A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle

Luke C. M. Mackinder[a], Moritz T. Meyer[b], Tabea Mettler-Altmann[c,1], Vivian K. Chen[a,d], Madeline C. Mitchell[b,2], Oliver Caspari[b], Elizabeth S. Freeman Rosenzweig[a,d], Leif Pallesen[a], Gregory Reeves[a,3], Alan Itakura[a,d], Robyn Roth[e], Frederik Sommer[c,4], Stefan Geimer[f], Timo Mühlhaus[c,4], Michael Schroda[c,4], Ursula Goodenough[e], Mark Stitt[c], Howard Griffiths[b], and Martin C. Jonikas[a,d,5]

[a]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305; [b]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom; [c]Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany; [d]Department of Biology, Stanford University, Stanford, CA 94305; [e]Department of Biology, Washington University in Saint Louis, St. Louis, MO 63130; and [f]Institute of Cell Biology, University of Bayreuth, 95440 Bayreuth, Germany

Edited by Paul G. Falkowski, Rutgers, The State University of New Jersey, New Brunswick, NJ, and approved April 7, 2016 (received for review November 20, 2015)

Biological carbon fixation is a key step in the global carbon cycle that regulates the atmosphere's composition while producing the food we eat and the fuels we burn. Approximately one-third of global carbon fixation occurs in an overlooked algal organelle called the pyrenoid. The pyrenoid contains the CO_2-fixing enzyme Rubisco and enhances carbon fixation by supplying Rubisco with a high concentration of CO_2. Since the discovery of the pyrenoid more that 130 y ago, the molecular structure and biogenesis of this ecologically fundamental organelle have remained enigmatic. Here we use the model green alga *Chlamydomonas reinhardtii* to discover that a low-complexity repeat protein, Essential Pyrenoid Component 1 (EPYC1), links Rubisco to form the pyrenoid. We find that EPYC1 is of comparable abundance to Rubisco and colocalizes with Rubisco throughout the pyrenoid. We show that EPYC1 is essential for normal pyrenoid size, number, morphology, Rubisco content, and efficient carbon fixation at low CO_2. We explain the central role of EPYC1 in pyrenoid biogenesis by the finding that EPYC1 binds Rubisco to form the pyrenoid matrix. We propose two models in which EPYC1's four repeats could produce the observed lattice arrangement of Rubisco in the *Chlamydomonas* pyrenoid. Our results suggest a surprisingly simple molecular mechanism for how Rubisco can be packaged to form the pyrenoid matrix, potentially explaining how Rubisco packaging into a pyrenoid could have evolved across a broad range of photosynthetic eukaryotes through convergent evolution. In addition, our findings represent a key step toward engineering a pyrenoid into crops to enhance their carbon fixation efficiency.

pyrenoid | Rubisco | carbon fixation | *Chlamydomonas reinhardtii* | CO_2-concentrating mechanism

Rubisco, the most abundant enzyme in the biosphere (1), fixes CO_2 into organic carbon that supports nearly all life on Earth (2, 3). Over the past 3 billion y, the enzyme became a victim of its own success as it drew down the atmospheric CO_2 concentration to trace levels (4) and as the oxygen-producing reactions of photosynthesis filled our atmosphere with O_2 (4). In today's atmosphere, O_2 competes with CO_2 at Rubisco's catalytic site, producing the toxic compound phosphoglycolate (5). Phosphoglycolate must be metabolized at the expense of energy and loss of fixed carbon and nitrogen (6). To overcome Rubisco's limitations, many photosynthetic organisms have evolved carbon-concentrating mechanisms (CCMs) (7, 8). CCMs increase the CO_2 concentration around Rubisco, decreasing O_2 competition and enhancing carbon fixation.

At the heart of the CCM of eukaryotic algae is an organelle known as the pyrenoid (9). The pyrenoid is a spherical structure in the chloroplast stroma, discovered more than 130 y ago (10–12). Pyrenoids have been found in nearly all of the major oceanic eukaryotic primary producers and mediate ∼28–44% of global carbon fixation (*SI Appendix*, Table S1) (3, 13–17). A pyrenoid typically consists of a matrix surrounded by a starch sheath and traversed by membrane tubules continuous with the photosynthetic thylakoid membranes (18). This matrix is thought to consist primarily of tightly packed Rubisco and its chaperone, Rubisco activase (19). In higher plants and non–pyrenoid-containing photosynthetic eukaryotes, Rubisco is instead soluble throughout the chloroplast stroma. The molecular mechanism by which Rubisco aggregates to form the pyrenoid matrix remains enigmatic.

Two mechanisms for Rubisco accumulation in the pyrenoid have been proposed: (*i*) Rubisco holoenzymes could bind each other directly through hydrophobic residues (20), or (*ii*) a linker protein may link Rubisco holoenzymes together (18, 20). The second model is based on analogy to the well-characterized prokaryotic carbon concentrating organelle, the β-carboxysome, where Rubisco aggregation is mediated by a linker protein consisting of repeats of a

Significance

Eukaryotic algae, which play a fundamental role in global CO_2 fixation, enhance the performance of the carbon-fixing enzyme Rubisco by placing it into an organelle called the pyrenoid. Despite the ubiquitous presence and biogeochemical importance of this organelle, how Rubisco assembles to form the pyrenoid remains a long-standing mystery. Our discovery of an abundant repeat protein that binds Rubisco in the pyrenoid represents a critical advance in our understanding of pyrenoid biogenesis. The repeat sequence of this protein suggests elegant models to explain the structural arrangement of Rubisco enzymes in the pyrenoid. Beyond advances in basic understanding, our findings open doors to the engineering of algal pyrenoids into crops to enhance yields.

Author contributions: L.C.M.M., M.T.M., T.M.-A., M. Schroda, M. Stitt, H.G., and M.C.J. designed research; L.C.M.M., M.T.M., T.M.-A., V.K.C., M.C.M., O.C., E.S.F.R., L.P., G.R., A.I., R.R., F.S., S.G., and T.M. performed research; L.C.M.M., M.T.M., T.M.-A., L.P., M. Schroda, and U.G. contributed new reagents/analytic tools; L.C.M.M., M.T.M., T.M.-A., M.C.M., F.S., S.G., T.M., and M.C.J. analyzed data; and L.C.M.M., M.T.M., T.M.-A., U.G., M. Stitt, H.G., and M.C.J. wrote the paper.

Conflict of interest statement: The authors wish to note that the Carnegie Institution for Science has submitted a provisional patent application on aspects of the findings.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]Present address: Cluster of Excellence in Plant Sciences and Institute of Plant Biochemistry, Heinrich-Heine University, 40225 Düsseldorf, Germany.

[2]Present address: Agriculture, Commonwealth Scientific and Industrial Research Organization, Canberra, ACT 2601, Australia.

[3]Present address: Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom.

[4]Present address: Institute of Molecular Biotechnology and Systems Biology, Technical University of Kaiserslautern, 67663 Kaiserslautern, Germany.

[5]To whom correspondence should be addressed. Email: mjonikas@carnegiescience.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522866113/-/DCSupplemental.

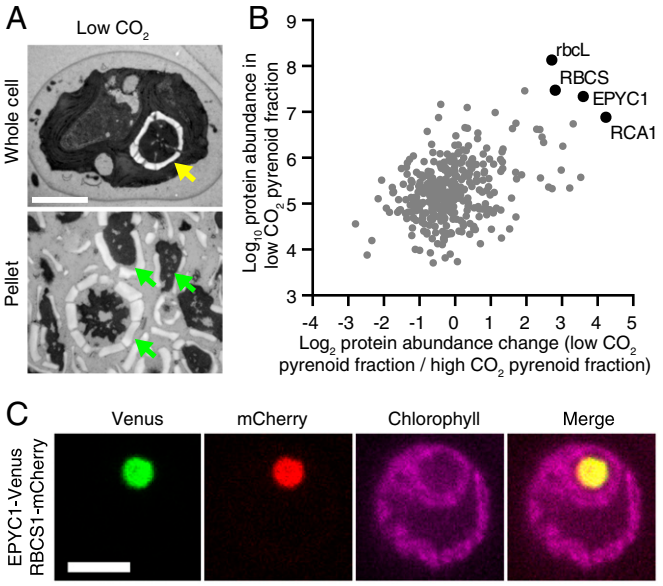


Fig. 1. EPYC1 is an abundant pyrenoid protein. (*A*) TEM images of *Chlamydomonas* whole cells and pyrenoid-enriched pellet fraction from cells grown at low CO_2. The yellow arrow indicates the pyrenoid, and green arrows indicate pyrenoid-like structures. (Scale bar: 2 μm.) (*B*) Mass spectrometry analysis of 366 proteins in pyrenoid-enriched pellet fractions from low- and high-CO_2–grown cells (mean of four biological replicates; raw data are provided in *SI Appendix* and Dataset S1). RbcL, RBCS, EPYC1, and RCA1 (black) are abundant in low-CO_2 pellets, as determined by iBAQ (*y*-axis). In addition, these proteins showed increased abundance in low-CO_2 pellets compared with high-CO_2 pellets, as determined by label-free quantification (LFQ; *x*-axis). (*C*) Confocal microscopy of EPYC1-Venus and RBCS1-mCherry coexpressed in WT cells. (Scale bar: 5 μm.)

domain resembling the Rubisco small subunit (21). Here we find that Rubisco accumulation in the pyrenoid of the model alga *Chlamydomonas reinhardtii* is mediated by a disordered repeat protein, which we term Essential Pyrenoid Component 1 (EPYC1). Our findings suggest a mechanism for aggregation of Rubisco in the pyrenoid matrix, and highlight similarities and differences between the mechanisms of assembly of the eukaryotic and prokaryotic organelles.

Results

EPYC1 Is an Abundant Pyrenoid Component. We hypothesized that the pyrenoid contains unidentified components that are important for its biogenesis. Therefore, we used mass spectrometry to analyze the protein composition of the pyrenoid of *Chlamydomonas*, before and after applying a stimulus that induces pyrenoid growth. When cells are transferred from high CO_2 (2–5% CO_2 in air) to low CO_2 (~0.04% CO_2 in air), the CCM is induced (22) and the pyrenoid increases in size (23). We developed a protocol for isolating largely intact pyrenoids by cell lysis and centrifugation, and applied this protocol to cells before and after a shift from high to low CO_2 (Fig. 1*A* and *SI Appendix*, Fig. S1 *A–C*). Mass spectrometry indicated that the most abundant proteins in the low-CO_2 pyrenoid fraction included the Rubisco large (rbcL) and small (RBCS) subunits, as well as Rubisco activase (RCA1) (Fig. 1*B*, *SI Appendix*, Fig. S1*D*, and Dataset S1).

Strikingly, a fourth protein, previously identified as a low-CO_2–induced nuclear-encoded protein (LCI5; Cre10.g436550) (24), was found in the low-CO_2 pyrenoid fraction with comparable abundance to Rubisco (Fig. 1*B*). Based on the data presented herein, we propose naming this protein Essential Pyrenoid Component 1 (EPYC1). Under low CO_2, the stoichiometry of EPYC1, estimated by intensity-based absolute quantification (iBAQ), was ~1:6 with rbcL and ~1:1 with RBCS (25). Consistent with EPYC1 being a component of the pyrenoid, the abundance of EPYC1 in the pyrenoid fraction was increased by ~12-fold after the shift from high to low CO_2 (Fig. 1*B* and *SI Appendix*, Fig. S1*D* and Dataset S1), an increase similar to that of rbcL (7-fold), RBCS (7-fold), and RCA1 (19-fold). To confirm the pyrenoid localization of EPYC1, we coexpressed fluorescently tagged EPYC1 and RBCS. Venus-tagged EPYC1 showed clear colocalization with mCherry-tagged RBCS in the pyrenoid (Fig. 1*C* and *SI Appendix*, Fig. S1*E*).

EPYC1 Is Essential for a Functional CCM. The high abundance of EPYC1 in the pyrenoid led us to ask whether EPYC1 is required for the CCM. We isolated a mutant in the 5′ UTR of the *EPYC1* gene (*SI Appendix*, Fig. S2*A* and Table S2), which contains markedly reduced levels of *EPYC1* mRNA (*SI Appendix*, Fig. S2*B* and Table S3) and EPYC1 protein (Fig. 2*A*), and lacks transcriptional regulation in response to CO_2 (*SI Appendix*, Fig. S2*B*). Similar to previously described mutants in other components of the CCM, the *epyc1* mutant showed defective photoautotrophic growth in low CO_2, which was rescued by high CO_2 and by reintroducing the *EPYC1* gene (Fig. 2*B* and *SI Appendix*, Fig. S2 *C–E*).

We further tested the CCM activity in the *epyc1* mutant by measuring whole-cell affinity for inorganic carbon, inferred from photosynthetic O_2 evolution. When grown under low CO_2, the *epyc1* mutant showed a reduced affinity for inorganic carbon (increased $K_{0.5}$) relative to WT ($P = 0.0055$, Student's t test; $n = 5$) (Fig. 2*C* and *SI Appendix*, Fig. S2*F* and Table S4). The affinity of the *epyc1* mutant under low CO_2 was slightly greater than that of WT at high CO_2, indicating a residual level of CCM activity. This activity may be due to trace levels of EPYC1 in the *epyc1* mutant (*SI Appendix*, Fig. S2 *A* and *B*), or a normal CO_2 concentration followed by inefficient capture by Rubisco.

EPYC1 Is Required for Normal Pyrenoid Size, Number, and Matrix Density. Given that EPYC1 is in the pyrenoid and is required for the CCM, we explored whether the *epyc1* mutant shows any visible defects in pyrenoid structure. Thin-section transmission electron microscopy (TEM) revealed that the *epyc1* mutant had smaller pyrenoids than WT at both low and high CO_2 (low CO_2: $n = 37–79$, $P < 10^{-19}$, Welch's t test; high CO_2: $n = 18–22$, $P < 10^{-5}$,

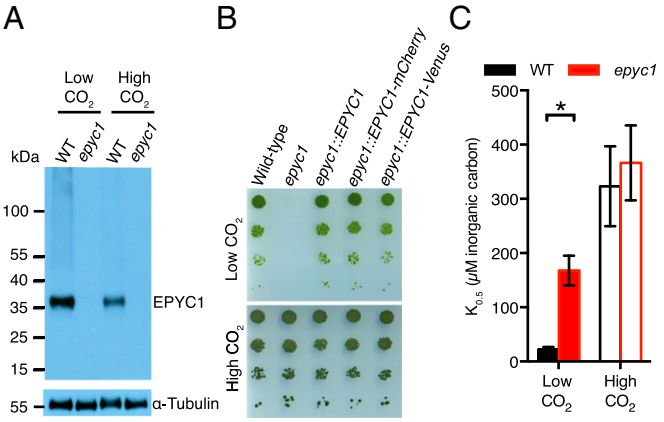


Fig. 2. EPYC1 is an essential component of the carbon-concentrating mechanism. (*A*) EPYC1 protein levels in WT and *epyc1* mutant cells grown at low and high CO_2 were probed by Western blot analysis with anti-EPYC1 antibodies. Anti-tubulin is shown as a loading control. (*B*) Growth phenotypes of WT, *epyc1*, and three lines complemented with *EPYC1*. Serial 1:10 dilutions of WT, *epyc1*, *epyc1*::*EPYC1*, *epyc1*::*EPYC1-mCherry*, and *epyc1*::*EPYC1-Venus* lines were spotted on TP minimal medium and grown at low and high CO_2 under 500 μmol photons m^{-2} s^{-1} illumination. (*C*) Inorganic carbon affinity of WT and *epyc1* cells. Cells were pregrown at low and high CO_2, and whole-cell inorganic carbon affinity was measured as the concentration of inorganic carbon at half-maximal O_2 evolution. Data are a mean of five low-CO_2 or three high-CO_2 biological replicates. Error bars represent SEM. *$P = 0.0055$, Student's t test.

Welch's t test) (Fig. 3 *A* and *B* and *SI Appendix*, Fig. S3 *A* and *B* and Fig. S4). *Chlamydomonas* typically has one pyrenoid per cell (17). The *epyc1* mutant showed a higher frequency of multiple pyrenoids; 13% of nondividing *epyc1* cells ($n = 231$) showed multiple pyrenoids, compared with 3% of WT cells ($n = 252$; $P = 0.00048$, Fisher's exact test of independence) (*SI Appendix*, Table S5). Higher-resolution quick-freeze deep-etch electron microscopy (QFDEEM) indicated a lower packing density of granular material in the pyrenoid matrix of the *epyc1* mutant (Fig. 3*C* and *SI Appendix*, Figs. S3*C* and S5). This defect was most noticeable when cells were grown in low CO_2, but was also visible at high CO_2.

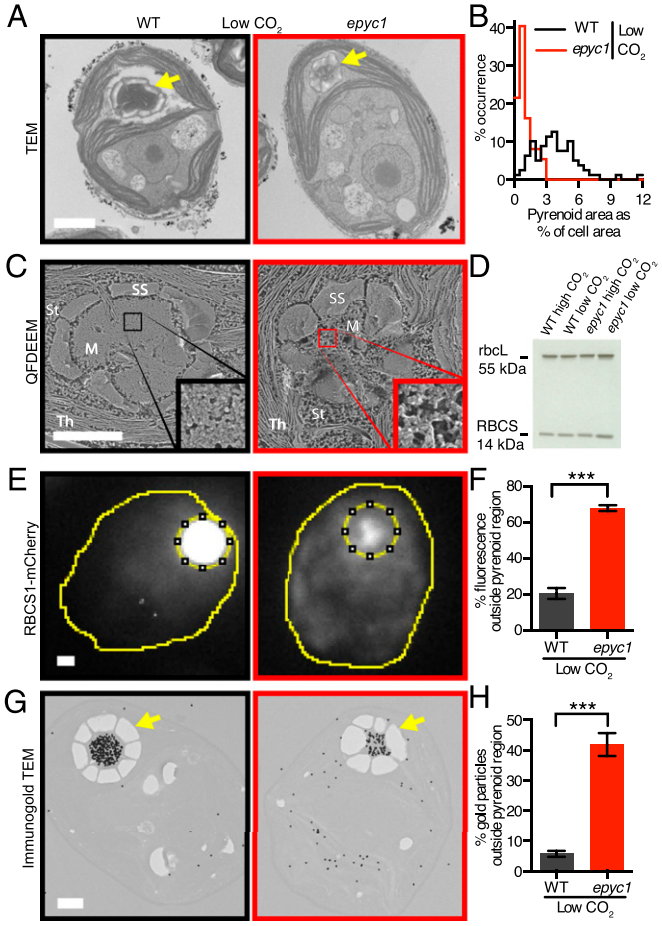


Fig. 3. EPYC1 is essential for Rubisco aggregation in the pyrenoid. (*A*) Representative TEMs of WT and *epyc1* cells grown at low CO_2. (*B*) Quantification of the pyrenoid area as a percentage of cell area of WT and *epyc1* cells grown at low CO_2. Data are from TEM images as represented in *A*. *epyc1*: $n = 37$; WT: $n = 79$. $P < 10^{-19}$, Welch's t test. (*C*) QFDEEM of the pyrenoid of WT and *epyc1* cells grown at low CO_2. M, pyrenoid matrix; St, stroma; Th, thylakoids; SS, starch sheath. (*Inset*) Four hundred percent zoom view of the pyrenoid matrix. (*D*) Rubisco protein levels in WT and *epyc1* cells grown at low and high CO_2 were probed by Western blot analysis. (*E*) Localization of Rubisco was determined by microscopy of WT and the *epyc1* mutant containing RBCS1-mCherry. The sum of the fluorescence signals from Z stacks was used for quantitation. (*F*) The fraction of RBCS1-mCherry signal from outside the pyrenoid region (inner dotted line in *E*) was quantified in WT and *epyc1* cells at low CO_2. *epyc1*: $n = 27$; WT: $n = 27$. ***$P < 10^{-15}$, Student's t test. (*G*) Representative images of anti-Rubisco immunogold labeling of WT and *epyc1* cells grown at low CO_2. Gold particles were enlarged 10× for visibility. (*H*) The fraction of immunogold particles outside the pyrenoid was quantified. WT: $n =$ 26 cells, 8,123 gold particles; *epyc1*: $n = 27$ cells, 2,708 gold particles. ***$P < 10^{-15}$, Student's t test. In *F* and *H*, data are mean values, with error bars indicating SEM. Yellow arrows indicate pyrenoids. (Scale bars: 1 μm.)

Interestingly, the *epyc1* mutant retains a number of canonical pyrenoid characteristics (17), including correct localization in the chloroplast, the presence of a starch sheath under low CO_2, and traversing membrane tubules, suggesting that normal levels of EPYC1 are not required for these characteristics. In addition, the *epyc1* mutant showed normal levels of the carbonic anhydrase CAH3, which is thought to be central in delivering CO_2 to Rubisco in the pyrenoid (*SI Appendix*, Fig. S2*G*).

EPYC1 Is Required for Rubisco Assembly into the Pyrenoid. Our observations of decreased pyrenoid size and apparent matrix density in the *epyc1* mutant could be explained by decreased whole-cell levels of Rubisco. However, Western blot analysis revealed no detectable difference in rbcL and RBCS abundance in *epyc1* relative to WT cells or between cells grown at low and high CO_2 levels (Fig. 3*D* and *SI Appendix*, Fig. S3*D*). This result led us to hypothesize that the localization of Rubisco was perturbed in the *epyc1* mutant. To test this hypothesis, we generated WT and *epyc1* cell lines expressing Rubisco tagged with mCherry, and determined the distribution of fluorescence signal by microscopy. Remarkably, a large fraction of Rubisco was found outside the pyrenoid in the *epyc1* mutant. In *epyc1* cells grown in low CO_2, 68% of fluorescence from Rubisco tagged with mCherry was found outside the pyrenoid region, compared with 21% in WT cells ($n = 27$; $P < 10^{-15}$, Student's t test) (Fig. 3 *E* and *F* and *SI Appendix*, Fig. S6). Immunogold-EM confirmed the mislocalization of Rubisco in *epyc1*. In pyrenoid-containing sections of low-CO_2–grown *epyc1* cells, 42% of anti-Rubisco immunogold particles were found outside the pyrenoid, whereas only 6% were found outside the pyrenoid in WT (WT: $n = 26$ cells, 8,123 gold particles; *epyc1*: $n = 27$ cells, 2,708 gold particles; $P < 10^{-15}$, Student's t test) (Fig. 3 *G* and *H* and *SI Appendix*, Fig. S7).

If EPYC1 functions in the recruitment of Rubisco to the pyrenoid solely at low CO_2 (23), then the *epyc1* mutant could be trapped in a "high-CO_2" state of Rubisco localization (23). However, the *epyc1* mutant showed a defect in Rubisco localization even under high CO_2 (*SI Appendix*, Fig. S3 *E* and *F* and Fig. S6), where the fraction of Rubisco-mCherry fluorescence outside the pyrenoid region increased to 80% in *epyc1*, compared with 68% in the WT (WT: $n = 20$; *epyc1*: $n = 20$; $P = 10^{-6}$, Student's t test). We conclude that EPYC1 is required for Rubisco localization to the pyrenoid not only at low CO_2, but also at high CO_2.

EPYC1 and Rubisco Are Part of the Same Complex. EPYC1 could promote the localization of Rubisco to the pyrenoid by a physical interaction. Thus, we immunoprecipitated EPYC1 and Rubisco, and probed the eluates by Western blot analysis (Fig. 4*A* and *SI Appendix*, Fig. S8*A*). Immunoprecipitation of tagged EPYC1 pulled down the Rubisco holoenzyme, and, reciprocally, tagged RBCS1 pulled down EPYC1. We conclude that EPYC1 and Rubisco are part of the same supramolecular complex in the pyrenoid. The high abundance of EPYC1 in the pyrenoid, EPYC1's physical interaction with Rubisco, and the dependence of Rubisco on EPYC1 for localization to the pyrenoid all suggest that EPYC1 plays a structural role in pyrenoid biogenesis.

The EPYC1 Protein Consists of Four Nearly Identical Repeats. To gain insight into how EPYC1 might contribute to pyrenoid biogenesis, we performed a detailed analysis of the EPYC1 protein sequence. This analysis indicated that EPYC1 consists of four nearly identical ~60-aa repeats (Fig. 4 *B*–*D*), flanked by short N and C termini, in contrast to a previous study suggesting only three repeats (26). We found that each repeat consists of a predicted disordered domain and a shorter, less disordered domain containing a predicted alpha helix (Fig. 4*C* and *SI Appendix*, Fig. S8 *B* and *C*). Given that these repeats cover >80% of the EPYC1 protein, it is likely that the Rubisco binding sites are contained within the repeats.

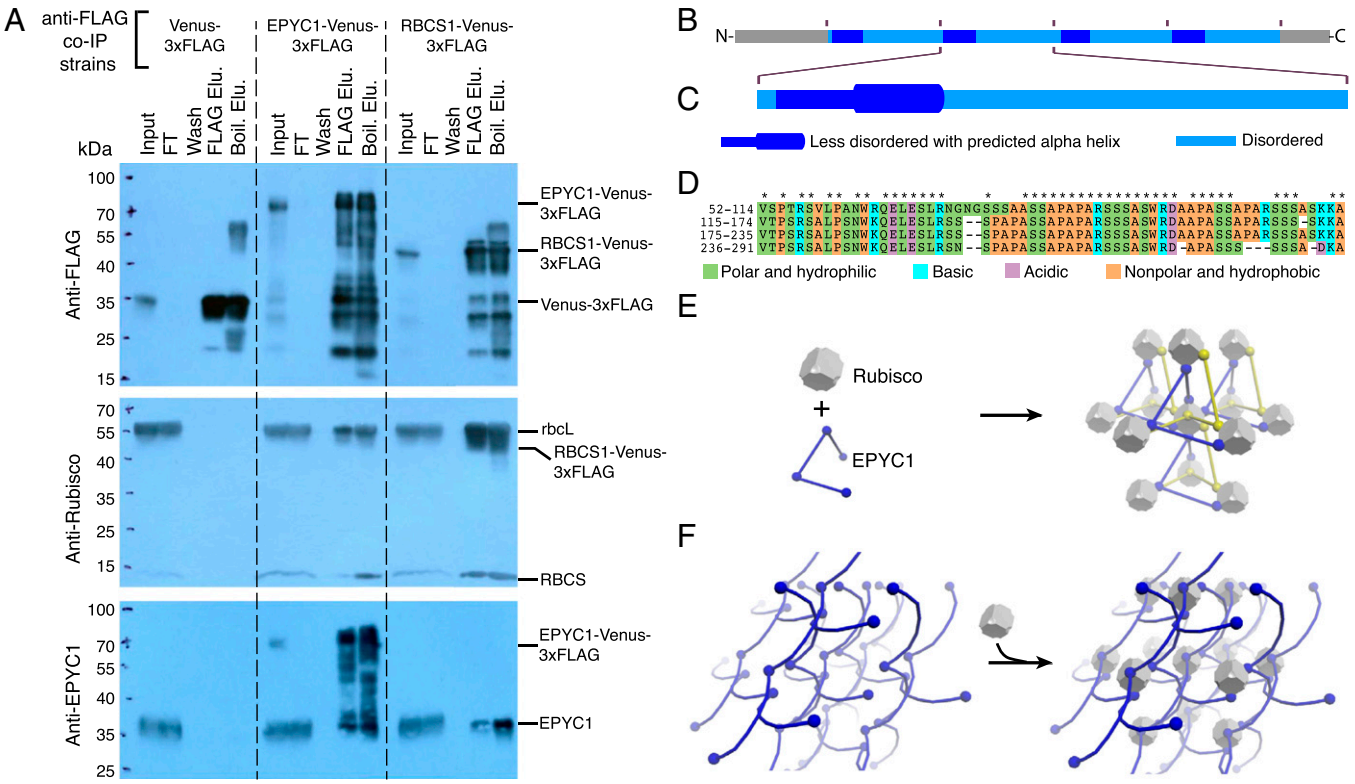


Fig. 4. EPYC1 forms a complex with Rubisco. (*A*) Anti-FLAG coimmunoprecipitation (co-IP) of WT cells expressing Venus-3×FLAG, EPYC1-Venus-3×FLAG, and RBCS1-Venus-3×FLAG. For each co-IP, the input, flow-through (FT), fourth wash (wash), 3×FLAG elution (FLAG Elu.), and boiling elution (Boil. Elu.) were probed with anti-FLAG, anti-Rubisco, or anti-EPYC1. Labels on the right show the expected sizes of proteins. (*B*) Analysis of the EPYC1 protein sequence showing that EPYC1 consists of four nearly identical repeats. (*C*) Each repeat has a highly disordered domain (light blue) and a less-disordered domain (dark blue) containing a predicted alpha-helix (thicker line) rich in charged residues. (*D*) Amino acid alignments of the four repeats. Asterisks indicate residues that are identical in all four repeats. (*E* and *F*) Two models illustrate how EPYC1 could bind the Rubisco holoenzyme in a manner compatible with the observed packing of Rubisco in the pyrenoid. (*E*) EPYC1 and Rubisco could form a codependent network. If each EPYC1 can bind four Rubisco holoenzymes, and each Rubisco holoenzyme can bind eight EPYC1s, eight EPYC1 proteins could connect each Rubisco to twelve neighboring Rubiscos. (*F*) EPYC1 could form a scaffold onto which Rubisco binds. Both arrangements could expand indefinitely in every direction. For clarity, the spacing between Rubisco holoenzymes was increased and EPYC1 is depicted in both yellow and blue.

We Propose Two Models for Rubisco Assembly into the Pyrenoid Matrix by EPYC1. If each repeat of EPYC1 binds Rubisco, then EPYC1 could link multiple Rubisco holoenzymes together to form the pyrenoid matrix. Multiple Rubisco binding sites on EPYC1 could arrange Rubisco into the hexagonal closely packed or cubic closely packed arrangement observed in recent cryoelectron tomography studies of the *Chlamydomonas* pyrenoid (18). EPYC1 and Rubisco could interact in one of two fundamental ways: (*i*) EPYC1 and Rubisco could form a codependent network (Fig. 4*E*), or (*ii*) EPYC1 could form a scaffold onto which Rubisco binds (Fig. 4*F*). Importantly, the 60-aa repeat length of EPYC1 is sufficient to span the observed 2- to 4.5-nm gap between Rubisco holoenzymes in the pyrenoid (18), and a stretched-out repeat could potentially span the observed 15-nm Rubisco center-to-center distance. A promising candidate for an EPYC1-binding site on Rubisco would be the two alpha-helices of the small Rubisco subunit. When these helices are exchanged for higher-plant alpha-helices, pyrenoids fail to form and the CCM does not function, but holoenzyme assembly is normal (20).

Proteins with Similar Physicochemical Properties to EPYC1 Are Present in a Diverse Range of Eukaryotic Algae. The primary sequences of disordered proteins like EPYC1 are known to evolve rapidly compared with those of structured proteins, but their physicochemical properties are under selective pressure and are evolutionarily maintained (27). Therefore, we searched for proteins with similar physicochemical properties (i.e., repeat number, length, high isoelectric point, disorder profile, and absence of transmembrane domains) across a broad range of algae (*SI Appendix*, Table S6). Excitingly, proteins with similar properties are found in most pyrenoid-containing algae and appear to be absent from pyrenoid-less algae, suggesting that EPYC1-like proteins may play similar roles in pyrenoids across eukaryotic algae. A thorough assessment of the generality of linker proteins will be enabled by future proteomic analyses of pyrenoids from a diverse range of algae.

Discussion

Our data provide strong support for the concept that Rubisco clustering into the pyrenoid is required for an efficient CCM in eukaryotic algae (9). Current models of the CCM (17, 28) suggest that CO_2 is released at a high concentration from the thylakoid tubules traversing the pyrenoid matrix. The mislocalization of Rubisco to the stroma of the *epyc1* mutant could decrease the efficiency of CO_2 capture by Rubisco, explaining the severe CCM defect observed in this mutant.

The observations presented here suggest that Rubisco packaging to form the matrix of the eukaryotic pyrenoid is achieved by a different mechanism than that used in the well-characterized prokaryotic β-carboxysome. In the β-carboxysome, aggregation of Rubisco is mediated by the protein CcmM (CO_2 concentrating mechanism protein M). CcmM contains multiple repeats of a domain resembling the Rubisco small subunit, and incorporation of these domains into separate Rubisco holoenzymes is thought to produce a link between Rubisco holoenzymes (21). Given that the

EPYC1 repeats show no homology to Rubisco and are highly disordered, it is likely that they bind to the surface of Rubisco holoenzymes rather than becoming incorporated in the place of small subunits. The simplicity of such a surface-binding mechanism potentially explains how Rubisco packaging into a pyrenoid could have evolved across a broad range of photosynthetic eukaryotes through convergent evolution (17, 29), leading to the dominant role of pyrenoids in aquatic CO_2 fixation. Such a surface-binding mechanism may even organize Rubisco in prokaryotic α-carboxysomes, where the intrinsically disordered Rubisco-binding repeat protein CsoS2 plays a poorly understood role in assembly (30).

In addition to being a key structural component, EPYC1 could regulate Rubisco partitioning to the pyrenoid or Rubisco kinetic properties. The Rubisco content of the pyrenoid changes in response to CO_2 (23 and our data), whereas total cellular Rubisco remains constant (Fig. 3*D*). Given that EPYC1 is required for Rubisco localization to the pyrenoid, changes in EPYC1 abundance and/or Rubisco-binding affinity could affect Rubisco partitioning to the pyrenoid. Consistent with this hypothesis, EPYC1 was previously found to be up-regulated at both the transcript and protein levels in response to light and low CO_2 (26), and our data further support this finding (Fig. 2*A* and *SI Appendix*, Fig. S2*A*). Moreover, previous studies have shown that EPYC1 becomes phosphorylated at multiple sites in response to low CO_2 (26, 31), potentially affecting its binding affinity for Rubisco.

Another mode of regulation of EPYC1–Rubisco binding could be through the methylation of Rubisco. Rubisco is methylated in multiple residues (32), and in *Chlamydomonas*, the predicted methyltransferase CIA6 is required for Rubisco localization to the pyrenoid (33). It is also possible that EPYC1 binding to Rubisco alters the kinetic properties of Rubisco to fine-tune its performance in the pyrenoid.

Along with advancing our understanding of the molecular mechanisms underlying global carbon fixation, our findings may contribute to the future engineering of crops with enhanced photosynthesis. There is great interest in introducing a CCM into C_3 plants, given that this enhancement is predicted to increase yields by up to 60% and to improve the efficiency of nitrogen and water use (34). Although much remains to be done to improve our understanding of the algal CCM, recent work suggests that algal components may be relatively easy to engineer into higher plants (35). Our discovery of a possible mechanism for Rubisco assembly to form the pyrenoid is a key step toward engineering an algal CCM into crops.

Materials and Methods

Strains and Culture Conditions. WT *Chlamydomonas* CC-1690 (36) was used for pyrenoid enrichment and proteomics. WT *Chlamydomonas* cMJ030 (CC-4533) (37) was used for all other experiments. The *epyc1* mutant was isolated from a collection of high-CO_2–requiring mutants generated by transformation of the pMJ016c mutagenesis cassette into cMJ030 (37). All experiments were performed under photoautotrophic conditions supplemented with high CO_2 (3% or 5% vol/vol CO_2-enriched air) or low CO_2 (air, ~0.04% vol/vol CO_2).

Proteomics. Pyrenoid enrichment was performed using a modified protocol based on previous studies (38, 39). In brief, cells were harvested by centrifugation (3,220 × *g* for 2 min at 4 °C), lysed by sonication, and then centrifuged again at 500 × *g* for 3 min at 4 °C to obtain a soluble fraction and a pellet fraction. Shotgun proteomics on the soluble and pellet fractions was performed as described by Mühlhaus et al. (40). Raw MS data files were processed with MaxQuant version 1.5.2.8 (41).

Cloning. *EPYC1* (Cre10.g436550) and *RBCS1* (Cre02.g120100) ORFs were amplified from gDNA and cloned into pLM005 (Venus) or pLM006 (mCherry) by Gibson assembly (42).

Transformation of *Chlamydomonas*. Constructs were transformed into the nuclear genome of WT and *epyc1* strains by electroporation as described by Zhang et al. (37). To screen for Venus- and mCherry-expressing colonies, transformation plates were imaged with a Typhoon Trio fluorescence scanner (GE Healthcare).

Microscopy. TEM images of the enriched pyrenoid fraction and whole cells before pyrenoid enrichment were prepared and taken according to Nordhues et al. (43). TEM imaging for pyrenoid area analysis and immunogold localization of Rubisco was based on methods described by Meyer et al. (20). QFDEEM was performed as described by Heuser (44). Fluorescence microscopy was performed using a spinning-disk confocal microscope (Leica DMI6000) with the following settings: Venus, 514 nm excitation with 543/22 nm emission; mCherry, 561 nm excitation with 590/20 nm emission; and chlorophyll, 561 nm excitation with 685/40 nm emission.

Quantitative Real-Time PCR. EPYC1 gene transcript levels were determined by qRT-PCR. CDNA was synthesized from total RNA, and relative gene expression was measured in real time using SYBR Green. Gene expression was calculated according to the method of Livak and Schmittgen (45), relative to RCK1 (Cre06.g278222) (46). The primers used are listed in *SI Appendix*, Table S2.

Western Blot Analysis. Protein levels of EPYC1 and CAH3 in WT and the *epyc1* mutant were measured according to Heinnickel et al. (47). Rubisco levels were measured as described by Meyer et al. (20).

O_2 Evolution and Spot Tests. Apparent affinity for inorganic carbon was determined using the oxygen evolution method described by Badger et al. (48). Spot tests were performed by spotting serially diluted WT, *epyc1*, and complemented cell lines onto Tris-phosphate (TP) plates. Plates were incubated in low or high CO_2 under 100 or 500 μmol photons m^{-2} s^{-1} of light for 7 d.

Coimmunoprecipitation. Cell lysate from WT cells expressing the bait proteins (Venus-3×FLAG, EPYC1-Venus-3×FLAG, or RbcS1-Venus-3×FLAG) was incubated with anti-FLAG M2 antibody (Sigma-Aldrich) bound to protein G Dynabeads (Life Technologies). Bait proteins with interaction partners were eluted by 3×FLAG competition, followed by boiling in 1× Laemmli buffer.

EPYC1 Sequence Analysis. To understand the intrinsic disorder of EPYC1, the full-length amino acid sequence was run through several structural disorder prediction programs, including VL3, VLTX (49), and GlobPlot 2 (50). To look for regions of secondary structure, the full-length and repeat regions of the EPYC1 amino acid sequence were analyzed by PSIPRED v3.3 (51) and Phyre2 (52).

Proteins with EPYC1-Like Physicochemical Properties in Other Algae. Complete translated genomic sequences from pyrenoid and non-pyrenoid algae were analyzed for tandem repeats using Xstream (53). The isoelectric point, disorder profile (54), and presence of transmembrane domains (55) of Xstream hits were calculated. Proteins with three or more repeats, a pI >8, an oscillating disorder profile with a frequency between 40 and 80 amino acids, and no transmembrane domains were classified as potential EPYC1-like Rubisco linker proteins.

More detailed information on the materials and methods used in this study is provided in *SI Appendix*, *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank J. N. Skepper, L. Carter, and M. Rütgers for TEM support, discussions on immunogold optimization, and ultramicrotomy; H. Cartwright for confocal microscopy support; S. Ramundo for technical advice with coimmunoprecipitation; W. Patena for help with data analysis; and W. Frommer, V. Walbot, P. Walter, and T. Cuellar for comments on the manuscript. The project was funded by National Science Foundation Grants EF-1105617 and IOS-1359682 (to L.C.M.M., L.P., G.R., and M.C.J.); the Carnegie Institution for Science (L.C.M.M. and M.C.J.); National Institutes of Health Grant T32GM007276 (to E.S.F.R., V.K.C., and A.I.); Biotechnology and Biological Sciences Research Council Grant BB/M007693/1 (to M.T.M. and H.G.); the Federal Ministry of Education and Research, Germany, within the frame of the *Go*FORSYS Research Unit for Systems Biology (Grant FKZ 0313924, to T.M.-A., F.S., M. Schroda, and M. Stitt); and the International Max Planck Research School of the Max Planck Society (T.M.-A. and T.M.).

1. Ellis RJ (1979) The most abundant protein in the world. *Trends Biochem Sci* 4(11): 241–244.
2. Falkowski PG, Barber RT, Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* 281(5374):200–207.
3. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281(5374):237–240.
4. Dismukes GC, et al. (2001) The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proc Natl Acad Sci USA* 98(5):2170–2175.
5. Somerville CR, Ogren WL (1979) A phosphoglycolate phosphatase-deficient mutant of *Arabidopsis*. *Nature* 280(5725):833–836.
6. Bauwe H, Hagemann M, Fernie AR (2010) Photorespiration: Players, partners and origin. *Trends Plant Sci* 15(6):330–336.

7. Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the evolution of C4 photosynthesis. *Annu Rev Plant Biol* 63:19–47.
8. Giordano M, Beardall J, Raven JA (2005) CO_2 concentrating mechanisms in algae: Mechanisms, environmental modulation, and evolution. *Annu Rev Plant Biol* 56: 99–131.
9. Badger MR, et al. (1998) The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. *Can J Bot* 76(6): 1052–1071.
10. Schmitz F (1882) *Die Chromatophoren der Algen: Vergleichende Untersuchungen über Bau und Entwicklung der Chlorophyllkörper und der analogen Farbstoffkörper der Algen* (M. Cohen & Sohn, Bonn, Germany).
11. Vaucher J-P (1803) *Histoire des Conferves D'eau Douce: Contenant Leurs Différents Modes De Reproduction, Et La Description De Leurs Principales Espèces, Suivie De L'histoire Des Trémelles Et Des Ulves D'eau Douce* (JJ Paschoud, Geneva, Switzerland).
12. Brown R (1967) Pyrenoid: Its structure distribution and function. *J Phycol* 3(Suppl 1): 5–7.
13. Behrenfeld MJ, et al. (2001) Biospheric primary production during an ENSO transition. *Science* 291(5513):2594–2597.
14. Rousseaux CS, Gregg WW (2013) Interannual variation in phytoplankton primary production at a global scale. *Remote Sens* 6(1):1–19.
15. Mann GD (1996) Chloroplast morphology, movements and inheritance in diatoms. *Cytology, Genetics and Molecular Biology of Algae*, eds Chaudhary BR, Agrawal SB (SPB Academic Publishing, Amsterdam), pp 249–274.
16. Thierstein HR, Young JR, eds (2004) *Coccolithophores: From Molecular Processes to Global Impact* (Springer, Heidelberg, Germany).
17. Meyer M, Griffiths H (2013) Origins and diversity of eukaryotic CO_2-concentrating mechanisms: Lessons for the future. *J Exp Bot* 64(3):769–786.
18. Engel BD, et al. (2015) Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron tomography. *eLife* 4:e04889.
19. McKay RML, Gibbs SP (1991) Composition and function of pyrenoids: Cytochemical and immunocytochemical approaches. *Can J Bot* 69(5):1040–1052.
20. Meyer MT, et al. (2012) Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. *Proc Natl Acad Sci USA* 109(47):19474–19479.
21. Long BM, Badger MR, Whitney SM, Price GD (2007) Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J Biol Chem* 282(40):29323–29335.
22. Berry J, Boynton J, Kaplan A, Badger M (1976) Growth and photosynthesis of *Chlamydomonas reinhardtii* as a function of CO_2 concentration. *Carnegie Institution of Washington Year Book* (Carnegie Institution of Washington, Washington, DC), Vol 75, pp 423–432.
23. Borkhsenious ON, Mason CB, Moroney JV (1998) The intracellular localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. *Plant Physiol* 116(4):1585–1591.
24. Miura K, et al. (2004) Expression profiling-based identification of CO_2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. *Plant Physiol* 135(3):1595–1607.
25. Schwanhäusser B, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473(7347):337–342.
26. Turkina MV, Blanco-Rivero A, Vainonen JP, Vener AV, Villarejo A (2006) CO_2 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*. *Proteomics* 6(9):2693–2704.
27. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18(6):756–764.
28. Wang Y, Stessman DJ, Spalding MH (2015) The CO_2-concentrating mechanism and photosynthetic carbon assimilation in limiting CO_2: How *Chlamydomonas* works against the gradient. *Plant J* 82(3):429–448.
29. Villarreal JC, Renner SS (2012) Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc Natl Acad Sci USA* 109(46):18873–18878.
30. Cai F, et al. (2015) Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. *Life (Basel)* 5(2):1141–1171.
31. Wang H, et al. (2014) The global phosphoproteome of *Chlamydomonas reinhardtii* reveals complex organellar phosphorylation in the flagella and thylakoid membrane. *Mol Cell Proteomics* 13(9):2337–2353.
32. Taylor TC, Backlund A, Bjorhall K, Spreitzer RJ, Andersson I (2001) First crystal structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*. *J Biol Chem* 276(51): 48159–48164.
33. Ma Y, Pollock SV, Xiao Y, Cunnusamy K, Moroney JV (2011) Identification of a novel gene, CIA6, required for normal pyrenoid formation in *Chlamydomonas reinhardtii*. *Plant Physiol* 156(2):884–896.
34. Long SP, Marshall-Colon A, Zhu XG (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161(1):56–66.
35. Atkinson N, et al. (2015) Introducing an algal carbon-concentrating mechanism into higher plants: Location and incorporation of key components. *Plant Biotechnol J*, 10.1111/pbi.12497.
36. Sager R (1955) Inheritance in the green alga *Chlamydomonas reinhardi*. *Genetics* 40(4):476–489.
37. Zhang R, et al. (2014) High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell* 26(4):1398–1409.
38. Kuchitsu K, Tsuzuki M, Miyachi S (1988) Changes of starch localization within the chloroplast induced by changes in CO_2 concentration during growth of *Chlamydomonas reinhardtii*: Independent regulation of pyrenoid starch and stroma starch. *Plant Cell Physiol* 29(8):1269–1278.
39. Kuchitsu K, Tsuzuki M, Miyachi S (1991) Polypeptide composition and enzyme activities of the pyrenoid and its regulation by CO_2 concentration in unicellular green algae. *Can J Bot* 69(5):1062–1069.
40. Mühlhaus T, Weiss J, Hemme D, Sommer F, Schroda M (2011) Quantitative shotgun proteomics using a uniform 15N-labeled standard to monitor proteome dynamics in time course experiments reveals new insights into the heat stress response of *Chlamydomonas reinhardtii*. *Mol Cell Proteom* 10(9):M110.004739.
41. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies, and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372.
42. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.
43. Nordhues A, et al. (2012) Evidence for a role of VIPP1 in the structural organization of the photosynthetic apparatus in *Chlamydomonas*. *Plant Cell* 24(2):637–659.
44. Heuser JE (2011) The origins and evolution of freeze-etch electron microscopy. *J Electron Microsc (Tokyo)* 60(Suppl 1):S3–S29.
45. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25(4):402–408.
46. Schloss JA (1990) A *Chlamydomonas* gene encodes a G protein β subunit-like polypeptide. *Mol Gen Genet* 221(3):443–452.
47. Heinnickel ML, et al. (2013) Novel thylakoid membrane GreenCut protein CPLD38 impacts accumulation of the cytochrome b6f complex and associated regulatory processes. *J Biol Chem* 288(10):7024–7036.
48. Badger MR, Kaplan A, Berry JA (1980) Internal inorganic carbon pool of *Chlamydomonas reinhardtii*: Evidence for a carbon dioxide-concentrating mechanism. *Plant Physiol* 66(3):407–413.
49. Obradovic Z, et al. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins* 53(Suppl 6):566–572.
50. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 31(13):3701–3708.
51. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41(Web Server issue, W1):W349–W357.
52. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858.
53. Newman AM, Cooper JB (2007) XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics* 8(1):382.
54. Romero P, et al. (2001) Sequence complexity of disordered protein. *Proteins* 42(1): 38–48.
55. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305(3):567–580.

Supporting Information Appendix

A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle

Luke C. M. Mackinder[a], Moritz T. Meyer[b], Tabea Mettler-Altmann[c,1], Vivian K. Chen[a,d], Madeline C. Mitchell[b,2], Oliver Caspari[b], Elizabeth S. Freeman Rosenzweig[a,d], Leif Pallesen[a], Gregory Reeves[a,3], Alan Itakura[a,d], Robyn Roth[e], Frederik Sommer[c,4], Stefan Geimer[f], Timo Mühlhaus[c,4], Michael Schroda[c,4], Ursula Goodenough[e], Mark Stitt[c], Howard Griffiths[b], Martin C. Jonikas[a,d,5]

[a]Carnegie Institution for Science, Department of Plant Biology, Stanford, California 94305, USA. [b]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. [c]Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany. [d]Department of Biology, Stanford University, Stanford, CA 94305, USA. [e]Department of Biology, Washington University, St. Louis, Missouri 63130, USA. [f]Cell Biology & Electron Microscopy, University of Bayreuth, 95440 Bayreuth, Germany.

Present addresses: [1]Cluster of Excellence in Plant Sciences and Institute of Plant Biochemistry, Heinrich-Heine-University, 40225 Düsseldorf, Germany; [2]Agriculture, Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT 2601, Australia; [3]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK; [4]Institute of Molecular Biotechnology and Systems Biology, Technical University of Kaiserslautern, 67663 Kaiserslautern, Germany.

[5]To whom correspondence should be addressed. Email: mjonikas@carnegiescience.edu. Telephone: +1 650 739 4216

SI Materials and Methods

Strains and culture conditions

Chlamydomonas reinhardtii strain CC-1690 was maintained at 22°C with 55 µmol photons m^{-2} s^{-1} light on tris-acetate-phosphate (TAP) (56) agar (1.4%) plates containing 0.4% Bacto-Yeast extract (Becton, USA). The cMJ030 wild-type (WT) and *epyc1* mutant were maintained in the dark or low light (~10 µmol photons m^{-2} s^{-1}) on 1.5% agar plates containing TAP with revised (57) or traditional Hutner's trace elements (58).

For proteomics analysis, a 50 mL pre-culture was grown mixotrophically in TAP on a rotatory shaker at 124 rpm, 22°C and under an illumination of 55 µmol photons m^{-2} s^{-1} for three days according to Mettler *et al.* (59). In brief, a second pre-culture of 500 mL was used to inoculate a 5-litre bioreactor BIOSTAT®B-DCU (Sartorius Stedim, Germany). The absence of contamination was monitored according to Mettler *et al.* (59). Cultures with a cell density of 3-5 x 10^6 cells mL^{-1} were grown photoautotrophically at 46 µmol photons m^{-2} s^{-1} light in air enriched with high CO_2 (5% CO_2) under constant turbidity for two days before the culture was aerated with low CO_2 (ambient air; 0.039% CO_2). The CO_2 level in the outlet air of the bioreactor was measured by an on-line multi-valve gas chromatograph (3000A MicroGC run by EZChromElute software, Agilent Technologies, USA). After switching from high to low CO_2, the CO_2 dropped from 4.5% to a constant 0.02% after 12 hours. Cells were harvested at 30 hours after the shift to low CO_2.

For mRNA levels, O_2 evolution, Rubisco content western blotting, pyrenoid size analysis by transmission electron microscopy (TEM) and Rubisco subcellular localization by immuno-gold labelling, strains were grown photoautotrophically in 50 mL tris-minimal medium (58) under constant aeration, shaking and illumination (150 rpm, 21°C, 50-65 µmols photons m^{-2} s^{-1}; Infors HT Multitron Pro, Switzerland). Briefly, starter cultures were inoculated from freshly re-plated cultures on TAP plates, to 0.3 µg chlorophyll (a+b) mL^{-1}, and aerated with high CO_2 (5% v/v CO_2 enriched air). When cell density reached mid-log (~3 µg chlorophyll (a+b) mL^{-1}), half of the cultures were harvested and analysed. The remaining half of the cultures were then switched to aeration with low CO_2 (air, 0.04% v/v CO_2) for induction of the CCM. For gene expression analysis and affinity for inorganic carbon, cells were air-adapted for 3 hours, corresponding to peak induction of CO_2-inducible genes (60-62). The state of CCM induction was controlled by measuring the mRNA accumulation of a highly CO_2-responsive gene, *LCI1* (Cre03.g162800). For TEM analysis of pyrenoid size and immuno-gold labelling of Rubisco, cells were adapted to low CO_2 for 24 hours.

For EPYC1 protein abundance and freeze fracture cryo-electron microscopy of WT and mutant cells, cultures were propagated continuously in tris-phospate (TP) (57) medium with 50 µmols photons m^{-2} s^{-1} light for ~1 week in a Multi-Cultivator (Photon Systems Instruments) with bubbling of high CO_2 (3% v/v CO_2). Cells were diluted every 24 hours to ensure they were kept in the log phase. 6 hours before sampling, half of the cultures were switched from high CO_2 to low CO_2 (air, ~0.04% v/v CO_2). The chlorophyll concentration at harvesting was ~3 µg chlorophyll (a+b) mL^{-1}.

For fluorescence microscopy and RbcS1-mCherry localization experiments, cells were grown in TP medium containing antibiotics used for selection of expression of the fluorescently labeled gene (Venus, paromomycin at 2 µg mL^{-1}; mCherry, hygromycin 6.25 µg mL^{-1}), bubbled with high CO_2 (3% v/v CO_2) at a 150 µmol photons m^{-2} s^{-1} light intensity. At ~2 x 10^6 cells mL^{-1}, after >6 doublings, cells were transferred to low CO_2 for 14 hours. For the RbcS1-mCherry localization experiments, samples were taken and imaged immediately before the switch to low CO_2 and after 14 hours at low CO_2.

For co-immunoprecipitation experiments, cells were grown in 50 mL of TAP at 150 µmol photons m^{-2} s^{-1} light until ~2-4 x 10^6 cells mL^{-1}, centrifuged at 1000 *g* for 4 min, resuspended in TP and used to inoculate 800 mL of TP. Cells were then bubbled with low CO_2 (air, ~0.04% v/v CO_2) at 150 µmol photons m^{-2} s^{-1} until ~2-4 x 10^6 cells mL^{-1} and harvested as indicated below. All liquid media contained 2 µg mL^{-1} paromomycin.

Cell concentrations were measured using a Z2 Coulter Counter (Beckman Coulter, USA).

Proteomics of pyrenoid-enriched fraction

Pyrenoid enrichment

10 mL algal material (3-5 x 10^6 cells mL^{-1}) were harvested by centrifugation for 2 min (3,220 x *g*, 4°C), immediately frozen in liquid nitrogen and extracted with extraction buffer (EB; 50 mM HEPES, 20µM leupeptin, 1 mM PMSF, 17.4% glycerol, 2% Triton). The samples were sonicated 6 x 15 s (6 cycles, 50% intensity, Sonoplus Bandelin Electronics, Germany) and kept on ice between cycles for 90 s. The samples were centrifuged at 500 *g* for 3 min to obtain a soluble and pellet fraction. This procedure resembled the first steps of a protocol used in previous studies (38, 39). The pellet was washed three times with 1 mL, 500 µL and 300 µL EB before resuspension in 100µL 50 mM ammonium bicarbonate. Protein concentrations were measured by Lowry assay using BSA as a standard (63).

SDS-PAGE

For SDS-PAGE, samples were resuspended in a buffer containing 50 mM dithiothreitol (DTT), 50 mM sodium-carbonate, 15% sucrose (w/v) and 2.5% SDS (w/v), heated 45 seconds at 95°C and spun down at 16,700 x *g* before applying 22 µg total protein to the polyacrylamide gel. The 14%-separating gel was stained with Coomassie Brilliant Blue (64).

Protein digestion and mass spectrometric analysis

For shotgun proteomics, samples were prepared and measured according to Mühlhaus *et al.* (40). In brief, 20 µg protein per sample was precipitated in 80% acetone at -20°C over night. The precipitated proteins were resuspended in 6 M urea and 2 M thiourea (in 50 mM ammonium hydrogen carbonate), reduced by DTT, carbamidomethylated with iodoacetamide, digested with endoproteinase LysC (Roche, Switzerland) and immobilised trypsin (Applied Biosystems, Thermo Fisher Scientific, USA), and subsequently desalted according to Mühlhaus *et al.* (40). The resuspended peptides were acidified with 1% acetic acid. Peptides were chromatographically separated by reverse

phase separation with a nanoUPLC (nanoACQUITY UPLC, Waters, Milford, USA) using a 10 cm x 75 µm BEH130 C18 1.7 µm particles (Waters) column for separation and a 2 cm x 180 µm Symmetry C18 5 µm particles (Waters) column for trapping. Peptides were analysed by a linear trap quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) according to Mühlhaus *et al.* (40).

Data processing and data analysis

Raw MS files were processed with MaxQuant (version 1.5.2.8) (41). Peak list files were searched against *Chlamydomonas reinhardtii* gene model JGIv4 from Phytozome 10.2 (http://phytozome.jgi.doe.gov/) including the organelle genome sequences. Maximum precursor and fragment mass deviations were set to 20 ppm and 0.5 Da. Peptides with at least six amino acids were considered for identification. The search included carbamidomethylation as a fixed modification and variable modifications for oxidation of methionine and protein N-terminal acetylation. The false discovery rate, determined by searching a reverse database, was set at 0.01 for both peptides and proteins. Identification across different replicates and treatments was achieved by enabling the "match between runs" option in MaxQuant within a time window of 2 min. For comparison of protein levels between samples, the label-free quantification (LFQ) intensity based method was used (41). For the estimation of protein stoichiometries within a sample, the intensity-based absolute quantification (iBAQ) (25) method was applied. Both values were calculated by the MaxQuant software. All statistical analyses were performed using Microsoft Excel.

Cloning of EPYC1 and RbcS1

EPYC1 (Cre10.g436550) and *RBCS1* (Cre02.g120100) ORFs were amplified from gDNA using Phusion Hotstart II polymerase (Thermo Scientific) with the respective EPYC1_ORF_F/R or RBCS1_ORF_F/R primer pairs (*SI Appendix*, Table S2). Gel purified PCR products, containing vector overlap regions, were cloned into pLM005 or pLM006 by Gibson assembly (42). Final pLM005 constructs are in frame with Venus-3xFLAG and contain the AphVIII gene for paromomycin resistance, final pLM006 constructs are in frame with mCherry-6xHIS and contain the AphVII gene for hygromycin resistance. Both pLM005 and pLM006 confer ampicillin resistance for bacterial selection. For complementation with untagged EPYC1, mCherry-6xHIS was removed from pLM006-EPYC1-mCherry-6xHIS by BglII restriction digestion, gel purified then re-ligated. All constructs were verified by Sanger sequencing. GenBank accession numbers are: pLM005-EPYC1-Venus, KX077944; pLM005-Venus-Cloning-Vector, KX077945; pLM006-EPYC1, KX077946; pLM006-EPYC1-mCherry, KX077947; pLM006-mCherry, KX077948; pLM006-mCherry-Cloning-Vector, KX077949; pLM006-RBCS1-mCherry, KX077950; pMJ016c, KX077951.

Transformation of *Chlamydomonas* for complementation and fluorescence localization of proteins

Transformation was performed by electroporation as in Zhang *et al.* (37). Nuclear transformation in *Chlamydomonas* occurs by non-homologous insertion of DNA into the genome, resulting in random integration.

For each transformation, 14.5 ng kbp^{-1} of EcoRV cut plasmid was mixed with 250 µL of 2 x 10^8 cells mL^{-1} at 16 °C and transformed immediately into WT or *epyc1* strains by electroporation. Cells we selected on TAP paromomycin (20 µg mL^{-1}) or hygromycin (25 µg mL^{-1}) plates and kept in low light (5-10 µmol photons m^{-2} s^{-1}) until picking or screening for fluorescence lines. In addition, for the complementation of the *epyc1* mutant, a second transformation was selected on TP plates, without antibiotics at low CO_2 (~0.04% v/v CO_2) under 500 µmol photons m^{-2} s^{-1} light.

To screen for Venus and mCherry expressing lines, transformations were spread on rectangular plates (Singer Instruments) containing 86 mL of TAP plus antibiotics. Once colonies were ~2-3 mm in diameter, plates were transferred to ~100 µmol photons m^{-2} s^{-1} light for 24-36 hours and then screened for colony fluorescence on a Typhoon Trio fluorescence scanner (GE Healthcare). Excitation and emission settings were: Venus, 532 excitation with 555/20 emission; mCherry, 532 excitation with 610/30 emission; and chlorophyll autofluorescence, 633 excitation with 670/30 emission. Dual-tag lines were generated sequentially by expressing pLM005-EPYC1-Venus-3xFLAG in WT then adding pLM006-RbcS1-mCherry-6xHIS. To confirm expression of both Venus and mCherry in dual-tag strains and to select for strains with equal fluorescence intensity for the analysis of RbcS1-mCherry localization in WT and *epyc1,* strains were also screened on a Tecan Infinite M1000 PRO (65).

Fluorescence microscopy and Rubisco-mCherry mislocalization in the *epyc1* mutant

All fluorescence microscopy was performed using a spinning disk confocal microscope (3i custom adapted Leica DMI6000) with samples imaged on poly-L-lysine coated plates (Ibidi). The following excitation and emission settings were used: Venus, 514 excitation with 543/22 emission; mCherry, 561 excitation with 590/20 emission; and chlorophyll, 561 excitation with 685/40 emission. Images were analysed using Fiji software. For RbcS1-mCherry localization in WT and the *epyc1* mutant, lines showing equal RbcS1-mCherry fluorescence intensity were chosen for analysis (see above). WT and *epyc1* lines were imaged using the above mCherry and chlorophyll settings. A Z-stack composed of 40 slices 0.3 µm apart was obtained for each field of view. To quantify the percentage of fluorescence signal from outside the pyrenoid region, raw images were analysed as follows: Pixel intensity in the mCherry channel was summed across the 40 Z-sections for cells that were fully sectioned. Using the chlorophyll channel as a reference a cell outline region of interest (ROI; varying between cells) and pyrenoid ROI (set at 2.8 µm in diameter for WT and mutant) were drawn. For each cell, background fluorescence was subtracted by taking the average of 4 measurements surrounding the cell, and autofluorescence was subtracted separately from the pyrenoid and whole cell ROIs by taking the average of 22 WT cells not expressing mCherry. Finally, the percentage RbcS1-mCherry signal from outside of the pyrenoid region was calculated as the (*total cell signal* - *pyrenoid signal*) / *total cell signal* x 100%.

Analysis of gene expression by qRT-PCR

Total RNA was extracted from 30 µg chlorophyll a+b (~10 mL mid-log cell suspension), using TRIzol Reagent, as per manufacturer's instructions (Life Technologies). Complementary DNA was synthesised from 500 ng of total RNA using SuperScript III Reverse Transcriptase (Life Technologies), RNaseOUT (Life Technologies), and

oligo(dT)18 primers (Thermo Scientific). Relative gene expression was measured in real time in a Rotor-Gene Q thermocycler (Qiagen). Reactions (10 µL) used SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich). Gene expression was calculated according to the method of Livak and Schmittgen (45), relative to the *Chlamydomonas* gene coding for the Receptor of Activated Protein Kinase C1 (RCK1, Cre06.g278222) (46), which is not significantly induced by low-CO_2 (61). All primers used are in *SI Appendix* (Table S2).

Screening for the *epyc1* mutant

The *epyc1* mutant was isolated from a collection of high CO_2 requiring mutants by a pooled screening approach. A collection of approximately 7,500 mutants on 79 plates, each with 96 colonies, was grown in liquid TAP in 96 well plates then pooled by well row, well column, whole plate row and whole plate column to give a total of 38 pools. Pooled cells were pelleted, DNA was extracted by phenol:chloroform:isoamyl alcohol (Phenol:CIA, 25:24:1; Sigma-Aldrich) and then screened by PCR for an EPYC1 mutant using a primer in the pMJ016c mutagenesis cassette (a modified pMJ013c cassette) (37) and a primer in the EPYC1 gene. The identified *epyc1* mutant has an insertion of the pMJ016c resistance cassette in the 5'UTR, the resistance cassette is 11 bp upstream of the ATG start codon, with the cassette having a 10 bp deletion at the 3' end. The upstream gDNA-cassette junction cannot be PCR amplified. However, PCR shows the full cassette is intact and that >397 bp upstream of the insertion site is also intact (Fig S2*A*). All primers used are in *SI Appendix* (Table S2).

Protein extraction and western blotting

For EPYC1 and CAH3 protein quantification in WT and the *epyc1* mutant, protein was extracted from unfrozen cells, normalised to chlorophyll, separated by SDS-PAGE and western blotted as described in Heinnickel *et al.* (47). Both the primary anti-EPYC1 and anti-CAH3 (Agrisera) antibodies were used at a 1:2,000 concentration and the secondary horseradish-peroxidase (HRP) conjugated goat anti-rabbit (Life Technologies) at a 1:10,000 concentration. To ensure even loading, membranes were stripped (Restore PLUS western blot stripping buffer, Thermo Scientific) and re-probed with anti-tubulin (1:25,000; Sigma) followed by HRP conjugated goat anti-mouse (1:10,000; Life Technologies). The anti-EPYC1 antibody was raised in rabbit to the C-terminal region of EPYC1 (KSKPEIKRTALPADWRKGL-cooh) by Yenzym Antibodies (South San Francisco, California, USA).

For Rubisco quantification in WT and *epyc1* mutant, total soluble proteins were extracted from 300 µg chlorophyll (a+b) (~100 mL mid-log cell suspension). Cells were harvested by centrifugation (13,000 *g*, 10 min, 4°C), re-suspended in ice cold 1.5 mL extraction buffer (50 mM Bicine, pH 8.0, 10 mM $NaHCO_3$, 10 mM $MgCl_2$, and 1 mM DTT), and lysed by sonication (6 x 30 second bursts of 20 microns amplitude, with 15 s on ice between bursts; Soniprep 150, MSE UK, London, UK). Lysis was checked by inspecting samples under a light microscope. Lysate was clarified by centrifugation (13,000 *g*, 20 min, 4°C). Protein content was determined using the Bradford method (Sigma Aldrich). Soluble proteins were separated on 12% (w/v) denaturing polyacrylamide gel. Sample loading was normalised by protein amount (10 µg per lane), and even loading was controlled by staining a gel with identical protein load (GelCode Blue, Life Technologies). After transfer onto a polyvinylidene difluoride membrane

(Amersham), Rubisco was immuno-detected with a polyclonal primary antibody raised against Rubisco (1:10,000) followed by a HRP conjugated goat anti-rabbit (1:20,000; GE Healthcare).

Chlorophyll concentration

Total pigments were extracted in 100% methanol, and the absorbance of the clarified supernatant (13,000 *g*, 1 min, 4°C) was measured at 470, 652, 665, 750 nm (UV 300 Unicam, Thermo Spectronic, Cambridge, UK). Concentration of chlorophyll (a+b) was calculated using the equation of Wellburn (66).

Spot tests

WT, *epyc1* and complemented cell lines were grown in TAP until ~$2x10^6$ cells mL^{-1}, washed once with TP, resuspended in TP to a concentration of $6.6x10^5$ cells mL^{-1}, then serially diluted 1:10 three times. 15 µL of each dilution was spotted onto four TP plates and incubated in low or high CO_2 with 100 or 500 µmol photons m^{-2} s^{-1} of light for seven days before imaging.

Oxygen evolution measurements

Apparent affinity for inorganic carbon was determined by oxygen evolution (48). Photoautotrophically grown liquid cultures were harvested by centrifugation (2,000 *g*, 5 min, 4°C) and re-suspended in 25 mM HEPES-KOH (pH 7.3) to a density of ~1.5 x 10^8 cells mL^{-1}, as determined by haemocytometer count. Aliquots of cells (1 mL) were added to a Clark-type oxygen electrode chamber (Rank Brothers, Bottisham, UK) attached to a circulating water bath set to 25°C. The chamber was closed for a light pre-treatment (200-300 µmol photons m^{-2} s^{-1} illumination for 10-25 min), to allow cells to deplete any internal inorganic carbon pool. When net oxygen evolution ceased, 10 µL of increasingly concentrated $NaHCO_3$ solution was added to the algal suspension at 30 second intervals, and the rate of oxygen evolution was recorded every second using a PicoLog 1216 data logger (Pico Technologies, St Neots, UK). Cumulative concentrations of sodium bicarbonate after each addition were as follows: 2.5, 5, 10, 25, 50, 100, 250, 500, 1,000, and 2,000 mM. Michaelis-Menten curves were fitted to plots of external inorganic carbon concentration versus the rate of O_2 evolution. The concentration of inorganic carbon required for half maximal rates of photosynthesis ($K_{0.5}$) was calculated from this curve.

Pyrenoid area analysis by transmission electron microscopy

To minimise the loss of biological signal during harvesting, fixative (glutaraldehyde, final 2.5%) was added to cell cultures immediately before harvesting. Cell suspensions containing ~5 x 10^7 cells in mid-log were pelleted (4,000 *g*, 5 min, 4°C) and fixed in 1 mL tris-minimal medium containing 2.5% glutaraldehyde and 1% H_2O_2 (30% w/v) for 1 hour on a tube rotator at 4°C. Unless otherwise specified, all following steps were performed at room temperature on a tube rotator. Cells were pelleted (4,000 *g* 5 min) and washed with ddH_2O (3X, 5 min). Cells were osmicated for 1 hour in 1 mL 1% (v/v) OsO_4 containing 1.5% (w/v) $K_3[Fe(CN)_6]$ and 2 mM $CaCl_2$. Cells were pelleted and washed with ddH_2O (4X, as above). Cells were stained for 1 hour in 1 mL 2% (w/v) uranyl acetate. After pelleting

and washing with ddH_2O (3X), cells were dehydrated in 70%, 95%, 100% ethanol, and 100% acetonitrile (2X). Cells were embedded in epoxy resin mix, containing Quetol 651, nonenyl succinic anhydride, methyl-5-norbornene-2,3-dicarboxylic anhydride, and dimethyl-benzylamine (all reagents from Agar Scientific, Stansted, UK), in the following proportions: 35%, 46%, 17%, 2%. Resin was refreshed 4 times over two days. Thin sections (50 nm) were prepared by the Cambridge Advanced Imaging Centre (Ms Lyn Carter) on a Leica Ultracut UCT Ultramicrotome and mounted onto 300 mesh copper grids. Samples were imaged with a Tecnai G2 transmission electron microscope (FEI, Hilsboro, OR, USA) at 200 kV. Image analysis (area measurements) was performed using ImageJ. Ten 54 μm^2 areas were randomly selected and all pyrenoid positive cells were imaged (WT low CO_2, 79 out of 271 cells displayed a pyrenoid; *epyc1* low CO_2, 37 out of 139 cells displayed a pyrenoid; WT high CO_2, 18 out of 196 cells displayed a pyrenoid; *epyc1* high CO_2, 22 out of 255 cells displayed a pyrenoid). Cell area was determined by outlining the plasma membrane. Pyrenoid area was taken as the area inside the starch sheath (generally visible in CCM-induced cells) or the electron dense area inside the chloroplast when no starch sheath was visible. Control immuno-gold labelling experiments using a high concentration of primary antibody (1:20) confirmed that these areas had dense concentrations of Rubisco. Pyrenoid area was expressed as a percentage of cell area, and data was ordained in classes of 0.5% increment.

Quick-freeze deep-etch EM (QFDEEM)

Sampling and fixation

It was ascertained in pilot experiments that pyrenoids fixed by the following procedure are indistinguishable in QFDEEM ultrastructure from unfixed controls. 150 mL of each of air-bubbled cultures and 75 mL of high CO_2-bubbled cultures were pelleted at 1,000 *g* for 10 min at RT to produce pellets of ~200 µL. The pellets were resuspended in 6 mL of ice-cold 10 mM HEPES buffer (pH 7) and transferred to a cold 25 mL glass flask. A freshly prepared solution of 4% glutaraldehyde (Sigma-Aldrich G7651) in 10 mM HEPES (pH 7) was added 100 µL at a time, swirling between drops, until 1.5 mL in total had been added. The mixture was then left on ice for 1 hour, with agitation every 10 min. The mixture was pelleted (1000 *g*, 5 min, 4° C), washed in cold HEPES buffer, pelleted again, and finally resuspended in 6 mL fresh HEPES. Samples were shipped overnight to St. Louis in 15 mL conical screw cap tubes maintained at 0-4° C.

Microscopy

QFDEEM was performed as in Heuser (44). Briefly, small samples of pelleted cells were placed on a cushioning material and dropped onto a liquid helium-cooled copper block; the frozen material was transferred to liquid nitrogen and then to an evacuated Balzers apparatus, fractured, etched at -80°C for 2 min, and platinum/carbon rotary-replicated. The replicas were examined with a JEOL electron microscope, model JEM 1400, equipped with an AMTV601 digital camera. The images are photographic negatives; hence, protuberant elements of the fractured/etched surface are more heavily coated with platinum and appear whiter.

Immunogold-localization of Rubisco

Resin embedded material previously used for ultra-structural characterization of the pyrenoid was re-cut and thin sections were mounted on nickel grids. Osmium removal and unmasking of epitopes was done by acid treatment (67). Grids were gently floated face down on a droplet (~30 µL) of 4% sodium meta-periodate (w/v in ddH_2O) for 15 min, and 1% periodic acid (w/v in ddH_2O) for 5 min. Each acid treatment was followed by several short washes in ddH_2O. Non-specific sites were blocked for 5 min in 1% BSA (w/v) dissolved in high-salt tris-buffered saline containing 500 mM NaCl, 0.05% Triton X-100 and 0.05% Tween 20 (hereafter abbreviated HSTBSTT). Salt, detergent, and surfactant concentrations were determined empirically to minimise background signal. Binding to primary antibody was done by incubating grids overnight in 1% BSA in HSTBSTT, with 1:1,000 dilution of the Rubisco antibody. Excess antibody was removed by 15 min washes (2X) in HSTBSTT and 15 min washes (2X) in ddH_2O. Incubation with secondary antibody (15 nm gold particle-conjugated goat anti-rabbit secondary antibody in 1% BSA in HSTBSTT, 1:250) was done at RT for 1 hr. Excess secondary antibody was removed by washing as above. Thin sections were prepared and imaged as for *Pyrenoid area analysis by transmission electron microscopy*, above. Randomisation was done as above (see TEM) with scoring capped to ~25 cells for each treatment. Non-specific labelling was taken as any particle on a free resin area, i.e. outside a cell. Non-specific density was subtracted from pyrenoid and chloroplast particle density. Fraction of particles in the pyrenoid was calculated as background-adjusted $n_{pyrenoid} / (n_{pyrenoid} + n_{stroma})$, where n_{stroma} is the number of particles in the stroma to the exclusion of the pyrenoid and the starch sheath. To improve the clarity of gold particles in Fig. 3g, particles were enlarged 10x using the image analysis software, Fiji. Briefly, images were thresholded to isolate individual gold particles, these were then enlarged 10x, and the new image overlaid on the original image with an opacity of 50%.

Co-Immunoprecipitations

WT cells expressing pLM005-Venus-3xFLAG, pLM005-EPYC1-Venus-3xFLAG or pLM005-RbcS1-Venus-3xFLAG were grown in 800 mL of TP plus 2 µg mL^{-1} paromomycin with continual bubbling at low CO_2 (0.04% CO_2) under 150 µmol photons m^{-2} s^{-1} of light until a cell density of ~2-4 x 10^6 cells mL^{-1}. Cells were then spun out (2,000 *g*, 4 min, 4°C), washed in 40 mL of ice cold TP, centrifuged then resuspended in a 1:1 (v/w) ratio of ice-cold 2xIP buffer (400 mM sorbitol, 100 mM HEPES, 100 mM KOAc, 4 mM $Mg(OAc)_2.4H_2O$, 2 mM $CaCl_2$, 2 mM NaF, 0.6 mM Na_3VO_4 and 1 Roche cOmplete EDTA-free protease inhibitor/ 25 mL) to cell pellet. This cell slurry was then added drop wise to liquid nitrogen to form small *Chlamydomonas* "popcorn" balls approximately 5 mm in diameter. These were stored at -70°C until needed.

Cells were lysed by grinding 1g (~500 mg of original cell pellet) of *Chlamydomonas* popcorn balls by mortar and pestle at liquid nitrogen temperatures, for 10 min. The ground cells were defrosted on ice, then dounced 20 times on ice with a Kontes Glass Co. Duall #21 homogenizer. Membranes were solubilized by incrementally adding an equal volume of ice-cold 1xIP buffer plus 2% digitonin (final concentration is 1%), then incubating at 4°C for 40 min with nutation. The lysate was then clarified by spinning for 30 min at full-speed in a table-top centrifuge at 4°C. The supernatant (Input) was then transferred to 225 µL of protein G Dynabeads (Life

Technologies) that had been incubated with anti-FLAG M2 antibody (Sigma) according to the manufacturer's instructions, except 1xIP buffer was used for the wash steps. The Dynabead-cell lysate was incubated for 2.5 hours on a rotating platform at 4°C, then the supernatant removed (Flow-through). The Dynabeads were washed 4 times with 1xIP buffer plus 0.1% digitonin followed by a 30 min elution with 50 µL of 1xIP buffer plus 0.25% digitonin and 2 µg/ µL 3xFLAG peptide (Sigma; 3xFLAG peptide elution) and a 10 min elution in 1x Laemmli buffer with 50 mM beta-mercaptoethanol at 70°C (Boiling elution). Samples were run on 10% SDS-PAGE gels, then silver stained or transferred to PVDF membrane and probed with anti-FLAG (1:2,000; secondary: 1:10,000 HRP goat anti-mouse), anti-Rubisco (1:10,000; secondary: 1:20,000 HRP goat anti-rabbit) or anti-EPYC1 (1:2,000; secondary: 1:10,000 HRP goat anti-rabbit).

EPYC1-Rubisco interaction model

We built a model of the EPYC1-Rubisco interaction using Blender (www.blender.org) based on the following logic: If each of the 4 EPYC1 repeats can bind a holoenzyme, the 2 internal repeats would have different linking properties from the 2 terminal repeats. If bound to an internal repeat, a holoenzyme would be directly linked through this EPYC1 protein to 2 other holoenzymes. In contrast, if bound to a terminal repeat, the holoenzyme would only be directly linked through this EPYC1 protein to one other holoenzyme. Therefore on average, each EPYC1 repeat would link one Rubisco holoenzyme to 1.5 other holoenzymes. Given the octameric structure of the Rubisco holoenzyme, a holoenzyme likely has 8 binding sites for EPYC1. Taken together, on average each holoenzyme would be bound to 12 other holoenzymes by 8 EPYC1 proteins, in an arrangement that could expand indefinitely in all directions. A perfect arrangement of this nature would require a stoichiometry of one EPYC1 polypeptide for every four Rubisco small or large subunits.

Analysis of other algal proteomes for EPYC1-like physicochemical properties

Complete translated genomic sequences from pyrenoid and non-pyrenoid algae were downloaded from Uniprot or Phytozome. Protein sequences were then analyzed for tandem repeats using Xstream (53) with default settings except the following were set to: Min Period, 40; Max Period, 80; Min Copy #, 3.0; Min TR Domain, 75; Min Seq Content, 0.7. The pI of the Xstream hits were then batch calculated using the Gene Infinity Protein Isoelectric Point calculator (http://www.geneinfinity.org/sms/sms_proteiniep.html), the disorder profile calculated using VLXT (54) and the presence of transmembrane domains using TMHMM v. 2.0 (55). Proteins with an oscillating disorder profile with a frequency between 40-80 were classified as potential EPYC1-like Rubisco linker proteins. By applying stringent parameters we have tried to reduce the number of false positive hits but we realize that our approach has several limitations, including: 1) Missing true linker proteins due to not all of the physicochemical properties of EPYC1 being essential for linker function. 2) Incomplete genome assembly of the investigated algae. 3) Incorrect gene models resulting in truncated, mis-spliced and frame-shifted proteins.

Statistical methods

When growing algal material in liquid medium, flasks were placed randomly throughout the orbital shaker/incubator. Placement was randomized after each sub-culturing to offset any differences in illumination quality. Manifold for air/CO_2 delivery had standardized tubing length and internal diameter for even aeration. Cells lysis via sonication required samples to be processed sequentially. Order of processing was randomized.

Sample size of O_2 evolution measurement was aligned to previously published work from the Griffiths Lab (20, 62). Sample size of electron microscopy related experiments (scoring of TEM thin sections and immunogold experiments) was validated by jackknife resampling.

Pre-established exclusion criteria for TEM image scoring were: (i) only grid areas fully covered with material (54 μm^2) were considered; (ii) sections through broken cells and cell sections with a cross area < 12.5 μm^2 (a circle with 2 μm radius served as a guide) were not scored.

Scoring of electron micrographs: images files were renamed with a random number (RANDBETWEEN function in Microsoft Excel), sorted from high to low, and scored blindly. The original filename appearing on the bottom left of each micrograph was masked during the on-screen processing in ImageJ. Randomly selected images were scored by a second experimenter for independent validation. No systematic bias (over- or underestimation) was measured, and measurements deviated on average only by a couple of percentage points.

Two-tailed Student's *t test* was used to compare affinities for inorganic carbon of WT and *epyc1*, as well as the mislocalization of Rubisco by fluorescence microscopy and EM, because this test is robust to non-normal distributions (68). Welch's *t test* was used to compare pyrenoid sizes, because the WT and mutant groups had substantially different standard deviations (68). Fisher's exact test of independence was used to compare the number of pyrenoids in WT and *epyc*, as this test is appropriate when there are two nominal variables (68).

SI Appendix References

56. Gorman DS & Levine R (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*. *Proceedings of the National Academy of Sciences of the United States of America* 54(6):1665-1669.
57. Kropat J*, et al.* (2011) A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *The Plant journal : for cell and molecular biology* 66(5):770-780.
58. Spreitzer RJ & Mets L (1981) Photosynthesis-deficient Mutants of *Chlamydomonas reinhardii* with Associated Light-sensitive Phenotypes. *Plant physiology* 67(3):565-569.
59. Mettler T*, et al.* (2014) Systems Analysis of the Response of Photosynthesis, Metabolism, and Growth to an Increase in Irradiance in the Photosynthetic Model Organism *Chlamydomonas reinhardtii*. *The Plant cell* 26(6):2310-2350.
60. Brueggeman AJ*, et al.* (2012) Activation of the carbon concentrating mechanism by CO_2 deprivation coincides with massive transcriptional restructuring in *Chlamydomonas reinhardtii*. *The Plant cell* 24(5):1860-1875.
61. Fang W*, et al.* (2012) Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator CIA5/CCM1. *The Plant cell* 24(5):1876-1893.
62. Mitchell MC, Meyer MT, & Griffiths H (2014) Dynamics of carbon-concentrating mechanism induction and protein relocalization during the dark-to-light transition in synchronized *Chlamydomonas reinhardtii*. *Plant physiology* 166(2):1073-1082.
63. Lowry OH, Rosebrough NJ, Farr AL, & Randall RJ (Protein measurement with the Folin phenol reagent.
64. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680-685.
65. Yang W*, et al.* (2014) Alternative acetate production pathways in *Chlamydomonas reinhardtii* during dark anoxia and the dominant role of chloroplasts in fermentative acetate production. *The Plant cell* 26(11):4499-4518.
66. Wellburn AR (1994) The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of plant physiology* 144(3):307-313.
67. Skepper J (2000) Immunocytochemical strategies for electron microscopy: choice or compromise. *Journal of microscopy* 199(1):1-36.
68. McDonald JH (2009) *Handbook of biological statistics* (Sparky House Publishing Baltimore, MD).
69. Flombaum P*, et al.* (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences* 110(24):9824-9829.
70. Not F*, et al.* (2004) A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. *Applied and Environmental Microbiology* 70(7):4064-4072.
71. Chrétiennot-Dinet M*, et al.* (1995) A new marine picoeucaryote: *Ostreococcus tauri* gen. et sp. nov.(Chlorophyta, Prasinophyceae). *Phycologia* 34(4):285-292.

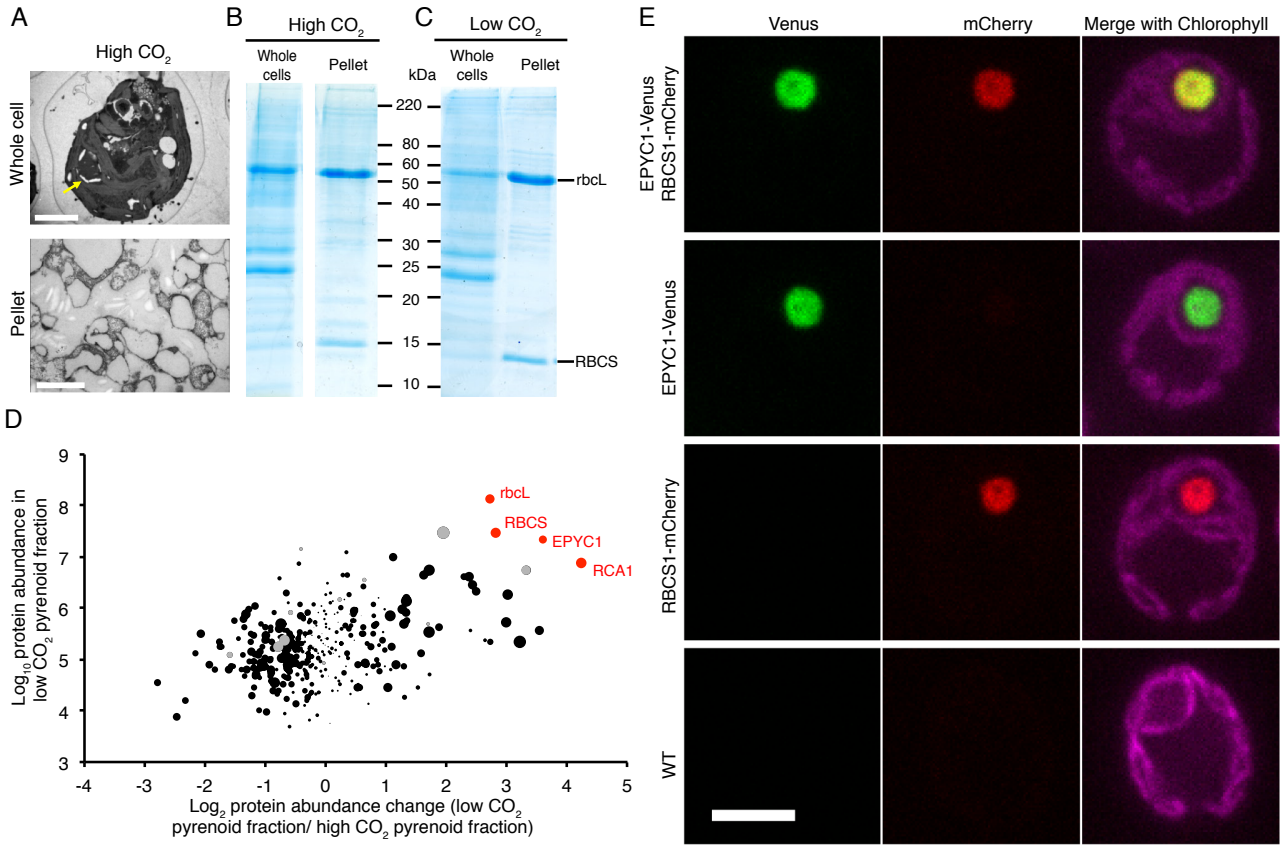


Fig. S1. EPYC1 is an abundant pyrenoid component. (*A*) TEM images of *Chlamydomonas reinhardtii* whole cells and pyrenoid-enriched pellet from cells grown at high CO_2. Yellow arrow indicates pyrenoid. (Scale bars: 2 μm.) *(B)* Coomassie-stained SDS-PAGE of whole cell and pyrenoid-enriched pellet at high CO_2. (*C*) Coomassie-stained SDS-PAGE of whole cell and pyrenoid-enriched pellet at low CO_2. (*D*) Additional analysis of the mass spectrometry data shown in Fig. 1*B* and Dataset S1. The x-axis is the label-free quantification (LFQ) enrichment in the low-CO_2 pellet fraction relative to the high CO_2 pellet fraction. The iBAQ given on the y-axis, represents the absolute protein abundance in the low-CO_2 pellet. Red data points highlight RbcL, RBCS, EPYC1 and RCA1. Grey circles depict sets of peptides represented by more than one protein due to high sequence similarities, whereas black circles are peptides representing a single protein. Dot sizes indicate the $\log_{10}$ *P*-value between low CO_2 and high CO_2 pellet fractions (Student´s *t* test). In total 366 proteins were identified in all four replicates of both the low- and high-CO_2-grown pellets. (*E*) Confocal microscopy of EPYC1-Venus and RBCS1-mCherry co-expressed and individually expressed in WT cells. All images were taken and processed with the same settings. Top panel is the same cell as shown in Fig. 1*C*. (Scale bar: 5 μm.)

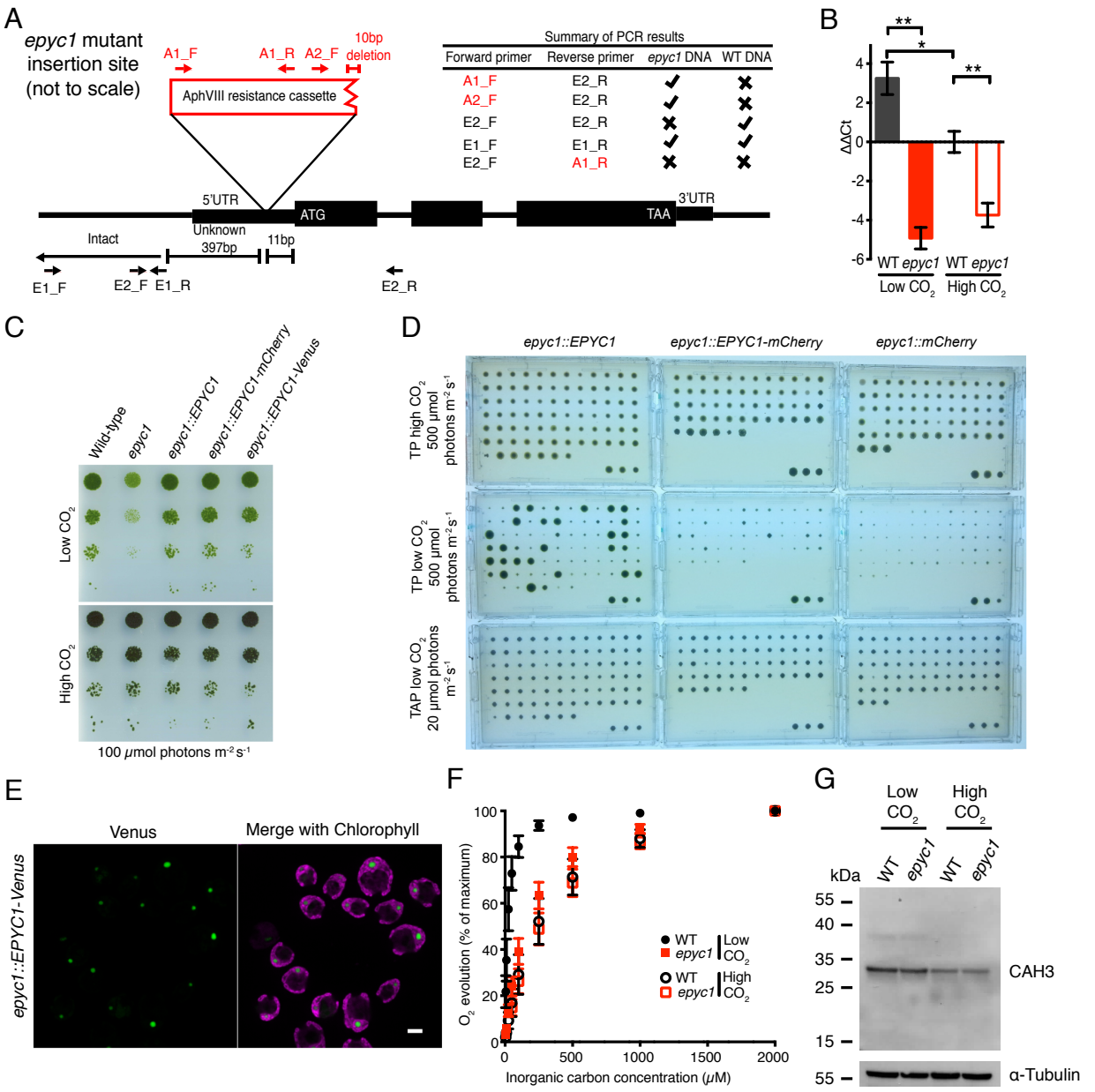


Fig. S2. Characterization and complementation of the *epyc1* mutant. (*A*) Cartoon of the *epyc1* mutant insertion site. The pMJ016c resistance cassette conferring paromomycin through the AphVIII gene is 11 bp upstream of the *EPYC1* ATG start codon. The resistance cassette has a known 10 bp deletion at the 3' end and is fully intact. The junction of the 3' cassette end and the *EPYC1* gene can be amplified with a forward primer annealing at the 5' end of the cassette (A1_F) and a reverse primer in the *EPYC1* gene (E2_R). The 5' end of the insert is still uncharacterized. Primers upstream of the insertion site fail to give PCR products when paired with reverse primers in the resistance cassette (e.g. E2_F and A1_R). It is known that insertion sites can undergo large insertions and/or deletions in *Chlamydomonas* (37). However, a large deletion upstream of the insertion site is ruled out due to the amplification of a region upstream of the insertion site (E1_F and E1_R) in the *epyc1* mutant. Note the cartoon is not to scale for clarity. (*B*) Quantification of *EPYC1* transcript levels in WT and the *epyc1* mutant at low and high CO_2 by qRT-PCR. Transcript levels are normalized to the reference gene *RCK1*, and plotted relative to WT at high CO_2. In the *epyc1* mutant, transcript levels were ~250-fold lower than in WT at low CO_2; and transcript levels were not significantly upregulated between low and high CO_2 ($P = 0.129$, Student's *t test*). Data is the mean of 3 biological replicates each measured in triplicate. Error bars: SEM * indicates $P < 0.05$, ** indicates $P < 0.005$, Student's *t* test. (*C*) Growth phenotypes of WT, *epyc1* and 3 *epyc1* complemented lines at 100 μmol photons m^{-2} s^{-1} light intensity. Serial 1:10 dilutions of WT, *epyc1*, *epyc1*::*EPYC1*, *epyc1*::*EPYC1-mCherry* and *epyc1*::*EPYC1-Venus* lines were spotted on TP minimal plates and grown at low and high CO_2 under 100 μmol photons m^{-2} s^{-1}. (*D*) Complementation screening of the *epyc1* mutant. The *epyc1* mutant was transformed with pLM006-EPYC1,

pLM006-EPYC1-mCherry or pLM006-mCherry and selected on TAP plates with hygromycin. Hygromycin resistant colonies were picked into a 96 format and propagated twice on TAP with hygromycin plates to allow even growth of all colonies. Colonies were then replicated onto TP or TAP plates and incubated as shown for 12 days before imaging. pLM006-EPYC1 fully rescued the *epyc1* mutant in 28% (22/79) of cases and partially rescued the mutant in 10% (8/79) of cases. pLM006-EPYC1-mCherry partially rescued the mutant in 20% (11/54) of cases. The negative control, pLM006-mCherry, failed to rescue the mutant (0/62). Partially rescued colonies were colonies that were visibly smaller in size than WT but larger in size than negative control colonies. The residual growth seen in all colonies at low CO_2 is carryover from the initial pinning from TAP plates. The 3 bottom right colonies are WT controls. (*E*) Confocal microscopy of the *epyc1::EPYC1-Venus* complemented line used for spot tests in Fig. 2*B*. Cells were grown mixotrophically in TAP media and imaged by confocal microscopy as in the materials and methods section. (Scale bar: 5 μm.) (*F*) The *epyc1* mutant has reduced inorganic carbon affinity. WT and *epyc1* cells were grown photoautotrophically at low and high CO_2, and whole cell inorganic carbon affinity was measured by O_2 evolution with step-wise increases in inorganic carbon. The $K_{0.5}$ values shown in Fig. 2*C* are derived from these curves. Data is a mean of 5 low CO_2 or 3 high CO_2 biological replicates. For clarity the data has been plotted as a percentage of maximum with the raw data provided in *SI Appendix* (Table S4). Error bars: SD. (*G*) CAH3 protein levels in WT and *epyc1* mutant cells grown at low and high CO_2 were probed by western blotting with anti-CAH3 antibodies. Anti-tubulin is shown as a loading control.

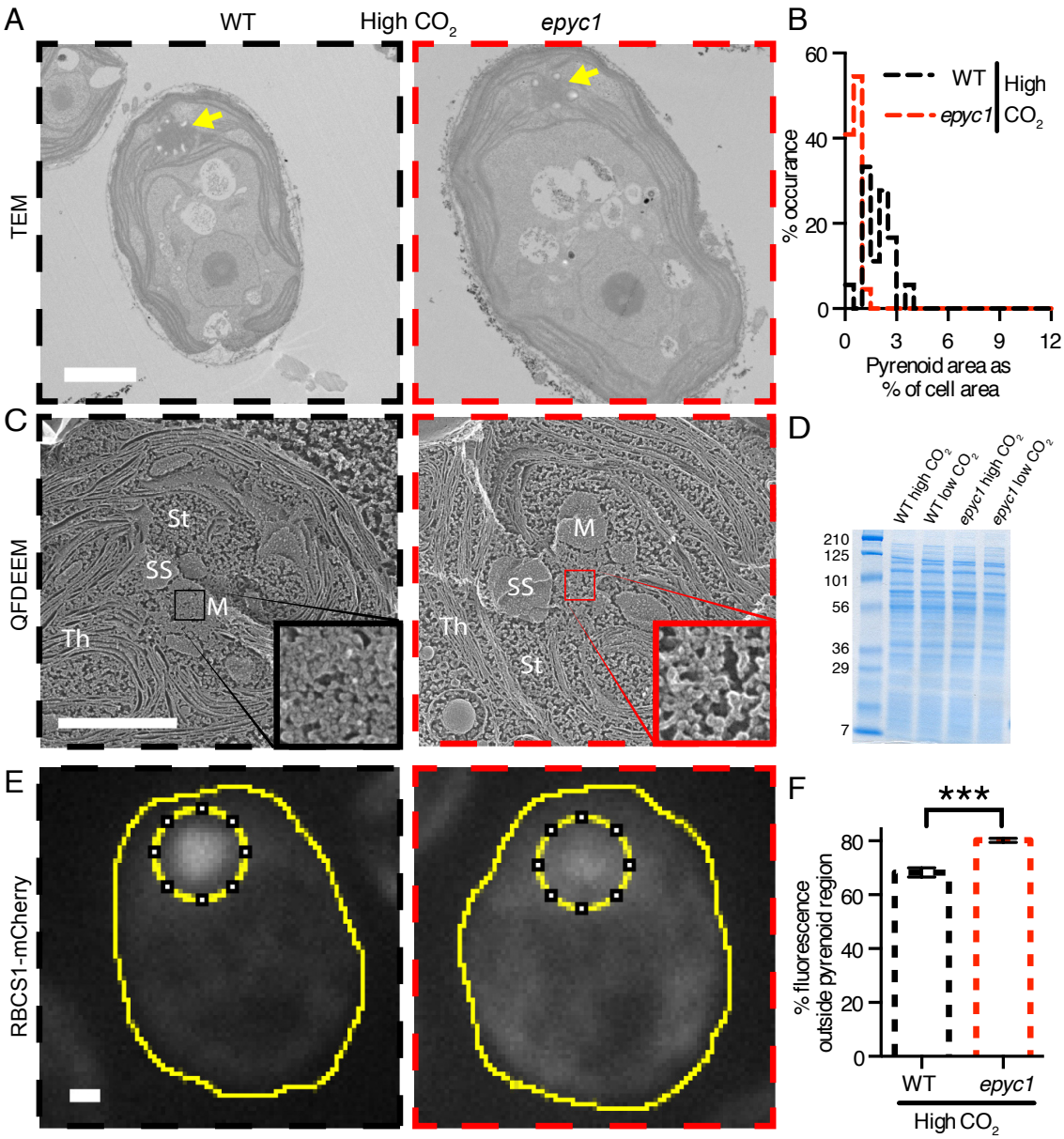


Fig. S3. Rubisco is mislocalized in the *epyc1* mutant at high CO_2. (*A*) Representative TEMs of WT and *epyc1* cells grown at high CO_2. Yellow arrows indicate pyrenoids. (*B*) Quantification of pyrenoid area as percentage of cell area of WT and *epyc1* cells grown at high CO_2 (WT: $n = 18$, *epyc1*: $n = 22$, $P < 10^{-5}$, Welch's *t test*). (*C*) Quick-Freeze Deep-Etch EM (QFDEEM) of the pyrenoid of WT and *epyc1* cells grown at high CO_2. M, pyrenoid matrix; St, stroma; Th, thylakoids; SS, starch sheath. Insets are a 400% zoom of the pyrenoid matrix. (*D*) Coomassie stained SDS-PAGE loading control gel for samples used in Fig. 3*D*. (*E*) The localization of Rubisco was determined by microscopy of WT and *epyc1* mutants containing RBCS1-mCherry at high CO_2. (*F*) The fraction of RBCS1-mCherry signal from outside the pyrenoid region (inner dotted line, *E*) was quantified in WT and *epyc1*. The sum of fluorescence signal from Z stacks is shown and was used for quantitation. WT: $n = 20$, *epyc1*: $n = 20$, *** represents $P = 10^{-6}$, Student's *t test*. (Scale bars: 1 μm.)

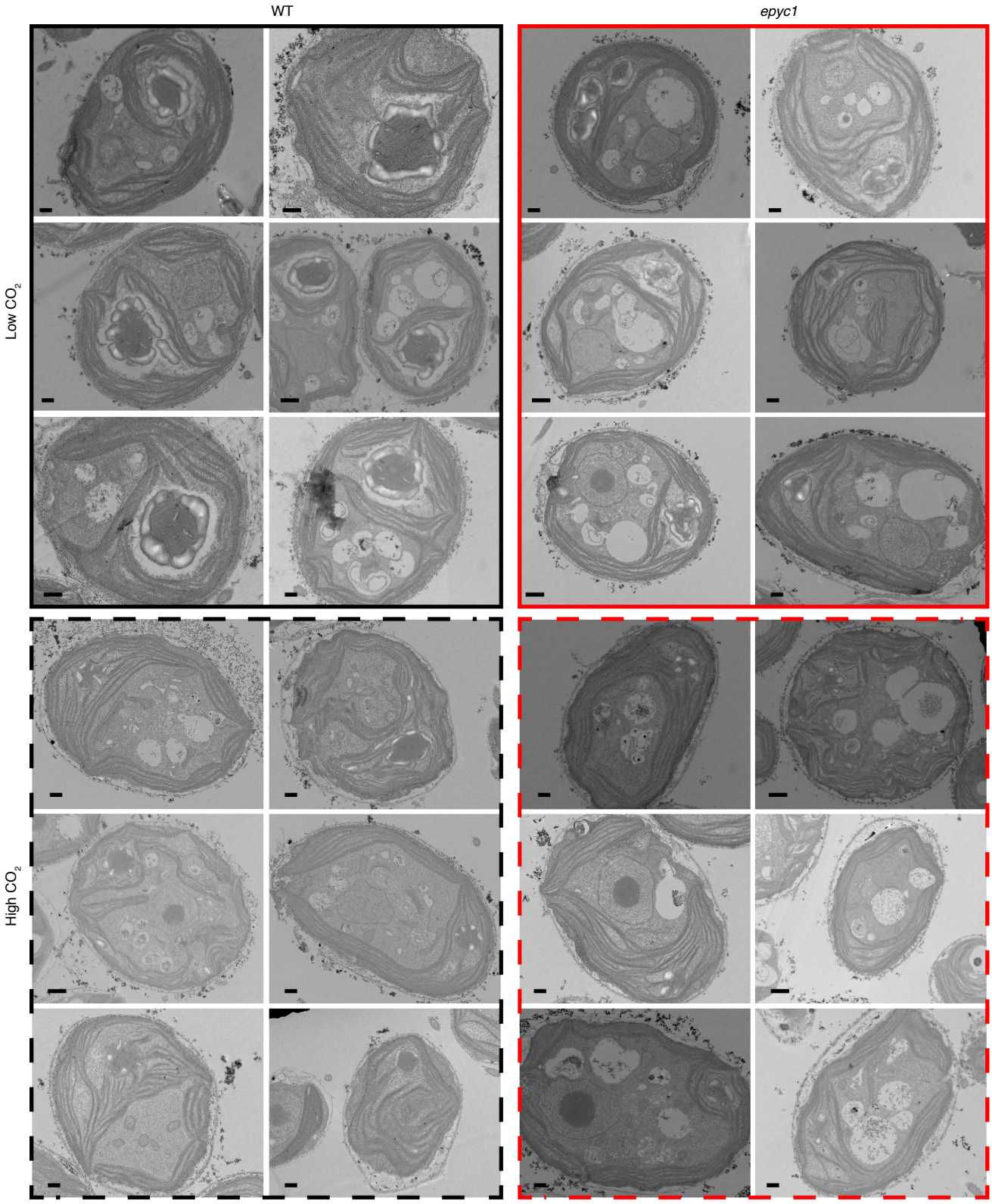


Fig. S4. Representative TEM images of WT and *epyc1* cells at low and high CO_2. A representative selection of TEM images used for pyrenoid area analysis. Cells were prepared and imaged for TEM as in the materials and methods. (Scale bars: 500 nm.)

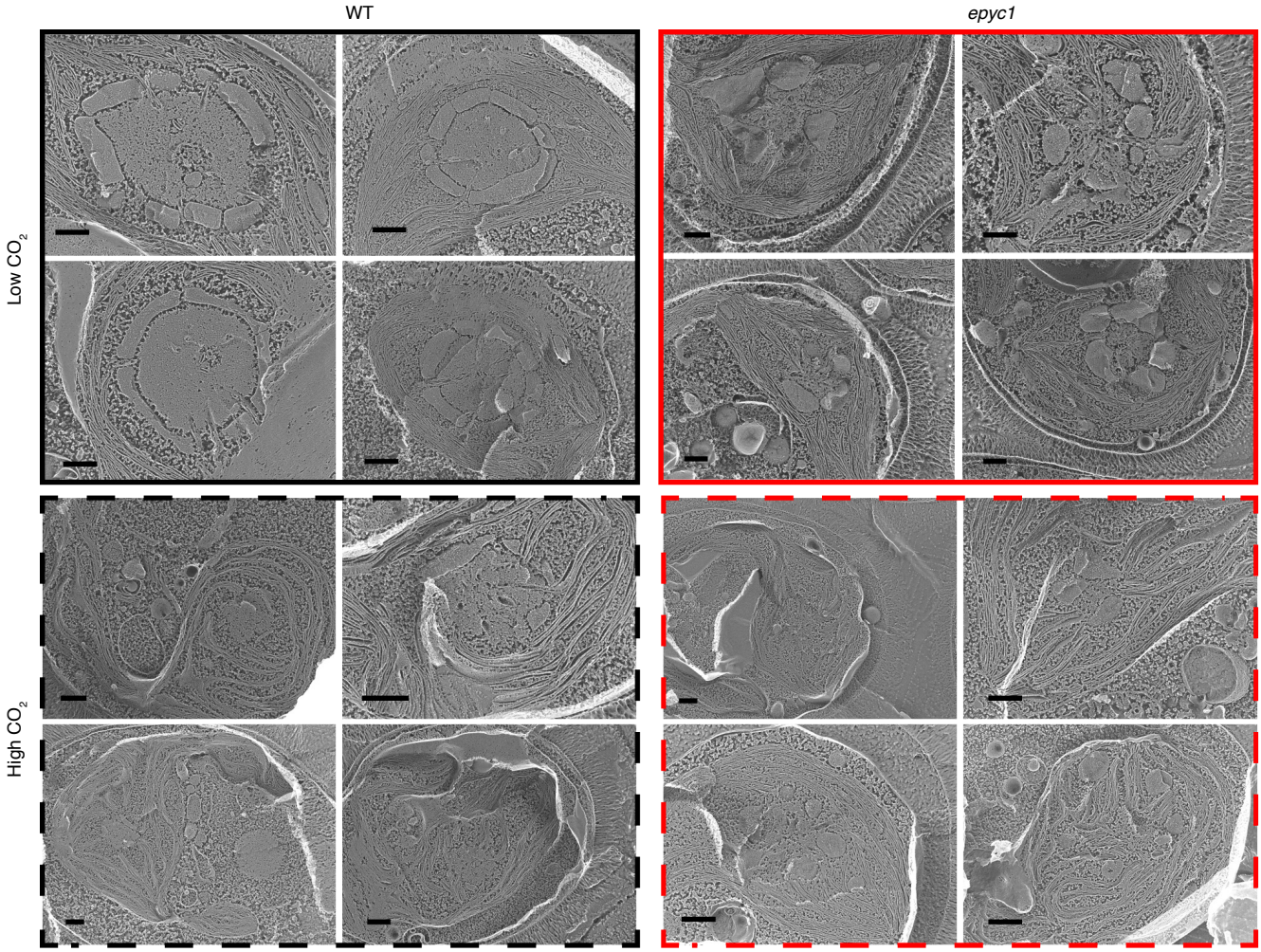


Fig. S5. Representative Quick-Freeze Deep-Etch EM (QFDEEM) images of WT and *epyc1* cells at low and high CO_2. Cells were prepared and imaged for QFDEEM as described in the materials and methods. (Scale bars: 500 nm.)

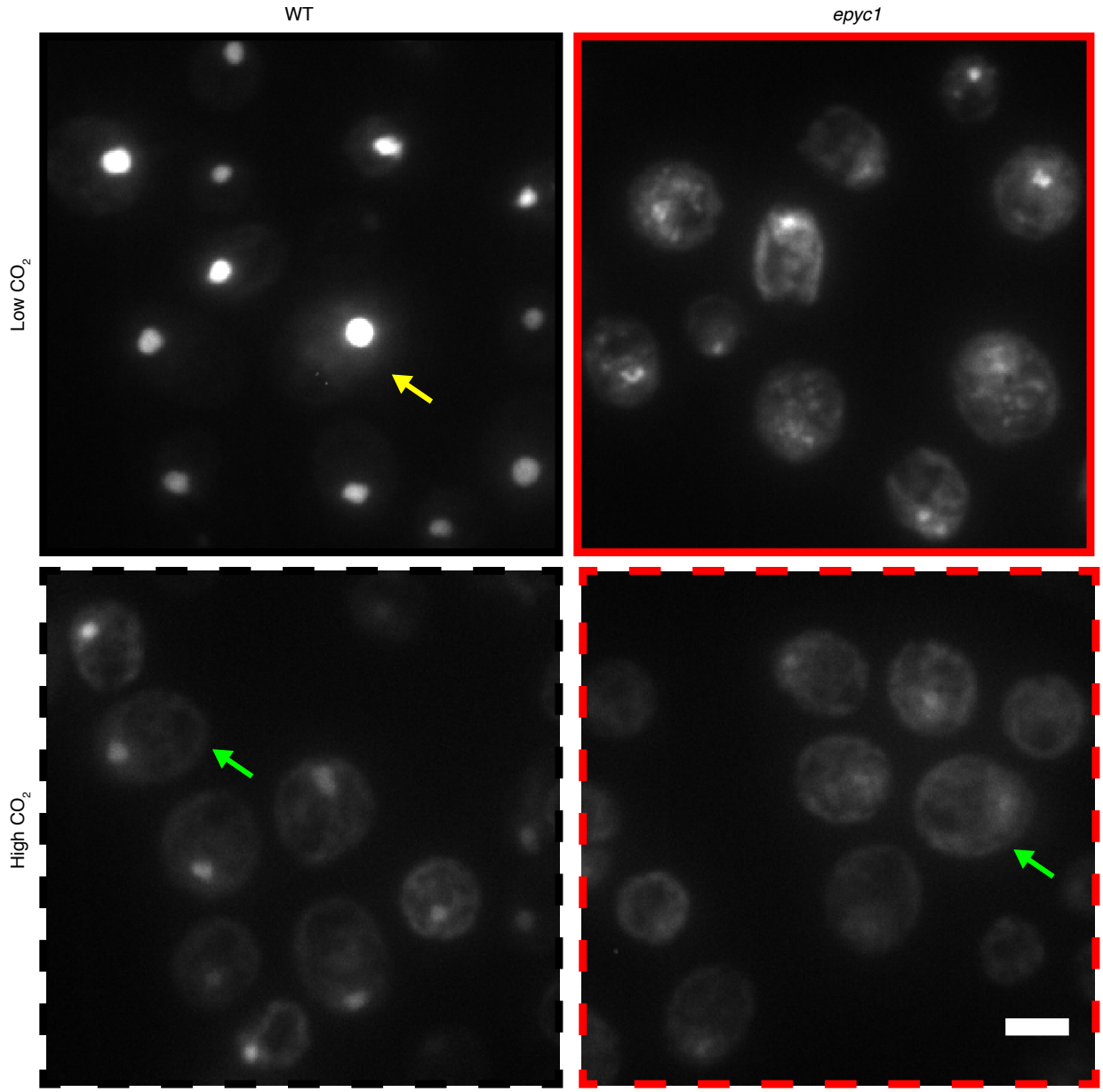


Fig. S6. Representative cells used for Rubisco-mCherry localization data. A representative field of view used for quantifying the mislocalization of Rubisco in the *epyc1* mutant. Images are summed z-stacks of 40 confocal sections 0.3 μm apart. Yellow arrow indicates the cell used for Fig. 3*E*. Green arrows indicate cells used for Fig. S3*E*. (Scale bar: 5 μm.)

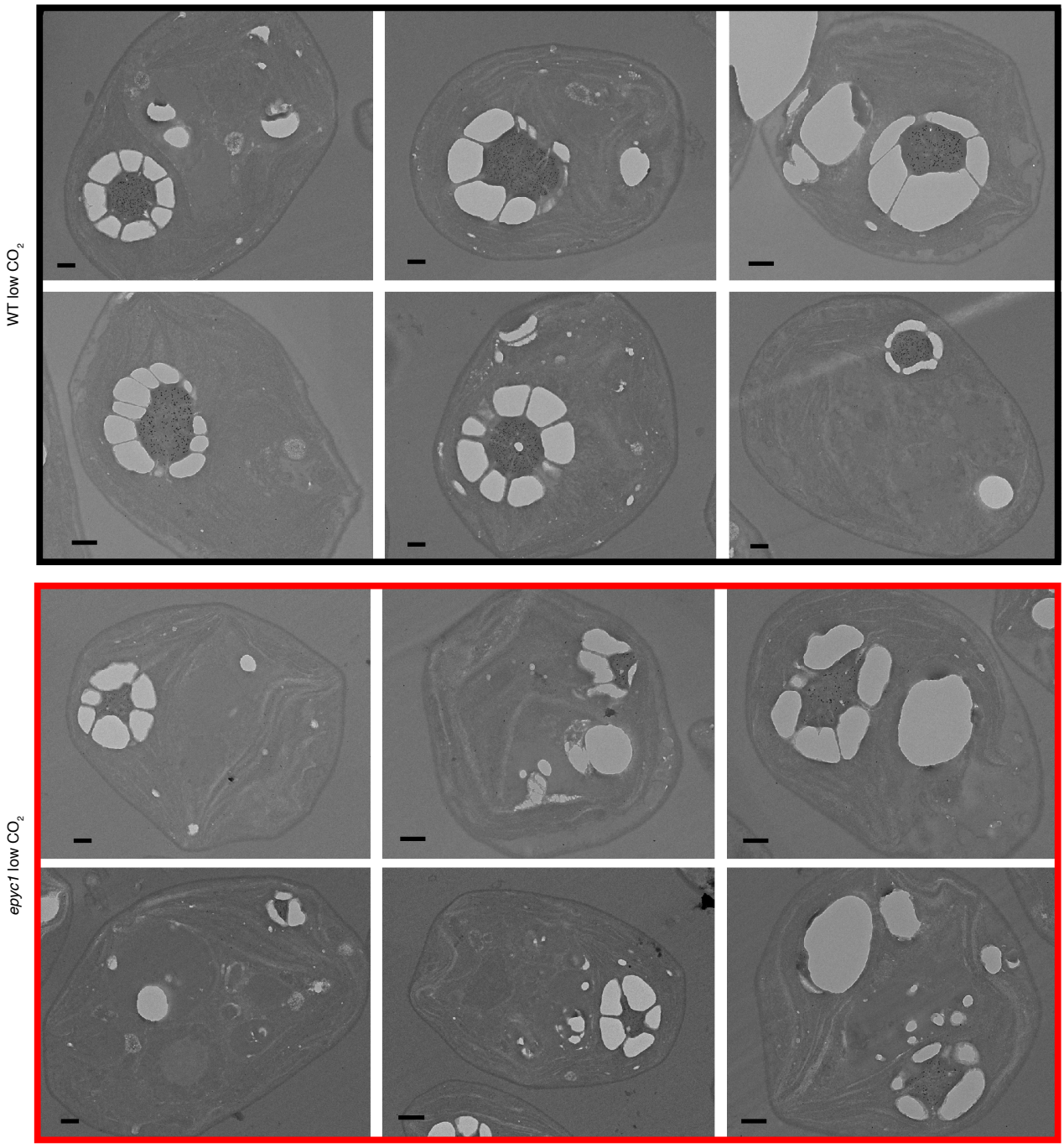


Fig. S7. Representative cells used for Rubisco immunogold labeling. A representative selection of immunogold-TEM images used for quantification of Rubisco levels outside the pyrenoid. The top left cell for each condition is the cell used in Fig. 3*G* before gold particle enlargement. Cells were prepared and imaged for immunogold-TEM as outlined in the materials and methods. (Scale bars: 500 nm.)

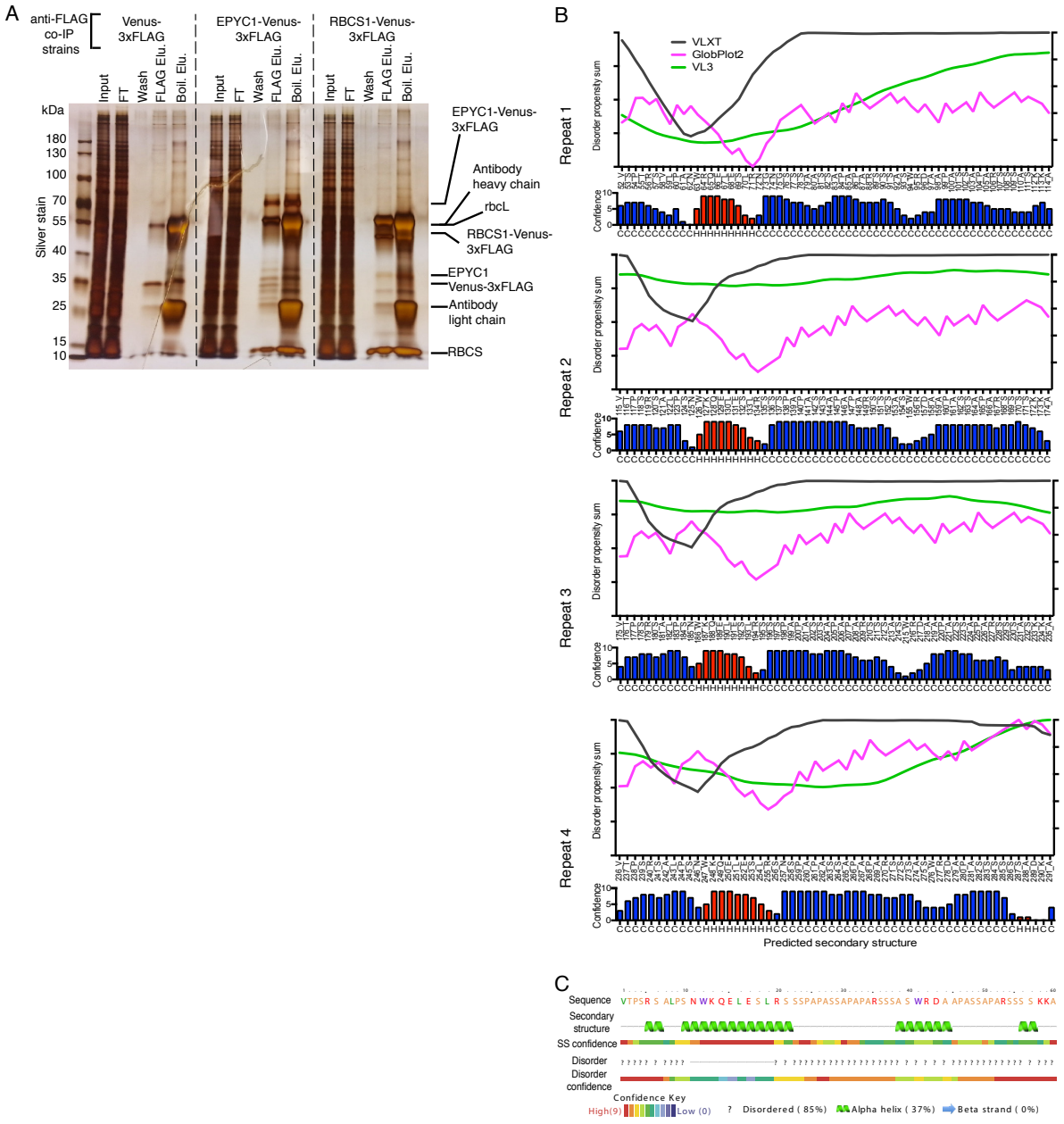


Fig. S8. EPYC1 interactions and sequence analysis. (*A*) Anti-FLAG co-immunoprecipitations (co-IPs) of WT cells expressing Venus-3xFLAG, EPYC1-Venus-3xFLAG and RBCS1-Venus-3xFLAG are shown. For each co-IP, the input, flow-through (FT), 4th wash (wash), 3xFLAG elution (FLAG Elu.) and boiling elution (Boil. Elu.) were run on an SDS-PAGE gel and silver stained. Right hand side labels show the expected sizes of proteins. (*B*) and (*C*) Analysis of the EPYC1 protein sequence. (*B*) To investigate the disorder of EPYC1, the full-length amino acid sequence was analyzed by VL3, VLTX and GlobPlot2 disorder prediction algorithms. The lower bar chart shows the PSIPRED v3.3 predicted secondary structure of full-length EPYC1, H = helix (red), C = coil (blue). Bar height indicates confidence value. (*C*) Analysis of the repeat region from 115-174 by Phyre2.

Table S1. The contribution of the pyrenoid to global net primary production (NPP)

	Percentage of ocean NPP	Percentage of algal group with a pyrenoid	Percentage of ocean NPP mediated by a pyrenoid	Percentage of total NPP
Global primary production				100%
Terrestrial				50 (13) - 54% (3)
Ocean				46 (3) - 50% (13)
Cyanobacteria (Prokaryotic)	10 (14) - 25% (69)			
Eukaryotic algae	75 (69) - 90% (14)			
Diatoms	42 (14, 69) - 50% (14)	100% (15)	42-50%	
Coccolithophores	17 (14, 69) -20% (14)	100% (16)	17-20%	
Chlorophytes	17 (14, 69) -20% (14)	10-90%*	2-18%	
Pyrenoid containing algae			61-88%	28 – 44%

*The majority of chlorophytes are known to have pyrenoids (17), with the pyrenoid containing *Micromonas pusilla* shown to be the dominant species in several oceanic and coastal regions (70). However, some ocean chlorophytes, including the abundant species *Bathycoccus prasinos,* appear to lack pyrenoids (17).

Table S2. Oligonucleotides used in this work

Primer name	Sequence
EPYC1_ORF_F	GCTACTCACAACAAGCCCAGTTATGGCCACTATCTCGTCGATGCGC
EPYC1_ORF_R	GAGCCACCCAGATCTCCGTTCAGGCCCTTGCGCCAGTCAGC
RBCS1_ORF_F	GCTACTCACAACAAGCCCAGTTATGGCCGCCGTCATTGCCAAGTC
RBCS1_ORF_R	GAGCCACCCAGATCTCCGTTCACGGAGCGCTTGTTGGCGGG
GBLP_F:	AACACCGTGACCGTCTCC
GBLP_R:	TGCTGGTGATGTTGAACTCG
EPYC1_F:	AAGCAGCTTGCCTAACCAGCAG
EPYC1_R:	ACATAACACACGCGTACCAAGGC
A1_F	GTTGGATGCACTAGTCACACGAGC
A2_F (EPYC1_Screen_pMJ016c_F)	GACGTTACAGCACACCCTTG
A1_R	GCACCAATCATGTCAAGCCT
E1_F	TCCTTCCGCACCAAAACATG
E2_F	CATAAGCTGTGAGCCGTTGA
E1_R	CAACTCAGTCAACGGCTCAC
E2_R (EPYC1_Screen_Gene_R)	ACAGTCGCATCAGAAAGGCT

Table S3. Raw qRT-PCR data

Gene	cDNA	Replica 1		Replica 2		Replica 3	
		Ct	Mean	Ct	Mean	Ct	Mean
Reference *RCK1/Cblp* (Cre06.g278222)	WT High CO_2	12.82	13.14	14.83	15.04	16.36	15.89
		13.3		15.35		15.6	
		13.31		14.93		15.7	
	epyc1 High CO_2	12.77	12.80	15.67	15.52	15.42	15.46
		12.82		15.41		15.38	
		12.81		15.48		15.58	
	WT Low CO_2	11.44	11.03	13.79	14.06	13.76	14.03
		11		14.48		14.53	
		10.65		13.92		13.81	
	epyc1 Low CO_2	11.28	11.54	14.55	14.67		14.66
		11.53		14.63		14.67	
		11.82		14.84		14.65	
EPYC1	WT High CO_2	18.28	18.52	19.56	19.63	21.87	21.80
		18.5		19.64		21.8	
		18.78		19.68		21.73	
	epyc1 High CO_2	22.19	22.43	23.74	23.62	24.64	24.81
		21.98		23.66		25.14	
		23.13		23.46		24.66	
	WT Low CO_2	13.42	13.30	14.74	14.75	17.08	17.22
		13.25		14.75		17.52	
		13.24		14.75		17.06	
	epyc1 Low CO_2	22.64	22.47	24.28	24.22	24.87	24.82
		22.34		23.96		24.62	
		22.44		24.43		24.98	

Table S4. Raw O_2 evolution data

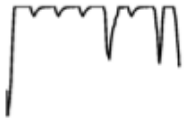
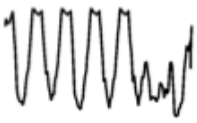
		Rate of O_2 evolution (µmol.mg $Chl^{-1}.h^{-1}$)										% of maximum													
		WT					*epyc*					WT							*epyc*						
	µM*Ci	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	mean	stdev	R1	R2	R3	R4	R5	mean	stdev
Low CO_2	2.5			5			3	17	11		2			3			3		2	11	9		2	6	5
	5	27	16	7	2	30	16		9	39	5	14	9	5	1	14	8	5	11		8		5	8	3
	10	63	53	57	70	39	24	22	6		9	32	29	39	45	18	33	10	17	15	5		10	12	5
	25	77	76	77	92	92	34		24	15	21	40	42	53	59	42	47	8	23		20	15	24	21	4
	50	150	117		138	119	25		32	13	18	77	65		89	54	71	15	17		27	13	21	20	6
	100	159	127	142	132	131	69	73	65	49	41	82	71	97	85	60	79	14	47	49	55	49	48	50	3
	250		170	147	133	129	74	79	80	68	66		95	100	86	59	85	18	50	54	68	68	76	63	11
	500	149		131	145	192	78	93	90	90	72	77		89	93	87	87	7	53	63	76	90	83	73	15
	1000	194	151	137	156	220	148	147	118	89	87	100	84	94	100	100	96	7	100	100	100	89	100	98	5
	2000		180	141	148	189		129		100	67		100	96	95	86	94	6		87		100	78	88	11
High CO_2	50	8	4	2			3	12	8			12	7	4			8	4	3	14	8			8	5
	100	31	7	8			38	23	15			44	12	13			23	18	36	26	17			26	10
	250	50	30	28			68	39	33			71	51	47			56	13	64	45	37			49	14
	500	57	45	52			80	67	69			79	76	88			81	6	76	76	77			76	0
	1000	71	55	60			100	74	82			100	94	100			98	4	95	84	91			90	6
	2000	69	58	55			106	88	90			97	100	92			96	4	100	100	100			100	0

*Ci: Inorganic carbon

Table S5. Quantification of number of cells with multiple pyrenoids

WT Low CO_2				*epyc1* Low CO_2			
Image name	Total cells scored	Cells with >1 pyrenoid	Cells with 1 pyrenoid	Image name	Total cells scored	Cells with >1 pyrenoid	Cells with 1 pyrenoid
2015Feb23Freq25	30	1	12	2015Feb23Freq25	26	0	13
2015Feb23Freq24	24	0	11	2015Feb23Freq24	22	1	8
2015Feb23Freq23	26	0	7	2015Feb23Freq23	21	0	5
2015Feb23Freq22	30	0	13	2015Feb23Freq22	28	1	8
2015Feb23Freq21	17	0	8	2015Feb23Freq21	21	0	10
2015Feb23Freq20	22	0	5	2015Feb23Freq20	26	1	8
2015Feb23Freq19	25	0	9	2015Feb23Freq19	31	3	14
2015Feb23Freq18	26	2	10	2015Feb23Freq18	29	2	12
2015Feb23Freq17	25	1	7	2015Feb23Freq17	23	1	12
2015Feb23Freq16	28	0	12	2015Feb23Freq16	26	0	9
2015Feb23Freq15	23	0	11	2015Feb23Freq15	21	0	6
2015Feb23Freq14	20	0	10	2015Feb23Freq14	25	0	4
2015Feb23Freq13	27	0	15	2015Feb23Freq13	20	0	6
2015Feb23Freq12	23	0	8	2015Feb23Freq12	25	2	8
2015Feb23Freq11	22	1	12	2015Feb23Freq11	23	2	10
2015Feb23Freq10	24	0	10	2015Feb23Freq10	22	0	9
2015Feb23Freq9	22	0	9	2015Feb23Freq9	30	4	15
2015Feb23Freq8	26	0	10	2015Feb23Freq8	25	0	7
2015Feb23Freq7	24	0	6	2015Feb23Freq7	24	1	8
2015Feb23Freq6	30	0	17	2015Feb23Freq6	29	1	11
2015Feb23Freq5	22	0	7	2015Feb23Freq5	34	1	10
2015Feb23Freq4	21	1	6	2015Feb23Freq4	30	3	11
2015Feb23Freq3	24	0	12	2015Feb23Freq3	26	1	9
2015Feb23Freq2	37	1	12	2015Feb23Freq2	29	3	8
2015Feb23Freq1	25	1	13	2015Feb23Freq1	18	2	10
TOTAL	623	8	252	TOTAL	634	29	231
% of cells with multiple pyrenoids			3.2%	% of cells with multiple pyrenoids			12.6%

Table S6. Analysis of pyrenoid positive and pyrenoid negative algae for proteins with EPYC1-like physicochemical properties

		Number of proteins with…					Protein characteristics					
Species (Phylum)	Pyrenoid	…>=3 repeats with a 40-80aa repeat length…	…and a pI >8…	…and an oscillating disorder profile…	…and no transmembrane domains	Uniprot or phytozome protein ID	Length	Repeat length	Repeat copy #	pI	Consensus repeat sequence from Xstream	Disorder profile*
Chlamydomonas reinhardtii (Chlorophyta)	Y	18	8	1	1	Cre10.g436550 (EPYC1)	318	61	3.84	11.8	VTPSRSALPSNWKQELESLRSSSPAPASSAPAPARSSSASWRDAAPASSAPARSSSASKKA	
Thalassiosira pseudonana (Heterokontophyta)	Y	4	1	1	1	B8CF53_THAPS	376	53	6.21	9.1	LSSKPSSAPFVRSEKPSSAPSDSPSASVAPTLETSFSPSSSGQPSPMTSESPS	
Phaeodactylum tricornutum (Heterokontophyta)	Y	12	1	1	1	B7GDW7_PHATC	380	46	7.17	9.9	TGPSMTGPSDSDDRRLRSPSSTGPSLTGPSMTGPSATGPSMTGPSM	
Emiliania huxleyi (Haptophyta)	Y	99	10	2	2	R1G412_EMIHU	353	70	4.10	12.1	PYLPISPARLARGSTSPHLSPSLPISPHISRTARSRFHIAPSLPISPHISPTAPHGFHEAPHLPISPHLS	
						R1D601_EMIHU	255	60	3.70	10.1	WTAADDALVKAGQEAGESWVDIAKRLPGRSADSVKSRSNRLKRQPDTSVKHEPVKRELVR	
Ostreococcus tauri (Chlorophyta)	Y/N†	3	3	2	2	A0A096PAN3_OSTTA	407	63	5.02	11.1	MAASKLGSKNASTRPTVGSTLDASALTPPSLRFTTENNIHSVPTAFGVADRPASRRVLRREDA	
						A0A090M8K8_OSTTA	470	63	6.02	11.2	MAASKLGSKNASTRPTVGSTLDASALTPPSLRFTTENNIHSVPTAFGVADRPASRRVLRREDA	
Micromonas pusilla (Chlorophyta)	Y	6	0	0	0							
Chlorella variabilis (Chlorophyta)	Y	3	2	1	0							
Chlorella prototheocoides (Chlorophyta)	N	1	0	0	0							
Cyanidioschyzon merolae (Rhodophyta)	N	0	0	0	0							
Galdieria sulphuraria (Rhodophyta)	N	2	0	0	0							
Nannochloropsis gaditana (Heterokontophyta)	N	3	0	0	0							

*Disordered profiles are a plot of disorder propensity (y axis; 0-1; 0 = ordered, 1 = disordered) against amino acid number (x-axis; 0-437). All profiles are on the same scale. †TEM images of *Ostreococcus tauri* show a singular starch deposit typical of a pyrenoid, however a Rubisco matrix has yet to be confirmed (71).