Supplementary Information

A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*

Alan K. Itakura[a,b1], Kher Xing Chan[c,1], Nicky Atkinson[d], Leif Pallesen[a], Gregory Reeves[a], Weronika Patena[a,f], Oliver Caspari[c], Robyn Roth[f], Ursula Goodenough[f], Alistair J. McCormick[d], Howard Griffiths[c,2], Martin C. Jonikas[a,2]

[1]A.K.I. and K.X.C. contributed equally to this work.

[2]To whom correspondence should be addressed.
Martin Jonikas (mjonikas@princeton.edu) or Howard Griffiths (hg230@cam.ac.uk)

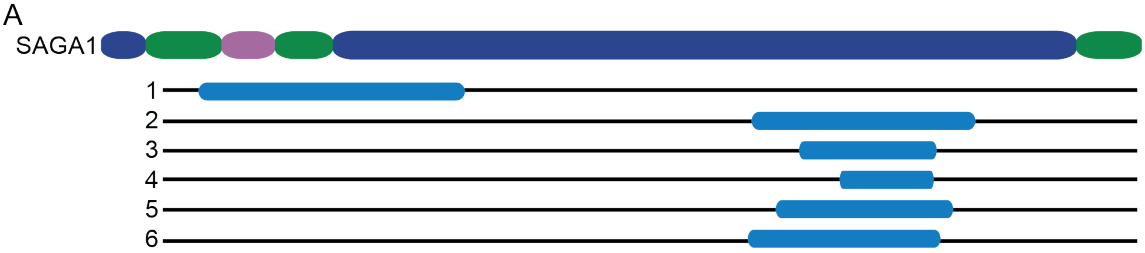
This PDF file includes:

- **Figs S1 to S13**
- **Tables S1 to S6**
- **SI Materials and Methods**
- **SI References**

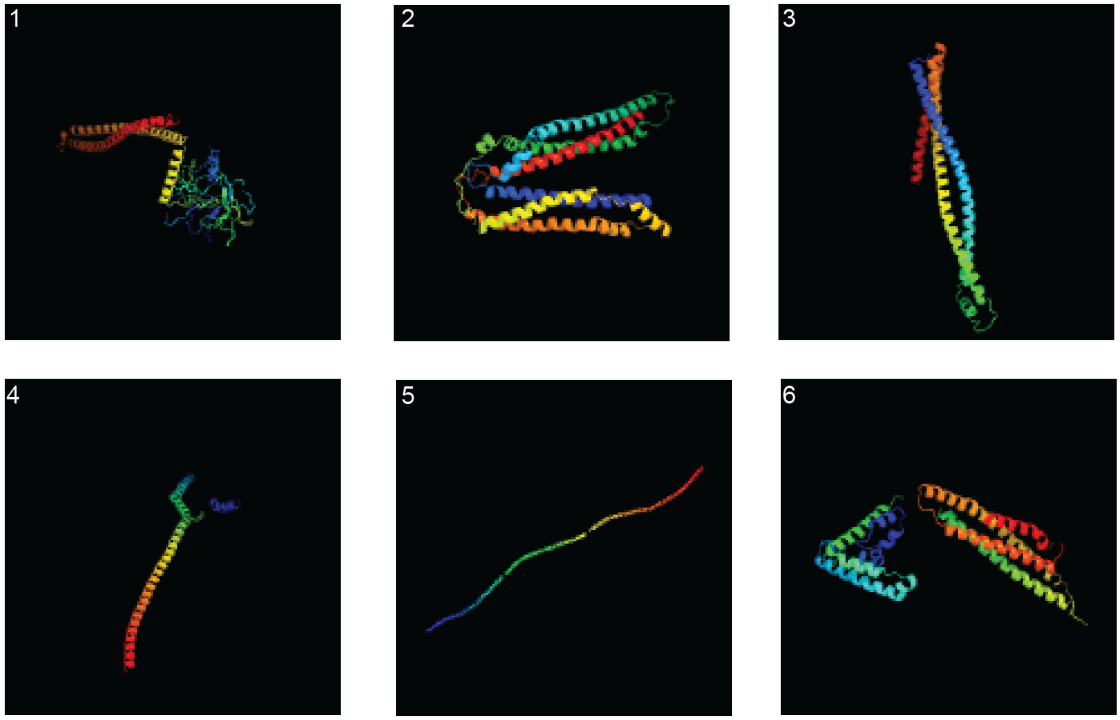
SI Figures

Fig. S1. SAGA1 contains a predicted starch binding domain. SAGA1 (query) was aligned to the conserved starch-binding site 1 of a range of proteins from several species (CDD, NCBI (52, 53); CBM20; E-value = 1.48e^{-08}). This site is suggested to act as an initial starch recognition site. 1CYG_A: Chain A, cyclodextrin glucanotransferase (E.C.2.4.1.19); 1DTU_A: Chain A, *Bacillus circulans* strain 251 cyclodextrin glycosyltransferase; 1ACZ_A: Chain A, glucoamylase; gi 23127960: Lysophospholipase L1 and related esterases [*Nostoc punctiforme* PCC73102]; gi 67926159 Glycoside hydrolase, starch-binding [*Crocosphaera watsonii* WH 8501]; gi 17227665: cyclomaltodextrin glucanotransferase [*Nostoc* sp. PCC 7120]; gi 87123854: Glycoside hydrolase, starch-binding [*Synechococcus* sp. RS9917]; gi 118364918 trehalose-6-phosphate synthase, putative [*Tetrahymena thermophila* SB210]; gi 145502819: hypothetical protein (macronuclear) [*Paramecium tetraurelia* strain d4-2]. Hash mark (#) with yellow highlights: amino acids involved in the starch recognition feature (Feature 1); Grey lower case: unaligned residues; Upper case: aligned residues used to generate PSSM (position-specific scoring matrix); Red to blue color scale: degree of conservation, with red designating highly conserved.

A



B



C

Panel	PDB Template	PDB Header	PDB Molecule	PDB Title	Confidence (%)	Identity (%)	Query range(aa)
1	c1jchC_	ribosome inhibitor, hydrolase	colicin e3	crystal structure of colicin e3 in complex with its immunity protein	98.6	18	36-450
2	c2oevA_	protein transport	programmed cell death 6-interacting protein	crystal structure of alix/aip1	98.4	10	893-1228
3	c4cgkA_	cell cycle	secreted 45 kda protein	crystal structure of the essential protein pcsb from streptococcus2 pneumoniae	98.3	11	950-1157
4	c3ojaB_	protein binding	anopheles plasmodium-responsive leucine-rich repeat protein	crystal structure of lrim1/apl1c complex	98.3	10	1024-1167
5	c1c1gA_	contractile protein	tropomyosin	crystal structure of tropomyosin at 7 angstroms resolution2 in the spermine-induced crystal form	98.2	16	920-1198
6	c1yvlB_	signalling protein	signal transducer and activator of transcription	structure of unphosphorylated stat1	98	9	875-1170

Fig. S2. SAGA1 is predicted to have structural homology to proteins with long alpha helical regions. (*A*) Using Phyre2, the last 1500 amino acids of SAGA1 (due to sequence length

submission limitation) were run in intensive mode. Shown is a schema of the top 6 structural homology results. In blue are the regions of SAGA1 that showed homology to PDB (Protein DataBase) templates. (*B*) The tertiary structure of the top 6 structural homology results are displayed. (*C*) Details of the top 6 structural homology results. PDB template refers to the template that SAGA1 was found to have structural homology with. Provided for each result is the PDB header, the PDB molecule, PDB title, the confidence (the probability that the match between SAGA1 and the PDB template is a true homology) and identity (coverage) of the query sequence that has this particular homology.

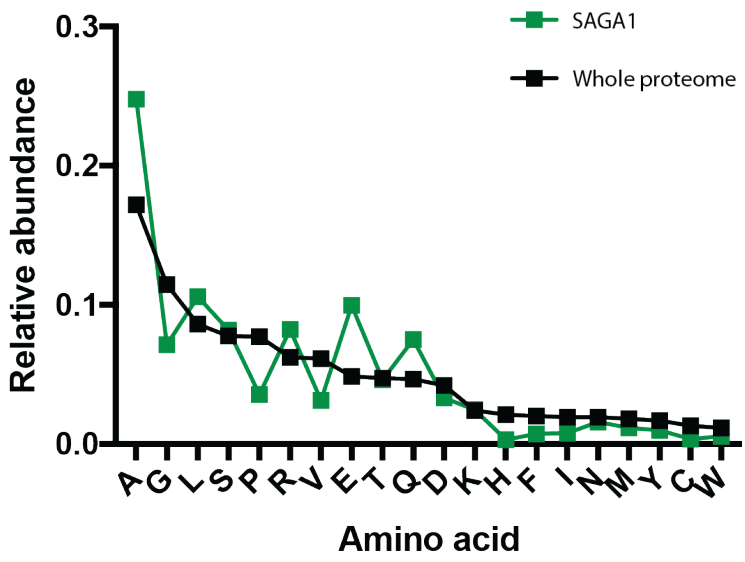


Fig. S3. SAGA1 has an unusual amino acid composition. The relative abundances of each amino acid of SAGA1 and the *Chlamydomonas* proteome were calculated as a fraction of the total number residues in each. Amino acids are ordered from highest to lowest abundance in the *Chlamydomonas* proteome.

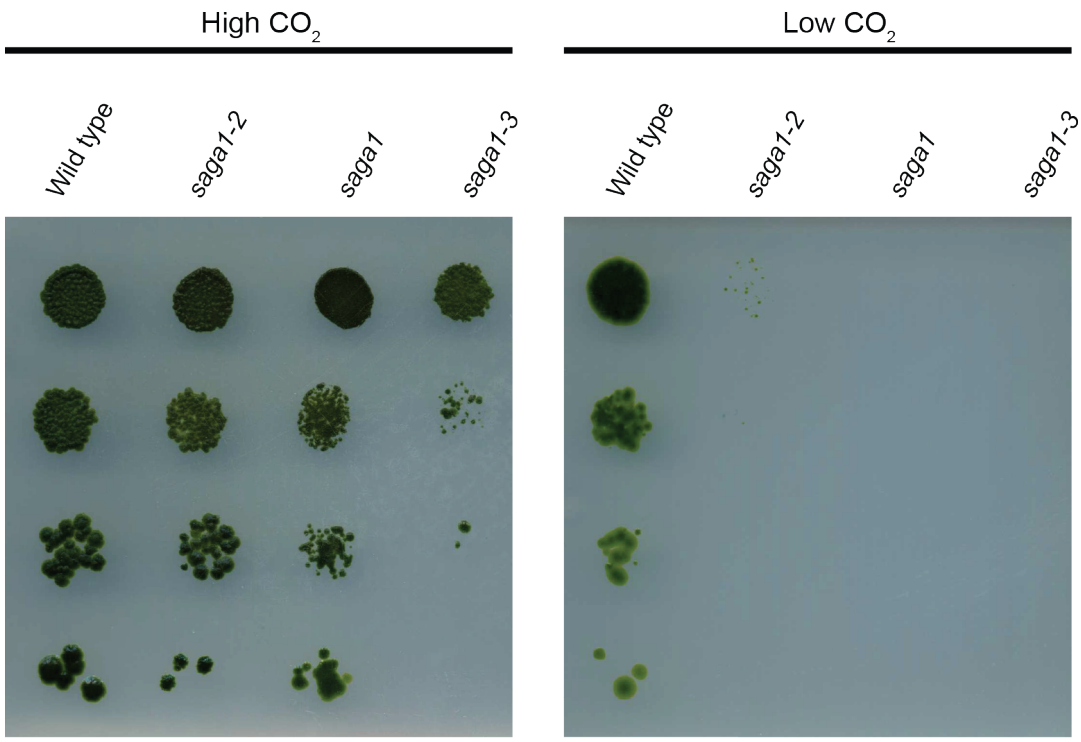


Fig. S4. Other mutant alleles of *saga1* also exhibit a growth defect in low CO_2. Serial 1:10 dilutions of wild type and 3 independent mutant alleles of *saga1* were spotted on TP minimal medium and grown at high and low CO_2 under 500 μmol photos $m^{-2}s^{-1}$ illumination.

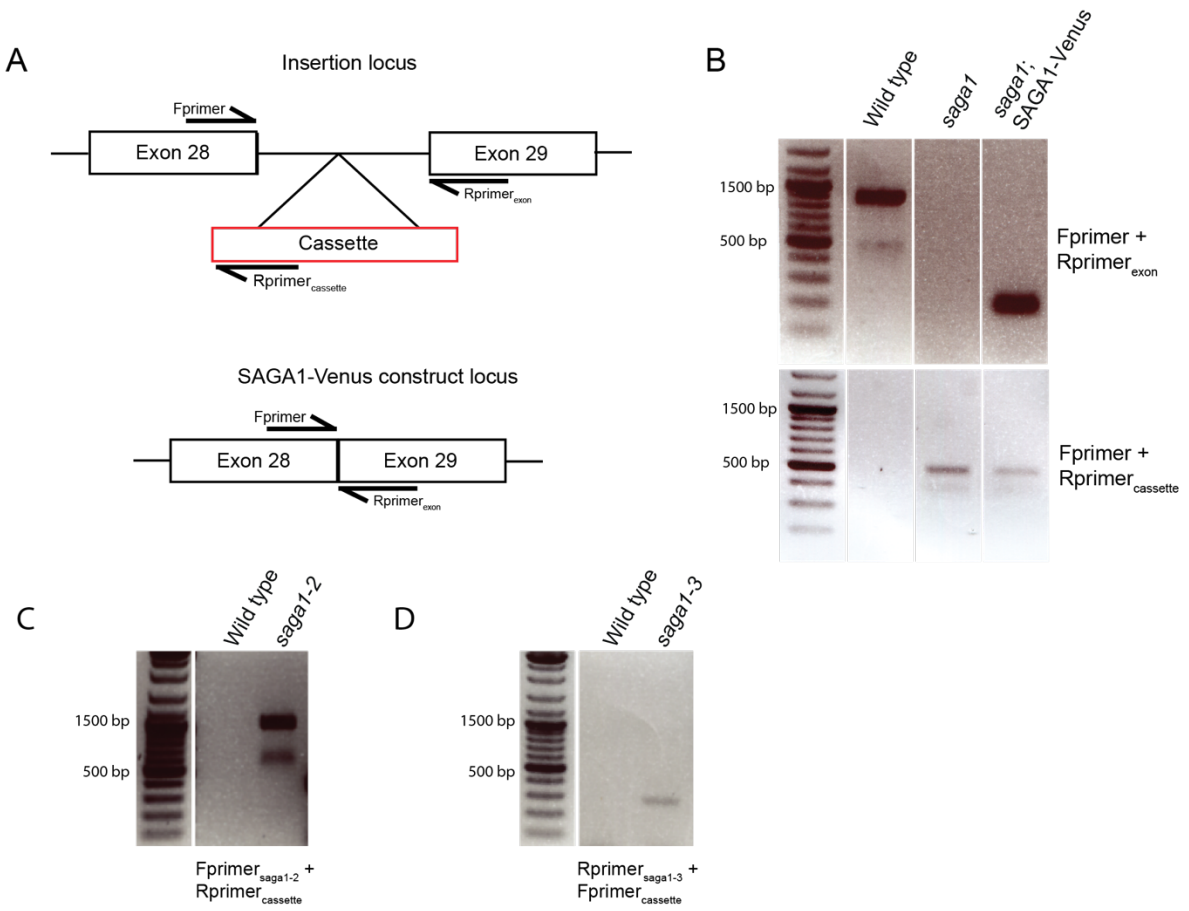
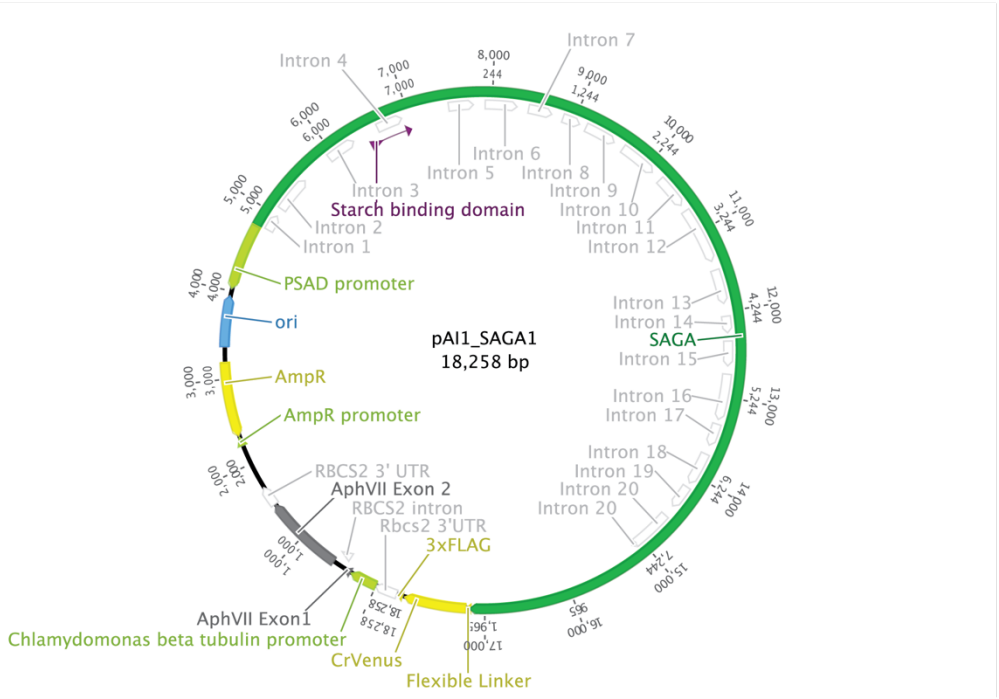


Fig. S5. The complemented *saga1* mutant contains both the insertion cassette and the SAGA1-Venus construct. (*A*) Schema depicting the PCR confirmation strategy of the *saga1* mutant and *saga1;SAGA1-Venus* complement. (*B*) The Fprimer and $Rprimer_{exon}$ yielded a 900 bp product in the wild type but did not produce a product in the *saga1* mutant due to the presence of the insertion cassette. Fprimer and $Rprimer_{cassette}$ yielded a 500 bp product in the *saga1* mutant and the *saga1;SAGA1-Venus* strain due to the presence of the insertion cassette. In the *saga1;SAGA1-Venus* strain, Fprimer and $Rprimer_{exon}$ yielded a 250 bp product, smaller than in the wild type because the complementation cassette lacks intron 28 (Fig. S13). (*C*) PCR confirming the presence and location of the insertion cassette in the *saga1-2* mutant. $Fprimer_{saga1\text{-}2}$ is upstream of the insertion cassette. The double bands indicate the presence of tandem cassette. The *saga1-2* insertion was mapped to intron 28. (*D*) PCR confirming the presence and location of the insertion cassette in the *saga1-3* mutant. $Rprimer_{saga1\text{-}3}$ is downstream of the insertion cassette. $Fprimer_{cassette}$ lies at the 3' end of the cassette. The insertion was mapped to intron 25. Primer sequences can be found in Table S6.



72

73 **Fig. S6**. The pRAM118-SAGA1 construct. The *SAGA1* gene includes the first 20 introns; the
74 remaining introns were omitted during gene synthesis. *SAGA1* is driven by a *PSAD* promoter
75 and is followed by a *CrVenus* tag and a *3xFLAG* tag. The construct contains *AmpR* cassette for
76 ampicillin resistance in *E. coli* and an *AphVII* cassette for hygromycin selection in *C.*
77 *reinhardtii*. The construct is derived from pLM005, but with an *AphVII* for hygromycin
78 resistance instead of *AphIII* for paromomycin resistance. The construct and sequence can be
79 found at the Chlamydomonas Resource Center as pRAM118_SAGA1_Venus_3xFLAG.

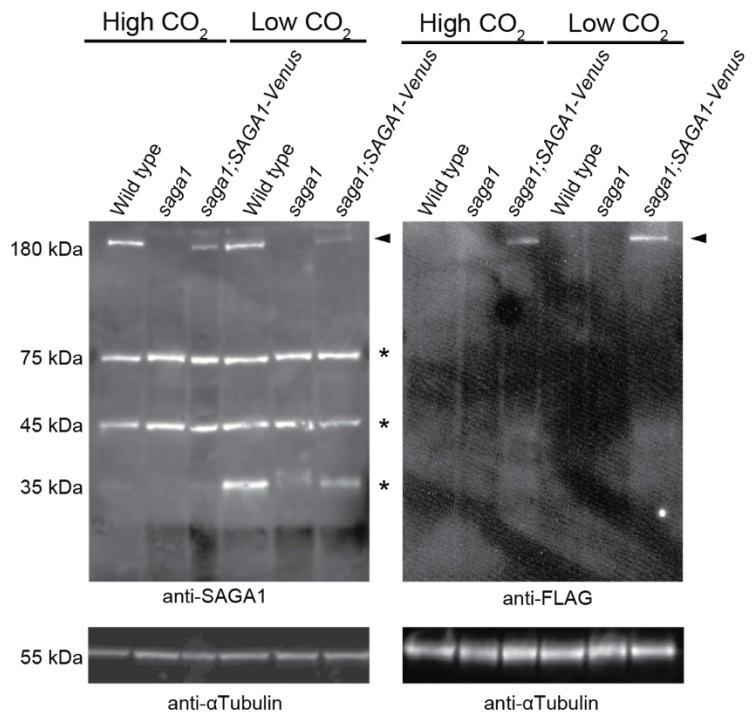


Fig. S7. Full membranes from Fig. 1D. SAGA1 protein levels in wild-type, *saga1*, and *saga1*;*SAGA1-Venus* cells grown at low and high CO_2 were probed with an anti-SAGA1 polyclonal antibody and with an anti-FLAG antibody. Arrow indicates the *SAGA1* protein product. The asterisks indicate non-specific bands. Anti-tubulin is shown as a loading control.

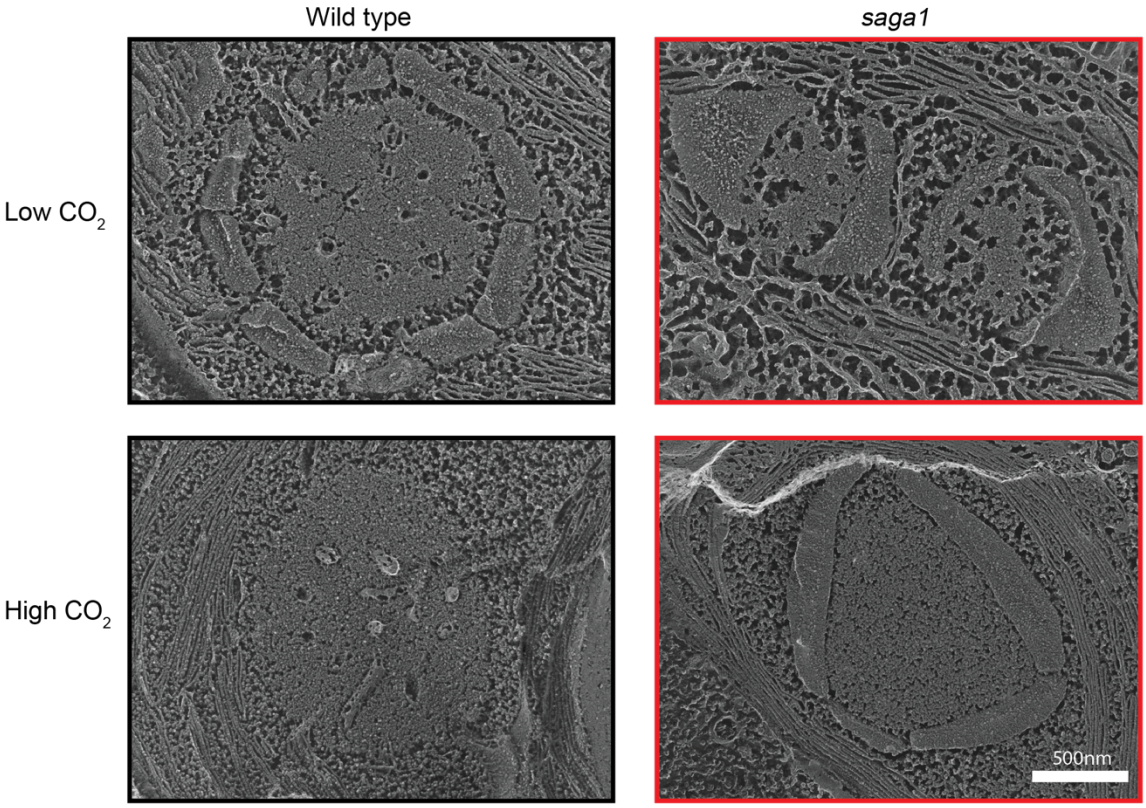


Fig. S8. Quick-freeze deep-etch cryo-electron microscopy reveals abnormal pyrenoid structure in the *saga1* mutant. Representative quick-freeze deep-etch cryo-electron microscopy images of cells grown in high and low CO_2 highlighting multiple pyrenoids, abnormal starch sheaths, and a decreased number of thylakoid tubules in the *saga1* mutant. Samples were fixed with glutaraldehyde prior to electron microscopy. Scale bar = 500nm.

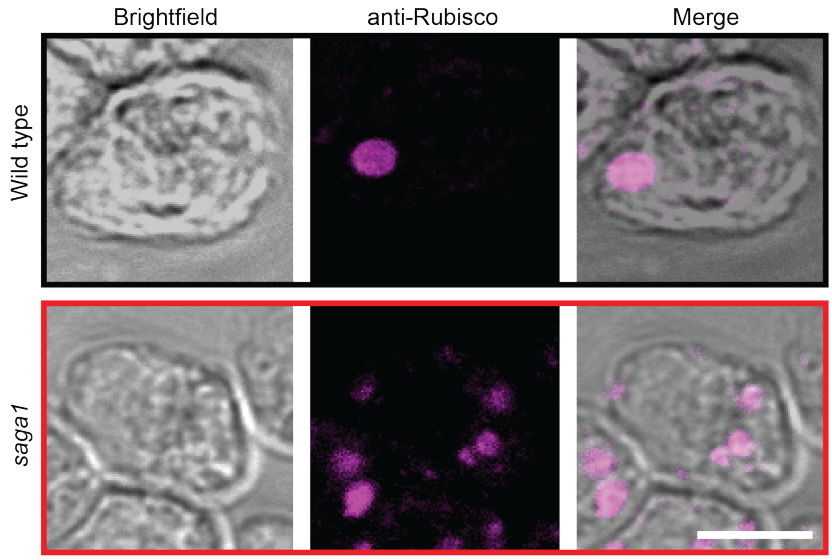


Fig. S9. Rubisco is localized to the pyrenoids of wild-type and *saga1* mutant *Chlamydomonas reinhardtii* cells grown in low CO_2. Subcellular localization of Rubisco in wild-type and *saga1* mutant cells shown by indirect immunofluorescence assay using an anti-Rubisco antibody. Multiple Rubisco localization sites could be observed in each *saga1* mutant cell, in contrast to a single site in each wild-type cell. Scale bar = 5 μm.

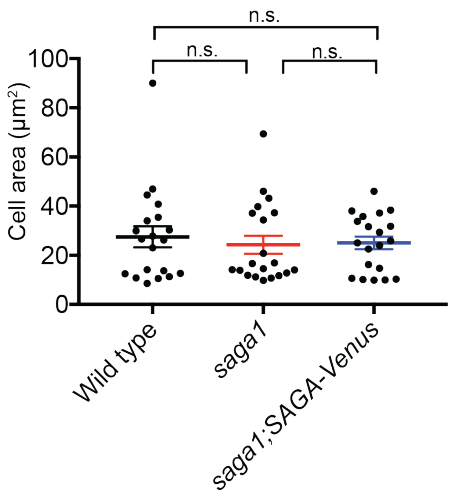


Fig. S10. *saga1* mutant cells are of similar size to wild-type cells. The areas of wild-type, *saga1*, and *saga1*;*SAGA1-Venus* cells imaged using TEM were quantified (N=10 cells; Mann Whitney *U* test; n.s., not significant). Error bars: SEM.

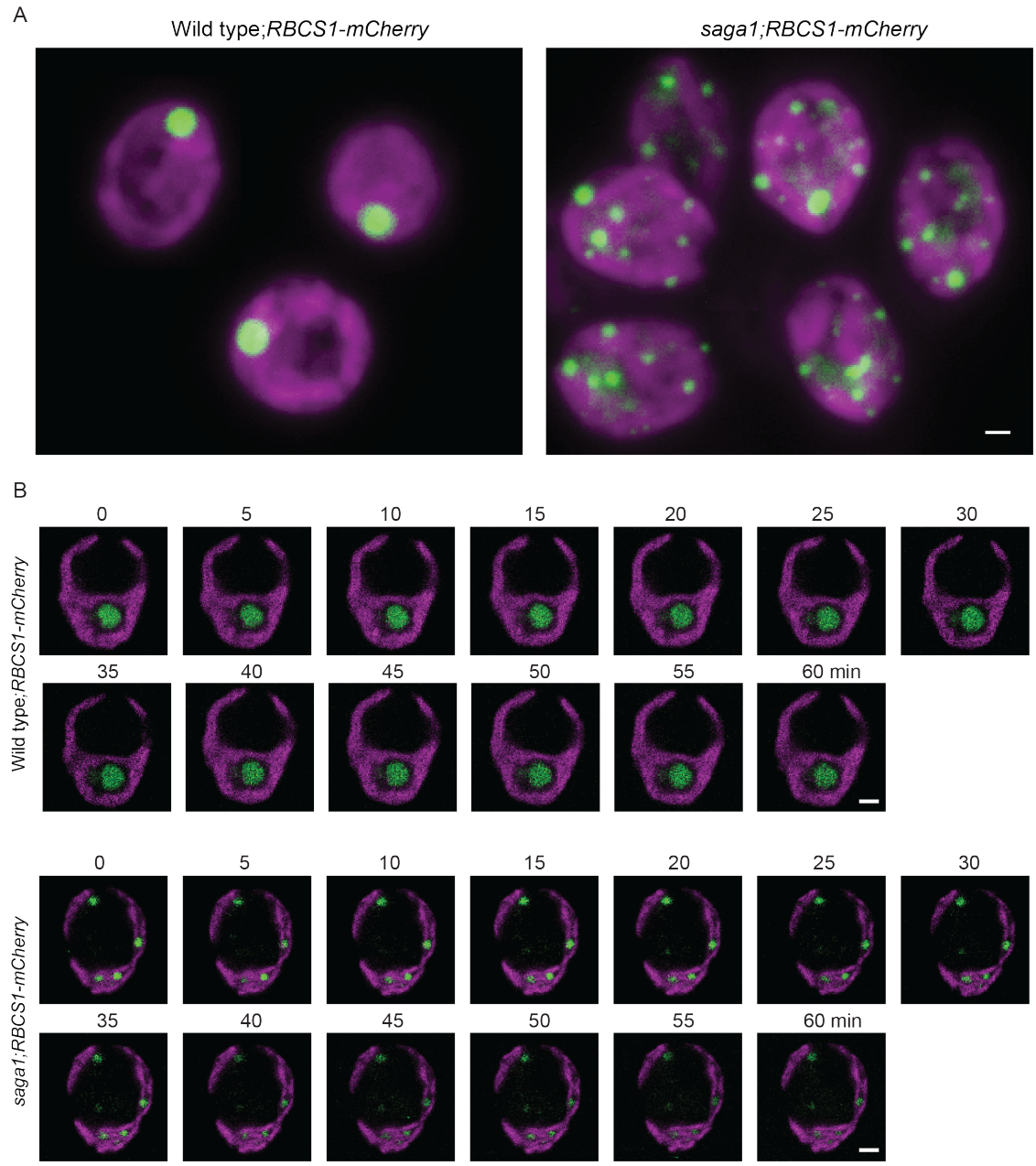
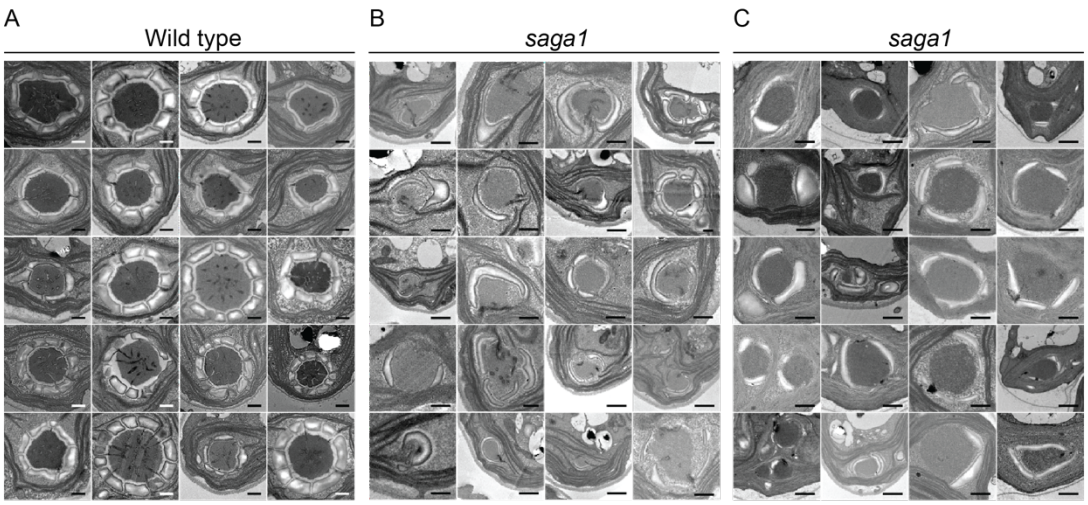


Fig. S11. The *saga1* mutant has multiple stable pyrenoids. (*A*) Representative summed z-stacks of a field of *saga1* and wild-type cells constitutively expressing RBCS1-mCherry grown in low CO_2. One cell from each field view is also shown in Fig. 3. (*B*) A 1-hour time course of *saga1* and wild-type cells expressing RBCS1-mCherry. Green is RBCS1-mCherry fluorescence and magenta is chlorophyll autofluorescence. Scale bar = 1μm.



107

108 **Fig. S12.** Pyrenoids in the *saga1* mutant occasionally contain thylakoid tubules. (*A*) TEM
109 images of wild-type pyrenoids. (*B*) TEM images of *saga1* mutant pyrenoids with thylakoid
110 tubules. (*C*) TEM images of *saga1* mutant pyrenoids without visible thylakoid tubules. Scale
111 bar = 500nm.

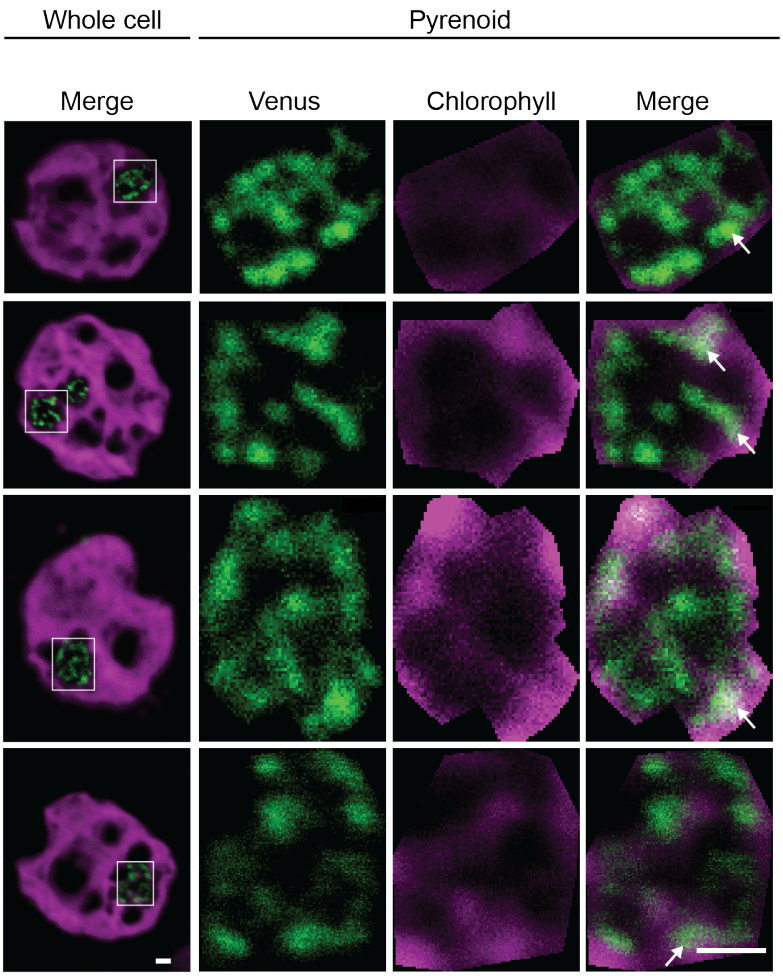


Fig. S13. SAGA1 partially colocalizes with chlorophyll autofluorescence in the pyrenoid. Representative confocal fluorescent microscopy images of SAGA1-Venus (green) constitutively expressed in *saga1* mutant cells grown in low CO_2. Magenta is chlorophyll autofluorescence. White arrows indicate white coloration arising from overlap of SAGA1-Venus and chlorophyll autofluorescence. Spearman rank rho values calculated from pixel intensities of the SAGA1-Venus fluorescence and the chlorophyll autofluorescence are listed in Table S5. Left: scale bar = 1 μm. Right: scale bar = 500nm.

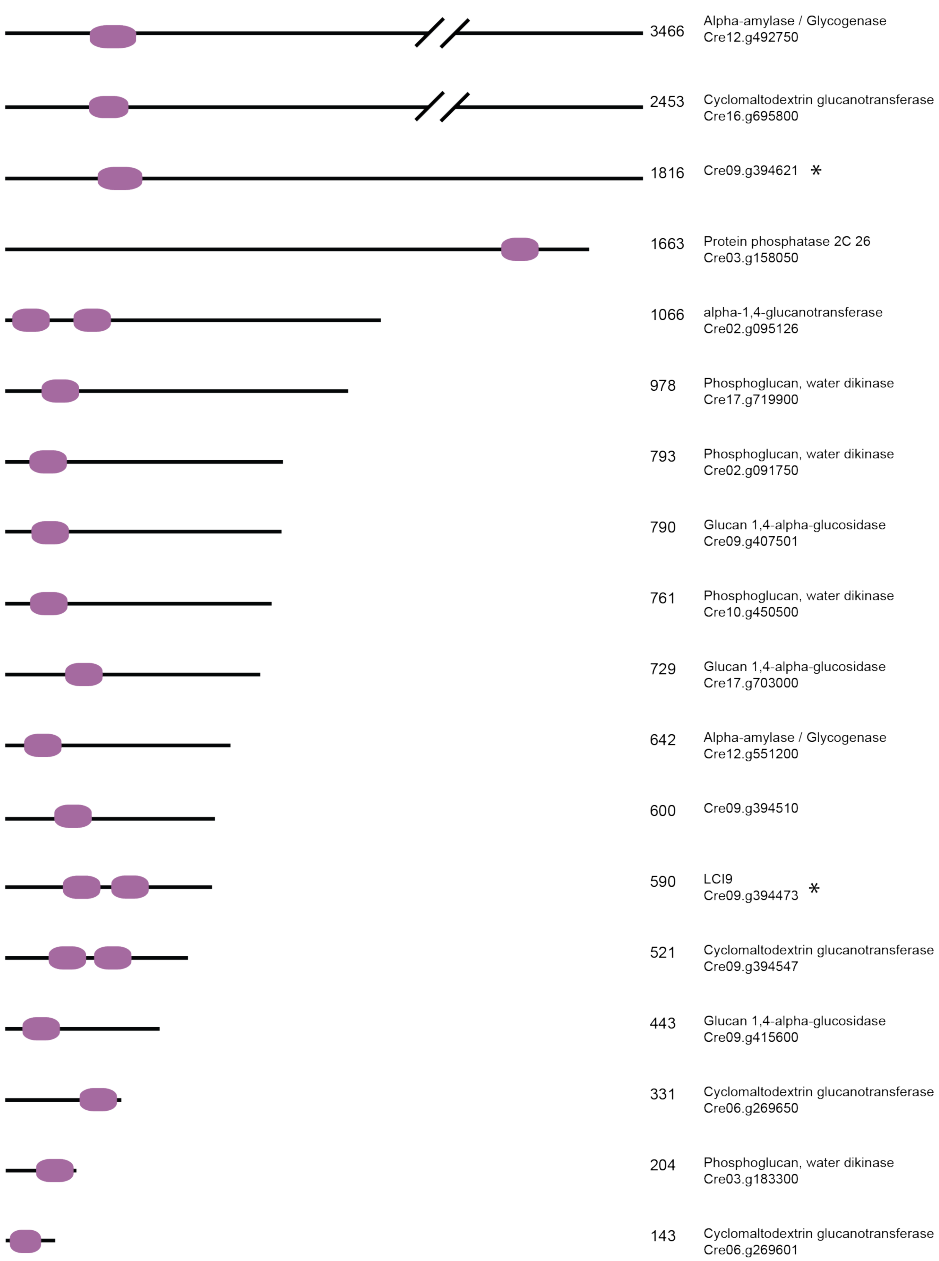


Fig. S14. *Chlamydomonas reinhardtii* proteins with putative starch binding domains. A cartoon depiction of the 18 *C. reinhardtii* proteins that have the CBM_20 motif (PFAM00686; Phytozome). Length of the protein is relative to Cre09.g394621. Included is the gene name and the description of the gene. The asterisks indicate proteins that interacted with components of the pyrenoid matrix by immunoprecipitation-mass spectrometry (26)

SI Tables

Table S1. Protein BLAST results using the SAGA1 starch binding domain (residues 212-280) as a query sequence.

Species	Description	Uniprot ID	Gene ID	Starch binding domain	Coiled-coil	Length	Query sequence	Identity	E-value	Score
Chlamydomonas reinhardtii	Uncharacterized protein	A0A2K3D7T6	CHLRE_11g467712v5	178-308	678	1626	SAGA1 SBD	100.00%	9.60E-47	389
Tetrabaena socialis	Uncharacterized protein	A0A2J7ZZR0	TSOC_007957	65-171	562	1462	SAGA1 SBD	59.40%	1.40E-19	207
Volvox carteri f. nagariensis	Uncharacterized protein	D8TYS9	VOLCADRAFT_105158	98-206, 232-354	403	1552	SAGA1 SBD	41.30%	2.90E-09	138
Gonium pectorale	Uncharacterized protein	A0A150GJ96	GPECTOR_19g330	78-203, 226-334	974	3,273	SAGA1 SBD	45.50%	6.40E-08	129
Gonium pectorale	Uncharacterized protein	A0A150G832	GPECTOR_49g527	14-117	167	672	SAGA1 SBD	39.10%	1.20E-07	127
Volvox carteri f. nagariensis	Uncharacterized protein	D8TPI1	VOLCADRAFT_88626	186-296, 325-432, 509-617	694	2801	SAGA1 SBD	41.50%	5.00E-07	123
Raphidocelis subcapitata	Uncharacterized protein	A0A2V0PDF5	Rsub_08056	42-143	449	1201	SAGA1 SBD	46.20%	7.00E-07	122
Chlamydomonas reinhardtii	alpha amylase?	A0A2K3D1W5	CHLRE_12g492750v5	210-323, 344-468, 491-599	1306	3466	SAGA1 SBD	42.20%	7.10E-07	122
Chlamydomonas reinhardtii	Uncharacterized protein	A0A2K3DEF8	CHLRE_09g394621v5	96-206, 221-343	410	1748	SAGA1 SBD	43.10%	9.9E-7	121
Gracilariopsis chorda	Kinesin-like protein KIN-14R	A0A2V3IYN9	BWQ96_03906	204-325, 669-996	99	1357	SAGA1 SBD	56.40%	2.8E-6	118
Chlorella variabilis	Uncharacterized protein	E1ZG60	CHLNCDRAFT_52636	119-233	49	541	SAGA1 SBD	45.30%	3.70E-06	117

Top 10 results from a Protein BLAST using the Uniprot BLAST portal (unitprot.org). Target database was Eukaryota. Predicted starch binding domains (CBM20) were identified using Interpro; listed are the corresponding a.a. residues. The total length of coiled-coil secondary structures was determined using UniProt's Automatic Annotation pipeline. Identity is a measurement of how similar the query sequence is to the target sequence. E-value is a statistical measure to estimate the number of expected matches in a random database. Score is a normalized metric for how

similar the sequences are, independent of sequence length and database size. E-Threshold was set at 10. BLOSUM62 matrix was used for alignment. No filtering was used. Gaps were permitted.

Table S2. Protein BLAST results using SAGA1 as a query sequence.

Species	Description	Uniprot ID	Gene ID	Starch binding domain	Coiled-coil (total a.a.)	Length (a.a.)	Query sequence	Identity	E-value	Score
Chlamydomonas reinhardtii	Uncharacterized protein	A0A2K3D7T6	CHLRE_11g467712v5	178-308	678	1,626	SAGA1	100.00%	0	7,807
Tetrabaena socialis	Uncharacterized protein	A0A2J7ZZR0	TSOC_007957	65-171	562	1,462	SAGA1	54.50%	0	3,603
Chlamydomonas reinhardtii	Uncharacterized protein	A8J9W7	CHLREDRAFT_151809	N/A	373	577	SAGA1	78.60%	0	1,992
Gonium pectorale	Uncharacterized protein	A0A150FZL5	GPECTOR_106g128	N/A	507	1,080	SAGA1	46.10%	0	1,746
*Volvox carteri**	Mitotic checkpoint protein MAD1	N/A	Vocar.0009s0363	192-305	~1000	1,639	SAGA1	37.60%	1.50E-97	1414
Chlorella sorokiniana	TPR repeat-containing protein	A0A2P6TEU7	C2E21_8179	N/A	343	4,188	SAGA1	27.10%	2.80E-82	801
Porphyra umbilicalis	Uncharacterized protein	A0A1X6PGC9	BU14_0071s0060	N/A	154	2,312	SAGA1	27.60%	7.10E-82	796
Chlorella sorokiniana	Kinesin K39 isoform A	A0A2P6TPQ7	C2E21_5045	N/A	956	2,826	SAGA1	27.70%	8.80E-82	796
Chlamydomonas reinhardtii	Uncharacterized protein	A0A2K3DKA0	CHLRE_07g336750v5	N/A	2450	3,869	SAGA1	28.50%	2.30E-80	785
Emiliania huxleyi	Uncharacterized protein	R1EUH6	EMIHUDRAFT_113389	N/A	1856	4,388	SAGA1	27.80%	3.20E-80	784
Gonium pectorale	Uncharacterized protein	A0A150G2C5	GPECTOR_75g720	N/A	1374	1,898	SAGA1	26.30%	1.30E-79	776

Top 10 results from a Protein BLAST using the Uniprot BLAST portal (unitprot.org). Target database was Eukaryota. Column titles are as in Table S1. * indicates that this result was identified using Phytozome (phytozome.jgi.doe.gov) Protein BLAST.

Table S3. Photosynthetic $K_{0.5}$ (Ci) of wild-type, *saga1* and *saga1;SAGA1-Venus-3xFLAG* cultures.

Sample	Conditions	$K_{0.5}$ (C_i) (μM)	p-value
Wild type	HC	115.8 ± 24.2	-
	LC	38.7 ± 11.8	-
saga1	HC	775.5 ± 88.7	1.0×10^{-5}
	LC	160.0 ± 19.1	0.00015
saga1;SAGA1-Venus-3xFLAG	HC	267.4 ± 84.3	0.20929
	LC	41.9 ± 26.3	0.91083

Data are shown as average ± standard error, which were obtained from at least three independent experiments. The paired t-test was used to determine the significance compared to wild type. C_i, inorganic carbon; HC, high CO_2; LC, low CO_2.

Table S4. Rubisco fraction in the pyrenoid or stroma of wild-type cells and the *saga1* mutant.

Sample	Average area of pyrenoid cross section (nm^2)	Average area of chloroplast cross section (nm^2)	Normalized immunogold particle count (particles nm^{-2}), N		Rubisco in pyrenoid (%)	Comparison to wild type (adjusted p-value)
			Pyrenoid, N_i	Stroma, N_j		
Wild type	9.25×10^6	120.11×10^6	140 ± 17.0	6 ± 1.3	93.8	-
saga1	7.91×10^6	113.21×10^6	50 ± 8.6	4 ± 0.8	90.3	n.s.

The percentage of Rubisco in the pyrenoid was calculated as described in the Supplementary Methods. The data shown are the means ± SEM.

Table S5. Spearman rank rho values calculated from pixel intensities of the SAGA1-Venus fluorescence and chlorophyll autofluorescence.

	Spearman rank rho	*p*-value	Expected mean ± SD
Image 1	0.0336	<0.0001	0.0028±0.0042
Image 2	0.038	<0.0001	0.0027±0.0034
Image 3	0.1128	<0.0001	$6.6577e^{-04}$±$9.2904e^{-04}$
Image 4	0.0699	<0.0001	0.0015±0.0022
Control	0.0023	ns	0.0016±0.0024

Expected Spearman-rank rho distributions (a nonparametric measure of the statistical dependence between the two rankings of two variables, ranging from 0 (no dependence) to 1 (complete dependence)) were created for each image using 100 independently scrambled pixel intensity values of the chlorophyll and Venus channels. The expected means and standard deviations were calculated from these simulated rho distributions. *p*-values were calculated by comparing the observed Spearman rank rho to the distribution of simulated Spearman-Rank rho values (Z-score test).

157 **Table S6.** Primers used to confirm the location of the *saga1* mutants.

158

Primer Name	Sequence
Fprimer (*saga1)*	GCATTGAGATCCGAGATGGT
Rprimer cassette	GCACCAATCATGTCAAGCCT
Rprimer exon (*saga1*)	AGTCCAGGCCGACTACTCC
Rprimer exon (*saga1-3*)	GTGTGAGTGGGATCGCATTCAT
Fprimer cassette (*saga1-3)*	GACGTTACAGCACACCCTTG
Fprimer (*saga1-2)*	CCCACCCCTCACATAAACAC
Rprimer cassette (*saga1-2)*	GCACCAATCATGTCAAGCCT

159

SI Materials and Methods

Screening for the *saga1* mutant

~7,500 mutants on 79 plates, each with 96 colonies were grown on solid TP media in high and low CO_2 in 100 µmol photons m^{-2} s^{-1} light. Mutants that required high CO_2 were sequenced as described in (54). Three mutant alleles mapped to the SAGA1 locus. *saga1:* ~12,590 nt (Intron 26); *saga1-2:* ~13,370 nt (Intron 28)**;** *saga1-3***:** ~12,070 nt (Intron 25). The mutants were confirmed by PCR using a primer that originated from the mutagenesis cassette and a primer in the SAGA1 gene flanking the insertion site. All primers used are in SI Appendix (Table S6).

SAGA1 Sequence Analysis

The full length amino acid sequence of SAGA1 was subjected to PSIPRED (55), NCBI CDD (52) and Phyre2 (56) analyses to identify predicted secondary structures, regions of disordered protein and putative domains.

Generation of the *saga1;SAGA1-Venus* line

Because of the challenge of amplifying across the entire gene, 3 fragments with ~40 nt overlap for Gibson assembly were synthesized (NeoScientific and GeneWiz). These 3 fragments and the resulting construct contained all the exons but only the first 20 introns of SAGA1. Introns 21-34 were removed due to highly repetitive sequences that made them difficult to synthesize and amplify. The 3 fragments were Gibson-assembled into an expression vector that includes a C-terminus Venus and FLAG tag, along with *AphVII* gene for hygromycin resistance under a beta2-tubulin promoter.

Using this construct, we transformed the *saga1* mutant by electroporation as in (45). For each transformation, 14.5 ng kb^{-1} of *Eco*RV-digested construct was mixed with 250 µL of 2 x 10^8 cells mL^{-1} and transformed immediately into *saga1* strains. Cells were selected for hygromycin resistant colonies on hygromycin TAP plates (25 µg mL^{-1}). These colonies were picked into 180 µl of TAP with hygromycin (25 µg mL^{-1}) in a 96-well plate and screened for fluorescence using a Tecan Infinite M1000 PRO plate reader. Excitation and emission settings were: Venus, 532 excitation with 555/20 emission; chlorophyll autofluorescence, 633 excitation with 670/30 emission.

Of the 9 colonies that exhibited a high Venus/chlorophyll fluorescence ratio, 2 showed pyrenoid localization. These 2 complements with pyrenoid Venus fluorescence were the only that rescued the CCM phenotype.

Western Blotting

Chlamydomonas cells were grown to mid-log phase (1 x 10^6 cells mL^{-1}) and then harvested by centrifugation at 5,000 x g. Total protein extraction, detection and quantification of SAGA1 in wild type and *saga1* mutant were performed using method described in (51). Protein was extracted from flash-frozen cells, normalized to chlorophyll absorbance, separated by SDS-PAGE, and western blots were performed with different antibodies.

The primary anti-SAGA1 antibody was used at a 1:10,000 dilution and the secondary horseradish-peroxidase (HRP) conjugated goat anti-rabbit (Life Technologies) at a 1:10,000 dilution. The primary anti-Rubisco antibody was used at a 1:50,000 dilution and the secondary horseradish-peroxidase (HRP) conjugated goat anti-rabbit (Life Technologies) or a goat anti-Rabbit IgG (1:10,000 dilution). 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) or anti-Histone H3 (1:10,000; abcam) followed by HRP conjugated goat anti-mouse (1:10,000; Life Technologies) or goat anti-Rabbit IgG (H+L) 800 CW (1:5,000; LiCOR).

The anti-SAGA1 antibody was raised in rabbit to the last 19 amino acids of C-terminal tail of SAGA1 (RGTGDSPTRRAFGDWRKNL-cooh) by Yenzym Antibodies (South San Francisco, California, USA). The anti-Rubisco polyclonal antibody was a gift from Prof. John Gray, University of Cambridge.

Oxygen Evolution

Chlamydomonas cells grown in Tris-phosphate medium were harvested by centrifugation at 3,000 x g for 3 min at 20°C and resuspended in 25 mM HEPES-KOH (pH 7.3) to the concentration of about 200 µg chlorophyll ml^{-1} using the equations from (57). Aliquots of cells (1 mL) were added to an OXYV1 Hansatech Oxyview System (Hansatech Instruments, King's Lynn, UK) maintained at 23°C using a circulating water bath. The chamber was sealed and illuminated with 200-300 µmol photons m^{-2} s^{-1} until the cells had depleted the internal inorganic carbon storage. When net oxygen evolution ceased, 10-µl aliquots of sodium bicarbonate were added to the cells at 30-s intervals. The cumulative concentrations of HCO_3^- after each addition were as follows: 2.5, 5, 10, 25, 50, 100, 250, 500, 1000 and 2000 µM. The rate of oxygen evolution was recorded per second using the PicoLog 1216 Data Logger (Pico Technologies, St. Neots, UK). $K_{0.5}$ values were calculated using the Michaelis-Menten equation.

Fixing and Embedding for TEM

Chlamydomonas cells were harvested by centrifugation at 5,000 x *g* and fixed for 1 h at room temperature with fixation buffer (2% (v/v) glutaraldehyde, 0.01% (v/v) hydrogen peroxide in TP medium)). Samples were then fixed and osmicated for 1 h at room temperature in an osmium mix (1% (v/v) OsO_4, 1.5% (w/v) $K_3[Fe(CN)_6]$, 2 mM $CaCl_2$). Bulk staining was performed with 2% (w/v) uranyl acetate for 1 h at room temperature. At the interval of each of the subsequent step, three 5 min washes with distilled water were performed. Samples were dehydrated progressively in 70% (v/v) and 95% (v/v) ethanol, followed by two washes in 100% ethanol and 100% acetonitrile.

The samples were then embedded in epoxy resin (34% (w/v) Quetol 651, 44% (w/v) nonenyl succinic anhydride, 20% (w/v) methyl-5-norbonene-2,3-dicarboxylic anhydride, and 2% (w/v) catalyst dimethylbenzylamine (Agar Scientific, Essex, UK)). The fixed cells were first mixed with a mixture of acetonitrile and epoxy resin at a ratio of 1:1 and left to settle overnight. Over the next two days, the cells were refreshed with 100% epoxy resin. Lastly, the samples were cured at 60°C for at least 24 h.

Sections of 50 nm thickness were prepared with a Leica Ultracut UCT (Leica Microsystems, Milton Keynes, UK), mounted on 300 mesh copper grids and counterstained with uranyl acetate followed by lead citrate (Ms. Lyn Carter, Cambridge Advanced Imaging Centre, UK).

Immunogold labeling

Samples mounted on 300 mesh nickel grids were treated with 4% (w/v) Na-meta-periodate for 15 min and 1% (w/v) periodic acid for 5 min to remove superficial osmium and unmask the epitopes. Then, the samples were blocked for 5 min in blocking buffer of 1% BSA (w/v) in high-salt Tris-buffered saline (HSTBSTT) (0.05% (v/v) Tween 20, 0.05% (v/v) Triton X-100 and 500 mM NaCl). The sections were then incubated overnight at room temperature with diluted antibody (1:1000 dilutions of rabbit anti-Rubisco antibody in blocking buffer) followed by two 5-min washes in HSTBSTT and two 5-min washes in dH_2O. To detect bound antibody, the sections were incubated with a secondary antibody (1:200 dilutions of goat anti-rabbit 15-nm gold conjugates (BBI Solutions, Cardiff, UK) in blocking buffer) for 1 h at room temperature followed by two 5-min washes in HSTBSTT and two 5-min washes in dH_2O. The sections were dried and kept for further observation.

Sections were examined using TEM and evaluation of immunogold labeling was made by using the cell counter and measurement functions of ImageJ2 (Fiji). The percentage of

aggregation of gold particles in pyrenoid and chloroplast stromal area was calculated using these equations:

Particle density in the pyrenoid,

$$d = \frac{n_i}{A_i} \quad (1)$$

Normalized number of particles in the pyrenoid,

$$N_i = (d_i - d_k) \times n_i \quad (2)$$

Calculated percentage of Rubisco aggregation in the pyrenoid,

$$\%N_i = \frac{N_i}{N_i + N_j} \times 100 \quad (3)$$

Where d = density, n = number of particles counted, N = normalized number of particles, A = measured area, i = pyrenoid, j = chloroplast stroma, k = background.

Calculation of particle density of the background of every image:

Non-pyrenoid/chloroplast area,

$$A_k = A_{all} - (A_i + A_j) \quad (4)$$

Particle density of the background of every image,

$$d_k = \frac{n_{other}}{A_k} \quad (5)$$

Where n_{other} is the number of particles counted in the non-pyrenoid/chloroplast area.

Normalized number of particles in the chloroplast stroma, j, was derived from equation (2):

$$N_j = (d_j - d_k) \times n_j$$

Indirect Immunofluorescence

10^6 cells were concentrated to a volume of 500 µl and fixed on the poly-L-lysine-coated slides for 10 min and the excess media was removed from the slides. The slides were placed in cold methanol and incubated at -20°C for about 10 min. After then, the dehydrated cells were rehydrated by incubating the slides in 1X PBS for 5 min. This step was repeated twice. After that, Permeating Solution (2% (v/v) of Triton X-100 in 1X PBS) was added to the jar and incubated for 10 min at room temperature followed by two washes in PBS-Mg solution (5 mM $MgCl_2$ in 1X PBS) for 10 min each then air-dried. About 80 µl of Blocking Solution (1% BSA, 1% cold water fish gelatin, 0.05% Triton X-100 and 0.05% Tween20 in 1X PBS) was pipetted onto each cover slip and the slides were inverted onto the cover slips followed by 30 min incubation at room temperature in the humid chambers. After that, about 80 µl of primary

antibody was added to the cells as described previously. The slides were then placed in the humid chambers and incubated overnight at 4°C.

The next day, the slides were washed thrice with 1X PBS for 5 min at room temperature. Secondary antibody was added to the slides as described previously and incubated at room temperature for 1 h. After the three washes with 1X PBS for 5 min each, the slides were then air-dried and mounted. First, about 25 µl of ProLong® Gold Antifade Reagent (Life Technologies) was added onto the slide and a cover slip was then inverted over the droplet. The mounted slides were left to dry overnight in the dark at 4°C.

Primary antibodies used for indirect immunofluorescence were anti-Rubisco (1:2,000) and anti-SAGA antibodies (1:2,000 dilution) while the secondary antibody used was the Alexa Fluor 488 Goat anti-Rabbit IgG (H+L) antibody (Invitrogen; 1:1,000).

Pyrenoid quantification using Rbcs1-mCherry fluorescence

15 Z-sections were taken per cell at 100x magnification. Cell boundaries were defined using chlorophyll autofluorescence. 3D pyrenoid reconstructions were generated from Z-stacks sections using Imaris software (Bitplane). Pyrenoids were defined as 'vesicle like objects' using mCherry fluorescence, and then quantified.

Quick-freeze deep-etch EM (QFDEEM)

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 previously described in (14, 58).

SI References

14. Mackinder LCM, et al. (2016) A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc Natl Acad Sci* 113(21):5958–5963.
45. 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* 26:1398–1409.
48. Schindelin J, et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9(7):676–682.
49. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
50. Rueden CT, et al. (2017) ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* 18(529):1–26.
51. 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.
52. Marchler-Bauer A, et al. (2017) CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45(D1):D200--D203.
53. Sorimachi K, Le Gal-Coëffet MF, Williamson G, Archer DB, Williamson MP (1997) Solution structure of the granular starch binding domain of Aspergillus niger glucoamylase bound to β-cyclodextrin. *Structure* 5(5):647–661.
54. Li X, et al. (2016) An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in *Chlamydomonas reinhardtii*. *Plant Cell* 28(2):367–387.
55. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202.
56. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat Protoc* 10(6):845–858.
57. Porra RJ, Thompson WA, Kriedemann PE (1989) Determination of accurate extinction coefficients and simulteneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim Biophys Acta* 975:384–394.
58. Heuser JE (2011) The origins and evolution of freeze-etch electron microscopy. *J Electron Microsc (Tokyo)* 60(Supplement 1):3–29.