

# Assembly of the algal CO<sub>2</sub>-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif

Moritz T. Meyer<sup>1</sup>, Alan K. Itakura<sup>2†</sup>, Weronika Patena<sup>1</sup>, Lianyong Wang<sup>1</sup>, Shan He<sup>1</sup>, Tom Emrich-Mills<sup>3</sup>, Chun S. Lau<sup>3</sup>, Gary Yates<sup>3</sup>, Luke C. M. Mackinder<sup>3</sup>, Martin C. Jonikas<sup>1\*</sup>.

<sup>1</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA.

<sup>2</sup>Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA.

<sup>3</sup>Department of Biology, University of York, York YO10 5DD, UK.

\*Correspondence to: [mjonikas@princeton.edu](mailto:mjonikas@princeton.edu).

†Current address: Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA.

**Abstract:** Approximately one-third of the Earth's photosynthetic CO<sub>2</sub> assimilation occurs in a pyrenoid, an organelle containing the CO<sub>2</sub>-fixing enzyme Rubisco. How constituent proteins are recruited to the pyrenoid, and how the organelle's sub-compartments - membrane tubules, a surrounding phase-separated Rubisco matrix, and a peripheral starch sheath - are held together is unknown. Using the model alga *Chlamydomonas reinhardtii*, we discovered that pyrenoid proteins share a sequence motif. We show that the motif is sufficient to target proteins to the pyrenoid and that the motif binds to Rubisco, suggesting a mechanism for targeting. The presence of the Rubisco-binding motif on proteins that localize to the tubules and on proteins that localize to the matrix-starch sheath interface suggests that the motif holds the pyrenoid's three sub-compartments together. Our findings advance our understanding of pyrenoid biogenesis and illustrate how a single protein motif can underlie the architecture of a complex multi-layered phase-separated organelle.

**One Sentence Summary:** A ubiquitous Rubisco-binding motif targets proteins to the pyrenoid and holds together the pyrenoid's three sub-compartments.

# **Main Text:**

CO<sub>2</sub> is the source of carbon for nearly the entire biosphere (1, 2), but its availability is limited in many environments. To overcome this limitation, many photosynthetic organisms use energy to locally concentrate CO<sub>2</sub> around the CO<sub>2</sub>-assimilating enzyme Rubisco (3–6). In eukaryotic algae, which perform a major fraction of primary production in the oceans (7), concentrated CO<sub>2</sub> is supplied to Rubisco inside a micro-compartment, the pyrenoid (8). The pyrenoid consists of a spheroidal protein matrix containing Rubisco, which in nearly all algae is traversed by membranous tubules through which CO<sub>2</sub> is delivered (9, 10). In green algae, the Rubisco matrix is additionally surrounded by a starch sheath (11) that limits CO<sub>2</sub> leakage out of the pyrenoid (12) (Fig. 1A).

In the model green alga *Chlamydomonas reinhardtii* (*Chlamydomonas* hereafter), two proteins are known to play central roles in pyrenoid assembly, EPYC1 (13) and SAGA1 (14) (Fig. 1B). EPYC1 is a ~35 kDa intrinsically disordered protein that phase-separates with Rubisco to form the pyrenoid matrix (13, 15, 16). SAGA1 is a ~180 kDa protein with a starch binding domain that localizes to the periphery of the matrix and is required for normal pyrenoid morphology, although the underlying molecular mechanism is unknown (14). Here, we present how further characterization of SAGA1 led us to discover that pyrenoid-localized proteins share a common Rubisco-binding protein sequence motif, providing insights into protein targeting to the pyrenoid and assembly of the pyrenoid's three sub-compartments.

In an effort to immunoprecipitate SAGA1 interacting proteins, we incubated a polyclonal SAGA1 antibody with clarified cell lysates from wildtype cells (Fig. 1C; Fig. S1). In addition to SAGA1, the antibody precipitated multiple proteins found in the pyrenoid proteome (17), suggesting at first that those proteins may interact with SAGA1. The precipitated proteins included EPYC1, SAGA2, RBMP1, RBMP2 and CSP41A (Fig. 1D and 1E; Table S1). The protein we named SAGA2 (Cre09.g394621) is 30% identical to SAGA1 (Cre11.g467712) and, like SAGA1, has a predicted starch-binding domain (Fig. S2A to S2C). The proteins we named Rubisco-Binding Membrane Proteins RBMP1 (Cre06.g261750) and RBMP2 (Cre09.g416850) have predicted transmembrane domains and were previously found to bind to Rubisco (18) (Fig. S2D to S2G). CSP41A (Cre10.g440050) is a predicted chloroplast epimerase (19) (Fig. S2H).

Surprisingly, we found that the precipitation of these pyrenoid proteome proteins was not mediated by SAGA1, but rather that the proteins were directly bound by the SAGA1 antibody.

Three lines of evidence supported this conclusion. Firstly, the same proteins were immunoprecipitated from a *saga1* mutant lysate (Fig. 1D and 1E; Table S1). Secondly, the predicted molecular weights of EPYC1, SAGA2, RBMP1, RBMP2, and CSP41A showed remarkable agreement with the multiple polypeptides recognized by the same SAGA1 antibody in immunoblots (14) (compare Fig. 1D and 1F). Thirdly, the ~35 kDa band was absent in anti-SAGA1 immunoblots of *epyc1* mutant cell extracts, strongly suggesting that EPYC1 was directly recognized by the SAGA1 antibody (Fig. 1F). We therefore hypothesized that our polyclonal SAGA1 antibody recognized similar sequences on all six pyrenoid proteins, raising questions about the nature and function of this sequence.

To identify potential sequences in the proteins that our antibody could bind, we searched their sequences for similarity to the nineteen C-terminal amino acids of SAGA1, against which our antibody had been raised (14). Sequence alignment revealed that all six proteins contain a common motif, with sequence [D/N]W[R/K]XX[L/I/V/A], present at their exact C-termini (Fig. 2A and 2B).

We identified an additional 14 variants of this motif at internal positions across all six proteins (Fig. 2A and 2B). Most internal occurrences of the motif are immediately followed by an aspartic acid (D) or a glutamic acid (E), both of which contain a carboxyl group. This observation suggests that when the motif is found at the C-terminus of the protein, the C-terminal carboxyl group of the protein is functionally important; and when the motif is found internally in the protein, this functionality is provided by the carboxyl group of the D or E side chains that follow the motif.

In summary, all six of the pyrenoid proteins share multiple copies of a common motif, which appears to have been coincidentally recognized by our SAGA1 antibody.

The prevalence of a common motif among pyrenoid-localized proteins led us to hypothesize that this motif mediates targeting of proteins to the pyrenoid. To test this hypothesis, we evaluated the impact of adding the motif to ferredoxin 1 (FDX1, Cre14.g626700), a small soluble protein that natively localizes throughout the chloroplast stroma. To increase the chances of observing the motif's effect, we chose to add three tandem copies of the motif to FDX1. Strikingly, addition of the motif re-localized fluorescently-tagged FDX1 exclusively to the pyrenoid matrix (Fig. 2C; Fig. S3A and S3B). We obtained a similar result with another chloroplast protein (Fig. S3C and S3E). These results demonstrate that the motif is sufficient to localize a chloroplast protein to the pyrenoid matrix.

Two observations led us to hypothesize that proteins bearing the motif are recruited to the pyrenoid via binding to Rubisco. First, the motif is present in each of the regions that were found to mediate EPYC1's binding to Rubisco in a parallel study (20). Second, the other five motif-containing proteins were also previously found to bind to Rubisco: SAGA1 by yeast two-hybrid (14), and SAGA2, RBMP1, RBMP2 and CSP41A by affinity purification-mass spectrometry (18).

To determine whether each variant of the motif can bind to Rubisco, we used surface plasmon resonance to measure the binding of synthetic peptides to Rubisco. With this method, 15 out of 20 peptides representing motif variants had a higher affinity to Rubisco than peptides with random sequences. Peptides with C-terminal motifs systematically showed higher affinity to Rubisco than peptides with internal motifs (Fig. 3A). Rubisco bound to all predicted internal motif sites when we incubated purified Rubisco with arrays of peptides tiling across the full-length proteins (Fig. 3B and 3C; Fig. S4A to S4D; C-terminal sites could not be assayed by this method). These results indicate that nearly all instances of the motif bind to Rubisco *in vitro*. Importantly, binding of the proteins to the SAGA1 antibody in the immunoprecipitation experiment (Fig. 1C and 1D) suggests that at least one motif on each protein is accessible for Rubisco binding when the proteins adopt their native folding *in vivo*.

Similar motifs are present in proteins other than the ones we studied here. We identified putative motifs in the proteome using a simple scoring scheme (Table S2). Cre10.g430350, a protein with a predicted motif identified by this search, localized to the pyrenoid matrix (Fig. S5A). High motif scores were modestly enriched among pyrenoid proteome proteins relative to other proteins in the chloroplast proteome (Fig. S5B). Not all pyrenoid proteome proteins have the motif; some proteins may be targeted to the pyrenoid via other mechanisms, for example by binding to a motif-containing protein. The motif is present in proteins that are not in the pyrenoid proteome; this could be due to these motifs not being accessible for interaction with Rubisco.

Serendipitously, in a parallel study (20) we determined where the motif binds on Rubisco. In that study, as part of an effort to understand how EPYC1 clusters Rubisco to form the pyrenoid matrix, we obtained a cryoelectron microscopy structure of Rubisco bound to a peptide from EPYC1. Remarkably, a Rubisco-binding motif is present at positions N<sub>62</sub>W<sub>63</sub>R<sub>64</sub>Q<sub>65</sub>E<sub>66</sub>L<sub>67</sub>E<sub>68</sub> on this EPYC1 peptide and plays a central role in the binding interface. One EPYC1 peptide binds to each of the eight Rubisco small subunits of the Rubisco holoenzyme. The motif-containing portion of the peptide adopts an alpha-helix that binds to the Rubisco surface. R<sub>64</sub> of the peptide,

corresponding to the [R/K] of the motif, forms a salt bridge with a glutamic acid of the Rubisco small subunit. Additionally, W<sub>63</sub> and L<sub>67</sub> of the peptide, corresponding to the W and [L/I/V/A] of the motif, respectively, contribute to a hydrophobic interaction with three hydrophobic residues of the Rubisco small subunit. The central role of the motif residues in the binding interface strongly suggests that all Rubisco-binding motifs studied here bind to Rubisco using the same mechanism.

Together, our findings suggest a mechanism for targeting proteins to the pyrenoid, involving the presence of a common motif which recruits its protein to the pyrenoid via direct binding interactions with Rubisco. It is possible that the mechanism operates via random diffusion of the motif-bearing protein through the chloroplast, followed by capture of the motif by Rubisco when the protein encounters the pyrenoid matrix.

Beyond providing a mechanism for targeting proteins to the pyrenoid matrix, the motif appears to play a role at the interfaces of the pyrenoid's three sub-compartments. Whereas Rubisco and EPYC1 are localized in the matrix (Fig. 4A; Fig. S6A and S6B), we observed that some motif-containing proteins localize to pyrenoid regions other than the matrix. Fluorescently tagged RBMP1 and RBMP2 localized to the tubules (Fig. 4A; Fig. S6C and S6D). SAGA1 (14) and SAGA2 localized to the interface between the Rubisco matrix and the starch sheath (Fig. 4A; Fig. S6E and S6F).

These observations together with previous work (9, 13-17, 20) are consistent with a model where the Rubisco-binding motif holds together the pyrenoid matrix, the traversing tubules, and the surrounding starch sheath (Fig. 4B). In this model, multiple copies of the motif on EPYC1 mediate cohesion of the matrix by bringing Rubisco holoenzymes together (20). Moreover, the presence of the same motif on the pyrenoid tubule-localized transmembrane proteins, RBMP1 and RBMP2, recruits Rubisco to the tubules and favors assembly of the matrix around them. Finally, the presence of the motif on proteins with starch-binding domains, SAGA1 and SAGA2, which localize to the pyrenoid periphery, mediates adherence of the starch sheath to the matrix.

Our microscopy data suggest that RBMP2 is confined to the central reticulated region of the tubules (9), whereas RBMP1 appears to localize to the more peripheral tubular regions to the exclusion of the reticulated region. These different localization patterns suggest that RBMP1 and RBMP2 could each promote Rubisco matrix binding to a different part of the tubules. As described previously (14), SAGA1 localized to puncta at the periphery of the matrix, likely at the interface between the Rubisco matrix and the starch sheath. We observed that SAGA2 also localized to that

interface but appeared to cover the surface of the matrix more homogeneously than SAGA1. Similarly to the RBMPs, the different localization patterns of the SAGAs suggest that they may each interface with different features on the starch sheath.

The model explains several previously puzzling observations. First, in a mutant lacking EPYC1, a pyrenoid-like structure still assembles around the tubules, containing some Rubisco enclosed by a starch sheath, even though the canonical matrix is absent (13). The presence of both Rubisco and starch in this structure can be explained by a layer of Rubisco that serves as a bridge between motif-containing proteins RBMP1 and RBMP2 on the tubules and motif-containing proteins SAGA1 and SAGA2 on the surrounding starch sheath (Fig. 4C). Second, point mutations that disrupt the EPYC1 binding site on Rubisco not only eliminate the matrix but additionally disrupt the association of the starch sheath with the pyrenoid (20). The additional disruption of the starch sheath can be explained by the same Rubisco binding site being required not only for binding to EPYC1 but also for binding the motif on other proteins that connect the starch and tubules to Rubisco (Fig. 4D).

Our work reveals a ubiquitous Rubisco-binding motif that is both sufficient for targeting proteins to the pyrenoid and also appears to mediate the overall assembly of the pyrenoid's three sub-compartments. The eight-fold symmetry of the Rubisco holoenzyme allows it to interact simultaneously with multiple binding partners via the motif, making Rubisco a central structural hub of the pyrenoid. The valency and binding strengths of the motif to Rubisco vary among the binding partners (Fig. 2A and Fig. 3), which could play a role in their relative priority of access to Rubisco, as observed for other phase-separated organelles (21–23).

The Rubisco-EPYC1 condensate can enhance CO<sub>2</sub> fixation only if it is anchored around the pyrenoid tubules, which are thought to provide concentrated CO<sub>2</sub> (8, 9, 24, 25). Identification in the present work of two tubule-localized proteins that bind to Rubisco provides a plausible explanation for how the *Chlamydomonas* pyrenoid forms preferentially around tubules, rather than anywhere else in the chloroplast. Considering that the tubules can form independently of the matrix (26, 27), whether the motif is required for RBMPs to localize to the tubules remains to be investigated. Our observation that RBMP1 and RBMP2 localize to different sub-domains of the tubules suggests that an additional localization mechanism may be at play, such as a preference for a specific membrane curvature.

Intriguingly, RBMP1 is predicted to be a  $\text{Ca}^{2+}$ -activated anion channel of the bestrophin family (Fig. S2D and S3F), of which several members are thought to supply  $\text{HCO}_3^-$  to the lumen of the tubules for conversion to  $\text{CO}_2$  (28). Whereas the previously described members localized primarily to membranes outside the pyrenoid, RBMP1 localizes exclusively to the tubules themselves, which may allow it to directly feed  $\text{HCO}_3^-$  to the tubules for conversion to  $\text{CO}_2$ .

We hypothesize that the organizational principle described here applies broadly to pyrenoids across the tree of life. Pyrenoids are thought to have evolved independently multiple times through convergent evolution (29, 30). Their convergent evolution may well have been facilitated by the possibility of using a common motif and binding site on Rubisco to perform three functions essential to all pyrenoids: clustering of Rubisco into a matrix, targeting of proteins to the matrix, and connecting the matrix to other structures.

Our findings advance the basic understanding of the biogenesis of the pyrenoid and provide a framework for engineering a pyrenoid into crops for improving yields (8, 31, 32). More broadly, the system presented here provides a remarkable example of how the architecture of a complex phase-separated organelle can be defined by a simple organizing principle.



# References:

1. J. A. Raven, Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56**, 177–192 (2009).
2. A. Bar-Even, E. Noor, R. Milo, A Survey of Carbon Fixation Pathways Through a Quantitative Lens. *J. Exp. Bot.* **63**, 2325–2342 (2012).
3. C. J. Still, J. A. Berry, G. J. Collatz, R. S. DeFries, Global distribution of C<sub>3</sub> and C<sub>4</sub> vegetation: Carbon cycle implications. *Global Biogeochem. Cycles.* **17**, 6-1-6–14 (2003).
4. E. Granum, J. A. Raven, R. C. Leegood, How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? *Can. J. Bot.* **83**, 898–908 (2005).
5. J. A. Raven, J. Beardall, M. Giordano, Energy Costs of Carbon Dioxide Concentrating Mechanisms in Aquatic organisms. *Photosynth. Res.* **121**, 111–124 (2014).
6. S. C. Maberly, B. Gontero, Ecological Imperatives for Aquatic CO<sub>2</sub>-Concentrating Mechanisms. *J. Exp. Bot.* **68**, 3797–3814 (2017).
7. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science.* **281**, 237–240 (1998).
8. J. H. Hennacy, M. C. Jonikas, Prospects for Engineering Biophysical CO<sub>2</sub> Concentrating Mechanisms into Land Plants to Enhance Yields. *Annu. Rev. Plant Biol.* **71** (2020), doi:10.1146/annurev-arplant-081519-040100.
9. B. D. Engel, M. Schaffer, L. Kuhn Cuellar, E. Villa, J. M. Plitzko, W. Baumeister, Native Architecture of the Chlamydomonas Chloroplast Revealed by in Situ Cryo-Electron Tomography. *Elife.* **4**, e11383 (2015).
10. M. T. Meyer, C. Whittaker, H. Griffiths, The Algal Pyrenoid: Key Unanswered Questions. *J. Exp. Bot.* **68**, 3739–3749 (2017).
11. J. D. Pickett-Heaps, *Green Algae: Structure, Reproduction and Evolution in Selected Genera* (Sinauer Associates, 1975).
12. C. Toyokawa, T. Yamano, H. Fukuzawa, Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO<sub>2</sub>-Concentrating Mechanism in Green Algae. *Plant Physiol.* **182**, 1883–1893 (2020).
13. L. C. M. Mackinder, M. T. Meyer, T. Mettler-Altmann, V. K. Chen, M. C. Mitchell, O. D.



- Caspari, E. S. Freeman Rosenzweig, L. Pallesen, G. Reeves, A. K. Itakura, R. Roth, F. Sommer, S. Geimer, T. Mühlhaus, M. Schroda, U. W. Goodenough, M. Stitt, H. Griffiths, M. C. Jonikas, A Repeat Protein Links Rubisco to Form the Eukaryotic Carbon-Concentrating Organelle. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5958–5963 (2016).
14. A. K. Itakura, K. X. Chan, N. Atkinson, L. Pallesen, L. Wang, G. Reeves, W. Patena, O. D. Caspari, R. Roth, U. W. Goodenough, A. J. McCormick, H. Griffiths, M. C. Jonikas, A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18445–18454 (2019).
15. E. S. Freeman Rosenzweig, B. Xu, L. Kuhn Cuellar, A. Martinez-Sanchez, M. Schaffer, M. Strauss, H. N. Cartwright, P. Ronceray, J. M. Plitzko, F. Förster, N. S. Wingreen, B. D. Engel, L. C. M. Mackinder, M. C. Jonikas, The Eukaryotic CO<sub>2</sub>-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell*. **171**, 148-162.e19 (2017).
16. T. Wunder, S. L. H. Cheng, S. K. Lai, H. Y. Li, O. Mueller-Cajar, The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat. Commun.* **9**, 5076 (2018).
17. Y. Zhan, C. H. Marchand, A. Maes, A. Mauries, Y. Sun, J. S. Dhaliwal, J. Uniacke, S. Arragain, H. Jiang, N. D. Gold, V. J. J. Martin, S. D. Lemaire, W. Zerges, Pyrenoid Functions Revealed by Proteomics in *Chlamydomonas Reinhardtii*. *PLoS One*. **13**, e0185039 (2018).
18. L. C. M. Mackinder, C. Chen, R. D. Leib, W. Patena, S. R. Blum, M. Rodman, S. Ramundo, C. M. Adams, M. C. Jonikas, A Spatial Interactome Reveals the Protein Organization of the Algal CO<sub>2</sub>-Concentrating Mechanism. *Cell*. **171**, 133-147.e14 (2017).
19. M. E. Baker, W. N. Grundy, C. P. Elkan, Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. *Biochem. Biophys. Res. Commun.* **248**, 250–254 (1998).
20. S. He, H.-T. Chou, D. Matthies, T. Wunder, M. T. Meyer, N. Atkinson, A. Martinez-Sanchez, P. D. Jeffrey, S. A. Port, W. Patena, G. He, V. K. Chen, F. M. Hughson, A. J. McCormick, O. Mueller-Cajar, B. D. Engel, Z. Yu, M. C. Jonikas, The structural basis of Rubisco phase separation in the pyrenoid.
21. S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, M. K. Rosen,

- Compositional Control of Phase-Separated Cellular Bodies. *Cell*. **166**, 651–663 (2016).
22. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49** (2020), doi:10.1146/annurev-biophys-121219-081629.
- 5 23. E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, T. Mittag, Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-Like Domains. *Science*. **367**, 694–699 (2020).
24. N. A. Pronina, V. E. Semenenko, Localization of membrane bound and soluble carbonic anhydrase in the Chlorella cell. *Fiziol. Rastenii*. **31**, 241–251 (1984).
- 10 25. J. A. Raven, CO<sub>2</sub>-concentrating mechanisms: a direct role for thylakoid lumen acidification? *Plant. Cell Environ.* **20**, 147–154 (1997).
26. U. W. Goodenough, R. P. Levine, Chloroplast Structure and Function in *ac-20*, a Mutant Strain of Chlamydomonas Reinhardtii. 3. Chloroplast Ribosomes and Membrane Organization. *J. Cell Biol.* **44**, 547–562 (1970).
- 15 27. O. D. Caspari, M. T. Meyer, D. Tolleter, T. M. Wittkopp, N. J. Cunniffe, T. Lawson, A. R. Grossman, H. Griffiths, Pyrenoid Loss in *Chlamydomonas Reinhardtii* Causes Limitations in CO<sub>2</sub> Supply, but Not Thylakoid Operating Efficiency. *J. Exp. Bot.* **68**, 3903–3913 (2017).
28. A. Mukherjee, C. S. Lau, C. E. Walker, A. K. Rai, C. I. Prejean, G. Yates, T. Emrich-Mills, S. G. Lemoine, D. J. Vinyard, L. C. M. Mackinder, J. V. Moroney, Thylakoid Localized Bestrophin-Like Proteins Are Essential for the CO<sub>2</sub> Concentrating Mechanism of *Chlamydomonas Reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16915–16920 (2019).
- 20 29. J. C. Villarreal, S. S. Renner, Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18873–18878 (2012).
- 25 30. J. A. Raven, J. Beardall, P. Sánchez-Baracaldo, The possible evolution and future of CO<sub>2</sub>-concentrating mechanisms. *J. Exp. Bot.* **68**, 3701–3716 (2017).
31. B. D. Rae, B. M. Long, B. Förster, N. D. Nguyen, C. N. Velanis, N. Atkinson, W. Y. Hee, B. Mukherjee, G. D. Price, A. J. McCormick, Progress and Challenges of Engineering a Biophysical CO<sub>2</sub>-concentrating Mechanism Into Higher Plants. *J. Exp. Bot.* **68**, 3717–3737 (2017).
- 30

32. L. C. M. Mackinder, The Chlamydomonas CO<sub>2</sub>-Concentrating Mechanism and Its Potential for Engineering Photosynthesis in Plants. *New Phytol.* **217**, 54–61 (2018).

**Acknowledgements:** We thank Christopher Dupont (JCVI, La Jolla, CA), as well as present members and alumni of the Jonikas laboratory for helpful discussions and feedback on the manuscript; the Princeton University Confocal Microscopy and Biophysics core facility managers, Drs. Gary Laevsky and Venu Gopal Vandavasi for instrumentation support; Ryan Leib and Christopher Adams at the Stanford University Mass Spectrometry facility. **Funding:** The project was funded by National Science Foundation (IOS-1359682 and MCB-1935444), National Institutes of Health (DP2-GM-119137), and Simons Foundation and Howard Hughes Medical Institute (55108535) grants to M.C.J.; and the UK Biotechnology and Biological Sciences Research Council (BB/R001014/1) grant to L.C.M. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.

**Author contributions:** M.T.M., A.I. and M.C.J. conceived the project. A.I. performed the coimmunoprecipitation experiment and prepared samples for analysis by mass spectrometry. M.T.M. purified Rubisco and performed western blots, surface plasmon resonance, protein re-localization experiments, and confocal microscopy. S.H. and M.T.M. performed the peptide tiling array experiment. T.E.-M., J.L., G.Y., L.W., M.T.M. and L.C.M. produced fluorescently tagged strains. W.P. conducted the bioinformatic analyses. All authors analyzed the data. M.T.M. and M.C.J. wrote the manuscript with input from all authors. M.C.J. supervised the work. **Competing interests:** Princeton University, Stanford University, and the University of York have submitted a provisional patent application on aspects of the findings. **Data and material availability:** all data are available in the main text or the supplementary materials.

## **Supplementary Materials:**

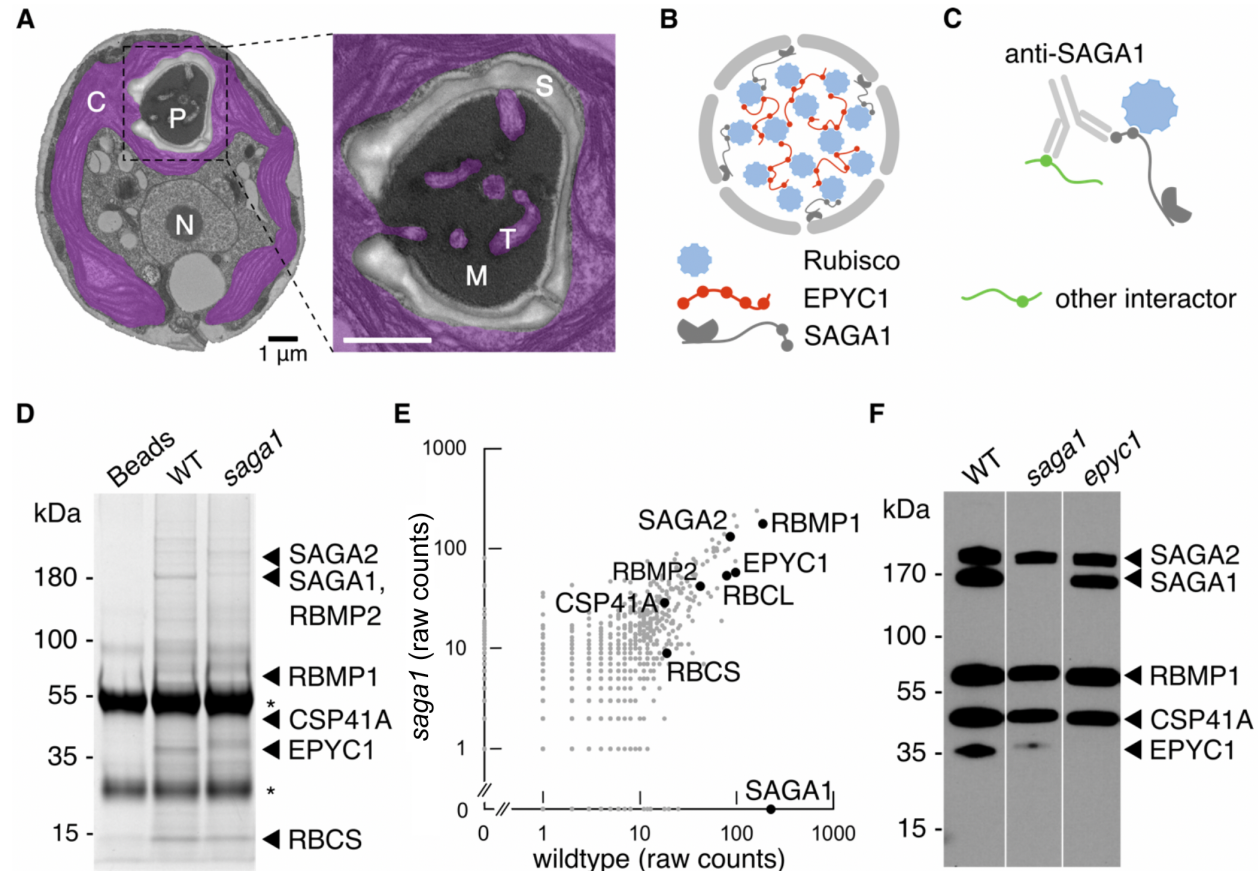
Materials and Methods

Figures S1-S6

Tables S1-S2

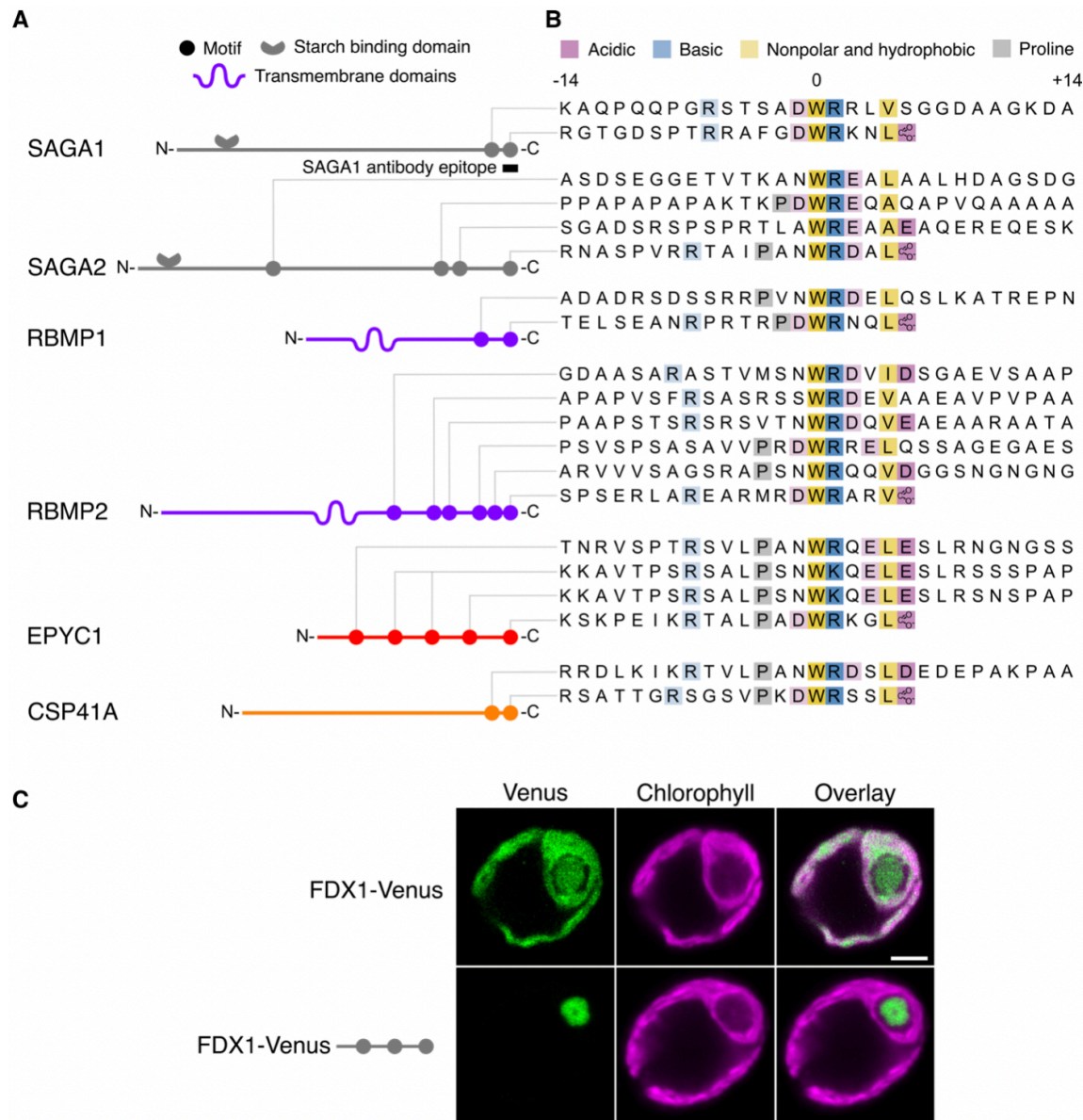
5      References (*1-45*)

## Figures:



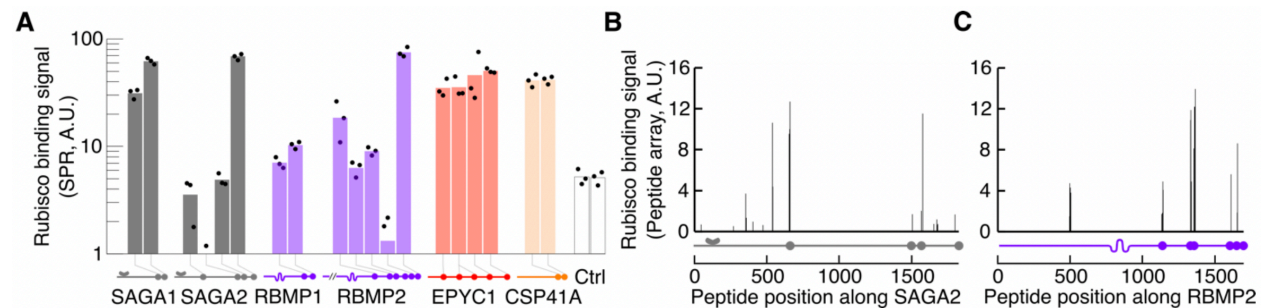
**Fig. 1. A polyclonal antibody raised against the pyrenoid protein SAGA1 interacts with at least five other pyrenoid proteins.** (A) Electron micrograph of a median plane section through an air acclimated *Chlamydomonas* cell. N: nucleus; C: chloroplast; P: pyrenoid; M: Rubisco matrix; T: tubules; S: starch sheath. Scale bar, 1  $\mu$ m. (B) Two proteins, the Rubisco linker EPYC1 and the starch sheath-binding protein SAGA1, have been previously characterized and localized to the pyrenoid. (C) An anti-SAGA1 antibody was incubated with cell lysate in an effort to coimmunoprecipitate proteins that bind to SAGA1. (D) Coomassie stained SDS-PAGE of proteins immunoprecipitated by the anti-SAGA1 antibody from wildtype (WT) and *saga1* mutant lysates. Immunoprecipitated proteins were eluted from anti-SAGA1 antibodies on beads by boiling; beads not incubated with lysate were also boiled for reference. (\*): heavy and light immunoglobulin chains. (E) Proteins immunoprecipitated by the SAGA1 antibody from wildtype and *saga1* were identified by mass spectrometry. Raw spectral counts are plotted on a log scale. (F) Anti-SAGA1 western blot on denatured protein extracted from wildtype, *saga1* and *epyc1*.





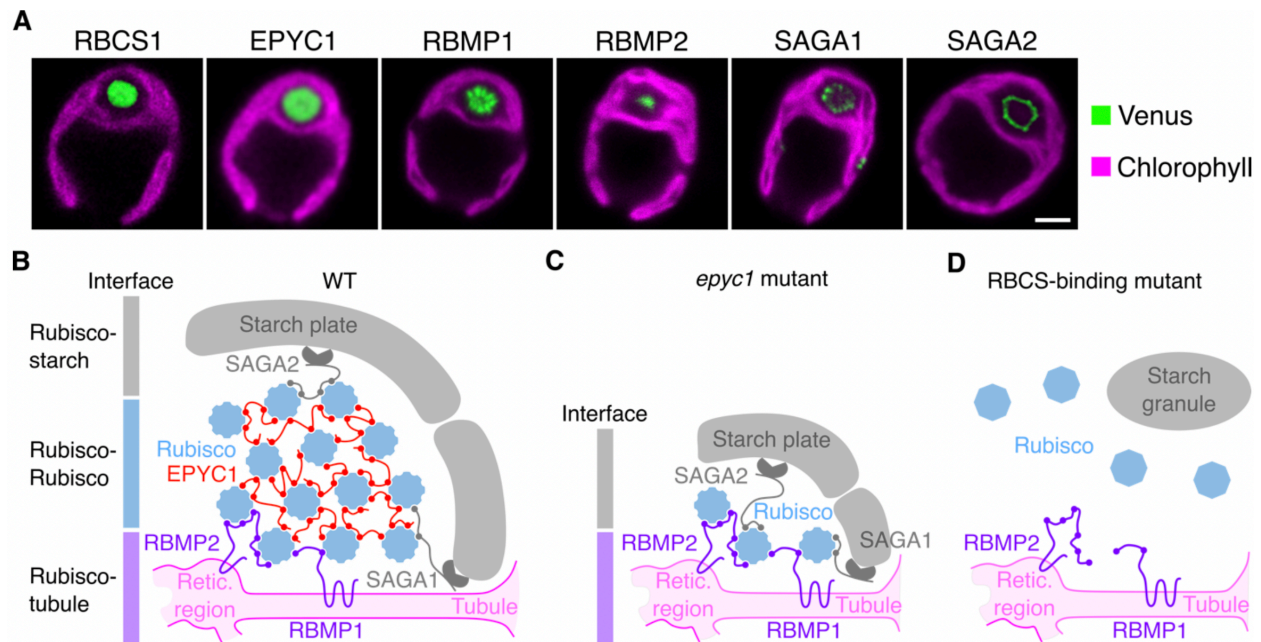
**Fig. 2. A motif found on pyrenoid proteins is sufficient for targeting proteins to the pyrenoid.**

(A) The location of motifs along the primary sequence of each protein is shown (not to scale). (B) Sequence alignment of protein regions containing the pyrenoid motif. Motif residues are colored by physicochemical properties. Intensity of coloring is proportional to frequency at a given amino acid position. Peptides with the sequences shown in B were synthesized and their binding to Rubisco was measured by surface plasmon resonance and peptide tiling array (see Fig. 3). (C) The localization of Venus-tagged ferredoxin 1 protein (FDX1) without and with the C-terminal addition of three copies of the C-terminal SAGA2 motif was determined by transforming the corresponding constructs into wildtype *Chlamydomonas*. Scale bar, 2µm.



**Fig. 3. The motif binds to Rubisco.** (A) Peptides containing motifs from the indicated proteins were synthesized and their binding to Rubisco was measured by surface plasmon resonance (A.U., arbitrary units). Two peptides not containing the motif were included as controls. Internal peptides were 29 amino acids; C-terminal peptides were 19 amino acids; the signal is normalized by peptide molecular weight. The positions of the predicted motifs are indicated below the graph (not to scale). (B) Array of 18 amino acid peptides tiling across the sequence of SAGA2 was synthesized and probed with Rubisco. (C) Tiling array for RBMP2. Binding signal in (B) and (C) is normalized to a control EPYC1 peptide known to bind to Rubisco (20) (one unit of binding). The positions of motifs are indicated to scale below each graph.





**Fig. 4. The motif orchestrates the architecture of the pyrenoid's three sub-compartments.**

(A) Representative confocal images of Venus-tagged proteins that possess the Rubisco-binding motif, and Rubisco small subunit (RBCS1). Chlorophyll autofluorescence delimits the chloroplast. Scale bar, 2  $\mu$ m. (B) Proposed model for how the motif mediates assembly of the pyrenoid's three sub-compartments in wildtype. The motif on tubule-localized transmembrane proteins RBMP1 and RBMP2 mediates Rubisco binding to the tubules (Retic. region = reticulated region of tubules (9)). Multiple copies of the motif on EPYC1 link Rubiscos to form the pyrenoid matrix (20). At the periphery of the matrix, the motif on starch-binding proteins SAGA1 and SAGA2 mediates interactions between the matrix and surrounding starch sheath. (C) The model in Fig. 4B explains the matrix-less phenotype observed in EPYC1-less mutants. (D) The model also explains the absence of matrix and starch plates in mutants where Rubisco's binding site for the motif has been disrupted (20).