White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/120655/

Version: Accepted Version

Article:

Mackinder, Luke, Chen, Chris, Leib, Ryan et al. (2017) A spatial interactome reveals the protein organization of the algal CO2 concentrating mechanism. Cell. e14. pp. 133-147. ISSN: 1097-4172

https://doi.org/10.1016/j.cell.2017.08.044

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS University of Sheffield UNIVERSITY of York

A Spatial Interactome Reveals the Protein Organization of the Algal CO_2 Concentrating Mechanism

Authors:

Luke C.M. Mackinder[1,2], Chris Chen[1,3], Ryan D. Leib[4], Weronika Patena[1], Sean R. Blum[5], Matthew Rodman[3], Silvia Ramundo[6], Christopher M. Adams[4], and Martin C. Jonikas[1,3,7,8*]

[1]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA

[2]Present address: Department of Biology, University of York, York, YO10 5DD, UK

[3]Department of Biology, Stanford University, Stanford, CA 94305, USA

[4]Stanford University Mass Spectrometry, Stanford University, Stanford, CA, USA

[5]Department of Biomolecular Engineering, UC Santa Cruz, Santa Cruz, CA 95064, USA

[6]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA

[7]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

[8]Lead Contact

[*]Correspondence: mjonikas@princeton.edu

18 **SUMMARY**

19 Approximately one-third of global CO_2 fixation is performed by eukaryotic algae. Nearly all algae
20 enhance their carbon assimilation by operating a CO_2-concentrating mechanism (CCM) built
21 around an organelle called the pyrenoid, whose protein composition is largely unknown. Here,
22 we developed tools in the model alga *Chlamydomonas reinhardtii* to determine the localizations
23 of 135 candidate CCM proteins, and physical interactors of 38 of these proteins. Our data reveal
24 the identity of 89 pyrenoid proteins, including Rubisco-interacting proteins, photosystem I
25 assembly factor candidates and inorganic carbon flux components. We identify three previously
26 un-described protein layers of the pyrenoid: a plate-like layer, a mesh layer and a punctate
27 layer. We find that the carbonic anhydrase CAH6 is in the flagella, not in the stroma that
28 surrounds the pyrenoid as in current models. These results provide an overview of proteins
29 operating in the eukaryotic algal CCM, a key process that drives global carbon fixation.
30

31 Keywords: CO_2-concentrating mechanism, CCM, carbon fixation, *Chlamydomonas reinhardtii*,
32 photosynthesis, pyrenoid, Rubisco, high-throughput fluorescence protein tagging, affinity
33 purification mass spectrometry

34 **INTRODUCTION**

35 Over the past three billion years, the carbon-fixing enzyme Rubisco drew down atmospheric
36 concentrations of CO_2 to trace levels (Dismukes et al., 2001), in effect starving itself of its
37 substrate. In parallel, the oxygenic reactions of photosynthesis have caused the appearance of
38 abundant O_2, which competes with CO_2 for the active site of Rubisco and results in a loss of
39 fixed CO_2 via photorespiration (Bauwe et al., 2010). To overcome these challenges of CO_2
40 assimilation in today's atmosphere, many photosynthetic organisms increase CO_2 levels in the
41 vicinity of Rubisco by operating CO_2 concentrating mechanisms (CCMs). Such mechanisms
42 increase the CO_2:O_2 ratio at the active site of Rubisco, enhancing CO_2 fixation and decreasing
43 photorespiration. CCMs are found in nearly all marine photoautotrophs, including cyanobacteria
44 and eukaryotic algae (Reinfelder, 2011), which together account for approximately 50% of
45 global carbon fixation (Field et al., 1998).

46 In cyanobacterial CCMs, inorganic carbon in the form of bicarbonate (HCO_3^-) is pumped
47 into the cytosol to a high concentration. This HCO_3^- is then converted into CO_2 in specialized
48 icosahedral compartments called carboxysomes, which are packed with Rubisco (Price and
49 Badger, 1989). The components of the cyanobacterial CCMs have largely been identified,
50 facilitated in part by the organization of the genes encoding them into operons (Price et al.,
51 2008). Knowledge of these components has enabled the detailed characterization of the
52 structure and assembly pathway of the beta carboxysome (Cameron et al., 2013).

53 Analogous to the cyanobacterial CCM, the eukaryotic green algal CCM concentrates
54 HCO_3^- in a microcompartment containing tightly-packed Rubisco, called the pyrenoid. The
55 pyrenoid is located in the chloroplast, surrounded by a starch sheath and traversed by
56 membrane tubules that are continuous with the surrounding photosynthetic thylakoid
57 membranes (Engel et al., 2015). Associated with the pyrenoid tubules is a carbonic anhydrase
58 that converts HCO_3^- to CO_2 for fixation by Rubisco (Karlsson et al., 1998). The mechanism of
59 delivery of HCO_3^- to the pyrenoid thylakoids remains unknown. In contrast to the prokaryotic

60 CCM, the protein composition of the eukaryotic algal CCM and the structural organization of the
61 pyrenoid remain largely uncharacterized.
62 In this study, we developed a high-throughput fluorescence protein tagging and affinity
63 purification mass spectrometry (AP-MS) pipeline for the model green alga *Chlamydomonas*
64 *reinhardtii* (Figure 1A). With this pipeline, we determined the localizations of 135 candidate CCM
65 proteins and the physical interactions of 38 core CCM components. Our microscopy data
66 reveals an unexpected localization for the carbonic anhydrase CAH6, identifies three previously
67 undescribed pyrenoid protein layers, and suggests that the pyrenoid shows size selectivity for
68 stromal proteins. The AP-MS data produce a spatially resolved protein-protein interaction map
69 of the CCM and pyrenoid, identifying novel protein complexes including a complex between
70 inorganic carbon transporters LCI1 and HLA3, and suggesting CCM functions for multiple
71 proteins. These results transform our basic knowledge of the eukaryotic CCM and advance the
72 prospects of transferring this system into higher plants to improve crop production (Atkinson et
73 al., 2016; Long et al., 2015).
74

75 **RESULTS AND DISCUSSION**

76 **We Developed a High-Throughput Pipeline for Systematic Localization of Proteins in**
77 ***Chlamydomonas***

78 To allow the parallel cloning of hundreds of genes, we designed an expression cassette that
79 enabled high-throughput seamless cloning via Gibson assembly (Gibson et al., 2009). Open
80 reading frames (ORFs) were amplified by PCR from genomic DNA and cloned in frame with a
81 C-terminal Venus YFP and a 3xFLAG epitope, driven by the strong PsaD promoter. These
82 constructs were transformed into wild-type *Chlamydomonas*, where they inserted into random
83 locations in the genome (Figure 1B). To allow dual tagging of different proteins in the same cell,
84 we developed a second expression vector with an mCherry fluorophore and a hygromycin
85 selection marker (Figure S1A). Potential caveats of our system include loss of the endogenous

86 transcriptional regulation of the protein, including information encoded in the promoter,
87 terminator and genomic locus. Additionally, the C-terminal protein tag could obscure subcellular
88 targeting signals or disrupt functional domains.
89

90 **Our Data Reveal Guidelines for Protein Localization in *Chlamydomonas***

91 Given the notorious difficulties with expressing tagged genes in *Chlamydomonas* (Fuhrmann et
92 al., 1999; Neupert et al., 2009), we started with the understanding that we would only succeed
93 in a fraction of cases, and sought to maximize the total number of proteins localized. We
94 selected target genes from three sources: 1) genes currently thought to be involved in the CCM
95 (See review: Wang et al., 2015); 2) candidate CCM genes, including those identified from both
96 transcriptomic (Brueggeman et al., 2012; Fang et al., 2012) and proteomic (Mackinder et al.,
97 2016) studies; and 3) organelle markers (Figure 1B and Table S1). We were able to determine
98 the localizations of 146 out of the 624 target genes (23%).

99 We sought to leverage the large scale of this study to uncover factors that may
100 contribute to cloning and tagging success in *Chlamydomonas*. We successfully cloned 298 of
101 the 624 target genes (48%). Our cloning success rate decreased with gene size (Figure S1B).
102 Intriguingly, cloning success was higher for genes with high expression levels (Figures S1C and
103 D; $P = 4 \times 10^{-13}$, Mann Whitney U test), suggesting that intrinsic properties of a gene that
104 influence endogenous expression may also affect PCR efficiency.

105 We successfully transformed and acquired protein localization data for 146 of the 298
106 cloned genes (49%). The two main factors correlated with our ability to obtain localization data
107 were: 1) high endogenous gene expression level (Figures S1E and F; $P = 6 \times 10^{-14}$, Mann
108 Whitney U test) and 2) absence of upstream in-frame ATGs (Figure S1G; Cross, 2016). The
109 failure to obtain localization data for genes with in-frame uATGs is likely due to the absence of
110 the correct translational start site in the cloned construct, resulting in a truncated protein that
111 can be functionally impaired, structurally unstable or lacking essential organelle targeting

112 sequence(s). These data suggest that transcript abundance is predictive for localization success
113 and that future protein expression studies will benefit substantially from improved annotation of
114 *Chlamydomonas* translation start sites.
115

116 **146 Tagged Proteins Show 29 Distinct Localization Patterns**

117 To aid in the classification of unknown proteins to subcellular regions, we tagged a series of
118 conserved, well-characterized organelle and cellular structure proteins (Table S1). We then
119 employed a decision tree (Figure 2A) to classify visually the localization of 135 additional
120 proteins into 29 distinct subcellular regions, representing nearly all of the known organelles and
121 cellular structures of *Chlamydomonas* (Figure 2B). The protein localizations from our study are
122 available at https://sites.google.com/site/chlamyspatialinteractome/.

123 Interestingly, 12 proteins were not confined to one organelle but were seen in multiple
124 compartments (Figure 2C and Table S2). If these multiple localizations are not artefacts of our
125 expression system, they may represent proteins that function in multiple compartments or are
126 involved in inter-organelle signalling. Additionally, we observed diverse cytosolic localizations,
127 with subtle differences between localization patterns (Figure S2A).
128

129 **Localization Assignments Agree with Previous Studies for 39/41 Proteins**

130 To evaluate the accuracy of our method, we compared our results with published localizations
131 of individual proteins. Our data shared 25 proteins with the validated "training" set of chloroplast,
132 mitochondria and secretory pathway proteins from Tardif et al. (2012). Nearly all (24/25)
133 matched our localization data, with the only exception being ACP2 (Cre13.g577100). Whereas
134 we saw ACP2 in the chloroplast (Figure 2D), Tardif et al. (2012) saw ACP2 in isolated
135 mitochondria. However, previous studies have either failed to detect ACP2 in mitochondria
136 (Atteia et al., 2009), or saw it in approximately equal abundances in isolated chloroplasts and
137 mitochondria (Terashima et al., 2010). Overall, the ambiguity in the published data leave open

138 the possibility that our ACP2 localization data may in fact be correct. We further compared our
139 data with previously published localizations of CCM components, and found that 15 of 16
140 localizations matched. The strong overlap with previously known localizations indicates that our
141 dataset is of high quality (>95% accurate) and that C-terminal tagging of *Chlamydomonas*
142 proteins results in minimal localization artefacts.
143
144 **CAH6 Localizes to the Flagella**
145 Carbonic anhydrases, which catalyse the reversible reaction of HCO_3^- to CO_2, play a critical role
146 in CCMs (Badger, 2003). Our successful localization of nine *Chlamydomonas* carbonic
147 anhydrases shows that they are found in a diverse range of cellular locations (Figure S2B). In all
148 current models of the CCM (Moroney et al., 2011; Wang et al., 2015), the carbonic anhydrase
149 CAH6 is in the chloroplast stroma, where it has been proposed to convert CO_2 to HCO_3^-.
150 Surprisingly, in our study, CAH6 localized to the flagella in two independent
151 transformation lines (Figure 2D and S2B), and produced no detectable signal in the chloroplast.
152 To exclude the possibility that our observation is due to an artefact (e.g. due to the C-terminal
153 Venus tag), we analysed the localization of CAH6 in existing proteomic datasets. CAH6 is
154 present in the flagellar proteome (Pazour et al., 2005) and has been shown to be an abundant
155 intraflagellar transport (IFT) cargo (Engel et al., 2012), providing independent validation of
156 CAH6 localization to the flagella. Additionally, CAH6 is absent from both the chloroplast
157 proteome (Terashima et al., 2010) and the mitochondrial proteome (Atteia et al., 2009), further
158 suggesting that levels in the chloroplast are low or non-existent.
159 Previous evidence for CAH6 in the stroma came from immunogold labeling experiments,
160 in which Mitra et al. (2004) found a 4.7 fold enrichment of gold particles associated with
161 chloroplast starch relative to control pre-immune serum. This could be an artefact due to cross-
162 reactivity of the immunized serum with another epitope. Alternatively, CAH6 may be an
163 abundant flagellar protein, but present at very low levels in the chloroplast.

164 The apparent absence of carbonic anhydrase in the stroma may be a requirement of the
165 *Chlamydomonas* CCM. A stromal carbonic anhydrase could risk short-circuiting the CCM by
166 promoting the release of CO_2 from HCO_3^- in areas that are not in close proximity to Rubisco. In
167 fact, it has been shown that the expression of carbonic anhydrase in the cyanobacterial cytosol,
168 the likely functional equivalent of the chloroplast stroma, results in the disruption of the
169 cyanobacterial CCM (Price and Badger, 1989).
170 Instead of directly participating in the CCM, CAH6 could be involved in inorganic carbon
171 sensing. *Chlamydomonas* was recently shown to chemotax towards HCO_3^- (Choi et al., 2016),
172 and carbonic anhydrases have been previously implicated in inorganic carbon sensing (Hu et
173 al., 2010). Localization of sensing machinery to the flagella, which are found at the leading edge
174 of swimming cells, could facilitate chemotaxis.
175
176 **PredAlgo is the Best Protein Localization Predictor for *Chlamydomonas***
177 The excellent agreement of our localization data with previous studies provided an opportunity
178 to test the accuracy of the two main localization prediction algorithms used for *Chlamydomonas*
179 proteins, PredAlgo (Tardif et al., 2012) and TargetP (Emanuelsson et al., 2000). For proteins
180 that we observed in the chloroplast, PredAlgo predicted a chloroplast localization for 90% of
181 them, whereas TargetP only predicted a chloroplast localization for 31% (Figure 2E). For
182 mitochondrial proteins, the accuracy dropped to 31% for PredAlgo and 15% for TargetP. For
183 secretory pathway proteins, the accuracy was 38% for PredAlgo and 24% for TargetP. These
184 results highlight that PredAlgo is the best localization predictor for *Chlamydomonas* proteins, but
185 its accuracy drops off significantly when proteins localize to compartments other than the
186 chloroplast.
187
188 **We Assigned 82 Proteins to 13 Sub-Chloroplast Locations**

189 Approximately 56% (82/146) of our proteins localized to the chloroplast. We assigned these 82
190 proteins to 13 sub-chloroplast locations (Table S1; Figures 2A and 3A). Chloroplast envelope
191 proteins showed three subcategories of localization: 1) envelope homogeneous (signal
192 observed evenly throughout the chloroplast envelope); 2) envelope non-homogenous and; 3)
193 envelope plus chloroplast homogenous (signal observed throughout the chloroplast in addition
194 to the envelope). LCIA (Low CO_2 Inducible A) and LCI20 both showed some homogeneous
195 chloroplast signal in addition to a clear envelope signal, suggesting the possibility that these
196 proteins are functional in both the chloroplast envelope and thylakoid membranes.

197 Three proteins produced similar patterns of punctate dots throughout the chloroplast
198 (Figure S3A): a protein with predicted 50S ribosome-binding GTPase activity (Cre12.g524950),
199 histone-like protein 1 (HLP1; Cre06.g285400) (Karcher et al., 2009), and the fatty acid
200 biosynthesis enzyme acetyl-CoA biotin carboxyl carrier (BCC2; Cre01.g037850). The similarity
201 of the localization patterns of these proteins suggests that chloroplast translation, chloroplast
202 DNA and fatty acid synthesis may be co-localized in the chloroplast.

203 We found that proteins with specific patterns of localization were often enriched in
204 certain physical properties. As expected, all eight chloroplast envelope proteins contained one
205 or more transmembrane domains (Table S1). Interestingly, proteins showing homogeneous
206 chloroplast localization (Figure 3B) were enriched in transmembrane domains, found in 9/14
207 homogeneous proteins vs 4/39 for chloroplast non-homogenous proteins (P = 0.0002, Fisher's
208 exact test). This observation suggests that proteins with homogeneous localization are most
209 likely thylakoid membrane-associated.

210

211 **The Pyrenoid Appears to Show Selectivity to Stromal Contents**

212 Because the pyrenoid is a non-membrane-bound organelle, its protein composition cannot be
213 regulated by a membrane translocation step. We therefore sought to understand whether
214 pyrenoid proteins are enriched for any specific physicochemical properties. We classified

215 chloroplast localized proteins into two groups: 1) pyrenoid depleted, where the signal from the
216 pyrenoid was weaker than the surrounding chloroplast and 2) not pyrenoid depleted, where the
217 signal from the pyrenoid was comparable to or brighter than the surrounding chloroplast.
218 Interestingly, the two groups showed different protein molecular weight distributions (P = 0.001,
219 Mann-Whitney U test). The 39 proteins that are not pyrenoid depleted are almost all smaller
220 than ~50 kDa (Figure 3C; the value of ~50 kDa excludes the Venus YFP region, therefore the
221 effective molecular weight is ~78 kDa), suggesting that the pyrenoid may exclude larger
222 proteins.
223
224 **We Identified Multiple New Pyrenoid Components**
225 Electron microscopy-based techniques have shown that the *Chlamydomonas* pyrenoid contains
226 a dense matrix of Rubisco surrounded by a starch sheath and traversed by membrane tubules
227 formed from merged thylakoids (Figure 4A; Engel et al., 2015). Currently, seven proteins have
228 been unambiguously localized to three different regions of the pyrenoid: the pyrenoid matrix,
229 periphery, and tubules. The pyrenoid matrix contains the Rubisco holoenzyme (RBCS/RbcL); its
230 chaperone Rubisco activase (RCA1); essential pyrenoid component 1 (EPYC1), a Rubisco
231 linker protein important for Rubisco packaging in the pyrenoid (Mackinder et al., 2016); and a
232 protein of unknown function (Cre06.g259100; Kobayashi et al., 2016). Under very low CO_2
233 conditions, the LCIB/LCIC complex, whose role is still uncertain (Jin et al., 2016), is known to
234 form puncta around the pyrenoid periphery (Yamano et al., 2010). Recently, a Ca^{2+}-binding
235 protein, CAS, has been shown to specifically localize to the pyrenoid tubules at low CO_2 (Wang
236 et al., 2016). Here, we identify seven additional pyrenoid-localized components and three
237 previously un-described sub-pyrenoid localization patterns (Figure 4B-D).
238
239 **The Pyrenoid Has at Least Four Distinct Outer Protein Layers**

240 Our data suggest that the pyrenoid is surrounded by at least four distinct outer protein layers: 1)
241 LCIB and LCIC localize to puncta around the periphery; 2) PSBP4 (photosystem II subunit P4)
242 localizes to a different set of puncta; 3) STA2 (starch synthase 2) and SBE3 (starch branching
243 enzyme 3) localize to plate-like structures; and 4) LCI9 localizes to a mesh-like structure (Figure
244 4C-E).

245 LCIB, LCIC and PSBP4 showed punctate outer pyrenoid patterns, whereas SBE3, STA2
246 and LCI9 showed a more homogeneous distribution around the pyrenoid periphery (Figure 4B).
247 LCIB and LCIC were co-localized (Figure 4D), supporting the previous finding that they are part
248 of the same complex in the stroma (Yamano et al., 2010).

249 PSBP4-Venus did not co-localize with LCIC-mCherry (Figure 4D), indicating that PSBP4
250 is in a different structure or complex. PPD1, the *Arabidopsis* homolog of PSBP4, has been
251 shown to be in the thylakoid lumen (Liu et al., 2012). Therefore, the PSBP4 puncta likely
252 represent proteins located in the thylakoid lumen. Consistent with this possibility, we also see a
253 small amount of PSBP4-Venus signal within the pyrenoid, and this signal forms a network-like
254 pattern reminiscent of pyrenoid tubules.

255 Our data suggest that both STA2 and SBE3 localize to the starch sheath. Co-localization
256 indicated that STA2 was localized within the perimeter described by LCIC (Figure 4D). STA2
257 formed a clearly defined plate-like pattern around the pyrenoid core (Figure 4C). SBE3 also
258 displayed this plate pattern, but was generally more diffuse than STA2 (Figure 4B).

259 LCI9 was tightly apposed to the pyrenoid matrix and, like STA2, also localized within the
260 perimeter described by LCIC (Figure 4D). However, analysis of Z-sections showed that unlike
261 STA2 and SBE3, LCI9 formed a mesh structure around the pyrenoid (Figure 4C). Intriguingly,
262 the complementary localizations of STA2 and LCI9 suggest that LCI9 may be part of a protein
263 layer that fills the gaps between the starch plates.

264

265 **A Putative Methyltransferase Localizes to the Pyrenoid Matrix**

266 We discovered that SMM7 (Cre03.g151650), a putative methyltransferase, localized to the
267 pyrenoid matrix. This is intriguing because another putative methyltransferase, CIA6
268 (Cre10.g437829), was found to be required for pyrenoid assembly (Ma et al., 2011), although its
269 localization was not determined. Unlike *CIA6*, *SMM7* is strongly transcriptionally upregulated
270 under low CO_2 conditions (Brueggeman et al., 2012; Fang et al., 2012). Identification of the
271 protein targets of CIA6 and SMM7 will likely provide critical insights into pyrenoid biogenesis
272 and regulation.
273
274 **Pyrenoid Tubules are Enriched in PSAH, a Component of Photosystem I**
275 Traversing the pyrenoid are pyrenoid tubules, which are thought to deliver CO_2 at a high
276 concentration to the matrix (Wang et al., 2015). Previous work using immunogold labeling and
277 photosystem (PS) I and PSII activity assays suggested that the pyrenoid tubules from several
278 different algal lineages contain active PSI components and are depleted in PSII components
279 (McKay and Gibbs, 1991). In contrast to these findings, we found that PSII components
280 (PSBP3, PSBQ, PSBR) showed similar pyrenoid localization patterns to those of PSI (PSAG,
281 PSAK and FDX1), cytochrome b_6f (CYC6) and ATP synthase (ATPC) components (Figure
282 S3B).
283 Strikingly, we found that unlike other PSI components, the PSI protein PSAH was
284 enriched within the pyrenoid tubules (Figure 4B). PSAH is a 130 amino-acid protein with a
285 single transmembrane helix that in land plants binds to the core PSI at the site where light
286 harvesting complex II (LHCII) docks in state transitions (Ben-Shem et al., 2003; Lunde et al.,
287 2000). The enrichment of PSAH in the pyrenoid tubules could indicate an additional, pyrenoid-
288 related, role for this protein in algae. Together, our localization data for pyrenoid components
289 allow us to propose a model for the spatial organization of the pyrenoid (Figure 4E).
290
291 **We Generated a Spatially Defined Protein-Protein Interaction Network of the CCM**

292 To understand the interconnectivity of the protein components of the CCM, we developed a
293 large-scale affinity purification mass spectrometry (AP-MS) approach. We chose 38 candidates
294 for AP-MS, focusing on proteins previously implicated in the CCM and on those we found in the
295 pyrenoid (Table S3). We affinity purified fusion proteins using their 3xFLAG tag.

296 To aid in filtering out nonspecific bait-prey interactions from true interactions, we used
297 ^{15}N labeling. We affinity purified baits and associated proteins from lines grown in ^{14}N media,
298 and, before mass spectrometry, we mixed each sample with affinity-purified Venus-3xFLAG and
299 associated proteins from lines grown in ^{15}N media. We quantified our confidence in each
300 protein-prey interaction with a modified WD-score (Behrends et al., 2010), which incorporates
301 the reproducibility, specificity and abundance of each interaction (Figure 5A; see STAR
302 Methods).

303 To identify high confidence interactions, we assumed that interactions between baits and
304 preys localized to different organelles in our study are nonspecific, and thus the distribution of
305 their WD-scores approximates the distribution of WD-scores for false positive interactions. We
306 took the highest WD-score value of 47.5 in this subset and used it as a cut-off. Approximately
307 3.8% of the interactions had WD-scores above this value, giving 513 interactions involving 398
308 proteins (Figure 5B and C). These proteins were considered high-confidence interacting
309 proteins (HCIPs). This method is more stringent than previous methods in which a simulated
310 dataset was used to determine a cut-off, resulting in approximately 5% of data being determined
311 as HCIPs (Behrends et al., 2010; Sowa et al., 2009). One inherent limitation of AP-MS is that it
312 cannot distinguish between direct and indirect interactions, for example this can result in large
313 protein complexes being affinity purified even though a bait protein only directly interacts with
314 one member of the complex.
315

316 **We Used Multiple Approaches to Validate the Network**

317 HCIPs of baits were enriched for proteins with the same PredAlgo predicted localizations
318 (Figure 5D and E). HCIPs recapitulated previously known physical interactions of Rubisco
319 subunits, EPYC1, LCIB and LCIC (Figure 5F). HCIPs of baits from a specific compartment (i.e.
320 chloroplast) are significantly enriched in Gene Ontology function and localization terms related
321 to that compartment (Figure 5G). Finally, as expected from tight transcriptional control of subunit
322 stoichiometry in most complexes (Jansen et al., 2002), most HCIPs were transcriptionally co-
323 regulated with their baits in response to high CO_2 (Figure S4).
324

325 **We Identified Many Novel Rubisco Interacting Proteins**

326 To identify novel protein complexes and new members of known complexes, we performed
327 hierarchical clustering on HCIPs (Figure 6; see Figure S5 for all bait-prey interactions with a
328 WD-score ≥1). The baits RBCS1 and RBCS2 clustered together and shared 15 HCIPs, four of
329 which were also HCIPs of EPYC1. RBCS1- and RBCS2-associated proteins were enriched in
330 uncharacterized proteins. Several of these interactors have homologs in other green algae but
331 lack any conserved domains (Cre01.g054700, Cre01.g054850, Cre02.g088950,
332 Cre16.g655050). We found that Cre16.g655050 contains a predicted N-terminal RbcX fold,
333 which is found in a class of Rubisco chaperones, and the rest of the protein is predicted to be
334 disordered (Figure S6). A BLAST analysis using Cre16.g655050 as the query showed that its
335 full sequence is conserved in the closely related species *Volvox carteri* and *Gonium pectorale.*
336 The N-terminal RbcX-like region is conserved in several more evolutionarily distant
337 Chlorophytes such as *Micromonas pusilla* (Table S4). Whether Cre16.g655050 is a chaperone
338 for Rubisco or performs an alternative function is unknown.

339 Carbohydrate binding domains were found in three Rubisco interactors, including the
340 two starch branching enzymes, SBE1 and SBE4, the latter of which also interacts with EPYC1.
341 Given the concave shape of the pyrenoid-surrounding starch sheaths, there may be variation in
342 starch synthesis and/or breakdown occurring between the two faces. One way to target a

343 subset of starch metabolic enzymes to the inner concave face would be through a binding
344 interaction with pyrenoid matrix proteins. The functional roles of the different SBE isoforms in
345 *Chlamydomonas* have yet to be determined.
346 Interestingly, RBCS1 and RBCS2 interact with an ATP binding cassette (ABC) family
347 transporter (Cre06.g271850). The specific role of this protein may help us elucidate
348 transmembrane transport processes occurring across pyrenoid tubules.
349
350 **EPYC1 Interacts with a Kinase and Two 14-3-3 Proteins**
351 The putative Rubisco linker protein EPYC1 is phosphorylated at low CO_2 (Turkina et al., 2006).
352 Interestingly, we see that EPYC1 associates with a predicted serine/threonine protein kinase
353 (KIN4-2; Cre03.g202000). Understanding the role of this kinase may shed light on post-
354 translational modifications associated with pyrenoid biogenesis and/or function.
355 EPYC1 interacts with two 14-3-3 proteins FTT1 and FTT2. 14-3-3 proteins are known to
356 bind phosphorylated proteins; hence the interaction of 14-3-3 proteins with EPYC1 could
357 potentially be regulated by the phosphorylation state of EPYC1. 14-3-3 proteins can influence
358 the stability, function, interactions and localization of their targets (Chevalier et al., 2009). It is
359 therefore possible that these 14-3-3 proteins are regulating an interaction between EPYC1 and
360 Rubisco, possibly by changing the availability of protein-binding domains.
361
362 **CAH3 Interacts with TAT proteins and STT7**
363 The carbonic anhydrase CAH3 is essential for the CCM (Karlsson et al., 1998) and is thought to
364 convert HCO_3^- to CO_2 in the thylakoid membranes that traverse the pyrenoid, supplying the
365 pyrenoid with a high concentration of CO_2. In our study, CAH3 associated with the TAT2 and
366 TAT3 proteins of the twin-arginine translocation (Tat) pathway (Figure 6 and 7; Table S5), which
367 delivers substrate proteins to the thylakoid lumen. This observation is consistent with work
368 showing that CAH3 contains a predicted Tat signal peptide (Benlloch et al., 2015) and with

369 previous biochemical studies suggesting that CAH3 localizes to the thylakoid lumen (Karlsson et
370 al., 1998).

371 At low CO_2, CAH3 is phosphorylated, and this phosphorylation correlates with increased
372 CA activity and localization to the pyrenoid (Blanco-Rivero et al., 2012). Here, we find that
373 CAH3 has a strong interaction (WD-score = 209) with the kinase STT7 (Figure 6). The role of
374 STT7 in LHCII phosphorylation and state transitions is well documented (Depège et al., 2003).
375 However, it is unlikely that STT7 is directly phosphorylating CAH3, because the kinase domain
376 of STT7 has been shown to be on the stromal side (Lemeille et al., 2009) and CAH3 is thought
377 to be localized in the lumen (Karlsson et al., 1998). A direct interaction between STT7 and
378 CAH3 may be occurring via the N-terminus of STT7, which is thought to be luminal via a single
379 membrane traversing domain (Lemeille et al., 2009).
380

381 **PSBP4 is in a Complex with PSI Assembly Factors**

382 PSBP4 is a PsbP domain (PPD)-containing protein whose *Arabidopsis* homolog is essential for
383 photosystem I assembly and function (Liu et al., 2012). In our data, PSBP4 interacted with four
384 proteins associated with PSI assembly: ycf3, ycf4, CGL71 and TAB2 (Heinnickel et al., 2016;
385 Rochaix et al., 2004), suggesting that PSBP4 and these factors form a PSI assembly complex.
386 PSBP4 also interacts with three uncharacterized conserved green lineage proteins (CGL30,
387 CGL59 and CPLD12) and nine other proteins of unknown function (Figure 7), indicating that
388 these proteins may have roles in PSI assembly and function. Notably, PSBP4's localization
389 suggests that PSI assembly occurs at the pyrenoid periphery.
390

391 **The LCIB/LCIC Complex Interacts with Two Bestrophin-Like Proteins**

392 Our data confirm that LCIB and LCIC, known stromal soluble proteins, are in a tight complex
393 (Yamano et al., 2010). The *lcib* mutant has an "air-dier" phenotype: it exhibits WT growth in
394 either very low CO_2 (0.01% CO_2 v/v) or high CO_2 (3% v/v), but dies in air levels of CO_2 (0.04%)

395 (Wang and Spalding, 2006). The functional role of the LCIB/C complex is still unknown. This
396 complex is hypothesized to either form a CO_2 leakage barrier at the pyrenoid periphery or to act
397 as a vectorial CO_2 to HCO_3^- conversion module to recapture CO_2 that escapes from the
398 pyrenoid (Wang et al., 2015). A role in the conversion of CO_2 to HCO_3^- is likely, as several
399 homologs of LCIB were recently shown to be functional β-carbonic anhydrases. However,
400 recombinant LCIB/C had no carbonic anhydrase function (Jin et al., 2016), suggesting that the
401 complex may be tightly regulated or may require additional factors for proper function.
402 Both LCIB and LCIC interact with LCI11 (Cre16.g663450), and LCIC also interacts with
403 Cre16.g662600 (Figure 6 and 7). Both LCI11 and Cre16.g662600 are putative bestrophins,
404 which typically transport chloride but have been shown to be permeable to HCO_3^- (Qu and
405 Hartzell, 2008). Furthermore, both proteins are upregulated at low CO_2 levels (Table S1 and
406 Figure S4). LCI11 and Cre16.g662600 directly interact, and both also interact with another
407 bestrophin-like protein, Cre16.g663400.
408
409 **LCI9 Interacts with PFK1, PFK2 and SBE3 to Form a Carbohydrate Metabolism Module**
410 As described above, LCI9 forms a mesh structure, likely in the gaps between starch plates.
411 LCI9 contains two CBM20 (carbohydrate binding module 20) domains and is predicted to
412 function as a glucan 1,4-α-glucosidase. Glucan 1,4-α-glucosidases hydrolyze glucosidic bonds,
413 releasing glucose monomers from glucan chains. Therefore, LCI9 most likely plays a role in
414 starch breakdown at the pyrenoidal starch plate junctions. AP-MS analysis shows that the
415 strongest HCIPs of LCI9 are PFK1 and PFK2 (phosphofructokinases 1 and 2). PFK is a key
416 regulator of glycolysis and is important for maintaining cellular ATP levels (Johnson and Alric,
417 2013). The exact metabolic role of an LCI9, PFK1 and PFK2 assemblage is still unclear. LCI9
418 also associates with SBE3, which in turn associates with STA3 and DPE2 (disproportionating
419 enzyme 2), a putative α-1,4-glucanotransferase. Because SBE3 and its HCIPs are involved in
420 starch synthesis and modification, enzymes catalysing starch breakdown and starch synthesis

421 are potentially in close proximity, allowing tight regulation of starch structure. It should be noted
422 that a caveat of performing AP-MS on proteins containing CBMs is that proteins could co-
423 precipitate due to binding a common carbohydrate substrate, not due to direct protein-protein
424 interactions.
425
426 **Bicarbonate Transporters LCI1 and HLA3 Form a Complex with a P-type ATPase**
427 HLA3 (high light activated 3) and LCI1 have both been implicated in HCO_3^- uptake at the plasma
428 membrane (Ohnishi et al., 2010; Yamano et al., 2015). HLA3 is an ABC transporter, and its
429 absence under low CO_2 conditions results in a reduced uptake of inorganic carbon by
430 *Chlamydomonas* cells (Yamano et al., 2015). HLA3 expressed in *Xenopus* oocytes showed
431 moderate uptake of HCO_3^- (Atkinson et al., 2016). LCI1 lacks any conserved functional or
432 structural domains and contains four predicted transmembrane regions. Knock-down of LCI1
433 protein resulted in a small reduction in inorganic carbon uptake (Ohnishi et al., 2010); however,
434 the function of LCI1 has not been demonstrated in a heterologous system.
435 Unexpectedly, we found that HLA3 and LCI1 are found together in a complex. The two
436 proteins showed a reciprocal, strong interaction, each having WD scores >125. In addition, they
437 appear to be in a complex with ACA4 (Autoinhibited Ca^{2+}-ATPase 4; Cre10.g459200), a P-type
438 ATPase/cation transporter. Alignment of ACA4 with functionally characterized P-type ATPases
439 shows that it is a member of the group IIIA family of P-type ATPases (Figure S7). Group IIIA
440 members are known H^+-exporting ATPases (Thever and Saier, 2009). ACA4 may be aiding
441 HCO_3^- uptake either by maintaining a H^+ gradient that HLA3 and/or LCI1 is using to drive HCO_3^-
442 uptake, or by generating localized cytosolic alkaline regions similar to those that form near anion
443 exchanger I during HCO_3^- uptake (Johnson and Casey, 2011). A localized alkaline region could
444 decrease HCO_3^- to CO_2 conversion and hence diffusion out of the cell.
445 The regulation of inorganic carbon transport is critical for the efficiency of the CCM.
446 Recent work has shown that Ca^{2+} signalling is key for proper regulation of the CCM, with the

447 Ca^{2+}-binding protein CAS1 transcriptionally regulating HLA3 and other components (Wang et
448 al., 2016). One HCIP of HLA3 is an EF-hand-containing Ca^{2+}/calmodulin-dependent protein
449 kinase (Cre13.g571700), which could potentially regulate HLA3 post-translationally. Additionally,
450 HLA3 physically interacts with an adenylate/guanylate cyclase (CYG63: Cre05.g236650).
451 Adenylate and guanylate cyclases are known to play a role in sensing inorganic carbon across a
452 broad range of taxa (Tresguerres et al., 2010). Thus, Cre13.g571700 and Cre05.g236650 may
453 represent another mode of CCM regulation, possibly by sensing inorganic carbon availability at
454 the plasma membrane.
455
456 **Perspective**
457 By developing an efficient fluorescent protein-tagging and AP-MS pipeline in *Chlamydomonas*,
458 we have generated a spatially defined network of the *Chlamydomonas* CCM. This large-scale
459 approach gives a comprehensive view of the CCM by revealing missing components, by
460 redefining the localization of others, and by identifying specific protein-protein interactions. Our
461 work also provides insight into the function and regulation of these known and newly discovered
462 CCM proteins, and represents a valuable resource for their further characterization.
463 Our observation that the pyrenoid matrix appears to exclude proteins larger than ~78
464 kDa may be related to the liquid-like nature of the matrix (Freeman Rosenzweig et al., 2017).
465 Interestingly, another liquid-like non-membrane organelle, the *C. elegans* P granule, shows size
466 exclusion of fluorescently labelled dextrans 70 kDa and larger (Updike et al., 2011). This
467 behavior may result from surface tension generated by the proteins that produce the liquid
468 phase (Bergeron-Sandoval et al., 2016).
469 Our results suggest changes to the existing model of inorganic carbon flux to the
470 pyrenoid (Figure 7). The apparent absence of carbonic anhydrase in the chloroplast stroma
471 aligns the *Chlamydomonas* CCM model more with the cyanobacterial model, in which the
472 absence of carbonic anhydrase in the cytosol is critical for inorganic carbon accumulation in the

473 form of HCO_3^- (Price and Badger, 1989; Price et al., 2008). The localization of the carbonic
474 anhydrase CAH6 in flagella suggests potential roles in inorganic carbon sensing. Furthermore,
475 the discovery that HLA3 and LCI1 form a complex and the identification of potential regulatory
476 factors of this complex will aid in the characterization and ultimately the reconstitution of this key
477 plasma membrane bicarbonate transport pathway.
478 Due to a rapidly rising global population and a finite agricultural land area, novel
479 approaches are essential to maintain food security. One potential approach for improving yields
480 is the transfer of a CCM into higher plants to increase CO_2 fixation rates (Long et al., 2015).
481 Recent work has found that nearly all algal CCM proteins localize correctly in higher plants with
482 no changes to their protein sequence, suggesting that the transfer of algal components could be
483 relatively straightforward (Atkinson et al., 2016). However, engineering efforts were constrained
484 by our limited knowledge of the components of the algal CCM. The work we present here
485 provides a detailed blueprint of the algal CCM, revealing dozens of new targets for transfer into
486 crop plants to improve carbon fixation, and enhancing our basic molecular understanding of a
487 fundamental cellular process that drives global biogeochemical cycles.
488
489 **SUPPLEMENTAL INFORMATION**
490 Supplemental Information includes 7 figures and 5 tables.
491
492 **AUTHOR CONTRIBUTIONS**
493 L.C.M.M. and M.C.J. designed and supervised the study. L.C.M.M., C.C. and M.R. performed
494 the cloning, L.C.M.M. did the microscopy and L.C.M.M. and C.C. carried out the AP-MS. S.R.
495 and L.C.M.M. developed the affinity purification protocol. W.P. and S.R.B. provided
496 bioinformatics support. R.L. and C.M.A. oversaw the mass spectrometry and peptide mapping.
497 L.C.M.M., C.C. and M.C.J. analysed and interpreted the data. L.C.M.M created the figures. C.C.
498 created the online viewing platform. L.C.M.M. and M.C.J. wrote the manuscript with input from

499 all authors.

500

501 **ACKNOWLEDGEMENTS**

502 We thank Jonikas laboratory members for helpful discussions, Z. Friedberg and R. Vasquez for
503 help with gene cloning, A. Okumu for mass spectrometry sample preparation, H. Cartwright for
504 microscopy support, and U. Goodenough, R. Milo, A. Smith, N. Wingreen, T. Silhavy, M. Meyer
505 and A. McCormick for comments on the manuscript. Stanford University Mass Spectrometry is
506 thankful to the NIH, Award Number S10RR027425 from the NCRR for assistance in purchasing
507 the mass spectrometer. The project was funded by NSF Grants EF-1105617 and IOS-1359682,
508 NIH Grant 7DP2GM119137-02, the Simons Foundation and HHMI grant #55108535, Princeton
509 University (M.C.J.); the University of York (L.C.M.M), and the Carnegie Institution for Science
510 (L.C.M.M. and M.C.J.). Conflict of interest statement: The authors wish to note that the Carnegie
511 Institution for Science has submitted a provisional patent application on aspects of the findings.

512

513 **REFERENCES**

514 Atkinson, N., Feike, D., Mackinder, L., Meyer, M.T., Griffiths, H., Jonikas, M.C., Smith, A.M., and
515 McCormick, A.J. (2016). Introducing an algal carbon-concentrating mechanism into higher
516 plants: location and incorporation of key components. Plant Biotechnol J 14, 1302–1315.

517 Atteia, A., Adrait, A., Brugière, S., Tardif, M., Van Lis, R., Deusch, O., Dagan, T., Kuhn, L.,
518 Gontero, B., and Martin, W. (2009). A proteomic survey of *Chlamydomonas reinhardtii*
519 mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of
520 the α-proteobacterial mitochondrial ancestor. Mol Biol Evol 26, 1533-1548.

521 Badger, M. (2003). The roles of carbonic anhydrases in photosynthetic CO_2 concentrating
522 mechanisms. Photosynthesis Res 77, 83-94.

523 Bauwe, H., Hagemann, M., and Fernie, A.R. (2010). Photorespiration: players, partners and
524 origin. Trends Plant Sci 15, 330-336.

525 Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the
526 human autophagy system. Nature 466, 68-76.

527 Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I.
528 Nature 426, 630-635.

529 Benlloch, R., Shevela, D., Hainzl, T., Grundstrom, C., Shutova, T., Messinger, J., Samuelsson,
530 G., and Sauer-Eriksson, A.E. (2015). Crystal structure and functional characterization of
531 photosystem II-associated carbonic anhydrase CAH3 in *Chlamydomonas reinhardtii*. Plant
532 Physiol 167, 950-962.

533 Bergeron-Sandoval, L.-P., Safaee, N., and Michnick, S.W. (2016). Mechanisms and
534 consequences of macromolecular phase separation. Cell 165, 1067-1079.

535 Blanco-Rivero, A., Shutova, T., Roman, M.J., Villarejo, A., and Martinez, F. (2012).
536 Phosphorylation controls the localization and activation of the lumenal carbonic anhydrase in
537 *Chlamydomonas reinhardtii*. PloS one 7, e49063.

538 Brueggeman, A.J., Gangadharaiah, D.S., Cserhati, M.F., Casero, D., Weeks, D.P., and
539 Ladunga, I. (2012). Activation of the carbon concentrating mechanism by CO_2 deprivation
540 coincides with massive transcriptional restructuring in *Chlamydomonas reinhardtii*. Plant Cell 24,
541 1860-1875.

542 Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013). Biogenesis of a bacterial
543 organelle: the carboxysome assembly pathway. Cell 155, 1131-1140.

544 Chevalier, D., Morris, E.R., and Walker, J.C. (2009). 14-3-3 and FHA domains mediate
545 phosphoprotein interactions. Annu Rev Plant Biol 60, 67-91.

546 Choi, H.I., Kim, J.Y.H., Kwak, H.S., Sung, Y.J., and Sim, S.J. (2016). Quantitative analysis of
547 the chemotaxis of a green alga, *Chlamydomonas reinhardtii*, to bicarbonate using diffusion-
548 based microfluidic device. Biomicrofluidics 10, 014121.

549 Cross, F.R. (2016). Tying down loose ends in the Chlamydomonas genome: functional
550 significance of abundant upstream open reading frames. G3: Genes|Genomes|Genetics 6, 435-
551 446.

552 Depège, N., Bellafiore, S., and Rochaix, J.-D. (2003). Role of chloroplast protein kinase Stt7 in
553 LHCII phosphorylation and state transition in *Chlamydomonas*. Science 299, 1572-1575.

554 Dismukes, G., Klimov, V., Baranov, S., Kozlov, Y.N., DasGupta, J., and Tyryshkin, A. (2001).
555 The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. PNAS
556 98, 2170-2175.

557 Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in
558 large-scale protein identifications by mass spectrometry. Nat Methods 4, 207-214.

559 Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular
560 localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300, 1005-
561 1016.

562 Engel, B.D., Ishikawa, H., Wemmer, K.A., Geimer, S., Wakabayashi, K.-i., Hirono, M., Craige,
563 B., Pazour, G.J., Witman, G.B., and Kamiya, R. (2012). The role of retrograde intraflagellar
564 transport in flagellar assembly, maintenance, and function. J Cell Biol 199, 151-167.

565 Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J.M., and Baumeister, W. (2015).
566 Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron
567 tomography. eLife 4, e04889.

568 Fang, W., Si, Y., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., Ladunga, I., Liu, P.,
569 and Spalding, M.H. (2012). Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene
570 expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator
571 CIA5/CCM1. Plant Cell 24, 1876-1893.

572 Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of
573 the biosphere: integrating terrestrial and oceanic components. Science 281, 237-240.

574 Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green
575 fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. Plant J 19, 353-
576 361.

577 Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O.
578 (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods
579 6, 343-345.

580 Heinnickel, M., Kim, R.G., Wittkopp, T.M., Yang, W., Walters, K.A., Herbert, S.K., and
581 Grossman, A.R. (2016). Tetratricopeptide repeat protein protects photosystem I from oxidative
582 disruption during assembly. PNAS 113, 2774-2779.

583 Hu, H., Boisson-Dernier, A., Israelsson-Nordström, M., Böhmer, M., Xue, S., Ries, A., Godoski,
584 J., Kuhn, J.M., and Schroeder, J.I. (2010). Carbonic anhydrases are upstream regulators of
585 CO_2-controlled stomatal movements in guard cells. Nat Cell Biol 12, 87-93.

586 Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data
587 with protein-protein interactions. Genome Res 12, 37-46.

588 Jin, S., Sun, J., Wunder, T., Tang, D., Cousins, A.B., Sze, S.K., Mueller-Cajar, O., and Gao, Y.-
589 G. (2016). Structural insights into the LCIB protein family reveals a new group of β-carbonic
590 anhydrases. PNAS 113, 14716-14721.

591 Johnson, D.E., and Casey, J.R. (2011). Cytosolic H^+ microdomain developed around AE1
592 during AE1-mediated Cl^-/HCO_3^- exchange. The Journal of physiology 589, 1551-1569.

593 Johnson, X., and Alric, J. (2013). Central carbon metabolism and electron transport in
594 *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and
595 starch. Eukaryot Cell 12, 776-793.

596 Karcher, D., Koster, D., Schadach, A., Klevesath, A., and Bock, R. (2009). The *Chlamydomonas*
597 chloroplast HLP protein is required for nucleoid organization and genome maintenance. Mol
598 Plant 2, 1223-1232.

599 Karlsson, J., Clarke, A.K., Chen, Z.Y., Hugghins, S.Y., Park, Y.I., Husic, H.D., Moroney, J.V.,
600 and Samuelsson, G. (1998). A novel α-type carbonic anhydrase associated with the thylakoid
601 membrane in *Chlamydomonas reinhardtii* is required for growth at ambient CO_2. EMBO J 17,
602 1208-1216.

603 Kobayashi, Y., Takusagawa, M., Harada, N., Fukao, Y., Yamaoka, S., Kohchi, T., Hori, K., Ohta,
604 H., Shikanai, T., and Nishimura, Y. (2016). Eukaryotic components remodeled chloroplast
605 nucleoid organization during the green plant evolution. Genome Biol Evo 8, 1-16.

606 Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant,
607 S.S., and Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and
608 growth rate in *Chlamydomonas reinhardtii*. Plant J 66, 770-780.

609 Lemeille, S., Willig, A., Depege-Fargeix, N., Delessert, C., Bassi, R., and Rochaix, J.D. (2009).
610 Analysis of the chloroplast protein kinase Stt7 during state transitions. PLoS Biol 7, e45.

611 Liu, J., Yang, H., Lu, Q., Wen, X., Chen, F., Peng, L., Zhang, L., and Lu, C. (2012). PsbP-
612 domain protein1, a nuclear-encoded thylakoid lumenal protein, is essential for photosystem I
613 assembly in Arabidopsis. Plant Cell 24, 4992-5006.

614 Long, S.P., Marshall-Colon, A., and Zhu, X.G. (2015). Meeting the global food demand of the
615 future by engineering crop photosynthesis and yield potential. Cell 161, 56-66.

616 Lunde, C., Jensen, P.E., Haldrup, A., Knoetzel, J., and Scheller, H.V. (2000). The PSI-H subunit
617 of photosystem I is essential for state transitions in plant photosynthesis. Nature 408, 613-615.

618 Ma, Y., Pollock, S.V., Xiao, Y., Cunnusamy, K., and Moroney, J.V. (2011). Identification of a
619 novel gene, CIA6, required for normal pyrenoid formation in *Chlamydomonas reinhardtii*. Plant
620 Physiol 156, 884-896.

621 Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O.,
622 Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., *et al.* (2016). A repeat protein
623 links Rubisco to form the eukaryotic carbon-concentrating organelle. PNAS 113, 5958-5963.

624 McKay, R.M.L., and Gibbs, S.P. (1991). Composition and function of pyrenoids: cytochemical
625 and immunocytochemical approaches. Canadian Journal of Botany 69, 1040-1052.

626 Mitra, M., Lato, S.M., Ynalvez, R.A., Xiao, Y., and Moroney, J.V. (2004). Identification of a new
627 chloroplast carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Physiol 135, 173-182.

628 Moroney, J.V., Ma, Y., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., DiMario, R.J., Yang,
629 J., and Mukherjee, B. (2011). The carbonic anhydrase isoforms of *Chlamydomonas reinhardtii*:
630 intracellular location, expression, and physiological roles. Photosynthesis Res 109, 133-149.

631 Morris, J.H., Knudsen, G.M., Verschueren, E., Johnson, J.R., Cimermancic, P., Greninger, A.L.,
632 and Pico, A.R. (2014). Affinity purification–mass spectrometry and network analysis to
633 understand protein-protein interactions. Nature protocols 9, 2539-2554.

634 Neupert, J., Karcher, D., and Bock, R. (2009). Generation of *Chlamydomonas* strains that
635 efficiently express nuclear transgenes. Plant J 57, 1140-1150.

636 Ohnishi, N., Mukherjee, B., Tsujikawa, T., Yanase, M., Nakano, H., Moroney, J.V., and
637 Fukuzawa, H. (2010). Expression of a low CO_2-inducible protein, LCI1, increases inorganic
638 carbon uptake in the green alga *Chlamydomonas reinhardtii*. Plant Cell 22, 3105-3117.

639 Pazour, G.J., Agrin, N., Leszyk, J., and Witman, G.B. (2005). Proteomic analysis of a eukaryotic
640 cilium. J Cell Biol 170, 103-113.

641 Price, G., and Badger, M. (1989). Expression of Human Carbonic Anhydrase in the
642 Cyanobacterium *Synechococcus* PCC7942 Creates a High CO_2-Requiring Phenotype Evidence
643 for a Central Role for Carboxysomes in the CO_2 Concentrating Mechanism. Plant Physiol 91,
644 505-513.

645 Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2008). Advances in understanding
646 the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci

647 transporters, diversity, genetic regulation and prospects for engineering into plants. J Exp Bot
648 59, 1441-1461.

649 Qu, Z., and Hartzell, H.C. (2008). Bestrophin Cl^- channels are highly permeable to HCO_3^-. Am J
650 Physiol Cell Physiol 294, C1371-C1377.

651 Reinfelder, J.R. (2011). Carbon concentrating mechanisms in eukaryotic marine phytoplankton.
652 Annual review of marine science 3, 291-315.

653 Rochaix, J.-D., Perron, K., Dauvillée, D., Laroche, F., Takahashi, Y., and Goldschmidt-
654 Clermont, M. (2004). Post-transcriptional steps involved in the assembly of photosystem I in
655 *Chlamydomonas*. Biochem Soc Trans 32, 567-570.

656 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch,
657 S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for
658 biological-image analysis. Nat Methods 9, 676-682.

659 Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human
660 deubiquitinating enzyme interaction landscape. Cell 138, 389-403.

661 Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugière, S., Hippler, M., Ferro, M.,
662 Bruley, C., and Peltier, G. (2012). PredAlgo: a new subcellular localization prediction tool
663 dedicated to green algae. Mol Biol Evol, mss178.

664 Terashima, M., Specht, M., Naumann, B., and Hippler, M. (2010). Characterizing the anaerobic
665 response of *Chlamydomonas reinhardtii* by quantitative proteomics. Mol Cell Proteomics 9,
666 1514-1532.

667 Thever, M.D., and Saier, M.H. (2009). Bioinformatic characterization of p-type ATPases
668 encoded within the fully sequenced genomes of 26 eukaryotes. J Membr Biol 229, 115-130.

669 Tresguerres, M., Buck, J., and Levin, L.R. (2010). Physiological carbon dioxide, bicarbonate,
670 and pH sensing. Pfluegers Arch/Eur J Physiol 460, 953-964.

671 Turkina, M.V., Blanco-Rivero, A., Vainonen, J.P., Vener, A.V., and Villarejo, A. (2006). CO_2
672 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas*
673 *reinhardtii*. Proteomics 6, 2693-2704.

674 Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear
675 pore complex environment in the *C. elegans* germ line. J Cell Biol 192, 939-948.

676 Wang, L., Yamano, T., Takane, S., Niikawa, Y., Toyokawa, C., Ozawa, S.-i., Tokutsu, R.,
677 Takahashi, Y., Minagawa, J., Kanesaki, Y., *et al.* (2016). Chloroplast-mediated regulation of
678 CO_2-concentrating mechanism by Ca^{2+}-binding protein CAS in the green alga *Chlamydomonas*
679 *reinhardtii*. PNAS 113, 12586-12591.

680 Wang, Y., and Spalding, M.H. (2006). An inorganic carbon transport system responsible for
681 acclimation specific to air levels of CO_2 in *Chlamydomonas reinhardtii*. PNAS 103, 10110-
682 10115.

683 Wang, Y., Stessman, D.J., and Spalding, M.H. (2015). The CO_2 concentrating mechanism and
684 photosynthetic carbon assimilation in limiting CO_2: how *Chlamydomonas* works against the
685 gradient. Plant J 82, 429-448.

686 Yamano, T., Sato, E., Iguchi, H., Fukuda, Y., and Fukuzawa, H. (2015). Characterization of
687 cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas*
688 *reinhardtii*. PNAS 112, 7315-7320.
689 Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S., Takahashi, Y., and Fukuzawa, H. (2010).
690 Light and low-CO_2-dependent LCIB-LCIC complex localization in the chloroplast supports the
691 carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant and Cell Physiology 51,
692 1453-1468.
693

694 **MAIN FIGURE LEGENDS**

695

696 **Figure 1. We Developed a High-Throughput Pipeline to Determine the Localization and**
697 **Physical Interactions of Algal Proteins.**

698 (A) A false-color transmission electron micrograph of a *Chlamydomonas reinhardtii* cell. The
699 chloroplast is highlighted in magenta and the pyrenoid matrix in orange.

700 (B) Tagging and mass spectrometry pipeline. Target genes were amplified by PCR and Gibson
701 assembled in frame with Venus-3xFLAG, under the constitutive *PSAD* promoter. Transformants
702 were screened for fluorescence using a scanner, and arrayed to allow robotic propagation.
703 Lines were either imaged using confocal microscopy to determine their spatial distribution or
704 batch cultured for affinity purification-mass spectrometry (AP-MS).

705 **Figure 2. Tagged Proteins Localized to a Diverse Range of Cellular Locations, and**

706 **Revealed That CAH6 Localizes to Flagella.**

707 (A) A decision tree was used to assign proteins to specific subcellular locations.

708 (B) Representative images of proteins localized to different cellular locations. The number of

709 different lines showing each localization pattern is in parentheses.

710 (C) Representative images of proteins that localized to more than one compartment. The solid

711 outer line inset in the Cre07.g337100 image is an overexposure of the region surrounded by a

712 dashed line, to highlight flagellar fluorescence.

713 (D) Comparison of our observations with published localizations. Images show the two proteins

714 that did not match their published locations. All scale bars: 5 μm.

715 (E) Comparison of our observations with localizations predicted by PredAlgo and TargetP.

716 **Figure 3. Chloroplast Proteins Show 13 different Localization Patterns.**

717 (A) Representative images of proteins localized to different chloroplast regions. The number of

718 proteins showing each pattern is in parentheses. Scale bar: 5 μm.

719 (B) The percentage of proteins with predicted transmembrane domains is shown for different

720 localization patterns. Bracket shows a significant difference using Fisher's exact test.

721 (C) Predicted molecular weight of proteins is shown as a function of pyrenoid signal intensity.

722 Cre01.g030900 that has a pyrenoid signal and is above the 50 kDa cut-off is labeled. Bracket

723 shows significant difference using a Mann-Whitney U test.

724 **Figure 4. Pyrenoid Proteins Show at Least Six Distinct Localization Patterns and Reveal**
725 **Three New Protein Layers.**
726 (A) A false-color transmission electron micrograph and deep-etched freeze-fractured image of
727 the pyrenoid highlight the pyrenoid tubules, starch sheath and pyrenoid matrix where the
728 principal carbon fixing enzyme, Rubisco, is located. Images courtesy of Moritz Meyer, Ursula
729 Goodenough and Robyn Roth.
730 (B) Proteins showing various localization patterns within the pyrenoid are illustrated. Scale bar:
731 5 μm.
732 (C) Confocal sections distinguish different localization patterns within the pyrenoid. Each end
733 panel is a space-filling reconstruction. Scale bars: 2 μm.
734 (D) Dual tagging refined the spatial distribution of proteins in the pyrenoid. Scale bar: 5 μm.
735 (E) A proposed pyrenoid model highlighting the distinct spatial protein-containing regions.

736 **Figure 5. The AP-MS Data are of High Quality.**

737 (A) Illustration of the influence of different AP-MS features (reproducibility, specificity, ratio and
738 outlier weighting) on the WD-score. R1 and R2 represent replica 1 and 2.

739 (B) To determine a WD-score cut-off value, a bait-prey matrix of WD-scores was formed
740 containing only baits and preys whose localizations were determined in this study. The WD-
741 scores from this matrix were then used to generate (C).

742 (C) A histogram of WD-scores for “All data,” “Different localization,” “Same localization.” A
743 conservative WD-score cut-off was chosen as the point where all data fell above the highest
744 “Different localization” WD-score. Proteins with a WD-score greater than the cut-off are
745 classified as high confidence interacting proteins (HCIPs).

746 (D) Protein-protein interaction network of baits and HCIPs. Bait proteins are grouped according
747 to their localization pattern as determined by confocal microscopy. Baits and preys are colored
748 based on their predicted localization by PredAlgo. Previously known interactions are indicated
749 by red arrows.

750 (E) Comparison of prey PredAlgo predictions with bait localization. C, chloroplast; SP, secretory
751 pathway; O, Other; M, mitochondria.

752 (F) Confirmation of known interactions from the literature (red arrows). Values are WD-scores.

753 (G) Significantly enriched gene ontology (GO) terms for interactors of baits localized to different
754 cellular structures.

755 **Figure 6. The AP-MS Data Reveals Previously Undescribed Physical Interactions,**
756 **Including That Inorganic Carbon Transporters LCI1 and HLA3 Form a Physical Complex.**
757 Hierarchical clustering of all 38 baits with 398 HCIP preys. Specific groups of interest are boxed
758 and highlighted below. Clustering of all baits and preys with interaction WD-scores ≥ 1 is
759 provided in Figure S5.

760 **Figure 7. Combining Localization, Protein-Protein Interaction and Protein Function Data**
761 **Reveals a Spatially Defined Interactome of the *Chlamydomonas* CCM.**
762 A spatially defined protein-protein interaction model of the CCM. Baits have a gradient fill, prey
763 have a solid fill. Each bait has a unique color. Prey are colored according to their bait, with
764 proteins that interact with multiple baits depicted as pies with each slice colored according to
765 one of their interacting baits. Interactors are connected to their bait by a dashed line
766 representing the direction of interaction. Baits are arranged based on their localization observed
767 in this study. Interactors with predicted transmembrane domains are placed on membranes.
768 Prey of membrane localized baits lacking transmembrane domains are arranged according to
769 their PredAlgo localization prediction. Solid black arrows indicate inorganic flux through the cell.
770 For clarity, a selection of interactors are not included in the map but are highlighted below. All
771 interaction data with corresponding WD-scores can be found in Table S5.

772 **Figure S1. Cloning Success Correlated with Short ORF Size and High Gene Expression;**
773 **and Protein Localization Success Correlated with Expression and Absence of Upstream**
774 **ATGs, Related to Figure 1**
775 (A) The pLM006 vector used for dual tagging of proteins with mCherry.
776 (B) Dependence of cloning success on open reading frame (ORF) size.
777 (C) Relationship of cloning success to the number of fragments per kilobase of transcript per
778 million mapped reads (FPKM) from phototrophic air-grown cells.
779 (D) Distribution of FPKM values of cloned genes and genes where cloning failed.
780 (E) Relationship of localization success to the FPKM from phototrophic air-grown cells.
781 (F) Distribution of FPKM values of cloned and localized genes vs. cloned and not localized
782 genes. (D) and (F) Brackets show significant difference using a Mann-Whitney U test.
783 (G) The relationship of localization success to presence of uATGs in transcripts. Asterisks
784 denote significant differences using Fisher's exact test: *** P <0.0001, ** P = 0.0025, * P =
785 0.025

786 **Figure S2. Diverse Cytosolic Patterns and Carbonic Anhydrase Localizations Were**
787 **Observed, Related to Figure 2**
788 (A) Representative confocal images demonstrating a diverse range of cytosolic localization
789 patterns.
790 (B) Confocal images of successfully tagged and localized carbonic anhydrases. *The cloned
791 construct was based on the CAH9 Augustus v5.0 gene model. Images for CAH5 and CAG1-3
792 are projected Z-stacks. (A) and (B) Scale bars: 5 μm.

793 **Figure S3. Proteins Forming Puncta Within the Chloroplast and Localization of Proteins**

794 **Associated with Photosynthetic Electron Transport, Related to Figure 3**

795 (A) Confocal images of proteins with signals in defined puncta within the chloroplast.

796 (B) Localization of Proteins Associated with Photosynthetic Electron Transport**.** The images for

797 PSBP4 and PSAH are the same as used in Figure 4B. (A) and (B) Scale bars: 5 μm.

798 **Figure S4. Transcriptional Regulation of the CCM Protein-Protein Interaction Network,**

799 **Related to Figure 5**

800 Log_2 fold changes of proteins upregulated (red) or downregulated (blue) in response to low CO_2

801 are overlaid onto the HCIP protein-protein interaction network.

802 **Figure S5. Hierarchical Clustering to Identify Protein Complexes, Related to Figure 6**

803 Hierarchical clustering of all 38 baits and preys having an interaction WD-score ≥1. Large

804 regions of blue across most/all baits correspond to clusters of non-specific interactors.

805 **Figure S6. Further Analysis of Cre16.g655050, Related to Figure 6**

806 Cre16.g655050 has a RbcX N-Terminal Domain and a Disordered C Terminus. Top: A

807 predicted Phyre2 structural model of Cre16.g655050. The table shows the ten best template

808 matches for Cre16.g655050 by Phyre2. The confidence score is the probability that the match

809 between Cre16.g655050 and the template is a true homology. The % ID shows the percentage

810 identity between Cre16.g655050 and the template.

811

812 **Figure S7. ACA4 Belongs to P-type ATPase Family IIIA, Related to Figure 6**

813 Phylogenetic tree analysis of 259 eukaryotic P-type ATPases, including functionally

814 characterized members representing the different P-type ATPase families. *Chlamydomonas*

815 ACA4 groups with family IIIA P-type ATPases, which are involved in H^+ pumping.

816 **Table S1. Summary of Target Gene Features, Cloning and Localization, Related to Figure**
817 **1**
818 (Attached excel spreadsheet)
819
820

821 **Table S2. Proteins That Showed Multiple Localizations, Related to Figure 2**

Phytozome v5.5 (Augustus u111.6) ID	Name	Localization	Predalgo predicted localization	Putative function	Predicted Mw
Cre09.g416800	-	Cytosol and chloroplast	O	-	22.78
Cre12.g552450	-	Cytosol and chloroplast	O	-	11.85
Cre16.g685000	-	Cytosol and chloroplast	C	-	25.67
Cre07.g334800	FDX4	Cytosol and chloroplast	C	Ferredoxin	14.05
Cre09.g396400	UBQ2	Flagella and cytosol	O	Bi-ubiquitin	17.2
Cre03.g204577	DNJ31	Flagella and cytosol	C	DnaJ-like protein	61.75
Cre07.g321800	-	Flagella and cytosol	O	-	24.88
Cre11.g467617	LCI19	Flagella and cytosol	O	Gamma hydroxybutyrate dehydrogenase	30.24
Cre16.g685050	LCI15	Flagella and cytosol	C	PRLI-interacting factor L	34.34
Cre03.g158000	GSA1	Flagella, chloroplast and cytosol	C	Glutamate-1-semialdehyde aminotransferase	49.23
Cre17.g725500	-	Flagella, chloroplast and cytosol	C	-	13.82
Cre07.g337100	-	Flagella and mitochondria	C	-	10.75

822

823

824

Table S3. Proteins Used as Baits for the AP-MS Study, Related to Figure 5

Phytozome JGI v5.5 (Augustus u111.6) ID	Name	Protein description	Localization	Replica type (Biological, BR; Affinity Purification AR; Mass Spectrometry, MR)	Replica 1 MS ID	Replica 2 MS ID	Number of Preys with WD-score >1	Number of HCIPs (WD-score >47.52)
Cre01.g051500	ULP1	Uncharacterized thylakoid lumenal polypeptide	Chloroplast not homogeneous with pyrenoid signal	MR	MAP30	T3B08	179	1
Cre01.g054850	-	-	Chloroplast not homogeneous with pyrenoid signal	AR	MAP17	T1D03	211	6
Cre02.g097800	HLA3	ABC transporter	Plasma membrane and late-secretory pathway	AR	MAP12	T2F06	372	36
Cre02.g120100	RBCS1	Rubisco small subunit 1	Pyrenoid matrix	BR	MAP23	T1D08	81	19
Cre02.g120150	RBCS2	Rubisco small subunit 2	Pyrenoid matrix	AR	MAP14	T2H06	104	20
Cre03.g151650	SMM7	-	Pyrenoid matrix	AR	MAP2	T1A12	303	5
Cre03.g162800	LCI1	Low-CO2-inducible membrane protein	Plasma membrane and late-secretory pathway	AR	MAP11	T1E12	266	7
Cre03.g179800	LCI24	Low-CO2-inducible membrane protein	Chloroplast homogeneous with pyrenoid signal	AR	MAP28	T2D06	249	3
Cre03.g191250	LCI34	Low-CO2-inducible protein	Chloroplast not homogeneous with pyrenoid signal	AR	MAP6	T2F03	239	5
Cre04.g223050	CAH2	Carbonic anhydrase, alpha type, periplasmic	ER	MR	MAP1	T1A07	323	34
Cre04.g223300	CCP1	Low-CO2-inducible chloroplast envelope protein	Mitochondria	AR	MAP22	T1C07	328	48
Cre04.g229300	RCA1	Rubisco activase	Pyrenoid matrix	AR	T3E8	T3E08	463	26
Cre05.g248450	CAH5	Mitochondrial carbonic anhydrase	Mitochondria	BR	MAP18	T1F02	290	19
Cre06.g283750	HST1	Homogentisate solanesyltransferase	Chloroplast not homogeneous with pyrenoid signal	AR	MAP32	T2F11	309	22
Cre06.g295450	HPR1	Hydroxypyruvate reductase	Mitochondria	MR	MAP21	T5D02	162	12
Cre06.g307500	LCIC	Low-CO2 inducible protein	Pyrenoid periphery punctate	MR	MAP4	T1C02	113	6
Cre06.g309000	LCIA	Anion transporter	Chloroplast envelope plus chloroplast homogeneous	AR	MAP9	T1C05	393	43
Cre07.g330250	PSAH	Subunit H of photosystem I	Pyrenoid tubules	AR	T2D2	T2D02	351	2
Cre08.g362900	PSBP4	Lumenal PsbP-like protein	Pyrenoid periphery punctate	AR	MAP35	T3F12	234	24
Cre08.g372450	PSBQ	Oxygen-evolving enhancer protein 3	Chloroplast not homogeneous with pyrenoid signal	AR	MAP20	T3B04	134	3
Cre09.g394473	LCI9	Low-CO2-inducible protein	Pyrenoid periphery mesh	MR	T1E6	T1E06	200	4
Cre09.g415700	CAH3	Carbonic anhydrase 3	Chloroplast homogeneous with pyrenoid signal	AR	T1E9	T1E09	500	11
Cre10.g436550	EPYC1/LCI5	Low-CO2-inducible protein	Pyrenoid matrix	BR	LCI5MAP	T1C11	146	9
Cre10.g444700	SBE3	Starch branching enzyme	Pyrenoid periphery spherical	AR	MAP7	T2F05	212	4
Cre10.g452800	LCIB	Low-CO2-inducible protein	Pyrenoid periphery punctate	AR	T1E11	T1E11	136	3
Cre12.g485050	CAH6	Carbonic anhydrase 6	Flagella	MR	MAP5	T1D07	190	4
Cre12.g507300	LCI30	Low-CO2-inducible protein	Nucleus	MR	MAP27	T2C11	320	34
Cre12.g509050	PSBP3	OEE2-like protein of thylakoid lumen	Chloroplast not homogeneous with pyrenoid signal	AR	MAP25	T3G08	245	10
Cre12.g519300	TEF9	Predicted protein	Chloroplast homogeneous with pyrenoid signal	MR	MAP31	T1D04	180	1
Cre12.g560950	PSAG	Photosystem I reaction center subunit V	Chloroplast not homogeneous with pyrenoid signal	AR	MAP33	T3B03	145	2
Cre13.g577100	ACP2	Acyl-carrier protein	Chloroplast not homogeneous with pyrenoid signal	MR	MAP19	T3B02	189	26
Cre14.g626700	Fd/FDX1	Ferredoxin	Chloroplast not homogeneous with pyrenoid signal	AR	MAP29	T2D10	199	20
Cre16.g651050	CYC6	Cytochrome c_6	Chloroplast not homogeneous with pyrenoid signal	AR	MAP10	T1D12	288	7
Cre16.g652800	-	-	Chloroplast homogeneous with pyrenoid signal	AR	MAP15	T1C01	281	2
Cre16.g662600	-	-	Chloroplast homogeneous pyrenoid disenriched	AR	MAP16	T1C04	371	23
Cre16.g663450	LCI11	Low-CO2-inducible membrane protein	Chloroplast homogeneous with pyrenoid signal	AR	MAP3	T1B12	284	6
Cre17.g721500	STA2	Granule-bound starch synthase I	Pyrenoid periphery spherical	AR	MAP13	T2F09	142	1
Cre17.g724300	PSAK	Photosystem I reaction center subunit psaK	Chloroplast not homogeneous with pyrenoid signal	AR	MAP34	T3B09	319	5
						Totals	9451	513

825

826 **Table S4. Cre16.g655050 BLAST Results, Related to Figure 6**

827 Query cover is the percentage of the query sequence that matches the hit sequence. E-value is

828 the expected value, the lower the E-value the more significant the hit.

Accession	Species	Query cover	E-value	Identity
XP_002950714.1	*Volvox carteri*	77%	4.0E-137	56%
KXZ52617.1	*Gonium pectorale*	71%	9.0E-127	56%
XP_005849673.1	*Chlorella variabilis*	31%	2.0E-53	56%
XP_005645512.1	*Coccomyxa subellipsoidea*	28%	2.0E-50	55%
XP_005847655.1	*Chlorella variabilis*	27%	5.0E-48	53%
XP_001698126.1	*Chlamydomonas reinhardtii*	10%	1.0E-29	100%
XP_013896920.1	*Monoraphidium neglectum*	16%	2.0E-29	60%
XP_002501227.1	*Micromonas commoda*	24%	1.0E-18	41%
XP_003062310.1	*Micromonas pusilla*	24%	2.0E-17	39%

829

830

831 **Table S5. Protein-Protein Interaction Data, Related to Figures 6 and 7**

832 All interactions with a WD-score ≥1 are shown. Rows highlighted in blue were classified as

833 HCIPs.

834

835 (Attached excel spreadsheet)

836 **STAR Methods**

837

838 **CONTACT FOR REAGENT AND RESOURCE SHARING**

839 Further information and requests for resources and reagents should be directed to and will be
840 fulfilled by the Lead Contact, Martin C. Jonikas (mjonikas@princeton.edu).

841

842 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

843 **Strains and Culturing**

844 The background *Chlamydomonas reinhardtii* strain for all experiments was wild-type (WT)
845 cMJ030 (CC-4533). WT cells were maintained on 1.5% Tris-acetate-phosphate (TAP) agar with
846 revised Hutner's trace elements (Kropat et al., 2011) at 22°C in low light (~10 μmol photons m^{-2}
847 s^{-1}). Lines harboring Venus-3xFLAG-tagged genes in the pLM005 plasmid were maintained in
848 the same conditions with solid media supplemented with 20 μg mL^{-1} paromomycin. For lines
849 also harbouring the pLM006 plasmid, the media was further supplemented with 25 μg mL^{-1}
850 hygromycin. During liquid growth for imaging and affinity purification mass spectrometry,
851 antibiotic concentrations were used at $1/10^{th}$ these concentrations.

852

853 **METHOD DETAILS**

854 **Plasmid Construct and Cloning**

855 For the tagging and AP-MS pipeline, we used the pLM005 plasmid, and for dual-tagging
856 experiments, we used the pLM006 plasmid (Mackinder et al., 2016). Open reading frames were
857 PCR amplified (Phusion Hotstart II polymerase, ThermoFisher Scientific) from genomic DNA,
858 gel purified (MinElute Gel Extraction Kit, Qiagen) and cloned in-frame with either a C-terminal
859 Venus-3xFLAG (pLM005) or an mCherry-6xHIS (pLM006) tag by Gibson assembly. Primers
860 were designed to amplify target genes from their predicted start codon up to, but not including,
861 the stop codon. To allow efficient assembly into *Hpa*I-cut pLM005 or pLM006, primers contained

862 the following adapters: Forward primers (5'-3'), GCTACTCACAACAAGCCCAGTT and reverse
863 primers (5'-3'), GAGCCACCCAGATCTCCGTT. To increase our success with larger genes, we
864 split some of these into multiple fragments that were reassembled following PCR amplification.
865 However, due to a multiplicative effect, the cloning efficiency dropped off rapidly: only a 20%
866 efficiency for two fragments (14/69) and 8% for three fragments (6/74). All junctions were
867 sequence verified by Sanger sequencing and constructs were linearized by either *Eco*RV or
868 *Dra*I prior to transformation into WT *Chlamydomonas reinhardtii*. For each transformation, 14.5
869 ng kbp^{-1} of cut plasmid was mixed with 250 µL of 2 x 10^8 cells mL^{-1} at 16 °C in a 0.4 cm gap
870 electroporation cuvette and transformed immediately into WT strains by electroporation using a
871 Gene Pulser II (Bio-Rad) set to 800V and 25uF. Cells transformed with plasmids containing the
872 pLM005 backbone were selected on TAP paromomycin (20 µg mL^{-1}) plates and kept in low light
873 (5-10 µmol photons m^{-2} s^{-1}) until screening for fluorescence. To generate dual-tag lines, lines
874 expressing Venus tagged proteins were sequentially transformed with target genes inserted in
875 the pLM006-mCherry-6xHIS plasmid and selected on TAP paromomycin (20 µg mL^{-1}) and
876 hygromycin (25 µg mL^{-1}) plates. Transformation plates were directly screened for fluorescence
877 using a Typhoon Trio fluorescence scanner (GE Healthcare) with the following excitation and
878 emission settings: Venus, 532 excitation with 555/20 emission; mCherry, 532 excitation with
879 610/30 emission; and chlorophyll autofluorescence, 633 excitation with 670/30 emission. For
880 each construct, three fluorescent colonies were isolated and maintained in 96 arrays using a
881 Singer Rotor propagation robot. A detailed, step by step protocol for cloning and AP-MS is
882 available at: https://sites.google.com/site/chlamyspatialinteractome/.
883
884 **Microscopy**
885 For microscopy of Venus-tagged lines, colonies were transferred from agar to Tris-phosphate
886 (TP) liquid medium (Kropat et al., 2011) in a 96-well microtiter plate and grown with gentle
887 agitation in air at 150 µmol photons m^{-2} s^{-1} light intensity (LumiBar LED lights, LumiGrow). After

888 ~2 days of growth, 15 µL of cells were pipetted onto a 96-well optical bottom plate (Brooks
889 Automation Inc.) and a 120 µL of 1% TP low-melting-point agarose at ~34°C was overlaid to
890 minimize cell movement. Lines grown for detailed Z-stack analysis and dual-tagged lines
891 containing proteins with both Venus and mCherry tags were grown in 80 mL of TP, bubbled with
892 0.01% CO_2 (with 21% O_2, balanced with N_2) for ~12 hours at 150 µmol photons m^{-2} s^{-1} light
893 intensity. 10-15 µL of cells were pipetted on poly-L-lysine coated plates (Ibidi) and overlaid with
894 1% TP agarose as above. All imaging was performed using a spinning-disk confocal
895 microscope (custom modified Leica DMI6000) with Slidebook software (3i). The following
896 excitation and emission settings were used: Venus, 514 excitation with 543/22 emission;
897 mCherry, 561 excitation with 590/20 emission; and chlorophyll, 561 excitation with 685/40
898 emission. All confocal microscopy images were analyzed using Fiji (Schindelin et al., 2012). 3D
899 pyrenoid reconstructions were generated from Z-sections using Imaris software (Bitplane).
900

901 **Affinity Purification**

902 Cell lines expressing Venus-3xFLAG-tagged proteins were grown in 50 mL of TAP media at 100
903 µmol photons m^{-2} s^{-1} light intensity until they reached a cell density of ~2-4 x 10^6 cells mL^{-1}.
904 Cells were then pelleted at 1000 *g* for 4 minutes, resuspended in TP medium and transferred to
905 800 mL of TP medium. They were then bubbled with air with constant stirring and 150 µmol
906 photons m^{-2} s^{-1} light intensity to a density of ~2-4 x 10^6 cells mL^{-1}. All liquid media contained 2
907 µg mL^{-1} paromomycin. In parallel, control strains expressing only the Venus-3xFLAG tag were
908 grown under identical conditions except that, during liquid growth, $^{14}NH_4Cl$, the sole nitrogen
909 source, was replaced with $^{15}NH_4Cl$. This ensured ^{15}N growth for at least eight generations.

910 Cells from Venus-3xFLAG-tagged protein lines and control lines were separately
911 harvested and affinity purified as follows: Cells were spun out (2,000 *g*, 4 minutes, 4°C), washed
912 in 40 mL of ice cold 1xIP buffer (200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM
913 $Mg(OAc)_2.4H_2O$, 1 mM $CaCl_2$, 1 mM NaF, 0.3 mM Na_3VO_4 and 1 cOmplete EDTA-free protease

914 inhibitor (Sigma-Aldrich)/ 50 mL), centrifuged then resuspended in a 1:1 (v/w) ratio of ice-cold
915 2xIP buffer to cell pellet. This cell slurry was then added drop wise to liquid nitrogen to form
916 small *Chlamydomonas* pellets approximately 5 mm in diameter. These were stored at -70°C
917 until needed.

918 Cells were lysed by grinding 1g of *Chlamydomonas* pellets by mortar and pestle at liquid
919 nitrogen temperatures. The ground cells were defrosted and dounced 20 times on ice with a
920 Kontes Duall #21 homogeniser (Kimble). Membranes were solubilised by incrementally adding
921 an equal volume of ice-cold 1xIP buffer plus 2% digitonin (final concentration is 1%; Sigma-
922 Aldrich), followed by a 40 minute incubation with nutation at 4°C. The lysate was then clarified
923 by spinning for 30 minutes at ~13,000 *g* in a table-top centrifuge at 4°C. The supernatant was
924 then transferred to 225 µL of protein G Dynabeads (ThermoFisher Scientific) that had been
925 incubated with anti-FLAG M2 antibody (Sigma-Aldrich) according to the manufacturer's
926 instructions, except 1xIP buffer was used for the wash steps. The Dynabead-cell lysate was
927 incubated for 1.5 hours on a rotating platform at 4°C, then the supernatant removed. The
928 Dynabeads were washed 4 times with 1xIP buffer plus 0.1% digitonin followed by a 30 minute
929 competitive elution with 50 µL of 1xIP buffer plus 0.25% digitonin and 2 µg/ µL 3xFLAG peptide
930 (Sigma-Aldrich). After elution samples were diluted 1:1 with 2X SDS-PAGE buffer (BioRad)
931 containing 50 mM β-mercaptoethanol and heat denatured for 10 minutes at 70°C. Tagged
932 protein and control denatured elutions were then mixed 1:1 (16µL:16µL), and 28 µL of sample
933 was partially purified by electrophoresing on a 10% Tris-glycine gel (Criterion TGX gel ; BioRad)
934 until the protein moved 1.8 to 2 cm (~40 minutes at 50V). Gel slices were then fixed in 1 mL of
935 10% acetic acid, 50% methanol, 40% deionised water for 1 hour, with a change of the fixing
936 solution after 15 minutes, 30 minutes and 1 hour. Gel slices were soaked twice in 1mL of
937 deionized water for 2 minutes, then stored in 1% acetic acid at 4°C until processing for mass
938 spectrometry.
939

940 **Mass Spectrometry**

941 Limited gel slices representing 3xFLAG AP eluates were diced into 1x1mM squares and then
942 incubated in 50 mM ammonium bicarbonate for ~15 minutes. After pH neutralization, the diced
943 gel slices were reduced with 5 mM DTT for 30 minutes at 55°C. The reducing buffer was
944 removed and samples were alkylated with 10 mM propionamide at 10 mM for 30 minutes at
945 room temperature. Gel samples were washed with multiple rounds of 1:1 acetonitrile:50mM
946 ammonium bicarbonate until the gels were free of all dye. 10 uL of 125 nanogram trypsin/lysC
947 (Promega) was added to each gel band and gels were allowed to swell for 10 minutes, followed
948 by the addition of 25 to 35uL 50 mM ammonium bicarbonate. The gels were digested overnight
949 at 37°C. Peptide extraction was performed in duplicate, and the peptide pools dried in a speed
950 vac until readied for LCMS/MS. Each peptide pool was reconstituted in 12.5 uL 0.1% formic
951 acid, 2% acetonitrile, 97.9% water and loaded onto a NanoAcquity UPLC (Waters). The mobile
952 phases were A: 0.585% acetic acid, 99.415% water and B: 0.585% acetic acid, 10% water,
953 89.415% acetonitrile. The analytical column was a picochip (New Objective) packed with 3 μM
954 C18 reversed phase material approximately 10.5cm in length. The flow rate was 600 nL/min
955 during the injection phase and 450 nL/min during the analytical phase. The mass spectrometer
956 was a orbitrap Elite, operated in a data-dependant acquisition (DDA) schema in which the
957 fifteen most intense multiply charged precursor ions were selected for fragmentation in the ion
958 trap. The precursor mass settings were a resolution of 120,000 and an ion target value of
959 750,000, max fill time 120 usec. The MS/MS settings were 50,000 ions and a maximum fill time
960 of 25 μsec.

961

962 **QUANTIFICATION AND STATISTICAL ANALYSIS**

963 **Mass Spectrometry Data Analysis**

964 *Peptide identification*

965 MS/MS data were analyzed using an initial screening by Preview for validation of data quality,
966 followed by Byonic v2.6.49 (Protein Metrics Inc.) for peptide identification and protein inference
967 against version 5.5 of the *Chlamydomonas reinhardtii* translated genome. In a typical analysis,
968 each data file was searched in two parallel Byonic analyses: one for the unlabeled peptides, and
969 one treating the incorporation of ^{15}N isotopic labels as a fixed modification. In both cases, these
970 data were restricted to 12 ppm mass tolerances for precursors, with 0.4 Da fragment mass
971 tolerances assuming up to two missed cleavages and allowing for only fully tryptic peptides.
972 These data were validated at a 1% false discovery rate using typical reverse-decoy techniques
973 as described previously (Elias and Gygi, 2007). The combined identified peptide spectral
974 matches and assigned proteins were then exported for further analysis using custom tools
975 developed in MatLab (MathWorks) to provide visualization and statistical characterization.
976
977 *Background to CompPASS analysis*
978 To identify *bona fide* interactions, we used an $^{14}N/^{15}N$ labeling strategy. Bait-Venus-3xFLAG
979 fusion proteins were grown in ^{14}N media in parallel to ^{15}N grown controls expressing only
980 Venus-3xFLAG. 3xFLAG affinity purification was performed for target and control lines in
981 parallel, proteins were eluted by 3xFLAG competition, and then target and control elutions were
982 mixed prior to SDS-PAGE purification and MS. In theory, this approach should control for non-
983 specific proteins interacting with the resin, 3xFLAG peptide, Venus and tubes and it should also
984 control for MS variation between runs, resulting in only large ratios for specific interactors.
985 However, analysis of the complete data set showed that using only $^{14}N/^{15}N$ ratios was
986 insufficient to identify real interactors from false positives. This is generally due to the spurious
987 nature of some preys, and in several cases the ratios diverged from 1 across all baits for some
988 preys. Therefore, to analyze our $^{14}N/^{15}N$ labeled dataset, we decided to adapt the CompPASS
989 method (Sowa et al., 2009), an approach previously developed to analyze AP-MS studies of this
990 size using unlabeled proteins.

991

992 *Identification of protein carry-over between MS runs*

993 Carry-over of proteins from previous MS runs is a common source of contamination, and
994 increases with protein abundance and hydrophobicity (Morris et al., 2014). To reduce carry-over
995 contamination, column wash steps and MS blanks were frequently included, and placed
996 between samples that were previously identified to be prone to carry-over. In addition, an *in*
997 *silico* filtering step was included to remove carry-over contamination prior to CompPASS
998 analysis. Data was sorted by MS run order and half-life-like patterns of decreasing raw values
999 were scanned for. To confirm contamination was due to carry-over and not true interactions,
1000 half-life-like patterns between MS replicas ran in a different order were compared. Raw values
1001 for carry-over contamination that showed the same patterns between replicas were set to zero.

1002

1003 *Generating WD-scores*

1004 The CompPASS method uses spectral counts and devises a score (WD-score) based on the
1005 specificity of the prey, spectral count number and reproducibility. Instead of using spectral
1006 counts, we used $^{14}N/^{15}N$ ratios. Using $^{14}N/^{15}N$ ratios helps clean out abundant common
1007 contaminants. Based on the CompPASS method, we generated WD-scores for each bait-prey
1008 interaction. First, we determined the $^{14}N/^{15}N$ ratios for the bait-prey interaction for each replica. If
1009 a protein had no spectral counts in one of the ^{14}N or ^{15}N, the spectral count was set to 1 to
1010 generate a ratio. If it was not detected in both the ^{14}N and ^{15}N, its $^{14}N/^{15}N$ ratio value was
1011 therefore 1. The ratios for each replica were then averaged to populate a stats table of 38 baits
1012 and 3251 preys.

1013

1014

Stats table

	Bait 1	Bait 2	Bait 3	Bait *k*	
Prey 1	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	$X_{k,1}$	$\overline{X}_1$
Prey 2	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{k,2}$	$\overline{X}_2$
Prey 3	$X_{1,3}$	$X_{2,3}$	$X_{3,3}$	$X_{k,3}$	$\overline{X}_3$
Prey *m*	$X_{1,m}$	$X_{2,m}$	$X_{3,m}$	$X_{k,m}$	$\overline{X}_m$

1015

1016 $X_{i,j}$ is the average $^{14}N/^{15}N$ ratio from two replicas (q and r) for prey *j* from bait *i* (Eq. S1).

1017

1018
$$X_{i,j} = \frac{(q_{i,j}^{14N}/q_{i,j}^{15N}) + (r_{i,j}^{14N}/r_{i,j}^{15N})}{2} \quad \text{(Eq. S1)}$$

1019

1020 *m* is the total number of unique prey proteins identified (3251).

1021 *k* is the total number of unique baits (38).

1022 We plugged the above values into the WD-score equation (Behrends et al., 2010), which is

1023 defined as follows (Eqs. S2-S4):

1024
$$WD_{i,j} = \sqrt{\left(\frac{k}{\sum_{j=1}^{i=k} f_{i,j}}\ \omega_j\right)^p X_{i,j}} \quad \text{(Eq. S2)}$$

1025
$$\omega_j = \left(\frac{\sigma_j}{\overline{X}_j}\right), \overline{X}_j = \frac{\sum_{i=1,j=n}^{i=k} X_{i,j}}{k}, n = 1,2, \ldots m, \begin{matrix} if\ \omega_j \le 1 \rightarrow \omega_j = 1 \\ if\ \omega_j > 1 \rightarrow \omega_j = \omega_j \end{matrix} \quad \text{(Eq. S3)}$$

1026
$$f_{i,j} = \{ \begin{matrix} 1; X_{i,j} > 0 \\ X_{i,j} \end{matrix} \quad \text{(Eq. S4)}$$

1027

1028
$$if\ 0.098 < \frac{q_{i,j}}{r_{i,j}} \le 10.2 \rightarrow p = 2 \quad \text{(Eq. S5)}$$

1029
$$if\ \frac{q_{i,j}}{r_{i,j}} \le 0.098\ or\ \frac{q_{i,j}}{r_{i,j}} > 10.2 \rightarrow p = 1 \quad \text{(Eq. S6)}$$

1030 $$if\ q_{i,j}^{14N} \leq 1\ or\ r_{i,j}^{14N} \leq 1\ \rightarrow p = 1 \qquad \text{(Eq. S7)}$$

1031

1032 The WD-score has 3 main components taking into account the uniqueness, the reproducibility
1033 and the $^{14}N/^{15}N$ ratio. $\frac{k}{\sum_{j=1}^{i=k} f_{i,j}}$ is a "uniqueness" measure that up-weights unique interactors and
1034 down-weights promiscuous interactors. It counts the number of baits that a given prey was
1035 detected in. Therefore, the less often the prey is seen across the baits, the larger the value. *k* is
1036 constant for all preys, in our case it is 38. Therefore, if a prey is unique to one bait, this term will
1037 equal 38 (38/1), whereas if is a prey is seen interacting with all baits this value would be 1
1038 (38/38). In addition to the uniqueness measurement is a weighting term, ω_j (Eq. S3). This term
1039 is only applied if the standard deviation is greater than the mean for a prey across all baits. It
1040 was introduced in Behrends et al. (2010) to offset the low uniqueness value for true interactors
1041 that are seen in many baits.

1042

1043 *p* is a reproducibility measure that upweights preys that are seen in both replicas. We modified
1044 the *p* weighting (Eqs. S5-S7) to only come into effect if the ratio averages were ≤10.2 fold of
1045 each other. We decided to add a "closeness" value of replica ratios, because spurious and
1046 general contaminant preys would be frequently detected in both replicas but would have a large
1047 $^{14}N/^{15}N$ ratio difference between replicas, whereas in true interactors $^{14}N/^{15}N$ ratios between
1048 replicas are generally very similar. To determine a cut-off, we looked at all preys that were only
1049 detected in one bait and which were also replicated in both MS runs (this gave 173 high-
1050 confidence true interactions). We then took the largest fold change between the replica $^{14}N/^{15}N$
1051 ratios where more than 1 spectral count was used to determine the ratio.

1052

1053 $X_{i,j}$ is the $^{14}N/^{15}N$ ratio. In Sowa et al. (2009), this is the average of total spectral counts for the
1054 replicas. In our case the $X_{i,j}$ is the average of the $^{14}N/^{15}N$ of both replicas. By using the $^{14}N/^{15}N$

1055 ratio we in effect have performed an initial clean up of the data, with background contaminants
1056 (seen in both the ^{14}N bait and ^{15}N control) down-weighted.
1057
1058 If the protein was not detected in either replica it was assigned a WD-score of 0.
1059
1060 *Determining the WD-score threshold*
1061 Due to the empirical nature of the WD-score, a cut-off must be determined. Sowa et al. (2009)
1062 generated a random dataset and used a cut-off value above which 5% of the random dataset
1063 fell. Interestingly, this also corresponded to ~5% of the real dataset, which they recommend as
1064 a suitable approximation for the threshold. Due to potential pitfalls in the generation of a random
1065 dataset, we decided to use an alternate approach to determine the WD-score cut-off. We made
1066 a new stats table that included all baits (38) and just preys (83) that we had obtained localization
1067 data for. We then made the assumption that interactions between baits and preys in spatially
1068 different regions (at the organelle level) were non-specific. We took the highest WD-score value
1069 in this new stats table and used it as the WD-score cut-off, which, in our case was 47.516.
1070 Approximately 3.78% of the data lies above this value, giving 513 interactions involving 398
1071 proteins. A WD-score >47.516 was thus considered a high confidence interacting protein
1072 (HCIP).
1073
1074 *Data visualization*
1075 WD-score analysis and bait-prey matrix assembly were performed in Microsoft Excel.
1076 Hierarchical clustering was done using Multi Experiment Viewer (http://mev.tm4.org/). Network
1077 visualization was done in Cytoscape (http://www.cytoscape.org/).
1078
1079 **Comparison of Localization Data with PredAlgo and TargetP**

1080 To allow the direct comparison of PredAlgo and TargetP predictions to our localization data, we
1081 classified our data as follows: Chloroplast (C) includes “Chloroplast,” "Cytosol and chloroplast,”
1082 and "Flagella, chloroplast and cytosol." Mitochondria (M) includes “Mitochondria,” "Flagella and
1083 mitochondria,” and "Unclear ER or mitochondria." Secretory pathway (SP) includes “Plasma
1084 membrane and late-secretory pathway,” “ER,” "Unclear ER or mitochondria,” “Golgi and
1085 secretory pathway,” “Cell wall,” and “Contractile vacuoles.” Other (O) includes “Cytosol,”
1086 “Flagella,” "Flagella and cytosol,” "Flagella and mitochondria,” "Flagella, chloroplast and
1087 cytosol,” and “Nucleus.” The data used for analysis excluded proteins used in the PredAlgo
1088 training set (Tardif et al., 2012).
1089

1090 **Gene Expression Values and Presence of Upstream ATGs**

1091 Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values were
1092 downloaded from Phytozome (https://phytozome.jgi.doe.gov/phytomine/begin.do). For analysis
1093 of cloning and localization success relative to transcript abundance, FPKM values for
1094 "photo.HighLight MidLog" from the GeneAtlas experiment group were used. These experiments
1095 were performed at ambient CO_2 levels (~400 ppm), a CO_2 concentration reflective of our
1096 experimental conditions. For an approximation of CCM induction, $\log_2$ FPKM changes were
1097 calculated by dividing FPKM values from photo.HighLight MidLog and hetero.Ammonia MidLog
1098 experiments of the GeneAtlas experiment group.

1099 An analysis of genes for upstream ATGs (uATGs) was recently performed on version 5.5
1100 of the *Chlamydomonas* genome (Cross, 2016). Comparison of our localization data to the
1101 presence of uATGs showed that localization success was 63% (89/141) in the absence of
1102 upstream ATGs (uATGs), relative to only 30% (17/57; Figure S1G) when uATGs were found in-
1103 frame to the annotated start site in the mRNA (Cross, 2016).

1104 Interestingly, localization success only rose to 40% for both cloned genes that contained
1105 an out-of-frame uATG (12/30) and cloned genes that contained an uATG followed by an in-

1106 frame stop codon (26/65). This suggests that in some cases out-of-frame uATGs may be the
1107 correct translation initiation sites due to unannotated splicing events. Our data is in general
1108 agreement with the analysis by Cross (2016), which proposed that ~10% of current transcript
1109 models would result in incorrect translation initiation and incorrect encoded peptides.
1110

1111 **P-Type ATPase Tree Assembly**

1112 Protein sequences of diverse P-type ATPases (Thever and Saier, 2009) were downloaded from
1113 the National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/). NCBI
1114 sequences were combined with six P-type ATPases found in *Chlamydomonas* for a total of 259
1115 sequences. Sequence alignment was performed using ClustalW and a phylogenetic tree
1116 created using FastTree2 (http://www.microbesonline.org/fasttree/).
1117

1118 **GO Term Analysis**

1119 HCIPs of baits that localized to either the chloroplast, mitochondria, nucleus, ER/extracellular or
1120 PM were analyzed for GO-term enrichment using the Cytoscape plugin, BINGO
1121 (https://www.psb.ugent.be/cbd/papers/BiNGO/Home.html). Preys also included some baits that
1122 were detected as HCIPs of other baits. The GO-term, "Generation of precursor metabolites and
1123 energy" was shortened to "metabolite and energy production" in Figure 5.
1124

1125 **Transmembrane and Protein Disorder Prediction**

1126 Protein transmembrane regions were predicted using TMHMM 2.0
1127 (http://www.cbs.dtu.dk/services/TMHMM/). The percentage of protein disorder was predicted
1128 using ESpritz v1.3 (http://protein.bio.unipd.it/espritz/) with the prediction type set to Disprot and
1129 decision threshold set to Best Sw.
1130

1131 **Pyrenoid Enrichment Analysis**

1132 To determine whether the pyrenoid showed selectivity regarding protein size we categorized
1133 chloroplast localized proteins into pyrenoid depleted or not pyrenoid depleted. The "all other
1134 localizations" included all non-chloroplast proteins.
1135
1136 **Statistical tests**
1137 All statistical tests were performed in SPSS or Microsoft Excel.
1138
1139 **DATA AND SOFTWARE AVAILABILITY**
1140 The computer code used for primer design is available at https://github.com/Jonikas-
1141 Lab/tagging_primer_design. The raw mass spectrometry data is available from PRIDE XXXX.
1142 Plasmid sequences in GenBank or Fasta format for the constructs generated in this study can
1143 be downloaded from: https://sites.google.com/site/chlamyspatialinteractome/ or Mendeley Data:
1144 http://dx.doi.org/10.17632/k5m9fd8nzw.1.
1145
1146 **ADDITIONAL RESOURCES**
1147 Protein localization images, z-stacks and an interactive protein-protein interaction network are
1148 available at: https://sites.google.com/site/chlamyspatialinteractome/.