Form 1D) rubiscos (Read and Tabita, 1994; Uemura et al., 1997; Whitney et al., 2001). C. merolae rubisco had a strong affinity for CO_2 (low KC), a poor affinity for O_2 (high KO), and a slow carboxylation rate (low kcatC) (Figure 2). Consistent with other studies, kcatC and KC were higher when measured at increased temperature, while KO was lower. Although KO is in the denominator of rubisco specificity (Sc/o) and Sc/o decreases with increased temperature, in vitro KO is observed to decrease with increased assay temperature in some species (Jordan and Ogren, 1984; Uemura et al., 1997; Prins et al., 2016). 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 KC vs. 24.9 ± 3.2 μM rubisco KC at 45 □C, p = 0.008 by two-sample t-test) (Figure 2). This result adds to the evidence of a CCM in C. merolae (Zenvirth, Volokita and Kaplan, 1985; Rademacher et al., 2017; Steensma, Shachar-Hill and Walker, 2023). 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 lacking structures like a pyrenoid or carboxysome (Figure 3). Cyanobacterial CCM models have also supported reduction to a simple model with only two compartments from the cell membrane inwards (Mangan and Brenner, 2014). Our results provided quantitative support for a CCM taking inorganic carbon from the environment solely through CO_2 diffusion into the cell without specialized compartments, which we term a "non-canonical" CCM due to its differences in structure and function from CCMs that have been characterized in detail. C. merolae has a different structure and environment than the "canonical" CCMs of Chlamydomonas reinhardtii and of model cyanobacteria, which allowed us to explore a biology and a parameter space which are different from those in previous CCM models. Though there is speculation that extremophilic red algae may use a C4-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 (Gehl and Colman, 1985; Fridlyand, 1997; Gross, 2000; Rademacher et al., 2016; Curien et al., 2021; Fei et al., 2022). Briefly, this mechanism would involve bicarbonate being 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. Various facilitated CO_2 uptake mechanisms exist in CCM-containing organisms, such as the NDH-I complexes in cyanobacteria and the periplasmic CA system in algae (Fridlyand, Kaplan and Reinhold, 1996; Moroney et al., 2011; Price, 2011). We here test a different model where inorganic carbon enters the cell solely by passive CO_2 diffusion into the cytosol, followed by the action of non-vectorial cytosolic carbonic anhydrase. In contrast to the well-studied cyanobacterial and algal systems, where growth under limiting CO_2 is supported by active bicarbonate uptake and the accumulation of cytosolic bicarbonate above equilibrium levels (Price and Badger, 1989; Price et al., 2004; Duanmu et al., 2009), our model functions as a CCM without taking any bicarbonate from the environment. Another unique feature of our model is the nature of the diffusion barrier surrounding rubisco. Cyanobacteria encapsulate rubisco in a proteinaceous shell called the carboxysome, which is thought to provide a diffusion barrier to CO_2 (Price et al., 2008). The model alga C. reinhardtii aggregates rubisco into an organelle called the pyrenoid, which in wild-type cells is surrounded by a starch sheath that may serve as a diffusion barrier. In contrast to the well-studied system of C. reinhardtii, there has been comparatively less investigation into algae which lack starch sheaths or lack pyrenoids entirely (Morita et al., 1999; Barrett, Girr and Mackinder, 2021). Thus, to broaden our knowledge of CCM anatomy, we modeled an arrangement where rubisco is diffuse within a series of concentric thylakoid membranes. This allowed us to further investigate whether membranes, which are thought to be highly permeable to CO_2 (Gutknecht, Bisson and Tosteson, 1977; Missner et al., 2008), could impact carbon-concentration, and how carbon-concentration could function without a carboxysome or pyrenoid. /