

1 The Structural Basis of Rubisco Phase Separation in the Pyrenoid

2

3 Shan He¹, Hui-Ting Chou², Doreen Matthies², Tobias Wunder³, Moritz T. Meyer¹, Nicky
4 Atkinson⁴, Antonio Martinez-Sanchez^{5,6}, Philip D. Jeffrey¹, Sarah A. Port¹, Weronika Patena¹,
5 Guanhua He¹, Vivian K. Chen⁷, Frederick M. Hughson¹, Alistair J. McCormick⁴, Oliver Mueller-
6 Cajar³, Benjamin D. Engel^{5,8}, Zhiheng Yu², Martin C. Jonikas^{1,*}

7

8 ¹Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. ²Janelia
9 Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA 20147,
10 USA. ³School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive,
11 Singapore 637551, Singapore. ⁴SynthSys & Institute of Molecular Plant Sciences, School of
12 Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, UK. ⁵Department of

13 Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried,
14 Germany. ⁶Institute of Neuropathology, University of Göttingen Medical Center, 37075
15 Göttingen, Germany. ⁷Department of Biology, Stanford University, Stanford, CA 94305, USA.
16 ⁸Helmholtz Pioneer Campus, Helmholtz Zentrum München, 85764 Neuherberg, Germany.

17 *Correspondence and Lead Contact: mjonikas@princeton.edu.

18

19 **Abstract**

20 Approximately one-third of global CO₂ fixation occurs in a phase separated algal organelle called
21 the pyrenoid. Existing data suggest that the pyrenoid forms by the phase-separation of the CO₂-
22 fixing enzyme Rubisco with a linker protein; however, the molecular interactions underlying this
23 phase-separation remain unknown. Here we present the structural basis of the interactions between
24 Rubisco and its intrinsically disordered linker protein EPYC1 (Essential Pyrenoid Component 1)
25 in the model alga *Chlamydomonas reinhardtii*. We find that EPYC1 consists of five evenly-spaced
26 Rubisco-binding regions that share sequence similarity. Single-particle cryo-electron microscopy
27 of one of these regions in complex with Rubisco indicates that each Rubisco holoenzyme has eight
28 binding sites for EPYC1, one on each Rubisco small subunit. Interface mutations disrupt binding,
29 phase separation, and pyrenoid formation. Cryo-electron tomography supports a model where
30 EPYC1 and Rubisco form a co-dependent multivalent network of specific low-affinity bonds,
31 giving the matrix liquid-like properties. Our results advance the structural and functional
32 understanding of the phase separation underlying the pyrenoid, an organelle that plays a
33 fundamental role in the global carbon cycle.

34 **Main Text**

35 The CO₂-fixing enzyme Rubisco drives the global carbon cycle, mediating the assimilation
36 of approximately 100 gigatons of carbon per year¹. The gradual decrease of atmospheric CO₂ over
37 billions of years² has made Rubisco's job increasingly difficult, to the point where CO₂
38 assimilation limits the growth rate of many photosynthetic organisms³. This selective pressure is
39 thought to have led to the evolution of CO₂ concentrating mechanisms, which feed concentrated
40 CO₂ to Rubisco to enhance growth⁴. Of these mechanisms, the most poorly understood relies on
41 the pyrenoid, a phase separated organelle⁵ found in the chloroplast of nearly all eukaryotic algae
42 and some land plants (Fig. 1a, b)^{6,7}. The pyrenoid enhances the activity of Rubisco by clustering
43 it around modified thylakoid membranes that supply Rubisco with concentrated CO₂^{8,9}.

44 For decades, the mechanism for packaging the Rubisco holoenzyme into the pyrenoid
45 remained unknown. Recent work showed that in the leading model alga *Chlamydomonas*
46 *reinhardtii* (Chlamydomonas hereafter), the clustering of Rubisco into the pyrenoid matrix
47 requires the Rubisco-binding protein EPYC1¹⁰. EPYC1 and Rubisco are the most abundant
48 components of the pyrenoid and bind to each other. Moreover, combining purified EPYC1 and
49 Rubisco together produces phase-separated condensates¹¹ that mix internally at a rate similar to
50 that observed for the matrix *in vivo*⁵, suggesting that these two proteins are sufficient to form the
51 structure of the pyrenoid matrix. The sequence repeats within EPYC1 and eight-fold symmetry of
52 the Rubisco holoenzyme led us to hypothesize that EPYC1 and Rubisco each have multiple
53 binding sites for the other, allowing the two proteins to form a co-dependent condensate (Fig. 1c)¹⁰.

54 Here, we determined the structural basis that underlies the EPYC1-Rubisco condensate.
55 Using biophysical approaches, we found that EPYC1 has five evenly spaced Rubisco-binding
56 regions that share sequence homology and can bind to Rubisco as short peptides. We obtained a
57 cryo-electron microscopy structure, which shows that each of EPYC1's Rubisco-binding regions
58 forms an α -helix that binds one of Rubisco's eight small subunits via salt bridges and hydrophobic
59 interactions. Mapping of these binding sites onto Rubisco holoenzymes within the native pyrenoid
60 matrix indicates that the linker sequences between Rubisco-binding regions on EPYC1 are
61 sufficiently long to connect together adjacent Rubisco holoenzymes. These discoveries advance
62 the understanding of the pyrenoid, and provide a high resolution structural view of a phase-
63 separated organelle.

64

65 **Results**

66 **EPYC1 has five nearly-identical Rubisco-binding regions**

67 We could not directly determine the structure of full-length EPYC1 bound to Rubisco because
68 mixing the two proteins together produces phase separated condensates¹¹. We thus aimed to first
69 identify Rubisco-binding regions on EPYC1, and subsequently to use a structural approach to
70 determine how these regions bind to Rubisco.

71 The intrinsically disordered nature of purified EPYC1¹¹ led us to hypothesize that the
72 Rubisco-binding regions of EPYC1 were short and could bind to Rubisco as peptides without a
73 need for tertiary folds. Therefore, to identify EPYC1 regions that bind to Rubisco, we synthesized
74 a peptide array consisting of 18 amino acid peptides tiling across the full length EPYC1 sequence
75 (Fig. 1d), and probed this array with native Rubisco purified from Chlamydomonas cells (Fig. 1e,
76 f).

77 Our tiling array revealed five evenly-spaced Rubisco-binding regions on EPYC1, each
78 consisting of a predicted α -helix and an upstream region (Fig. 1g, h). We confirmed the binding
79 regions using surface plasmon resonance (SPR; Extended Data Fig. 1b, c). Sequence alignment
80 guided by the five binding regions revealed that mature EPYC1 consists entirely of five sequence
81 repeats (Fig. 1i), in contrast to the previously defined four repeats and two termini¹⁰ (Extended
82 Data Fig. 1a). Our alignment indicates that the previously defined EPYC1 N- and C- termini, which
83 at the time were not considered part of the repeats, actually share sequence homology with the
84 central repeats.

85 The presence of a Rubisco-binding region on each of the previously defined EPYC1 repeats
86 (Extended Data Fig. 1a) explains our yeast two-hybrid observations¹² that a single EPYC1 repeat
87 can interact with Rubisco, that knocking out the α -helix in an EPYC1 repeat disrupts this
88 interaction, and that decreasing the number of EPYC1 repeats leads to a proportional decrease in
89 EPYC1 interaction with Rubisco. It also explains our observations that decreasing the number of
90 EPYC1 repeats leads to a proportional decrease in the tendency of EPYC1 and Rubisco to phase
91 separate together¹¹.

92

93 **EPYC1 binds to Rubisco small subunits**

94 The sequence homology of the five Rubisco-binding regions suggests that each region binds to
95 Rubisco in a similar manner. To identify the binding site of EPYC1 on Rubisco, we performed

96 single-particle cryo-electron microscopy on a complex of Rubisco and a peptide corresponding to
97 the first Rubisco-binding region of EPYC1 (Fig. 2a). We selected this region of EPYC1 because
98 in preliminary experiments it had the highest affinity to Rubisco, which was still low by protein
99 interaction standards ($K_D \sim 3$ mM; Extended Data Fig. 1d, e). This low affinity meant that
100 millimolar concentrations of peptide were required to approach full occupancy of peptide bound
101 to Rubisco, leading to challenges with peptide insolubility and high background signal in the
102 electron micrographs. Despite these challenges, we successfully obtained a 2.62 Å structure of the
103 complex (~2.9 Å EPYC1 peptide local resolution; Fig. 2, Extended Data Fig. 2 and 3; Extended
104 Data Table 1). For reference purposes, we also obtained a 2.68 Å cryo-electron density map of
105 Rubisco in the absence of EPYC1 peptide (Extended Data Fig. 2), which was nearly identical to
106 the previously published X-ray crystallography structure¹³, with minor differences likely due to
107 the absence of the substrate analog 2-CABP in the active site of Rubisco in our sample¹⁴ (Extended
108 Data Fig. 4).

109 The Rubisco holoenzyme consists of a core of eight catalytic large subunits in complex
110 with eight small subunits, four of which cap each end of the holoenzyme (Fig. 2b-e). In our
111 structure, an EPYC1 peptide was clearly visible bound to each Rubisco small subunit, suggesting
112 that each Rubisco holoenzyme can bind up to eight EPYC1s (Fig. 2b-e).

113

114 **Binding is mediated by salt bridges and a hydrophobic interface**

115 The EPYC1 peptide forms an extended chain that sits on top of the Rubisco small subunit's two
116 α -helices (Fig. 3a, b). The structure explains our previous observations that mutations in the
117 Rubisco small subunit α -helices disrupted yeast two-hybrid interactions between EPYC1 and the
118 Rubisco small subunit¹² and prevented Rubisco's assembly into a pyrenoid *in vivo*¹⁵. The C-
119 terminal region of the EPYC1 peptide (NWRQELESLRN) is well-resolved and forms an α -helix
120 that runs parallel to helix B of the Rubisco small subunit (Fig. 3a, b). The peptide's N-terminus
121 extends the trajectory of the helix and follows the surface of the Rubisco small subunit (Fig. 2b-e,
122 3a-b and Extended Data Fig. 5a). The side chains of the peptide's N-terminus could not be well
123 resolved, suggesting that this region is more conformationally flexible.

124 Our atomic model based on the density map suggests that binding is mediated by salt
125 bridges and a hydrophobic interface. Three residue pairs likely form salt bridges (Fig. 3c, d and
126 g): EPYC1 residues R64 and R71 interact with E24 and D23, respectively, of Rubisco small

127 subunit α -helix A, and EPYC1 residue E66 interacts with R91 of Rubisco small subunit α -helix B.
128 Furthermore, a hydrophobic interface is formed by W63, L67 and L70 of EPYC1 and M87, L90
129 and V94 of Rubisco small subunit helix B (Fig. 3e-g).

130

131 **Interface residues are required for binding and phase separation *in vitro***

132 To determine the importance of individual EPYC1 residues for binding, we investigated the impact
133 on Rubisco binding of every possible single amino acid substitution for EPYC1's first Rubisco-
134 binding region by using a peptide array (Fig. 4a) and SPR (Extended Data Fig. 5b). Consistent
135 with our structural model, the peptide array indicated that EPYC1 salt bridge-forming residues
136 R64, R71 and E66 and the hydrophobic interface residues W63, L67 and L70 were all required for
137 normal EPYC1 binding to Rubisco. The strong agreement of our mutational analysis suggests that
138 our structural model correctly represents EPYC1's Rubisco-binding interface.

139 To determine the importance of EPYC1's Rubisco-binding regions for pyrenoid matrix
140 formation, we assayed the impact of mutations in these regions on formation of phase separated
141 droplets by EPYC1 and Rubisco *in vitro*. The phase boundary was shifted by mutating R64 in the
142 first Rubisco-binding region and the corresponding K or R in the other four Rubisco-binding
143 regions of EPYC1 (Fig. 4b and Extended Data Fig. 5c-e), suggesting that the Rubisco-binding
144 regions mediate condensate formation.

145

146 **The binding interface is required for pyrenoid matrix formation *in vivo***

147 We validated the importance of Rubisco residues for binding to EPYC1 by yeast two-hybrid assays
148 (Fig. 5a and Extended Data Fig. 6). Rubisco small subunit D23A mutation, which eliminates the
149 charge of that residue, had a severe impact on Rubisco small subunit interaction with EPYC1, as
150 expected from the contribution of that residue to a salt bridge with R71 of EPYC1. Likewise, E24A
151 and R91A each showed a moderate defect, consistent with the contributions of those residues to
152 salt bridges with R64 and E66 of EPYC1, respectively. Additionally, M87D and V94D, which
153 convert hydrophobic residues to bulky charged residues, each had a severe impact on interaction,
154 as expected from the participation of those residues in the hydrophobic interface. Combinations of
155 these mutations abolished the interactions completely (Extended Data Fig. 6).

156 To evaluate the importance of the binding interface *in vivo*, we generated Chlamydomonas
157 strains with point mutations in the binding interface. Rubisco small subunit mutations D23A/E24A

158 or M87D/V94D caused a growth defect under conditions requiring a functional pyrenoid (Fig. 5b,
159 Extended Data Fig. 7a-b). Furthermore, the mutants lacked a visible pyrenoid matrix (Fig. 5c, d
160 and Extended Data Fig. 7c), indicating that those Rubisco small subunit residues are required for
161 matrix formation *in vivo*. The Rubisco mutants retained pyrenoid tubules¹⁶, as previously observed
162 in other matrix-deficient mutants^{10,17-19}.

163 Together, our data demonstrate that EPYC1's Rubisco-binding regions bind to the Rubisco
164 small subunit α -helices via salt-bridge interactions and a hydrophobic interface, enabling the
165 condensation of Rubisco into the phase separated matrix.

166

167 **The spacing between EPYC1's Rubisco-binding regions allows linking of adjacent Rubisco 168 holoenzymes in the native pyrenoid matrix**

169 The presence of multiple Rubisco-binding regions along the EPYC1 sequence supports a model
170 where consecutive Rubisco-binding regions on the same EPYC1 polypeptide can bind to different
171 Rubisco holoenzymes and thus hold them together to form the pyrenoid matrix. If this model is
172 correct, we would expect that the ~40 amino acid "linker" regions between consecutive Rubisco-
173 binding regions on EPYC1 (Fig. 1g, i) would be sufficient to span the distance between EPYC1-
174 binding sites on neighboring Rubisco holoenzymes in the pyrenoid matrix. To test this aspect of
175 the model, we combined our atomic structure of the EPYC1-Rubisco interaction with the precise
176 positions and orientations of Rubisco holoenzymes within the pyrenoid matrix of native cells that
177 we had previously obtained by *in-situ* cryo-electron tomography⁵ (Fig. 6a, b). We mapped the
178 positions of EPYC1 binding sites onto Rubisco holoenzymes in the matrix and measured the
179 distances between nearest neighbor EPYC1 binding sites on adjacent holoenzymes (Fig. 6c). The
180 observed distances ranged from ~2 nm to ~7 nm, with a median distance of ~4 nm (Fig. 6d).

181 A "linker" region of 40 amino acids is unlikely to be stretched to its maximum possible
182 length of 14 nm *in vivo* due to the high entropic cost of this configuration. To determine whether
183 a linker region can span the observed distances between nearest neighbor binding sites on adjacent
184 Rubisco holoenzymes, we used a simple physics model to calculate the energy required to stretch
185 a 40 amino acid chain to any given distance (Fig. 6d; see Methods). The model indicates that
186 stretching the chain to ~7 nm requires an energy of $3 k_B T$ (where k_B is the Boltzmann constant and
187 T is the temperature), which could reasonably be borrowed from thermal fluctuations. Thus, our
188 data suggest that the linker region between consecutive Rubisco-binding sites on the EPYC1

189 polypeptide can span the distance between adjacent Rubisco holoenzymes to hold the pyrenoid
190 matrix together *in vivo*. It also appears likely that, in addition to bridging neighboring Rubisco
191 holoenzymes, consecutive Rubisco-binding regions on a given EPYC1 may bind multiple sites on
192 one Rubisco holoenzyme, as the distance between the nearest binding sites on the same
193 holoenzyme is < 9 nm.

194

195 **Discussion**

196 **Our data explain the structural basis of Rubisco condensation into a pyrenoid matrix**

197 In this study, we have determined the structural basis for pyrenoid matrix formation for the first
198 time in any species. We found that in the model alga Chlamydomonas, the intrinsically disordered
199 protein EPYC1 has five regions of similar sequence that can bind to Rubisco as short peptides.
200 These EPYC1 regions form an α -helix that binds to the Rubisco small subunit α -helices via salt
201 bridges and hydrophobic interactions. EPYC1's Rubisco-binding regions are spaced by linker
202 sequences that are sufficiently long to span the distance between binding sites on adjacent Rubisco
203 holoenzymes within the pyrenoid, allowing EPYC1 to serve as a molecular glue that clusters
204 Rubisco together to form the pyrenoid matrix (Fig. 6e).

205 The multivalency of EPYC1 and the high K_D (~3 mM; Extended Data Fig. 1e) of individual
206 Rubisco-binding regions are consistent with the emerging principle that cellular phase separation
207 is mediated by weak multivalent interactions²⁰. The high dissociation rate constant (>1/s; Extended
208 Data Fig. 1d) of individual Rubisco-binding regions explains how the pyrenoid matrix can mix
209 internally on the time scale of seconds⁵ despite the multivalency of EPYC1. The even spacing of
210 the five Rubisco-binding regions across EPYC1 is noteworthy and may be an indication of
211 selective pressure for an optimal distance between binding regions, and thus of an optimal spacing
212 between Rubisco holoenzymes in the matrix.

213 Knowledge of the EPYC1-Rubisco binding mechanism now opens doors to the molecular
214 characterization of the regulation of this interaction, which may govern the dissolution and
215 condensation of the matrix during cell division⁵ and in response to environmental factors²¹. For
216 example, phosphorylation of EPYC1²² may provide a mechanism to rapidly change the binding
217 affinity of EPYC1 to Rubisco. Inactivation of one of EPYC1's five Rubisco-binding regions would
218 yield four binding regions, which would allow two EPYC1 molecules to form a mutually satisfied
219 complex with each Rubisco, a configuration that is predicted to favor dissolution of the matrix⁵.

220

221 **The Rubisco-EPYC1 structure explains how other key pyrenoid proteins bind to Rubisco**

222 In a parallel study (Meyer et al., please see unpublished manuscript provided as reference
223 material), we recently discovered that a common sequence motif is present on many pyrenoid-
224 localized proteins. The motif binds Rubisco, enabling recruitment of motif-containing proteins to
225 the pyrenoid and mediating adhesion between the matrix, pyrenoid tubules, and starch sheath. This
226 motif, [D/N]W[R/K]XX[L/I/V/A], is serendipitously present in EPYC1's Rubisco-binding
227 regions, and the motif residues mediate key binding interactions with Rubisco. In our structure,
228 the R/K of the motif is represented by R64 of EPYC1, which forms a salt bridge with E24 of the
229 Rubisco small subunit. The XX of the motif almost always includes a D or E; in our structure this
230 feature is represented by E66 of EPYC1, which forms a salt bridge with R91 of the Rubisco small
231 subunit. Finally, the W and [L/I/V/A] of the motif are represented by W63 and L67 of EPYC1,
232 which contribute to the hydrophobic interactions with M87, L90 and V94 of the Rubisco small
233 subunit. The key roles of the motif residues in the interface presented here strongly suggest that
234 the structure we have obtained for one motif from EPYC1 also explains where and how all other
235 variants of the motif, including those found on the key pyrenoid proteins SAGA1, SAGA2,
236 RBMP1, RBMP2 and CSP41A, bind to Rubisco. Our observation that the Rubisco small subunit
237 D23A/E24A and M87D/V94D mutants exhibit a more severe disruption of the pyrenoid than the
238 *epyc1* mutant¹⁰ supports the idea that this region of Rubisco interacts not only with EPYC1, but
239 also with other proteins required for pyrenoid biogenesis, making this binding interaction a central
240 hub of pyrenoid biogenesis.

241

242 **There are structural similarities and differences between the pyrenoid matrix and bacterial**
243 **carboxysomes**

244 Although α - and β -carboxysomes are morphologically, functionally and evolutionarily distinct
245 from the pyrenoid, their Rubisco is also thought to be clustered by linker proteins. Like EPYC1,
246 the α -carboxysome linker protein CsoS2²³ is intrinsically disordered and is proposed to bind
247 Rubisco as an unfolded peptide²⁴. In contrast, the β -carboxysome linker protein CcmM has been
248 proposed to bind Rubisco using folded globular domains^{25,26}. The use of an unfolded peptide as in
249 the case of EPYC1 and CsoS2 may provide the benefit of requiring fewer amino acids for
250 achieving the desired binding function. A notable difference is the location of the binding site on

251 Rubisco: whereas both carboxysomal linker proteins bind to the interface between two Rubisco
252 large subunits^{24,26}, EPYC1 binds to the Rubisco small subunit. It remains to be seen whether this
253 difference in binding site has functional consequences, such as impacts on the three-dimensional
254 packing of Rubisco.

255

256 **Our findings advance the ability to engineer a pyrenoid into crops**

257 There is currently significant interest in engineering Rubisco condensates into monocotyledonous
258 crops such as wheat and rice to enhance yields²⁷⁻³⁰. Binding of EPYC1 to the Rubisco small subunit
259 presents a promising route for engineering a Rubisco condensate, as the Rubisco small subunit is
260 encoded in the nuclear genome, making it more easily amenable to genetic modification in those
261 crops than the chloroplast-encoded Rubisco large subunit³¹. Knowledge of the binding mechanism
262 now allows engineering of minimal sequence changes into native crop Rubiscos to enable binding
263 to EPYC1 and to other key proteins required to reconstitute a functional pyrenoid.

264

265 **Our work provides insights into pyrenoid matrix formation in other species**

266 Pyrenoids appear to have evolved independently in different lineages through convergent
267 evolution^{7,32}. EPYC1, its Rubisco-binding sequences, and the amino acid residues that form the
268 EPYC1 binding site on the surface of Rubisco are conserved across the order Volvocales, as
269 evidenced from the genome sequences of *Tetraebaena socialis*, *Gonium pectorale* and *Volvox*
270 *carteri* (Extended Data Table 2). While the molecular mechanisms of matrix formation in other
271 lineages remain to be uncovered, candidate linker proteins have been identified based on similarity
272 of sequence properties to EPYC1¹⁰. We hypothesize that the matrix in other lineages is formed
273 based on similar principles to those we observed in Chlamydomonas. Our experimental approach
274 for characterizing the binding interaction provides a roadmap for future structural studies of
275 pyrenoids across the tree of life.

276

277 **This study provides a high-resolution structural view of a phase separated organelle**

278 The pyrenoid matrix presents an unusual opportunity to study a two-component molecular
279 condensate where one of the components, Rubisco, is large and rigid, and the other component,
280 EPYC1, is a simple intrinsically disordered protein consisting of nearly identical tandem repeats.
281 The rigidity and size of Rubisco holoenzymes previously enabled the determination of their

282 positions and orientations within the pyrenoid matrix of native cells by cryo-electron tomography⁵.
283 The identification of EPYC1 binding sites on Rubisco in the present work and the modeling of
284 linker regions between EPYC1's Rubisco binding regions now make the Chlamydomonas
285 pyrenoid matrix one of the most structurally well-defined phase separated organelles. Thus,
286 beyond advancing our structural understanding of pyrenoids, organelles that play a central role in
287 the global carbon cycle, we hope that the findings presented here will also more broadly enable
288 advances in the biophysical understanding of phase separated organelles.

289 **Methods**

290 **Strains and culture conditions**

291 Chlamydomonas wild-type (WT) strain cMJ030 was maintained in the dark or low light (~10 μmol
292 photons $\text{m}^{-2} \text{s}^{-1}$) on 1.5% agar plates containing Tris-Acetate-Phosphate medium with revised trace
293 elements³³. For Rubisco extraction, 500 mL Tris-Acetate-Phosphate medium in a 1 L flask was
294 inoculated with a loopful of cells and the culture was grown to 4×10^6 cells/mL at 22°C, shaking
295 at 200 rpm under ~100 μmol photons $\text{m}^{-2} \text{s}^{-1}$ white light in 3% CO₂. Chlamydomonas mutant T60-
296 3³⁴ ($\Delta rbcS$; containing a deletion of both *RBCS* genes) was used for generating Rubisco small
297 subunit point mutants and a wild-type control in the same background. This strain was maintained
298 on agar in the dark or low light (~10 μmol photons $\text{m}^{-2} \text{s}^{-1}$).

299

300 **Protein extraction**

301 Rubisco was purified from Chlamydomonas strain cMJ030³⁵. Cells were disrupted by
302 ultrasonication in lysis buffer (10 mM MgCl₂, 50 mM Bicine, 10 mM NaHCO₃, 1 mM
303 dithiothreitol, pH 8.0) supplemented with Halt Protease Inhibitor Cocktail, EDTA-Free (Thermo
304 Fisher Scientific). The soluble lysate was fractionated by ultracentrifugation on a 10-30% sucrose
305 gradient in a SW 41 Ti rotor at a speed of 35,000 rpm for 20 hours at 4°C. Rubisco-containing
306 fractions were applied to an anion exchange column (MONO Q 5/50 GL, GE Healthcare) and
307 eluted with a linear salt gradient from 30 to 500 mM NaCl in lysis buffer.

308

309 **Peptide arrays**

310 Peptide arrays were purchased from the MIT Biopolymers Laboratory (Cambridge, MA). The
311 tiling array was composed of 18-amino-acid peptides that tiled across the full-length EPYC1
312 sequence with a step size of one amino acid. Each peptide was represented by at least two spots
313 on the array, and these replicates were averaged during data analysis. The locations of peptides on
314 the array were randomized. In the substitution arrays, peptides were designed to represent every
315 possible one-amino-acid mutation of the indicated region on EPYC1 by substitution with one of
316 the other 19 amino acids. The arrays were activated by methanol, then washed 3x10 min in binding
317 buffer (50 mM HEPES, 50 mM KOAc, 2 mM Mg(OAc)₂.4H₂O, 1 mM CaCl₂ and 200 mM sorbitol,
318 pH 6.8). The arrays were then incubated at 4°C with 1 mg purified Rubisco overnight. The arrays
319 were washed in binding buffer to remove any unbound Rubisco. Using a semi-dry transfer

320 apparatus (BIO-RAD), bound Rubisco was transferred onto an Immobilon-P PVDF membrane
321 (Millipore Sigma). Rubisco was immuno-detected with a polyclonal primary antibody raised
322 against Rubisco¹⁵ (1:10,000) followed by a HRP conjugated goat anti-rabbit (1:20,000;
323 Invitrogen). Arrays were stripped with RestoreTM Western Blot Stripping Buffer before re-use
324 (Thermo Fisher Scientific).

325

326 **Surface plasmon resonance (SPR) experiments**

327 All the surface preparation experiments were performed at 25°C using a Biacore 3000 instrument
328 (GE Healthcare). Purified Rubisco was immobilized on CM5 sensor chips using a Biacore Amine
329 Coupling Kit according to the manufacturer's instructions. Briefly, the chip surface was activated
330 by an injection of 1:1 N-hydroxysuccinimide (NHS)/1-ethyl-3-(3-
331 dimethylaminopropyl)carbodiimide hydrochloride (EDC). Rubisco was diluted to ~100 µg/mL in
332 10 mM acetate (pH 4.5; this pH had been previously optimized using the immobilization pH
333 scouting wizard) and was injected over the chip surface. Excess free amine groups were then
334 capped with an injection of 1 M ethanolamine. Typical immobilization levels were 8,000 to 10,000
335 resonance units (RU), as recommended for binding experiments of small molecules. For kinetic
336 experiments (for determining the binding affinities), the typical immobilization levels were ~5,000
337 RU. The control surfaces were prepared in exactly the same manner as the experimental surfaces
338 except that no Rubisco was injected. For immobilizations, the running buffer was the Biacore
339 HBS-EP Buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20).

340 All the binding assays were performed using the Biacore PBS-P+ Buffer (20 mM
341 phosphate buffer, 2.7 mM KCl, 137 mM NaCl and 0.05% Surfactant P20, pH 6.8) as a running
342 buffer, as recommended for small molecule analysis in Biacore systems. The analytes, consisting
343 of EPYC1 peptides synthesized by Genscript (Piscataway, New Jersey), were dissolved in the
344 same running buffer and diluted to 1 mM. The analytes were injected over the control surface and
345 experimental surfaces at a flow rate of 26 µL/min for 2.5 minutes, followed by 2.5 minutes of the
346 running buffer alone to allow for dissociation. The surfaces were then regenerated using running
347 buffer at a flow rate of 30 µL/min for 10 minutes. In all cases, binding to the control surface was
348 negligible.

349 For determining the K_D of EPYC1 peptide, the kinetic assays were performed with a
350 running buffer consisting of 200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM

351 Mg(OAc)₂.4H₂O and 1 mM CaCl₂ at pH 6.8 (the same buffer as the peptide array assay). The
352 EPYC1 peptide was dissolved in the same running buffer as the assay and the serial dilutions were
353 also made in the same buffer. The analytes were injected over the control surface and experimental
354 surfaces at a flow rate of 15 μ L/min for 2 minutes, followed by 10 minutes with the running buffer
355 alone to allow for dissociation. The surfaces were then regenerated by the running buffer at a flow
356 rate of 30 μ L/min for 10 minutes. In all cases, binding to the blank chip was negligible. The fitting
357 and modeling were performed with the BIAevaluation software.

358

359 **Single-particle cryo-electron microscopy data collection and image processing**

360 Rubisco and peptide with the final concentration of 1.69mg/ml (=3.02 μ M) and 7.5 mM were
361 incubated on ice for 20 minutes in buffer consisting of 200 mM sorbitol, 50 mM HEPES, 50 mM
362 KOAc, 2 mM Mg(OAc)₂.4H₂O and 1 mM CaCl₂ at pH 6.8 (the same buffer as the peptide array
363 assay and the SPR binding assay). For both apo Rubisco and Rubisco incubated with peptide,
364 similar cryo grid-making procedures were used. 400-mesh Quantifoil 1.2/1.3 Cu grids (Quantifoil,
365 Großlobichau, Germany) were made hydrophilic by glow discharging for 60 seconds with a
366 current of 15 mA in a Pelico EasiGlow system. Samples on cryo grids were plunge-frozen using
367 an FEI Mark IV Vitrobot (FEI company, part of Thermo Fisher Scientific, Hillsboro, OR). The
368 chamber of the Vitrobot was kept at 4°C and 100% relative humidity. 3 μ l of sample was applied
369 to the glow-discharged grid, blotted with filter paper for 3 seconds with the equipment-specific
370 blotting force set at 3. After blotting, the grid was rapidly plunge-frozen into a liquid ethane bath.

371 Cryo grids were loaded into a 300 kV FEI Titan Krios cryo electron microscope (FEI
372 Company) at HHMI Janelia Research Campus, Janelia Krios2, equipped with a Gatan K2 Summit
373 camera. After initial screening and evaluation, fully automated data collection was carried out
374 using SerialEM. The final exposure from each collection target was collected as a movie utilizing
375 dose fractionation on the K2 Summit camera operated in super-resolution mode. The movie was
376 collected at a calibrated magnification of 38,168x, corresponding to 1.31 \AA per physical pixel in
377 the image (0.655 \AA per super-resolution pixel). The dose rate on the specimen was set to be 5.82
378 electrons per \AA^2 per second and total exposure time was 10 s, resulting in a total dose of 58.2
379 electrons per \AA^2 . With dose fractionation set at 0.2 s per frame, each movie series contained 50
380 frames and each frame received a dose of 1.16 electrons per \AA^2 . The spherical aberration constant

381 of the objective lens is 2.7 mm and an objective aperture of 100 μ m was used. The nominal defocus
382 range for the automated data collection was set to be between -1.5 μ m and -3.0 μ m.

383 The movies were 2x binned and motion corrected using MotionCor2³⁶ and CTF was
384 estimated using CTFFIND³⁷ in Relion 3.0³⁸. The particles were selected using cisTEM³⁹ and
385 1,809,869 peptide bound Rubisco particles and 677,071 particles in the apo state were extracted
386 with a box size of 192x192pixels. 2D classification was performed using cisTEM 2D. The classes
387 presenting detailed features in class averages were chosen for 3D classification on cryoSPARC^{40,41}
388 for peptide-bound Rubisco and on Relion for the apo state. The 3D class showing clear secondary
389 structures was chosen for 3D auto-refine first without symmetry and then with D4 symmetry
390 imposed. After CTF refinement and Bayesian polishing in Relion, the reconstructed map
391 resolution is 2.68 \AA for the apo state and 2.62 \AA for the peptide bound state. Details for single-
392 particle cryo-EM data collection and image processing are included in the Extended Data Table 1.
393

394 **Single-particle cryo-electron microscopy model building, fitting, and refinement**

395 A full model for Rubisco from Chlamydomonas was produced from an X-ray structure¹³ (PDB
396 entry 1GK8) and used for rigid body fitting into a local resolution filtered cryo-EM map with an
397 average resolution of 2.62 \AA using UCSF Chimera⁴². After rigid body fitting of the full complex,
398 initial flexible fitting was performed in COOT⁴³ by manually going through the entire peptide
399 chain of a single large and small Rubisco subunit before applying the changes to the other seven
400 large and small subunits. The sequence of the peptide was used to predict secondary structure
401 elements using JPred4⁴⁴ which resulted in the prediction that the C-terminal region (NWRQELES)
402 is α -helical. Guided by this prediction, the peptide was built manually into the density using
403 COOT. Additional real space refinement of the entire complex was performed using Phenix⁴⁵.
404 Models were subjected to an all-atom structure validation using MolProbity⁴⁶. Figures were
405 produced using UCSF Chimera.

406

407 **Liquid–liquid phase separation assay**

408 Proteins used in the liquid–liquid phase separation assay were obtained and stored essentially as
409 described previously¹¹. Briefly, Rubisco was purified from *C. reinhardtii* cells (CC-2677 cw15
410 nit1-305 mt-5D, Chlamydomonas Resource Center) grown in Sueoka’s high-salt medium⁴⁷, using
411 a combination of anion exchange chromatography and gel filtration.

412 The EPYC1 full-length gene (encoding amino acids 1-317) and corresponding R/K mutant
413 (EPYC1^{R64A/K127A/K187A/K248A/R314A}) were synthesized by GenScript and cloned between the SacII
414 and HindIII site of the pHue vector⁴⁸. Proteins were produced in the *E. coli* strain BL21 (DE3)
415 harbouring pBADES⁴⁹ for co-expression of the *E. coli* chaperonin GroEL/S. The purification
416 was conducted with minor changes (dialysis for removal of high imidazol concentrations was
417 skipped by running the gel-filtration column before the second IMAC). After the first IMAC step
418 and cleavage⁵⁰ of the N-terminal His₆-ubiquitin tag, proteins were separated by gel filtration.
419 Finally, the peak fraction was passed a second time through an IMAC column, collecting EPYC1
420 from the flow through.

421 EPYC1-Rubisco condensates were reconstituted *in vitro* in a buffer containing 20 mM
422 Tris-HCl (pH 8.0) and NaCl concentrations as indicated. 5 µl reactions were incubated for 3 min
423 at room temperature before monitoring the droplet formation by differential interference contrast
424 (DIC) microscopy. DIC images were acquired with a Nikon Eclipse Ti Inverted Microscope using
425 a 60× oil-immersion objective after allowing the droplets to settle on the coverslip (Superior
426 Marienfeld, Germany) surface for about 3 min. For droplet sedimentation assays 10 µl reactions
427 were incubated for 3 min at 20°C before separating the droplets from the bulk phase by spinning
428 for 3 min at 21,000xg and 4°C. Pelleted droplets and supernatant fractions were analyzed using
429 Coomassie-stained SDS-PAGE.

430

431 **Yeast two-hybrid assay**

432 Yeast two-hybrid to detect interactions between EPYC1 and RbcS1 was carried out as described
433 previously¹². EPYC1 was cloned into the two-hybrid vector pGBK7 to create a fusion with the
434 GAL4 DNA binding domain. Point mutations were introduced by PCR into RbcS1, which was
435 then cloned in the pGADT7 to create a fusion with the GAL4 activation domain. Yeast cells were
436 then co-transformed with binding and activation domain vectors. Successful transformants were
437 cultured, diluted to an optical density at 600 nm (OD600) of 0.5 or 0.1, and plated onto SD-L-W
438 and SD-L-W-H containing increasing concentrations of the HIS3 inhibitor triaminotriazole (3-
439 AT). Plates were imaged after 3 days. Spots shown in Fig. 5a were grown at 5 mM 3-AT from a
440 starting OD600 of 0.5; they are a subset of the full dataset shown in Extended Data Fig. 6.

441

442 **Cloning of Rubisco small subunit point mutants**

443 The plasmid pSS1-ITP⁵¹ which contains Chlamydomonas *RBCS1* including UTRs and introns 1
444 and 2 was used as a starting point for generating plasmids pSH001 and pSH002, which encode
445 *RBCS1*^{D23A/E24A}, and *RBCS1*^{M87D/V94D}, respectively. The point mutations were generated by
446 Gibson assembly⁵² of gBlocks (synthesized by Integrated DNA Technologies) containing the
447 desired mutations into pSS-ITP that had been enzyme digested by restriction endonucleases (XcmI
448 and BbvCI for the D23A/E24A mutations and BbvCI and BlpI for the M87D/V94D mutations).
449 All constructs were verified by Sanger sequencing.

450 The fragment for making pSH001 (containing the D23A/E24A Rubisco small subunit
451 mutant) had the following sequence:

452 GCAGGGCTGCCCGGCTCAGGCCAACAGATGATGGTCTGGACCCGGTCAACAAAC
453 AAGATGTTCGAGACCTCTCCTACCTGCCTCCTTGACCGCCGCGCAGATGCCGCC
454 CAGGTCGACTACATCGTCGCCAACGGCTGGATCCCCTGCCTGGAGTCGCTGAGGCC
455 GACAAGGCCTACGTGTCCAAC

456 The fragment for making pSH002 (containing the M87D/V94D Rubisco small subunit
457 mutant) had the following sequence:

458 CTGCCTGGAGTCGCTGAGGCCGACAAGGCCTACGTGTCCAACGAGTCGGCCATCC
459 GCTTCGGCAGCGTGTCTGCCTGTACTACGACAACCGCTACTGGACCATGTGGAAGC
460 TGCCCATGTTGGCTGCCCGACCCCGACCAGGTGCTGCGCGAGATCGACGCCCTGCA
461 CCAAGGCCTTCCCCGATGCCTACGTGCGCCTGGTGGCCTTCGACAACCAGAAGCAG
462 GTGCAGATCATGGGCTCCTGGTCCAGCGCCCCAAGACTGCCCGCGACTTCCAGCCC
463 GCCAACAAAGCGCTCCGTAAATGGAGGCGCTCGATCTGAGCCGTGTGATGT
464 TTGTTGGTGTGAGCGAGTGCAATGAGAGTGTGTGTGTTGGTGTG
465 GCTAAGCCAAGCGTGATCGC

466

467 **Transformation of Chlamydomonas to make the Rubisco small subunit point mutants**

468 Chlamydomonas strains *Δrbcs;RBCS^{WT}*, *Δrbcs;RBCS^{D23A/E24A}*, and *Δrbcs;RBCS^{M87D/V94D}* were
469 generated by transforming pSS1-ITP, pSH001, and pSH002 (encoding Rubisco small subunit
470 constructs) into the Rubisco small subunit deletion mutant T60 (*Δrbcs*) by electroporation as
471 described previously⁵³. For each transformation, 29 ng kbp⁻¹ of KpnI linearized plasmid was mixed
472 with 250 μL of 2 x 10⁸ cells mL⁻¹ at 16°C and electroporated immediately. Transformant colonies
473 were selected on Tris-Phosphate plates without antibiotics at 3% v/v CO₂ under ~50 μmol photons

474 $\text{m}^{-2} \text{ s}^{-1}$ light. The sequence of RbcS in the transformants was verified by PCR amplification and
475 Sanger sequencing.

476

477 **Spot tests**

478 *Δrbcs;RBCS^{WT}*, *Δrbcs;RBCS^{D23A/E24A}*, and *Δrbcs;RBCS^{M87D/V94D}* were grown in Tris-Phosphate
479 medium at 3% CO₂ until $\sim 2 \times 10^6$ cells mL⁻¹. Cells were diluted in Tris-Phosphate medium to a
480 concentration of 8.7×10^7 cells mL⁻¹, then serially diluted 1:10 three times. 7.5 μL of each dilution
481 was spotted onto four TP plates and incubated in air or 3% CO₂ under 20 or 100 μmol photons m⁻
482 $^2 \text{s}^{-1}$ white light for 9 days before imaging.

483

484 **Transmission electron microscopy**

485 Samples for electron microscopy were fixed for 1 hour at room temperature in 2.5% glutaraldehyde
486 in Tris-Phosphate medium (pH 7.4), followed by 1 hour at room temperature in 1% OsO₄, 1.5%
487 K₃Fe(CN)₃, and 2 mM CaCl₂. Fixed cells were then bulk stained for 1 hour in 2% uranyl acetate,
488 0.05 M maleate buffer at pH 5.5. After serial dehydration (50%, 75%, 95%, and 100% ethanol,
489 followed by 100% acetonitrile), samples were embedded in epoxy resin containing 34% Quetol
490 651, 44% nonenyl succinic anhydride, 20% methyl-5- norbornene-2,3-dicarboxylic anhydride, and
491 2% catalyst dimethylbenzylamine. Ultramicrotomy was done by the Core Imaging Lab, Medical
492 School, Rutgers University. Imaging was performed at the Imaging and Analysis Center, Princeton
493 University, on a CM100 transmission electron microscope (Philips, Netherlands) at 80 kV.

494

495 **Measurement of nearest-neighbor distances between EPYC1-binding sites on Rubisco 496 holoenzymes within pyrenoids**

497 For detailed descriptions of the Chlamydomonas cell culture, vitrification of cells onto EM grids,
498 thinning of cells by cryo-focused ion beam milling, 3D imaging of native pyrenoids by cryo-
499 electron tomography, tomographic reconstruction, and subtomogram averaging, see our previous
500 study⁵. In that study, we measured the distances between the center positions of Rubisco
501 complexes within tomograms of five pyrenoids. The spatial parameters determined in that study
502 were combined with the EPYC1-binding sites resolved here by cryo-EM single-particle analysis
503 to measure the nearest-neighbor distances between EPYC1-binding sites on adjacent Rubisco
504 complexes within the native pyrenoid matrix.

505 The *in situ* subtomogram average EMD-3694⁵ was used as the reference for the Rubisco
506 model. We extracted the isosurface from this density using the 0.5 contour level recommended in
507 the Electron Microscopy Data Bank entry. We then fit the atomic model of EPYC1-bound Rubisco
508 (Fig. 2) within the EMD-3694 density, and for each EPYC1-binding site, we marked the closest
509 point on the isosurface to define the EPYC1 binding sites on this model. The positions and
510 orientations previously determined by subtomogram averaging were used to place each Rubisco
511 model and its corresponding binding sites into the pyrenoid tomograms using the PySeg program⁵⁴.

512 To compute the nearest-neighbor distances between EPYC1-binding sites on two different
513 Rubisco complexes, first, linkers were drawn between each EPYC1 binding site and all other
514 binding sites within 25 nm. Binding sites on the same Rubisco complex were ignored. Next, the
515 linkers were filtered by length (defined as the Euclidean distance between the two binding sites),
516 and only the shortest linker was retained for each binding site. To prevent edge effects, linkers
517 were discarded if they had a binding site <12 nm from the masked excluded volume (grey in Fig.
518 6b), which marks the border of the analyzed pyrenoid matrix. Finally, linker distances were plotted
519 in a histogram to show the distribution of lengths (normalized to 100%).

520

521 **Modeling of the energy required to stretch EPYC1-linker regions**

522 The energy required to stretch the linker regions between EPYC1's Rubisco-binding regions was
523 determined as follows. The force F required to stretch a 40 amino acid linker region to any given
524 length z was approximated using a wormlike chain model⁵⁵:

$$525 F(z) = \frac{k_B T}{4L_p} \left[\frac{1}{(1 - z/L_0)^2} - 1 + \frac{4z}{L_0} \right]$$

526 In the above equation, k_B is the Boltzmann constant, T is the temperature, L_p is the persistence
527 length (assumed to be 1 nm, a representative value for disordered proteins), and L_0 is the contour
528 length (estimated as 40 amino acids * 0.36 nm/amino acid). The energy required to stretch the
529 linker to a length x is given by:

$$530 E(x) = \int_0^x F(z) dz$$

531 This energy was calculated and plotted in Fig. 6d.

532

533 References

534 1 Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the
535 biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237-240 (1998).

536 2 Hessler, A. M., Lowe, D. R., Jones, R. L. & Bird, D. K. A lower limit for atmospheric carbon
537 dioxide levels 3.2 billion years ago. *Nature* **428**, 736-738 (2004).

538 3 Ainsworth, E. A. & Long, S. P. What have we learned from 15 years of free-air CO₂
539 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy
540 properties and plant production to rising CO₂. *New Phytol* **165**, 351-371, doi:10.1111/j.1469-
541 8137.2004.01224.x (2005).

542 4 Raven, J. A., Cockell, C. S. & De La Rocha, C. L. The evolution of inorganic carbon
543 concentrating mechanisms in photosynthesis. *Philosophical transactions of the Royal Society
544 of London. Series B, Biological sciences* **363**, 2641-2650, doi:10.1098/rstb.2008.0020 (2008).

545 5 Freeman Rosenzweig, E. S. *et al.* The Eukaryotic CO₂-Concentrating Organelle Is Liquid-
546 like and Exhibits Dynamic Reorganization. *Cell* **171**, 148-162 e119,
547 doi:10.1016/j.cell.2017.08.008 (2017).

548 6 Badger, M. R. *et al.* The diversity and coevolution of Rubisco, plastids, pyrenoids, and
549 chloroplast-based CO₂-concentrating mechanisms in algae. *Can J Bot* **76**, 1052-1071,
550 doi:10.1139/b98-074 (1998).

551 7 Villarreal, J. C. & Renner, S. S. Hornwort pyrenoids, carbon-concentrating structures, evolved
552 and were lost at least five times during the last 100 million years. *Proc Natl Acad Sci U S A*
553 **109**, 18873-18878, doi:10.1073/pnas.1213498109 (2012).

554 8 Wang, Y., Stessman, D. J. & Spalding, M. H. The CO₂ concentrating mechanism and
555 photosynthetic carbon assimilation in limiting CO₂ : how Chlamydomonas works against the
556 gradient. *The Plant journal : for cell and molecular biology* **82**, 429-448,
557 doi:10.1111/tpj.12829 (2015).

558 9 Raven, J. A. CO₂-concentrating mechanisms: A direct role for thylakoid lumen acidification?
559 *Plant Cell Environ* **20**, 147-154, doi:Doi 10.1046/J.1365-3040.1997.D01-67.X (1997).

560 10 Mackinder, L. C. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-
561 concentrating organelle. *Proc Natl Acad Sci U S A* **113**, 5958-5963,
562 doi:10.1073/pnas.1522866113 (2016).

563 11 Wunder, T., Cheng, S. L. H., Lai, S. K., Li, H. Y. & Mueller-Cajar, O. The phase separation
564 underlying the pyrenoid-based microalgal Rubisco supercharger. *Nature communications* **9**,
565 5076, doi:10.1038/s41467-018-07624-w (2018).

566 12 Atkinson, N. *et al.* The pyrenoidal linker protein EPYC1 phase separates with hybrid
567 Arabidopsis-Chlamydomonas Rubisco through interactions with the algal Rubisco small
568 subunit. *Journal of experimental botany* **70**, 5271-5285, doi:10.1093/jxb/erz275 (2019).

569 13 Taylor, T. C., Backlund, A., BJORHALL, K., Spreitzer, R. J. & Andersson, I. First crystal structure
570 of Rubisco from a green alga, Chlamydomonas reinhardtii. *J Biol Chem* **276**, 48159-48164,
571 doi:10.1074/jbc.M107765200 (2001).

572 14 Duff, A. P., Andrews, T. J. & Curmi, P. M. The transition between the open and closed states
573 of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J Mol Biol*
574 **298**, 903-916, doi:10.1006/jmbi.2000.3724 (2000).

575 15 Meyer, M. T. *et al.* Rubisco small-subunit α -helices control pyrenoid formation in
576 Chlamydomonas. *Proc Natl Acad Sci U S A* **109**, 19474-19479,
577 doi:10.1073/pnas.1210993109. (2012).

578 16 Engel, B. D. *et al.* Native architecture of the *Chlamydomonas* chloroplast revealed by in situ
579 cryo-electron tomography. *eLife* **4**, DOI: 10.7554/eLife.04889, doi:10.7554/eLife.04889
580 (2015).

581 17 Goodenough, U. W. & Levine, R. P. Chloroplast structure and function in ac-20, a mutant
582 strain of Chlamydomonas reinhardi. 3. Chloroplast ribosomes and membrane organization. *J
583 Cell Biol.* **44**, 547-562, doi:10.1083/jcb.44.3.547 (1970).

584 18 Ma, Y., Pollock, S. V., Xiao, Y., Cunnusamy, K. & Moroney, J. V. Identification of a novel
585 gene, CIA6, required for normal pyrenoid formation in Chlamydomonas reinhardtii. *Plant
586 Physiol* **156**, 884-896, doi:10.1104/pp.111.173922 (2011).

587 19 Caspari, O. D. *et al.* Pyrenoid loss in Chlamydomonas reinhardtii causes limitations in CO₂
588 supply, but not thylakoid operating efficiency. *Journal of experimental botany* **68**, 3903-3913,
589 doi:10.1093/jxb/erx197 (2017).

590 20 Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**,
591 336-340, doi:10.1038/nature10879 (2012).

592 21 Borkhsenious, O. N., Mason, C. B. & Moroney, J. V. The Intracellular Localization of
593 Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase in Chlamydomonas reinhardtii. *Plant
594 Physiol* **116**, 1585-1591, doi:10.1104/pp.116.4.1585 (1998).

595 22 Turkina, M. V., Blanco-Rivero, A., Vainonen, J. P., Vener, A. V. & Villarejo, A. CO₂
596 limitation induces specific redox-dependent protein phosphorylation in Chlamydomonas
597 reinhardtii. *Proteomics* **6**, 2693-2704, doi:10.1002/pmic.200500461 (2006).

598 23 Cai, F. *et al.* Advances in Understanding Carboxysome Assembly in Prochlorococcus and
599 Synechococcus Implicate CsoS2 as a Critical Component. *Life (Basel)* **5**, 1141-1171,
600 doi:10.3390/life5021141 (2015).

601 24 Oltrogge, L. M. *et al.* Multivalent interactions between CsoS2 and Rubisco mediate alpha-
602 carboxysome formation. *Nat Struct Mol Biol* **27**, 281-287, doi:10.1038/s41594-020-0387-7
603 (2020).

604 25 Long, B. M., Badger, M. R., Whitney, S. M. & Price, G. D. Analysis of carboxysomes from
605 Synechococcus PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins
606 CcmM and CcaA. *J Biol Chem* **282**, 29323-29335, doi:10.1074/jbc.M703896200 (2007).

607 26 Wang, H. *et al.* Rubisco condensate formation by CcmM in beta-carboxysome biogenesis.
608 *Nature* **566**, 131-135, doi:10.1038/s41586-019-0880-5 (2019).

609 27 Hennacy, J. H. & Jonikas, M. C. Prospects for Engineering Biophysical CO₂ Concentrating
610 Mechanisms into Land Plants to Enhance Yields. *Annual review of plant biology*,
611 doi:10.1146/annurev-arplant-081519-040100 (2020).

612 28 Long, B. M. *et al.* Carboxysome encapsulation of the CO₂-fixing enzyme Rubisco in tobacco
613 chloroplasts. *Nature communications* **9**, 3570, doi:10.1038/s41467-018-06044-0 (2018).

614 29 Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. & Hanson, M. R. A faster Rubisco
615 with potential to increase photosynthesis in crops. *Nature* **513**, 547-550,
616 doi:10.1038/nature13776 (2014).

617 30 Atkinson, N. *et al.* Introducing an algal carbon-concentrating mechanism into higher plants:
618 location and incorporation of key components. *Plant Biotechnol J* **14**, 1302-1315,
619 doi:10.1111/pbi.12497 (2016).

620 31 Hanson, M. R., Gray, B. N. & Ahner, B. A. Chloroplast transformation for engineering of
621 photosynthesis. *Journal of experimental botany* **64**, 731-742, doi:10.1093/jxb/ers325 (2013).

622 32 Raven, J. A., Beardall, J. & Sanchez-Baracaldo, P. The possible evolution and future of CO₂-
623 concentrating mechanisms. *J Exp Bot* **68**, 3701-3716, doi:10.1093/jxb/erx110 (2017).

624 33 Kropat, J. *et al.* A revised mineral nutrient supplement increases biomass and growth rate in
625 *Chlamydomonas reinhardtii*. *The Plant journal : for cell and molecular biology* **66**, 770-780,
626 doi:10.1111/j.1365-313X.2011.04537.x (2011).

627 34 Khrebukova, I. & Spreitzer, R. J. Elimination of the Chlamydomonas gene family that
628 encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Proc Natl
629 Acad Sci U S A* **93**, 13689-13693, doi:10.1073/pnas.93.24.13689 (1996).

630 35 Zhang, R. *et al.* High-Throughput Genotyping of Green Algal Mutants Reveals Random
631 Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming
632 DNA. *Plant Cell* **26**, 1398-1409, doi:10.1105/tpc.114.124099 (2014).

633 36 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved
634 cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

635 37 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron
636 micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

637 38 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination
638 in RELION-3. *eLife* **7**, doi:10.7554/eLife.42166 (2018).

639 39 Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle
640 image processing. *eLife* **7**, doi:10.7554/eLife.35383 (2018).

641 40 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for
642 rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296,
643 doi:10.1038/nmeth.4169 (2017).

644 41 Punjani, A., Brubaker, M. A. & Fleet, D. J. Building Proteins in a Day: Efficient 3D Molecular
645 Structure Estimation with Electron Cryomicroscopy. *IEEE Trans Pattern Anal Mach Intell*
646 **39**, 706-718, doi:10.1109/TPAMI.2016.2627573 (2017).

647 42 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and
648 analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

649 43 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta
650 Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

651 44 Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure
652 prediction server. *Nucleic Acids Res* **43**, W389-394, doi:10.1093/nar/gkv332 (2015).

653 45 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular
654 structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221,
655 doi:10.1107/S0907444909052925 (2010).

656 46 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular
657 crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21,
658 doi:10.1107/S0907444909042073 (2010).

659 47 Sueoka, N. Mitotic replication of deoxyribonucleic acid in *Chlamydomonas reinhardtii*. *Proc
660 Natl Acad Sci U S A* **46**, 83-91, doi:10.1073/pnas.46.1.83 (1960).

661 48 Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient
662 system for high-level expression and easy purification of authentic recombinant proteins.
663 *Protein Sci* **13**, 1331-1339, doi:10.1110/ps.04618904 (2004).

664 49 Ewalt, K. L., Hendrick, J. P., Houry, W. A. & Hartl, F. U. In vivo observation of polypeptide
665 flux through the bacterial chaperonin system. *Cell* **90**, 491-500, doi:10.1016/s0092-
666 8674(00)80509-7 (1997).

667 50 Baker, R. T. *et al.* in *Ubiquitin and Protein Degradation, Part A Methods in Enzymology*
668 540-554 (2005).

669 51 Genkov, T., Meyer, M., Griffiths, H. & Spreitzer, R. J. Functional hybrid rubisco enzymes
670 with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in
671 chlamydomonas. *J Biol Chem* **285**, 19833-19841, doi:10.1074/jbc.M110.124230 (2010).

672 52 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases.
673 *Nature methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).

674 53 Li, X. *et al.* A genome-wide algal mutant library and functional screen identifies genes
675 required for eukaryotic photosynthesis. *Nat Genet* **51**, 627-635, doi:10.1038/s41588-019-
676 0370-6 (2019).

677 54 Martinez-Sanchez, A. *et al.* Template-free detection and classification of membrane-bound
678 complexes in cryo-electron tomograms. *Nat Methods*, doi:10.1038/s41592-019-0675-5
679 (2020).

680 55 Cheng, S., Cetinkaya, M. & Grater, F. How sequence determines elasticity of disordered
681 proteins. *Biophys J* **99**, 3863-3869, doi:10.1016/j.bpj.2010.10.011 (2010).

682

683

684

685 **Acknowledgements**

686 We thank Jianping Wu, Nieng Yan, Luke Mackinder, Cliff Brangwynne and members of the
687 Jonikas laboratory for helpful discussions; Ned Wingreen, Silvia Ramundo, Jessi Hennacy, and
688 Eric Franklin for constructive feedback on the manuscript; Wolfgang Baumeister and Jürgen
689 Plitzko for providing support and cryo-ET instrumentation; and Miroslava Schaffer for help with
690 acquiring the cryo-ET data, previously published in Freeman Rosenzweig *et al.*, 2017. This project
691 was funded by National Science Foundation (IOS-1359682 and MCB-1935444), National
692 Institutes of Health (DP2-GM-119137), and Simons Foundation and Howard Hughes Medical
693 Institute (55108535) grants to M.C.J., Deutsche Forschungsgemeinschaft grant (EN 1194/1-1 as
694 part of FOR2092) to B.D.E., Ministry of Education (MOE Singapore) Tier 2 grant (MOE2018-
695 T2-2-059) to O.M.-C., UK Biotechnology and Biological Sciences Research Council
696 (BB/S015531/1) and Leverhulme Trust (RPG-2017-402) grants to A.J.M and N.A., NIH grant
697 R01GM071574 to F.M.H., Deutsche Forschungsgemeinschaft fellowship (PO2195/1-1) to S.A.P.,
698 and National Institute of General Medical Sciences of the National Institutes of Health
699 (T32GM007276) training grant to V.K.C.. The content is solely the responsibility of the authors
700 and does not necessarily represent the official view of the National Institutes of Health.

701

702 **Author contributions**

703 S.H., P.D.J., V.C., F.M.H., T.W., O.M.-C., B.D.E., and M.C.J. designed experiments. S.H.
704 identified EPYC1's Rubisco-binding regions on EPYC1 by peptide tiling array and SPR. S.H. and
705 S.A.P. prepared the Rubisco and EPYC1 peptide sample for single-particle cryo-EM; S.H., S.A.P.
706 and G.H. prepared the Rubisco samples for peptide tiling array and surface plasmon resonance.
707 H.-T.C., D.M. and Z.Y. performed Cryo-EM grid preparation, sample screening, data acquisition,
708 image processing, reconstruction and map generation. D.M. and P.D.J. carried out single-particle
709 model building and fitting and refinement. S.H., H.-T.C., D.M., P.D.J., F.M.H. and M.C.J.
710 analyzed the structures. S.H. and W.P. analyzed EPYC1 binding to Rubisco by peptide substitution
711 array and SPR. T.W. performed in vitro reconstitution phase separation experiments. N.A. and
712 A.J.M. performed yeast two-hybrid experiments. S.H. and M.T.M. made Rubisco small subunit
713 point mutants. S.H. performed spot test experiments. M.T.M. performed TEM. A.M.-S. performed
714 the cryo-ET data analysis and modeling. S.H. and M.C.J. wrote the manuscript. All authors read
715 and commented on the manuscript.

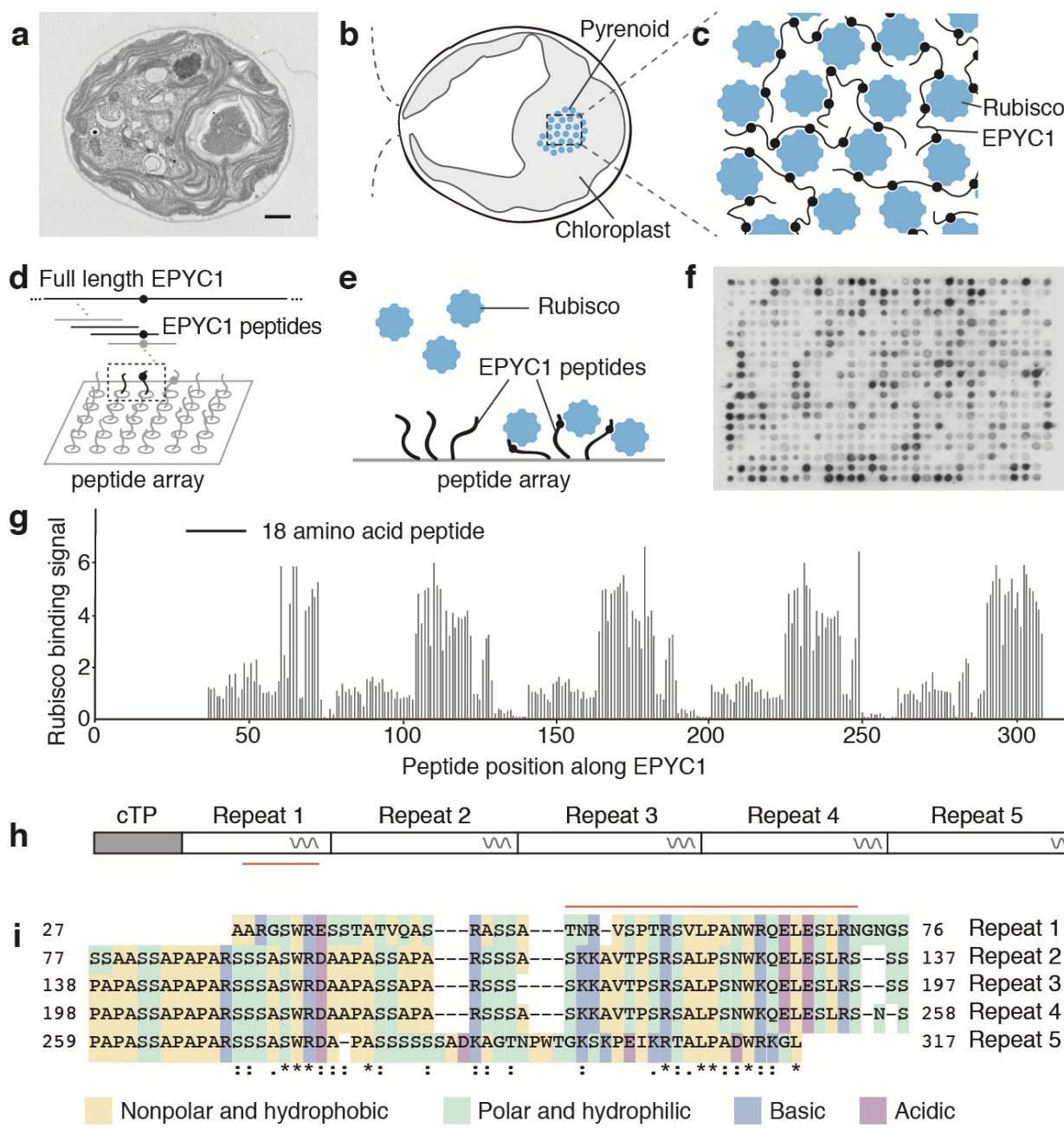
716

717 **Conflict of interest statement**

718 Princeton University and HHMI have submitted a provisional patent application on aspects of
719 the findings.

720 **Figures**

721

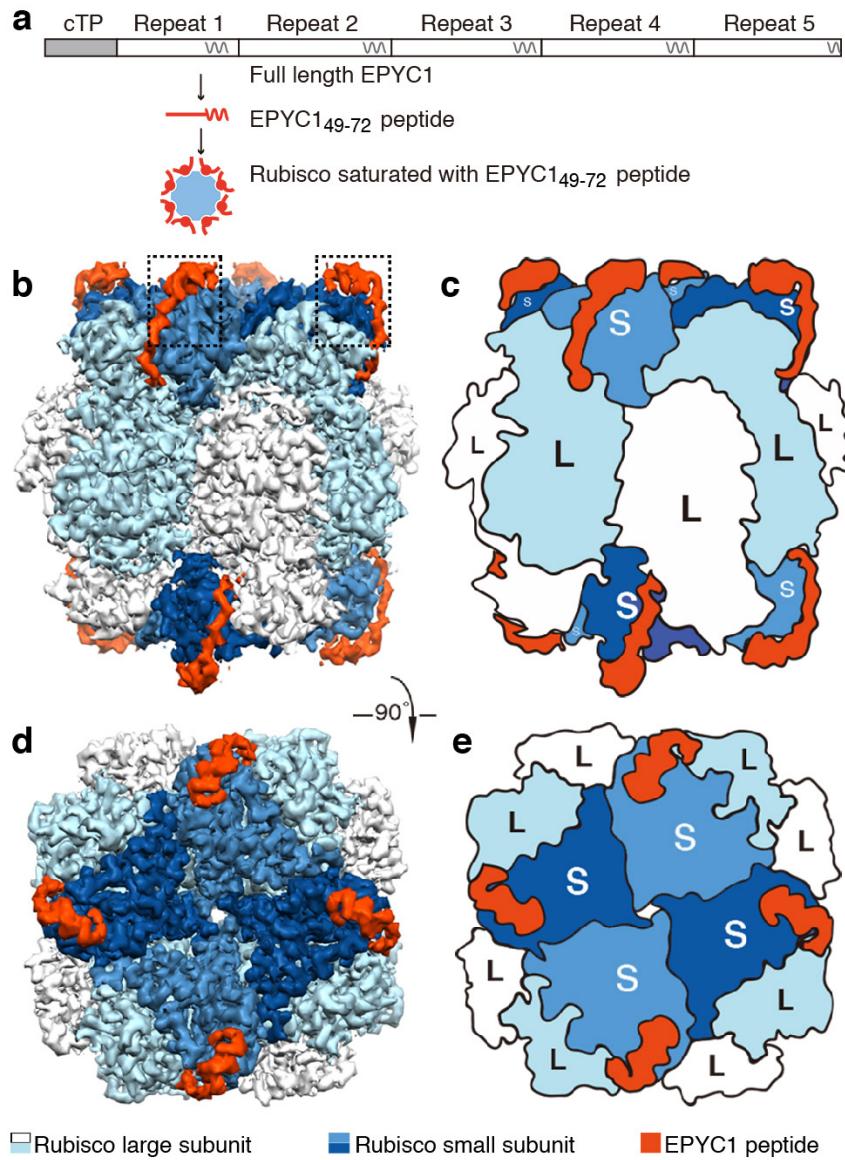


722

723

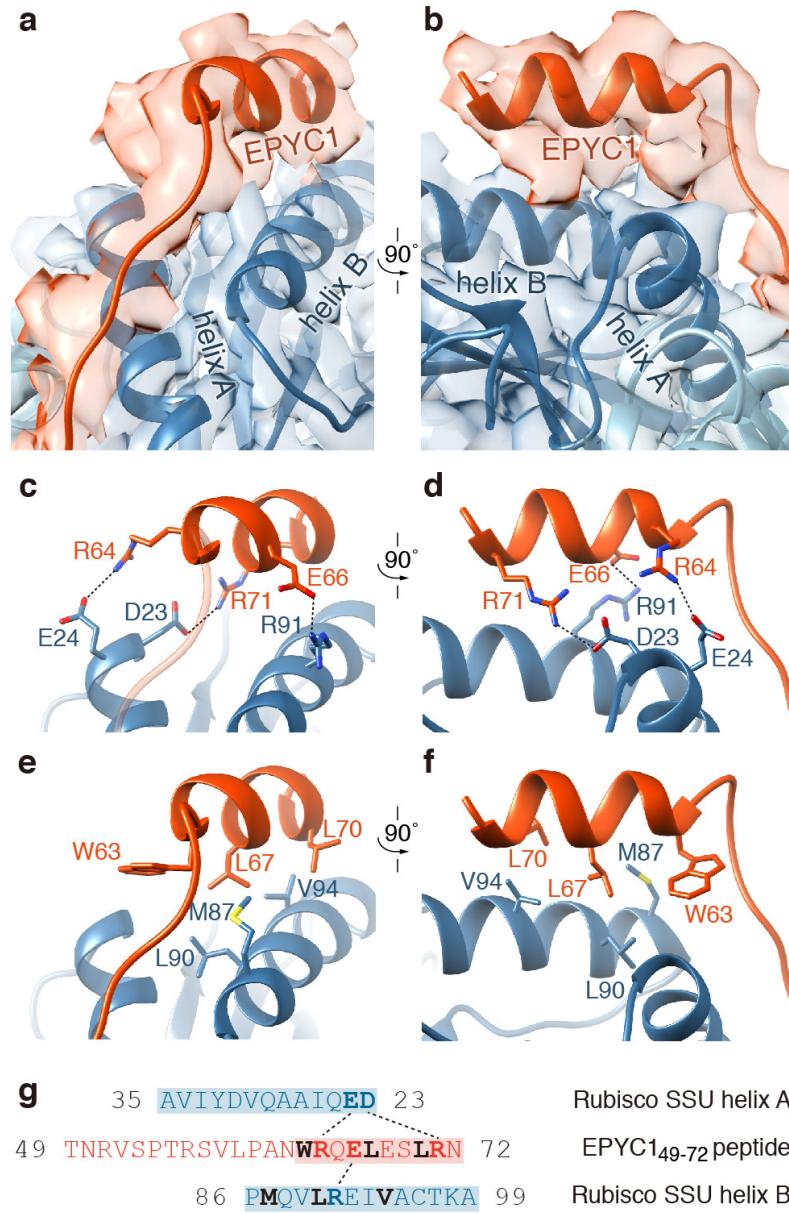
724 **Fig. 1 | EPYC1 consists of five tandem sequence repeats, each of which contains a Rubisco-
725 binding region.** **a**, Transmission electron microscopy (TEM) image of a Chlamydomonas cell.
726 Scale bar = 1 μ m. **b**, Cartoon depicting the chloroplast and pyrenoid in the image shown in panel
727 a. The blue dots indicate the location of Rubisco enzymes clustered in the pyrenoid matrix. **c**, We

728 hypothesized that pyrenoid matrix formation is mediated by multivalent interactions between
729 Rubisco and the intrinsically disordered protein EPYC1. **d**, We designed an array of 18 amino acid
730 peptides tiling across the full length EPYC1 sequence. **e**, Incubation of the array with purified
731 Rubisco allows identification of peptides that bind to Rubisco. **f**, Image of the Rubisco binding
732 signal from the peptide tiling array. **g**, The Rubisco binding signal was quantified and plotted for
733 each peptide as a function of the position of the middle of the peptide along the EPYC1 sequence.
734 The initial 26 amino acids of EPYC1 correspond to a chloroplast targeting peptide (cTP), which
735 is not present in the mature protein¹². Results are representative of three independent experiments.
736 **h**, The positions of EPYC1's five sequence repeats are shown to scale with panel g. Predicted α -
737 helical regions are shown as wavy lines. **i**, Primary sequence of EPYC1, with the five sequence
738 repeats aligned. In panels h and i, the region used for structural studies (EPYC1₄₉₋₇₂) is indicated
739 by a red line.



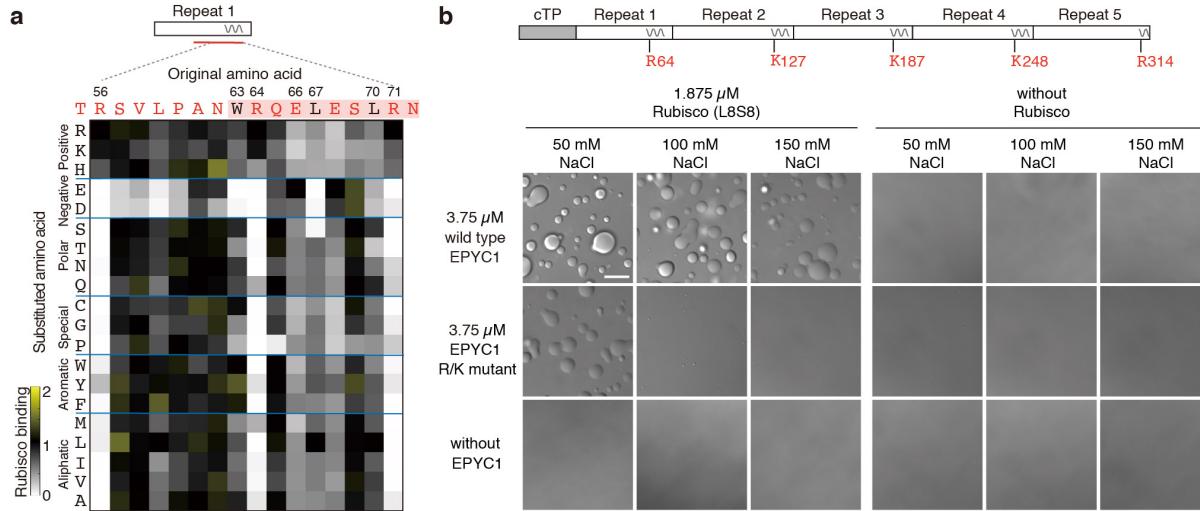
742 **Fig. 2 | EPYC1 binds to Rubisco small subunits.** **a**, Peptide EPYC1₄₉₋₇₂, corresponding to the
743 first Rubisco-binding region of EPYC1, was incubated at saturating concentrations with Rubisco
744 prior to single particle cryo-electron microscopy. **b-e**, Density maps (b, d) and cartoons (c, e)
745 illustrate the side views (b, c) and top views (d, e) of the density map of the EPYC1 peptide-
746 Rubisco complex. Dashes in panel b indicate regions shown in Fig.3a-3f.

747



750 **Fig. 3 | EPYC1 binds to Rubisco small subunit α -helices via salt bridges and a hydrophobic**
 751 **pocket.** **a-b**, Front (a) and side (b) views of the EPYC1 peptide (red) bound to the two α -helices
 752 of the Rubisco small subunit (blue). **c-d**, Three pairs of residues form salt bridges between the
 753 helix of the EPYC1 peptide and the helices on the Rubisco small subunit. Shown are front (c) and
 754 side (d) views as in panel a and panel b. The distances from EPYC1 R64, R71 and E66 to Rubisco
 755 small subunit E24, D23 and R91 are 3.06 Å, 3.23 Å, and 3.13 Å, respectively. **e-f**, A hydrophobic
 756 pocket is formed by three residues of the EPYC1 peptide and three residues of helix B of the
 757 Rubisco small subunit. Shown are front (e) and side (f) views as in panel a and panel b. **g**, Summary

758 of the interactions observed between the EPYC1 peptide and the two α -helices of the Rubisco
759 small subunit. Helices are highlighted; the residues mediating interactions are bold; salt bridges
760 are shown as dotted lines; residues contributing to the hydrophobic pocket are shown in black.

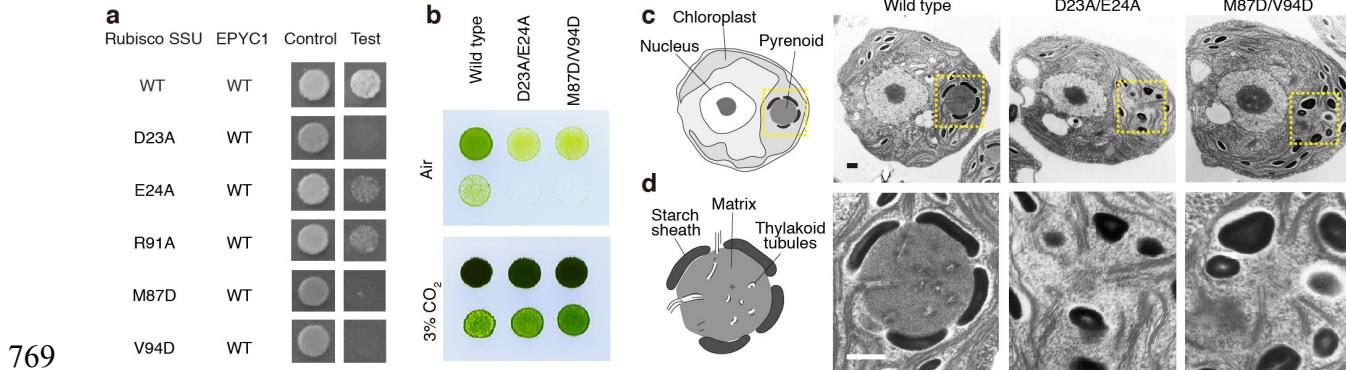


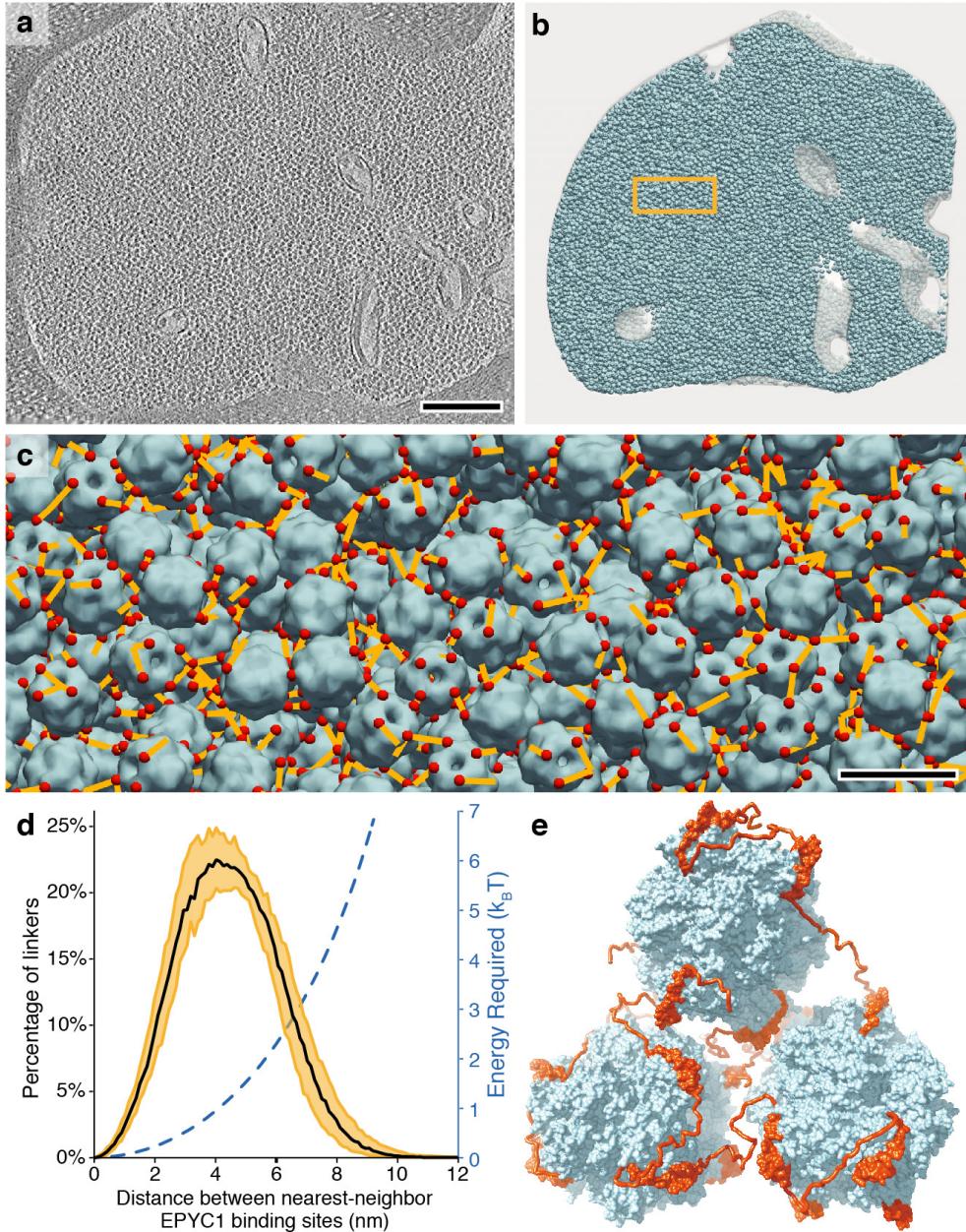
761

762

763 **Fig. 4 | Interface residues on EPYC1 are required for binding and phase separation of**
764 **EPYC1 and Rubisco *in vitro*. a**, Rubisco binding to a peptide array representing every possible

765 single amino acid substitution for amino acids 56-71 of EPYC1. The binding signal was
766 normalized by the binding signal of the original sequence. **b**, The effect of mutating the central R
767 or K in each of EPYC1's Rubisco-binding regions on *in vitro* phase separation of EPYC1 with
768 Rubisco. Scale bar = 10 μ m.





780

781

782 **Fig. 6 | A model for matrix structure consistent with *in situ* Rubisco positions and**
783 **orientations.** a, The pyrenoid matrix was imaged by cryo-electron tomography⁵. An individual

784 slice through the three-dimensional volume is shown. Scale bar = 200 nm. b, The positions and

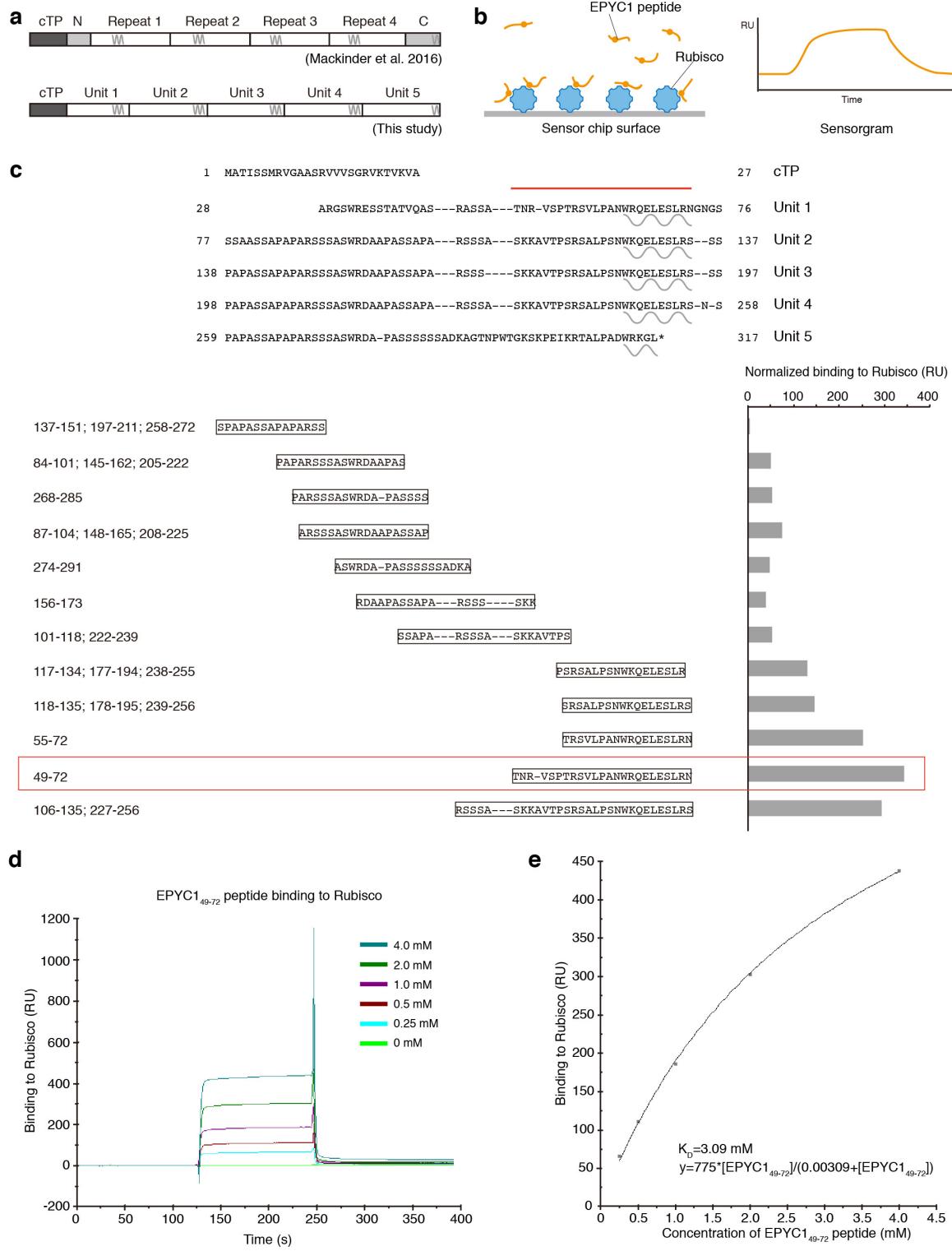
785 orientations of individual Rubisco holoenzymes (blue) were determined by subtomogram

786 averaging and fit into the tomogram volume. c, The distances (yellow) between the nearest

787 EPYC1-binding sites (red) on neighboring Rubisco holoenzymes (blue) were measured. The view

788 is from inside the matrix; in some cases the nearest EPYC1 binding site is on a Rubisco that is out

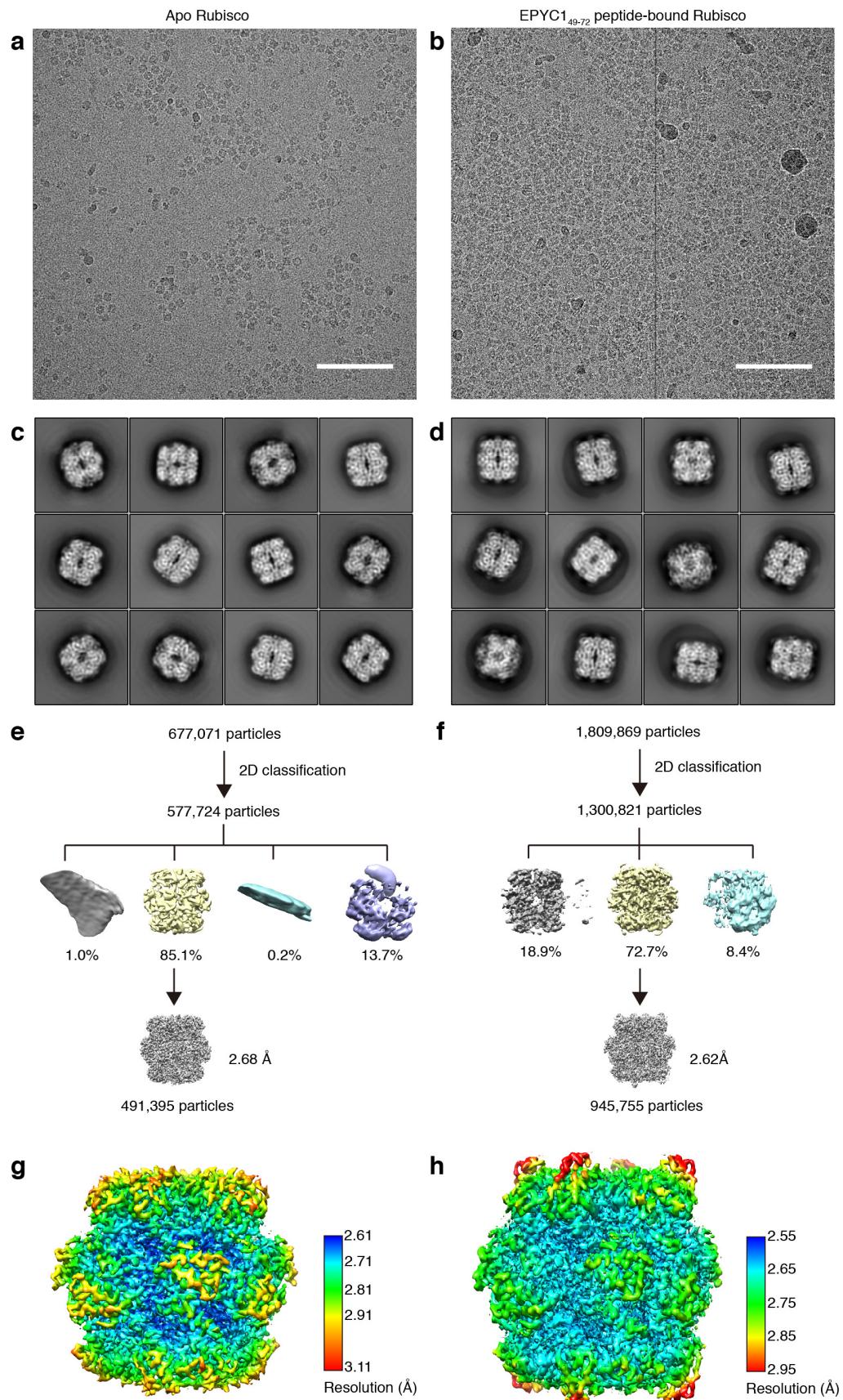
789 of the field of view, causing some yellow lines to appear unconnected in this image. Scale bar =
790 20 nm. **d**, Histogram showing the distances between the nearest EPYC1 binding sites on
791 neighboring Rubisco holoenzymes. Shading indicates 95% confidence interval based on data from
792 five independent tomograms. The estimated energy required for stretching a chain of 40 amino
793 acids a given distance is shown in blue. **e**, A 3D model illustrates how EPYC1 (red) could crosslink
794 multiple Rubisco holoenzymes (blue) to form the pyrenoid matrix. The conformations of the
795 intrinsically disordered linkers between EPYC1 binding sites were modeled hypothetically.



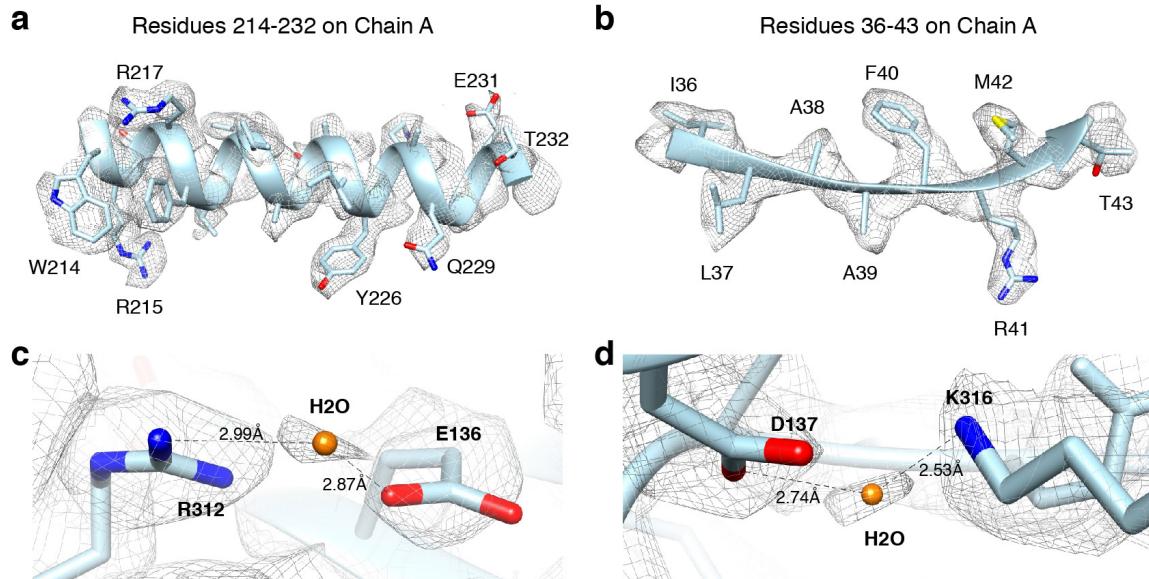
796

797 **Extended Data Fig. 1 | The EPYC1 peptide with the highest binding affinity to Rubisco was**
 798 **chosen for structural studies.** a, Diagram indicating the differences between the previously
 799 defined sequence repeats¹⁰ and the newly defined sequence repeats on full-length EPYC1. b, To

800 verify the Rubisco-binding regions on EPYC1, surface plasmon resonance (SPR) was used to
801 measure the binding of EPYC1 peptides to Rubisco. Purified Rubisco was immobilized on a sensor
802 surface, and the EPYC1 peptides in solution were injected over the surface. The binding activity
803 was recorded in real time in a sensorgram. **c**, The peptides used in SPR experiments are shown
804 aligned to the sequence as shown in Fig. 1. The Rubisco-binding signal from the SPR experiment
805 of each peptide is shown after normalization to the peptide's molecular weight. EPYC1₄₉₋₇₂ was
806 chosen for structural studies based on its reproducible high Rubisco binding signal. **d**, The
807 Rubisco-binding response of the EPYC1₄₉₋₇₂ peptide at different concentrations was measured by
808 SPR. **e**, The binding responses shown in (d) were fitted to estimate the K_D of EPYC1₄₉₋₇₂ peptide
809 binding to Rubisco.



811 **Extended Data Fig. 2 | Cryo-EM data collection and image processing procedure.** a,
812 Representative micrograph of the apo Rubisco sample. Scale bar = 100 nm. b, Representative
813 micrograph of Rubisco-EPYC1₄₉₋₇₂ complexes. Scale bar = 100 nm. c, Representative 2D class
814 averages of the apo Rubisco sample. d, Representative 2D class averages of Rubisco-EPYC1₄₉₋₇₂
815 complexes. e, Overview of the workflow for single particle data processing for the apo Rubisco
816 sample. f, Overview of the workflow for single particle data processing for the Rubisco-EPYC1₄₉₋
817 ₇₂ sample. g, Local resolution estimation diagram of the final refined apo Rubisco map. h, Local
818 resolution estimation diagram of the final refined Rubisco-EPYC1₄₉₋₇₂ complexes map.

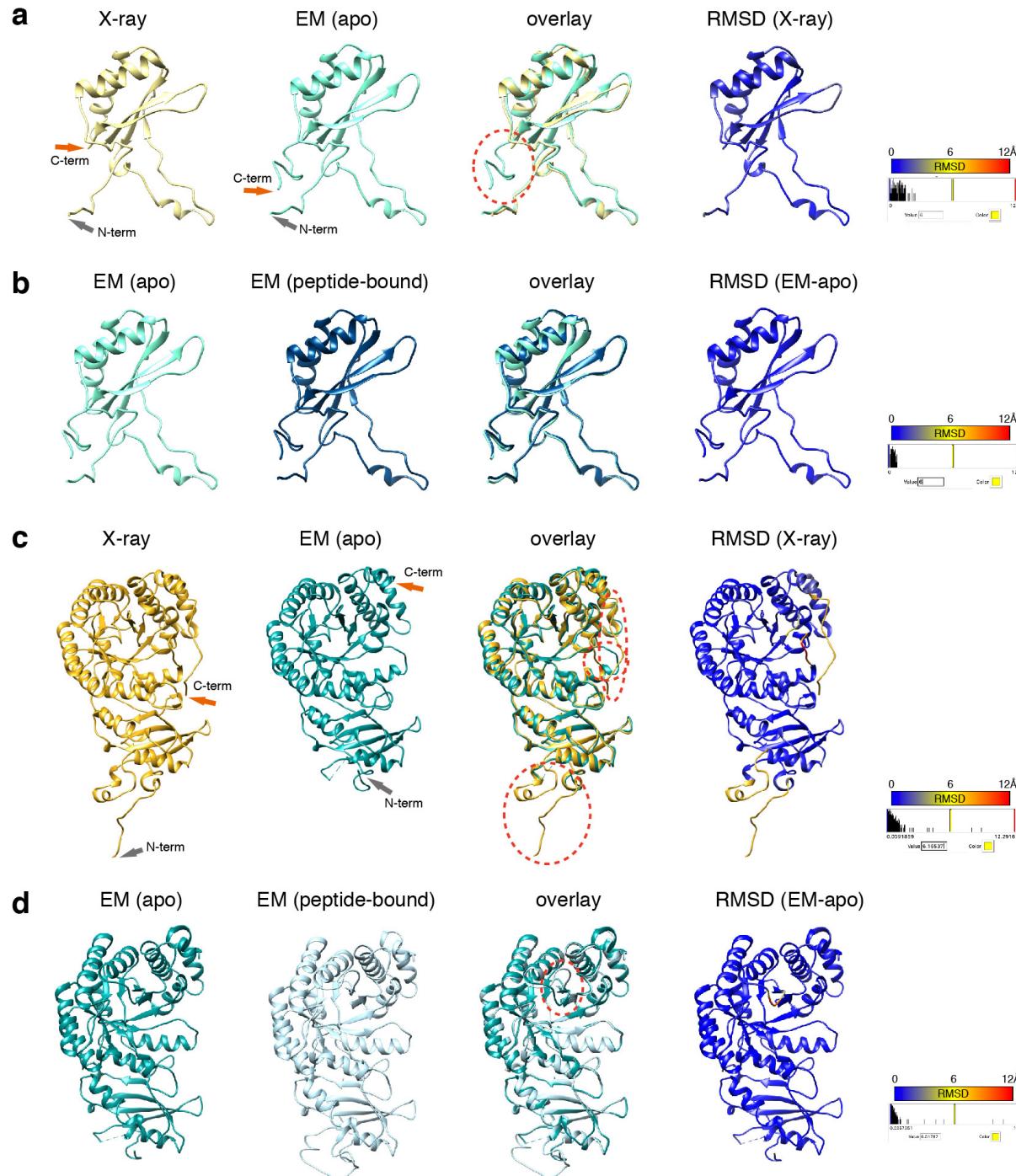


819

820

821 **Extended Data Fig. 3 | Cryo-EM analysis and resolution of Rubisco-EPYC1 peptide**
822 **complexes in this study.** **a-b**, Representative cryo-EM density quality showing an α -helix of
823 residues 214-232 in chain A (a) (one of the Rubisco large subunits) and a β -sheet of residues 36-
824 43 in chain A (b). The densities are shown as meshwork in gray. The backbones of the structural
825 model are in ribbon representation, and side chains are shown in stick representation. **c-d**,
826 Representative cryo-EM density quality showing water molecules as orange spheres. One water
827 molecule between R312 and E136 on chain A is shown in panel c, and another water molecule
828 between D137 and K316 on chain A is shown in panel d.

829



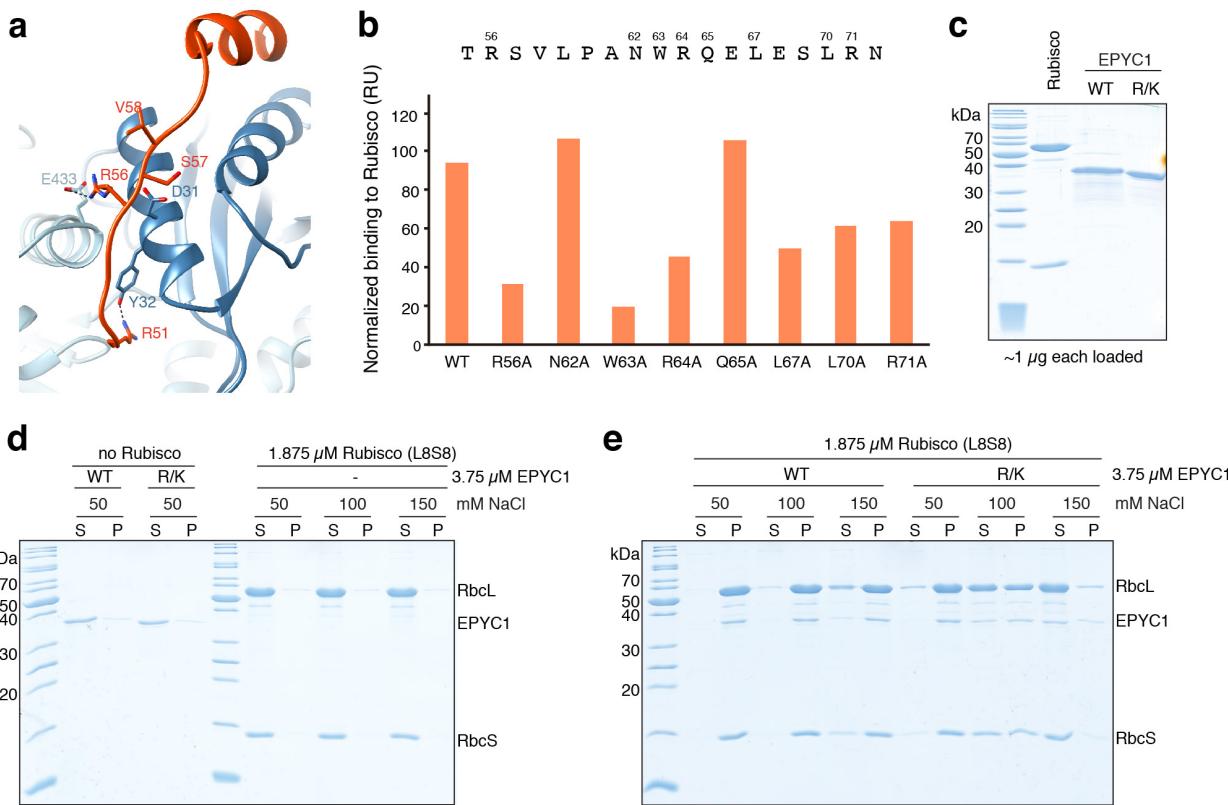
830

831 **Extended Data Fig. 4 | Comparison of our EM structure and the published X-ray**
832 **crystallography structure (1gk8) of Rubisco purified from *Chlamydomonas reinhardtii*¹³, and**
833 **comparison of our EM structure of native Rubisco and Rubisco bound with EPYC149-72**
834 **peptide.** a, Comparison of the structure of the small subunit of apo Rubisco obtained here by EM

835 with 1gk8. The EM structure has additional C-terminus density past residue 126, circled by a red

836 dashed line. b, Comparison of our two EM structures of the small subunit: from apo Rubisco and
837 from EPYC1 peptide-bound Rubisco. c, Comparison of the structure of the large subunit of apo
838 Rubisco obtained here by EM with 1gk8. The three major differences found between the X-ray
839 structure and the EM structure of the large subunit are circled with red dashed lines. d, Comparison
840 of our two EM structures of the large subunit: from apo Rubisco and from EPYC1 peptide-bound
841 Rubisco. The major difference found between the peptide-bound structure and the apo EM
842 structure was the loop between K175 and L180 of the large subunit, which is shown circled by a
843 red dashed line.

844



845

846

847 **Extended Data Fig. 5 | Interface residues on EPYC1 identified by Cryo-EM are important**
 848 **for binding and phase separation of EPYC1 and Rubisco.** **a**, Our structure suggests that some

849 residues in addition to the ones shown in Fig. 4 may also contribute to the interaction between

850 EPYC1 and Rubisco. R56 of the EPYC1 peptide may interact with both D31 of the Rubisco small

851 subunit and E433 of the Rubisco large subunit. R51 of the EPYC1 peptide may form a salt bridge

852 with Y32 of the Rubisco small subunit. Residues S57 and V58 of the EPYC1 peptide are close to

853 D31 in the structure, which may explain why replacing either of these residues with a negatively

854 charged residue disrupts binding (Fig. 4a). **b**, The wild-type (WT) EPYC1 peptide or EPYC1

855 peptides with the indicated point mutations were synthesized, and their Rubisco-binding signal

856 was measured by surface plasmon resonance. **c**, SDS-PAGE analysis of purified proteins used for

857 *In vitro* phase separation experiments. WT = wild-type EPYC1; R/K =

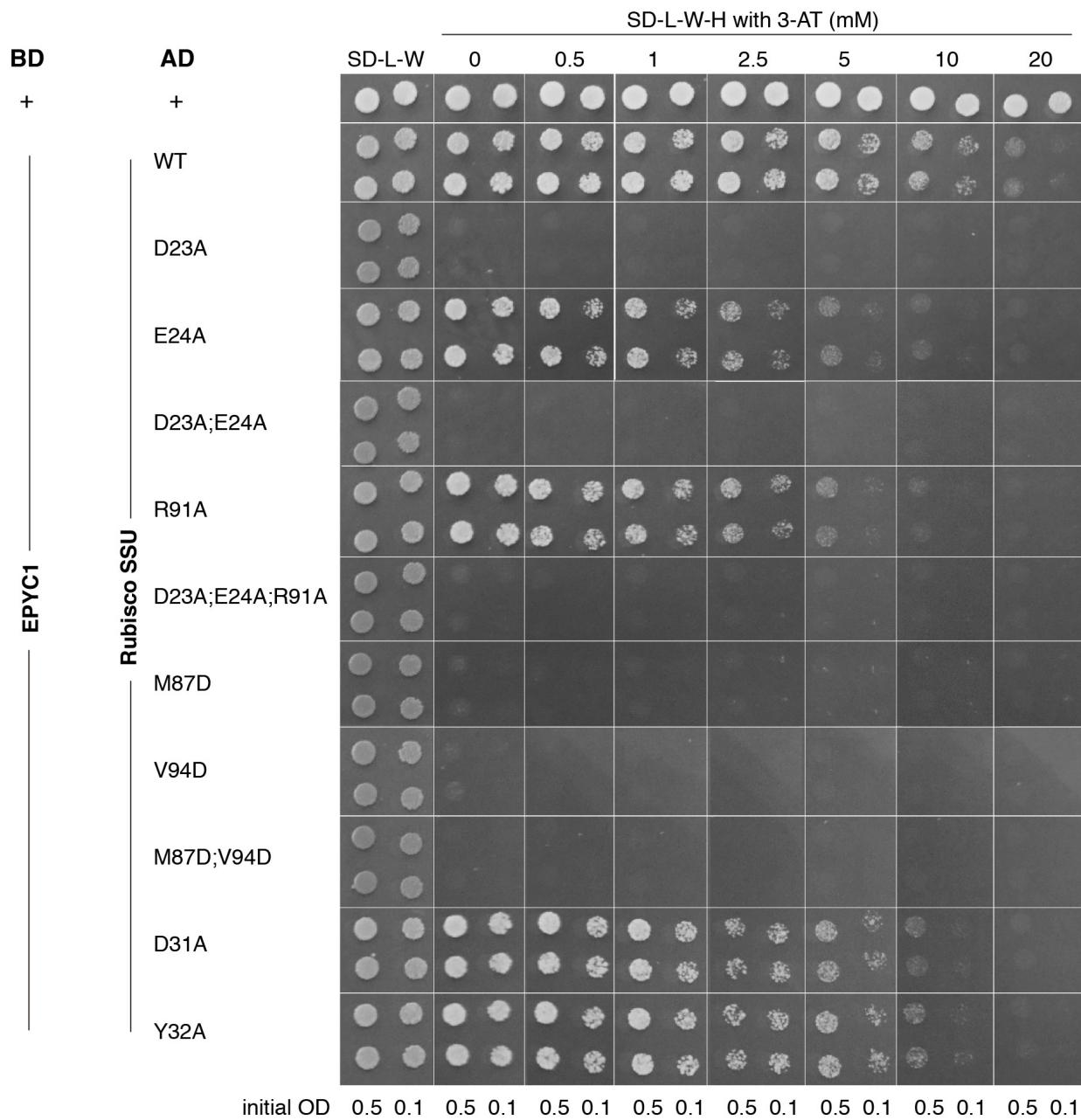
858 EPYC1^{R64A/K127A/K187A/K248A/R314}. **d-e**, A droplet sedimentation assay was used as a readout of phase

859 separation complementary to the microscopy analyses shown in Fig. 4b. Proteins at indicated

860 concentrations were mixed and incubated for 10 minutes, then condensates were pelleted by

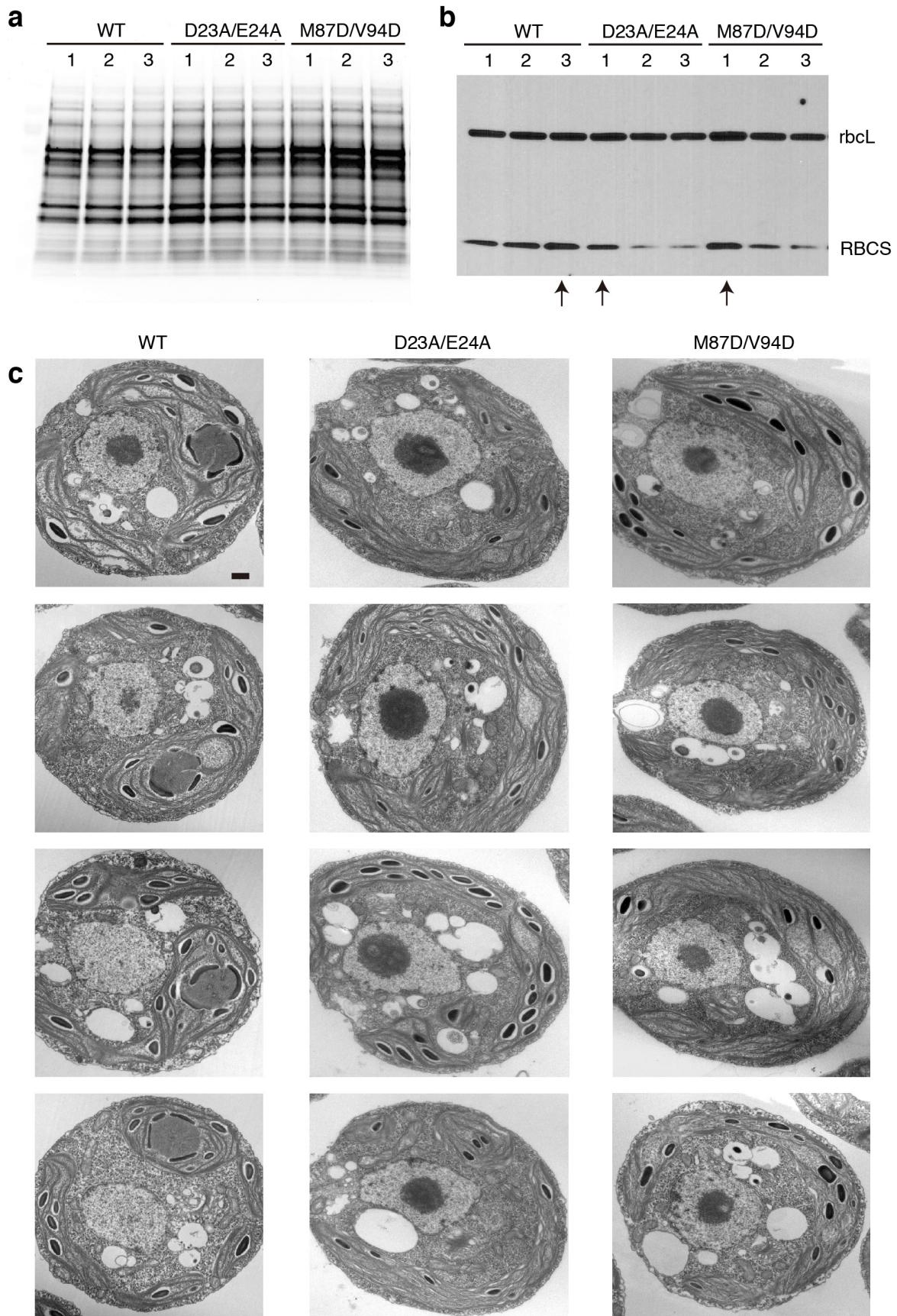
861 centrifugation. Supernatant (S) and pellet (P) fractions were run on a denaturing gel. The negative
862 controls with no Rubisco or with no EPYC1 are shown in (d), and the wild-type Rubisco with
863 wild-type EPYC1 or mutant EPYC1 are shown in (e).

864



866

867 **Extended Data Fig. 6 | Yeast Two-Hybrid assays of interactions between EPYC1 and wild-
868 type or mutated Rubisco small subunit.** Colonies are shown after 3 days' growth on plates. A
869 subset of the data shown in this figure are shown in Fig. 5a.



871 **Extended Data Fig. 7 | Selection of the Rubisco small subunit mutant strains for phenotype**
872 **analysis.** **a**, The Rubisco small subunit-less mutant T60 (*Δrbcs*) was transformed with DNA
873 encoding wild-type and mutant Rubisco small subunits (RBCS) to produce candidate
874 transformants with the genotypes *Δrbcs;RBCS^{WT}*, *Δrbcs;RBCS^{D23A/E24A}*, and *Δrbcs;RBCS^{M87D/V94D}*.
875 Total protein extracts for three strains from each transformation were separated on a
876 polyacrylamide gel. **b**, The gel shown in A was probed by Western Blot using a polyclonal
877 antibody mixture that detects both large and small Rubisco subunits. The candidate transformants
878 with highest RBCS expression level from each genotype are indicated by an arrow below the lanes
879 and were used for the subsequent phenotypic analyses shown in Fig. 5 and panel c. **c**, Additional
880 representative TEM images of whole cells of the strains expressing wild-type, D23A/E24A, and
881 M87D/V94D Rubisco small subunit. Scale bar = 500 nm.

882 **Extended Data Table 1 | Cryo-EM data collection and refinement.**

883

	#1 Apo Rubisco (EMDB-xxxx) (PDB xxxx)	2# EPYC1 ₄₉₋₇₂ peptide-bound Rubisco (EMDB-xxxx) (PDB xxxx)
Voltage (kV)	300	300
Magnification	22,500	22,500
Defocus range (μm)	-1.5 to -3.0	-1.5 to -3.0
Pixel size (Å)	1.31	1.31
Exposure time (s)	10	10
No. movie frames	50	50
Electron dose (e ⁻ /Å ²)	58	58
No. micrographs	2,500	2,500
No. initial particles	677,071	1,809,869
No. final particle	491,395	945,755
Symmetry	D4	D4
Resolution (Å)	2.68	2.62
Map sharpening <i>B</i> factor (Å ²)	-147.233	-154.161

884

885

886 **Extended Data Table 2 | The amino acid residues that form the Rubisco-binding regions on**
887 **EPYC1 homologs, and the EPYC1 binding site on the surface of Rubisco, appear to be**
888 **conserved across the order Volvocales.** Residues with roles in the binding interface are bolded.
889 Residues that are different from the *Chlamydomonas reinhardtii* sequence are highlighted in grey.
890

Species	First Rubisco-binding region on EPYC1 homolog	Rubisco SSU helix A	Rubisco SSU helix B
<i>Chlamydomonas reinhardtii</i>	TRSVLPAN WRQELESLRN	D EQIAAQVVDYIVA	PMQVL R EIVACTKA
<i>Tetrabaena socialis</i>	TRSVLPAN WRQELESLRG	D EQIAAQVVDYIVA	PMQVL R EIV S CTRA
<i>Gonium pectorale</i>	TRSVLPAN WRQELESLRN	D EQIAAQVVDYIVA	PMQVL R EIVACTKA
<i>Volvox carteri</i>	TRSVLPAN WRQELESLRN	D EQIAAQVVDYIVA	PMQVL R EIVACTKA

891