UNIVERSITY of York

This is a repository copy of *The eukaryotic CO2-concentrating organelle is liquid-like and exhibits dynamic reorganization*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/120656/

Version: Accepted Version

Article:

Rosenzweig, Elizabeth S Freeman, Xu, Bin, Cuellar, Luis Kuhn et al. (11 more authors) (2017) The eukaryotic CO2-concentrating organelle is liquid-like and exhibits dynamic reorganization. Cell. 148-162.e19. ISSN 1097-4172

https://doi.org/10.1016/j.cell.2017.08.008

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.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The Eukaryotic CO_2-Concentrating Organelle is Liquid-Like and Exhibits Dynamic Reorganization

Elizabeth S. Freeman Rosenzweig[1,2], Bin Xu[3,11], Luis Kuhn Cuellar[4,11], Antonio Martinez-Sanchez[4], Miroslava Schaffer[4], Mike Strauss[5], Heather N. Cartwright[2], Pierre Ronceray[6], Jürgen M. Plitzko[4], Friedrich Förster[4,9], Ned S. Wingreen[7,8]*, Benjamin D. Engel[4]*, Luke C. M. Mackinder[2,10,12], Martin C. Jonikas[1,2,8,12,13]*

[1] Department of Biology, Stanford University, Stanford, California 94305, USA;
[2] Department of Plant Biology, Carnegie Institution for Science, Stanford, California 94305, USA;
[3] Department of Physics, Princeton University, Princeton, New Jersey 08544, USA;
[4] Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany;
[5] Cryo-EM facility, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany;
[6] Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA
[7] Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544;
[8] Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544;
[9] Present address: Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands;
[10] Present address: Department of Biology, University of York, York, YO10 5DD, UK.
[11] These authors contributed equally
[12] These authors contributed equally
[13] Lead Contact

* Correspondence: mjonikas@princeton.edu, engelben@biochem.mpg.de, and wingreen@princeton.edu

Summary

Approximately 30-40% of global CO_2 fixation occurs inside a poorly understood organelle called the pyrenoid, which is found in the chloroplasts of most eukaryotic algae. The pyrenoid matrix is densely packed with the CO_2-fixing enzyme Rubisco, and is thought to be a crystalline or amorphous solid. Here, we show that the pyrenoid matrix of the unicellular alga *Chlamydomonas reinhardtii* is not crystalline, but behaves as a liquid that dissolves and condenses during cell division. Furthermore, we show that the pyrenoid is inherited both by fission and *de novo* assembly. A model inspired by this system predicts the existence of a "magic number" effect that influences phase separation in liquid-like organelles. This view of the pyrenoid matrix as a phase-separated compartment provides a new paradigm for understanding its structure, biogenesis, and regulation. More broadly, our findings expand our understanding of the principles that govern the architecture and inheritance of liquid-like organelles.

38

39 **Introduction**

40 Although pyrenoids mediate approximately one third of global carbon fixation
41 (Mackinder *et al.*, 2016), the molecular structure and biogenesis of these biogeochemically
42 fundamental organelles remain largely unknown. Pyrenoids are non-membrane-bound,
43 proteinaceous structures that contain a matrix packed with the CO_2-fixing enzyme Rubisco. In
44 many species, the matrix is traversed by membrane tubules that are continuous with the
45 photosynthetic thylakoid membranes (Gibbs, 1962; Ohad *et al.*, 1967; Griffiths, 1970; Engel *et*
46 *al.*, 2015). Pyrenoids are a central feature of the algal CO_2 concentrating mechanism, which
47 supplies Rubisco with a high concentration of its substrate CO_2, enabling more efficient carbon
48 capture than that of most land plants (Badger *et al.*, 1998; Meyer *et al.*, 2012; Wang *et al.*, 2015).
49 Pyrenoids are found within the chloroplasts of a diverse array of photosynthetic eukaryotes,
50 including nearly all freshwater and marine algae, as well as a group of non-vascular plants
51 (Brown *et al.*, 1967; Griffiths, 1980; Wang *et al.*, 2015). Although pyrenoids were among the
52 first organelles to be scientifically documented (Vaucher, 1803), they have remained largely
53 uncharacterized at a molecular level due to the limited availability of genetic tools for algae.

54 Based on results in the model unicellular green alga *Chlamydomonas reinhardtii*
55 (*Chlamydomonas* hereafter), we recently proposed that a protein called Essential Pyrenoid
56 Component 1 (EPYC1; also known as LCI5) links Rubisco holoenzymes together to form the
57 pyrenoid matrix (Mackinder *et al.*, 2016). EPYC1 localizes to the pyrenoid matrix, is of similar
58 abundance to Rubisco, and is required for Rubisco's localization in the pyrenoid matrix,
59 supporting a possible structural role. EPYC1 binds Rubisco, and the EPYC1 protein sequence
60 consists of four nearly identical ~60 amino acid repeats, suggesting a model in which each of

61 EPYC1's repeats contains a Rubisco binding site, allowing EPYC1 to link multiple Rubisco
62 holoenzymes together. The molecular details of the EPYC1-Rubisco binding interaction, and the
63 packing organization of EPYC1 and Rubisco in the matrix, remain unknown.

64 Electron micrographs of several species suggest that the densely-packed pyrenoid matrix
65 is crystalline (Holdsworth, 1968; Kowallik, 1969; Bertagnolli and Nadakavukaren, 1970), while
66 micrographs of other species suggest the matrix is amorphous (Griffiths, 1970; Meyer *et al*.,
67 2012). Our recent cryo-electron tomography (cryo-ET) study of *Chlamydomonas* revealed that
68 the average packing of Rubisco molecules within the pyrenoid matrix resembles a hexagonal
69 close packed lattice (Engel *et al.*, 2015). We have proposed that EPYC1 could link Rubisco
70 holoenzymes together into this arrangement (Mackinder *et al.*, 2016). However, if Rubisco is
71 immobilized within a rigid lattice, it is not clear how Rubisco Activase, which regenerates
72 Rubisco's active sites (Pollock *et al.*, 2003; Portis, 2003) but is significantly less abundant than
73 Rubisco in pyrenoids (Mckay *et al.*,1991; Wienkoop *et al.*, 2010; Mackinder *et al.*, 2016), could
74 reach enough Rubisco molecules to maintain efficient carbon fixation.

75 Moreover, pyrenoid biogenesis has not been definitively observed in living cells. Until
76 recently, the only tool available for visualization of the pyrenoid matrix was electron
77 microscopy, which precludes direct observation of dynamics. Classic electron microscopy
78 studies on fixed cells have suggested that pyrenoids may arise *de novo* (Bisalputra and Weier,
79 1964; Brown *et al.*, 1967; Retallack and Butler, 1970) in some species and by fission (Buffaloe,
80 1958; Brown *et al.*, 1967; Goodenough, 1970) in others.

81 Here, we describe our study of pyrenoid inheritance in living cells, and how our findings
82 inspired us to investigate the *Chlamydomonas* pyrenoid as a liquid. Examination of pyrenoid
83 ultrastructure with improved *in situ* cryo-ET revealed that the pyrenoid matrix is not crystalline,

84 but exhibits liquid-like local order. Further live-cell fluorescence experiments showed that the
85 matrix mixes internally and disperses into the surrounding stroma during cell division,
86 demonstrating that the matrix exhibits liquid-like properties. Finally, using a simple model
87 inspired by the pyrenoid, we predict that phase transitions in some multicomponent, liquid-like
88 biological systems are governed by changes in the ratio of available multivalent binding sites on
89 the constituent macromolecules. This work provides insights into both algal CO_2 concentrating
90 mechanisms and liquid-like biological systems.
91
92 **Results**
93
94 **The Pyrenoid Exhibits Both Fission and *de novo* Assembly**
95 To enable the first observations of pyrenoid matrix dynamics in living cells, we expressed
96 pyrenoid matrix proteins tagged with the fluorescent protein Venus and imaged them in 3D with
97 fluorescence time-lapse microscopy during photoautotrophic growth. We tracked inheritance of
98 the pyrenoid matrix by monitoring Venus-tagged Rubisco small subunit 1 (RbcS1) or Venus-
99 tagged EPYC1, and recorded chlorophyll autofluorescence to follow cellular orientation and
100 chloroplast division. The fusion proteins localized to the pyrenoid (Figure 1A,D), and did not
101 measurably perturb pyrenoid function (Figure S1A-B). During our observation, cells often
102 divided twice in rapid succession, which is typical for *Chlamydomonas* (Cross and Umen, 2015).
103 We observed that roughly two thirds of daughter cells inherited their pyrenoid by
104 elongation and then fission of the mother cell's pyrenoid (Figure 1; Figure 2E-F; Movie S1;
105 Figure S2). The duration of pyrenoid fission, defined as the time between visible elongation of
106 the mother pyrenoid and separation of the daughter pyrenoids, was ~7 minutes, much faster than

107 chloroplast division (~30-80 minutes, Figure 1F). All pyrenoid fissions occurred during the final
108 minutes of chloroplast division (Figure 1G-H).
109 In all 88 cell divisions in which pyrenoid fission was observed, the completion of
110 pyrenoid division was quickly followed by a gap in the chlorophyll signal between the daughter
111 pyrenoids, generated by the chloroplast cleavage furrow (Figure 1E,G-H). Conversely, in the 44
112 cell divisions in which the furrow did not bisect the pyrenoid, pyrenoid fission did not occur.
113 These findings support the hypothesis that pyrenoid fission is driven by the constrictive force of
114 the chloroplast division furrow (Goodenough, 1970; Hirakawa and Ishida, 2015).
115 In cases where pyrenoid fission did not occur, one of the daughter cells typically
116 inherited the mother cell's whole pyrenoid (~16-21% of daughter cells; Figure 2A-F; Movie S2),
117 leaving the other daughter without a pyrenoid. Some of the cells that failed to inherit a pyrenoid
118 remained without a visible pyrenoid for the duration of observation (~8-9% of all daughter cells;
119 Figure 2A,C,E-F; Movie S2). In other cases, one or more fluorescent puncta appeared *de novo*
120 and grew or coalesced into an apparent pyrenoid (~6-7% of all daughter cells; Figure 2B,D,E-I;
121 Movie S3). The proportions of each inheritance pattern in RbcS1-Venus and EPYC1-Venus were
122 not significantly different (Figure 2E-F), indicating that these patterns are not artifacts related to
123 a particular construct.
124 Several characteristics of pyrenoid behavior during division are reminiscent of liquids.
125 Towards the end of fission, a "bridge" of matrix material connecting the two lobes of a dividing
126 pyrenoid is briefly visible (Figure 3). After the bridge ruptures, daughter pyrenoids quickly
127 revert to spherical shapes, similar to the behavior of liquid droplets (Stone, 1994; Yanashima *et*
128 *al.*, 2012). Furthermore, during apparent *de novo* pyrenoid formation, we typically observed that
129 smaller puncta shrank while larger ones grew until the cell contained a single pyrenoid (Figure

130 2B,D,G-I) – much like Ostwald ripening in systems containing multiple liquid droplets (Hyman
131 *et al.*, 2014), indicating that components likely exchange between the puncta. Both fission and *de*
132 *novo* formation has been observed in established liquid-like organelles such as *C. elegans* P
133 granules (Brangwynne *et al.*, 2009; Brangwynne, 2011; Saha *et al.*, 2016) and *Xenopus* oocyte
134 nucleoli (Brangwynne *et al.*, 2011; Weber and Brangwynne, 2015). The similarity of the
135 pyrenoid's behavior to that of such phase-separated biological liquid droplets suggested the
136 possibility that pyrenoids may also be liquid, rather than static solids or crystals, as has been
137 proposed (Holdsworth 1968; Kowallik 1969; Bertagnolli & Nadakavukaren 1970). We therefore
138 decided to investigate the fine-scale structure and potential liquid-like dynamics of the pyrenoid
139 matrix.
140

141 **The Matrix is not Crystalline, and has Liquid-Like Organization**

142 The different reports of crystalline and amorphous pyrenoid matrices emerged from
143 classical electron micrographs (Holdsworth, 1968; Kowallik, 1969; Bertagnolli and
144 Nadakavukaren, 1970; Griffiths, 1970, 1980). However, these micrographs have limited
145 resolution, and the native arrangement of Rubisco may be compromised by artifacts from the
146 sample preparation procedure, which involves chemical fixation, dehydration, plastic
147 embedding, and heavy metal staining (Crang and Klomparens, 1988).

148 To avoid such artifacts, we analyzed the positions of individual Rubisco holoenzymes
149 within the *Chlamydomonas* pyrenoid matrix by *in situ* cryo-ET (Figure 4A, Movie S4), which
150 provides 3D views of the native cellular environment at molecular resolution (Asano *et al.*, 2016;
151 Schaffer *et al.*, 2017). In order to achieve the complete freezing vitrification required for high-
152 quality cryo-ET, we examined *mat3-4* cells (Umen and Goodenough, 2001), which are smaller

153 due to premature cell division but have normal pyrenoid morphology and a functional CO_2
154 concentrating mechanism. The high fidelity of direct electron detector imaging enabled us to
155 localize ~97.5% of the Rubisco holoenzymes within five pyrenoid tomograms, totaling 192,100
156 particles (Figure 4B-C; Figures S3, S4A-C).

157 As a quantitative metric for the packing of Rubisco within the matrix, we calculated the
158 local concentration of holoenzymes in an expanding shell around every Rubisco. Our analysis
159 sampled only the pyrenoid matrix, avoiding potential effects from the traversing membrane
160 tubules and pyrenoid borders. Quasiperiodic near-neighbor peaks of Rubisco density eroded and
161 then vanished as distance increased from the reference particle (Figure 4D), suggesting that the
162 pyrenoid lacks the long-range order characteristic of a crystal.

163 We previously reported that the average local neighborhood in the pyrenoid matrix
164 resembles a lattice of hexagonal close packing (HCP) with ~15 nm between particle centers
165 (Engel *et al.*, 2015). However, due to limitations in imaging, this conclusion was not reached by
166 determining the positions of individual Rubisco holoenzymes, but rather by averaging together
167 larger volumes that contained neighborhoods of multiple Rubisco particles. The improved
168 quality of our new tomograms enabled us to pinpoint the positions of each Rubisco, revealing
169 that the "HCP lattice" is actually an average of a much more heterogeneous environment. In
170 HCP, the 12 nearest neighbors to any given particle are equidistant. However, when we plotted
171 the center-to-center distances between holoenzymes and their 12 nearest neighbors, we observed
172 a specific peak corresponding to 4.4 ± 1.8 neighbors that were 13.9 ± 1.5 nm away, with most of
173 the remaining neighbors at a range of longer distances, for an average distance of 15.9 ± 2.9 nm
174 (Figure 4E). The diameter of Rubisco is ~10-13 nm depending on orientation (Taylor *et al.*,
175 2001); thus, there is ~1-4 nm between a Rubisco and its nearest neighbors in the specific peak.

176 To directly compare the observed Rubisco packing to that of a crystal, we performed the
177 same local density analysis on simulated data in which the exact volume of the pyrenoid matrix
178 from the tomograms was replaced with one of two crystalline matrix models: one based on HCP
179 with 13.9 nm between particle centers, and the other on the structure observed in crystallized
180 Rubisco (Taylor *et al.*, 2001) (Figure 4F; Figure S4E-F). Compared to the experimental data,
181 both crystalline simulated data sets produced much taller short-range peaks along with
182 pronounced long-range peaks. Furthermore, the Rubisco concentration observed in the pyrenoid
183 matrix (377,856 ± 37,823 Rubiscos/μm^3; 628 ± 63 μM) was 28% lower than in the HCP
184 simulated data (526,039 ± 709 Rubiscos/μm^3; 874 ± 1 μM) (Figure S3). Therefore, we conclude
185 that the arrangement of Rubisco in the pyrenoid is neither HCP nor that observed in crystallized
186 Rubisco.
187 To explore whether the pyrenoid matrix has liquid-like organization, we next compared
188 the observed Rubisco packing to the distribution of molecules within known fluids. The local
189 density of Rubisco within the pyrenoid matrix fit well to the radial distribution function of a
190 simple model for liquid molecular interactions, the Lennard-Jones fluid (Johnson *et al.*, 1993)
191 (Figure 4G), determined both analytically (Morsali *et al.*, 2005) and by molecular dynamics
192 simulations (Plimpton, 1995). Additionally, similar radial distribution functions have been
193 measured for colloidal liquids (Gu *et al.*, 2010; Thorneywork *et al.*, 2014). Therefore, we
194 conclude that the radial distribution of Rubisco within the matrix resembles that of a liquid.
195 We sought to determine whether effects at the macromolecular scale could produce the
196 liquid-like organization that we observed in the pyrenoid matrix. We first examined whether
197 random packing of Rubisco particles could yield the local density seen in our experimental data.
198 Using Monte Carlo simulations, we randomly placed the same number of Rubisco particles as

199 the experimental data throughout identical pyrenoid matrix volumes (Figure 4H, Figure S4G).
200 The local density of this simulated data fit our experimental data poorly, showing that the
201 organization of pyrenoid matrix is not random.
202 We next randomly mapped Rubisco particles into the matrix volumes while imposing
203 specific distance constraints between either linked pairs of particles or interconnected linked
204 networks (Figure 4H, Figure S4H-J). As we increased the number of constraints on particle
205 spacing from the single constraint between paired particles to the numerous constraints between
206 particles in a network, the local density of Rubisco much more closely resembled the observed
207 distribution in the pyrenoid matrix. This result suggests that holoenzymes could be linked in a
208 network with a favored Rubisco-to-Rubisco distance. EPYC1 may be the molecular structure
209 that enforces this spacing.
210
211 **The Pyrenoid Mixes Internally**
212 The Rubisco packing distribution we observed is consistent with either a liquid or an
213 amorphous solid state. In order to discriminate between these states, a dynamic assay is
214 necessary. The most direct test of liquid-like behavior is whether the constituent components
215 exhibit internal mixing (Brangwynne *et al.*, 2009; Hyman *et al.*, 2014), which is traditionally
216 assayed *in vivo* by fluorescence recovery after photobleaching (FRAP) (Reits and Neefjes, 2001;
217 van Royen *et al.*, 2009; Ishikawa-Ankerhold *et al.*, 2012; Hyman *et al.*, 2014).
218 Strikingly, our FRAP experiments revealed that the pyrenoid matrix mixes internally
219 (Figure 5). After bleaching approximately half the volume of the matrix (Figure S5A-C), we
220 observed signal re-homogenization on a timescale of ~20 seconds (Figure 5A,C,G,I-L; Movie
221 S5), similar to that of established liquid-like compartments (Brangwynne *et al.*, 2009; Li *et al.*,

222 2012; Hubstenberger *et al.*, 2013; Kroschwald *et al.*, 2015; Nott *et al.*, 2015; Patel *et al.*, 2015;
223 Saha *et al.*, 2016). During recovery, the signal from the unbleached region decreased as the
224 signal in the bleached region increased, and a "wave" of fluorescence could be seen moving from
225 the unbleached to the bleached region (Figure 5C,E), suggesting that recovery is primarily due to
226 internal rearrangements, rather than import of stromal RbcS1-Venus. In contrast, pyrenoids in
227 chemically crosslinked cells showed no recovery (Figure 5B,D,F,G; Movie S5).

228 Further FRAP experiments demonstrated that the other major components of the
229 pyrenoid matrix, EPYC1 (Mackinder *et al.*, 2016) and Rubisco Activase (Mckay *et al.*, 1991;
230 Mackinder *et al.*, 2016) (RCA1), are also mobile (Figure 5G-L, Figure S5D-F). Differences in
231 the initial recovery rates of the three matrix components (Figure 5H) suggest that they may have
232 distinct mobilities within the matrix, apparently inversely proportional to their molecular weights
233 (Rubisco holoenzymes are ~540 kDa; RCA1 hexamers are ~270 kDa; EPYC1 is ~35 kDa). The
234 different mobilities suggest that although these matrix components can bind to each other, the
235 major mobile unit is not a complex between EPYC1 and Rubisco; thus, EPYC1-Rubisco binding
236 interactions (Mackinder *et al.*, 2016) are transient. Alongside our observations of liquid-like
237 fission (Figures 1,3) and local order (Figure 4), the internal mixing of matrix components
238 demonstrates that the *Chlamydomonas* pyrenoid matrix behaves like a liquid.

239

240 **The Pyrenoid Disperses During Cell Division**

241 A remarkable property of liquid-like protein compartments is their ability to transition
242 between an aggregated liquid phase and a dispersed soluble phase (Brangwynne *et al.*, 2009; Li
243 *et al.*, 2012; Weber and Brangwynne, 2012). Surprisingly, our experiments revealed that the
244 *Chlamydomonas* pyrenoid matrix also appears to undergo such a phase transition during

245 division. A portion of the RbcS1-Venus and EPYC1-Venus signals rapidly dispersed from the
246 pyrenoid matrix into the stroma for ~20 minutes near the end of chloroplast and pyrenoid
247 division (Figures 6, S6). During this time, the Venus signal in the chloroplast stroma rose
248 dramatically, while the signal in the pyrenoid dropped by ~35-50%. The Venus intensity in the
249 pyrenoid subsequently recovered, coincident with a reduction in the stromal signal (Figure 6B-
250 C). The increased stromal signal during chloroplast division is unlikely to be due to new protein
251 synthesis because these transitions occurred on the timescale of 1-5 minutes, and were not
252 associated with an increase in the total fluorescence signal (Figure 6B, Figure S6). Pyrenoid
253 fission nearly always occurred during the time of increased stromal signal (Figure 6B,D; Figure
254 S6), raising the intriguing possibility that the partial dispersal may be associated with a decrease
255 in matrix surface tension or viscosity, which could facilitate the progression of the cleavage
256 furrow through the pyrenoid.

257 We observed pyrenoid component dispersal regardless of the mode of pyrenoid
258 inheritance (Figure S6). During dispersal, puncta of matrix material often appeared transiently
259 throughout the stroma (Figures 1D, S1C-D; Movie S3). These *de novo* matrix protein puncta
260 may correspond to the "dense regions" that were observed in electron micrographs of dividing
261 *Chlamydomonas* cells (Goodenough, 1970). In some daughter cells that failed to inherit part or
262 all of the maternal pyrenoid during chloroplast division, we observed that these puncta grew into
263 apparent new pyrenoids (Figure 2B,D,G-I). Our observations suggest a model in which the
264 dispersal of the building blocks of the matrix serves as a redundant mechanism to fission,
265 facilitating equal partitioning of the pyrenoid matrix to daughter chloroplasts. The inheritance of
266 dispersed pyrenoid components enables rapid *de novo* pyrenoid assembly in daughter
267 chloroplasts that fail to inherit a pyrenoid by other means.

268

269 **A "Magic Number" Effect in Multicomponent Phase-Separation**

270 The liquid-like nature of the pyrenoid fits well with the principle that biological liquid
271 compartments are formed by weak multivalent binding between constituent proteins (Li *et al.*,
272 2012). The two major constituents of the pyrenoid, Rubisco and EPYC1, likely have multiple
273 binding sites for each other. The octameric symmetry of the Rubisco holoenzyme makes it
274 plausible that Rubisco has eight binding sites for EPYC1. Additionally, the four repeats of the
275 EPYC1 protein suggest that EPYC1 has four binding sites for Rubisco (Mackinder *et al.*, 2016).
276 We have previously proposed that dynamic regulation of the availability of EPYC1's binding
277 sites could cause a change in the aggregation state of the pyrenoid (Mackinder *et al.*, 2016).

278 In the model liquid-like system involving binding between the SRC homology 3 (SH3)
279 domain and a proline-rich motif ligand (PRM), it has been established that decreasing the
280 number of binding sites on flexible linker proteins promotes a phase transition from an
281 aggregated to a dispersed state (Li *et al.*, 2012). Surprisingly, modeling of the EPYC1-Rubisco
282 interaction predicts the existence of an additional effect that governs phase transitions in
283 multicomponent liquid-like biological systems.

284 Because the binding mechanism between Rubisco and EPYC1 is uncharacterized, we
285 developed a simplified computational model to investigate how interactions between these two
286 components may drive pyrenoid aggregation and dissolution. In our model, "Rubisco
287 holoenzymes" and "EPYC1 molecules" occupy a 2D square grid. Each Rubisco holoenzyme is
288 modeled as a 2x4 rectangle of 8 EPYC1 binding sites, and each EPYC1 is modeled as a flexible
289 chain of 4 Rubisco binding sites. Each binding site on Rubisco can bind to a single EPYC1 site,
290 and *vice versa.* Binding occurs when a Rubisco site and an EPYC1 site occupy the same grid

291 position. While our model is simplified and abstract, it allows the observation of certain
292 fundamental behaviors.

293 Our simulation produced the expected increased aggregation when we increased the
294 number of binding sites on EPYC1 from 4 to 5 (Figure 7B-C). However, contrary to the
295 established paradigm for liquid-like systems, we also observed increased aggregation when we
296 decreased the number of binding sites on EPYC1 from 4 to 3 (Figure 7A-B).

297 Further investigation into this previously unappreciated behavior revealed a “magic
298 number” effect in the simulated EPYC1-Rubisco system. Magic numbers occur in various
299 contexts in chemistry and physics (Sakurai *et al.*, 1999; Steppenbeck *et al.*, 2013), in which a
300 certain number of particles form an unusually stable state, such as filled electronic shells in
301 atoms. However, to our knowledge, magic numbers have not previously been observed in a
302 biological system. In our model, the magic number effect arises because all 8 binding sites of one
303 Rubisco can be exactly saturated by two EPYC1s with 4 binding sites each to form a stable
304 trimer, resulting in minimal aggregation. When the magic number stoichiometry is broken either
305 with more or fewer binding sites on EPYC1, larger aggregates form.

306 To ensure that the magic number effect is not an artifact of the lattice geometry or of two
307 dimensions, we also simulated the system using a more computationally intensive but more
308 realistic three-dimensional, off-lattice representation. We modeled each Rubisco as a sphere with
309 4 small spherical binding sites on each end, and each EPYC1 molecule as a polymer with 3, 4, or
310 5 binding sites (Figure 7G-I, Movie S6). When the EPYC1 and Rubisco binding sites
311 overlapped, specific bonds were formed. Similar to the lattice simulation results, we observed
312 more aggregation when we increased the number of binding sites on EPYC1 from 4 to 5 (Figure
313 7H, I) and when we decreased it from 4 to 3 (Figure 7G, H). Additionally, small trimer

314 complexes containing one Rubisco and two EPYC1s were commonly observed when the number
315 of binding sites on EPYC1 was 4, but these complexes were rarely seen when the number of
316 binding sites on EPYC1 was 3 or 5. Thus, the magic number effect persists in three-dimensions
317 and off-lattice. Furthermore, the effect is robust for a wide range of protein concentrations and
318 model parameters (Figure 7D-F, Figure S7, Table S1, Movie S6).

319 To explore the generality of this effect, we used the 2D lattice model to consider an even
320 wider range of the number of Rubisco binding sites on EPYC1 (Figure 7J-L), and found
321 additional magic numbers at 2 and 8 binding sites per EPYC1. When the number of binding sites
322 on EPYC1 is a magic number, small heterocomplexes are favored; consequently, higher protein
323 concentrations are required in order to form large aggregates (Figure 7J-L). These results are
324 robust to constraints such as restricting EPYC1 to only bind one end of each Rubisco (Figure
325 S7). Thus, in addition to the previously established general trend that increasing the number of
326 binding sites in a flexible linker protein promotes phase separation (Li *et al.*, 2012), there are
327 strong magic number effects that impact the phase diagram.

328

329 Discussion

330

331 Phase Separation Helps Explain Known Pyrenoid Physiology

332 Phase separation provides a new framework for understanding the structure, biogenesis,
333 and inheritance of the pyrenoid. Our observations suggest that the re-localization of Rubisco
334 from the pyrenoid to the stroma in response to high CO_2 (Borkhsenious *et al.*, 1998) or darkness
335 (Mitchell *et al.*, 2014) is a phase transition of the pyrenoid matrix. The use of phase transitions
336 could allow rapid reorganization of Rubisco to enhance CO_2 fixation in fluctuating
337 environmental conditions. Furthermore, the fluidity of the pyrenoid matrix resolves the paradox

338 of how Rubisco Activase chaperones can efficiently access the more abundant Rubisco active
339 sites throughout the matrix.
340

341 **EPYC1-Rubisco Interactions May Drive Pyrenoid Phase Separation**

342 The pyrenoid's liquid-like nature may be mediated by binding between Rubisco and
343 EPYC1's sequence repeats. Each EPYC1 repeat displays low complexity, with 73% of the repeat
344 regions consisting of just alanine, serine, proline, and arginine (Mackinder *et al.*, 2016). Low
345 complexity domains and repeat regions have been widely implicated in mediating the liquid-like
346 properties and phase separation of nonmembrane-bound organelles (Li *et al.*, 2012; Fromm *et*
347 *al.*, 2014; Bergeron-Sandoval *et al.*, 2016; Mitrea and Kriwacki, 2016). The exact nature of the
348 binding between Rubisco and EPYC1 is unknown, but it is likely that surface-exposed α-helices
349 on the Rubisco small subunit play a role in the binding interaction because a specific sequence in
350 these helices is required for Rubisco aggregation in *Chlamydomonas* (Meyer *et al.*, 2012). Weak
351 multivalent interactions have been implicated in mediating biological phase separation in other
352 liquid-like organelles (Li *et al.*, 2012; Hyman *et al.*, 2014). Thus, if EPYC1-Rubisco interactions
353 are weak, then these two binding partners could explain both the structure and the fluidity of the
354 matrix.
355

356 **A Magic Number Effect Could Facilitate Phase Transitions**

357 We speculate that a magic number effect could help explain the rapid phase transitions of
358 the pyrenoid matrix that we observed experimentally (Figure 6). The rapid partial dissolution and
359 reorganization of the pyrenoid suggests regulation at the level of EPYC1-Rubisco interactions.
360 The EPYC1 phosphorylation state changes under conditions that require a pyrenoid (Turkina *et*

361 *al.*, 2006; Wang *et al.*, 2014), suggesting that phosphorylation could be a mechanism to alter the
362 availability of Rubisco binding sites on EPYC1. In our simulations, the magic number effect
363 produces a dramatic phase change in response to the addition or removal of a single binding site,
364 from aggregation with 3 binding sites on EPYC1 to dissolution with 4 binding sites and back to
365 aggregation with 5 binding sites (Figure 7). Thus, as a result of the magic number effect, a
366 change in the affinity of only a single binding site on each EPYC1 may trigger a phase change.
367 Both lattice and off-lattice models indicate that clustering can be strongly suppressed if
368 the number of binding sites on one species (*e.g.*, Rubisco) is an integral multiple of the number
369 of binding sites on the other (*e.g.*, EPYC1), as this favors the assembly of small oligomers in
370 which all binding sites are saturated. In general, we expect this magic number effect to pertain to
371 multicomponent, multivalent binding systems where bonds are one-to-one and saturable.
372 Moreover, the effect requires these specific bonds to have an energy of several k_BT, strong
373 enough for most small oligomers to be fully bonded without defects (Figure 5 D-F). Since such
374 binding energies are quite common, magic number effects can be predicted broadly for
375 interacting protein pairs such as SH3-PRM (Li *et al.*, 2012) and SIM-SUMO (Banani *et al.*,
376 2016) as well as for RNA-protein droplets (Lin *et al.*, 2015). Additional insights could be gained
377 in the future by investigating the influence of polymer flexibility, stoichiometry, and mixed
378 valency populations on the magic number effect described in this work.
379

380 **Interactions Between the Pyrenoid Matrix and Other Chloroplast Structures Remain to be**
381 **Explored**

382 In *Chlamydomonas*, the pyrenoid matrix is located at the base of the chloroplast, is
383 traversed by tubules that are continuous with the thylakoid membranes, and is surrounded by a

384 starch sheath (Griffiths, 1970, 1980; Meyer and Griffiths, 2013; Engel *et al.,* 2015). It is not yet
385 known what controls the location of the pyrenoid, but we hypothesize that the matrix's liquid-
386 like nature may promote its localization around cellular components that have affinity for the
387 matrix, including possibly the pyrenoid tubules. Future investigations into the division and
388 possible *de novo* formation of the starch sheath and thylakoid tubules may provide insights into
389 the principles that underlie the interface between liquid-like organelles and other cellular
390 structures.
391
392 **Other Species May Also Have Liquid-Like Pyrenoids**
393 Our results suggest that the pyrenoid "regression" (reduction in size and disappearance)
394 observed in some algae during chloroplast division in zoosporogenesis (Brown and Arnott, 1970)
395 may be a phase transition from an aggregated to a soluble phase. Previous reports suggested that
396 different species of algae undergo either pyrenoid fission or pyrenoid regression during cell
397 division (Brown et al., 1967; Griffiths, 1970, 1980). However, our observations demonstrate that
398 both phenomena can occur simultaneously in the same cell.
399 The observation of pyrenoid fission and regression in other species of algae (Brown,
400 1964; Brown and Arnott, 1970; Griffiths, 1970, 1980), combined with the amorphous appearance
401 of the matrix of many species (Evans, 1966; Griffiths, 1970, 1980; Meyer *et al.*, 2012), suggests
402 that the liquid-like nature may be a general property of all pyrenoids. However, our data does not
403 exclude the possibility that the pyrenoid matrix of some species is more permanently crosslinked
404 or crystalline. As new genetic tools are developed, it will be particularly interesting to perform
405 experiments similar to those presented here on the diatom *Achnanthes brevipes*, where some of
406 the most striking crystalline matrix packing was observed (Holdsworth, 1968). If solid pyrenoids

407 do exist, it will be interesting to understand what benefits the solid phase can provide that

408 outweigh the advantages of a liquid for inheritance and protein mobility.

409 Pyrenoids are not only broadly distributed in the Chloroplastida, but are also found in

410 five out of the seven supergroups of the eukaryotic tree of life [Excavates, Stramenopiles,

411 Alveolates, Rhizaria and Archaeplastids (Burki, 2014)], and it is thought that pyrenoids have

412 evolved multiple times (Villarreal and Renner, 2012; Meyer and Griffiths, 2013). Intriguingly,

413 across sequenced algae, the presence of a pyrenoid is correlated with the presence of a protein

414 with EPYC1-like physiochemical properties, such as a low-complexity repeat sequence, a high

415 pI, and no transmembrane domains (Mackinder *et al.*, 2016). If the only requirement for

416 aggregating diffuse Rubisco into a liquid-like pyrenoid matrix is the evolution of a linker protein

417 with multiple weak binding sites for Rubisco, pyrenoids could be relatively simple to evolve.

418

419 **Pyrenoid Behavior is Distinct from that of Carboxysomes**

420 Cyanobacteria have a functionally similar structure to the pyrenoid called the

421 carboxysome. Like the pyrenoid, carboxysomes contain aggregated Rubisco and are not bound

422 by a membrane. Unlike the *Chlamydomonas* pyrenoid, carboxysomes are not traversed by

423 membrane tubules or surrounded by a starch sheath, but are instead surrounded by an icosahedral

424 protein shell. There are two classes of carboxysomes, α and β, which are thought to have arisen

425 through convergent evolution (Rae *et al.,* 2013).

426 Rubisco in β-carboxysomes is thought to be linked together by CcmM, a protein

427 containing multiple Rubisco small subunit-like domains, each of which is thought to be

428 incorporated into a different Rubisco holoenzyme (Long *et al.*, 2010). Such incorporation would

429 be expected to preclude internal mixing. Indeed, in contrast to our results in pyrenoids,

430 fluorescently-tagged Rubisco in β-carboxysomes does not recover after photobleaching (Chen *et*
431 *al.*, 2013).

432 Less is known about the arrangement of Rubisco in α-carboxysomes. The proposed α-
433 carboxysome Rubisco linker protein CsoS2 (Cai *et al.*, 2015) is a disordered repeat protein,
434 raising the possibility that the α-carboxysome may have liquid-like characteristics. Whereas we
435 have observed pyrenoid biogenesis by both fission and *de novo* assembly, biogenesis of new α-
436 and β-carboxysomes has only been observed to occur *de novo* (Iancu *et al.*, 2010; Cameron *et al.*,
437 2013; Chen *et al.*, 2013; Kerfeld and Melnicki, 2016). It will be interesting to see whether α-
438 carboxysomes do exhibit liquid-like behavior, or whether the underlying structural and
439 biogenesis principles are a general distinguishing feature between carboxysomes and pyrenoids.
440

441 **The Pyrenoid Provides Insights into Liquid-Like Organelle Biology**

442 To our knowledge, the pyrenoid is now the first example of a liquid-like organelle
443 specific to photosynthetic organisms. Like other previously-characterized liquid-like organelles,
444 new pyrenoids can be generated by either *de novo* aggregation or fission into two daughters
445 (Brangwynne, 2011; Brangwynne *et al.*, 2011; Weber and Brangwynne, 2015; Saha *et al.*, 2016).
446 The first characterized liquid-like organelle, the *C. elegans* P granule, leverages the properties of
447 phase transitions to promote its own asymmetric inheritance by preferentially condensing at one
448 end of the cell before cell division (Brangwynne *et al.*, 2009; Saha *et al.*, 2016). In contrast, the
449 pyrenoid appears to leverage the liquid-like properties of partial dissolution and fission to
450 promote symmetric inheritance, allowing both daughter cells to inherit a pyrenoid.
451
452

Author Contributions

E.S.F.R. performed the FRAP and division experiments and data analysis. L.C.M.M. generated the fluorescent constructs and made the initial FRAP observation. E.S.F.R., L.C.M.M., and M.C.J. designed and interpreted the FRAP and division work. H.N.C. assisted confocal data acquisition and analysis. M. Schaffer and B.D.E. performed cryo-ET data acquisition. L.K.C. and M. Strauss performed the cryo-ET image analysis, while A.M.S. performed the cryo-ET statistical analysis and simulated data generation. A.M.S. and B.D.E performed the Lennard-Jones analytical fit, and B.X. performed the Lennard-Jones simulation fit. J.M.P., F.F., and B.D.E. supervised the cryo-ET work. B.X. and N.S.W. designed and interpreted, and B.X. performed, the on-lattice EPYC1-Rubisco interaction modeling work, and B.X. and P.R. performed the 3D off-lattice simulation. E.S.F.R., B.X., N.S.W., B.D.E., and M.C.J. wrote the text with input from all authors.

Acknowledgements

We thank W. Baumeister, J. Berry, D. Ehrhardt, J. Feldman, A. Grossman, Y. Meir, M.B. Mudgett, M. Scott, H. Stone, V. Walbot, and members of the Jonikas laboratory for helpful discussion and suggestions; and U. Goodenough, A. Grossman, A. McCormick, and M.T. Meyer for critical reading of the manuscript. This work was supported by National Science Foundation Grants EF-1105617 and IOS-1359682 (M.C.J.), and PHY-1305525 (N.S.W.); the Carnegie Institution for Science (L.C.M.M. and M.C.J.); National Institutes of Health Grant T32GM007276 (E.S.F.R.); the Simons Foundation and HHMI grant #55108535 and Princeton University (M.C.J.); a CONACyT-DAAD Graduate Scholarship (L.K.C); a Fundación Séneca Postdoctoral Fellowship (A.M); an Alexander von Humboldt Foundation Postdoctoral

476 Fellowship (B.D.E); and Deutsche Forschungsgemeinschaft Grant FO 716/4-1 (F.F.). The

477 authors declare no competing financial interests.

478 **References**

479 Asano, S., Engel, B.D. and Baumeister, W. (2016). *In situ* cryo-electron tomography: a post-

480 reductionist approach to structural biology. J. Mol. Biol. 4*28*, 332–343.

481 Badger, M.R., Andrews, T.J., Whitney, S.M., Ludwig, M., Yellowlees, D.C., Leggat, W. and

482 Price, G.D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and

483 chloroplast-based CO_2-concentrating mechanisms in algae. Can. J. Bot. *76*, 1052–1071.

484 Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016).

485 Compositional control of phase-separated cellular bodies. Cell *166*, 651-663.

486 Bergeron-Sandoval, L P., Safaee, N. and Michnick, S.W. (2016). Mechanisms and consequences

487 of macromolecular phase separation. Cell *165*, 1067–1079.

488 Bertagnolli, B.L. and Nadakavukaren, M.J. (1970). An ultrastructural study of pyrenoids from

489 *Chlorella pyrenoidosa*. J. Cell Sci. *7*, pp. 623–630.

490 Bharat, T.A.M. and Scheres, S.H.W. (2016). Resolving macromolecular structures from electron

491 cryo-tomography data using subtomogram averaging in RELION. Nat. Protoc. *11*, 2054–

492 2065.

493 Bisalputra, T. and Weier, T.E. (1964). The pyrenoid of *Scenedesmus quadricauda*. Am. J. Bot.

494 *51*, 881–892.

495 Borkhsenious, O.N., Mason, C.B. and Moroney, J.V. (1998). The intracellular localization of

496 ribulose-1,5-bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. Plant

497 Phys. *116*, 1585–91.

498 Brangwynne, C.P. (2011). Soft active aggregates: mechanics, dynamics and self-assembly of

499 liquid-like intracellular protein bodies. Soft Matter *7*, 3052.

500 Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J.,

501 Jülicher, F. and Hyman, A.A. (2009). Germline P granules are liquid droplets that
502 localize by controlled dissolution/condensation. Science *324*, 1729–1732.
503 Brangwynne, C.P., Mitchison, T.J. and Hyman, A.A. (2011). Active liquid-like behavior of
504 nucleoli determines their size and shape in *Xenopus laevis* oocytes. PNAS *108*, 4334–
505 4339.
506 Brown, R.M. Jr. (1964). Comparative studies of the algal genera Tetracystis and Chlorococcum.
507 PhD Thesis (The University of Texas).
508 Brown, R.M. Jr. and Arnott, H.J. (1970). Structure and function of the algal pyrenoid. I.
509 Ultrastructure and cytochemistry during zoosporogenesis of *Tetracystis excentrica*. J.
510 Phycol. *6*, 14–22.
511 Brown, R.M. Jr., Arnott, H.J., Bisalputra, T. and Hoffman, L.R. (1967). The pyrenoid: Its
512 structure, distribution, and function. J. Phycol. 3, supp. 5–6.
513 Buffaloe, N.D. (1958). A comparative cytological study of four species of *Chlamydomonas*.
514 Bull. Torrey Bot. Club *85*, 157–178.
515 Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. Cold
516 Spring Harb. Perspect. Biol. *6*, a016147.
517 Cai, F., Dou, Z., Bernstein, S., Leverenz, R., Williams, E., Heinhorst, S., Shively, J., Cannon, G.
518 and Kerfeld, C. (2015). Advances in understanding carboxysome assembly in
519 *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. Life *5,*
520 1141-1171.
521 Cameron, J.C., Wilson, S.C., Bernstein, S.L. and Kerfeld, C.A. (2013). Biogenesis of a bacterial
522 organelle: the carboxysome assembly pathway. Cell *155*, 1131–1140.
523 Chen, A.H., Robinson-Mosher, A., Savage, D.F., Silver, P.A. and Polka, J.K. (2013). The

524 bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo.
525 PLoS ONE *8*, e76127.
526 Cheng, S., Cetinkaya, M., and Gräter, F. (2010). How sequence determines elasticity of
527 disordered proteins. Biophys. J. *99*, 3863-3869.
528 Conduit, P.T., Brunk, K., Dobbelaere, J., Dix, C.I., Lucas, E.P. and Raff, J.W. (2010). Centrioles
529 regulate centrosome size by controlling the rate of Cnn incorporation into the PCM. Curr.
530 Biol. *20*, 2178–2186.
531 Conduit, P.T., Wainman, A., Novak, Z.A., Weil, T.T. and Raff, J.W. (2015). Re-examining the
532 role of *Drosophila* Sas-4 in centrosome assembly using two-colour-3D-SIM FRAP. eLife
533 *4*, e08483.
534 Crang, R.F.E. and Klomparens, K.L. (1988) *Artifacts in biological electron microscopy*. (Plenum
535 Press).
536 Cross, F.R. and Umen, J.G. (2015). The *Chlamydomonas* cell cycle. Plant J. *82*, 370–392.
537 Eichel, K., Jullié, D. and von Zastrow, M. (2016). β-Arrestin drives MAP kinase signalling from
538 clathrin-coated structures after GPCR dissociation. Nat. Cell Biol. *18*, 303–310.
539 Engel, B.D., Schaffer, M., Cuellar, L.K., Villa, E., Plitzko, J.M. and Baumeister, W. (2015).
540 Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron
541 tomography. eLife *4*, e04889.
542 Evans, L.V. (1966). Distribution of pyrenoids among some brown algae. J. Cell Sci. *1*, 499-454.
543 Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W.,
544 Pappu, R.V. and Brangwynne, C.P. (2016). Coexisting liquid phases underlie nucleolar
545 subcompartments. Cell *165*, 1686–1697.
546 Förster, F., Han, B. and Beck, M. (2010). Visual proteomics. Method. Enzymol. *483*, 215–253.

547 Förster, F., Pruggnaller, S., Seybert, A. and Frangakis, A.S. (2008). Classification of cryo-
548 electron sub-tomograms using constrained correlation. J. Struct. Biol. *161*, 276–286.
549 Fromm, S.A., Kamenz, J., Nöldeke, E.R., Neu, A., Zocher, G. and Sprangers, R. (2014). *In vitro*
550 reconstitution of a cellular phase-transition process that involves the mRNA decapping
551 machinery. Angew. Chem. Int. Ed. *53*, 7354–7359.
552 Gibbs, S.P. (1962). The ultrastructure of the pyrenoids of green algae. J. Ultrastruct. Res. *7*, 262–
553 272.
554 Goodenough, U.W. (1970). Chloroplast division and pyrenoid formation in *Chlamydomonas*
555 *reinhardtii*. J. Phycol. *6*, 1–6.
556 Griffiths, D.J. (1970). The pyrenoid. Bot. Rev. *36*, 29–58.
557 Griffiths, D.J. (1980). The pyrenoid and its role in algal metabolism. Sci. Prog. *66*, 537–553.
558 Gu, L., Xu, S., Sun, Z. and Wang, J.T. (2010). Brownian dynamics simulation of the
559 crystallization dynamics of charged colloidal particles. J. Colloid Interface Sci. *350*, 409–
560 416.
561 Hirakawa, Y. and Ishida, K. (2015). Prospective function of FtsZ proteins in the secondary
562 plastid of chlorarachniophyte algae. *BMC plant biology*. BMC Plant Biol. *15*, 276.
563 Hofmann, H., Soranno, A., Borgia, A., Gast, K., Nettels, D., and Schuler, B. (2012). Polymer
564 scaling laws of unfolded and intrinsically disordered proteins quantified with single-
565 molecule spectroscopy. PNAS *109*, 16155-16160.
566 Holdsworth, R.H. (1968). The presence of a crystalline matrix in pyrenoids of the diatom,
567 *Achnanthes brevipes*. JCB *37*, 831–837.
568 Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L.K., Mangold, A. and Förster, F. (2012). PyTom: A
569 python-based toolbox for localization of macromolecules in cryo-electron tomograms and

570 subtomogram analysis. J. Struct. Biol. *178*, 177–188.
571 Hubstenberger, A., Noble, S.L., Cameron, C. and Evans, T.C. (2013). Translation repressors, an
572 RNA helicase, and developmental cues control RNP phase transitions during early
573 development. Dev. Cell *27*, 161–173.
574 Hyman, A.A., Weber, C.A. and Jülicher, F. (2014). Liquid-liquid phase separation in biology.
575 Ann. Rev. Cell Dev. Biol. *30*, 39–58.
576 Iancu, C.V., Morris, D.M., Dou, Z., Heinhorst, S., Cannon, G.C. and Jensen, G.J. (2010).
577 Organization, structure, and assembly of α-carboxysomes determined by electron
578 cryotomography of intact cells. J. Molec. Biol. *396*, 105–117.
579 Ishikawa-Ankerhold, H.C., Ankerhold, R. and Drummen, G.P.C. (2012). Advanced fluorescence
580 microscopy techniques-FRAP, FLIP, FLAP, FRET and FLIM. Molecules *17*, 4047–4132.
581 Johnson, J.K., Zollweg, J.A. and Gubbins, K.E. (1993). The Lennard-Jones equation of state
582 revisited. Mol. Phys. *78*, 591–618.
583 Kerfeld, C.A. and Melnicki, M.R. (2016). Assembly, function and evolution of cyanobacterial
584 carboxysomes. Curr. Op. Plant Biol. *31*, 66–75.
585 Keown, J.R., Griffin, M.D., Mertens, H.D., & Pearce, F.G. (2013). Small oligomers of ribulose-
586 bisphosphate carboxylase/oxygenase (Rubisco) activase are required for biological
587 activity. JBC *288*, 20607-20615.
588 Klammt, C., Novotná, L., Li, D.T., Wolf, M., Blount, A., Zhang, K., Fitchett, J.R. and
589 Lillemeier, B.F. (2015). T cell receptor dwell times control the kinase activity of Zap70.
590 Nat. Immunol. *16*, 961–969.
591 Kowallik, K. (1969). The crystal lattice of the pyrenoid matrix of *Prorocentrum micans*. J. Cell
592 Sci. *5*, 251–269.

593 Kremer, J.R., Mastronarde, D.N. and McIntosh, J.R. (1996). Computer visualization of three-
594 dimensional image data using IMOD. J. Struct. Biol. *76*, 71–76.
595 Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S.
596 S. and Malasarn, D. (2012). A revised mineral nutrient supplement increases biomass and
597 growth rate in *Chlamydomonas reinhardtii*. Plant J. *66*, 770–780.
598 Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nüske, E., Poser, I., Richter, D. and
599 Alberti, S. (2015). Promiscuous interactions and protein disaggregases determine the
600 material state of stress-inducible RNP granules. eLife *4*, e06807.
601 Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V,
602 King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.-X., Nixon, B.T. and Rosen, M.K. (2012).
603 Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.
604 Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and maturation of phase-
605 separated liquid droplets by RNA-binding proteins. Mol. Cell *60*, 208-219.
606 Long, B.M., Tucker, L., Badger, M.R. and Price, G.D. (2010). Functional cyanobacterial beta-
607 carboxysomes have an absolute requirement for both long and short forms of the CcmM
608 protein. Plant Phys. *153*, 285–93.
609 Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O.,
610 Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., Roth, R., *et al.* (2016).
611 A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle.
612 PNAS *113*, 5958–5963.
613 Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction
614 of specimen movements. J. Struct. Biol. *152*, 36–51.
615 Mckay, R.M.L., Gibbs, S.P. and Vaughn, K.C. (1991). RuBisCo activase is present in the

616 pyrenoid of green algae. Protoplasma *162*, 38–45.

617 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953).

618 Equation of state calculations by fast computing machines. J. Chem. Phys. *21*, 1087–

619 1092.

620 Meyer, M.T., Genkov, T., Skepper, J.N., Jouhet, J., Mitchell, M.C., Spreitzer, R.J. and Griffiths,

621 H. (2012). Rubisco small-subunit α-helices control pyrenoid formation in

622 *Chlamydomonas*. PNAS *109*, 19474–19479.

623 Meyer, M.T. and Griffiths, H. (2013). Origins and diversity of eukaryotic CO_2-concentrating

624 mechanisms: lessons for the future. J. Exp. Bot. *64*, 769–786.

625 Mitchell, M.C., Griffiths, H. and Meyer, M.T. (2014). Dynamics of carbon concentrating

626 mechanism induction and protein re-localisation during the dark to light transition in

627 synchronised *Chlamydomonas*. Plant Phys. *166*, 1073-1082.

628 Mitrea, D.M. and Kriwacki, R.W. (2016). Phase separation in biology; functional organization of

629 a higher order. CCS *14*, 1.

630 Morsali, A., Goharshadi, E.K., Mansoori, G.A. and Abbaspour, M. (2005). An accurate

631 expression for radial distribution function of the Lennard-Jones fluid. *Chem. Phys. 310*,

632 11–15.

633 Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-

634 Jones, D.P., Pawson, T., Forman-Kay, J.D. and Baldwin, A.J. (2015). Phase transition of

635 a disordered nuage protein generates environmentally responsive membraneless

636 organelles. Molec. Cell *57*, 936–947.

637 Ohad, I., Siekevitz, P. and Palade, G.E. (1967). Biogenesis of chloroplast membranes I. Plastid

638 dedifferentiation in a dark-grown algal mutant (*Chlamydomonas reinhardi*). JCB *35*,

639 521–552.
640 Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid,
641 J., Saha, S., Franzmann, T.M., *et al.* (2015). A liquid-to-solid phase transition of the ALS
642 protein FUS accelerated by disease mutation. Cell *162*, 1066–1077.
643 Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. J. Comput.
644 Phys. *117*, 1–19. http://lammps.sandia.gov
645 Pollock, S.V, Colombo, S.L., Prout, D.L.J., Godfrey, A.C. and Moroney, J.V. (2003). Rubisco
646 activase is required for optimal photosynthesis in the green alga *Chlamydomonas*
647 *reinhardtii* in a low-CO_2 atmosphere. Plant Phys. *133*, 1854–1861.
648 Portis, A.R. (2003). Rubisco activase - Rubisco's catalytic chaperone. Photosynth. Res. *75*, 11–
649 27.
650 Rae, B.D., Long B.M., Whitehead L.F., Förster B., Badger M.R., Price G.D. (2013).
651 Cyanobacterial carboxysomes: microcompartments that facilitate CO_2 fixation. J. Mol.
652 Microbiol. Biotechnol. *23*, 300–307.
653 Reits, E.A.J. and Neefjes, J.J. (2001). From fixed to FRAP: measuring protein mobility and
654 activity in living cells. Nat. Cell Biol. *3*, E145–E147.
655 Retallack, B. and Butler, R.D. (1970). The development and structure of pyrenoids in
656 *Bulbochaete hiloensis*. J. Cell Sci. *6*, 229–241.
657 van Royen, M.E., Farla, P., Mattern, K.A., Geverts, B., Trapman, J. and Houtsmuller, A.B.
658 (2009). Fluorescence recovery after photobleaching (FRAP) to study nuclear protein
659 dynamics in living cells. Methods Mol. Biol. *464*, 363–385.
660 Saha, S., Weber, C.A., Nousch, M., Eckmann, C.R., Hyman, A.A., Saha, S., Weber, C.A.,
661 Nousch, M., Adame-arana, O., Hoege, C., *et al.* (2016). Polar positioning of phase-

662 separated liquid compartments in cells regulated by an mRNA competition mechanism.
663 Cell *166*, 1572–1584.
664 Sakurai, M., Watanabe, K., Sumiyama, K. and Suzuki, K. (1999). Magic numbers in transition
665 metal (Fe, Ti, Zr, Nb, and Ta) clusters observed by time-of-flight mass spectrometry. J.
666 Chem. Phys. *111*, 235–238.
667 Schaffer, M., Engel, B.D., Laugks, T., Mahamid, J. and Plitzko, J.M. (2015). Cryo-focused ion
668 beam sample preparation for imaging vitreous cells by cryo-electron tomography. Bio-
669 Protoc. *5*, e1575.
670 Schaffer, M., Mahamid, J., Engel, B.D., Laugks, T., Baumeister, W. and Plitzko, J.M. (2017).
671 Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies
672 of membrane proteins. J. Struct. Biol. *197*, 73–82.
673 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch,
674 S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for
675 biological-image analysis. Nat. Methods *9*, 676–682.
676 Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., Toettcher, J.E. and Brangwynne, C.P. (2016).
677 Spatiotemporal control of intracellular phase transitions using light-activated
678 optodroplets. Cell *168*, 430–436.
679 Steppenbeck, D., Takeuchi, S., Aoi, N., Doornenbal, P., Matsushita, M., Wang, H., Baba, H.,
680 Fukuda, N., Go, S., Honma, M., *et al.* (2013). Evidence for a new nuclear “magic
681 number” from the level structure of ^{54}Ca. Nature *502*, 207–210.
682 Stone, H.A. (1994). Dynamics of drop deformation and breakup in viscous fluids. Annu. Rev.
683 Fluid Mech. *26*, 65–102.
684 Taylor, T.C., Backlund, A., Bjorhall, K., Spreitzer, R.J. and Andersson, I. (2001). First crystal

685 structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*. JBC *276*, 48159–64.
686 Thévenaz, U.E. Ruttimann, M.U. (1998). A pyramid approach to subpixel registration based on
687 intensity. IEEE Trans. Image Process. *7*, 27–41.
688 Thorneywork, A.L., Roth, R., Aarts, D.G.A.L. and Dullens, R.P.A. (2014). Communication:
689 Radial distribution functions in a two-dimensional binary colloidal hard sphere system. J.
690 Chem. Phys. *140*, 161106 .
691 Turkina, M.V, Blanco-Rivero, A., Vainonen, J.P., Vener, A.V and Villarejo, A. (2006). CO_2
692 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas*
693 *reinhardtii*. Proteomics *6*, 2693–704.
694 Umen, J.G. and Goodenough, U.W. (2001). Control of cell division by a retinoblastoma protein
695 homolog in *Chlamydomonas*. Genes Dev. *15*, 1652–1661.
696 Vaucher, J.-P. (1803). Histoire des conferves d'eau douce, contenant leurs différens modes de
697 reproduction, et la description de leurs principales espèces, suivie de l'histoire des
698 trémelles et des ulves d'eau douce. (J.J. Paschoud).
699 Villarreal, J.C. and Renner, S.S. (2012). Hornwort pyrenoids, carbon-concentrating structures,
700 evolved and were lost at least five times during the last 100 million years. PNAS *109*,
701 18873–18878.
702 Wales, D.J. and Doye, J.P.K. (1997). Global optimization by basin-hopping and the lowest
703 energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem.
704 *101*, 5111–5116.
705 Wang, H., Gau, B., Slade, W. O., Juergens, M., Li, P. and Hicks, L.M. (2014). The global
706 phosphoproteome of *Chlamydomonas reinhardtii* reveals complex organellar
707 phosphorylation in the flagella and thylakoid membrane. Molec. Cell. Proteom. *13*, 2337-

708 2353.
709 Wang, Y., Stessman, D.J. and Spalding, M.H. (2015). The CO_2 concentrating mechanism and
710 photosynthetic carbon assimilation in limiting CO_2: how *Chlamydomonas* works against
711 the gradient. Plant J. *82*, 429–448.
712 Weber, S.C. and Brangwynne, C.P. (2012). Getting RNA and protein in phase. Cell *149*, 1188–
713 1191.
714 Weber, S.C. and Brangwynne, C.P. (2015). Inverse size scaling of the nucleolus by a
715 concentration-dependent phase transition. Curr. Biol. *25*, 641–646.
716 Wiegand, T. and Moloney, K.A. (2004). Rings, circles, and null-models for point pattern
717 analysis in ecology. Oikos *104*, 209–229.
718 Wienkoop, S., Weiss, J., May, P., Kempa, S., Irgang, S., Recuenco-Munoz, L., Pietzke, M.,
719 Schwemmer, T., Rupprecht, J., Egelhofer, V. and Weckwerth, W. (2010). Targeted
720 proteomics for *Chlamydomonas reinhardtii* combined with rapid subcellular protein
721 fractionation, metabolomics and metabolic flux analyses. Molec. BioSys. *6*, 1018–31.
722 Yanashima, R., Garcia, A.A., Aldridge, J., Weiss, N., Hayes, M.A. and Andrews, J.H. (2012).
723 Cutting a drop of water pinned by wire loops using a superhydrophobic surface and knife.
724 PLoS ONE *7*, e45893.
725 Zhang, R., Patena, W., Armbruster, U., Gang, S.S., Blum, S.R. and Jonikas, M.C. (2014). High-
726 throughput genotyping of green algal mutants reveals random distribution of mutagenic
727 insertion sites and endonucleolytic cleavage of transforming DNA. Plant Cell *26*, 1398–
728 1409.
729
730

731 **Main Figure Titles and Legends**

732

733 **Figure 1. The Pyrenoid is Inherited Primarily by Fission.**

734 (**A** and **B)** Confocal Z-sum images of pyrenoid divisions by fission, with chlorophyll

735 autofluorescence shown in magenta, and RbcS1-Venus in green; *t*=0 is the first observation of a

736 gap in chlorophyll between the daughter pyrenoids in the first division shown. Dashed curves

737 represent approximate chloroplast outlines in the mother (white) and daughter (pink) cells.

738 (**C**) A cartoon of the approximate locations of the pyrenoid (green), chloroplast (magenta), and

739 cell membrane (black outline).

740 (**D**) Example of pyrenoid fission in EPYC1-Venus, annotated as in (B).

741 (**E**) Example of the progressing chloroplast cleavage furrow (arrows) appearing to separate

742 daughter pyrenoids. Images are 2D snapshots of 3D Z-stack reconstructions.

743 (**F**) Average and standard deviation of the durations of chloroplast (magenta) and pyrenoid

744 (green) fissions in RbcS1-Venus (left; *n* = 28 1st and 2nd divisions) and EPYC1-Venus (right; *n* =

745 22 1st and 25 2nd divisions).

746 (**G** and **H**) Duration and relative timing of chloroplast (magenta) and pyrenoid (green) division

747 for the pyrenoid fissions plotted in (F). Each bar represents a different division.

748 See also Figures S1-2, and Movie S1.

749

750 **Figure 2. Pyrenoids Can Also be Inherited by Other Means.**

751 (**A** – **D**) Examples of other types of pyrenoid inheritance patterns observed in RbcS1-Venus (A,
752 B) and EPYC1-Venus (C, D) cells. (A, C) One daughter (blue) inherits an entire pyrenoid from
753 the mother cell (white) and another daughter (orange) inherits neither a pyrenoid nor puncta. (B,
754 D) One daughter (blue) inherits the entire pyrenoid, and puncta appear in the other daughter
755 (yellow) and coalesce into a new pyrenoid.

756 (**E** – **F**) Proportion of observed RbcS1-Venus (E) and EPYC1-Venus (F) daughter cells that
757 exhibited each observed inheritance pattern; the distribution of inheritance patterns in EPYC1-
758 Venus cells was not significantly different from that of RbcS1-Venus cells (Chi-square test, p =
759 0.8).

760 (**G** – **I**) Stills from time course image captures in which pyrenoids were observed to grow or
761 coalesce from puncta that appeared in the chloroplast stroma during division. (G – H) RbcS1-
762 Venus; (I) EPYC1-Venus. Images are 2D projections of the sum of pixel values in each channel
763 in a Z-stack through the whole cell at each time point. The chloroplast of the dividing cell of
764 interest in each series is outlined in white. Arrows point to growing or coalescing pyrenoids. t=0
765 is defined as the first minute at which the daughter chloroplasts are observed to be distinct in 3D.
766 See also Figures S1-2, and Movies S2-3.

767

768 **Figure 3. A "Bridge" of Matrix Material Connects Nascent Daughter Pyrenoids During**

769 **Fission.**

770 (**A** – **E**) Examples of pyrenoid fissions in five RbcS1-Venus cells. Magenta is chlorophyll

771 autofluorescence; green is RbcS1-Venus. Images are 2D projections of the sum of pixel values in

772 each channel from a Z-stack through the whole cell.

773 (**F**) Example of pyrenoid fission in an EPYC1-Venus cell. Magenta is chlorophyll

774 autofluorescence; green is EPYC1-Venus. Images are 2D projections of the sum of pixel values

775 in each channel from a Z-stack through the whole cell.

776 See also Movie S1.

777

778 **Figure 4. The Pyrenoid Matrix is not Crystalline but Exhibits Short-Range Liquid-Like**

779 **Order.**

780 (**A**) Slice through a tomographic volume of the native *Chlamydomonas* pyrenoid.

781 (**B**) Segmentation of the tomogram shown in (A) with localized positions of 46,567 Rubisco

782 holoenzymes (magenta) mapped into the volume. Green and yellow: pyrenoid tubule

783 membranes.

784 (**C**) *In situ* subtomogram average of Rubisco (16.5 Å resolution; Figure S4A) generated from

785 30,000 particles extracted from the tomogram shown in (A).

786 (**D**) The local density of neighbor Rubisco particles as a function of the distance from the

787 reference particle. Each line represents a separate tomogram, showing the sum of the local

788 densities around every Rubisco. The distances to peaks of high local Rubisco concentration are

789 indicated.

790 (**E**) Histogram of distances from reference particles to their nearest neighbors (NN), summed

791 from all five tomograms. Red dashed line: Gaussian distribution fit to the 13.9 nm NN peak.

792 Light blue bars: distance to the 12 NN within 1 standard deviation (<1 SD) of the 13.9 nm peak,

793 dark blue bars: distance to the 12 NN beyond 1 standard deviation (>1 SD) from the 13.9 nm

794 peak, grey bars: distance to further (13+ NN) neighbors. Mean distance to the 12 NN = 15.9 nm.

795 Inset: distribution of the number of neighbors per reference particle (mean = 4.4 neighbors) that

796 are <1 SD from the 13.9 nm peak.

797 (**F**) The normalized local density of neighbor particles (local density divided by the global

798 density), showing the mean value ± 99% CI of the experimental data (black) compared to

799 crystalline simulated data generated within the same tomogram volumes (Figure S4E,F): crystal

800 structure packing (Taylor *et al.*, 2001) (red), 13.9 nm-spaced HCP (blue).

801 (**G**) The mean value ± 99% CI of the experimental data's normalized local density (black) fit
802 with the radial distribution function of a Lennard-Jones fluid generated by an analytical model
803 (Morsali *et al.*, 2005) (red) and by molecular dynamics simulations (Plimpton, 1995) (blue).
804 (**H**) The normalized local density of neighbor particles, showing the mean value ± 99% CI of the
805 experimental data (black) compared to random simulated data generated within the same
806 tomogram volumes (Figure S4G-J): single particles (red), pairs linked by 13.9 ± 1.5 nm
807 (yellow), and linked networks (blue).
808 See also Figures S3-4, and Movie S4.
809

810 **Figure 5. Pyrenoid Matrix Components Mix Internally.**

811 (**A** and **B**) FRAP in live (A) and fixed (B) RbcS1-Venus pyrenoids. Cartoons depict the

812 approximate bleached region (dark gray). Different intensity display scales are used in the pre-

813 and post-bleach image sets.

814 (**C – D**) Kymographs of the pyrenoids shown in parts (A-B), respectively. From left to right: the

815 pyrenoid cartoons from (A-B), showing the region used to create the kymographs (dashed

816 rectangle); the pre-bleach section of the kymographs; and the post-bleach kymographs.

817 (**E – F**) Fluorescence recovery occurs from within the pyrenoid in live pyrenoids (E), but does

818 not occur in fixed pyrenoids (F). The *x*-axis is µm along the dashed regions in (C- D).

819 (**G**) Average fluorescence recovery profiles ± SEM for pyrenoids in live RbcS1-Venus (blue),

820 RCA1-Venus (red), or EPYC1-Venus (yellow) cells, and in fixed RbcS1-Venus cells (gray).

821 (**H**) Average recovery rates ± SEM over the first 12 seconds in (G). ** $p < 0.005$; * $p < 0.05$

822 (one-way ANOVA & post-hoc Bonferroni means comparison;).

823 (**I** – **L**) Examples of half-pyrenoid FRAP in live RCA-Venus (I, K) and EPYC1-Venus (J, L)

824 cells, with images from the recovery time-courses (I, J) and corresponding kymographs (K, L) as

825 shown in (A – D).

826 See also Figure S5 and Movie S5.

827 **Figure 6. The Pyrenoid Matrix Disperses and Re-Aggregates During Division.**

828 (**A**) Heat maps of the RbcS1-Venus signal during the divisions in Figure 1A-C. Times are

829 defined as in Figure 1**.**

830 (**B**) Raw signal from (A) plotted over time by regions of interest, representing the sum through

831 the whole Z-stack in each masked region over time. Times of pyrenoid divisions are highlighted

832 in gray.

833 (**C**) The average signal in the pyrenoid during division is significantly lower than that 15 minutes

834 later (Wilcoxon Matched-Pairs Signed-Ranks Test; * $p \leq 1.50^{-6}$; $n = 31$), shown ± SEM.

835 (**D**) Timeline of an average cell division with pyrenoid fission. Chloroplast division (magenta),

836 pyrenoid dissolution (cyan), and pyrenoid fission (green) are displayed relative to the moment

837 the chloroplast division furrow passes between the daughter pyrenoids (t=0). Cartoons depicting

838 the aggregation state of the pyrenoid matrix are shown above each stage, with the chloroplast

839 outlined in black, aggregated matrix components shown as filled black circles, and partially

840 dispersed matrix components as speckles.

841 See also Figure S6, and Movies S1-S3.

842 **Figure 7. Simulations of an EPYC1-Rubisco System Reveal an Effect of Binding Site**
843 **Stoichiometry on the Aggregation State.**
844 (**A** – **C**)**,** Snapshots of simulations with 3 (A), 4 (B), and 5 (C) Rubisco binding sites on EPYC1.
845 "Rubiscos" (blue rectangles) and "EPYC1s" (red polymers) bind when they occupy the same
846 sites in a 2D grid. Snapshots are from simulations with 10 k_BT specific bonds and 0.1 k_BT
847 nearest-neighbor non-specific bonds.
848 (**D – F**) Heat maps of the fraction of Rubiscos that are in clusters of >10 Rubiscos connected by
849 EPYC1s with 3 (D), 4 (E), or 5 (F) Rubisco binding sites. The fraction of grid sites occupied by
850 Rubiscos (*x*-axis) is varied from ~10% to ~70%, with an equal fraction of grid sites occupied by
851 EPYC1s. The specific bond energy (*y*-axis) is varied from 0 to 10 k_BT, while the nonspecific
852 bond energy is fixed at 0.1 k_BT. Red dots indicate the parameters used for the snapshots in (A –
853 C).
854 **(G – I),** Snapshots of off-lattice 3D simulations with 3 (G), 4 (H), and 5 (I) Rubisco binding sites
855 on EPYC1 for Rubisco. "Rubiscos" (blue spheres with 4 binding sites on each end) and
856 "EPYC1s" (red polymers) bind when their binding sites overlap. Inset in (I): zoom-in with one
857 Rubisco and one EPYC1 with 5 binding sites; 4 of the 5 binding sites of the EPYC1 are in
858 specific bonds with the Rubisco. Snapshots are from simulations with 10 k_BT specific bonds and
859 a Lennard-Jones nonspecific interaction with $\epsilon = 0.1$ k_BT. The molecules occupy ~2% of the
860 total space in these simulations, with equal total numbers of EPYC1 and Rubisco binding sites.
861 (**J**) Fraction of Rubiscos in clusters of >10 Rubiscos for EPYC1s with different numbers of
862 binding sites, in the 2D model. The specific bond energy is 10 k_BT and the nonspecific bond
863 energy is 0.1 k_BT.

864 (**K**) The concentration of Rubisco at which clustering begins for systems with different numbers
865 of EPYC1 binding sites in the 2D model. The onset is determined from the curves in (J) (see
866 Figure S7).
867 (**L**) Heat map of the distribution of cluster sizes for different numbers of EPYC1 binding sites in
868 the 2D model. Each column depicts the normalized cluster-size distribution at [Rubisco] = 0.15,
869 with 10 $k_B T$ specific bond energy and 0.1 $k_B T$ nonspecific bond energy.
870 (**M**) Schematic of a possible mechanism by which magic numbers could regulate the formation
871 and dissolution of pyrenoids: EPYC1s with <4 binding sites favor Rubisco clustering in the
872 pyrenoid (left), while EPYC1s with <4 binding sites form stable 2:1 complexes of
873 EPYC1:Rubisco that dissolve into the chloroplast stroma (right).
874 See also Figure S7, Table S1, and Movie S6.
875
876
877

878 **Supplemental Figure Titles and Legends**

879

880 **Figure S1. Fluorescently-Tagged Lines Grow Normally, and Specks Can Appear**

881 **Transiently Even During Fission; Related to Figures 1-3 and 5-6.**

882 (**A** – **B**) Spot tests on cells grown in photoautotrophic conditions that require a functional

883 pyrenoid (A) versus heterotrophic conditions that do not require a functional pyrenoid (B) show

884 that tagging pyrenoid matrix proteins (RbcS1, EPYC1, RCA1) with the fluorescent protein

885 Venus results in wild-type-like growth. In comparison, the *epyc1* mutant, which has perturbed

886 pyrenoid structure (Mackinder *et al.*, 2016), displays a growth defect under conditions that

887 require a functional pyrenoid. TP air cells were imaged ~5 days after spotting; TAP dark cells

888 were imaged ~12 days after spotting.

889 (**C** – **D**) Additional example of pyrenoid fission in an EPYC1-Venus cell, with transient puncta

890 appearing in the stroma. Images are summed through Z, with chlorophyll autofluorescence

891 shown in magenta, and RbcS1-Venus in green; t=0 is the first observation of a gap in chlorophyll

892 between the daughter pyrenoids in the first division shown. Dashed curves represent approximate

893 chloroplast outlines in the mother (white) and daughter (pink) cells.

894 **Figure S2. The Mode and Lineage of Pyrenoid Inheritance were Tracked for RbcS1-Venus**
895 **and EPYC1-Venus Daughter Cells; Related to Figures 1-3 and 6.**
896 (**A – B**) Pyrenoid inheritance was observed in daughter cells resulting from the first division
897 (middle columns) or second division (right columns) of 17 original RbcS1-Venus cells, resulting
898 in 86 daughters (A) and 41 original EPYC1-Venus cells, resulting in 178 daughters (B). Colors
899 correspond to the pattern of pyrenoid inheritance in that cell. As in Figures 1 and 2, pyrenoid
900 fissions are displayed in magenta, and inheritance of an entire previously existing pyrenoid is
901 blue. Cases in which no pyrenoid was inherited and puncta appeared are shown in yellow; cases
902 in which neither a pyrenoid nor puncta were inherited are shown in orange. In some lineages, the
903 mother cell contained two pyrenoids at the start of observation, allowing both daughters to be
904 labeled in blue. Some observed cells only divided once. The observed proportions of these
905 inheritance patterns are displayed in Figure 2. Images from selected divisions in this figure are
906 presented elsewhere as follows: Fig 1A-C, Figure 6A-B, and Movie S1 are RbcS1-Venus
907 original cell 7 divisions 1 and 2; Figure 1D is EPYC1-Venus original cell 9 division 1; Figure 1E
908 is RbcS1-Venus cell 9 division 1; Figure 2A and Movie S2 are RbcS1-Venus cell 14 division 2;
909 Figure 2B is RbcS1-Venus cell 9 division 2; Figure 2C is EPYC1-Venus cell 19 division 2;
910 Figure 2D is EPYC1-Venus cell 40 division 1; Figure 2G-I are RbcS1-Venus cell 15 division 2,
911 RbcS1-Venus cell 10 division 2, and EPYC1-Venus cell 8 division 2, respectively; Figure 3A is
912 RbcS1-Venus cell 12 division 1, B is RbcS1-Venus cell 6 division 1, C is RbcS1-Venus cell 5
913 division 1, D is RbcS1-Venus cell 13 division 1, E is RbcS1-Venus cell 3 division 1, and F is
914 EPYC1-Venus cell 19 division 1; Figure S1C-D are EPYC1-Venus cell 29 division 1; Figure
915 S6A is RbcS1-Venus cell 5 divisions 1 and 2; Movie S3 and Figure S6B are RbcS1-Venus cell 9
916 divisions 1 and 2; Figure S6C is RbcS1-Venus cell 14 divisions 1 and 2, D is RbcS1-Venus cell

917 13 divisions 1 and 2, E is RbcS1-Venus cell 12 divisions 1 and 2, and H is EPYC1-Venus cell 5

918 divisions 1 and 2.

919 **Figure S3. *In situ* Localization of Rubisco Holoenzymes Within Tomograms; Related to**
920 **Figure 4.**
921 (**A**) Tomogram volumes were masked to separate the pyrenoid matrix (yellow), which was
922 template matched for Rubisco particles, from the rest of the volume (blue), which was excluded
923 from the search. 2D projection images of the 3D masks are shown. The volumes contained
924 within the matrix masks are underlined in yellow.
925 (**B**) Following template matching, particles were exhaustively extracted and plotted in
926 histograms by their cross-correlation coefficient (CCC) to the Rubisco template. Clear true-
927 positive peaks with higher CCC values than the false-positive noise were apparent for all
928 tomograms. Each true-positive peak was fit with a Gaussian distribution (red dashed line), and a
929 two standard deviation (2s) cutoff was used to separate particles to be processed further (yellow,
930 ~97.5% of true positives) from those to be discarded (blue). The numbers of particles that were
931 kept are underlined in yellow.
932 (**C**) Following 3D subvolume alignment, the particles were subjected to hierarchical
933 classification to remove contaminating false positives. Between 6.5% and 11.8% of particles
934 were removed per tomogram (blue averages), leaving only clean true-positives (yellow
935 averages). The averages were all filtered to 30 Å resolution for clear comparison. The numbers
936 of particles remaining after classification are underlined in yellow.
937 (**D**) Calculation of the concentration of Rubisco particles within each masked pyrenoid matrix
938 region, adjusted for the ~97.5% sampling of true positives in (B). For comparison to lower-
939 resolution light microscopy measurements, concentrations that do not omit the volume of the
940 pyrenoid tubules were also calculated. Concentrations for the HCP simulated data generated
941 within the same masked matrix volumes (Figure S4E) are also listed.

942 **Figure S4. Assessment of the Rubisco *in situ* Subtomogram Average and Generation of the**

943 **Crystalline and Randomly Packed Simulated Pyrenoid Data; Related to Figure 4.**

944 (**A**) Fourier shell correlation (FSC) resolution estimates for the subtomogram average shown in

945 Figure 4C. Resolution was calculated both by cross-resolution of the full dataset to the crystal

946 structure (Taylor *et al.*, 2001) (blue line, 15.5 Å, 0.3 cutoff) and by randomly splitting the dataset

947 and comparing the consistency of the two half-sets (red line, 16.5 Å, 0.143 cutoff).

948 (**B**) The Rubisco crystal structure (Taylor *et al.*, 2001) fitted into the subtomogram average.

949 (**C**) A difference map between the crystal structure and the subtomogram average. There is only

950 a very small significant difference (green volume: 2.5s variance) near the Rubisco small

951 subunits. Half of the subtomogram average has been cut away to more clearly show the position

952 of the difference density relative to the crystal structure.

953 (**D – I**) Experimental Rubisco positions (D), HCP with 13.9 nm spacing between Rubisco centers

954 (E), Rubisco packing within a Rubisco crystal (F), randomly positioned single Rubisco particles

955 (G), randomly positioned linked pairs of Rubisco with 13.9 ± 1.5 nm between pair centers (H),

956 and a randomly positioned network of linked Rubisco (I), each mapped into the same tomogram

957 volume (corresponding slices are shown for each volume). In (G – I), the rotation symbols with

958 "?^{o}" indicate that the particles have random orientations in addition to their random positions.

959 (**J**) Schematic for the generation of the random linked network in (I), illustrated from left to

960 right. A seed Rubisco template is randomly placed into the matrix volume, and a zone for the

961 potential placement of neighbors (yellow) is defined in a shell occupying the volume that is 5.5-7

962 nm from the seed particle surface. A second particle is randomly placed within the neighbor zone

963 (center position of particle indicated by "X"), and the zone for potential neighbor placement is

964 then updated to occupy a shell of 5.5-7 nm around both particles. The rounds of random neighbor

965 placement and zone updating are iteratively repeated to grow the random linked network. After a

966 maximum of 8 particles have been added to a network, a new seed is randomly placed elsewhere

967 in the matrix volume and another network is formed.

968 **Figure S5. The Bleached Region During FRAP Experiments Extends Vertically Through**
969 **the Pyrenoid Volume, and RCA1-Venus and EPYC1-Venus are also Mobile Within the**
970 **Pyrenoid; Related to Figure 5.**
971 (**A** – **B**) Serial sections through a fixed RbcS1-Venus pyrenoid before (A) and after (B) a bleach
972 event. Sections are 0.23 μm apart.
973 (**C**) Average total intensity and volume of the Venus signal after bleaching, as a percent of the
974 initial value, in 3D reconstructions of 10 fixed RbcS1-Venus pyrenoids before and after a bleach
975 event. Error bars are standard deviation.
976 (**D**) Average FRAP recovery profiles ± SEM for 28 pyrenoids in live *epyc1* mutants
977 complemented with EPYC1-Venus. The *y*-axis is intra-pyrenoid homogeneity.
978 (**E – F**) Examples of FRAP in an *epyc1* EPYC1-Venus pyrenoid. Cartoons depict the
979 approximate bleached region (dark gray). Image stills from the recovery time-course (E) and
980 corresponding kymograph (F) as shown in Figure 5.
981

982 **Figure S6. Pyrenoid Matrix Components Undergo Relocalization out of and into the**
983 **Pyrenoid During Cell Division; Related to Figures 1, 2, and 6.**
984 (**A** – **E**) Examples of the changing localization of the RbcS1-Venus signal during divisions in
985 five lineages. As in Figure 6A-B, the plots represent the sum of the fluorescence through the
986 whole Z-stack in each masked region over time. The time window during which the pyrenoid is
987 undergoing fission is highlighted in gray; note that the pyrenoids in the second division of (E) do
988 not undergo fission, and thus lack a gray highlight. t=0 is the minute in which a chlorophyll gap
989 first appears between daughter pyrenoids in the first division. Underneath each plot, snapshots of
990 the RbcS1-Venus signal at several different points are displayed in two ways: on the top, the
991 RbcS1-Venus signal is false-colored based on intensity, with the approximate chloroplast
992 boundaries outlined in white, and signal from outside the chloroplast masked out. Beneath that,
993 an overexposed version of each image is displayed, with the daughter cells colored to match the
994 mode of pyrenoid inheritance in that cell. The second division in (C) is also shown in Figure 2A;
995 the second division in (E) is also shown in Figure 2B.
996 (**F**) Duration and relative timing of the RbcS1-Venus chloroplast (magenta) and pyrenoid (green)
997 division for the pyrenoid fissions, as in Figure 1G, with the duration and relative timing of
998 pyrenoid dissolution superimposed in blue. Left: 1st divisions; right: 2nd divisions. t=0 is the
999 first observation of a gap in chlorophyll between the daughter pyrenoids. We observed only one
1000 instance in which pyrenoid dissolution and fission appeared to be temporally distinct; in this
1001 case, the mother cell contained two pyrenoids (which is rare, but does occur), and one daughter
1002 cell inherited both an entire pyrenoid and a pyrenoid by fission.

1003 (**G**) Average and standard deviation of the durations of chloroplast (magenta) and pyrenoid

1004 (green) fissions, as well as pyrenoid dissolution (blue) for the RbcS1-Venus divisions shown in

1005 (F) (n = 28).

1006 (**H**) Example of the changing localization of Venus signal during a series of divisions in EPYC1-

1007 Venus cells, displayed as in (A – E).

1008 **Figure S7. Moves in Monte Carlo Simulations, Determination of Onset of Clustering, and**
1009 **Snapshots of Simulations with Binding of Each EPYC1 Restricted to one end of a Rubisco;**
1010 **Related to Figure 7.**

1011 (**A** – **I**) Schematics of Rubisco and EPYC1 moves in Monte Carlo simulations.

1012 (**J**) Determination of clustering onsets in Monte Carlo simulations. Data points are from the
1013 simulation data in Figure 7J. Each curve for a fixed number of EPYC1 interacting sites is fitted
1014 with a 4th order polynomial, and the highest zero-crossing is taken as the onset of clustering in
1015 Figure 7H.

1016 (**K**) Fraction of Rubiscos in clusters of >10 Rubiscos for EPYC1s with 3, 4, or 5 binding sites in
1017 the off-lattice 3D simulation. The specific bond energy is 10 $k_B T$ and the Lennard-Jones
1018 nonspecific interaction energy is $\epsilon = 0.1\ k_B T$.

1019 (**L** – **N**) Snapshots of an EPYC1-Rubisco system with 8 binding sites on Rubisco holoenzymes
1020 for EPYC1 with four binding sites, with the constraint that a given EPYC1 cannot bind to both
1021 ends of one Rubisco. Percentage of occupied grid sites is 20% in each layer.

1022 **STAR★METHODS**

1023

1024 **CONTACT FOR REAGENT AND RESOURCE SHARING**

1025

1026 Please contact the Lead Contact, Martin C. Jonikas (mjonikas@princeton.edu), with any
1027 requests regarding reagents used in this study.

1028

1029 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

1030

1031 **RbcS1-, RCA1-, epyc1 EPYC1-, and EPYC1-Venus *Chlamydomonas reinhardtii* strain**
1032 **generation and culture conditions.**

1033 All strains expressing fluorescent proteins were generated in Mackinder, *et al.* (2016), as
1034 follows: the DNA encoding the protein of interest (RbcS1, RCA1, or EPYC1) was amplified
1035 from *Chlamydomonas reinhardtii* gDNA and cloned behind the PsaD promoter with a Venus-
1036 3xFLAG on the C-terminus, in a construct containing the *aphVIII* gene for paromomycin
1037 resistance. Vector sequences have been deposited at GenBank under accession numbers
1038 KY550376 (pLM005-RBCS1-Venus), KX077944 (pLM005-EPYC1-Venus), and KY550375
1039 (pLM005-RCA1-Venus). Linearized constructs were transformed into wild-type (cMJ030; also
1040 known as CC-4533 cw15; mating type minus) or *epyc1* (also in the cMJ030 background)
1041 (Mackinder *et al.*, 2016) *Chlamydomonas reinhardtii* by electroporation, which results in random
1042 integration into the nuclear genome (Zhang *et al.*, 2014; Mackinder *et al.*, 2016). Strains
1043 expressing the construct were selected by growth on agar plates containing Tris-Acetate-
1044 Phosphate (TAP) plus paromomycin, and subsequent fluorescence screening for Venus on a

1045 Typhoon Trio fluorescence scanner (GE Healthcare) with excitation at 532 nm and emission at
1046 555/20 nm. Strains containing RbcS1- and EPYC1-Venus were authenticated by Western Blots
1047 in Mackinder *et al.* (2016). The above strains have been deposited at the Chlamydomonas
1048 Resource Center under accession numbers CC-5357 (RbcS1-Venus), CC-5359 (EPYC1-Venus),
1049 CC-5358 (RCA1-Venus), CC-5360 (*epyc1*), and CC-5361 (*epyc1* EPYC1-Venus).
1050 For microscopy, photoheterotrophic pregrowth cultures were inoculated from ~two-
1051 week-old TAP plates into 50 mL liquid TAP media in 250 mL glass Erlenmeyer flasks, and
1052 grown under continuous white light (~40 μmol photons m^{-2} s^{-1}) while shaking at 125 rpm until
1053 cells reached log phase growth (2-5 x 10^6 cells/mL; 3-4 days after inoculation). Photoautotrophic
1054 cultures were inoculated by centrifuging the pregrowth cultures at 1000 g for 5 minutes at room
1055 temperature and then resuspending the pellet in 50 mL liquid minimal Tris-Phosphate (TP)
1056 media (Kropat *et al.*, 2012). Photoautotrophic cultures were grown in 50 mL liquid TP in glass
1057 tubes bubbled with air (~0.04% CO_2) under continuous light (150 μmol photons m^{-2} s^{-1} red and
1058 blue LumiGrow LumiBar LED lights). After several rounds of dilution and growth in TP, cells
1059 were imaged when they reached a density of ~1-2x10^6 cells/mL.
1060
1061 **Vitrification, cryo-FIB, cryo-ET, and tomogram reconstruction**
1062 *Chlamydomonas mat3-4* cells (Umen and Goodenough, 2001), acquired from the
1063 Chlamydomonas Resource Center (Univ. of Minnesota, CC-3994, mt+), were used because they
1064 vitrify better than wild-type cells due to their smaller size (~5 μm) but have normal pyrenoid
1065 morphology and a functional carbon concentrating mechanism, shown by growth in low-CO_2
1066 conditions (Umen and Goodenough, 2001), which inhibit carbon concentrating-deficient strains
1067 such as the *epyc1* mutant (Mackinder *et al.*, 2016). Cells were grown in TAP media with

1068 constant light (~90 µmol photons $m^{-2} s^{-1}$) and aeration with normal atmosphere until mid-log

1069 phase.

1070

1071 **METHOD DETAILS**

1072

1073 **Division microscopy & analysis**

1074 *Image acquisition*

1075 Strains were grown and prepared for microscopy as described above. Cells were imaged

1076 at room temperature every minute for 8-16 hours on a Leica TCS SP8 laser scanning confocal

1077 microscope in resonant scanning mode with LASX software and a 63X NA 1.4 oil objective,

1078 while illuminated with a 626 nm red LED circle affixed to the condenser (LED Angel Eye

1079 Headlight Accent Light Kit; SuperBrightLeds.com, AE80-RGB12). Venus fluorescence was

1080 imaged by exciting with a white light laser at 488 nm and collecting emission from 499-551 nm

1081 on a HyD SMD hybrid detector (Leica) with lifetime gate filter (1-6 ns) to reduce background

1082 due to chlorophyll autofluorescence. Chlorophyll autofluorescence was imaged simultaneously

1083 with the same 488 nm excitation, and emission was collected from 680-750 nm on a PMT

1084 (Leica). Images were collected with 8-line averaging, 0.3 µm steps through the entire cell

1085 volume, and adaptive focus control. Displayed division analysis data represent images collected

1086 over at least two independent biological replicates for each strain. Sample sizes were chosen

1087 based on similar published studies (Conduit *et al.*, 2010; Chen *et al.*, 2013; Weber and

1088 Brangwynne, 2015; Eichel *et al.*, 2016; Feric *et al.*, 2016).

1089 *Image analysis*

1090 4D TIFFs were exported to FIJI as virtual stacks and dividing cells were manually

1091 marked in MaxZ projections over time. The pyrenoid inheritance pattern and timing of division

1092 was manually noted for each dividing cell in 3D in Imaris (Bitplane). If a cell exhibited more
1093 than one mode of inheritance (for instance, both pyrenoid fission and puncta), then fission took
1094 precedence for categorization. The duration of chloroplast division and pyrenoid fission (as
1095 defined in the text) were measured for all observed RbcS1-Venus pyrenoid fission events (n = 14
1096 first divisions and 14 second divisions) and randomly-selected EPYC1-Venus fission events (n =
1097 22 first divisions and 25 second divisions). The extent of pyrenoid dissolution and increase in
1098 stromal Venus signal was measured in FIJI as follows: For each dividing cell, a SumZ projection
1099 was created through the whole cell. Then, any saturated pixels were masked out, and intensity-
1100 based masks were created from SumZ projections of the chlorophyll and Venus channels to mask
1101 out any signal in the Venus channel from outside the chloroplast or outside the pyrenoid,
1102 respectively. Finally, for each masked region of each cell, the total intensity (sum of the
1103 unmasked pixels in the entire image; RawIntDen) was measured at each time point and exported
1104 to Excel. The Wilcoxon Matched-Pairs Signed-Ranks Test was used to quantify the difference in
1105 the average pyrenoid signal between maximal dimming and 15 minutes beforehand in 31
1106 divisions (with the error bars denoting SEM) as described below.
1107
1108 **Vitrification, cryo-FIB, cryo-ET, and tomogram reconstruction**
1109 Using a Vitrobot Mark 4 (FEI), cells were blotted onto carbon-coated EM grids
1110 (Quantifoil Micro Tools; 4 µL cell culture per grid) and plunge-frozen into a liquid
1111 ethane/propane mixture. Cryo-FIB milling was performed on a dual-beam Quanta3D FIB/SEM
1112 microscope (FEI) by scanning the cells with Gallium ions, as previously described (Schaffer *et*
1113 *al.*, 2015, 2017). Thinned samples were transferred to a Titan Krios 300 kV transmission electron
1114 microscope (FEI) equipped with a 968 Quantum post-column energy filter (Gatan), and imaged

1115 with a K2 Summit direct detector (Gatan) operated in movie mode at 12-17 frames per second.
1116 Using SerialEM software (Mastronarde, 2005), single-axis tilt-series were recorded from -60° to
1117 +60° (in two halves separated at 0°) at 2° increments, with an object pixel size of 3.42 Å, a target
1118 defocus of -4 or -5 mm, and a cumulative electron dose of ~100 electrons/$Å^2$. Image frames were
1119 aligned using in-house developed software to correct for beam-induced motion. Using IMOD
1120 software (Kremer *et al.*, 1996), tilt-series were aligned by patch-tracking and reconstructed by
1121 weighted back projection. The biggest gain in image quality compared to the previous study
1122 (Engel *et al.*, 2015) was the result of using a direct electron detector instead of a CCD camera.
1123
1124 **Localization of Rubisco within tomograms and subtomogram averaging**
1125 *Tomogram masking and template matching*
1126 Tomograms were binned to a pixel size of 13.68 Å and subjected to template matching
1127 using the PyTom software (Hrabe *et al.*, 2012). The template was generated from the X-ray
1128 crystallography structure of *Chlamydomonas* Rubisco (Taylor *et al.*, 2001) (PDB entry: 1GK8)
1129 by lowpass filtering the electron density map to a resolution of 33 Å, a value determined by the
1130 approximate first zero of the contrast transfer function in the tomograms (Förster *et al.*, 2010).
1131 Tomogram masks were manually segmented in Amira software (FEI), restricting the search area
1132 to only the pyrenoid matrix (Figure S3A). For each tomogram, the template matching cross-
1133 correlation function was filtered by the tomogram mask, and the remaining cross-correlation
1134 peaks were exhaustively extracted with a mask radius of 9.5 nm. Histograms of cross-correlation
1135 scores from the extracted peaks revealed Gaussian distributions of true-positive particles at the
1136 high values. A cutoff was set to two standard deviations towards the low-valued tail of each
1137 Gaussian, and all particles with scores below this cutoff were discarded (Figure S3B).

1138 *Subtomogram averaging and classification*

1139 Subvolumes corresponding to the extracted peaks were binned to a pixel size of 6.84 Å
1140 and aligned in PyTom using a real-space expectation maximization method that implements
1141 gold-standard alignment (Hrabe *et al.*, 2012). This alignment procedure did not impose D_4
1142 symmetry. The number of iterations was set to 10, the initial angular increment to 3°, and the
1143 angular shells to 3. Next, subtomogram classification was performed using constrained principal
1144 component analysis (CPCA) (Förster *et al.*, 2008). PyTom was used to calculate similarity
1145 matrices and for hierarchical clustering, while CPCA and k-means clustering were performed in
1146 Matlab (MathWorks). Subtomograms were lowpass filtered to 38 Å, 5 eigenvectors were used,
1147 and the number of classes was set between 99 and 165, depending on the number of particles in
1148 each tomogram. Classes were then merged by hierarchical clustering, using constrained cross-
1149 correlation as distance measure, yielding a "positive" and "negative" class for each tomogram
1150 (Figure S3C).

1151 *High-resolution subtomogram averaging*

1152 Unbinned subtomogram volumes (3.42 Å pixel size) from "tomogram 1" corresponding
1153 to the 30,000 highest template matching cross-correlation scores in the "positive" class were
1154 used to generate the average shown in Figure 4C. The subvolumes were corrected for the
1155 contrast transfer function by phase-flipping in IMOD, then aligned with imposed D_4 symmetry
1156 using gold-standard alignment in RELION (Bharat and Scheres, 2016). This procedure was
1157 restricted to a local angular search, using initial angle and offset sampling rates of 3.7° and 0.34
1158 nm, respectively, while the maximal offset value was set to 1.7 nm. The initial reference was a
1159 subtomogram average filtered to 4 nm resolution. In the final average, the B-factor variable was

1160 set to 4 Å^2 for each 1 electron/Å^2, and the cumulative electron dose of the subtomograms was

1161 restricted by only using the central 60° of the tilt-series (-30° to +30°).

1162

1163 **Local density analysis and generation of simulated data**

1164 *Measurement of radial local Rubisco density*

1165 Pyrenoid matrix organization was quantified by Radial Local Density (RLD) estimation,

1166 defined as:

$$\lambda(r) = \frac{\sum_{i=0}^{n} C\left(S_h(\vec{x}_{p_i}, r)\right)}{V\left(S_h(\vec{x}_{p_i}, r)\right)}$$

1167 The coordinates of the reference particle are represented by the vector $\vec{x}_{p_i}$. Function C counts the

1168 number of particles contained within the S_h subvolume, which is a radial shell subset of the S_P

1169 total masked matrix volume ($S_h \subset S_P$). Function V computes the volume of S_h. Since all

1170 particles in S_P are used, then $\{\vec{x}_{p_0}, \ldots, \vec{x}_{p_n}\} \in S_P$. The definition of the local subvolume S_h

1171 associated to a particle $\vec{x}_{p_i}$ and a radius r is:

$$S_h(\vec{x}_{p_i}, r) = \left\{\vec{x} \in S_P | \{d(\vec{x}, \vec{x}_{p_i}) \geq (r - \Delta r/2)\} \wedge \{d(\vec{x}, \vec{x}_{p_i}) \leq (r + \Delta r/2)\}\right\}$$

1172 where d is the Euclidean distance function. In an unbounded space, S_h would correspond with a

1173 spherical shell centered at $\vec{x}_{p_i}$, with radius r (distance to reference particle p_i) and a shell

1174 thickness of $\Delta r > 0$. However, because the pyrenoid matrix volume S_P is actually a finite

1175 irregularly bounded space, $V(S_h)$ cannot be accurately estimated by analytical formulas. Thus,

1176 we used a numerical estimator analogous to Wiegand and Moloney (2004), but adapted for 3D

1177 volumes instead of 2D areas. This approach restricts S_h to the masked pyrenoid matrix volume,

1178 eliminating edge effects that would have arisen from the inclusion of the membrane tubules and

1179 areas outside of the pyrenoid. For all RLD measurements in Figure 4 (experimental and

1180 simulated data in panels D, F and H, experimental data in panel G), we set Δr = 1.4 nm as a

1181 compromise between precision and graph smoothness.

1182 As the global pyrenoid matrix Rubisco density, ρ, is slightly different for each pyrenoid

1183 (Figure 4D, Figure S3D), we normalized each RLD by the global density so that RLD shape

1184 could be directly compared between multiple tomograms: $\lambda' = \lambda/\rho$.

1185 *Generation of simulated tomogram volumes*

1186 For direct comparison to the experimental data, all of the simulated tomograms were

1187 created by placing Rubisco particles into the same pyrenoid matrix volumes as those in the real

1188 tomograms (Figures S3A, S4D-I), as defined by the manual masking step described above.

1189 *Crystal*

1190 This data was generated by propagating the unit cell of crystalized *Chlamydomonas*

1191 Rubisco (Taylor *et al.*, 2001) throughout the matrix volumes. Despite their "noisy" appearance,

1192 the crystalline profiles have less variance than the experimental data – their 99% confidence

1193 intervals are almost too small to see in Fig. 4F. Two factors cause the simulated data to look

1194 "noisy": 1) Unlike the heterogeneous organization of the experimental data, the crystalline

1195 packing is exact, leading to dramatic peaks. 2) RLD is sampled with a spherical shell (of 1.4 nm

1196 width), whereas the crystalline organization does not propagate in a spherical pattern. This

1197 mismatch between the spherical RLD measurement and the crystalline packing causes the

1198 "noisy" jagged appearance of the plots.

1199 *Hexagonal close packing*

1200 Within the matrix volumes, an HCP lattice was generated with 13.9 nm between the

1201 centers of all nearest neighbors, a distance based on the preferred nearest neighbor spacing

1202 measured in Figure 4E.

1203 *Random singles*

1204 A binary Rubisco template was generated by thresholding our Rubisco subtomogram
1205 average (Figure 4C) to allow the minimum near neighbor distance found in our experimental
1206 data (Figure 4E). Using Monte Carlo simulations, these templates were sequentially mapped into
1207 the matrix volumes, using random positions and orientations but forbidding overlap between
1208 templates. The procedure was stopped upon reaching the same number of Rubisco particles as
1209 the experimental data.

1210 *Random linked pairs*

1211 Pairs of two Rubisco templates, each with random orientation, were placed together with
1212 an inter-center distance defined by the experimental data's distribution (mean = 13.9 nm,
1213 standard deviation = 1.5 nm; Figure 4E). Pairs were sequentially mapped into the matrix
1214 volumes, using random positions and orientations, while avoiding overlap with previously
1215 inserted particles. These steps were repeated until the same number of particles as the
1216 experimental data was reached.

1217 *Random linked network*

1218 First, a seed Rubisco template was randomly placed into the matrix volume, while
1219 avoiding overlap with already inserted particles. Second, a zone for the potential placement of
1220 neighbors was defined as a shell around the seed template in a range of 5.5-7 nm from the seed
1221 particle surface (Figure S4J). Third, a neighbor Rubisco particle was randomly placed with its
1222 center inside the zone. Fourth, the zone for potential neighbor placement was updated to
1223 encompass a range of 5.5-7 nm around both particles. The rounds of random neighbor placement
1224 and zone updating were sequentially repeated until a maximum of 8 networked particles were
1225 placed. After 8 particles were placed or the network failed to place a particle due to lack of

1226 space, a new seed was randomly created within the matrix volume and the procedure was
1227 repeated. When two networks encountered each other, they merged their neighbor placement
1228 zones. The procedure was stopped when the same number of Rubisco particles as the
1229 experimental data was reached.

1230 *Fitting Lennard-Jones fluid models to the experimental data*

1231 RLD is a robust estimator of the Radial Distribution Function (RDF) for proteins in a
1232 finite, bounded, and irregularly shaped space like the pyrenoid matrix. Thus, analytical models
1233 for RDF can also be applied to RLD results. We fit our experimental data with the Lennard-
1234 Jones fluid analytical model proposed in Morsali *et al.* (2005). This RDF model is constrained by
1235 65 constants that Morsali *et al.* (2005) calculated from 353 molecular dynamics simulations of
1236 argon atoms interacting via the Lennard-Jones potential under a range of state variables. We set
1237 the Lennard-Jones length parameter (σ = 13.9 nm) to scale the x-axis to reduced r^*, and
1238 optimized the variables for reduced temperature (T^* = 17.296) and reduced density (ρ^* = 1.143)
1239 by minimizing the mean squared error between the model and our data with the Python SciPy
1240 library's global optimization algorithm described in Wales and Doye (1997). With these values
1241 for σ , T^* and ρ^*, the Morsali RDF model fit our data well (root mean squared error: 0.0935,
1242 mean squared error: 0.0087, maximum squared error: 0.5308, standard deviation of squared
1243 error: 0.05819). Our optimized value for T^* fell outside the range of molecular dynamics data
1244 that Morsali *et al.* (2005) used to develop their model; therefore, we also determined the
1245 Lennard-Jones RDF for these conditions using a molecular dynamics simulation (Plimpton,
1246 1995).

1247 The simulation was performed using the LAMMPS Molecular Dynamics Simulator

1248 (Plimpton, 1995). The simulation space was a 10x10x10 box with hard-wall boundary conditions

1249 and $N = 1000$ particles. We employed a Lennard-Jones interaction between particles, defined by:

1250 $$V_{LJ} = \epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right].$$

1251 We set the interaction strength at $\epsilon = 1$, and the Lennard-Jones length scale σ was chosen to

1252 reproduce the reduced density via $\rho^* = N\sigma^3/V$, where V is the system volume. The reduced

1253 density ($\rho^* = 1.143$) and reduced temperature ($T^* = 17.296$) were set to the values determined by

1254 the analytical fit of the experimental RLD to the Lennard-Jones fluid RDF formula from Morsali

1255 *et al.* (2005) as described above. To avoid boundary artifacts, the RDF was computed from the

1256 particle-to-particle distances around a set of reference particles occupying a small 2x2x2 box at

1257 the center of the full simulation volume. To compare the simulated RDF with the experimental

1258 RLD curve, we set the vertical scale by matching the far distance amplitude, and then set the

1259 horizontal scale by a least-squares fit to the experimental data.

1260 There is an important geometrical difference between the Lennard-Jones RDF models

1261 and our experimental measurements of the pyrenoid's RLD. Lennard-Jones models consider

1262 spherical particles [such as argon (Morsali *et al.*, 2005)] that always have one minimum-energy

1263 distance between particle centers (r_m). However, Rubisco particles are not spherical, and have a

1264 minimum diameter of ~10 nm and a maximum diameter of ~13 nm. Thus, unlike spherical

1265 particles, a single mimimum-energy distance between the surfaces of neighboring Rubiscos (as

1266 shown in our linked-network model, Figure S4J) yields a range of minimum-energy distances

1267 between particle centers (a distribution of r_m values) instead of a single, discrete distance. This

1268 may explain why our experimental data had a broader first peak than that observed in Lennard-

1269 Jones models.

1270 The Lennard-Jones potential describes the attractive and repulsive forces between small
1271 molecules (our models used argon), the balance of which results in a single preferred distance
1272 between neighbors. Rubisco particles exist on a much larger size scale than argon atoms, and
1273 thus are subject to different molecular forces such as protein interactions. However, we reasoned
1274 that this simple Lennard-Jones fluid might nonetheless serve as an informative analogy for liquid
1275 organization. Interestingly, the Lennard-Jones distribution appears to be robust to changes in
1276 scale, matching well not only to ~12 nm pyrenoid Rubisco and 120 nm colloidal particles (Gu *et*
1277 *al.*, 2010), but also to 4 μm colloidal particles (Thorneywork *et al.*, 2014). Due to differences in
1278 the underlying molecular forces, the analogy between the pyrenoid and a Lennard-Jones fluid is
1279 limited to the interpretation that the pyrenoid matrix may be liquid; the analogy is not intended
1280 for the application of other Lennard-Jones descriptors to properties of the pyrenoid.
1281
1282 **Spot tests**
1283 Pre-growth cultures were inoculated in TAP as described above. After 4 days, 50 mL of
1284 each culture were washed of acetate twice by centrifuging for 5 minutes at 2000 g and
1285 resuspending the pellet in 20 mL TP. After washing, cells were counted three times (Countess II
1286 Automated Cell Counter; ThermoFisher Scientific) and diluted serially to contain 10^4, 10^3, 10^2,
1287 or 10 cells per 15 μL. 15 uL of each strain in each concentration were spotted in replicate onto
1288 TAP and TP plates and left to dry in the dark for one hour. TAP plates were then wrapped in
1289 Parafilm and aluminum foil and kept at room temperature in the dark for 12 days before imaging;
1290 TP plates were incubated in custom containers with filtered air flow (~0.04% CO_2) with 40 hours
1291 of acclimation at 50 μmol photons m^{-2} s^{-1} red and blue LED light, and then 3 days in ~100 μmol

1292 photons m^{-2} s^{-1} before imaging. Representative samples out of 4 total replicates (2 biological
1293 replicates, each with two technical replicates) per condition are shown.
1294
1295 **Fluorescence Recovery After Photobleaching microscopy & analysis**
1296 *Slide preparation*
1297 For live cell imaging, 300 µL of photoautotrophic culture at ~2x10^6 cell/mL were plated
1298 onto poly-L-lysine-coated 8 well µ-slides (ibidi, 80824) and allowed to settle for ~5 min. The
1299 supernatant was removed, and the remaining cells were coated with 300 µL of TP in 1.5%
1300 UltraPure Low Melting Point Agarose (Invitrogen) at ~40°C and allowed to cool for ~20
1301 minutes. For fixed cell imaging, cells were fixed in 10% glutaraldehyde as follows: ~70 x 10^6
1302 cells were pelleted by centrifugation for 10 minutes at 1000 g at room temperature, resuspended
1303 in 6 mL of 10 mM HEPES-KOH (pH 7) at 4°C, and transferred to a cold 20 mL glass beaker. 1.5
1304 mL of 10% glutaraldehyde solution (300 uL of 50% glutaraldehyde in 1.2 mL of 10 mM
1305 HEPES-KOH, pH 7) was added in successive 100 uL drops every five seconds while swirling
1306 vigorously. The mixture was incubated on ice and agitated every 10 minutes for one hour, after
1307 which the fixed cells were centrifuged (1000 g, 5 minutes, 4°C), resuspended in 6 mL cold 10
1308 mM HEPES-KOH (pH 7), centrifuged again, and resuspended in 1 mL of 10 mM HEPES-KOH
1309 (pH 7). 300 µL were plated for microscopy, as described above.
1310 *FRAP image acquisition*
1311 Pyrenoids were imaged at mid-plane at room temperature on a spinning disk confocal
1312 microscope (Leica DMI6000B custom-adapted with a Yokogawa CSU-X1 spinning disk head; a
1313 Photometrics Evolve 512 camera; and Intelligent Imaging Innovations SlideBook software,
1314 Vector FRAP, LaserStack, and mSAC spherical aberration systems), with a 100X oil objective

1315 (HCX Pl APO, 1.4-0.7 NA; Leica). Venus fluorescence was imaged by excitation at 514 nm and
1316 emission with a YFP 540/15 filter (Semrock) and 445/514/561 nm Yokogawa dichroic
1317 beamsplitter (Semrock) under the following conditions: 100 ms exposure every 3 seconds for
1318 100 images, with <2 mW laser power (measured at the fiber), and a camera gain of 3 and EM
1319 gain of 300, with Adaptive Focus Control active before every acquisition. FRAP experiments
1320 were conducted one by one on individual pyrenoids centered in the field of view and far from
1321 previously bleached pyrenoids on the slide. Vector was used to direct the 514 nm laser at full
1322 power for photobleaching (~18 mW, measured at the Vector fiber), which took place between
1323 the third and fourth image captures. on a sub-resolution region of interest positioned at the edge
1324 of each selected pyrenoid; this bleached ~⅓-½ of the cross-section of the pyrenoid. For the
1325 comparative recoveries of RbcS1-, RCA1-, and EPYC1-Venus graphed in Figure 5, the
1326 photobleaching event consisted of one repetition of a 4 ms exposure on a 2x2 pixel region of
1327 interest. Due to alterations in the light path of the microscope, subsequent acquisition and bleach
1328 conditions were altered to achieve the same cross-sectional proportion of bleaching: For the
1329 images of live and fixed RbcS1-Venus FRAP pyrenoids shown in Figure 5, the photobleaching
1330 event consisted of 2 repetitions of a 10 ms exposure directed to a 4x4 pixel region of interest; for
1331 *epyc1* EPYC1-Venus FRAP, 1 repetition of a 10 ms exposure on a 2x2 pixel region of interest
1332 was used. Displayed results are from independent experiments conducted at least three (live
1333 RbcS1-, RCA1-, and EPYC1-Venus), two (fixed RbcS1-Venus), or one (*epyc1* EPYC1-Venus)
1334 times. Sample sizes were chosen based on similar published studies (Conduit *et al.*, 2015;
1335 Klammt *et al.*, 2015; Eichel *et al.*, 2016).
1336 *FRAP quantitative image analysis*

1337 FRAP images were analyzed in FIJI software (Schindelin *et al.*, 2012). The StackReg
1338 translation registration plug-in (Thévenaz and Ruttimann, 1998) was used to align image sets that
1339 shifted in XY during imaging. For each bleached pyrenoid that was analyzed, a kymograph was
1340 generated using the "Reslice" function on a 3-pixel-wide, 3.5 μm-long rectangle that spanned the
1341 bleached and unbleached regions (shown in Figure 5), avoiding any saturated pixels. To plot the
1342 fluorescence recoveries, the signal along 3-pixel-wide rectangles over the bleached and
1343 unbleached regions of each kymograph were measured, respectively; these signals were exported
1344 to Excel and compared to calculate fluorescence recovery, shown as "intra-pyrenoid
1345 homogeneity" over time for each bleached pyrenoid, as follows: for each time point, the signal
1346 from the bleached region of the kymograph was divided by that from the unbleached region.
1347 The signal ratio at the fourth time point ($t = 0$; immediately post-bleach) was defined as y
1348 $= 0$ for each pyrenoid by subtraction. Each recovery plot was then normalized to the average of
1349 the ratios of the three pre-bleach time points, which was defined as $y = 1$. This normalized
1350 recovery was averaged over the stated number of pyrenoids and displayed with the standard error
1351 of the mean for each strain.
1352 Thus, it is important to note that our FRAP curves represent homogeneity over the
1353 bleached and unbleached regions, correcting for signal loss due to repeated measurements. In our
1354 plots, therefore, a return to $y = 1$ is not a return to the initial signal intensity, but a return to the
1355 initial signal homogeneity. This is different from how FRAP recovery is often plotted, and
1356 results in seemingly "higher" recoveries than, for example, those in Shin *et al.* (2016). However,
1357 our FRAP images and kymographs show that the final signal intensities are less than the initial
1358 intensities, but that the signal disparity between the bleached and unbleached regions relaxes,

1359 and our FRAP kymographs resemble those of other liquid-like organelles (Kroschwald *et al.*,

1360 2015; Patel *et al.*, 2015).

1361 Additionally, plotting "homogeneity" allows us to control for potential variations in the

1362 proportion of the pyrenoid that was bleached in each event. Because the fluorescence recovery in

1363 the pyrenoid cannot be said to draw from an infinite pool, variation in the ratio of the bleached to

1364 unbleached pools could produce artifacts in the observed final recovery.

1365 To calculate the fraction of the total pyrenoid volume that was bleached, Z-stacks of

1366 glutaraldehyde-fixed RbcS1-Venus pyrenoids were captured before and after a bleaching event;

1367 in live cells, fluorescence recovered too quickly to section through the entire pyrenoid. 23 planes

1368 were imaged with a 0.23-μm step size (spanning 5.06 μm) before and after bleaching, and pre-

1369 and post-bleach 3D volumes were reconstructed for 10 pyrenoids in Imaris (Bitplane), using the

1370 same intensity thresholds for pre- and post-bleach volumes. The total fluorescence intensity and

1371 volume of each reconstructed pre- and post-bleach pyrenoid was exported and analyzed in Excel.

1372

1373 **"Magic number" modeling**

1374 *Model parameters*

1375 Simulations were performed using a square grid system of 50x50 sites with periodic

1376 boundary condition. In the model, each Rubisco is represented as a 4 by 2 rectangle, and each

1377 EPYC1 occupies several connected (nearest neighbor) grid sites. Rubisco and EPYC1 are self-

1378 avoiding. However, each grid site can be simultaneously occupied by EPYC1 and Rubisco, and

1379 if so, the two are considered to form a specific bond. This scheme allows for stoichiometric

1380 bonding between EPYC1s and Rubiscos without the artifacts due to crowding that would occur if

1381 the two were prevented from occupying the same sites. The simulations also include a weak non-

1382 specific attractive interaction between all occupied nearest-neighbor sites, EPYC1-EPYC1,

1383 EPYC1-Rubisco, and Rubisco-Rubisco.

1384 We performed Markov-Chain Monte Carlo simulations using the Metropolis algorithm

1385 (Metropolis *et al.*, 1953). Briefly, in each simulation step we randomly propose a move of the

1386 EPYC1-Rubisco configuration. The move is always accepted if it reduces system energy, and

1387 accepted with probability $e^{-(E_f-E_i)/k_BT}$, where E_f and E_i are the final and initial energies, if the

1388 move increases system energy. Three categories of moves are proposed: single-EPYC1 moves,

1389 single-Rubisco moves and EPYC1-Rubisco joint moves (Figure S7). Single-EPYC1 moves are

1390 standard lattice-polymer local moves: the end-point move, the corner move, and the reptation

1391 move. Single-Rubisco moves consist of one-step translations in the four cardinal directions and a

1392 90-degree rotation around the Rubisco's center. In the regime of strong specific bonds, EPYC1s

1393 and Rubiscos are typically held together by multiple specific bonds, which leads to dynamical

1394 freezing. To enable the system to better explore configuration space, we include EPYC1-Rubisco

1395 joint moves such that connected clusters of EPYC1s and Rubiscos move together, without

1396 breaking any specific bonds. The joint moves consist of translating a connected cluster of

1397 EPYC1s and Rubiscos together or rotating the whole cluster by 90-degrees around any point. To

1398 obtain thermalized ensembles, we follow a simulated two-step procedure: we keep k_BT constant

1399 and gradually increase bond strength. We first increase the non-specific bond from 0 to 0.1 k_BT

1400 in 0.005 k_BT increments, keeping the specific bond energy at 0 k_BT. Then the specific bond

1401 energy is increased from 0 to 11 k_BT in 0.04 k_BT increments, while the non-specific bond energy

1402 is kept at 0.1 k_BT. Each step of annealing is simulated with 50,000-150,000 Monte-Carlo steps

1403 and results are averaged over 20-100 of the resulting thermalized snapshots.

1404 *Analysis of clustering in the 2D simulations*

1405 To assess the extent of clustering of Rubiscos, we consider a cluster to be a group of
1406 Rubiscos that are connected by EPYC1s via specific bonds. To quantify the fraction of Rubiscos
1407 in large clusters, we employ a cluster-size cutoff of 10+ Rubiscos (Figure 7D-F). This avoids
1408 ambiguities due to smaller clusters that form independent units with all specific bonds satisfied
1409 (*e.g.*, 3 Rubiscos with 8 EPYC1s each of 3 binding sites). To determine the onset of clustering,
1410 we fit the fraction of Rubiscos in large clusters with a degree 4 spline (Figure S7); for the case of
1411 EPYC1s with 8 binding sites, because of the delayed onset of clustering we only use the data for
1412 concentrations > 40% for the fit.

1413 *Three-dimensional off-lattice model*

1414 Molecular dynamics simulations were performed using the LAMMPS Molecular
1415 Dynamics Simulator (Plimton *et al.*, 1995). We simulate a cubic box of 120 nm in each
1416 dimension, with periodic boundary conditions. In the simulation, particles representing Rubisco
1417 and EPYC1 interact both non-specifically and via specific EPYC1-Rubisco bonds. Particles
1418 denoted by A and B interact non-specifically with each other through the Lennard-Jones
1419 potential with a cutoff

$$V(r) = 4\epsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r}\right)^{12} - \left(\frac{\sigma_{AB}}{r}\right)^{6}\right] - 4\epsilon\left[\left(\frac{\sigma_{AB}}{r_c}\right)^{12} - \left(\frac{\sigma_{AB}}{r_c}\right)^{6}\right], r < r_c$$

1420 where $\sigma_{AB} = \sigma_A + \sigma_B$ is the sum of the effective radii of the two particles, ϵ_{AB} is the interaction
1421 strength between the two particles, and r_c is the cutoff of the interaction range.

1422 Each Rubisco in the simulation is represented by a sphere of radius $R_R = 5$ nm, which is
1423 the radius of gyration of Rubisco (Keown, *et al.,* 2013). Since the specific Rubisco binding sites
1424 for EPYC1 are not yet known, we employ a simple model in which each Rubisco has 4 binding
1425 sites, forming a rigid square of edge length 3.4 nm, on each of its two ends. The radius of each
1426 binding site is 0.9 nm and the center of the binding site is 0.9 nm away from the Rubisco sphere

1427 (i.e., 5.9 nm from the center of the Rubisco sphere). Each EPYC1 is represented by 3, 4, or 5
1428 connected binding sites: the binding sites are spheres of radius $R_E = 0.9$ nm, which is the radius
1429 of a compact region of 18 amino acids representing the repeat region of EPYC1. We model the
1430 unstructured chain of 34 amino acids separating these repeats as harmonic springs with zero rest
1431 length and stiffness 0.24 k_BT/nm^2, reflecting the entropic elasticity of a worm-like polymer chain
1432 consisting of 34 units of size 0.35 nm (the approximate size of an amino acid) with a persistence
1433 length of 0.5 nm (a rough consensus for polypeptide chains [Hofmann *et al.*, 2012, Cheng *et al.*,
1434 2010]).

1435 We set the non-specific Lennard-Jones interactions between two EPYC1 binding sites,
1436 between two Rubisco cores, and between one EPYC1 binding site and one Rubisco core to be
1437 the Lennard-Jones potential with a cutoff with interaction energy $\epsilon = 0.1$ k_BT and σ the sum of
1438 the radii of the two interacting particles. We set $r_c = 1.4\sigma$ as a cutoff so that the system does not
1439 include long-range interactions.

1440 In addition to the non-specific interaction, we model the attractive specific interaction
1441 between Rubisco and EPYC1 binding sites as a soft potential

$$V(r) = \epsilon_b\left(1 + \cos\left(\frac{\pi r}{r_c}\right)\right), r < r_c \; \#(1)$$

1442 with binding energy $\epsilon_b = -10$ k_BT, r the distance between binding site centers, and a cutoff
1443 distance $r_c = 0.45$nm. The minimum of energy for this interaction corresponds to fully
1444 overlapping binding sites. The one-to-one character of EPYC1-Rubisco specific bonds is
1445 guaranteed by the Lennard-Jones repulsion between EPYC1 binding sites, which prevents them
1446 from binding to the same Rubisco binding site, and by an additional soft repulsion (with the
1447 same functional form as Equation (1) and $\epsilon = +50$ k_BT) between two Rubisco binding sites,
1448 which prevents them from binding to the same EPYC1 binding site. There is no Lennard-Jones

1449 interaction between a Rubisco binding site and an EPYC1 binding site, in order to allow them to

1450 overlap and form a specific bond.

1451 We performed molecular dynamics simulations with Langevin dynamics in the NVT

1452 ensemble. For convenience, we attributed equal mobilities to all particles, chose a simulation unit

1453 length of 1 nm and a time step of 0.008 in LJ units. Each simulation consisted of a total of

1454 150,000,000 time steps. The first 100,000,000 steps were used to let the system reach thermal

1455 equilibrium and a snapshot was taken every 10,000 steps after that for clustering analysis.

1456 *Analysis of clustering in the 3D simulations*

1457 To identify specific bonds between EPYC1 and Rubisco binding sites, we first computed

1458 the distance distribution histogram between all EPYC1 and Rubisco binding sites, and found a

1459 clear gap around 0.5 nm. Below this gap, the EPYC1 and Rubisco binding sites form a specific

1460 bond; above this gap, the binding sites are farther away from each other than the binding site

1461 diameter, and thus do not form a specific bond. As in the 2D case, for each snapshot we

1462 tabulated the Rubiscos connected by EPYC1s through specific bonds and plotted the fraction of

1463 Rubiscos in clusters with 10+ Rubiscos (Figure S10K).

1464 *Robustness of the magic-number effect*

1465 To confirm that the magic-number effect is robust with respect to our choice of

1466 interaction parameters, we performed additional 3D simulations in which we set the EPYC1 and

1467 Rubisco binding-site radii to be the same, and equal to either 0.8 nm, 1 nm, or 1.25 nm (i.e.

1468 factor of two volume changes), with the distance between the center of the Rubisco core and the

1469 Rubisco binding site as 5.8 nm, 6 nm, or 6.25 nm, correspondingly. We also varied the non-

1470 specific Lennard-Jones interaction energy and the EPYC1 inter-binding-site spring stiffness by a

1471 factor of 2 (increase or decrease) for a fixed concentration of Rubisco (2% volume fraction) and

1472 an equal number of Rubisco and EPYC1 binding sites. In Table S1, we report the resulting

1473 percentage of Rubiscos in clusters with 10+ Rubiscos for EPYC1s with 3, 4, or 5 binding sites

1474 for each of these parameter sets.

1475

1476

1477 **QUANTIFICATION AND STATISTICAL ANALYSIS**

1478

1479 **Division image analysis**

1480 Chi-square statistic calculations on the relative proportions of each observed inheritance

1481 pattern in RbcS1- and EPYC1-Venus were performed using an online calculator

1482 (http://www.socscistatistics.com/tests/chisquare2/Default2.aspx).

1483 The Wilcoxon Matched-Pairs Signed-Ranks Test was used to quantify the difference in

1484 the pyrenoid signal during division using an online calculator

1485 (http://www.fon.hum.uva.nl/Service/Statistics/Signed_Rank_Test.html). This nonparametric test

1486 assumes data are paired, but that pairs are independent of each other. In this case, each pair was

1487 made up of the signal from the pyrenoind of a single mother cell at 1) the minimal value during

1488 dimming, and 2) the value 15 minutes after that point (including both daughters); thus, different

1489 pairs represent different mother cells. Sample size and error definitions can be found in the

1490 relevant figure legends.

1491

1492 **Tomogram analysis**

1493 Nine tomograms were screened for their reconstruction quality based on IMOD patch-

1494 tracking scores and the resolution of Rubisco subtomogram averages produced from each

1495 tomogram's extracted particles. The final dataset used for quantitative analysis was five
1496 tomograms, representing five cells from three separate liquid cultures.
1497 Following the localization of Rubisco holoenzymes within pyrenoid tomograms by
1498 volume masking, template matching, subtomogram averaging and classification (described
1499 above), Rubisco concentrations for each tomogram were calculated by dividing the number of
1500 particles in the "positive" class by the masked volume of the pyrenoid matrix. Error values
1501 displayed in the text are standard deviations, unless marked otherwise.
1502

1503 **FRAP image analysis**

1504 Images were screened for quality before quantitative analysis: any pyrenoids with
1505 indistinct bleach regions, many saturated pixels, or loss of focus during recovery were discarded.
1506 If a pyrenoid contained 1-2 saturated pixels, those pixels were not included in quantification.
1507 Average recovery rates for n = 44 (EPYC1-Venus), 48 (RCA1-Venus), 42 (live RbcS1-Venus), 8
1508 (fixed RbcS1-Venus), or 28 (*epyc1* EPYC1-Venus) pyrenoids are shown with error bars or
1509 shading representing standard error of the mean (SEM). FRAP data were collected over 1 (*epyc1*
1510 EPYC1-Venus) or 3 (RbcS1-, EPYC1-, and RCA1-Venus) biological replicates. For comparison
1511 of recovery rates between strains, a one-way ANOVA with post-hoc Bonferroni means
1512 comparison was performed using OriginPro software; the variances for these recovery rates were
1513 6.36×10^{-4} (RbcS1-Venus), 6.58×10^{-4} (RCA1-Venus), and 8.88×10^{-4} (EPYC1-Venus). A one-way
1514 ANOVA test assumes normal distributions, independent samples, and equal variances, all of
1515 which appeared to be met in our data
1516

1517 **DATA AND SOFTWARE AVAILABILITY**

1518 The computer code used in modeling Rubisco-EPYC1 aggregation and fitting the
1519 Lennard-Jones RDF to experimental data will be made available through
1520 https://github.com/binarybin/RubiscoSimulation and https://github.com/anmartinezs/LJRDFfit/,
1521 respectively.

1522 The datasets generated during and/or analyzed during the current study are available from
1523 the corresponding authors upon reasonable request. Sequence information for the fluorescent
1524 constructs is available from GenBank under accession numbers KY550376 (pLM005-RBCS1-
1525 Venus), KX077944 (pLM005-EPYC1-Venus), and KY550375 (pLM005-RCA1-Venus). The *in*
1526 *situ* subtomogram average of *Chlamydomonas* Rubisco has been deposited in the EMDataBank
1527 with accession code EMD-3694

1528

1529

1530

1531

1532 **Supplemental Item Titles and Legends**

1533

1534 **Movie S1. Example of Pyrenoid Inheritance by Fission; Related to Figures 1 and 6.**

1535 Movie of the division shown in Figure 1A-C and Figure 6A-B. Left: an overlay of the Venus
1536 (green) and chlorophyll autofluorescence (magenta) channels, with saturated pixels masked out.
1537 Right: heat map of the Venus channel alone, with the scale identical to that in Figure 6A. Images
1538 are 2D projections of the sum of pixel values in each channel in a Z-stack through the whole cell
1539 at each time point. Scale bar = 2 μm. Time stamps correspond to Figures 1A-B and 6A-B.

1540

1541 **Movie S2. Example of Whole Pyrenoid Inheritance and Inheritance of no Apparent**
1542 **Pyrenoid; Related to Figure 2.**

1543 Movie of the division shown in Figure 2A. Left: an overlay of the Venus (green) and chlorophyll
1544 autofluorescence (magenta) channels, with saturated pixels masked out. Right: heat map of the
1545 Venus channel alone, with the scale identical to that in Figure 6A. Images are 2D projections of
1546 the sum of pixel values in each channel in a Z-stack through the whole cell at each time point.
1547 Scale bar = 2 μm. Time stamps correspond to Figure 2A.

1548

1549 **Movie S3. Example of Whole Pyrenoid Inheritance and *de novo* Pyrenoid Formation;**
1550 **Related to Figure 2.**

1551 The second divisions in this Movie are highlighted in Figure 2B; first division (not shown in
1552 Figure 2B, but shown in Figure S6B) exhibits pyrenoid fission; Left: an overlay of the Venus
1553 (green) and chlorophyll autofluorescence (magenta) channels, with saturated pixels masked out.
1554 Right: heat map of the Venus channel alone, with the scale identical to that in Figure 6A. Images

1555 are 2D projections of the sum of pixel values in each channel in a Z-stack through the whole cell

1556 at each time point. Scale bar = 2 μm. Time stamps correspond to Figure 2B.

1557

1558 **Movie S4. Sub-Nanometer Localization of Rubisco Holoenzymes within the Native**

1559 ***Chlamydomonas* Pyrenoid Matrix by *in situ* Cryo-Electron Tomography; Related to Figure**

1560 **4.**

1561 Sequential sections through the tomogram displayed in Figure 4A, followed by a reveal of the

1562 segmented membranes shown in Figure 4B, corresponding to the pyrenoid tubules and

1563 minitubules (green and yellow, respectively). Sequential sections are then shown through a

1564 binary volume with white spheres indicating localized Rubisco positions, followed by a reveal of

1565 every Rubisco position within the tomogram (magenta). Finally, there is a short tour of

1566 segmented volume, showing the interconnected pyrenoid membranes embedded within the

1567 pyrenoid matrix. Scale bar = 200 nm.

1568

1569 **Movie S5. Fluorescence Recovery After Photobleaching in Live and Fixed Pyrenoids;**

1570 **Related to Figure 5.**

1571 Movies of the RbcS1-Venus FRAP experiments on the live (top) and fixed (bottom) pyrenoids

1572 shown in Figure 5A-F. Left: images scaled to pre-bleach values. Right: images scaled to post-

1573 bleach values. Color scaling is displayed in Figure 5. Each image is the RbcS1-Venus signal in a

1574 single mid-volume plane in the pyrenoid. Scale bar = 1 μm. Time stamps correspond to Figure 5.

1575

1576 **Movie S6. "Magic number" EPYC1-Rubisco Aggregation in Simulated 3D Systems with**

1577 **EPYC1s with 3, 4, or 5 Binding Sites; Related to Figure 7.**

1578 Top: Examples of off-lattice 3D-modeled aggregation over time in EPYC1-Rubisco systems,
1579 starting from random dispersed initial configurations, in which EPYC1 contains 3 (left), 4
1580 (middle), or 5 (right) binding sites for Rubisco. Bottom: Plots of the fraction of Rubiscos in a
1581 cluster of 10 or more through time in the above movies.
1582

1583 **Table S1: Robustness of the Magic-Number Effect in 3D Off-Lattice Simulations; Related**
1584 **to Figure 7.**

1585 The EPYC1 and Rubisco binding-site radii are set as 0.8 nm, 1 nm, and 1.25 nm, and the non-specific
1586 Lennard-Jones interaction energy and the EPYC1 inter-binding-site spring stiffness are varied by a factor
1587 of 2 (increase or decrease) for a fixed concentration of Rubisco (2% volume fraction) and an equal
1588 number of Rubisco and EPYC1 binding sites. Reported is the resulting fraction of Rubiscos in clusters
1589 with 10+ Rubiscos for EPYC1s with 3, 4, or 5 binding sites for each of these parameter sets. The magic-
1590 number effect, *i.e.* strong reduction of clustering for EPYC1s with 4 binding sites, is consistently
1591 observed, except at the stiffest value of spring constant (0.48 kBT/nm2), where strong clustering is
1592 always observed. At this high stiffness, adjacent EPYC1 binding sites cannot readily bind to the same
1593 Rubisco but can still bridge different Rubiscos, favoring cluster formation.
1594
1595