The structural basis of rubisco phase separation in the pyrenoid

The Structural Basis of Rubisco Phase Separation in the Pyrenoid

Shan He[1], Hui-Ting Chou[2,10], Doreen Matthies[2], Tobias Wunder[3], Moritz T. Meyer[1], Nicky Atkinson[4], Antonio Martinez-Sanchez[5,6], Philip D. Jeffrey[1], Sarah A. Port[1], Weronika Patena[1], Guanhua He[1], Vivian K. Chen[7], Frederick M. Hughson[1], Alistair J. McCormick[4], Oliver Mueller-Cajar[3], Benjamin D. Engel[5,8,9], Zhiheng Yu[2], Martin C. Jonikas[1,*]

[1]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. [2]Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA 20147, USA. [3]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore. [4]SynthSys & Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, UK. [5]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. [6]Institute of Neuropathology, University of Göttingen Medical Center, 37075 Göttingen, Germany. [7]Department of Biology, Stanford University, Stanford, CA 94305, USA. [8]Helmholtz Pioneer Campus, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [9]Department of Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany. [10]Present address: Department of Therapeutic Discovery, Amgen Discovery Research, Amgen Inc., South San Francisco, CA 94080, USA.

[*]Correspondence and Lead Contact: mjonikas@princeton.edu.

Abstract

Approximately one-third of global CO_2 fixation occurs in a phase-separated algal organelle called the pyrenoid. Existing data suggest that the pyrenoid forms by the phase separation of the CO_2-fixing enzyme Rubisco with a linker protein; however, the molecular interactions underlying this phase separation remain unknown. Here we present the structural basis of the interactions between Rubisco and its intrinsically disordered linker protein EPYC1 (Essential Pyrenoid Component 1) in the model alga *Chlamydomonas reinhardtii*. We find that EPYC1 consists of five evenly-spaced Rubisco-binding regions that share sequence similarity. Single-particle cryo-electron microscopy of one of these regions in complex with Rubisco indicates that each Rubisco holoenzyme has eight binding sites for EPYC1, one on each Rubisco small subunit. Interface mutations disrupt binding, phase separation, and pyrenoid formation. Cryo-electron tomography supports a model where EPYC1 and Rubisco form a co-dependent multivalent network of specific low-affinity bonds, giving the matrix liquid-like properties. Our results advance the structural and functional understanding of the phase separation underlying the pyrenoid, an organelle that plays a fundamental role in the global carbon cycle.

Main Text

The CO_2-fixing enzyme Rubisco drives the global carbon cycle, mediating the assimilation of approximately 100 gigatons of carbon per year[1]. The gradual decrease of atmospheric CO_2 over billions of years[2] has made Rubisco's job increasingly difficult, to the point where CO_2 assimilation limits the growth rate of many photosynthetic organisms[3]. This selective pressure is thought to have led to the evolution of CO_2 concentrating mechanisms, which feed concentrated CO_2 to Rubisco to enhance growth[4]. Of these mechanisms, the most poorly understood relies on the pyrenoid, a phase separated organelle[5] found in the chloroplast of nearly all eukaryotic algae and some land plants (Fig. 1a, b)[6,7]. The pyrenoid enhances the activity of Rubisco by clustering it around modified thylakoid membranes that supply Rubisco with concentrated CO_2[8,9].

For decades, the mechanism for packaging the Rubisco holoenzyme into the pyrenoid remained unknown. Recent work showed that in the leading model alga *Chlamydomonas reinhardtii* (Chlamydomonas hereafter), the clustering of Rubisco into the pyrenoid matrix requires the Rubisco-binding protein EPYC1[10]. EPYC1 and Rubisco are the most abundant components of the pyrenoid and bind to each other. Moreover, combining purified EPYC1 and Rubisco together produces phase-separated condensates[11] that mix internally at a rate similar to that observed for the matrix *in vivo*[5], suggesting that these two proteins are sufficient to form the structure of the pyrenoid matrix. The sequence repeats within EPYC1 and eight-fold symmetry of the Rubisco holoenzyme led us to hypothesize that EPYC1 and Rubisco each have multiple binding sites for the other, allowing the two proteins to form a co-dependent condensate (Fig. 1c)[10].

Here, we determined the structural basis that underlies the EPYC1-Rubisco condensate. Using biophysical approaches, we found that EPYC1 has five evenly spaced Rubisco-binding regions that share sequence homology and can bind to Rubisco as short peptides. We obtained cryo-electron microscopy structures showing that each of EPYC1's Rubisco-binding regions forms an α-helix that binds one of Rubisco's eight small subunits via salt bridges and hydrophobic interactions. Mapping of these binding sites onto Rubisco holoenzymes within the native pyrenoid matrix indicates that the linker sequences between Rubisco-binding regions on EPYC1 are sufficiently long to connect together adjacent Rubisco holoenzymes. These

discoveries advance the understanding of the pyrenoid and provide a high-resolution structural view of a phase-separated organelle.

Results

We could not directly determine the structure of full-length EPYC1 bound to Rubisco because mixing the two proteins together produces phase-separated condensates[11]. We thus aimed to first identify Rubisco-binding regions on EPYC1, and subsequently to use a structural approach to determine how these regions bind to Rubisco.

EPYC1 has five nearly identical Rubisco-binding regions

The intrinsically disordered nature of purified EPYC1[11] led us to hypothesize that the Rubisco-binding regions of EPYC1 were short and could bind to Rubisco as peptides without a need for tertiary folds. Therefore, to identify EPYC1 regions that bind to Rubisco, we synthesized a peptide array consisting of 18 amino acid peptides tiling across the full length EPYC1 sequence (Fig. 1d), and probed this array with native Rubisco purified from Chlamydomonas cells (Fig. 1e, f).

Our tiling array revealed five evenly-spaced Rubisco-binding regions on EPYC1, each consisting of a predicted α-helix and an upstream region (Fig. 1g, h and Supplementary Table 3). We confirmed the binding regions using surface plasmon resonance (SPR; Extended Data Fig. 1b, c). Sequence alignment guided by the five binding regions revealed that mature EPYC1 consists entirely of five sequence repeats (Fig. 1i), in contrast to the previously defined four repeats and two termini[10] (Extended Data Fig. 1a). Our alignment indicates that the previously defined EPYC1 N- and C- termini, which at the time were not considered part of the repeats, actually share sequence homology with the central repeats.

The presence of a Rubisco-binding region on each of the previously defined EPYC1 repeats (Extended Data Fig. 1a) explains our yeast two-hybrid observations[12] that a single EPYC1 repeat can interact with Rubisco, that knocking out the α-helix in an EPYC1 repeat disrupts this interaction, and that decreasing the number of EPYC1 repeats leads to a proportional decrease in EPYC1 interaction with Rubisco. It also explains our observations that

decreasing the number of EPYC1 repeats leads to a proportional decrease in the tendency of EPYC1 and Rubisco to phase separate together[11].

EPYC1 binds to Rubisco small subunits

The sequence homology of the five Rubisco-binding regions suggests that each region binds to Rubisco in a similar manner. To identify the binding site of EPYC1 on Rubisco, we determined three structures by using single-particle cryo-electron microscopy. The first structure consists of Rubisco in complex with peptide $EPYC1_{49-72}$, representing the first Rubisco-binding region of EPYC1 (2.13 Å overall resolution; ~2.5 Å EPYC1 peptide local resolution; Fig. 2, Extended Data Fig. 2-5; Supplementary Table 1). The second structure consists of Rubisco in complex with a second peptide, $EPYC1_{106-135}$, representing the second, third and fourth Rubisco-binding regions of EPYC1 (2.06 Å overall resolution, ~2.5 Å EPYC1 peptide local resolution, Extended Data Fig. 6). The affinities of these peptides to Rubisco were low by protein interaction standards (K_D ~3 mM; Extended Data Fig. 1d, e); thus, millimolar concentrations of peptide were required to approach full occupancy of peptide bound to Rubisco. For reference purposes, we also obtained a third structure of Rubisco in the absence of EPYC1 peptide (2.68 Å; Extended Data Fig. 2 and 3), which was nearly identical to the previously published X-ray crystallography structure[13], with minor differences likely due to the absence of the substrate analog 2-CABP in the active site of Rubisco in our sample[14] (Extended Data Fig. 4).

The structures of Rubisco in complex with $EPYC1_{49-72}$ and of Rubisco in complex with $EPYC1_{106-135}$ were remarkably similar, indicating that these two peptides and the corresponding four regions of EPYC1 each bind to the same site on the Rubisco holoenzyme. The Rubisco holoenzyme consists of a core of eight catalytic large subunits in complex with eight small subunits, four of which cap each end of the holoenzyme (Fig. 2b-e). In each structure, an EPYC1 peptide was clearly visible bound to each Rubisco small subunit, suggesting that each Rubisco holoenzyme can bind up to eight EPYC1s (Fig. 2b-e and Extended Data Fig. 6b, c).

Salt bridges and a hydrophobic interface mediate binding

Both the $EPYC1_{49-72}$ and $EPYC1_{106-135}$ peptides formed an extended chain that sits on top of the Rubisco small subunit's two α-helices (Fig. 3a, b, Extended Data Fig. 6d, e). This binding site explains our previous observations that mutations in the Rubisco small subunit α-helices

disrupted yeast two-hybrid interactions between EPYC1 and the Rubisco small subunit[12] and prevented Rubisco's assembly into a pyrenoid *in vivo*[15]. The C-terminal regions of the $EPYC1_{49-72}$ and $EPYC1_{106-135}$ peptides (NW[R/K]QELESLR[N/S]) are well-resolved; each forms an α-helix that runs parallel to helix B of the Rubisco small subunit (Fig. 3a, b). The peptides' N-termini extend the trajectory of the helix and follow the surface of the Rubisco small subunit (Fig. 2b-e, 3a-b and Extended Data Fig. 5, Extended Data Fig. 6b, c). The side chains of the peptides' N-termini could not be well resolved, suggesting that these regions are more conformationally flexible.

Our atomic models based on the density maps suggest that binding is mediated by salt bridges and a hydrophobic interface. Three residue pairs of $EPYC1_{49-72}$ likely form salt bridges (Fig. 3c, d and g): EPYC1 residues R64 and R71 interact with E24 and D23, respectively, of Rubisco small subunit α-helix A, and EPYC1 residue E66 interacts with R91 of Rubisco small subunit α-helix B. Furthermore, a hydrophobic interface is formed by W63, L67 and L70 of EPYC1 and M87, L90 and V94 of Rubisco small subunit helix B (Fig. 3e-g). Similar interactions were observed for the corresponding residues in $EPYC1_{106-135}$ (Extended Data Fig. 6f-j).

Binding and phase separation require interface residues

To determine the importance of individual EPYC1 residues for binding, we investigated the impact on Rubisco binding of every possible single amino acid substitution for EPYC1's first Rubisco-binding region by using a peptide array (Fig. 4a and Supplementary Table 4) and SPR (Extended Data Fig. 7). Consistent with our structural model, the peptide array indicated that EPYC1 salt bridge-forming residues R64, R71 and E66 and the hydrophobic interface residues W63, L67 and L70 were all required for normal EPYC1 binding to Rubisco. The strong agreement of our mutational analysis suggests that our structural model correctly represents EPYC1's Rubisco-binding interface.

To determine the importance of EPYC1's Rubisco-binding regions for pyrenoid matrix formation, we assayed the impact of mutations in these regions on formation of phase separated droplets by EPYC1 and Rubisco *in vitro*. The phase boundary was shifted by mutating R64 in the first Rubisco-binding region and the corresponding K or R in the other four Rubisco-binding regions of EPYC1 (Fig. 4b and Extended Data Fig. 8), suggesting that the Rubisco-binding regions mediate condensate formation.

Pyrenoid matrix formation requires interface residues

We validated the importance of Rubisco residues for binding to EPYC1 by yeast two-hybrid assays (Fig. 5a and Extended Data Fig. 9). A Rubisco small subunit D23A mutation, which eliminates the charge of the aspartate residue, had a severe impact on Rubisco small subunit interaction with EPYC1, as expected from the contribution of that residue to a salt bridge with R71 and homologous residues of EPYC1's Rubisco-binding regions. Likewise, E24A and R91A each showed a moderate defect, consistent with the contributions of those residues to salt bridges with R64 and E66 (and homologous residues) of EPYC1, respectively. Additionally, M87D and V94D, mutations which convert hydrophobic residues to bulky charged residues, each had a severe impact on interaction, as expected from the participation of those residues in the hydrophobic interface. Combinations of these mutations abolished the interactions completely (Extended Data Fig. 9).

To evaluate the importance of the binding interface *in vivo*, we generated Chlamydomonas strains with point mutations in the binding interface. Rubisco small subunit mutations D23A/E24A or M87D/V94D caused a growth defect under conditions requiring a functional pyrenoid (Fig. 5b, Extended Data Fig. 10a-b). Furthermore, the mutants lacked a visible pyrenoid matrix (Fig. 5c, d and Extended Data Fig. 10c), indicating that those Rubisco small subunit residues are required for matrix formation *in vivo*. The Rubisco mutants retained pyrenoid tubules[16], as previously observed in other matrix-deficient mutants[10,17-19].

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

A structural model for pyrenoid matrix formation

The presence of multiple Rubisco-binding regions along the EPYC1 sequence supports a model where consecutive Rubisco-binding regions on the same EPYC1 polypeptide can bind to different Rubisco holoenzymes and thus hold them together to form the pyrenoid matrix. If this model is correct, we would expect that the ~40 amino acid "linker" regions between consecutive Rubisco-binding regions on EPYC1 (Fig. 1g, i) would be sufficient to span the distance between EPYC1-binding sites on neighboring Rubisco holoenzymes in the pyrenoid matrix. To test this

aspect of the model, we combined our atomic structure of the EPYC1-Rubisco interaction with the precise positions and orientations of Rubisco holoenzymes within the pyrenoid matrix of native cells that we had previously obtained by *in-situ* cryo-electron tomography[5] (Fig. 6a, b). We mapped the positions of EPYC1 binding sites onto Rubisco holoenzymes in the matrix and measured the distances between nearest neighbor EPYC1 binding sites on adjacent holoenzymes (Fig. 6c, binding sites on the same holoenzyme were excluded in this analysis). The observed distances ranged from ~2 nm to ~7 nm, with a median distance of ~4 nm (Fig. 6d and Supplementary Table 5).

A "linker" region of 40 amino acids is unlikely to be stretched to its maximum possible length of 14 nm *in vivo* due to the high entropic cost of this configuration. To determine whether a linker region can span the observed distances between nearest neighbor binding sites on adjacent Rubisco holoenzymes, we used a simple physics model to calculate the energy required to stretch a 40 amino acid chain to any given distance (Fig. 6d; see Methods). The model indicates that stretching the chain to ~7 nm requires an energy of 3 k_BT (where k_B is the Boltzmann constant and T is the temperature), which could reasonably be derived from thermal fluctuations. Thus, our data suggest that the linker region between consecutive Rubisco-binding sites on the EPYC1 polypeptide can span the distance between adjacent Rubisco holoenzymes to hold the pyrenoid matrix together *in vivo*. It also appears likely that, in addition to bridging neighboring Rubisco holoenzymes, consecutive Rubisco-binding regions on a given EPYC1 can bind multiple sites on one Rubisco holoenzyme, as the distance between the nearest binding sites on the same holoenzyme is < 9 nm.

Discussion

In this study, we have determined the structural basis for pyrenoid matrix formation for the first time in any species. We found that in the model alga Chlamydomonas, the intrinsically disordered protein EPYC1 has five regions of similar sequence that can bind to Rubisco as short peptides. These EPYC1 regions form an α-helix that binds to the Rubisco small subunit α-helices via salt bridges and hydrophobic interactions. EPYC1's Rubisco-binding regions are spaced by linker sequences that are sufficiently long to span the distance between binding sites on adjacent

Rubisco holoenzymes within the pyrenoid, allowing EPYC1 to serve as a molecular glue that clusters Rubisco together to form the pyrenoid matrix (Fig. 6e).

The multivalency of EPYC1 and the high K_D (~3 mM; Extended Data Fig. 1e) of individual Rubisco-binding regions are consistent with the emerging principle that cellular phase separation is mediated by weak multivalent interactions[20]. The high dissociation rate constant (>1/s; Extended Data Fig. 1d) of individual Rubisco-binding regions explains how the pyrenoid matrix can mix internally on the time scale of seconds[5] despite the multivalency of EPYC1. The even spacing of the five Rubisco-binding regions across EPYC1 is noteworthy and may be an indication of selective pressure for an optimal distance between binding regions, and thus of an optimal spacing between Rubisco holoenzymes in the matrix.

Knowledge of the EPYC1-Rubisco binding mechanism now opens doors to the molecular characterization of the regulation of this interaction, which may govern the dissolution and condensation of the matrix during cell division[5] and in response to environmental factors[21]. For example, phosphorylation of EPYC1[22] may provide a mechanism to rapidly change the binding affinity of EPYC1 to Rubisco. Inactivation of one Rubisco-binding region on EPYC1 would yield four binding regions, which would allow two such EPYC1 molecules (each containing four Rubisco-binding regions, for a total of eight Rubisco-binding regions) to form a mutually satisfied complex with each Rubisco holoenzyme (containing eight EPYC1-binding sites), a configuration that is predicted to favor dissolution of the matrix[5].

Our structures explain how other proteins bind to Rubisco

In a parallel study (Meyer et al., please see unpublished manuscript provided as reference material), we recently discovered that a common sequence motif is present on many pyrenoid-localized proteins. The motif binds Rubisco, enabling recruitment of motif-containing proteins to the pyrenoid and mediating adhesion between the matrix, pyrenoid tubules, and starch sheath. This motif, [D/N]W[R/K]XX[L/I/V/A], is present in EPYC1's Rubisco-binding regions as defined in the present study, and the motif residues mediate key binding interactions with Rubisco. In our structures, the R/K of the motif is represented by R64 and K127 of EPYC1, each of which forms a salt bridge with E24 of the Rubisco small subunit. The XX of the motif almost always includes a D or E; in our structures this feature is represented by E66 and E129 of EPYC1, each of which forms a salt bridge with R91 of the Rubisco small subunit. Finally, the W

and [L/I/V/A] of the motif are represented by W63/W126 and L67/L130 of EPYC1, which contribute to the hydrophobic interactions with M87, L90 and V94 of the Rubisco small subunit. The key roles of the motif residues in the interface presented here strongly suggest that the structures we have obtained for motifs from EPYC1 also explain where and how all other variants of the motif, including those found on the key pyrenoid proteins SAGA1, SAGA2, RBMP1, RBMP2 and CSP41A, bind to Rubisco. Our observation that the Rubisco small subunit D23A/E24A and M87D/V94D mutants exhibit a more severe disruption of the pyrenoid than the *epyc1* mutant[10] supports the idea that this region of Rubisco interacts not only with EPYC1, but also with other proteins required for pyrenoid biogenesis, making this binding interaction a central hub of pyrenoid biogenesis.

Structural differences between the pyrenoid and carboxysomes

Although α- and β-carboxysomes are morphologically, functionally and evolutionarily distinct from the pyrenoid, their Rubisco is also thought to be clustered by linker proteins. Like EPYC1, the α-carboxysome linker protein CsoS2[23] is intrinsically disordered and is proposed to bind Rubisco as an unfolded peptide[24]. In contrast, the β-carboxysome linker protein CcmM has been proposed to bind Rubisco using folded globular domains[25,26]. The use of an unfolded peptide as in the case of EPYC1 and CsoS2 may provide the benefit of requiring fewer amino acids for achieving the desired binding function. A notable difference is the location of the binding site on Rubisco: whereas both carboxysomal linker proteins bind to the interface between two Rubisco large subunits[24,26], EPYC1 binds to the Rubisco small subunit. It remains to be seen whether this difference in binding site has functional consequences, such as impacts on the three-dimensional packing of Rubisco.

Our findings advance the ability to engineer a pyrenoid

There is currently significant interest in engineering Rubisco condensates into monocotyledonous crops such as wheat and rice to enhance yields[27-30]. Binding of EPYC1 to the Rubisco small subunit presents a promising route for engineering a Rubisco condensate, as the Rubisco small subunit is encoded in the nuclear genome, making it more easily amenable to genetic modification in those crops than the chloroplast-encoded Rubisco large subunit[31]. Knowledge of the binding mechanism now allows engineering of minimal sequence changes into

native crop Rubiscos to enable binding to EPYC1 and to other key proteins required to reconstitute a functional pyrenoid.

Insights into pyrenoid matrix formation in other species

Pyrenoids appear to have evolved independently in different lineages through convergent evolution[7,32]. EPYC1, its Rubisco-binding sequences, and the amino acid residues that form the EPYC1 binding site on the surface of Rubisco are conserved across the order Volvocales, as evidenced from the genome sequences of *Tetrabaena socialis*, *Gonium pectorale* and *Volvox carteri* (Supplementary Table 2). While the molecular mechanisms of matrix formation in other lineages remain to be uncovered, candidate linker proteins have been identified based on similarity of sequence properties to EPYC1[10]. We hypothesize that the matrix in other lineages is formed based on similar principles to those we observed in Chlamydomonas. Our experimental approach for characterizing the binding interaction provides a roadmap for future structural studies of pyrenoids across the tree of life.

We provide a structural view of a phase-separated organelle

The pyrenoid matrix presents an unusual opportunity to study a two-component molecular condensate where one of the components, Rubisco, is large and rigid, and the other component, EPYC1, is a simple intrinsically disordered protein consisting of nearly identical tandem repeats. The rigidity and size of Rubisco holoenzymes previously enabled the determination of their positions and orientations within the pyrenoid matrix of native cells by cryo-electron tomography[5]. The identification of EPYC1 binding sites on Rubisco in the present work and the modeling of linker regions between EPYC1's Rubisco binding regions now make the Chlamydomonas pyrenoid matrix one of the most structurally well-defined phase-separated organelles. Thus, beyond advancing our structural understanding of pyrenoids, organelles that play a central role in the global carbon cycle, we hope that the findings presented here will also more broadly enable advances in the biophysical understanding of phase-separated organelles.

Methods

Strains and culture conditions

Chlamydomonas wild-type (WT) strain cMJ030 was maintained in the dark or low light (~10 μmol photons m^{-2} s^{-1}) on 1.5% agar plates containing Tris-Acetate-Phosphate medium with revised trace elements[33]. For Rubisco extraction, 500 mL Tris-Acetate-Phosphate medium in a 1 L flask was inoculated with a loopful of cells and the culture was grown to 4 x 10^6 cells/mL at 22°C, shaking at 200 rpm under ~100 μmol photons m^{-2} s^{-1} white light in 3% CO_2. Chlamydomonas mutant T60-3[34] (Δ*rbcs*; containing a deletion of both *RBCS* genes) was used for generating Rubisco small subunit point mutants and a wild-type control in the same background. This strain was maintained on agar in the dark or low light (~10 μmol photons $m^{-2}s^{-1}$).

Protein extraction

Rubisco was purified from Chlamydomonas strain cMJ030[35]. Cells were disrupted by ultrasonication in lysis buffer (10 mM $MgCl_2$, 50 mM Bicine, 10 mM $NaHCO_3$, 1 mM dithiothreitol, pH 8.0) supplemented with Halt Protease Inhibitor Cocktail, EDTA-Free (Thermo Fisher Scientific). The soluble lysate was fractionated by ultracentrifugation on a 10-30% sucrose gradient in a SW 41 Ti rotor at a speed of 35,000 rpm for 20 hours at 4˚C. Rubisco-containing fractions were applied to an anion exchange column (MONO Q 5/50 GL, GE Healthcare) and eluted with a linear salt gradient from 30 to 500 mM NaCl in lysis buffer.

Peptide arrays

Peptide arrays were purchased from the MIT Biopolymers Laboratory (Cambridge, MA). The tiling array was composed of 18-amino-acid peptides that tiled across the full-length EPYC1 sequence with a step size of one amino acid. Each peptide was represented by at least two spots on the array, and these replicates were averaged during data analysis. The locations of peptides on the array were randomized. In the substitution arrays, peptides were designed to represent every possible one-amino-acid mutation of the indicated region on EPYC1 by substitution with one of the other 19 amino acids. The arrays were activated by methanol, then washed 3x10 min in binding buffer (50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2.4H_2O$, 1 mM $CaCl_2$ and 200 mM sorbitol, pH 6.8). The arrays were then incubated at 4˚C with 1 mg purified Rubisco

overnight. The arrays were washed in binding buffer to remove any unbound Rubisco. Using a semi-dry transfer apparatus (BIO-RAD), bound Rubisco was transferred onto an Immobilon-P PVDF membrane (Millipore Sigma). The Rubisco was detected by one of two methods: western blotting, or fluorescent labeling. While replicates using the two methods gave similar results, toward the end of the project we found that fluorescent labeling had a lower background, so we show fluorescent labeling data in Figure 1; all other data were obtained by western blotting and chemiluminescence. For fluorescent labeling, Rubisco was labeled with Alexa Fluor™ 680 dye (Thermo Fisher Scientific) and detected by Typhoon Scanner (GE Healthcare). For western blotting, Rubisco was immuno-detected with a polyclonal primary antibody raised against Rubisco[15] (1:10,000) followed by a HRP conjugated goat anti-rabbit (1:20,000; Invitrogen), and the chemiluminescence was detected by ImageQuant LAS 4000 (GE Healthcare). Images were analyzed with ImageQuant TL (GE Healthcare). Arrays were stripped with Restore™ Western Blot Stripping Buffer before re-use (Thermo Fisher Scientific).

For both types of arrays, values for identical sequences present multiple times were averaged. For tiling arrays, the average value for each sequence was plotted at its position in EPYC1 (or at multiple positions for sequences present multiple times). For amino acid substitution arrays, the ratio of each substitution sequence to the corresponding wildtype sequence was calculated and arrayed by amino acid and position.

Surface plasmon resonance (SPR) experiments

All the surface preparation experiments were performed at 25˚C using a Biacore 3000 instrument (GE Healthcare). Purified Rubisco was immobilized on CM5 sensor chips using a Biacore Amine Coupling Kit according to the manufacturer's instructions. Briefly, the chip surface was activated by an injection of 1:1 N-hydroxysuccinimide (NHS)/1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC). Rubisco was diluted to ~100 µg/mL in 10 mM acetate (pH 4.5; this pH had been previously optimized using the immobilization pH scouting wizard) and was injected over the chip surface. Excess free amine groups were then capped with an injection of 1 M ethanolamine. Typical immobilization levels were 8,000 to 10,000 resonance units (RU), as recommended for binding experiments of small molecules. For kinetic experiments (for determining the binding affinities), the typical immobilization levels were ~5,000 RU. The control surfaces were prepared in exactly the same manner as the

experimental surfaces except that no Rubisco was injected. For immobilizations, the running buffer was the Biacore HBS-EP Buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20).

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

For determining the K_D of EPYC1 peptide, the kinetic assays were performed with a running buffer consisting of 200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2•4H_2O$ and 1 mM $CaCl_2$ at pH 6.8 (the same buffer as the peptide array assay). The EPYC1 peptide was dissolved in the same running buffer as the assay and the serial dilutions were also made in the same buffer. The analytes were injected over the control surface and experimental surfaces at a flow rate of 15 μL/min for 2 minutes, followed by 10 minutes with the running buffer alone to allow for dissociation. The surfaces were then regenerated by the running buffer at a flow rate of 30 μL/min for 10 minutes. In all cases, binding to the blank chip was negligible. The fitting and modeling were performed with the BIAevaluation software.

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

Rubisco and $EPYC1_{49-72}$ peptides with the final concentrations of 1.69 mg/ml (=3.02 μM) and 7.5 mM, respectively, were incubated together on ice for 20 minutes in buffer consisting of 200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2•4H_2O$ and 1 mM $CaCl_2$ at pH 6.8 (the same buffer as the peptide array assay and the SPR binding assay). Rubisco and $EPYC1_{106-135}$ peptides with the final concentrations of 1.75 mg/ml (=3.13 μM) and 10 mM, respectively, were incubated together on ice for 20 minutes in the same buffer described above. For apo Rubisco and Rubisco incubated with peptides, similar cryo grid-making procedures were used. 400-mesh Quantifoil 1.2/1.3 Cu grids (Quantifoil, Großlöbichau, Germany) were made

hydrophilic by glow discharging for 60 seconds with a current of 15 mA in a PELCO easiGlow system. Cryo grids were produced using a FEI Mark IV Vitrobot (FEI company, part of Thermo Fisher Scientific, Hillsboro, OR). The chamber of the Vitrobot was kept at 4˚C and 100% relative humidity. 3 µl of sample was applied to the glow-discharged grid and blotted with filter paper for 3 seconds with the equipment-specific blotting force set at 3. After blotting, the grid was rapidly plunge-frozen into a liquid ethane bath.

For apo Rubisco and Rubisco incubated with $EPYC1_{49-72}$ peptide, cryo grids were loaded into a 300 kV FEI Titan Krios cryo electron microscope (FEI Company) at HHMI Janelia Research Campus, Janelia Krios2, equipped with a Gatan K2 Summit camera. After initial screening and evaluation, fully automated data collection was carried out using SerialEM. The final exposure from each collection target was collected as a movie utilizing dose fractionation on the K2 Summit camera operated in super-resolution mode. The movies were collected at a calibrated magnification of 38,168x, corresponding to 1.31 Å per physical pixel in the image (0.655 Å per super-resolution pixel). The dose rate on the specimen was set to be 5.82 electrons per $Å^2$ per second and total exposure time was 10 s, resulting in a total dose of 58.2 electrons per $Å^2$. With dose fractionation set at 0.2 s per frame, each movie series contained 50 frames and each frame received a dose of 1.16 electrons per $Å^2$. The spherical aberration constant of the objective lens is 2.7 mm and an objective aperture of 100 µm was used. The nominal defocus range for the automated data collection was set to be between -1.5 µm and -3.0 µm. For Rubisco incubated with $EPYC1_{106-135}$ peptide, the final exposure was collected on Janelia Krios1 equipped with a C_S-corrector, a Gatan Bioquantum energy filter and a post-filter K3 camera. The movies were collected at a nominal magnification of 81,000x, corresponding to 0.844 Å per physical pixel in the image (0.422 Å per super-resolution pixel). The dose rate on the specimen was set to be 12 electrons per pixel per second and total exposure time was 3.56 s, resulting in a total dose of 60 electrons per $Å^2$. Each movie series contained 60 frames and each frame received a dose of 1 electron per $Å^2$. The nominal defocus range for the automated data collection was set to be between -1 µm and -1.6 µm.

The movies were 2x binned and motion corrected using MotionCor2[36] and CTF was estimated using CTFFIND[37] in Relion[38]. 1,809,869 $EPYC1_{49-72}$ peptide-bound Rubisco particles, 2,257,131 $EPYC1_{106-135}$ peptide-bound Rubisco particles, and 677,071 Rubisco particles in the apo state were selected using cisTEM[39]. 2D classification was performed using cisTEM. The

448 classes presenting detailed features in class averages were chosen for 3D classification on
449 cryoSPARC[40,41] and on Relion. The 3D class showing clear secondary structures was chosen for
450 3D auto-refine first without symmetry and then with D4 symmetry imposed. After CTF
451 refinement and Bayesian polishing in Relion, the reconstructed map resolution is 2.68 Å for the
452 apo state, 2.62 Å for the EPYC1$_{49\text{-}72}$ peptide-bound state, and 2.06 Å for the EPYC1$_{106\text{-}135}$
453 peptide-bound state. The EPYC1$_{49\text{-}72}$ peptide-bound particles at super-resolution pixel size were
454 further subjected to CTF refinement and polishing, resulting in map at 2.13 Å resolution. Details
455 for single-particle cryo-EM data collection and image processing are included in the
456 Supplementary Table 1.
457
458 **Single-particle cryo-electron microscopy model building, fitting, and refinement**
459 A full model for Rubisco from Chlamydomonas was produced from an X-ray structure[13] (PDB
460 entry 1GK8) and used for rigid body fitting into a local resolution filtered apo or EPYC1$_{49\text{-}72}$
461 peptide-bound Rubisco cryo-EM map using UCSF Chimera[42]. After rigid body fitting of the full
462 complex, initial flexible fitting was performed in COOT[43] by manually going through the entire
463 peptide chain of a single large and small Rubisco subunit before applying the changes to the
464 other seven large and small subunits. The C-terminal part of the small subunit was built manually
465 and the sequences updated to the RBCS2 sequences. The sequence of the EPYC1$_{49\text{-}72}$ peptide
466 was used to predict secondary structure elements using JPred4[44], which gave the prediction that
467 the C-terminal region (NWRQELES) is α-helical. Guided by this prediction, the peptide was
468 built manually into the density using COOT. Additional maps like the initial 2.62 Å from the
469 binned data and maps filtered to different resolutions with various applied B-factors were also
470 used to help with model building in unclear regions. Additional real space refinement of the
471 entire complex was performed using Phenix[45]. The EPYC1$_{106\text{-}135}$ peptide bound map was used to
472 build a model of the EPYC1$_{106\text{-}135}$ peptide. First rigid body fitting of the EPYC1$_{49\text{-}72}$ peptide-
473 bound Rubisco model into a local resolution filtered EPYC1$_{106\text{-}135}$ peptide-bound Rubisco cryo-
474 EM map was performed using UCSF Chimera. Then the sequence of the peptide was mutated to
475 the EPYC1$_{106\text{-}135}$ peptide sequence, followed by flexible fitting to slightly adjust the PDB to the
476 density. Models were subjected to an all-atom structure validation using MolProbity[46]. Figures
477 were produced using UCSF Chimera.
478

Liquid–liquid phase separation assay

Proteins used in the liquid–liquid phase separation assay were obtained and stored essentially as described previously[11]. Briefly, Rubisco was purified from *Chlamydomonas reinhardtii* cells (CC-2677 cw15 nit1-305 mt-5D, Chlamydomonas Resource Center), grown in Sueoka's high-salt medium[47], using a combination of anion exchange chromatography and gel filtration.

The EPYC1 full-length gene (encoding amino acids 1-317) and corresponding R/K mutant (EPYC1$^{R64A/K127A/K187A/K248A/R314A}$) were synthesized by GenScript and cloned between the SacII and HindIII site of the pHue vector[48]. Proteins were produced in the *E. coli* strain BL21 (DE3) harbouring pBADESL[49] for co-expression of the *E. coli* chaperonin GroEL/S. The purification was conducted with minor changes (dialysis for removal of high immidazol concentrations was skipped by running the gel-filtration column before the second IMAC). After the first IMAC step and cleavage[50] of the N-terminal His_6–ubiquitin tag, proteins were separated by gel filtration. Finally, the peak fraction was passed a second time through an IMAC column, collecting EPYC1 from the flow through.

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

Yeast two-hybrid assay

Yeast two-hybrid to detect interactions between EPYC1 and RbcS1 was carried out as described previously[12]. EPYC1 was cloned into the two-hybrid vector pGBKT7 to create a fusion with the GAL4 DNA binding domain. Point mutations were introduced by PCR into RbcS1, which was then cloned in the pGADT7 to create a fusion with the GAL4 activation domain. Yeast cells were then co-transformed with binding and activation domain vectors. Successful transformants were cultured, diluted to an optical density at 600 nm (OD600) of 0.5 or 0.1, and plated onto SD-

L-W and SD-L-W-H containing increasing concentrations of the HIS3 inhibitor triaminotriazole (3-AT). Plates were imaged after 3 days. Spots shown in Fig. 5a were grown at 5 mM 3-AT from a starting OD600 of 0.5; they are a subset of the full dataset shown in Extended Data Fig. 6.

Cloning of Rubisco small subunit point mutants

The plasmid pSS1-ITP[51] which contains Chlamydomonas *RBCS1* including UTRs and introns 1 and 2 was used as a starting point for generating plasmids pSH001 and pSH002, which encode RBCS1$^{D23A/E24A}$, and RBCS1$^{M87D/V94D}$, respectively. The point mutations were generated by Gibson assembly[52] of gBlocks (synthesized by Integrated DNA Technologies) containing the desired mutations into pSS-ITP that had been enzyme digested by restriction endonucleases (XcmI and BbvCI for the D23A/E24A mutations and BbvCI and BlpI for the M87D/V94D mutations). All constructs were verified by Sanger sequencing.

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

GCAGGGCTGCCCCGGCTCAGGCCAACCAGATGATGGTCTGGACCCCGGTCAACAAC
AAGATGTTCGAGACCTTCTCCTACCTGCCTCCTCTGACCGCCGCGCAGATCGCCGCC
CAGGTCGACTACATCGTCGCCAACGGCTGGATCCCCTGCCTGGAGTTCGCTGAGGCC
GACAAGGCCTACGTGTCCAAC

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

CTGCCTGGAGTTCGCTGAGGCCGACAAGGCCTACGTGTCCAACGAGTCGGCCATCC
GCTTCGGCAGCGTGTCTTGCCTGTACTACGACAACCGCTACTGGACCATGTGGAAGC
TGCCCATGTTCGGCTGCCGCGACCCCGACCAGGTGCTGCGCGAGATCGACGCCTGCA
CCAAGGCCTTCCCCGATGCCTACGTGCGCCTGGTGGCCTTCGACAACCAGAAGCAG
GTGCAGATCATGGGCTTCCTGGTCCAGCGCCCCAAGACTGCCCGCGACTTCCAGCCC
GCCAACAAGCGCTCCGTGTAAATGGAGGCGCTCGTCGATCTGAGCCGTGTGTGATGT
TTGTTGGTGTTTGAGCGAGTGCAATGAGAGTGTGTGTGTGTGTTGTTGGTGTGTG
GCTAAGCCAAGCGTGATCGC

Both the plasmids pSH001 and pSH002 have been submitted to the Chlamydomonas Resource Center (www.chlamycollection.org).

Transformation of Chlamydomonas to make the Rubisco small subunit point mutants

Chlamydomonas strains *Δrbcs*;$RBCS^{WT}$, *Δrbcs*;$RBCS^{D23A/E24A}$, and *Δrbcs*;$RBCS^{M87D/V94D}$ (The accession numbers of these strains in Chlamydomonas Resource Center are CC-5616, CC-5617 and CC-5618, respectively.) were generated by transforming pSS1-ITP, pSH001, and pSH002 (encoding Rubisco small subunit constructs) into the Rubisco small subunit deletion mutant T60 (*Δrbcs*) by electroporation as described previously[53]. For each transformation, 29 ng kbp^{-1} of KpnI linearized plasmid was mixed with 250 µL of 2 x 10^8 cells mL^{-1} at 16°C and electroporated immediately. Transformant colonies were selected on Tris-Phosphate plates without antibiotics at 3% v/v CO_2 under ~50 µmol photons m^{-2} s^{-1} light. The sequence of RbcS in the transformants was verified by PCR amplification and Sanger sequencing.

Spot tests

Δrbcs;$RBCS^{WT}$, *Δrbcs*;$RBCS^{D23A/E24A}$, and *Δrbcs*;$RBCS^{M87D/V94D}$ (The accession numbers of these strains in Chlamydomonas Resource Center are CC-5616, CC-5617 and CC-5618, respectively.) were grown in Tris-Phosphate medium at 3% CO_2 until ~2x10^6 cells mL^{-1}. Cells were diluted in Tris-Phosphate medium to a concentration of 8.7 x10^7 cells mL^{-1}, then serially diluted 1:10 three times. 7.5 µL of each dilution was spotted onto four TP plates and incubated in air or 3% CO_2 under 20 or 100 µmol photons m^{-2} s^{-1} white light for 9 days before imaging.

Transmission electron microscopy

Samples for electron microscopy were fixed for 1 hour at room temperature in 2.5% glutaraldehyde in Tris-Phosphate medium (pH 7.4), followed by 1 hour at room temperature in 1% OsO_4, 1.5% $K_3Fe(CN)_3$, and 2 mM $CaCl_2$. Fixed cells were then bulk stained for 1 hour in 2% uranyl acetate, 0.05 M maleate buffer at pH 5.5. After serial dehydration (50%, 75%, 95%, and 100% ethanol, followed by 100% acetonitrile), samples were embedded in epoxy resin containing 34% Quetol 651, 44% nonenyl succinic anhydride, 20% methyl-5- norbornene-2,3-dicarboxylic anhydride, and 2% catalyst dimethylbenzylamine. Ultramicrotomy was done by the Core Imaging Lab, Medical School, Rutgers University. Imaging was performed at the Imaging and Analysis Center, Princeton University, on a CM100 transmission electron microscope (Philips, Netherlands) at 80 kV.

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

For detailed descriptions of the Chlamydomonas cell culture, vitrification of cells onto EM grids, thinning of cells by cryo-focused ion beam milling, 3D imaging of native pyrenoids by cryo-electron tomography, tomographic reconstruction, template matching, and subtomogram averaging, see our previous study[5]. In that study, we measured the distances between the center positions of Rubisco complexes within tomograms of five pyrenoids. The spatial parameters determined in that study were combined with the EPYC1-binding sites resolved here by cryo-EM single-particle analysis to measure the nearest-neighbor distances between EPCY1 binding sites on adjacent Rubisco complexes within the native pyrenoid matrix.

The *in situ* subtomogram average EMD-3694[5] was used as the reference for the Rubisco model. We extracted the isosurface from this density using the 0.5 contour level recommended in the Electron Microscopy Data Bank entry. We then fit the atomic model of EPYC1-bound Rubisco (Fig. 2) within the EMD-3694 density, and for each EPYC1 binding site, we marked the closest point on the isosurface to define the EPYC1 binding sites on this model. The positions and orientations previously determined by subtomogram averaging were used to place each Rubisco model and its corresponding binding sites into the pyrenoid tomograms using the PySeg program[54].

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

Regarding the accuracy of the Rubisco localization in the previous study[5], we used template matching, subtomogram alignment, and hierarchical classification to identify 97.5% of the Rubisco complexes in each of the five pyrenoid volumes. The Rubisco average was determined at a resolution of 16.5 Å. This could be interpreted to mean that the Rubisco complexes were localized with 1.65 nm precision. However, this resolution is not a simple

603 reflection of translational and rotational accuracy, but is also greatly limited by the contrast
604 transfer function and pixel size of the tomographic data. Furthermore, with this localization
605 uncertainty being random for each particle, it would not change the ~4 nm peak distance between
606 neighboring EPYC1 binding sites measured in our current study.
607
608 **Modeling of the energy required to stretch EPYC1-linker regions**
609 The energy required to stretch the linker regions between EPYC1's Rubisco-binding regions was
610 determined as follows. The force F required to stretch a 40 amino acid linker region to any given
611 length z was approximated using a wormlike chain model[55]:

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

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

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

616 This energy was calculated and plotted in Fig. 6d.
617

Data availability

All data generated or analyzed during this study are included in this Article, the Extended Data and the Supplementary tables. The single-particle cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMDB-22401, EMDB-22308 and EMDB-22462. The atomic models have been deposited in the Protein Data Bank under accession codes PDB 7JN4, PDB 7JFO and PDB 7JSX. The raw datasets have been deposited in the Electron Microscopy Public Image Archive with accession codes EMPIAR-10503, EMPIAR-10502 and EMPIAR-10501. The subtomogram average of Rubisco has been deposited in the Electron Microscopy Data Bank with accession code EMD-3694.

627
628 ## References

629 1 Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of
630 the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237-240
631 (1998).

632 2 Hessler, A. M., Lowe, D. R., Jones, R. L. & Bird, D. K. A lower limit for atmospheric
633 carbon dioxide levels 3.2 billion years ago. *Nature* **428**, 736-738 (2004).

634 3 Ainsworth, E. A. & Long, S. P. What have we learned from 15 years of free-air CO2
635 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy
636 properties and plant production to rising CO2. *New Phytol* **165**, 351-371,
637 doi:10.1111/j.1469-8137.2004.01224.x (2005).

638 4 Raven, J. A., Cockell, C. S. & De La Rocha, C. L. The evolution of inorganic carbon
639 concentrating mechanisms in photosynthesis. *Philosophical transactions of the Royal*
640 *Society of London. Series B, Biological sciences* **363**, 2641-2650,
641 doi:10.1098/rstb.2008.0020 (2008).

642 5 Freeman Rosenzweig, E. S. *et al.* The Eukaryotic CO2-Concentrating Organelle Is
643 Liquid-like and Exhibits Dynamic Reorganization. *Cell* **171**, 148-162 e119,
644 doi:10.1016/j.cell.2017.08.008 (2017).

645 6 Badger, M. R. *et al.* The diversity and coevolution of Rubisco, plastids, pyrenoids, and
646 chloroplast-based CO2-concentrating mechanisms in algae. *Can J Bot* **76**, 1052–1071,
647 doi:10.1139/b98-074 (1998).

648 7 Villarreal, J. C. & Renner, S. S. Hornwort pyrenoids, carbon-concentrating structures,
649 evolved and were lost at least five times during the last 100 million years. *Proc Natl Acad*
650 *Sci U S A* **109**, 18873-18878, doi:10.1073/pnas.1213498109 (2012).

651 8 Wang, Y., Stessman, D. J. & Spalding, M. H. The CO2 concentrating mechanism and
652 photosynthetic carbon assimilation in limiting CO2 : how Chlamydomonas works against
653 the gradient. *The Plant journal : for cell and molecular biology* **82**, 429-448,
654 doi:10.1111/tpj.12829 (2015).

655 9 Raven, J. A. CO2-concentrating mechanisms: A direct role for thylakoid lumen
656 acidification? *Plant Cell Environ* **20**, 147-154, doi:Doi 10.1046/J.1365-3040.1997.D01-
657 67.X (1997).

658 10 Mackinder, L. C. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-
659 concentrating organelle. *Proc Natl Acad Sci U S A* **113**, 5958-5963,
660 doi:10.1073/pnas.1522866113 (2016).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

27 Hennacy, J. H. & Jonikas, M. C. Prospects for Engineering Biophysical CO_2 Concentrating Mechanisms into Land Plants to Enhance Yields. *Annual review of plant biology*, doi:10.1146/annurev-arplant-081519-040100 (2020).

28 Long, B. M. *et al.* Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts. *Nature communications* **9**, 3570, doi:10.1038/s41467-018-06044-0 (2018).

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

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

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

32 Raven, J. A., Beardall, J. & Sanchez-Baracaldo, P. The possible evolution and future of CO_2-concentrating mechanisms. *J Exp Bot* **68**, 3701-3716, doi:10.1093/jxb/erx110 (2017).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Acknowledgements

We thank Jianping Wu, Nieng Yan, Luke Mackinder, Cliff Brangwynne and members of the Jonikas laboratory for helpful discussions; Ned Wingreen, Silvia Ramundo, Jessi Hennacy, Eric Franklin and Alexandra Wilson for constructive feedback on the manuscript; Wolfgang Baumeister and Jürgen Plitzko for providing support and cryo-ET instrumentation; and Miroslava Schaffer for help with acquiring the cryo-ET data, previously published in Freeman Rosenzweig et al., 2017. This 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., Deutsche Forschungsgemeinschaft grant (EN 1194/1-1 as part of FOR2092) to B.D.E., Ministry of

Education (MOE Singapore) Tier 2 grant (MOE2018-T2-2-059) to O.M.-C., UK Biotechnology and Biological Sciences Research Council (BB/S015531/1) and Leverhulme Trust (RPG-2017-402) grants to A.J.M and N.A., NIH grant R01GM071574 to F.M.H., Deutsche Forschungsgemeinschaft fellowship (PO2195/1-1) to S.A.P., and National Institute of General Medical Sciences of the National Institutes of Health (T32GM007276) training grant to V.K.C.. 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

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

Conflict of interest statement

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

Figures

Fig. 1 | EPYC1 consists of five tandem sequence repeats, each of which contains a Rubisco-binding region. a, A representative (N=15) transmission electron microscopy (TEM) image of a Chlamydomonas cell. Scale bar = 1 μm. **b**, Cartoon depicting the chloroplast and pyrenoid in the image shown in panel a. The blue dots indicate the location of Rubisco enzymes clustered in the pyrenoid matrix. **c**, We hypothesized that pyrenoid matrix formation is mediated by multivalent interactions between Rubisco and the intrinsically disordered protein EPYC1. **d**, We designed an array of 18 amino acid peptides tiling across the full length EPYC1 sequence. **e**, Incubation of the array with purified Rubisco allows identification of peptides that bind to Rubisco. **f**, Image of the Rubisco binding signal from the peptide tiling array. **g**, The Rubisco binding signal was quantified and plotted for each peptide as a function of the position of the middle of the peptide along the EPYC1 sequence. The initial 26 amino acids of EPYC1 correspond to a chloroplast targeting peptide (cTP), which is not present in the mature protein[12]. Results are representative of three independent experiments. **h**, The positions of EPYC1's five sequence repeats are shown to scale with panel g. Predicted α-helical regions are shown as wavy lines. **i**, Primary sequence of EPYC1, with the five sequence repeats aligned. In panels h and i, the regions represented by peptides subsequently used for structural studies are underlined with red lines ($EPYC1_{49-72}$) and pink lines ($EPYC1_{106-135}$). $EPYC1_{106-135}$ is an exact match to the underlined sequence of Repeats 2 and 4, and has a one-amino acid difference from the corresponding region in Repeat 3 (dashed underline).

847 **Fig. 2 | EPYC1 binds to Rubisco small subunits. a**, Peptide EPYC1$_{49-72}$, corresponding to the
848 first Rubisco-binding region of EPYC1, was incubated at saturating concentrations with Rubisco
849 prior to single-particle cryo-electron microscopy. **b-e**, Density maps (b, d) and cartoons (c, e)
850 illustrate the side views (b, c) and top views (d, e) of the density map of the EPYC1 peptide-
851 Rubisco complex. Dashed boxes in panel b indicate regions shown in Fig. 3a-f.
852

Fig. 3 | EPYC1 binds to Rubisco small subunit α-helices via salt bridges and a hydrophobic pocket. a-b, Front (a) and side (b) views of the EPYC1$_{49-72}$ peptide (red) bound to the two α-helices of the Rubisco small subunit (blue). **c-d**, Three pairs of residues form salt bridges between the helix of the EPYC1$_{49-72}$ peptide and the helices on the Rubisco small subunit. Shown are front (c) and side (d) views as in panel a and panel b. The distances from EPYC1 R64, R71 and E66 to Rubisco small subunit E24, D23 and R91 are 3.06 □, 2.78 □, and 2.79 □, respectively. **e-f**, A hydrophobic pocket is formed by three residues of the EPYC1$_{49-72}$ peptide and three residues of helix B of the Rubisco small subunit. Shown are front (e) and side (f) views as in panel a and panel b. **g**, Summary of the interactions observed between the EPYC1$_{49-72}$ peptide and the two α-helices of the Rubisco small subunit. Helices are highlighted; the residues mediating interactions are bold; salt bridges are shown as dotted lines; residues contributing to the hydrophobic pocket are shown in black.

Fig. 4 | Interface residues on EPYC1 are required for binding and phase separation of EPYC1 and Rubisco *in vitro*. a, Rubisco binding to a peptide array representing every possible single amino acid substitution for amino acids 56-71 of EPYC1. The binding signal was normalized by the binding signal of the original sequence. **b**, The effect of mutating the central R or K in each of EPYC1's Rubisco-binding regions on *in vitro* phase separation of EPYC1 with Rubisco. Scale bar = 10 μm. For each condition, the experiment was performed twice independently with similar results.

Fig. 5 | Interface residues on Rubisco are required for yeast two-hybrid interactions between EPYC1 and Rubisco, and for pyrenoid matrix formation *in vivo*. a, The importance of Rubisco small subunit residues for interaction with EPYC1 was tested by mutagenesis in a yeast two-hybrid experiment. **b**, The Rubisco small subunit-less mutant T60 (*Δrbcs*) was transformed with wild-type, D23A/E24A or M87D/V94D Rubisco small subunits. Serial 1:10 dilutions of cell cultures were spotted on TP minimal medium and grown in air or 3% CO_2. **c-d**, Representative electron micrographs of whole cells (c) and corresponding pyrenoids (d) of the strains expressing wild-type, D23A/E24A, and M87D/V94D Rubisco small subunit. Dashes in panel c indicate regions shown in panel d. Scale bars = 500 nm. At least 25 cells were imaged for each strain; additional representative images are shown in Extended Data Fig. 10c.

Fig. 6 | A model for matrix structure consistent with *in situ* Rubisco positions and orientations. a, The pyrenoid matrix was imaged by cryo-electron tomography[5]. An individual slice through the three-dimensional volume is shown. Scale bar = 200 nm. **b**, The positions and orientations of individual Rubisco holoenzymes (blue) were determined with high sensitivity and specificity (97.5% positive identification) by template matching, subtomogram averaging, and classification and then mapped back into the tomogram volume shown in panel a. The yellow box indicates the region shown in panel c. Scale bar = 200 nm. **c**, The distances (yellow) between the nearest EPYC1-binding sites (red) on neighboring Rubisco holoenzymes (blue) were measured. The view is from inside the matrix; in some cases the nearest EPYC1 binding site is on a Rubisco that is out of the field of view, causing some yellow lines to appear unconnected in this image. Scale bar = 20 nm. The data shown in panels a-c are representative of the five independent tomograms used for this study. **d**, Histogram showing the distances between the nearest EPYC1 binding sites on neighboring Rubisco holoenzymes. The black line indicates the median, and the yellow shading indicates 95% confidence interval based on data from five independent tomograms. The estimated energy required for stretching a chain of 40 amino acids a given distance is shown in blue. **e**, A 3D model illustrates how EPYC1 (red) could crosslink multiple Rubisco holoenzymes (blue) to form the pyrenoid matrix. The conformations of the intrinsically disordered linkers between EPYC1 binding sites were modeled hypothetically.

900

1. Extended Data

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper.
Extended Data Fig. 1	The EPYC1 peptides with the highest binding affinities to Rubisco were chosen for structural studies.	Extended_Data_Fig. 1-20200921.jpg	a, Diagram indicating the differences between the previously defined sequence repeats[10] and the newly defined sequence repeats on full-length EPYC1. b, To verify the Rubisco-binding regions on EPYC1, surface plasmon resonance (SPR) was used to measure the binding of EPYC1 peptides to Rubisco. Purified Rubisco was immobilized on a sensor surface, and the EPYC1 peptides in solution were injected over the surface. The binding activity was recorded in real time in a sensorgram. c, The peptides used in SPR experiments are shown aligned to the sequence as shown in Fig. 1. The Rubisco-binding signal from the SPR experiment of each peptide is shown after normalization to the peptide's molecular weight. $EPYC1_{49-72}$ (boxed in red) and $EPYC1_{106-135}$ (boxed in pink) were chosen for structural studies based on their reproducible high Rubisco binding signal. d, The Rubisco-binding response of the $EPYC1_{49-72}$ peptide at different concentrations was measured by SPR. e, The binding responses shown in (d) were fitted to estimate the K_D of $EPYC1_{49-72}$ peptide binding to Rubisco.
Extended Data Fig. 2	Single-particle cryo-EM data collection and image processing procedure.	Extended_Data_Fig. 2-20200919.jpg	a-c, Representative micrographs of the apo Rubisco sample (a), the Rubisco-$EPYC1_{49-72}$ complex (b) and the Rubisco-$EPYC1_{106-135}$ complex (c). Scale bars = 100 nm. d-f, Representative 2D class averages of the apo Rubisco sample (d), the Rubisco-$EPYC1_{49-72}$ complexes (e) and the Rubisco-

			EPYC1$_{106-135}$ complexes (f). g-i, Overview of the workflow for single-particle data processing for the apo Rubisco sample (g), the Rubisco-EPYC1$_{49-72}$ sample (h) and the Rubisco-EPYC1$_{106-135}$ sample (i). j-l, Local resolution estimation of the final refined apo Rubisco map (j), the final refined Rubisco-EPYC1$_{49-72}$ complex map (k) and the final refined Rubisco-EPYC1$_{106-135}$ complex map (l).
Extended Data Fig. 3	Cryo-EM analysis and resolution of apo Rubisco and Rubisco-EPYC1 peptide complexes in this study.	Extended_Data_Fig. 3-20200728.jpg	a-b, Representative cryo-EM density quality showing an α-helix of residues 214-232 in chain A (one of the Rubisco large subunits) (a) and a β-sheet of residues 36-43 in chain A (b) of the Rubisco-EPYC1$_{49-72}$ density map and structural model. The densities are shown as meshwork in gray. The backbones of the structural model are in ribbon representation, and side chains are shown in stick representation. c-d, Representative cryo-EM density quality showing water molecules as orange spheres. One water molecule between R312 and E136 on chain A is shown in panel c, and another water molecule between D137 and K316 on chain A is shown in panel d. e, Fourier shell correlation (FSC) curves of the final density maps of apo Rubisco and the Rubisco-EPYC1 peptide complexes.
Extended Data Fig. 4	Comparison of our EM structure of apo Rubisco and the published X-ray crystallography structure (1gk8) of Rubisco purified from *Chlamydomonas reinhardtii*[13], and comparison of our EM structure of apo Rubisco and Rubisco bound with EPYC1$_{49-72}$ peptide.	Extended_Data_Fig. 4-20200527.jpg	a, Comparison of the structure of the small subunit of apo Rubisco obtained here by EM with 1gk8. The EM structure has additional C-terminus density past residue 126, circled by a red dashed line. b, Comparison of our two EM structures of the small subunit: from apo Rubisco and from EPYC1$_{49-72}$ peptide-bound Rubisco. c, Comparison of the structure of the large subunit of apo Rubisco obtained here by EM with 1gk8. The three major differences found between the X-ray structure and the EM structure of the large subunit are circled with red dashed lines. d, Comparison of our two EM structures of the large subunit: from apo Rubisco and from EPYC1$_{49-72}$ peptide-bound Rubisco. The major difference found between the EPYC1$_{49-72}$ peptide-bound structure and the apo EM structure was the loop between K175 and L180 of the large

			subunit, which is shown circled by a red dashed line.
Extended Data Fig. 5	Additional residues may contribute to the interaction between EPYC1 and Rubisco.	Extended_Data_Fig. 5-20200804-4.jpg	Our Rubisco-EPYC1$_{49-72}$ peptide structure suggests that R56 of the EPYC1$_{49-72}$ peptide may interact with D31 of the Rubisco small subunit and E433 of the Rubisco large subunit (the atoms of the backbone of E433 are also shown to display the possible interaction). R51 of the EPYC1$_{49-72}$ peptide may form a salt bridge with Y32 of the Rubisco small subunit. Residues S57 and V58 of the EPYC1$_{49-72}$ peptide are close to D31 in the structure, which may explain why replacing either of these residues with a negatively charged residue disrupts binding (Fig. 4a).
Extended Data Fig. 6	The EPYC1$_{106-135}$ peptide binds to Rubisco small subunit α-helices via salt bridges and a hydrophobic pocket in a similar manner to the EPYC1$_{49-72}$ peptide.	Extended_Data_Fig. 6-20200730.jpg	a, The EPYC1$_{106-135}$ peptide represents the second, third and fourth Rubisco-binding regions of EPYC1 indicated by pink lines and dash line (the peptide is a perfect match to the second and fourth Rubisco-binding regions, and there is a one-amino acid difference between the peptide and the third repeat). b-c, Side view (b) and top view (c) of the density map of the EPYC1$_{106-135}$ peptide-Rubisco complex. Dashes in panel b indicate regions shown in panels d-i. d-e, Front (d) and side (e) views of the EPYC1$_{106-135}$ peptide (red) bound to the two α-helices of the Rubisco small subunit (blue). f-g, Three pairs of residues form salt bridges between the helix of the EPYC1$_{106-135}$ peptide and the helices on the Rubisco small subunit. Shown are front (f) and side (g) views as in panel d and panel e. The distances from EPYC1 K127, R134 and E129 to Rubisco small subunit E24, D23 and R91 are 2.96 Å, 3.17 Å, and 2.68 Å, respectively. h-i, A hydrophobic pocket is formed by three residues of the EPYC1$_{106-135}$ peptide and three residues of helix B of the Rubisco small subunit. Shown are front (h) and side (i) views as in panel d and panel e. j, Summary of the interactions observed between the EPYC1$_{106-135}$ peptide and the two α-helices of the Rubisco small subunit. Helices are highlighted; the residues mediating interactions are bold; salt bridges are shown as dotted lines; residues contributing to the

			hydrophobic pocket are shown in black. k, Color keys used in this figure.
Extended Data Fig. 7	Surface plasmon resonance analysis of binding of point mutants of $EPYC1_{55-72}$ to Rubisco.	Extended_Data_Fig. 7-20200722.jpg	The wild-type (WT) peptide or peptides with the indicated mutations were synthesized, and their Rubisco-binding signal was measured by surface plasmon resonance.
Extended Data Fig. 8	Interface residues on EPYC1 identified by cryo-EM are important for binding and phase separation of EPYC1 and Rubisco.	Extended_Data_Fig. 8-20200722.jpg	a, SDS-PAGE analysis of purified proteins used for *in vitro* phase separation experiments. WT = wild-type EPYC1; R/K = $EPYC1^{R64A/K127A/K187A/K248A/R314}$. b-c, A droplet sedimentation assay was used as a readout of phase separation complementary to the microscopy analyses shown in Fig. 4b. Proteins at indicated concentrations were mixed and incubated for 10 minutes, then condensates were pelleted by centrifugation. Supernatant (S) and pellet (P) fractions were run on a denaturing gel. The negative controls with no Rubisco or with no EPYC1 are shown in (b), and the wild-type Rubisco with wild-type EPYC1 or mutant EPYC1 are shown in (c). Data shown here are representative of two independent replicates.
Extended Data Fig. 9	Yeast two-hybrid assays of interactions between EPYC1 and wild-type or mutated Rubisco small subunit.	Extended_Data_Fig. 9-20200803.jpg	Colonies are shown after 3 days growth on plates. A subset of the data shown in this figure is shown in Fig. 5a.
Extended Data Fig. 10	Selection of the Rubisco small subunit mutant strains for phenotype analysis.	Extended_Data_Fig. 10-20200920.jpg	a, The Rubisco small subunit-less mutant T60 (*Δrbcs*) was transformed with DNA encoding wild-type and mutant Rubisco small subunits (RBCS) to produce candidate transformants with the genotypes *Δrbcs*;$RBCS^{WT}$, *Δrbcs*;$RBCS^{D23A/E24A}$, and *Δrbcs*;$RBCS^{M87D/V94D}$. Total protein extracts for three strains from each transformation were separated on a polyacrylamide gel. b, The gel shown in panel a was probed by Western blot using a polyclonal antibody mixture that detects both large and small Rubisco subunits. The experiments shown in panel a and b were performed once for selecting the candidate transformants with the

			highest RBCS expression level from each genotype, in case any phenotype may be caused by low expression level of Rubisco. Selected strains are indicated by an arrow below the lanes and were used for the subsequent phenotypic analyses shown in Fig. 5 and panel c. c, Additional representative TEM images of whole cells of the strains expressing wild-type, D23A/E24A, and M87D/V94D Rubisco small subunit. Scale bar = 500 nm. For each strain, at least 25 images (one image for one cell) were taken and showing similar results.

2. Supplementary Information:

A. Flat Files

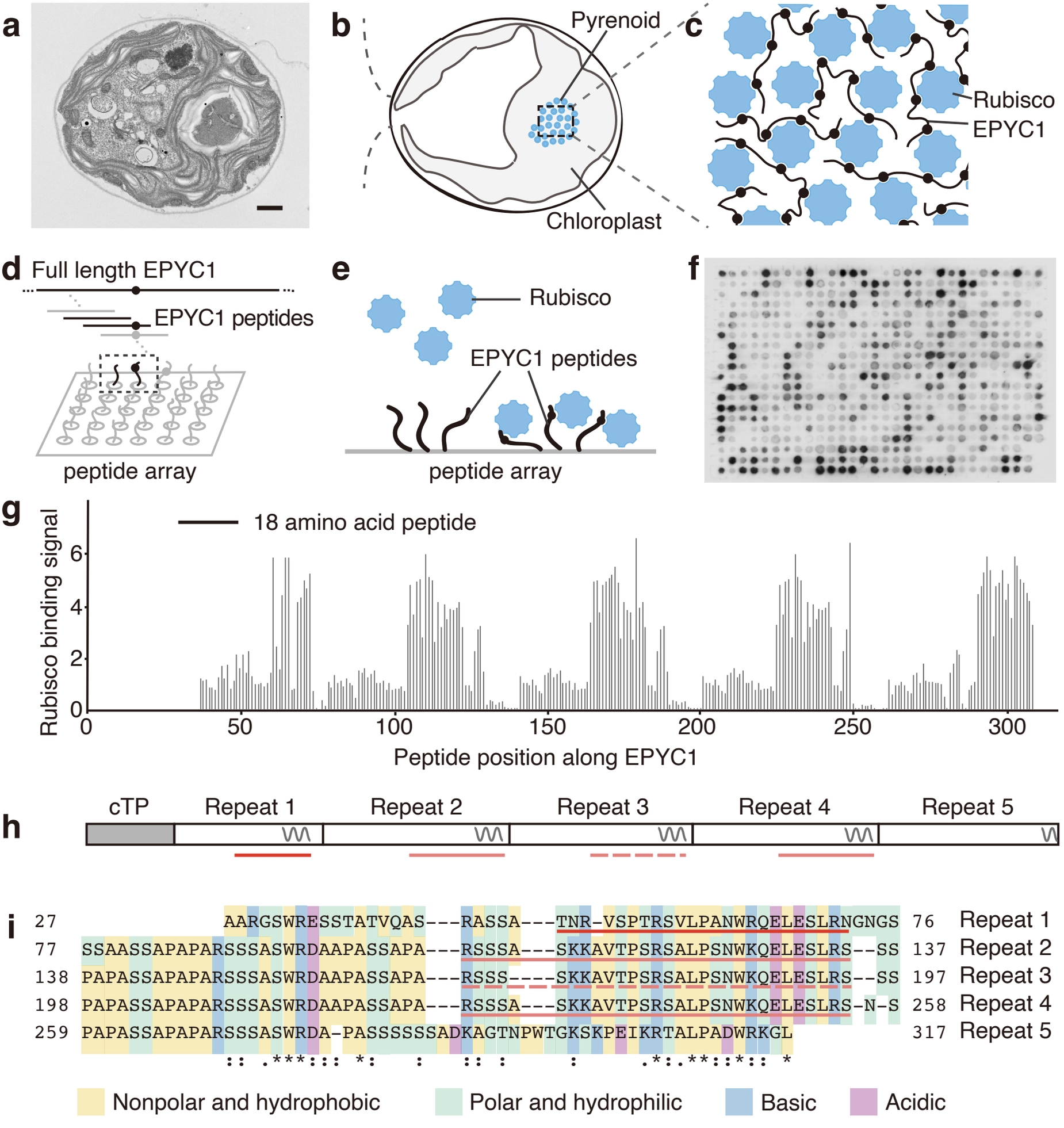
Item	**Present?**	**Filename** **This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf**	**A brief, numerical description of file contents.** **i.e.:** ***Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.***
Supplementary Information	Yes	Supplementary_Table 1-2_20201016.pdf	Supplementary Tables 1-2
Reporting Summary	Yes	200923_nr-reporting-summary_SH.pdf	

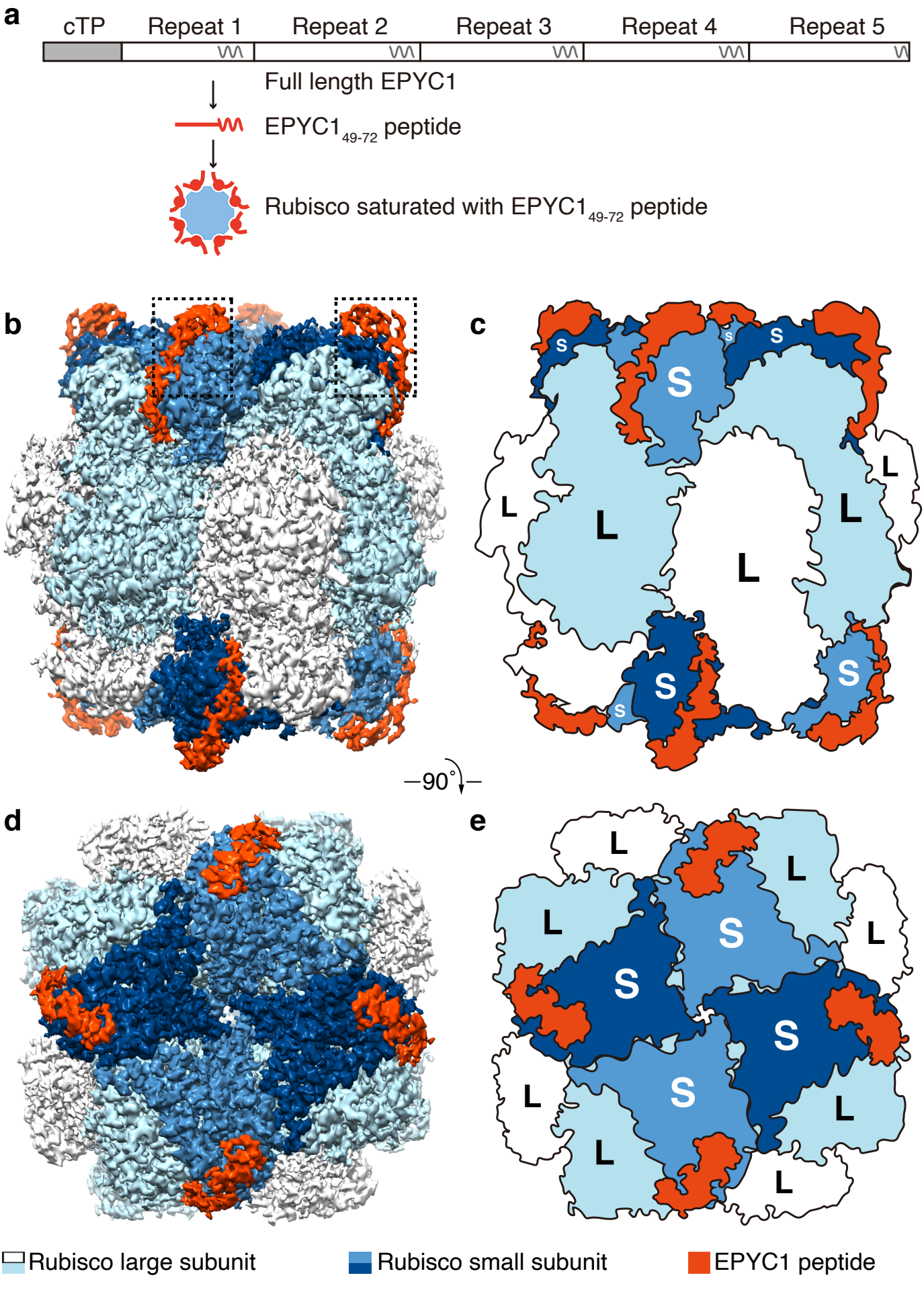
11 ## B. Additional Supplementary Files

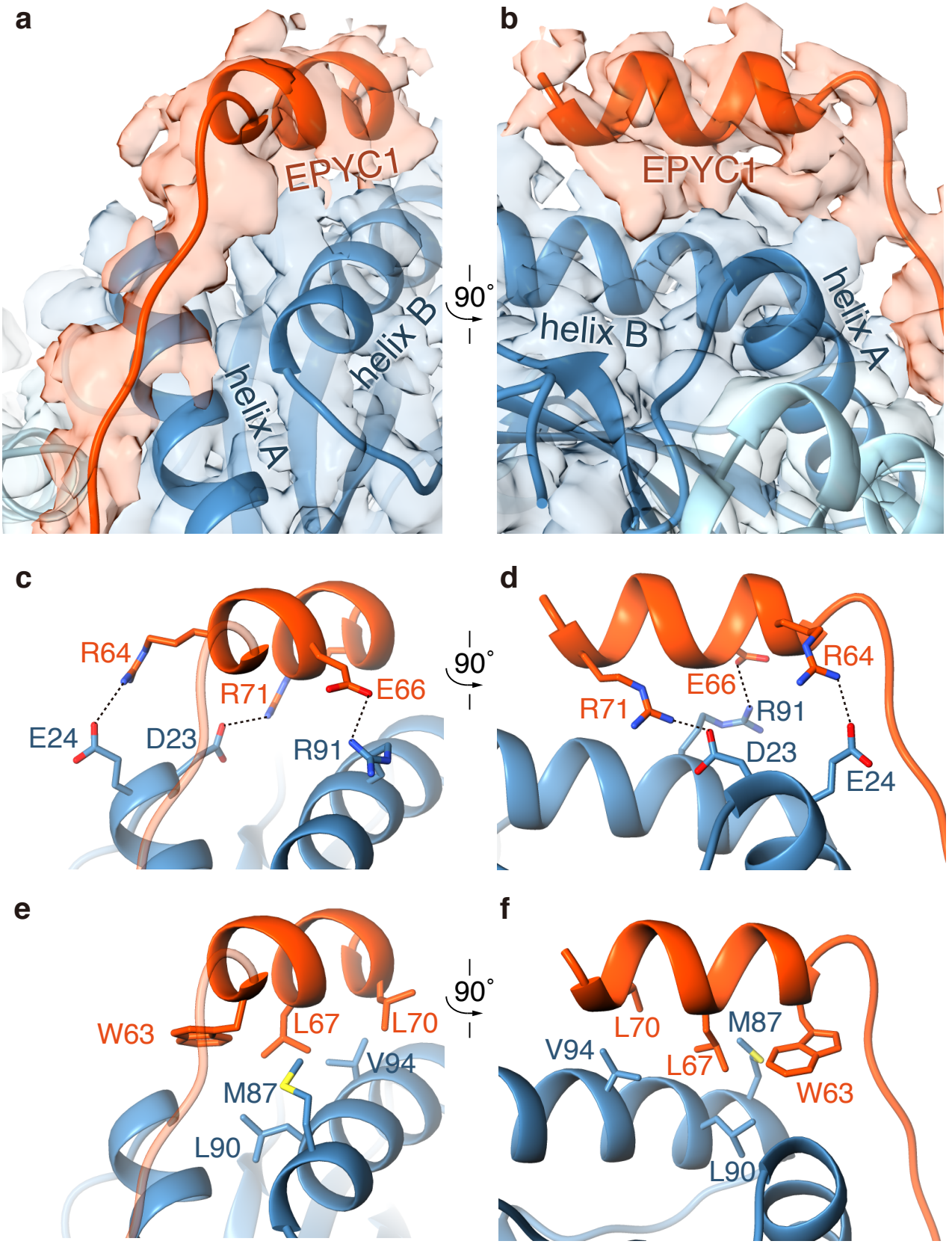
12

13

Type	**Number** **If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.**	**Filename** **This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_ Supplementary_Video_1.mov***	**Legend or Descriptive Caption** **Describe the contents of the file**
Supplementary Table	1	Supplementary_Table_3-5_20201016.xlsx	Supplementary Tables 3-5


a
b
Pyrenoid
Chloroplast
c
Rubisco
EPYC1
d
Full length EPYC1
EPYC1 peptides
peptide array
e
Rubisco
EPYC1 peptides
peptide array
f
g
18 amino acid peptide
Rubisco binding signal
0
2
4
6
0
50
100
150
200
250
300
Peptide position along EPYC1
h
cTP
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Repeat 5
i
27 AARGSWRESSTATVQAS---RASSA---TNR-VSPTRSVLPANWRQELESLRNGNGS 76 Repeat 1
77 SSAASSAPAPARSSSASWRDAAPASSAPA---RSSSA---SKKAVTPSRSALPSNWKQELESLRS--SS 137 Repeat 2
138 PAPASSAPAPARSSSASWRDAAPASSAPA---RSSS----SKKAVTPSRSALPSNWKQELESLRS--SS 197 Repeat 3
198 PAPASSAPAPARSSSASWRDAAPASSAPA---RSSSA---SKKAVTPSRSALPSNWKQELESLRS-N-S 258 Repeat 4
259 PAPASSAPAPARSSSASWRDA-PASSSSSSADKAGTNPWTGKSKPEIKRTALPADWRKGL 317 Repeat 5
Nonpolar and hydrophobic
Polar and hydrophilic
Basic
Acidic


a
cTP
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Repeat 5
Full length EPYC1
EPYC1 49-72 peptide
Rubisco saturated with EPYC1 49-72 peptide
b
c
S
L
−90°
d
e
Rubisco large subunit
Rubisco small subunit
EPYC1 peptide

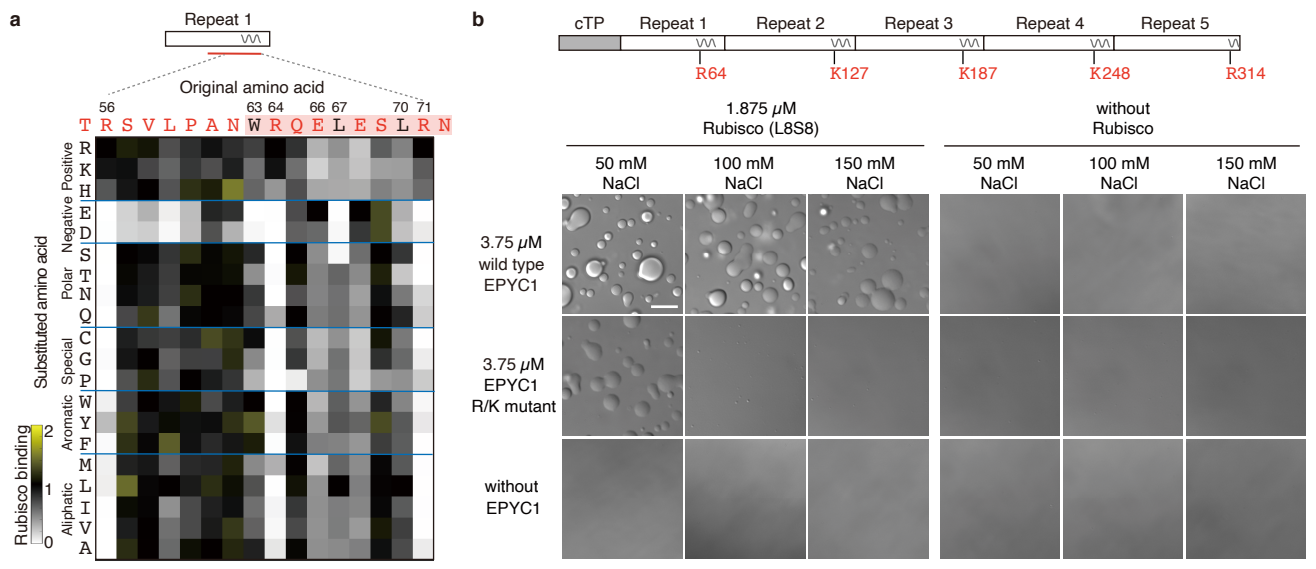


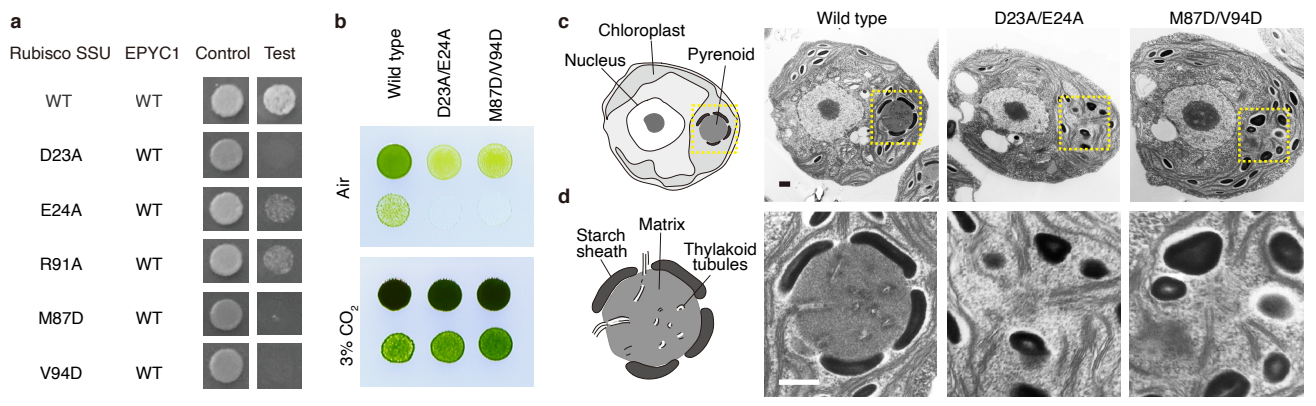
g

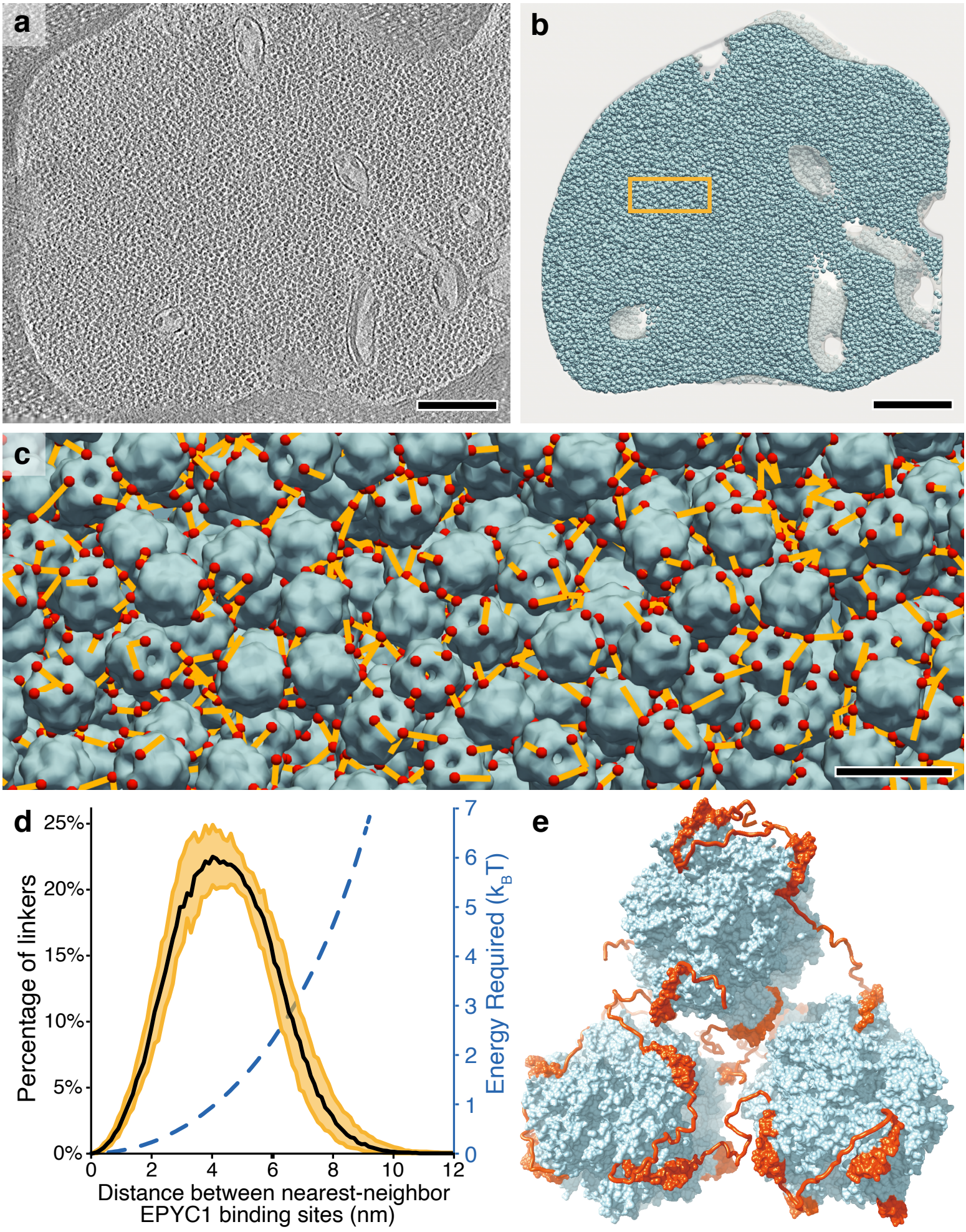
35 AVIYDVQAAIQED 23 — Rubisco SSU helix A

49 TNRVSPTRSVLPANWRQELESLRN 72 — EPYC1$_{49\text{-}72}$ peptide

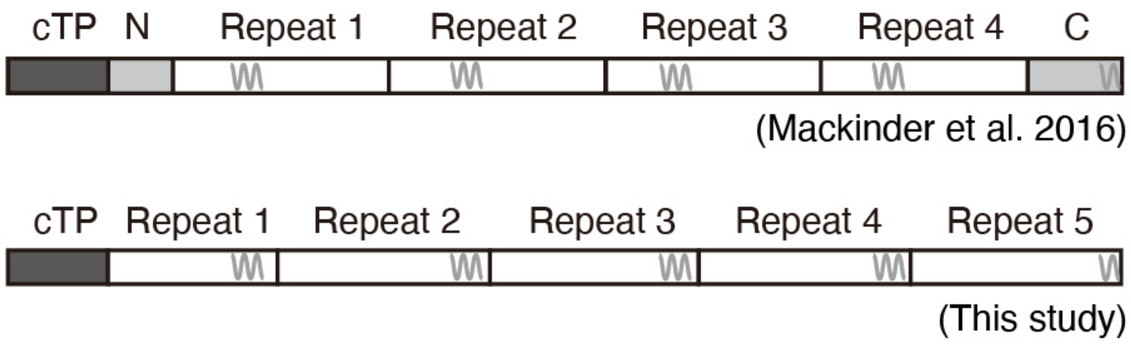
86 PMQVLREIVACTKA 99 — Rubisco SSU helix B


a
Repeat 1
Original amino acid
56
63 64
66 67
70 71
T R S V L P A N W R Q E L E S L R N
Substituted amino acid
Positive
Negative
Polar
Special
Aromatic
Aliphatic
R
K
H
E
D
S
T
N
Q
C
G
P
W
Y
F
M
L
I
V
A
Rubisco binding
2
1
0
b
cTP
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Repeat 5
R64
K127
K187
K248
R314
1.875 μM
Rubisco (L8S8)
without
Rubisco
50 mM
NaCl
100 mM
NaCl
150 mM
NaCl
50 mM
NaCl
100 mM
NaCl
150 mM
NaCl
3.75 μM
wild type
EPYC1
3.75 μM
EPYC1
R/K mutant
without
EPYC1

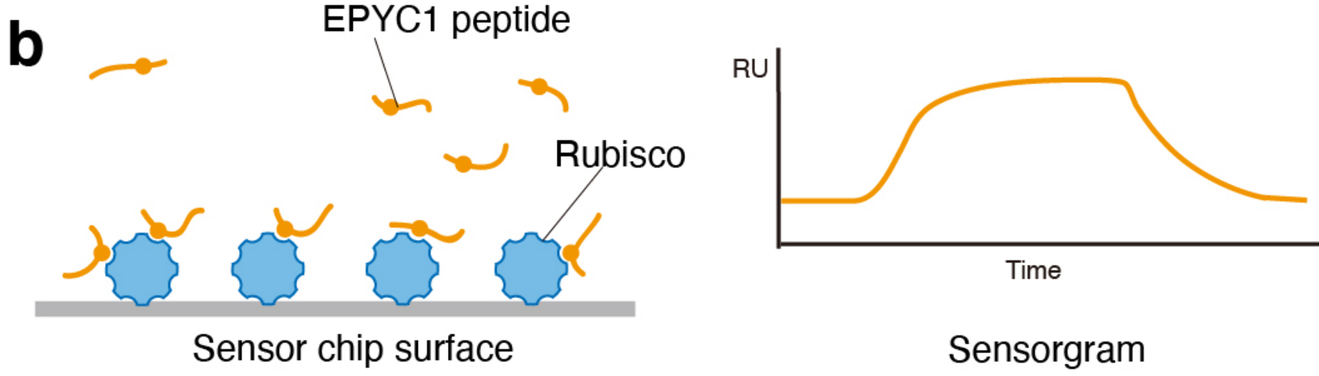

a
Rubisco SSU
EPYC1
Control
Test
WT
WT
D23A
WT
E24A
WT
R91A
WT
M87D
WT
V94D
WT
b
Wild type
D23A/E24A
M87D/V94D
Air
3% CO2
c
Chloroplast
Nucleus
Pyrenoid
Wild type
D23A/E24A
M87D/V94D
d
Starch sheath
Matrix
Thylakoid tubules


a
b
c
d
25%
20%
15%
10%
5%
0%
Percentage of linkers
0
2
4
6
8
10
12
Distance between nearest-neighbor
EPYC1 binding sites (nm)
7
6
5
4
3
2
1
0
Energy Required ($k_B T$)
e

a



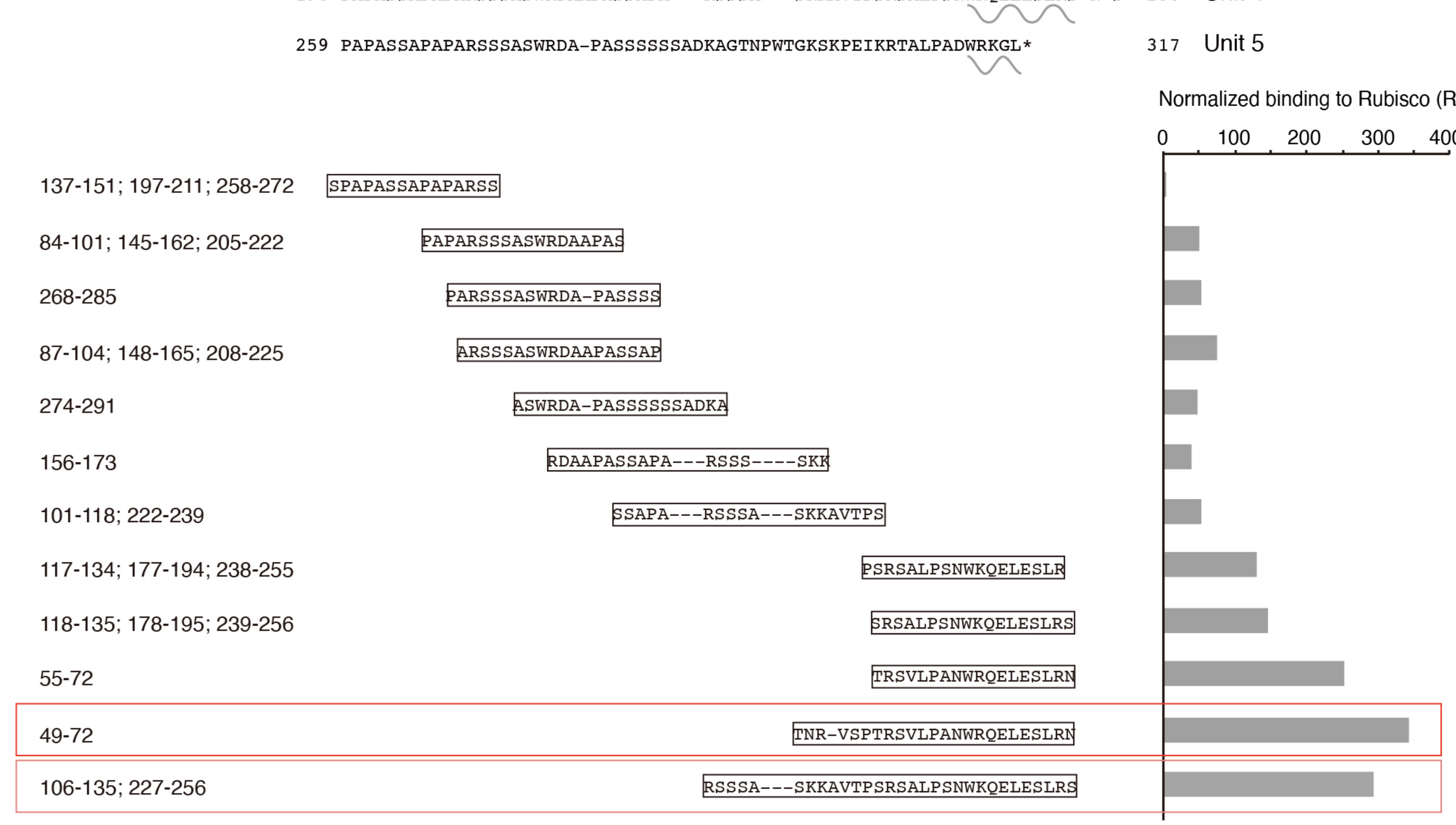
b



c

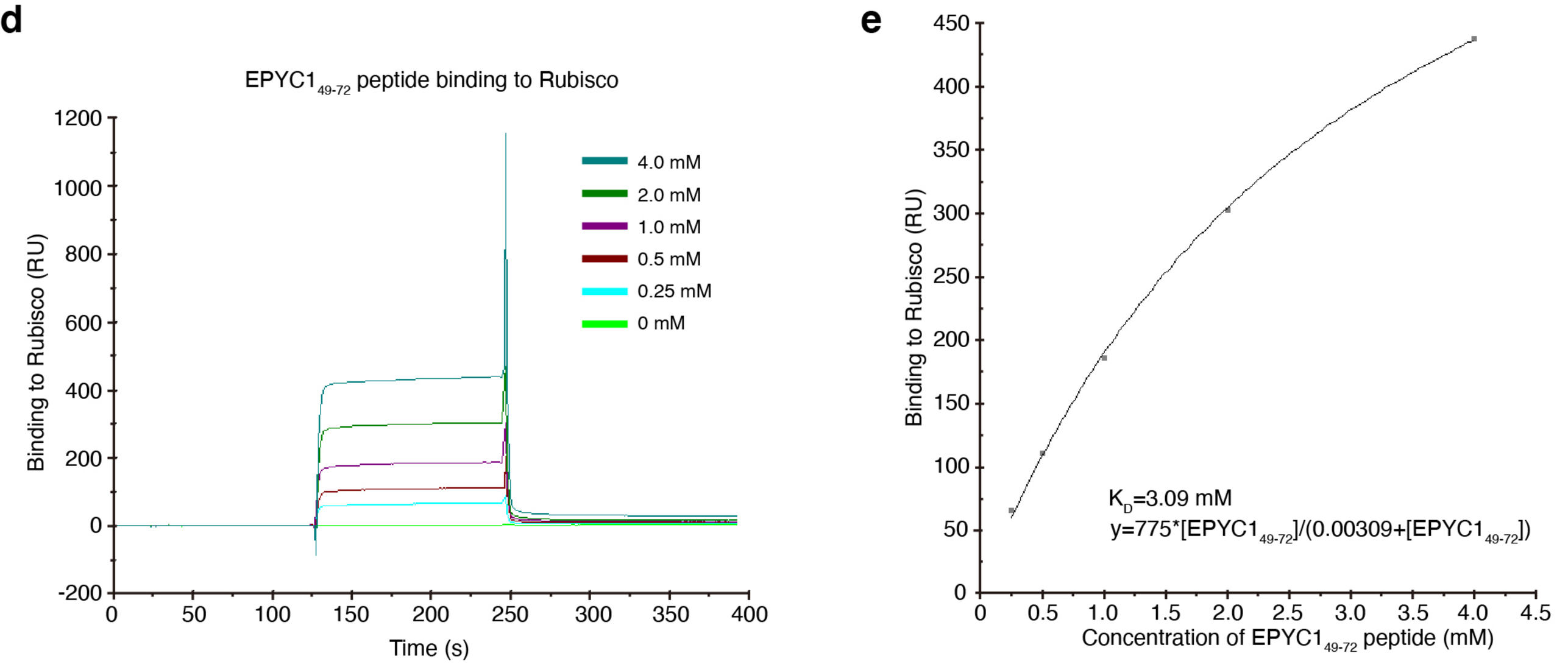
```
  1 MATISSMRVGAASRVVVSGRVKTVKVA                                          27
28                ARGSWRESSTATVQAS---RASSA---TNR-VSPTRSVLPANWRQELESLRNGNGS  76
77 SSAASSAPAPARSSSASWRDAAPASSAPA---RSSSA---SKKAVTPSRSALPSNWKQELESLRS--SS 137
138 PAPASSAPAPARSSSASWRDAAPASSAPA---RSSS----SKKAVTPSRSALPSNWKQELESLRS--SS 197
198 PAPASSAPAPARSSSASWRDAAPASSAPA---RSSSA---SKKAVTPSRSALPSNWKQELESLRS-N-S 258
259 PAPASSAPAPARSSSASWRDA-PASSSSSSADKAGTNPWTGKSKPEIKRTALPADWRKGL*          317
```

cTP
Unit 1
Unit 2
Unit 3
Unit 4
Unit 5



d

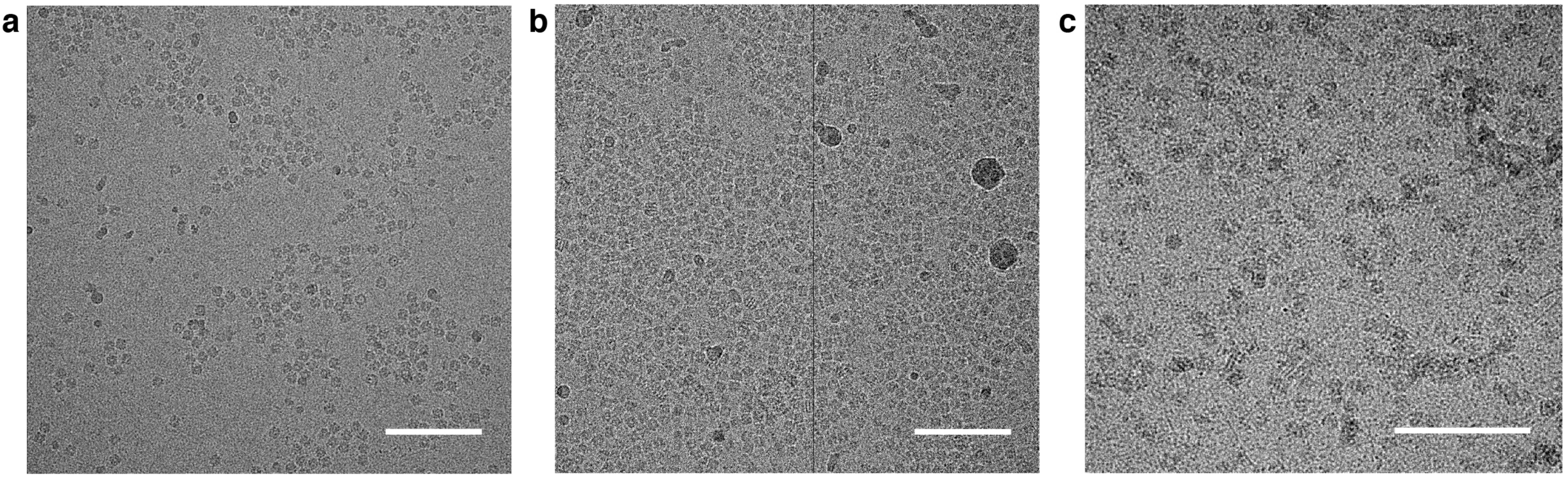
e



Apo Rubisco
(1,834 micrographs)

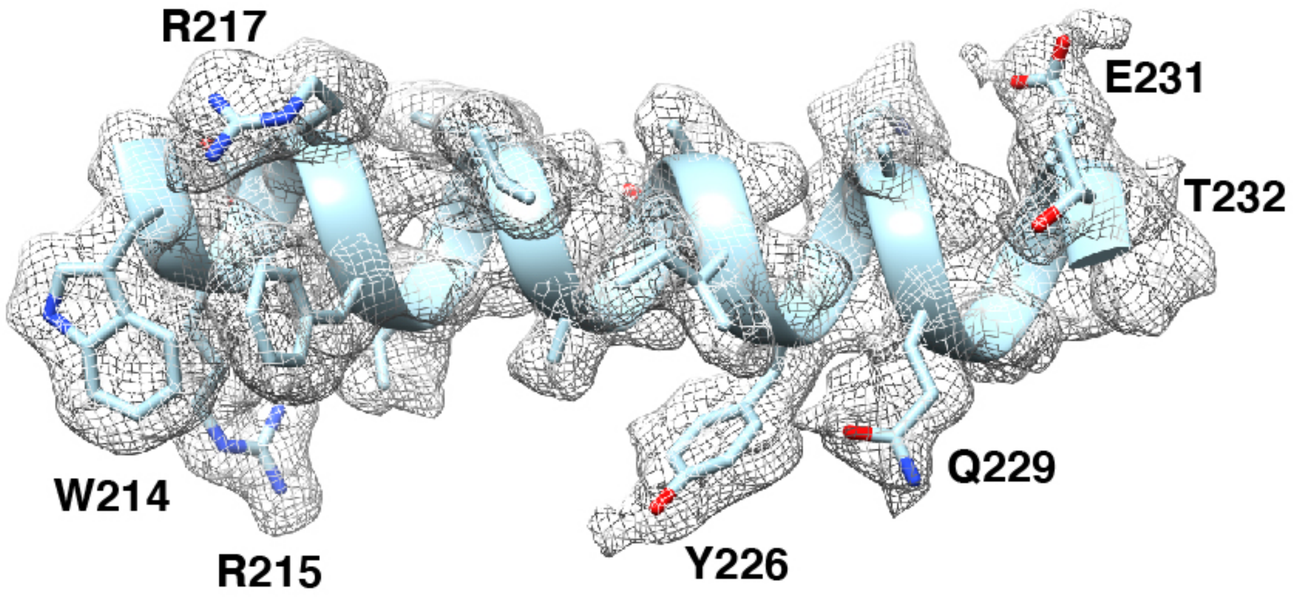
$EPYC1_{49-72}$ peptide-bound Rubisco
(2,500 micrographs)

$EPYC1_{106-135}$ peptide-bound Rubisco
(13,727 micrographs)



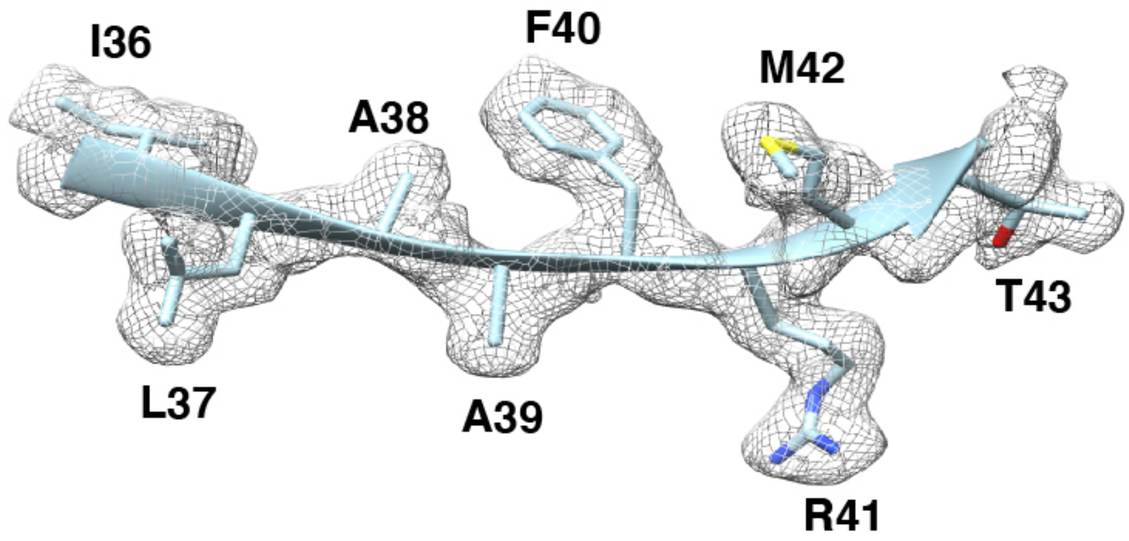
a

Residues 214-232 on Chain A

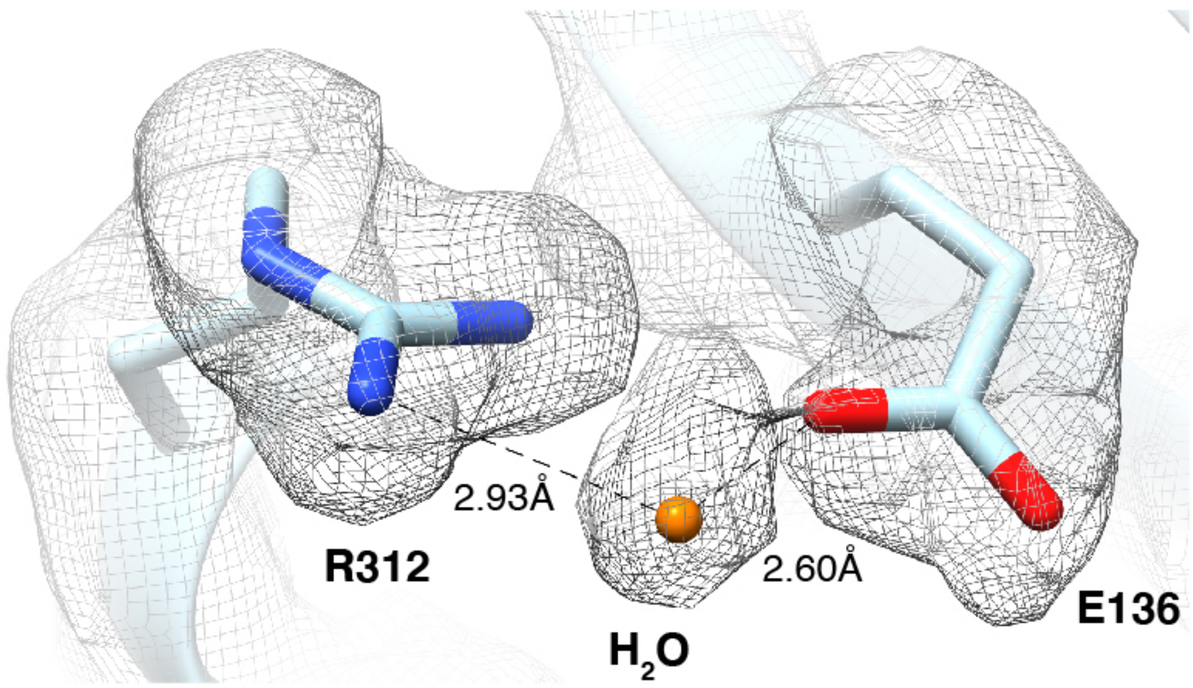


b

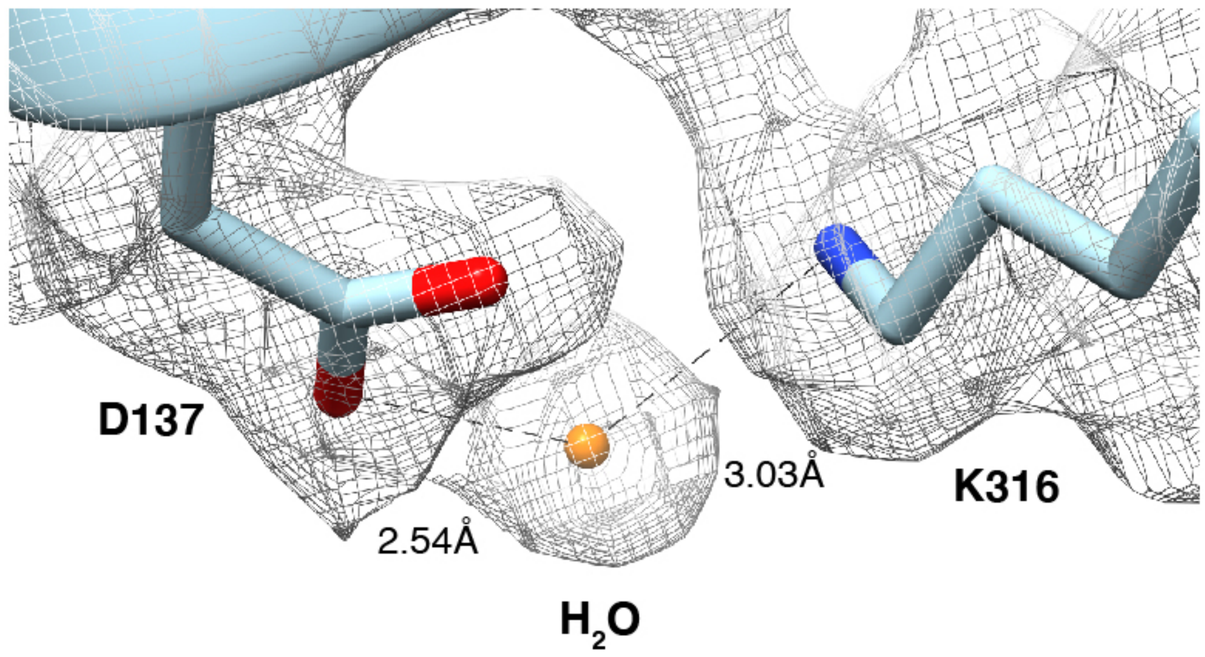
Residues 36-43 on Chain A



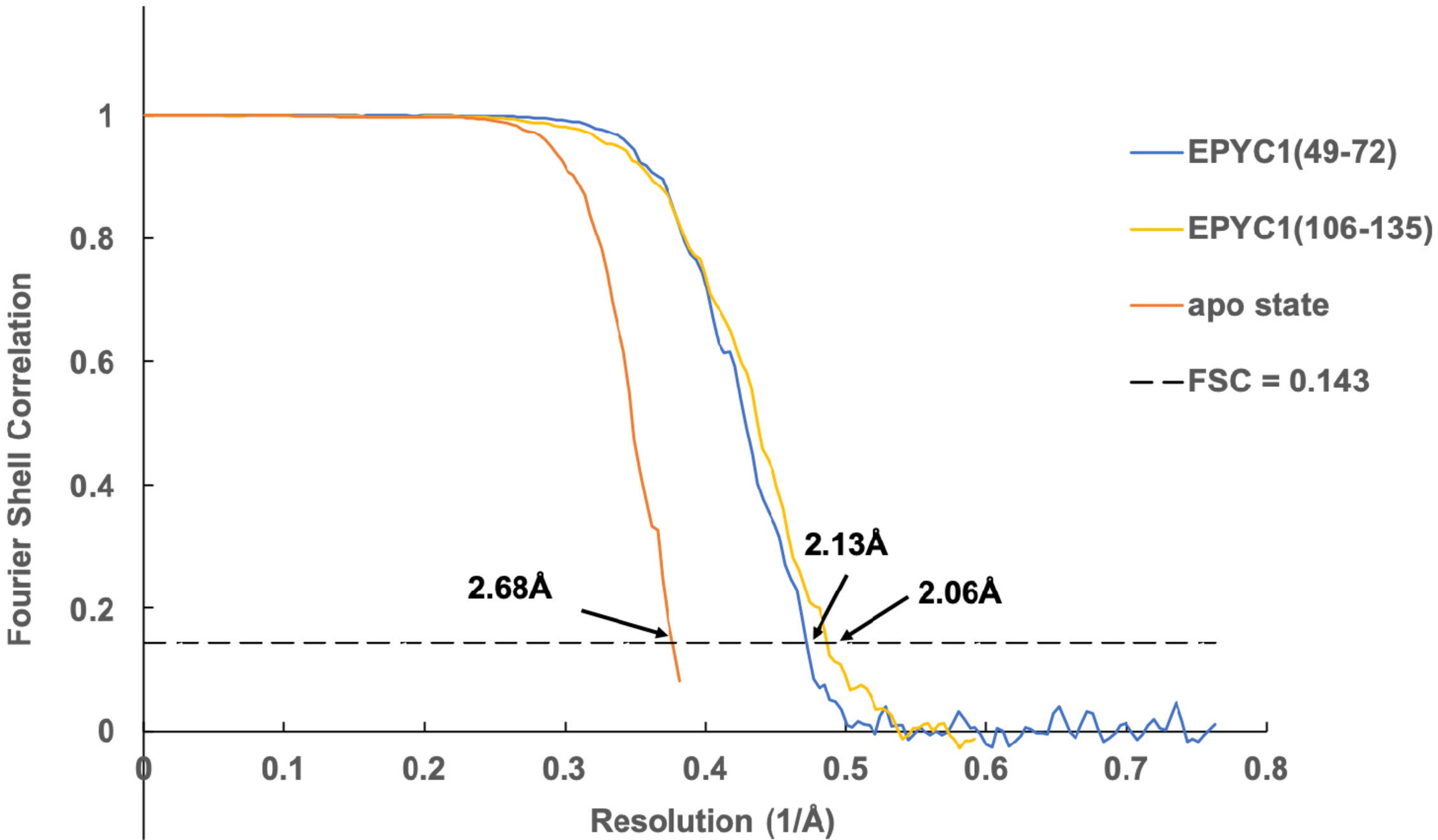
c



d

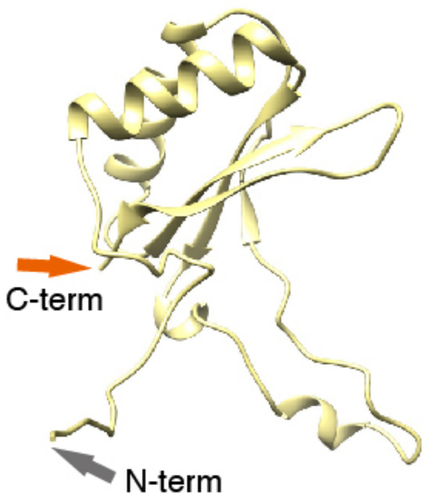


e

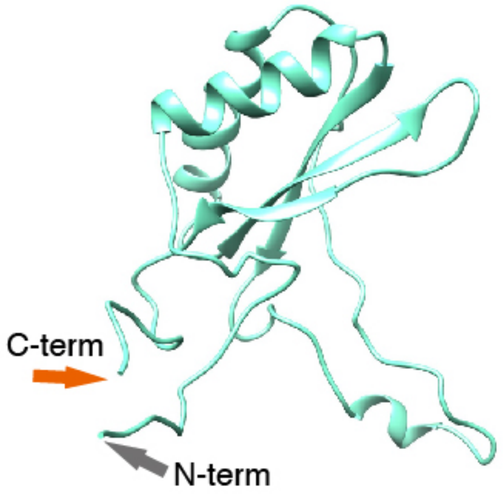


a

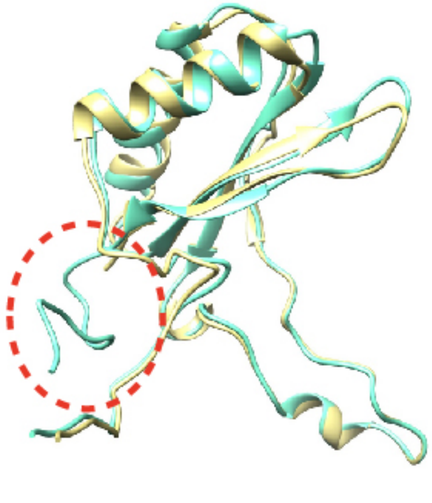
X-ray



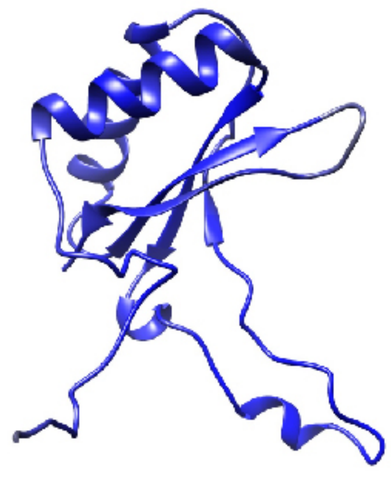
EM (apo)

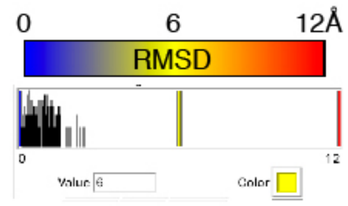


overlay



RMSD (X-ray)

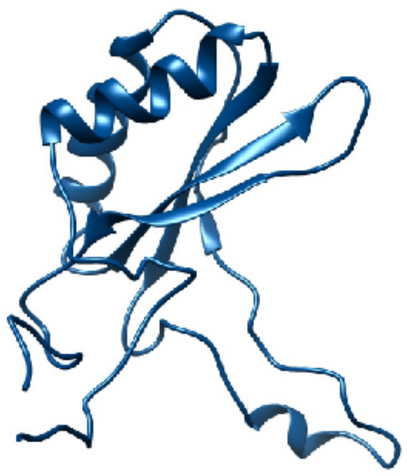




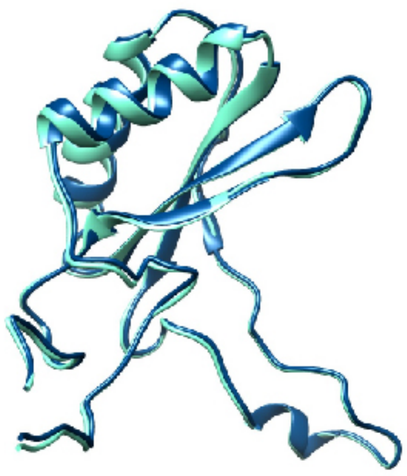
b

EM (apo)

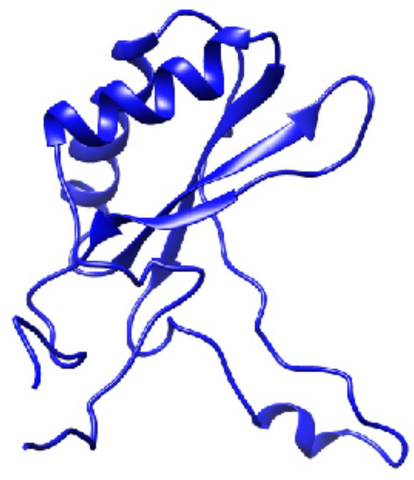
EM (peptide-bound)

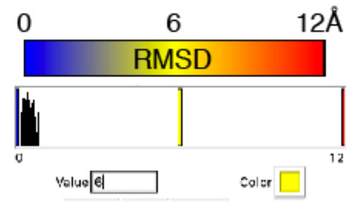


overlay



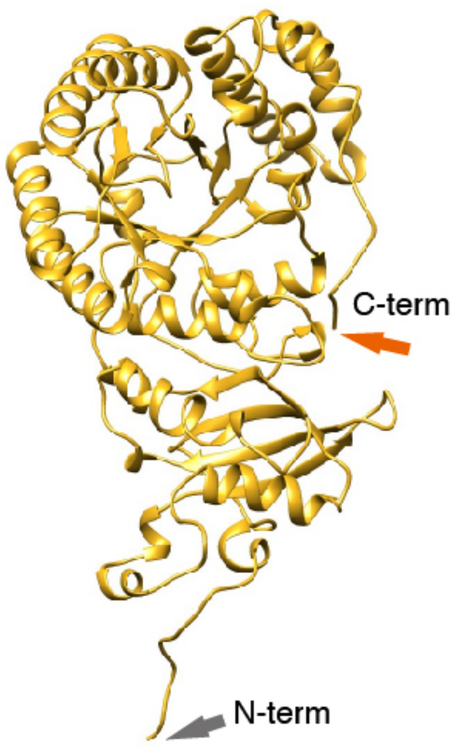
RMSD (EM-apo)





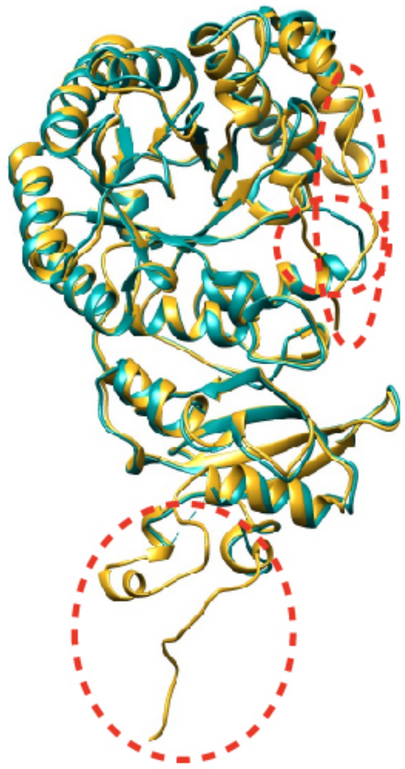
c

X-ray

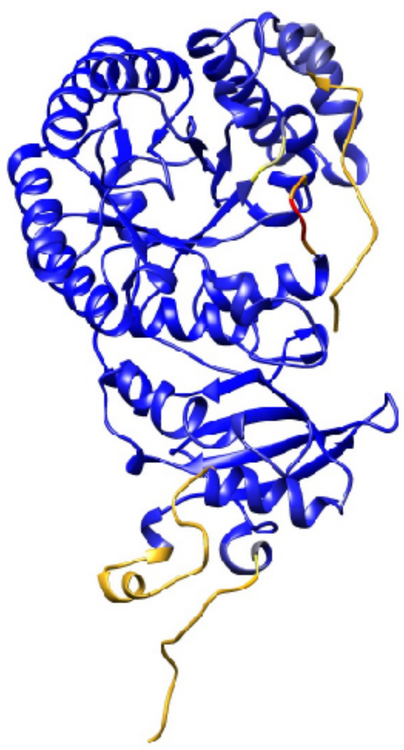


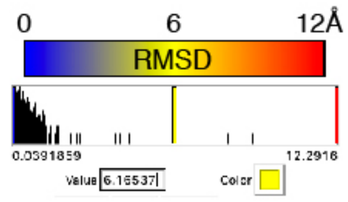
EM (apo)

overlay



RMSD (X-ray)





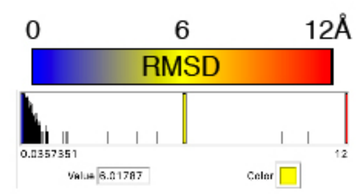
d

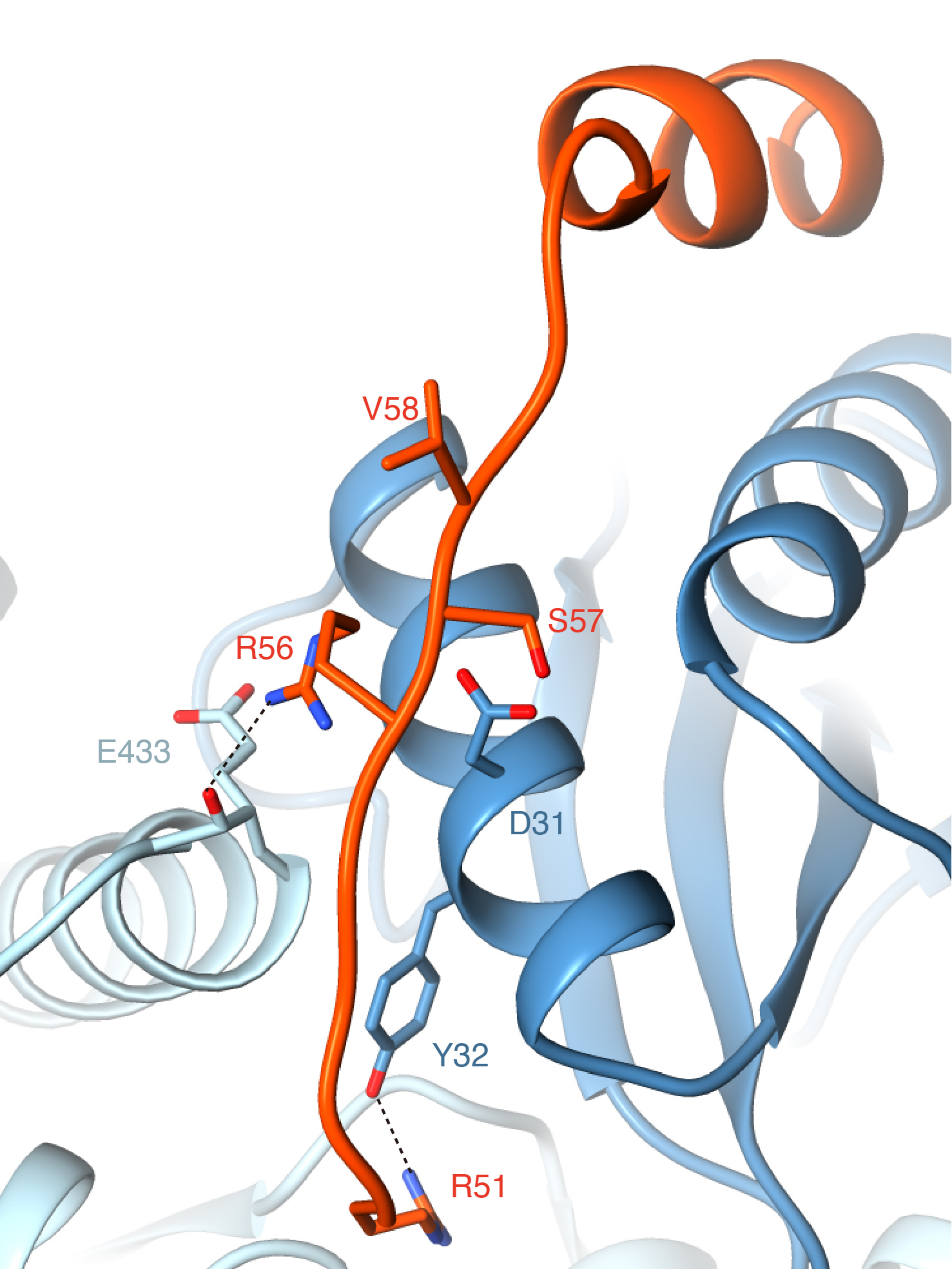
EM (apo)

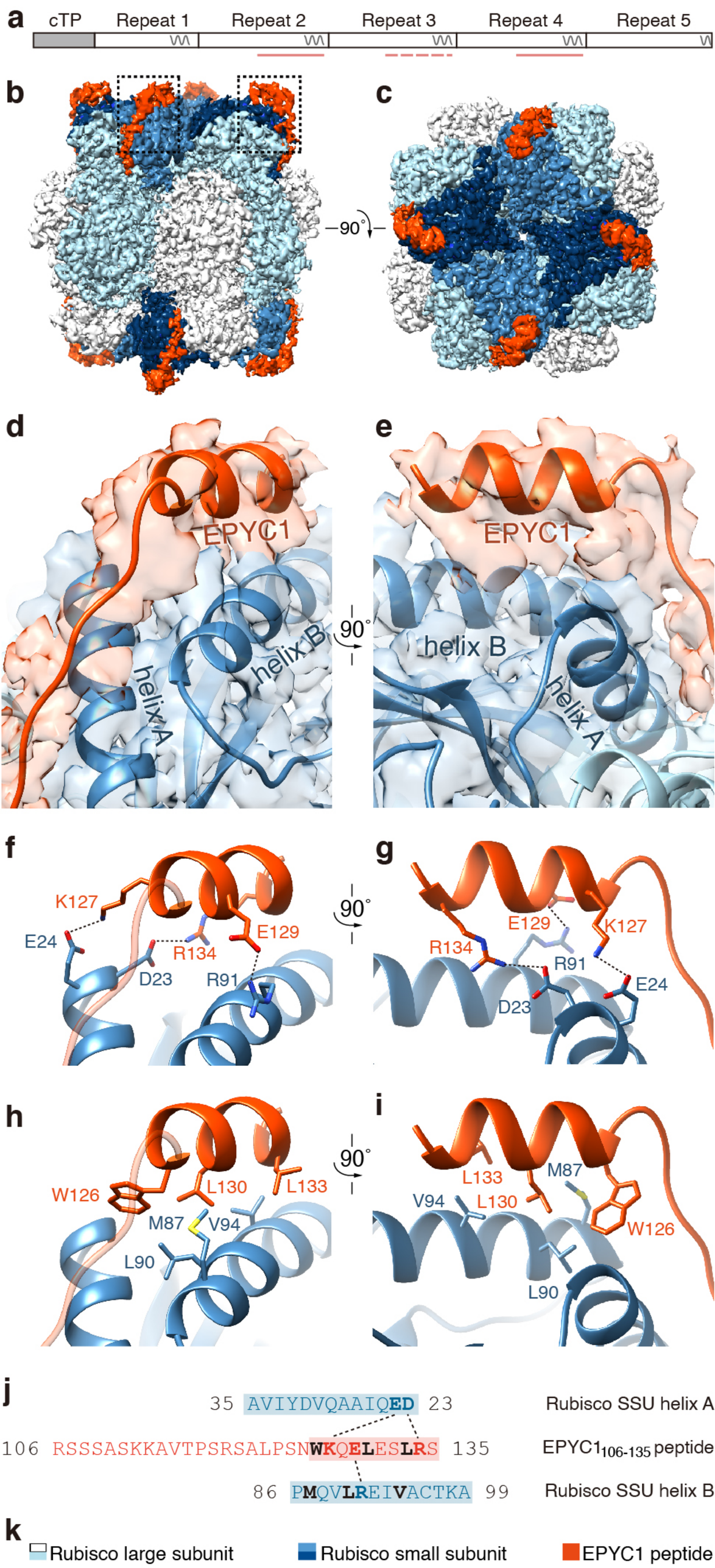
EM (peptide-bound)

overlay

RMSD (EM-apo)




V58
S57
R56
E433
D31
Y32
R51


a
cTP
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Repeat 5
b
c
90˚
d
EPYC1
helix B
helix A
90˚
e
EPYC1
helix B
helix A
f
K127
E24
R134
E129
D23
R91
90˚
g
E129
K127
R134
R91
E24
D23
h
W126
L130
L133
M87
V94
L90
90˚
i
L133
M87
V94
L130
W126
L90
j
35 AVIYDVQAAIQED 23
Rubisco SSU helix A
106 RSSSASKKAVTPSRSALPSNWKQELESLRS 135
EPYC1106-135 peptide
86 PMQVLREIVACTKA 99
Rubisco SSU helix B
k
Rubisco large subunit
Rubisco small subunit
EPYC1 peptide

56 62 63 64 65 67 70 71

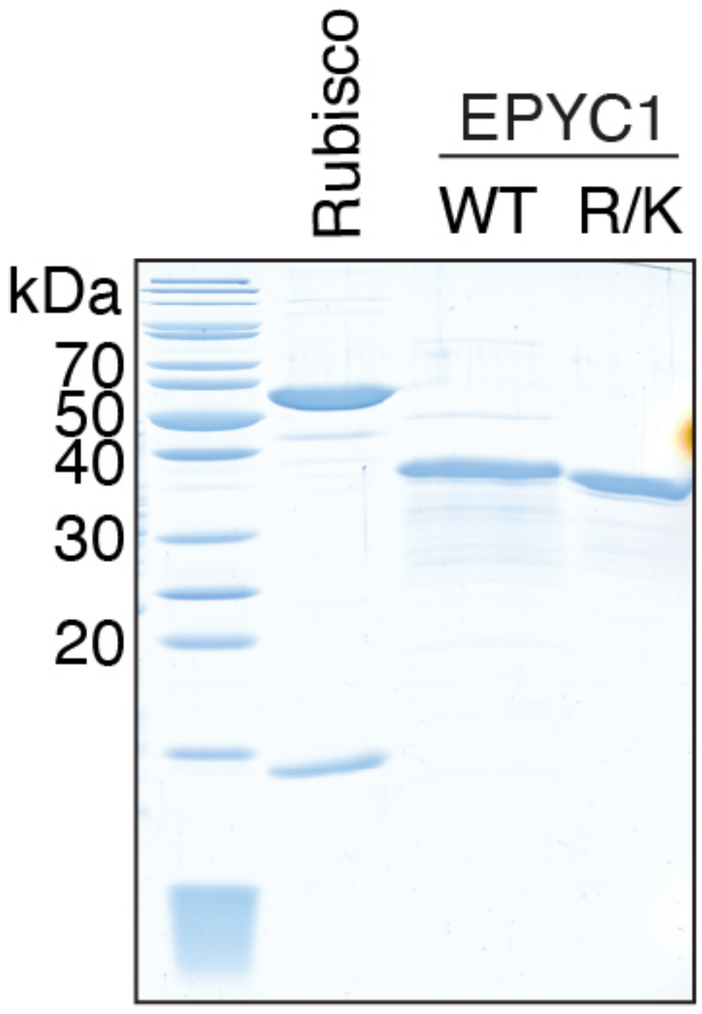
T R S V L P A N W R Q E L E S L R N

Normalized binding to Rubisco (RU)

120
100
80
60
40
20
0

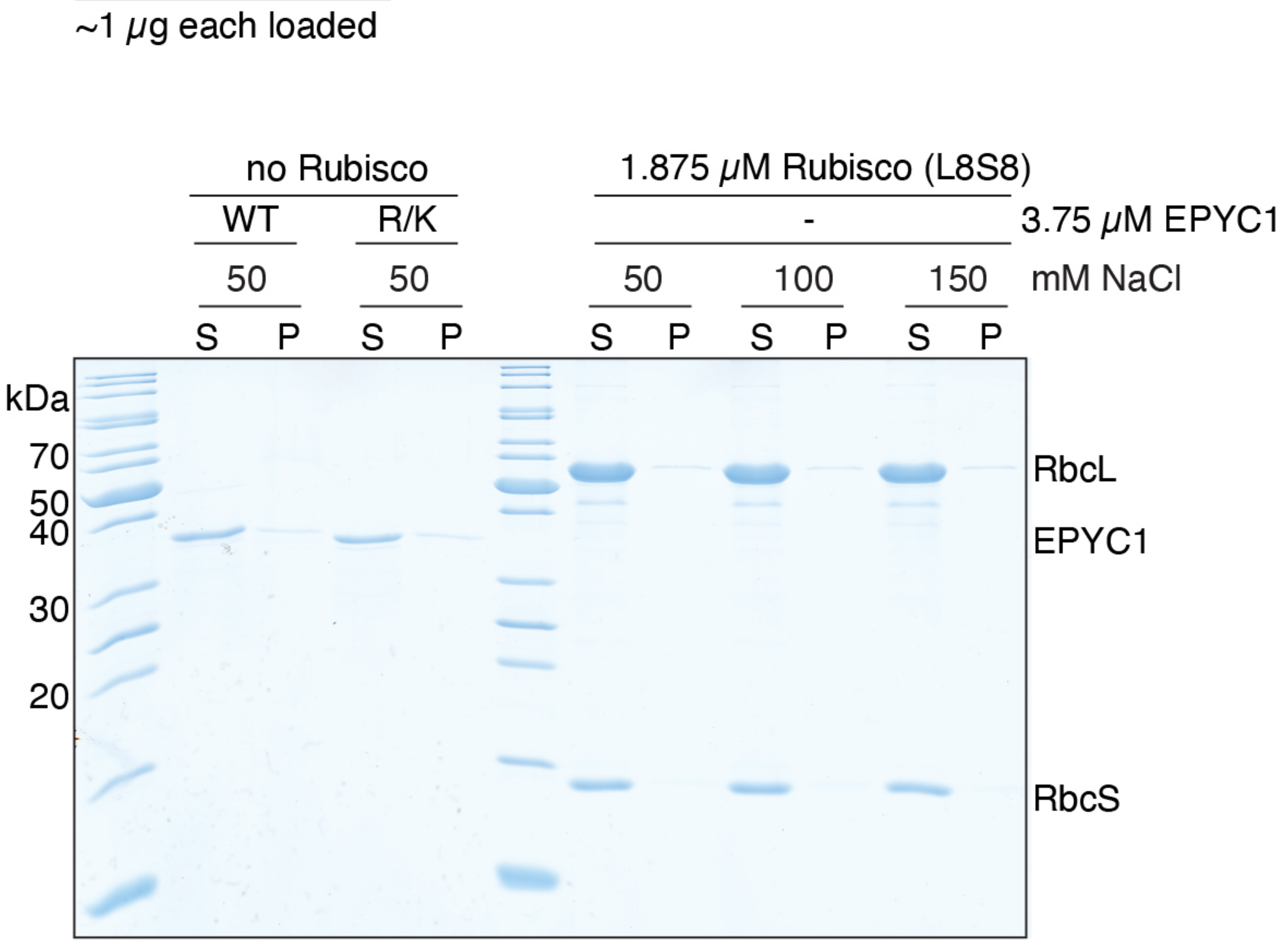
WT R56A N62A W63A R64A Q65A L67A L70A R71A

a

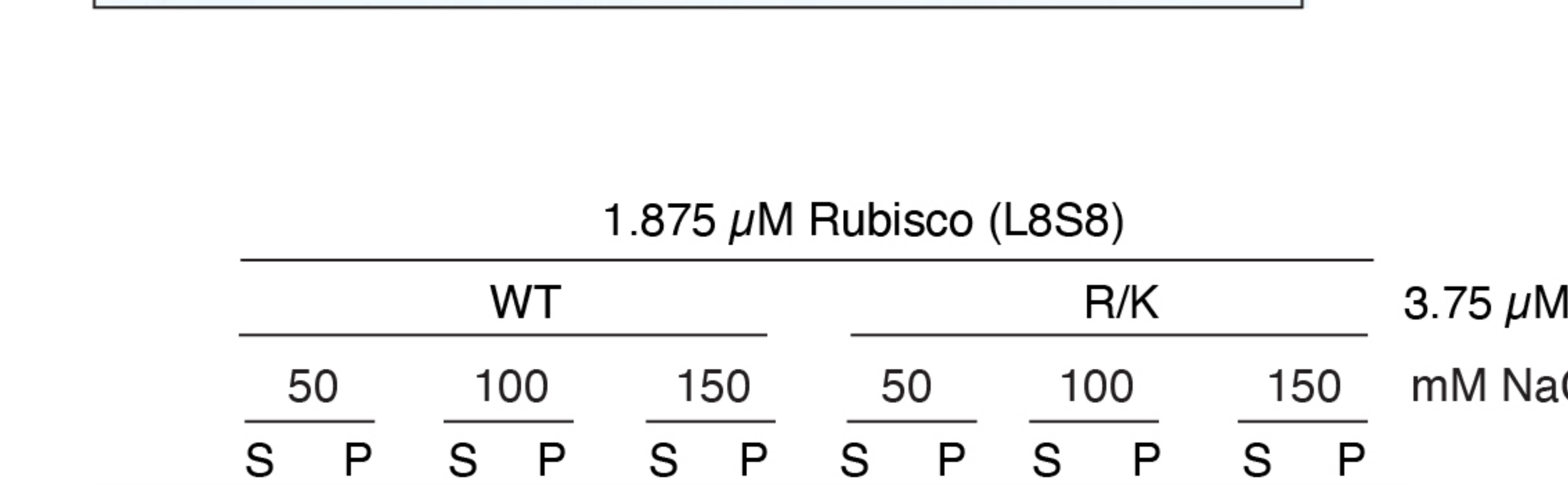


~1 μg each loaded

b



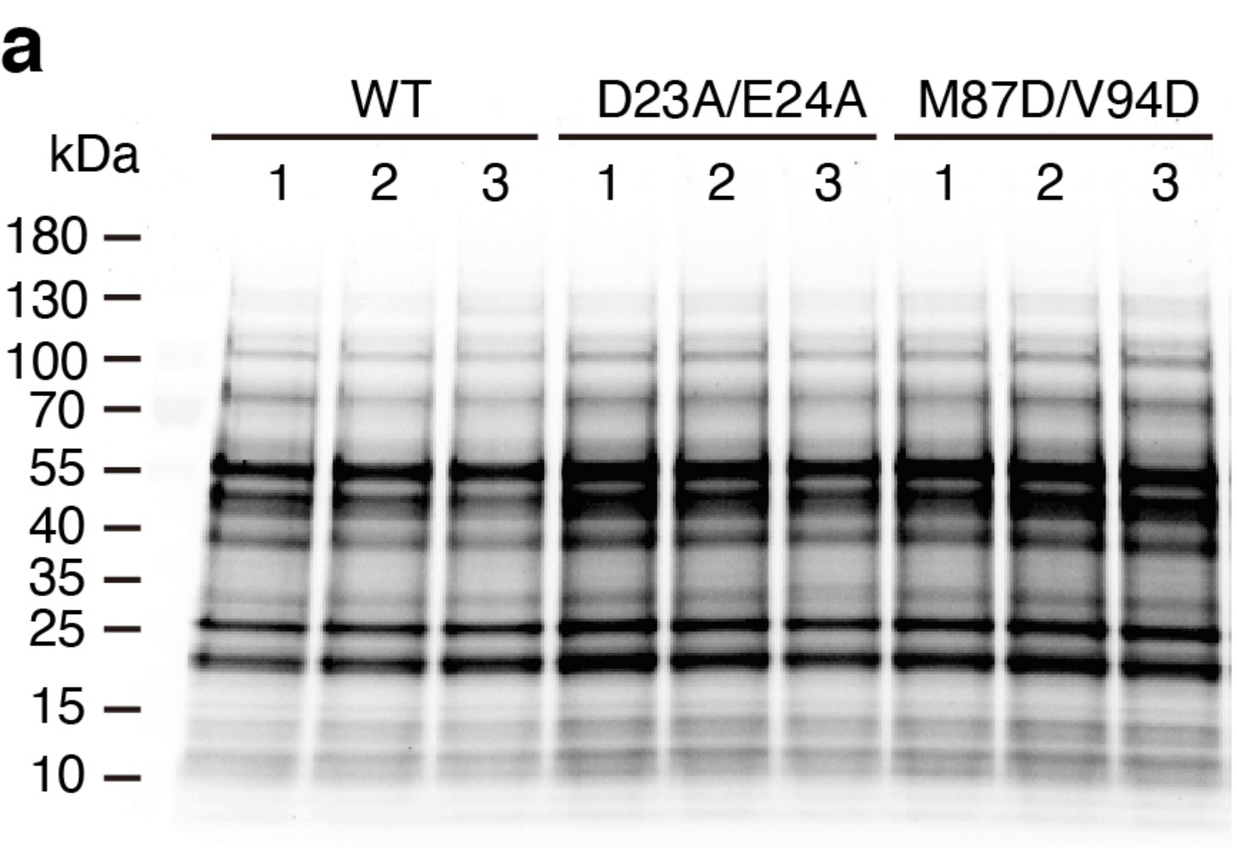
c

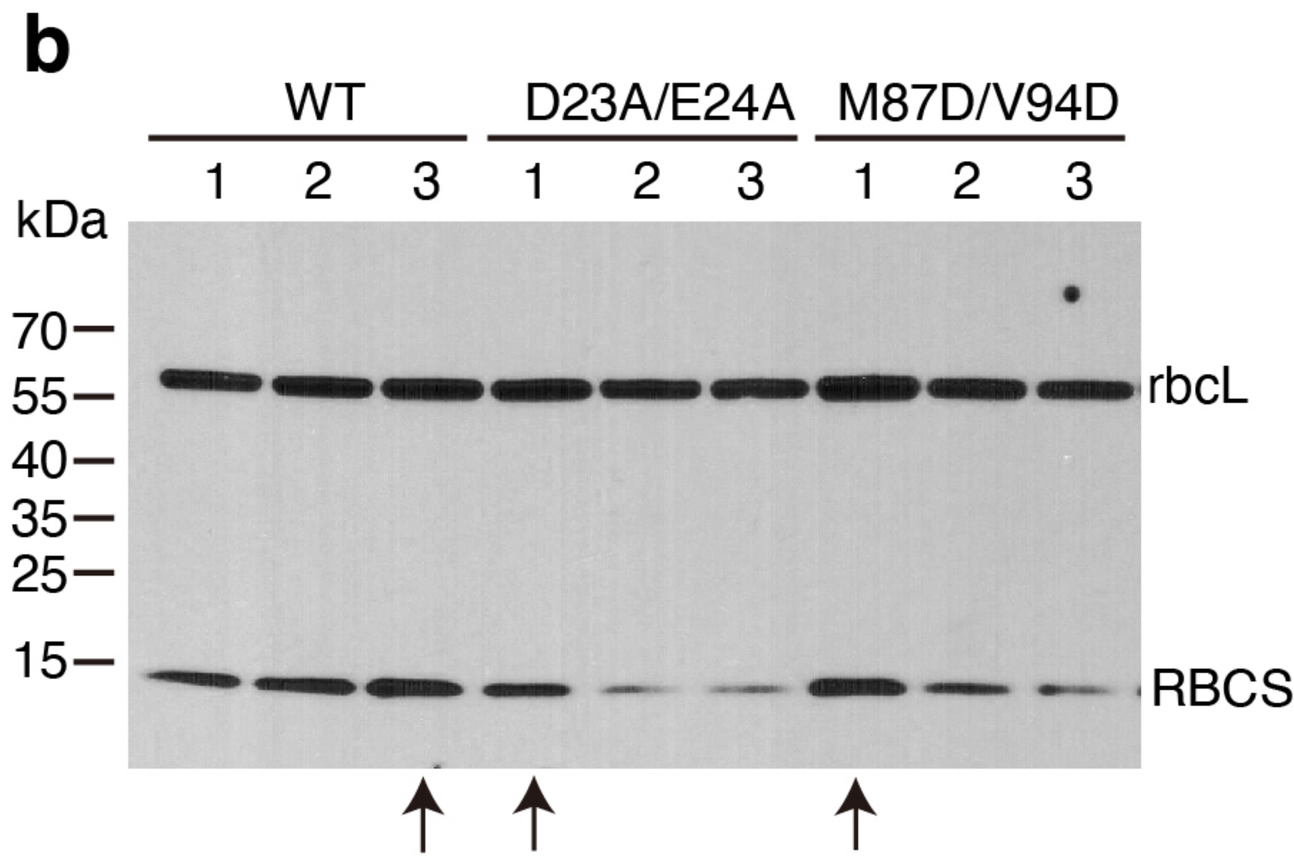


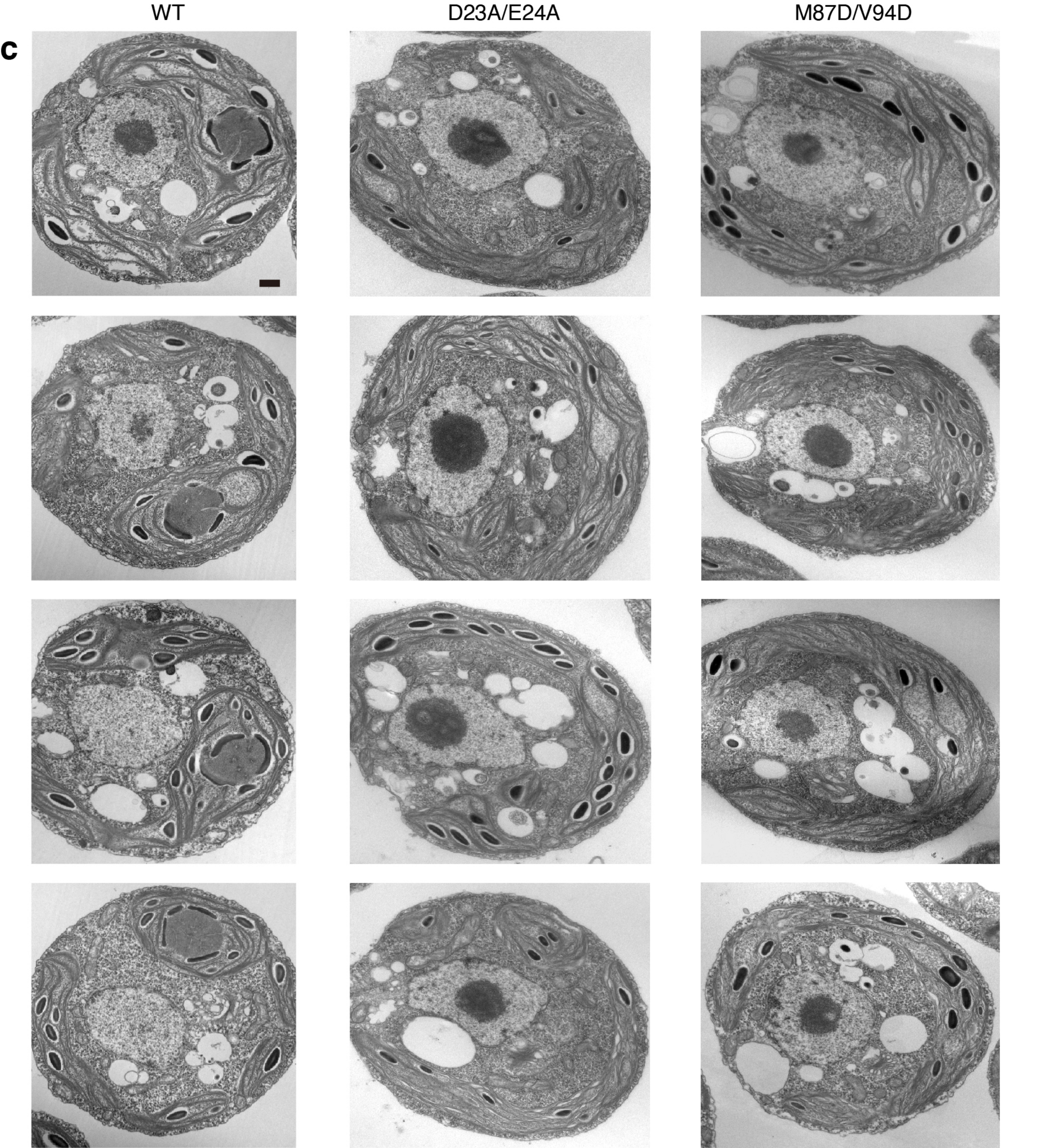
SD-L-W-H with 3-AT (mM)

BD	AD	SD-L-W	0	0.5	1	2.5	5	10	20
+	+								
EPYC1	Rubisco SSU WT								
	D23A								
	E24A								
	D23A;E24A								
	R91A								
	D23A;E24A;R91A								
	M87D								
	V94D								
	M87D;V94D								
	D31A								
	Y32A								

initial OD 0.5 0.1 0.5 0.1 0.5 0.1 0.5 0.1 0.5 0.1 0.5 0.1 0.5 0.1 0.5 0.1


a
WT
D23A/E24A
M87D/V94D
kDa
1 2 3 1 2 3 1 2 3
180
130
100
70
55
40
35
25
15
10


b
WT
D23A/E24A
M87D/V94D
1 2 3 1 2 3 1 2 3
kDa
70
55
40
35
25
15
rbcL
RBCS


c
WT
D23A/E24A
M87D/V94D