nature plants

ARTICLES

https://doi.org/10.1038/s41477-020-0665-8

Check for updates

Molecular basis for the assembly of RuBisCO assisted by the chaperone Raf1

Ling-Yun Xia[1,2,3], Yong-Liang Jiang[1,2,3]✉, Wen-Wen Kong[1,2,3], Hui Sun[1], Wei-Fang Li[1], Yuxing Chen[1]✉ and Cong-Zhao Zhou[1,2]✉

The folding and assembly of RuBisCO, the most abundant enzyme in nature, needs a series of chaperones, including the RuBisCO accumulation factor Raf1, which is highly conserved in cyanobacteria and plants. Here, we report the crystal structures of Raf1 from cyanobacteria *Anabaena* sp. PCC 7120 and its complex with RuBisCO large subunit RbcL. Structural analyses and biochemical assays reveal that each Raf1 dimer captures an RbcL dimer, with the C-terminal tail inserting into the catalytic pocket, and further mediates the assembly of RbcL dimers to form the octameric core of RuBisCO. Furthermore, the cryo-electron microscopy structures of the RbcL–Raf1–RbcS assembly intermediates enable us to see a dynamic assembly process from $RbcL_8Raf1_8$ to the holoenzyme $RbcL_8RbcS_8$. In vitro assays also indicate that Raf1 can attenuate and reverse CcmM-mediated cyanobacterial RuBisCO condensation. Combined with previous findings, we propose a putative model for the assembly of cyanobacterial RuBisCO coordinated by the chaperone Raf1.

Photosynthesis is a fundamental biological process on Earth that provides the energy and sugars for most living organisms. In this process, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which is the most abundant enzyme in nature[1,2], is responsible for converting the atmospheric CO_2 into organic carbon. It catalyses the first, often rate-limiting, step of carbon fixation by carboxylating the five-carbon sugar substrate ribulose-1,5-bisphosphate (RuBP) to produce two molecules of 3-phosphoglyceric acid (3-PGA)[3]. The RuBisCO enzymes are generally classified into three forms[4]. Form I, which is the most common form widespread in plants, algae, cyanobacteria and proteobacteria, is composed of eight large (RbcL, ~53 kDa) and eight small (RbcS, ~15 kDa) subunits, whereas form II and form III RuBisCO consist of only one or more RbcL dimers. Notably, RuBisCO is considered to be inefficient owing to its slow catalytic rate and unavoidable inhibition by O_2, which impairs its CO_2/O_2 specificity[5]. Thus, engineering RuBisCO to improve the carboxylation efficiency and CO_2/O_2 specificity is viewed as a promising strategy to increase crop growth and yield[6–9].

The biogenesis of form I RuBisCO is a complicated process that needs a series of chaperones[10,11]. In cyanobacteria, the initial folding of RbcL subunit is generally assisted by the chaperonin GroEL and its cofactor GroES (corresponding to the homologues Cpn60 and Cpn10 in plants, respectively)[12]. Afterwards, the RbcL subunits form dimers that are further assembled into the octameric core $RbcL_8$, mainly assisted by individual molecular chaperones, such as RbcX[13–15] and the RuBisCO accumulation factor Raf1 (ref. [16]). Notably, the assembly of cyanobacterium *Synechococcus* PCC 6301 RbcL octameric core in *Escherichia coli* can occur independent of RbcX and Raf1 (ref. [17]). Once the $RbcL_8$ core is formed, docking of RbcS—which displaces the chaperones RbcX[15] and/or Raf1 (ref. [18])—eventually enables the formation of the holoenzyme $RbcL_8RbcS_8$ (hereafter, L_8S_8). The scaffold protein CcmM can further engage with cyanobacterial RuBisCO to form functional condensates of aggregated proteins that form the inner protein matrix of the carboxysome[19,20], which is a subcellular microcompartment that utilizes a CO_2-concentrating mechanism to achieve a higher efficiency of carbon fixation[21]. Notably, in vitro assembly of plant RuBisCO requires four chloroplast chaperones, RbcX and Raf1 in addition to the plant-specific Raf2 and BSD2, as well as the chaperonins Cpn60 and Cpn20 (ref. [22]).

The chaperone Raf1 was first identified in maize as a nucleus-encoded chloroplast protein that is required for efficient RuBisCO biogenesis[23], and overexpression of RuBisCO large and small subunits with Raf1 resulted in a greater than 30% increase in RuBisCO content in maize[24]. Sequence analysis showed that Raf1 proteins are highly conserved in plants and cyanobacteria, sharing a similar domain organization with an N-terminal α-helical domain (Raf1α) and a C-terminal β-sheet domain (Raf1β) connected by a flexible linker of around 23 residues. Crystal structures of the two individual domains Raf1α and Raf1β of *Arabidopsis thaliana* Raf1, in combination with biochemical and biophysical analysis of the interfaces between RbcL and Raf1, show that Raf1 brackets the antiparallel RbcL dimer to promote the $RbcL_8$ formation[16]. Furthermore, it was found that Raf1 participates in the RuBisCO degradation pathway in *Synechocystis* sp. PCC 6803 under sulfur starvation, indicating that Raf1 also has a role in RuBisCO recycling and homeostasis[25].

Extensive studies on the structure and catalytic mechanism of RuBisCO in the past decades[4,26] largely advanced our knowledge on its biogenesis. However, mechanistic details on the assembly and maturation of RuBisCO coordinated by a series of chaperones are still lacking. Here we report the crystal structures of full-length Raf1 from *Anabaena* sp. PCC 7120 and its complex with RbcL, as well as four cryo-electron microscopy (cryo-EM) structures of RbcL–Raf1–RbcS assembly intermediates. Using structural analyses combined with biochemical assays, we determined the molecular mechanism of the dynamic assembly process of RuBisCO. Moreover, we found that Raf1 can antagonize CcmM-mediated condensation of RuBisCO in vitro. These findings, in combination with the previous

[1]Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, China. [2]Innovation Academy for Seed Design, Chinese Academy of Sciences, Hefei, China. [3]These authors contributed equally: Ling-Yun Xia, Yong-Liang Jiang, Wen-Wen Kong. ✉e-mail: jyl@ustc.edu.cn; cyxing@ustc.edu.cn; zcz@ustc.edu.cn

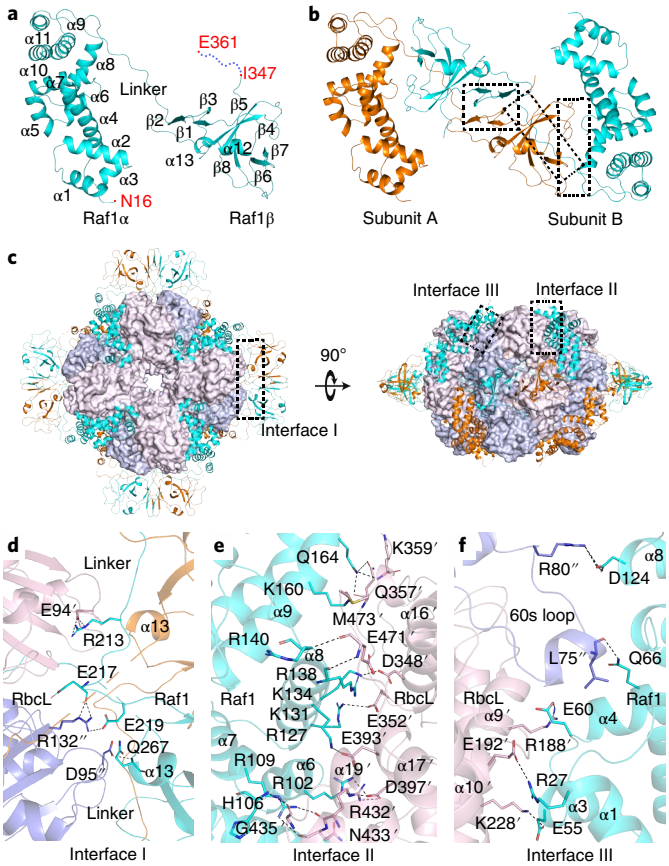


Fig. 1 | Crystal structures of Raf1 in apo- and RbcL-bound forms. a, Ribbon representation of the Raf1 subunit. The secondary structural elements are labelled sequentially. The residues Asn16 and Ile347 at the N and C termini in the structure, as well as the C-terminal-most Glu361 of the Raf1 sequence, are labelled. **b**, The dimeric structure of Raf1, with the two subunits coloured in cyan and orange. The three interfaces are indicated by dashed boxes. **c**, The overall structure of L_8F_8 in two orientations rotated by 90°. The four Raf1 dimers are shown as cartoons; the two subunits of each dimer are coloured in cyan and orange. The RbcL octameric core is shown as surface, with the two subunits of each dimer coloured in blue and pink. **d**–**f**, The three interfaces between RbcL and Raf1. The RbcL subunits are shown in blue and pink as semi-transparent cartoons, whereas the Raf1 subunits are shown in cyan and orange as semi-transparent cartoons. The interacting residues are shown as sticks, with polar interactions indicated as dashed lines. Residues from two subunits of an RbcL dimer are labelled with a prime and a double prime, respectively. **d**, The interface I between the interdomain linker regions of Raf1 and the area around the active-site pockets of RbcL dimer. **e**, The interface II between the central concave surface of Raf1α and the convex surface of the C-terminal TIM barrel domain of RbcL. **f**, The interface III between the convex surface of Raf1α and the neighbouring RbcL dimer.

reports, enable us to propose a model of chaperone-assisted assembly of cyanobacterial RuBisCO.

Results

The full-length structure of Raf1 dimer. Here we solved the full-length structure of *Anabaena* sp. PCC 7120 Raf1 at a resolution of 2.85 Å. Each asymmetric unit consists of one Raf1 molecule, which covers the residues from Asn16 to Ile347 (Fig. 1a). Symmetric operation indicated that Raf1 further assembles into a swapped dimer (Fig. 1b). Size-exclusion chromatography coupled with multi-angle static light scattering (SEC–MALS) data suggested that Raf1 also exists as a dimer in solution (Supplementary Fig. 1).

Each Raf1 subunit consists of three segments—the N-terminal α-helical domain Raf1α (residues 16–197), the C-terminal β-sheet domain Raf1β (residues 221–347), and a linker that connects the Raf1α and Raf1β domains (residues 198–220) with a length of about 61 Å (Fig. 1a). The full-length Raf1 adopts an extended V-shaped structure (Fig. 1a), with an interdomain angle of 39°. The two domains Raf1α and Raf1β, which have no direct interactions, are structurally similar to those of *A. thaliana* Raf1, with root-mean-square deviation (RMSD) values of 1.4 and 1.9 Å over 147 and 110 Cα atoms, respectively (Supplementary Fig. 2).

Two subunits of Raf1 form an N-shaped dimeric structure, in which the two Raf1β domains are swapped and pack against each other to form a tight junction, flanked by a Raf1α domain on each side (Fig. 1b). Raf1β of one subunit interacts with Raf1α, Raf1β and the linker of the symmetric subunit through three interfaces with a total surface area of approximately 3,700 Å^2. The β-hairpin β2–β3 of one Raf1β protrudes towards the symmetric Raf1β and packs against its counterpart, therefore forming a double-layered β-sheet at the centre with an interface area of around 1,400 Å^2 (Extended Data Fig. 1a). The hydrophobic linker region of one subunit runs along the hydrophobic cleft of Raf1β of the symmetric subunit (Extended Data Fig. 1b), resulting in the swapped Raf1β domains; Raf1β of one subunit loosely interacts with Raf1α of the symmetric subunit through a small interface of approximately 450 Å^2 (Extended Data Fig. 1c). Notably, the linker region consists of an N-terminal moiety (residues 198–202) of high flexibility and a C-terminal moiety (residues 203–220) that interacts with Raf1β of the symmetric subunit. Sequence analysis revealed that the residues of the long linker, except for a couple of residues at the C terminus, are relatively divergent (Extended Data Fig. 2).

Overall structure of the $RbcL_8Raf1_8$ complex. Previous findings showed that Raf1 facilitates the assembly of RuBisCO by forming the $RbcL_8Raf1_8$ assembly intermediate[16,18,23]. Indeed, SEC–MALS analyses suggested that coexpression of *Anabaena* Raf1 and RbcL in *E. coli* enabled the production of a protein complex of 780 kDa (Supplementary Fig. 3), which is comparable to the theoretical molecular weight of $RbcL_8Raf1_8$ (hereafter, L_8F_8). We next crystallized the L_8F_8 complex and solved its structure at a resolution of 3.00 Å using molecular replacement. Each asymmetric unit contains two RbcL dimers and two Raf1 dimers, which further form the L_8F_8 complex by symmetric operation.

In the complex, eight RbcL subunits form an octameric core consisting of four tightly packed antiparallel dimers (Fig. 1c). Seeing along the four-fold axis, the four Raf1 dimers embrace the exposed surface of the $RbcL_8$ core and, therefore, shield further lateral contacts (Fig. 1c). Each Raf1 dimer acts as a tweezer to clamp an RbcL dimer, in which each dimeric Raf1β (the junction of the tweezer) positions at the equator of the RbcL dimer with the two-fold axes coinciding with each other (Fig. 1c). By contrast, each Raf1α (the arm of the tweezer), which is docked to the interface cleft between two neighbouring RbcL dimers, embraces one side of the RbcL dimer (Fig. 1c). The overall structure of the $RbcL_8$ core is quite similar to that in the RuBisCO holoenzyme[15,27], with an RMSD value of 0.80 Å over 848 Cα atoms for each RbcL dimer. Notably, the overall architecture of L_8F_8 resembles the previously published 21-Å negative-stain EM map of L_8F_8 from *Synechococcus elongatus* PCC 6301 (ref. [16]).

In the L_8F_8 complex, each Raf1 dimer clamps to an RbcL dimer through two different interfaces (Fig. 1c, Supplementary Table 1). The interface I, which is between the junction of dimeric Raf1β and the area around the catalytic pockets of RbcL dimer (Fig. 1d), possesses a total buried interface area of around 1,000 Å^2. More extensive contacts come from the 1,600-Å^2 interface II, which is formed by the central concave surface of one Raf1α and the convex surface of the C-terminal TIM-barrel domain of an RbcL subunit (Fig. 1e).

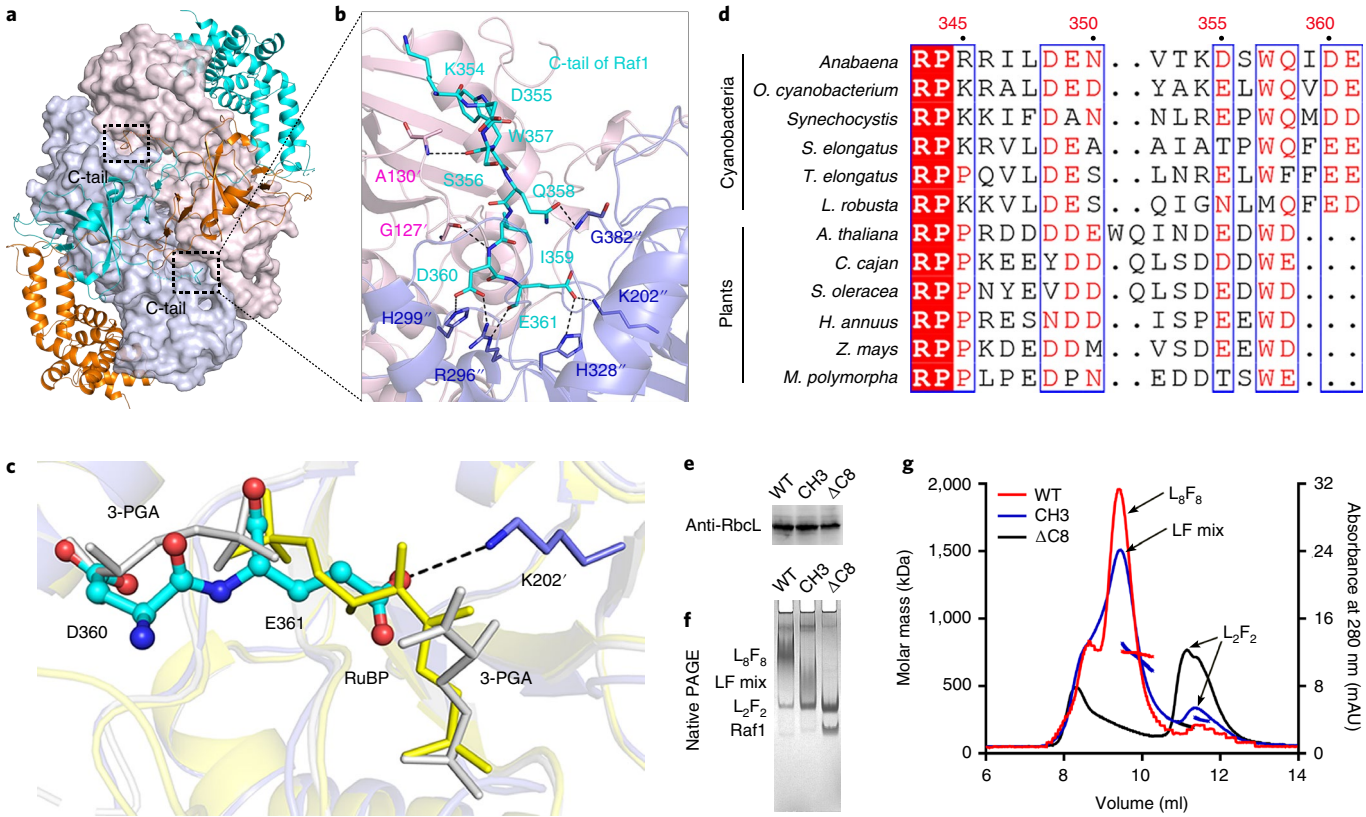


Fig. 2 | The C-tail of Raf1 inserting into the catalytic pocket of RbcL and its contribution to the RbcL octamer assembly. a, Overall view of the C-tail of Raf1 inserting into the catalytic pocket of RbcL. The two C-tails of the Raf1 dimer are indicated by dashed boxes. The two subunits of Raf1 are coloured in cyan and orange, whereas the two subunits of RbcL are shown as surface and coloured in pink and blue. **b**, The interaction network between the Raf1 C-tail and the catalytic pocket of RbcL. The Raf1 C-tail and interacting residues of RbcL are shown as sticks, whereas the RbcL dimer is shown as a semi-transparent cartoon. **c**, The Raf1 C-tail (cyan) occupies the spaces that correspond to the substrate RuBP (yellow) or the product 3-PGA (grey). Glu 361 of Raf1 forms a salt bridge with the catalytic residue Lys 202 of RbcL (corresponding to Lys 201 in plant RbcL). **d**, Multiple-sequence alignment of Raf1 C-tail sequences from cyanobacteria and plants. *C. cajan*, *Cajanus cajan*; *H. annuus*, *Helianthus annuus*; *L. robusta*, *Limnoraphis robusta*; *M. polymorpha*, *Marchantia polymorpha*; *O. cyanobacterium*, *Oscillatoriales cyanobacterium*; *Z. Mays*, *Zea Mays*. **e**, The yield of His-tagged RbcL proteins in the soluble fraction by coexpressing with wild-type Raf1 (WT) or Raf1 C-tail mutants in *E. coli*, as analysed by immunoblotting using anti-His-tag antibodies. A total of ~40 μg protein was loaded per lane in the SDS–PAGE gel. CH3, extension of the three histidine residues at the C terminus of Raf1; ΔC8, truncation of the C-terminal 8 residues. **f**,**g**, The assembly states of RbcL coexpressed with wild-type Raf1 or C-tail mutants (CH3 and ΔC8) in *E. coli*, as analysed using native PAGE (**f**) and SEC–MALS (**g**). The bands corresponding to different assembly intermediates are indicated to the left of the gel, in which the LF mix represents a mix of varying RbcL–Raf1 assembly intermediates in the smear of the CH3 mutant lane. The SEC–MALS profiles for RbcL coexpressed with wild-type Raf1 or the Raf1 mutants (CH3 and ΔC8 are shown in red, blue and black, respectively). The different assembly intermediates in the peaks are indicated by arrows.

Interface II is stabilized by multiple salt bridges between the basic residues on the helices α6–α9 of Raf1 and the acidic residues on the helices α16, α17 and α19 and the nearby loops of RbcL, in addition to several hydrogen bonds. In addition to the above two interfaces, there exists a 900-Å^2 interface III between the convex surface of Raf1α and an RbcL subunit from the neighbouring RbcL dimer (Fig. 1f, Supplementary Table 1). The helices α1, α3, α4 and α8 of Raf1α interact with α9 and α10, in addition to the 60s loop (residues 64–83) of RbcL (Fig. 1f), which is a loop that forms part of the catalytic site[28]. In contrast to the relatively small interface I, interfaces II and III—which show substantial complementarity in shape and electrostatic potential (Extended Data Fig. 3)—contribute to the majority of interactions between Raf1 and RbcL. Notably, the RbcL residues at the interfaces are exclusively conserved, owing to the functional conservation of RuBisCO. Despite the fact that Raf1 proteins exhibit more divergent sequences, most interface residues are conserved (Extended Data Fig. 2, interfaces I, II and III), suggesting that Raf1 and its homologues share a similar interaction pattern with RbcL. These data also indicate that the RbcL and Raf1 proteins coevolved across cyanobacteria and plants, in agreement with a previous phylogenetic analysis[29]. Notably, the previously proposed residues Arg 97, Arg 104, Lys 126, Lys 129, Arg 155 and Glu 159 of *S. elongatus* PCC 7942 Raf1α[16] are indeed at interface II in our L_8F_8 structure. As interface II is between Raf1α and its captured RbcL dimer, the present complex structure also explained why previous mutations at this interface might decrease the binding to RbcL dimer, but have a minor effect on the assembly of the RbcL octamer[16].

Compared with the apo-form dimeric Raf1 structure, the individual domains of Raf1 in the L_8F_8 complex undergo slight conformational changes in the overall structure, with an RMSD value of 1.04 Å over 175 Cα atoms for Raf1α (Extended Data Fig. 4a) and 0.36 Å over 137 Cα atoms for Raf1β (Extended Data Fig. 4b). However, after binding to RbcL, each Raf1α rotates against the Raf1β junction by around 75°, forming a tweezer-like Raf1 dimer (Extended Data Fig. 4c). Notably, residues Arg 213, Glu 217 and Glu 219 at the C-terminal moiety of the linker interact with Glu 94 and Arg 132 of RbcL, mediating the interactions between Raf1β and RbcL. By contrast, the N-terminal moiety of the linker, which is rather flexible and invisible in the L_8F_8 structure, allows Raf1α to freely rotate against Raf1β.

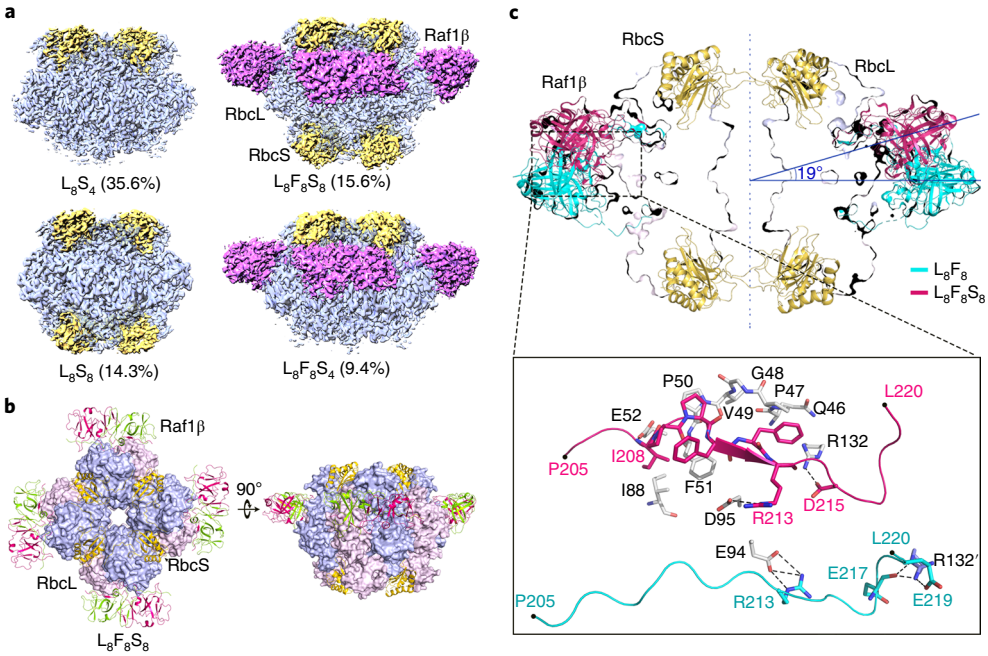


Fig. 3 | Cryo-EM structures of RuBisCO assembly intermediates. a, Cryo-EM maps of different RuBisCO assembly intermediates containing RbcL, RbcS and Raf1. The maps for RbcL, RbcS and Raf1 are coloured in blue, yellow and purple, respectively. The approximate ratios of the intermediates in the total amounts of particles are shown in brackets. **b**, The overall structure of $L_8F_8S_8$, showing in two orientations rotated by 90°. The RbcL octamer core is shown as surface. The two RbcL subunits within a dimer are coloured in blue and pink. The subunits of the Raf1β dimer are shown as green and magenta, whereas RbcS is shown as a yellow cartoon. **c**, Structural comparisons of Raf1β and the linker of Raf1 in the structures of L_8F_8 and $L_8F_8S_8$. The RbcL structure is shown as a surface section, and RbcS in $L_8F_8S_8$ is shown as a yellow cartoon. The Raf1β dimers in L_8F_8 and $L_8F_8S_8$ are shown as cyan and purple cartoons. Compared with the L_8F_8 structure, the Raf1β dimer of $L_8F_8S_8$ rotates ~19° against RbcL along the central axis. Inset: magnification of the detailed interactions between the loop of Raf1 and RbcL in the structures of L_8F_8 and $L_8F_8S_8$. The interacting residues are shown as sticks; polar interactions are indicated by dashed lines.

The C-terminal tail of Raf1 inserting into the catalytic pocket of RbcL also contributes to the RbcL octamer assembly. During the refinement of the L_8F_8 structure, an extra density buried in the catalytic pocket of RbcL was eventually assigned to the eight residues K^{354}DSWQIDE^{361} (Supplementary Fig. 4) at the C-terminal end of Raf1 (termed C-tail for short), which are missing in the structure of the apo-form Raf1 dimer (Fig. 1a). This C-tail inserts deeply into the catalytic pocket of RbcL (Fig. 2a), forming extensive interactions of hydrogen bonds and salt bridges (Fig. 2b). Notably, in the closed-state structure of the holoenzyme L_8S_8, the entrance of the catalytic pocket is covered by the so-called loop 6 (residues 329–336) of RbcL, which is involved in regulating substrate access to the active site[28]. However, in the complex L_8F_8, the C-tail of Raf1 partially occupies the corresponding space that accommodates loop 6 in the holoenzyme (Extended Data Fig. 5).

As shown in the structural superposition (Fig. 2c), the two C-terminal residues Asp 360 and Glu 361 of Raf1 in the structure of L_8F_8 adopt an orientation that is similar to the orientation of the substrate RuBP or the product 3-PGA in the holoenzyme structures from *Spinacia oleracea*[30,31]. These two acidic residues exactly mimic the substrate/product in electrostatic potential. Multiple-sequence alignment showed that Asp 360 and Glu 361 are strictly conserved in cyanobacteria, but are absent in plant Raf1 proteins, which usually have a slightly shorter tail and harbour only one acidic residue at the C terminus (Fig. 2d). Furthermore, Glu 361 of Raf1 forms a salt bridge with the conserved catalytic residue Lys 202 of RbcL (corresponding to Lys 201 in plant RbcL), the carbamylation of which is necessary for the catalysis[26,32]. The structure of the chimeric complex comprising *Thermosynechococcus elongatus* RbcL and *A. thaliana* BSD2 chaperone showed that the RbcL catalytic pocket is occupied by the BSD2 C-terminal tail ending with two conserved acidic residues[22]. As BSD2 is absent in cyanobacteria, it is feasible that cyanobacterial Raf1 may serve the dual functions of both plant Raf1 and BSD2.

A previous study showed that in vitro assembly of *Methanococcoides burtonii* RuBisCO dimers into decamers is enhanced by the binding of either the substrate RuBP or the inhibitor carboxyarabinitol-1,5-bisphosphate[33]. It is reminiscent of the possibility that the insertion of the Raf1 C-tail into the catalytic pocket is involved in the assembly of RbcL octamer. To test this hypothesis, we performed coexpression assays of *Anabaena* His-tagged RbcL, GroEL–ES with wild-type Raf1 or C-tail mutants in *E. coli*. Immunoblot analysis showed a comparable yield of RbcL proteins in the soluble fractions (Fig. 2e), indicating that Raf1 C-tail mutations did not affect the solubility and stability of recombinant RbcL. We next purified the RbcL-containing complexes in the soluble fraction using nickel-affinity purification, and assessed the assembly states using native polyacrylamide gel electrophoresis (native PAGE). This showed that the majority of RbcL proteins assembled into L_8F_8 in the presence of wild-type Raf1, with a minor portion of L_2F_2 (Fig. 2f, lane 1; Supplementary Fig. 5), which was further confirmed using SEC–MALS analysis (Fig. 2g). Moreover, addition of RbcS gradually triggered the formation of higher-molecular-mass products (HMM), corresponding to the ternary complex RbcL–Raf1–RbcS (Extended Data Fig. 6a,b). By contrast, either extension of Raf1 C-tail with three extra histidine residues (termed CH3) or truncation of the C-terminal 8 residues (termed ΔC8) resulted in a substantial decrease in L_8F_8 yield, accompanied by an increase in L_2F_2 (Fig. 2f, lanes 2 and 3; Supplementary Fig. 5), which was further validated using SEC–MALS analysis (Fig. 2g). Notably, the assembly of

RbcL exhibited a smear of varying RbcL–Raf1 intermediates when coexpressing with Raf1-CH3 mutant, whereas almost all RbcL proteins assembled into L_2F_2 when coexpressing with Raf1-ΔC8 mutant (Fig. 2f, lanes 2 and 3), further confirming the role of the Raf1 C-tail in the assembly of the RbcL octamer. Despite the fact that the addition of RbcS to the coexpressed RbcL and C-terminal mutants of Raf1 still triggers the formation of HMM complexes from L_2F_2, these complexes possess a much lower composition ratio of Raf1, most of which exist as the free form (Extended Data Fig. 6a,c,d). The results suggested that the Raf1 C-tail inserting into the catalytic pocket of RbcL also contributes to the assembly of the RbcL octamer from the dimers.

Cryo-EM structures of multiple intermediates of the RbcL–Raf1–RbcS complex. Structural comparison between the L_8F_8 complex and the holoenzyme L_8S_8 revealed that Raf1α and RbcS bind to the regions around the cleft between two adjacent RbcL dimers. These two binding regions on RbcL largely overlap with each other, sharing several binding residues—Trp 71, Leu 74, Leu 75, Glu 192, Lys 228 and Asn 433 (Extended Data Fig. 7). Moreover, most of the RbcS-binding sites on RbcL, including the so-called 60s loop[28], in the structures of L_8F_8 and L_8S_8 share similar conformations (Extended Data Fig. 8). The overlapped binding regions indicate that Raf1α and RbcS most likely bind exclusively to RbcL. However, coexpression of *Anabaena* GroEL–ES, RbcL, RbcS and Raf1 in *E. coli* unexpectedly yielded a complex that contains all three components—RbcL, RbcS and Raf1—which was further confirmed using native-PAGE analysis (Supplementary Fig. 6a,b, lane LFS). To elucidate the interaction pattern in this purified complex, we solved the structures using single-particle analysis with cryo-EM. The 3D classifications revealed several distinct classes of particles, which were separately processed for 3D refinement by imposing C_4 symmetry (Supplementary Fig. 7). Eventually we obtained four structures of complexes representing the snapshots of different assembly intermediates, including $L_8F_8S_8$, $L_8F_8S_4$, L_8S_4 and the holoenzyme L_8S_8 at 3.37 Å, 3.73 Å, 3.37 Å and 3.67 Å, which account for 15.6%, 9.4%, 35.6% and 14.3% of all particles, respectively (Fig. 3a, Supplementary Fig. 8).

In the structure of $L_8F_8S_8$, the Raf1β domain of Raf1 is clearly assigned, whereas the density of Raf1α is highly dispersed and is not traceable in the map (Fig. 3b). Generally, 8 RbcL and 8 RbcS subunits assemble into L_8S_8 at the core, which was embraced by 8 Raf1β domains. Compared with the L_8F_8 structure, Raf1β in $L_8F_8S_8$ rotates ~19° as a rigid body against the equator of L_8S_8 (Fig. 3c) and binds to L_8S_8 in an asymmetric pattern. In L_8F_8, the Raf1β dimer and the linker region symmetrically position at the equator of the RbcL dimer, with an interface area of ~1,000 Å^2, whereas flipping Raf1β away from the equator makes the linker region of Raf1 move towards RbcL and form more interactions with RbcL in the structure of $L_8F_8S_8$ (Fig. 3c), yielding an interface area of ~2,000 Å^2. A similar binding pattern of Raf1 was also observed in the structure of $L_8F_8S_4$; however, only four RbcS subunits bind to one hemisphere of L_8S_8. Notably, as the two affinity tags were fused with RbcL and RbcS, respectively, there exists more particles of L_8S_4 and L_8S_8 in the tandem-purified complex. These structures of RbcL–Raf1–RbcS intermediates enabled us to observe, at the molecular level, sequential displacement of Raf1 by RbcS, suggesting that the assembly process of RuBisCO is dynamic.

Raf1 is a major individual chaperone that assists the assembly of the RuBisCO holoenzyme. In addition to Raf1, other chaperones, such as GroEL–ES and RbcX, are required to assist the assembly of cyanobacterial RuBisCO holoenzyme[4,11,34,35]. To compare these chaperones in the process of *Anabaena* sp. PCC 7120 RuBisCO assembly, we coexpressed the His-tagged RbcL or holoenzyme L_8S_8 in the presence of various combinations of chaperones in *E. coli*.

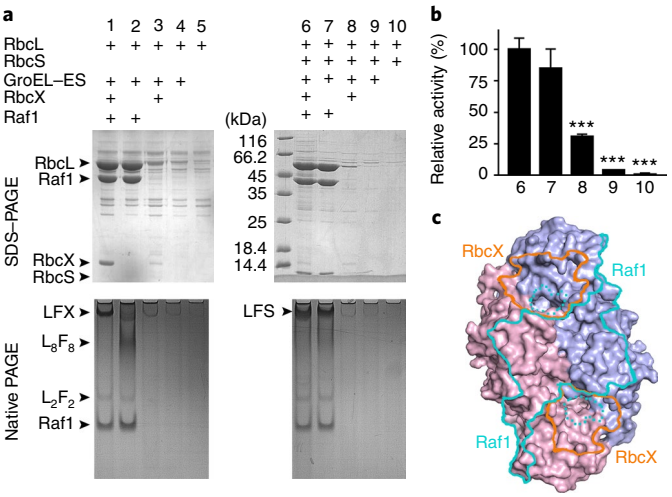


Fig. 4 | Assembly of RuBisCO assisted by Raf1 and other chaperones. **a**, Coexpression of RbcL or RbcL and RbcS in *E. coli* with various combinations of chaperones. The results were analysed using SDS–PAGE and native PAGE. The soluble fractions that contain the same amount of total proteins were applied to nickel-affinity purification. The eluted proteins of 6 ml were then concentrated to a final volume of 400 µl. Each sample of 10 µl was loaded into corresponding lanes of SDS–PAGE and native-PAGE gels. The assembly intermediates are indicated to the left of the native PAGE, in which LFX and LFS represent the ternary complexes of RbcL–Raf1–RbcX and RbcL–Raf1–RbcS, respectively. **b**, The relative activities corresponding to RuBisCO coexpressed with different chaperones in lanes 6–10 of **a**. Statistical analysis was performed using two-tailed Student's *t*-tests; ***$P < 0.001$. **c**, RbcX and Raf1 bind to different regions on RbcL. The subunits of the RbcL dimer are shown as blue and pink surface. The Raf1- and RbcX-binding regions are indicated by cyan and orange circles, respectively.

After nickel-affinity purification, the yield of soluble RbcL was estimated using denaturing polyacrylamide gel electrophoresis (SDS–PAGE) and the assembly states were assessed using native PAGE. Coexpression of RbcL with all of these chaperones gave a high yield of HMM complexes (labelled with LFX) in native PAGE (Fig. 4a, lane 1), which contains RbcL, Raf1 and RbcX, as confirmed by SDS–PAGE (Supplementary Fig. 9). This indicated that RbcL could be properly folded and assembled under the coordinated assistance of this full combination of tested chaperones. Using this full combination as a positive control, a similar amount of RbcL, most of which in the form of L_8F_8, was yielded in the absence of RbcX (Fig. 4a, lane 2; Supplementary Fig. 9), consistent with a previous report[16]. By contrast, only a small amount of RbcL was obtained in the supernatant in the absence of Raf1 (Fig. 4a, lane 3), suggesting that Raf1 is a major individual chaperone for RbcL folding. Moreover, removal of both Raf1 and RbcX resulted in no detectable soluble RbcL (Fig. 4a, lane 4), similar to the combination without any chaperones (Fig. 4a, lane 5).

Similarly, co-expressing RbcL and RbcS with the full combination of tested chaperones in *E. coli* resulted in a large amount of HMM complexes that contain RbcL, RbcS and Raf1 (Fig. 4a, LFS, lane 6; Supplementary Fig. 9). A comparable amount of HMM complex was also observed when coexpressing with the combination of GroEL–ES and Raf1 (Fig. 4a, lane 7; Supplementary Fig. 9), whereas almost no RbcL/RbcS or the HMM complex could be detected in the absence of Raf1 (Fig. 4a, lanes 8 and 9).

To further detect whether the recombinant RuBisCO was enzymatically active, we performed activity assays with the crude extract from *E. coli*. Coexpressing RuBisCO in the presence of all

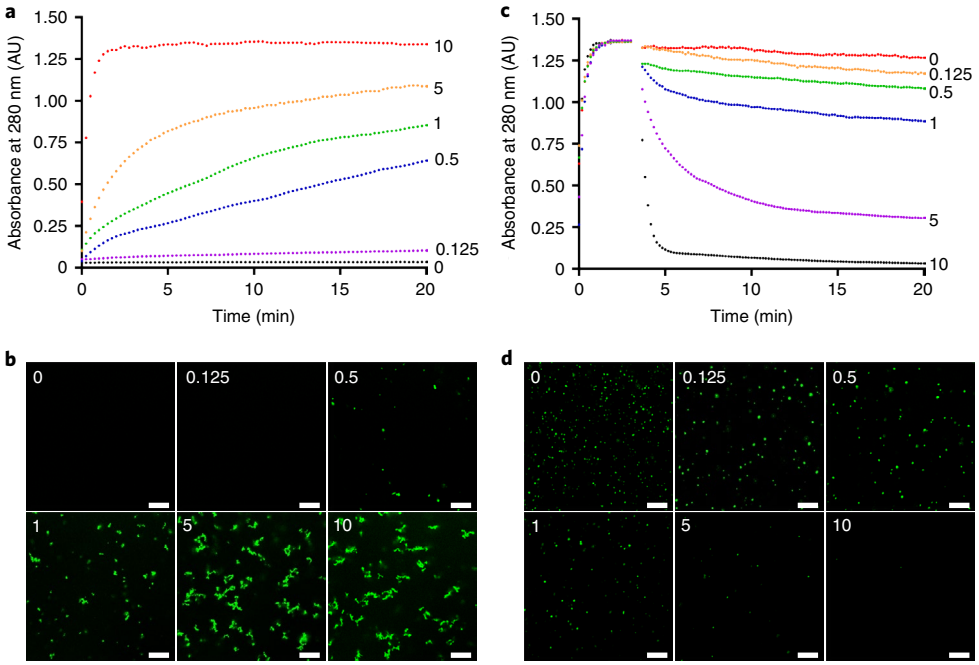


Fig. 5 | Raf1 antagonizes CcmM35-mediated condensation of RuBisCO. All of the assays were performed using the *S. elongatus* PCC 7942 RbcL proteins with N-terminally fused eGFP (eGFP–RbcL). **a**, Plots of condensate formation from L_8F_8 pre-incubated with RbcS, after the addition of CcmM35. RbcS at various concentrations (0 μM, 0.25 μM, 1 μM, 2 μM, 10 μM and 20 μM), corresponding to a molar ratio of 0, 0.125, 0.5, 1, 5 and 10 RbcS to eGFP–RbcL, respectively, was added to 0.25 μM L_8F_8 and incubated for 30 min; 2 μM CcmM35 was then added, and the turbidity was monitored at 340 nm. **b**, Confocal microscopy images of the final condensates corresponding to **a**. Scale bars, 20 μm. **c**, Plots of RuBisCO condensate solubilization after adding Raf1. The RuBisCO condensates were formed by incubating 0.25 μM L_8S_8 and 2 μM CcmM35 for 3 min. Raf1 at various concentrations (0 μM, 0.25 μM, 1 μM, 2 μM, 10 μM and 20 μM), corresponding to a molar ratio of 0, 0.125, 0.5, 1, 5 and 10 Raf1 to eGFP–RbcL, respectively, was then added to the turbid solution; turbidity was then monitored at 340 nm over time. **d**, Confocal microscopy images of the final condensates corresponding to **c**. Scale bars, 20 μm.

chaperones exhibits a maximal activity (Fig. 4b, column 6). Removal of RbcX resulted in a slight decrease in activity (Fig. 4b, column 7). By contrast, the absence of Raf1 led to a ~60% decrease in activity (Fig. 4b, column 8). These data suggest that both Raf1 and RbcX are necessary for the proper assembly and full activity of RuBisCO; however, Raf1 exhibits a more substantial contribution.

To elucidate the structural basis for the different contributions, we compared the binding surfaces of Raf1 and RbcX on RbcL on the basis of our L_8F_8 structure and the previously reported chimeric $RbcL_8(RbcX_2)_8$ structure[15]. In the L_8F_8 structure, each Raf1 dimer covers a large portion of exposed surface on RbcL (Fig. 4c). However, in the $RbcL_8(RbcX_2)_8$ structure, each RbcX dimer covers a rather smaller patch crossing the interface of RbcL dimer (Fig. 4c). Distinct from RbcX, which stabilizes only one RbcL dimer, Raf1 also bridges two neighbouring RbcL dimers through interface III (Fig. 1f). Notably, Raf1 covers a surface region on RbcL that is different from those of two RbcX dimers, suggesting that the two chaperones could simultaneously bind to RbcL. Indeed, coexpression of RbcL, GroEL–ES, Raf1 and RbcX yielded a ternary complex of RbcL–Raf1–RbcX, as confirmed by native PAGE (Supplementary Fig. 6a,b, lane LFX). In addition to cooperating with RbcX in RbcL folding, Raf1 is the majority contributor to the assembly of RuBisCO, especially from RbcL dimers to octamers, in agreement with our biochemical assays that coexpressing with Raf1 yields more soluble RbcL (Fig. 4a) and active RuBisCO (Fig. 4b) compared with that with RbcX.

In vitro condensation of RuBisCO mediated by CcmM35 could be antagonized by Raf1. The cyanobacterial RuBisCO holoenzymes L_8S_8 can further form condensates that are cross-linked by the scaffold protein CcmM35 (refs. [19,36]). In *S. elongatus* PCC 7942, the *ccmM* gene produces both a full-length CcmM58 and a truncated CcmM35 protein by alternative translation initiation[20]. CcmM35 possesses three RuBisCO small-subunit-like (SSUL) modules, whereas CcmM58 contains an extra carbonic-anhydrase-like domain at the N terminus. The cryo-EM structure of *S. elongatus* PCC 7942 RuBisCO in complex with the first SSUL module (SSUL1) revealed that SSUL1 does not replace RbcS, but binds to the equatorial region of RuBisCO between two neighbouring RbcL dimers[36,37]. Structural analysis showed that Raf1α and SSUL1 possess a slightly overlapped binding region on RbcL (Extended Data Fig. 9); however, simultaneous binding of Raf1α and SSUL to RbcL is impossible owing to steric hindrance.

To test this hypothesis, we applied in vitro turbidity assays, using RbcL proteins fused to enhanced green fluorescent protein (eGFP) on the N terminus, to compare CcmM35-mediated condensation of RuBisCO in the presence or absence of Raf1. After pre-incubating the L_8F_8 complex with various ratios of RbcS in molarity, addition of equal molarity of CcmM35 in the solution triggers a substantial increase in turbidity (Fig. 5a, Supplementary Fig. 10a). With more RbcS present in the pre-incubated solution, the turbidity increases at a much faster velocity and, ultimately, reaches the maximum with a half time $t_{1/2}$ of about 15 s at a molar ratio of RbcS to RbcL of 10:1 (Fig. 5a). Notably, in the absence of RbcS, CcmM35 could not trigger the condensation of L_8F_8, which is reasonable from the viewpoint of physiology to avoid the condensation of immature RuBisCO. Moreover, the confocal fluorescence spectroscopy assays also confirmed that addition of RbcS eventually triggers the formation of RuBisCO condensates in the presence of CcmM35 (Fig. 5b), which is most likely due to the gradual displacement of Raf1α by RbcS from RbcL. These results were also in agreement with our cryo-EM structures of multiple intermediates in the process of RuBisCO assembly.

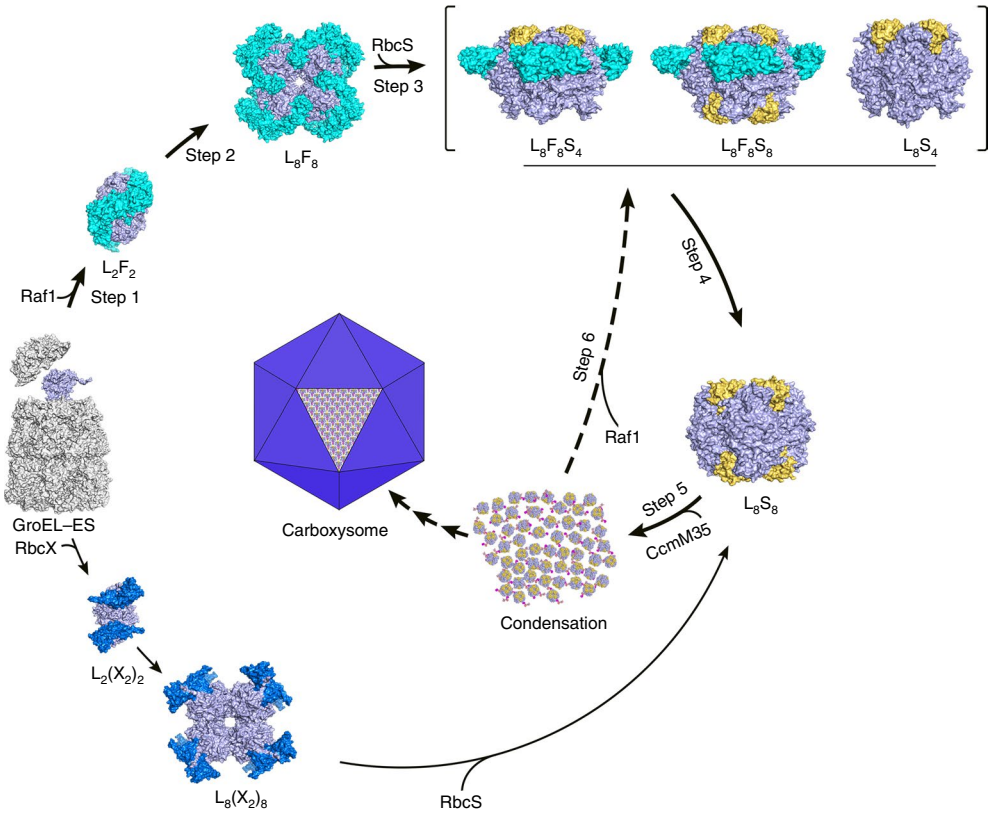


Fig. 6 | A putative model of chaperone-assisted assembly of cyanobacterial RuBisCO. The dissected steps of Raf1-mediated assembly and condensation of RuBisCO are sequentially labelled. The previously identified process of RbcX-assisted RuBisCO assembly is also included in the model. The proteins RbcL, RbcS, Raf1, RbcX and GroEL–ES are shown as surfaces and are coloured in light blue, yellow, cyan, dark blue and grey, respectively. The putative step 6 is indicated by the dashed line.

Given the high affinity of Raf1 towards RbcL and the dynamic interactions among RbcL, Raf1 and RbcS, we tested whether the RuBisCO condensates could be solubilized by Raf1 in vitro. As predicted, the addition of Raf1 triggers the gradual disassembly of the condensates, resulting in a decrease in turbidity over time (Fig. 5c, Supplementary Fig. 10b). The more Raf1 that was added, the faster the condensates were solubilized. After adding tenfold Raf1 in molarity to that of RbcL, the solution became clear in about 20 min (Fig. 5c), indicating that the RuBisCO condensates were almost completely solubilized. The results of confocal fluorescence spectroscopy also showed the in vitro solubilization of RuBisCO condensates by Raf1 (Fig. 5d). As the binding region of Raf1α on RbcL is largely overlapped to that of RbcS (Extended Data Fig. 6), compared with the slightly overlapped region with the SSUL domain (Extended Data Fig. 9), displacement of RbcS from RbcL most likely contributes a majority to the antagonism of CcmM35-mediated RuBisCO condensation.

These in vitro experiments indicate that Raf1 might not only be a fine modulator that controls the condensation velocity of RuBisCO holoenzymes, but also a solubilizer that disassembles the condensates of RuBisCO. As the cyanobacterial condensates of RuBisCO are usually encapsulated in the carboxysome, we hypothesized that Raf1 might also participate in the disassembly of RuBisCO condensates during the dismantling of carboxysomes. In fact, a previous report showed that *T. elongatus* Raf1 could even dissociate the RbcL octamer and form the RbcL–Raf1 intermediate[18]. It was also found that *Synechocystis* sp. PCC 6803 Raf1 participates in the RuBisCO homeostasis, as *raf1* deletion showed a decreased susceptibility under sulfur depletion, accompanied with slower degradation of RbcL[25]. However, the physiological function of Raf1 on the condensation/disassembly of RuBisCO remains unclear.

Discussion

It was previously proposed that the assembly of the RuBisCO holoenzyme generally consists of two stages—the formation of an RbcL octameric core followed by docking of RbcS subunits[4,38]. The fine mechanism for the RbcX-mediated assembly of RuBisCO holoenzyme has been proposed on the basis of a series of reports[13–15]. However, the dissected steps and molecular mechanism driven by the major individual chaperone Raf1 remain limited. Our findings here in combination with previous reports[13–16,36] enabled us to propose a multiple-step model of Raf1-assisted assembly of cyanobacterial RuBisCO (Fig. 6). At the very beginning, the folding of nascent RbcL polypeptide is initiated by the general chaperonin system GroEL–ES[12]. Capture of the flexible C-terminal tail of RbcL by RbcX was proposed to facilitate the escape of RbcL from the GroEL–ES cycle[13–15]. In fact, our L_8F_8 structure showed that Raf1 also interacts with the C-terminal tail of RbcL through several hydrogen bonds as well as van der Waals contacts (Extended Data Fig. 10). Thus, we propose that both Raf1 and RbcX participate in the release of folded RbcL subunit from GroEL–ES chaperonin. Once released, two subunits form an antiparallel RbcL dimer, which is embraced by the tweezer-like Raf1 dimer, with the two-fold axis of the swapped Raf1β dimer perfectly aligned with the two-fold axis of RbcL dimer (Fig. 6, step 1). The convex surfaces of Raf1α at interface III, in addition to the Raf1 C-tail inserting into the catalytic pocket of RbcL, then further mediate the tetramerization of L_2F_2 to form the L_8F_8 complex (Fig. 6, step 2). During the

recruitment of RbcS to the octameric RbcL core, we observed a series of intermediates, including $L_8F_8S_8$, $L_8F_8S_4$ and L_8S_4 (Fig. 6, step 3). This dynamic step is most likely initiated by the replacement of Raf1α by RbcS, followed by the dissociation of Raf1β domain from RbcL and, eventually, stops at the formation of holoenzyme L_8S_8 (Fig. 6, step 4). Notably, our biochemical and structural data showed that, compared with RbcX, Raf1 is a more efficient chaperone for RuBisCO holoenzyme formation, consistent with previous research showing that the deletion of RbcX in *S. elongatus* PCC 7942 had no detectable effect on RuBisCO production[39]. Along with the intracellular accumulation of RuBisCO holoenzymes, CcmM35 mediates the condensation of RuBisCO (Fig. 6, step 5) using the SSUL modules as the cross-linker[36]. Usually, these RuBisCO condensates are further stacked against each other to constitute the inner core of carboxysome. Notably, our in vitro assays suggested that, as a putative antagonist of CcmM35, Raf1 might be able to regulate the condensation rate or even reverse the condensation process (Fig. 6, step 6). However, more investigations are needed to elucidate the in vivo regulation of Raf1 on the RuBisCO condensation and/or carboxysome biogenesis.

In summary, here we present the crystal structures of the full-length Raf1 and its complex with RbcL, in addition to the cryo-EM structures of a series of assembly intermediates that eventually form the RuBisCO holoenzyme. Structural analysis combined with biochemical assays enabled us to figure out the fine molecular mechanism of Raf1-assisted cyanobacterial RuBisCO assembly.

Methods

Cloning, plasmids and strains. The genes encoding RbcL, RbcS, GroEL–ES, Raf1, RbcX and CcmM35 were amplified by PCR from the cyanobacteria *Anabaena* sp. PCC 7120 or *S. elongatus* PCC 7942, and were cloned into the pET19 and/or pCDFduet vectors using homologous recombination methods. A list of cyanobacterial strains, plasmids and protein sequences used in this study is provided in Supplementary Table 2.

Protein expression and purification. The pET19-His-Raf1 (*Anabaena* sp. PCC 7120) plasmid was transformed into *E. coli* strain BL21 cells (Novagen), which were precultured in 20 ml Luria-Bertani (LB) medium (10 g l^{-1} NaCl, 10 g l^{-1} Bacto Tryptone and 5 g l^{-1} yeast extract) overnight at 37 °C. The precultured cells were transferred to 1 l LB medium containing ampicillin of 50 μg ml^{-1} for further growing. When the cells reached an optical density at 600 nm (OD_{600}) of ~0.8, protein expression was induced by adding isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 0.2 mM. After growing at 16 °C for ~20 h, the cell pellets were collected and resuspended in 50 ml buffer A (20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 5% glycerol). After lysis using an Ultrasonic Cell Disruptor (Sonics), the cell lysate was cleared by centrifugation at 12,000*g* for 30 min and the supernatant was loaded onto a 5-ml Ni-NTA column (Qiagen) that was pre-equilibrated with buffer A. The target proteins were eluted with buffer A containing 500 mM imidazole, and further purified using size-exclusion chromatography (Superdex 200, GE Healthcare). The peak fractions containing Raf1 were collected by monitoring the absorbance at 280 nm, and concentrated to 20 mg ml^{-1} for crystallization. A part of the purified protein was flash-frozen in liquid nitrogen and stored at −80 °C for further use. The recombinant RbcS proteins were expressed with the pET19-His-RbcS plasmid in *E. coli* strain BL21 cells, and were purified using a method that was similar to the method used for Raf1. All protein concentrations were determined using a NanoDrop (Thermo Fisher Scientific) and the purity was assessed using denaturing polyacrylamide gel electrophoresis (SDS–PAGE).

Coexpression of *Anabaena* sp. PCC 7120 RbcL–Raf1 complex (termed L_8F_8 for short) was performed by cotransforming the plasmids of pET19-His-RbcL-Raf1 and pCDFduet-GroEL-GroES into *E. coli* strain BL21 cells. The cell pellets of 3 l culture were resuspended in 50 ml buffer B (50 mM Tris-HCl, pH 8.0, 20 mM NaCl, 5 mM $MgCl_2$ and 5 mM phenylmethylsulfonyl fluoride). The procedures for protein expression and purification of L_8F_8 were the same as those described for Raf1. The concentrated proteins at 10 mg ml^{-1} supplemented with 10% glycerol were flash-frozen in liquid nitrogen and stored at −80 °C for further use.

The complex of *Anabaena* sp. PCC 7120 RbcL–Raf1–RbcS was obtained by cotransforming the plasmids of pET19-His-RbcL-Raf1 and pCDFduet-GroEL-GroES-FLAG-RbcS into *E. coli* strain BL21 cells. The cells in 1 l culture were collected by centrifugation at 8,000*g* for 10 min and resuspended in 50 ml buffer C (20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 5 mM $MgCl_2$). The cells were disrupted by sonication for 30 min, followed by centrifugation at 12,000*g* for 30 min. The clarified cell supernatant was then pooled and processed for tandem-affinity purification. The supernatant fraction was first loaded onto a 5 ml Ni-NTA column (Qiagen) that was pre-equilibrated with buffer C. The target complex of RbcL–RbcS–Raf1 was eluted with 300 mM imidazole, and further incubated with the anti-Flag M2 affinity gel (Sigma) on ice for 1 h. The resin was then washed five times, each with 5 ml of buffer C. The target protein complex that was eluted with the buffer containing 200 μg ml^{-1} Flag peptide was further purified by size-exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Healthcare). The peak fraction was pooled and frozen in liquid nitrogen or concentrated for cryo-EM analysis.

The ternary complex of *Anabaena* sp. PCC 7120 RbcL–Raf1–RbcX was obtained by co-transforming the plasmids of pET19-His-RbcL and pCDFduet-GroEL-GroES-Raf1-RbcX into *E. coli* strain BL21 cells. The procedures of cell culture, protein expression and purification were as described for the L_8F_8 complex.

The *S. elongatus* PCC 7942 RuBisCO holoenzyme was expressed in *E. coli* strain BL21 cells with transformed plasmids of pET19-His-RbcL-Raf1 and pCDFduet-GroEL-GroES-RbcS. The procedures of cell culture and protein expression were the same as those described for *Anabaena* sp. PCC 7120 Raf1. The RuBisCO holoenzyme was purified using ion-exchange chromatography (Hitrap QHP, GE Healthcare) followed by size-exclusion chromatography (Superdex 200 Increase 10/300 column, GE Healthcare). The *S. elongatus* PCC 7942 RuBisCO fused with eGFP on the N terminus of RbcL was expressed in *E. coli* strain BL21 cells by cotransforming the plasmids of pET19-His-eGFP-RbcL-Raf1 and pCDFduet-GroEL-GroES. The protein was purified using a similar method to that used for RuBisCO. The *S. elongatus* PCC 7942 eGFP–RbcL in complex with Raf1 was expressed by cotransforming the plasmids of pET19-His-eGFP-RbcL-Raf1 and pCDFduet-GroEL-GroES into *E. coli* strain BL21 cells. The procedures for protein expression and purification were the same as those for *Anabaena* sp. PCC 7120 L_8F_8.

The recombinant proteins of *S. elongatus* PCC 7942 RbcS, Raf1 and CcmM35 were expressed in *E. coli* strain BL21 cells by transforming the individual plasmids pET19-His-RbcS, pET19-His-Raf1 or pET19-His-CcmM35, respectively. The procedures for cell culture, protein expression and purification were the same as those described for *Anabaena* sp. PCC 7120 Raf1.

Crystallization, data collection and processing. Crystals of *Anabaena* sp. PCC 7120 full-length Raf1 were grown at 14 °C using the hanging-drop vapour-diffusion method, with a drop of 1 μl protein at ~20 mg ml^{-1} mixed with an equal volume of reservoir solution. Crystals were initially obtained against the reservoir solution containing 2.4 M sodium acetate trihydrate, pH 7.0. After several rounds of optimization, crystals grown in 1.6–2.0 M sodium acetate trihydrate (pH 7.0) were applied to X-ray diffraction. Crystals of L_8F_8 were grown at 14 °C by mixing 1.5 μl protein sample at ~10 mg ml^{-1} with 1.5 μl reservoir solution containing 8~10% 2-methyl-2,4-pentanediol and 0.1 M 2-ethanesulfonic acid (pH 6.0). All of the crystals were transferred to the cryoprotectant containing the mother liquor supplemented with 35% (v/v) glycerol, and were then flash-frozen in the liquid nitrogen. X-ray diffraction data were collected at 100 K in a liquid nitrogen stream using beamlines BL17U and BL19U at the Shanghai Synchrotron Radiation Facility (SSRF)[40]. The diffraction data of Raf1 were integrated and scaled using XDS[41], whereas the L_8F_8 data were processed using HKL2000 (ref. [42]). Crystals of Raf1 and L_8F_8 belong to the space groups of $P3_221$ and $C222_1$, respectively.

Crystal structure determination and refinement. The crystal structure of Raf1 was solved by molecular replacement with Molrep in CCP4i[43] using structures of Raf1α (PDB, 4WT3) and Raf1β (PDB, 4WT4) from *A. thaliana* as the search models. After individually placing Raf1α and Raf1β by molecular replacement, the calculated difference map showed extra electron density, which enabled us to manually build the linker region. After several rounds of manual building using Coot[44] with refinement in REFMAC5 (ref. [45]) implemented in CCP4i[43], the model was greatly improved and perfectly matches the map, in which R_{factor} and R_{free} values are converged to 22.4% and 25.7%, respectively. The structure of L_8F_8 was also determined by molecular replacement using the structures of RbcL (PDB, 1RBL) from *Synechococcus* sp. PCC 6301 and two Raf1 domains from *Anabaena* sp. PCC 7120. All of the structures were refined using REFMAC5 in CCP4i and then rebuilt interactively using Coot. The final models showed well geometry and were evaluated using MolProbity[46] (http://molprobity.biochem.duke.edu). The interface areas were calculated using PDBePISA[47]. All of the structure figures were generated using PyMOL (https://pymol.org/2/). A list of the parameters of data collection, processing, structure determination and refinement is provided in Supplementary Table 3.

Cryo-EM sample preparation, data collection and processing. The purified sample containing *Anabaena* sp. PCC 7120 RbcL–Raf1–RbcS complex was concentrated to ~1.6 mg ml^{-1}. An aliquot of 3.5 μl sample was applied to a glow-discharged Quantifoil R1.2/1.3 300-mesh Cu Holey Carbon Grids. The grids were blotted for 5 s with a blot force of 2 s and a wait time of 20 s, and were then plunged into liquid ethane using a Vitrobot Mark IV (FEI) at 4 °C and 100% humidity. The Cryo-EM datasets were collected using a 300 keV Titan Krios electron microscope (FEI) at the Institute of Biophysics (IBP), Chinese Academy

of Sciences (CAS). A total of 4,284 video stacks (32 frames, each 0.17 s, 9 e $Å^{-2}$ s^{-1}, total dose ~50 e $Å^{-2}$) were recorded using an electron-counting K2 detector (Gatan) in super-resolution mode at a nominal magnification of ×22,500 with a defocus range from −1.5 µm to −2.0 µm. All of the video stacks were motion-corrected and dose-weighted using MotionCor2 (ref. [48]), and were binned two-fold to yield a pixel size of 1.04 Å. The defocus values were estimated using CTFFIND4 (ref. [49]).

After removing the bad images by manual checking, a total of 3,640 images were processed for particle picking using RELION[50]. Approximately 1,400,000 particles were boxed and binned four-fold for 2D classification. All good classes containing 957,600 good particles were selected for the first round 3D classification with six classes by C_1 symmetry. The dominant class of ~50% particles represented the complex of RbcL–RbcS, whereas other two classes showed clear features of the ternary complex of RbcL–RbcS–Raf1, which were combined together and binned two-fold for the second round 3D classification with four classes by imposing C_4 symmetry. Among the four classes, two of them represented the complex of $L_8F_8S_8$, and account for 15.6% of the good particles, which were combined together for the final 3D refinement, yielding a resolution of up to 3.37 Å after PostProcess, as determined by Golden standard Fourier shell correlation using the 0.143 threshold. Another class of 9.4% particles represented the complex of $L_8F_8S_4$, which was refined to 3.73 Å. Moreover, the dominant class of the complex RbcL–RbcS was processed for the second round of 3D classification by imposing C_4 symmetry. Finally, two classes, which represented the complexes of L_8S_4 (35.6%) and L_8S_8 (14.3%), were further refined to 3.37 Å and 3.67 Å, respectively.

Model building of $L_8F_8S_8$ and $L_8F_8S_4$ was performed using Chimera[51] by manually fitting the L_8F_8 structure into the map. The Raf1α domains were not modelled in the final structure due to the dispersed density. Moreover, the Raf1β domains were manually adjusted to best fit the map. The RbcS structure was then manually fitted into the extra density of the map. Similarly, the structures of L_8S_8 and L_8S_4 were manually built using Chimera by fitting the *Synechococcus* sp. PCC 6301 L_8S_8 structure (PDB, 1RBL) into the map. All of the models were then manually refined using Coot[44], followed by the iterative positional and *B*-factor refinement in real space using PHENIX[52]. The final structures showed good geometry and were further evaluated using MolProbity[46] (http://molprobity.biochem.duke.edu). A list of the parameters of cryo-EM data collection, processing, structure determination and refinement is provided in Supplementary Table 4.

SEC–MALS analysis. The complexes of RbcL–Raf1 (wild-type Raf1 or mutants), RbcL–Raf1–RbcX and RuBisCO were analysed using static light scattering using the AKTA Pure system (GE Healthcare). The sample was loaded onto a SEC column (Superdex 200 10/300 GL column, GE Healthcare) at a flow rate of 0.4 ml min^{-1}. The system was coupled with an eight-angle static light scattering detector (DAWN HELEOS II, Wyatt Technology) and a differential refractometer (Optilab T-rEX, Wyatt Technology). The molecular mass of the complex was calculated using ASTRA v.7.0.1.

Coexpression of RbcL or L_8S_8 with different chaperones in *E. coli*. Coexpression of *Anabaena* sp. PCC 7120 RbcL with wild-type Raf1 or mutants (CH3 and ΔC8) was performed in *E. coli* strain BL21 by cotransforming the plasmid pCDFduet-GroEL-GroES with the plasmid pET19-His-RbcL-Raf1, pET19-His-RbcL-Raf1-CH3 or pET19-His-RbcL-Raf1-ΔC8. Protein expression was induced by adding 0.2 mM IPTG at 16 °C for 20 h. Then, 1 l cells at OD_{600} of 1.6 were collected and lysed using ultrasonication. After centrifugation at 12,000*g* for 30 min at 4 °C, the total protein yields in the soluble fractions were quantified using the BCA Protein Assay Kit (Boster), and were further normalized for the following assays. The yield of soluble His-tagged RbcL proteins was detected by immunoblotting with anti-His-tag antibodies (Proteintech).

Co-expression of *Anabaena* sp. PCC 7120 RbcL or RbcL–RbcS with different chaperones was conducted in the *E. coli* strain BL21 (DE3) cells by transforming the corresponding combinations of plasmids. Supernatants containing the same amounts of total proteins were incubated in a 5 ml Ni-NTA column (GE Healthcare) on ice for 1 h. The column was then washed three times, each with 20 ml of buffer B. Protein was eluted with 6 ml of wash buffer plus 300 mM imidazole. The eluted proteins were concentrated to a final volume of 400 µl using a 100-kDa cut-off Centricon (Millipore). Each sample of 10 µl was then mixed with the loading buffer and analysed using SDS–PAGE (15% Tris-HCl) and native PAGE (6% Bis-Tris and boric acid). The protein compositions in the corresponding bands of native PAGE were further analysed using SDS–PAGE.

The RuBisCO activity assays. The RuBisCO activity assays were performed using the supernatants in the cell lysate, which were prepared using the same method as described for the coexpression of RuBisCO with different chaperones. The amounts of total proteins in the supernatant were quantified and further normalized using the BCA Protein Assay Kit (Boster). The 20 µl supernatant was incubated with 225 µl activation buffer (50 mM Tris-HCl, pH 8.2, 20 mM $MgCl_2$, 1 mM EDTA, 40 mM $NaHCO_3$) for 20 min at 25 °C. The reaction was initiated after adding the substrates of 1 mM RuBP and 0.24 mM radioactive $NaH^{14}CO_3$. The reactions lasted for 1 min, and were terminated by adding 10% (v/v) formic acid. The reaction mixture was then dried at 95 °C to remove the unfixed carbon. The pellet was resuspended by adding 500 µl distilled water, and was transferred to a 5 ml scintillation solution. The amount of fixed carbon was quantified using a HITACHI AccuFLEX LSC-8000 scintillation counter.

Turbidimetric assays. The turbidity of the solution was measured by monitoring the absorbance at 340 nm using a Beckman DU800 spectrophotometer. All of the assays were performed using *S. elongatus* PCC 7942 RbcL proteins fused to eGFP at the N terminus. The eGFP–RuBisCO and eGFP–L_8F_8 proteins were expressed and purified using the same methods as those described for *S. elongatus* PCC 7942 RuBisCO and L_8F_8, respectively. The purified eGFP–RuBisCO, eGFP–L_8F_8, CcmM35, Raf1 and RbcS were concentrated to 10 mg ml^{-1}, 16 mg ml^{-1}, 6 mg ml^{-1}, 20 mg ml^{-1} and 5 mg ml^{-1}, respectively, for the assays. All of the measurements were performed at 25 °C in buffer D (20 mM Tris-HCl, pH 8.0, 50 mM NaCl and 5 mM $MgCl_2$). Two groups of experiments were designed to detect the effect of Raf1 on RuBisCO condensates formation. First, 0.25 µM eGFP–L_8F_8 was added to the solution containing RbcS at various concentrations (0 µM, 0.25 µM, 1 µM, 2 µM, 10 µM and 20 µM). After incubation for about 30 min, 2 µM CcmM35 was added to the solution, and the absorbance at 340 nm was monitored. Second, RuBisCO condensation was triggered by mixing 0.25 µM eGFP–RuBisCO and 2 µM CcmM35. The turbidity was monitored at 340 nm over time. After the condensates reached the maximum turbidity in ~3 min, Raf1 at various concentrations (0 µM, 0.25 µM, 1 µM, 2 µM, 10 µM and 20 µM) was added to the solution, and turbidity was detected at 340 nm. The purity and stability of proteins in all of the reaction mixtures were analysed using SDS–PAGE.

Laser-scanning confocal microscopy. Two groups of condensation experiments were conducted in the same manner as described for the turbidimetric assays. The reaction mixtures containing the eGFP–RbcL proteins were imaged using a laser-scanning confocal microscope (ZEISS LSM710). The 20 µl samples were transferred to an uncoated chambered coverslip and excited with a laser at 488 nm (laser intensity, 15%) for fluorescence imaging. Images were recorded by focusing on the bottom of the plate using Axio Observer Z1 microscope with a Plan-Apochromat ×63/1.40 NA oil-immersion DIC M27 objective.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The structural factors and atomic coordinates of Raf1 and its complex with RbcL have been deposited at PDB (Raf1, 6KKN; L_8F_8, 6KKM). The cryo-EM structures of $L_8F_8S_8$ and L_8S_4 have been deposited at PDB ($L_8F_8S_8$, 6LRR; L_8S_4, 6LRS). The cryo-EM density maps of $L_8F_8S_8$, L_8S_4, $L_8F_8S_4$ and L_8S_8 have been deposited at the Electron Microscopy Data Bank (EMDB-0959–EMDB-0962, respectively). Source Data for Figs. 2 and 4–6 are provided with the paper.

Received: 18 September 2019; Accepted: 14 April 2020;
Published online: 25 May 2020

References

1. Bar-On, Y. M. & Milo, R. The global mass and average rate of RuBisCO. *Proc. Natl Acad. Sci. USA* **116**, 4738–4743 (2019).
2. Ellis, R. J. The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244 (1979).
3. Bracher, A., Whitney, S. M., Hartl, F. U. & Hayer-Hartl, M. Biogenesis and metabolic maintenance of RuBisCO. *Ann. Rev. Plant Biol.* **68**, 29–60 (2017).
4. Andersson, I. & Backlund, A. Structure and function of RuBisCO. *Plant Physiol. Biochem.* **46**, 275–291 (2008).
5. Whitney, S. M., Houtz, R. L. & Alonso, H. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, RuBisCO. *Plant Physiol.* **155**, 27–35 (2011).
6. Parry, M. A. et al. RuBisCO activity and regulation as targets for crop improvement. *J. Exp. Bot.* **64**, 717–730 (2013).
7. 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 (2014).
8. Erb, T. J. & Zarzycki, J. Biochemical and synthetic biology approaches to improve photosynthetic CO_2-fixation. *Curr. Opin. Chem. Biol.* **34**, 72–79 (2016).
9. Sharwood, R. E. Engineering chloroplasts to improve RuBisCO catalysis: prospects for translating improvements into food and fiber crops. *New Phytol.* **213**, 494–510 (2017).
10. Hauser, T., Popilka, L., Hartl, F. U. & Hayer-Hartl, M. Role of auxiliary proteins in RuBisCO biogenesis and function. *Nat. Plants* **1**, 15065 (2015).
11. Wilson, R. H. & Hayer-Hartl, M. Complex chaperone dependence of RuBisCO biogenesis. *Biochemistry* **57**, 3210–3216 (2018).
12. Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44–47 (1989).

13. Saschenbrecker, S. et al. Structure and function of RbcX, an assembly chaperone for hexadecameric RuBisCO. *Cell* **129**, 1189–1200 (2007).
14. Liu, C. et al. Coupled chaperone action in folding and assembly of hexadecameric RuBisCO. *Nature* **463**, 197–202 (2010).
15. Bracher, A., Starling-Windhof, A., Hartl, F. U. & Hayer-Hartl, M. Crystal structure of a chaperone-bound assembly intermediate of form I RuBisCO. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011).
16. Hauser, T. et al. Structure and mechanism of the RuBisCO-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22**, 720–728 (2015).
17. Andrews, T. J. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* **263**, 12213–12219 (1988).
18. Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M. & Szczepaniak, A. RuBisCO accumulation factor 1 from *Thermosynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and in vitro. *FEBS J.* **281**, 3920–3932 (2014).
19. Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
20. Long, B. M., Tucker, L., Badger, M. R. & Price, G. D. Functional cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol.* **153**, 285–293 (2010).
21. Turmo, A., Gonzalez-Esquer, C. R. & Kerfeld, C. A. Carboxysomes: metabolic modules for CO_2 fixation. *FEMS Microbiol. Lett.* **364**, fnx176 (2017).
22. Aigner, H. et al. Plant RuBisCO assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
23. Feiz, L. et al. Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor 1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
24. Salesse-Smith, C. E. et al. Overexpression of RuBisCO subunits with RAF1 increases RuBisCO content in maize. *Nat. Plants* **4**, 802–810 (2018).
25. Kolesinski, P., Rydzy, M. & Szczepaniak, A. Is RAF1 protein from *Synechocystis sp*. PCC 6803 really needed in the cyanobacterial RuBisCO assembly process? *Photosynth. Res.* **132**, 135–148 (2017).
26. Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C. & Lorimer, G. H. Mechanism of RuBisCO: the carbamate as general base. *Chem. Rev.* **98**, 549–562 (1998).
27. Newman, J., Branden, C. I. & Jones, T. A. Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus* PCC 6301. *Acta Crystallogr. D* **49**, 548–560 (1993).
28. 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 (2000).
29. Whitney, S. M., Birch, R., Kelso, C., Beck, J. L. & Kapralov, M. V. Improving recombinant RuBisCO biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc. Natl Acad. Sci. USA* **112**, 3564–3569 (2015).
30. Taylor, T. C. & Andersson, I. Structure of a product complex of *spinach* ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry* **36**, 4041–4046 (1997).
31. Taylor, T. C. & Andersson, I. The structure of the complex between RuBisCO and its natural substrate ribulose 1,5-bisphosphate. *J. Mol. Biol.* **265**, 432–444 (1997).
32. Hartman, F. C. & Harpel, M. R. Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu. Rev. Biochem.* **63**, 197–234 (1994).
33. Alonso, H., Blayney, M. J., Beck, J. L. & Whitney, S. M. Substrate-induced assembly of *Methanococcoides burtonii* D-ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. *J. Biol. Chem.* **284**, 33876–33882 (2009).
34. Hayer-Hartl, M. From chaperonins to RuBisCO assembly and metabolic repair. *Protein Sci.* **26**, 2324–2333 (2017).
35. Vitlin Gruber, A. & Feiz, L. RuBisCO assembly in the chloroplast. *Front. Mol. Biosci.* **5**, 24 (2018).
36. Wang, H. et al. RuBisCO condensate formation by CcmM in β-carboxysome biogenesis. *Nature* **566**, 131–135 (2019).
37. Ryan, P. et al. The small RbcS-like domains of the beta-carboxysome structural protein CcmM bind RuBisCO at a site distinct from that binding the RbcS subunit. *J. Biol. Chem.* **294**, 2593–2603 (2019).
38. Joshi, J., Mueller-Cajar, O., Tsai, Y. C., Hartl, F. U. & Hayer-Hartl, M. Role of small subunit in mediating assembly of red-type form I RuBisCO. *J. Biol. Chem.* **290**, 1066–1074 (2015).
39. Emlyn-Jones, D., Woodger, F. J., Price, G. D. & Whitney, S. M. RbcX can function as a RuBisCO chaperonin, but is non-essential in *Synechococcus* PCC 7942. *Plant Cell Physiol.* **47**, 1630–1640 (2006).
40. Wang, Q. S. et al. The macromolecular crystallography beamline of SSRF. *Nucl. Sci. Tech.* **26**, 12–17 (2015).
41. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
42. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
43. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
45. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
46. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
47. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
48. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
49. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
50. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
51. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
52. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

Acknowledgements

We thank the staff at the Shanghai Synchrotron Radiation Facility (SSRF) for X-ray diffraction data collection; L. Chen and B. Zhu for technical support with cryo-EM data collection at the Center for Biological Imaging at the Institute of Biophysics (IBP), Chinese Academy of Sciences; and the staff at the Core Facility Center for Life Sciences at University of Science and Technology of China for technical assistance. This research was supported by the National Natural Science Foundation of China (http://www.nsfc.gov.cn; grant numbers 31630001 and 31621002), the Strategic Priority Research Program of the Chinese Academy of Sciences (http://www.cas.cn; grant numbers XDA24020302 and XDB37020301) and the Ministry of Science and Technology of China (http://www.most.gov.cn; project number 2016YFA0400900). Y.-L.J. thanks the Youth Innovation Promotion Association of Chinese Academy of Sciences for their support.

Author contributions

C.-Z.Z., Y.-L.J. and Y.C. conceived, designed and supervised the project. Y.-L.J., C.-Z.Z., Y.C., W.-F.L. and L.-Y.X. analysed the data. Y.-L.J. and C.-Z.Z. wrote the manuscript. L.-Y.X., W.-W.K. and H.S. performed the molecular cloning, protein expression and purification. L.-Y.X. and W.-W.K. performed protein crystallization and optimization. Y.-L.J. and L.-Y.X. conducted the X-ray and cryo-EM data collection, structure determination and model refinement. L.-Y.X. and W.-W.K. performed the biochemical assays. All of the authors discussed the data and read the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-020-0665-8.

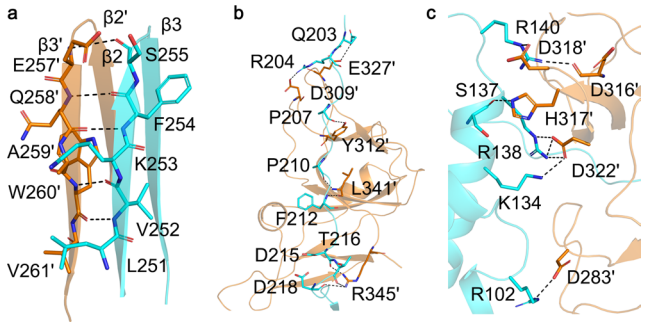
Supplementary information is available for this paper at https://doi.org/10.1038/s41477-020-0665-8.

Correspondence and requests for materials should be addressed to Y.-L.J., Y.C. or C.-Z.Z.

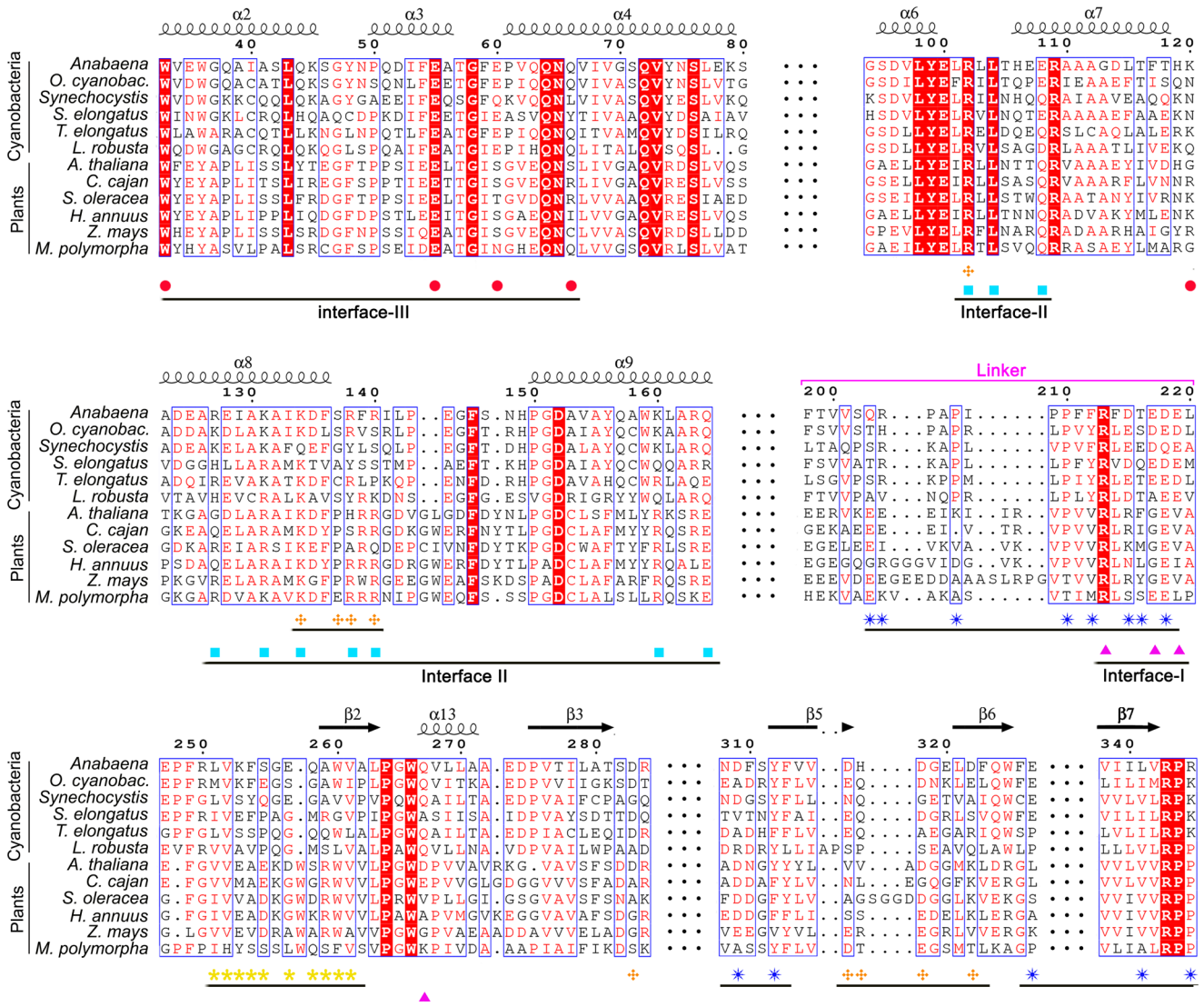
Peer review information: *Nature Plants* thanks Oliver Martin Mueller-Cajar, Spencer Whitney and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

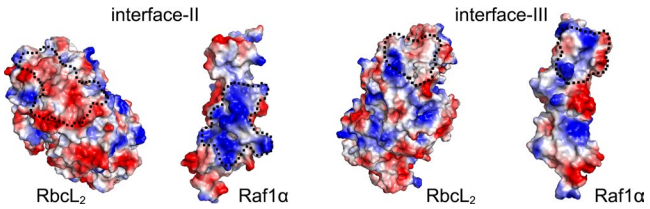
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



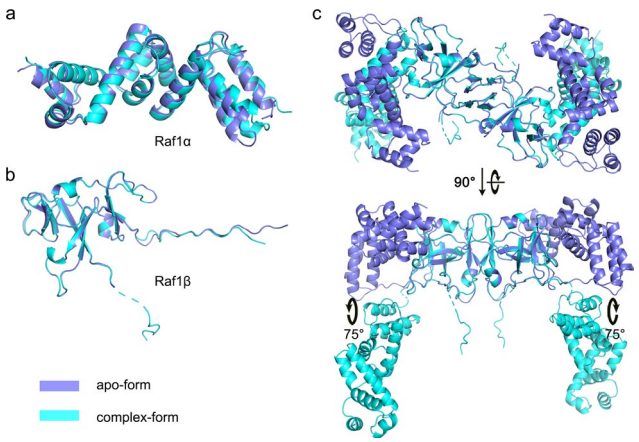
Extended Data Fig. 1 | Three interfaces between two subunits of Raf1 dimer. The two Raf1 subunits (orange and cyan) are shown as semi-transparent cartoons, whereas the interacting residues are shown as sticks, with polar interactions indicated as dashed lines. **a,** The interface between two Raf1β domains. **c,** The interface between the hydrophobic linker of one subunit and Raf1β of the symmetric subunit. **c,** The interface between Raf1α of one subunit and Raf1β of the symmetric subunit.



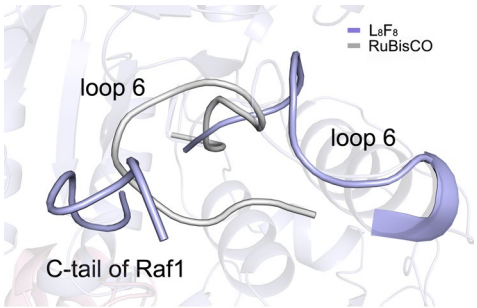
Extended Data Fig. 2 | Multiple-sequence alignment of Raf1 homologs in cyanobacteria and plants. Amino acid sequences of Raf1 homologs from cyanobacteria and plants are aligned using MultAlin (http://multalin.toulouse.inra.fr/multalin/). The secondary structural elements of *Anabaena* sp. PCC 7120 Raf1 are labeled at the top. Three interfaces (interface-I, II, III) between Raf1 and RbcL are labeled at the bottom, with the interacting residues at the three interfaces indicated by pink triangles, cyan squares, and red circles, respectively. Residues in the three interfaces between the subunits of Raf1 dimer are labeled with orange, yellow and blue stars, respectively. The linker region connecting the two Raf1 domains is marked by a pink line on the top of the sequences. The NCBI accession codes for the sequences are: *Anabaena* sp. PCC 7120, WP_010999374; *Oscillatoriales cyanobacterium*, TAF57237; *Synechocystis sp. PCC 6803*, WP_010873864; *Synechococcus elongatus* PCC 7942, WP_011377752; *Thermosynechococcus elongatus* BP-1, WP_011057603; *Limnoraphis robusta*, WP_046279487; *Arabidopsis thaliana*, NP_198202; *Cajanus cajan*, XP_020203342; *Spinacia oleracea*, XP_021866119; *Helianthus annuus*, XP_021973841; *Zea mays*, NP_001140763; *Marchantia polymorpha*, PTQ50472.



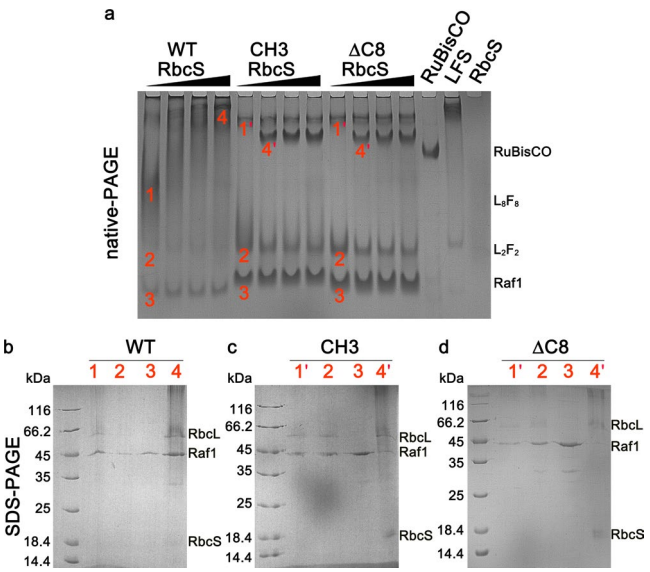
Extended Data Fig. 3 | The complementary electrostatic surfaces presentation of the RbcL dimer and Raf1, showing the areas at interface-II and III. The interacting areas on RbcL and Raf1 are highlighted as black dashed lines. The interface-II and III show a substantial complementarity in shape and charge between RbcL and Raf1α.



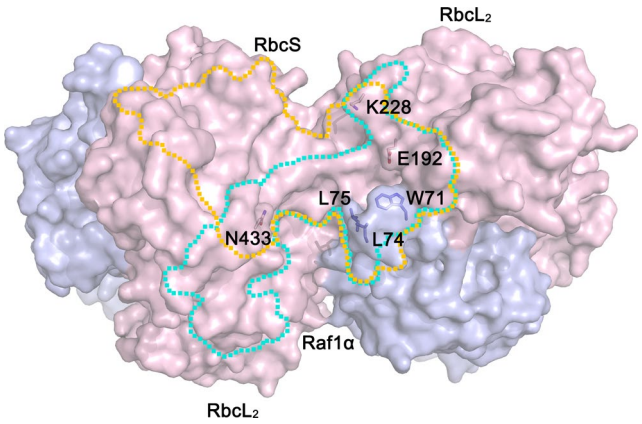
Extended Data Fig. 4 | Superposition of Anabaena sp. PCC 7120 Raf1 in apo and RbcL-bound forms. Superposition of individual (**a**) Raf1α and (**b**) Raf1β domains in apo (blue) and RbcL-bound forms (cyan). **c,** Structural comparison of Raf1 dimer in the apo and RbcL-bound forms, shown in two orientations rotated by 90°. The Raf1β domains were aligned together, in which Raf1α domains rotates against the swapped Raf1β dimer by ~75°.



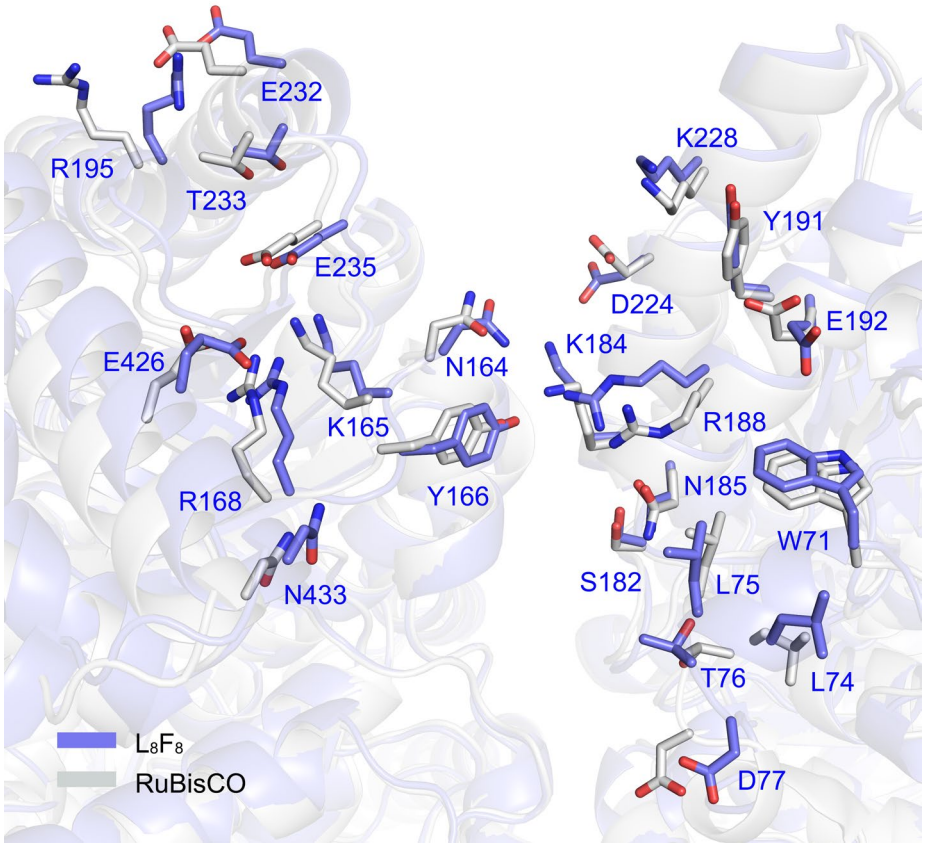
Extended Data Fig. 5 | Comparison of the 'loop 6' in the structures of L8F8 and RuBisCO holoenzyme L8S8. The Raf1 C-tail partially occupies the space, which is held by the 'loop 6' in the RuBisCO.



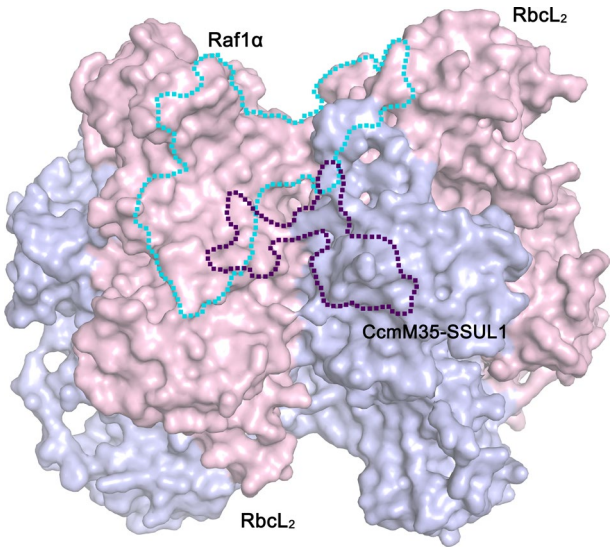
Extended Data Fig. 6 | Native- and SDS-PAGE analysis of altered RbcL-containing complex formation upon mutations of Raf1 C-tail. a, Native-PAGE analysis of RbcL-Raf1 complex with addition of RbcS at various concentrations (0, 4, 8 and 20 μM, with the molar ratio of 0, 0.5, 1 and 2.5 to RbcL, respectively). WT, CH3 and ΔC8 represent the wild-type Raf1 or Raf1 mutants, in which CH3 stands for extension of three histidine residues at the C-terminus whereas ΔC8 represents the truncation of the C-terminal eight residues. The numerically labeled bands in panel **a** are cut off for further analysis by SDS-PAGE in the lanes of (**b**) WT, (**c**) CH3 and (**d**) ΔC8, respectively.Source data



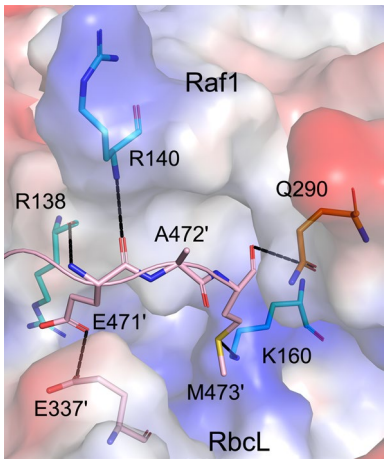
Extended Data Fig. 7 | Raf1α and RbcS share a largely overlapped binding regions on RbcL. The RbcL structures are shown as surface. The binding regions of Raf1α and RbcS on RbcL are circled by dashed lines in cyan and gold, respectively. The shared binding residues on RbcL are shown in sticks and labeled.



Extended Data Fig. 8 | Comparison of the RbcS-binding residues of RbcL in the structures of L8F8 and the holoenzyme L8S8. RbcS-binding residues are shown as blue and gray sticks for L_8F_8 and L_8S_8, respectively.



Extended Data Fig. 9 | Raf1α and SSUL1 possess a slightly overlapped binding region on RbcL. Superposition of L_8F_8 onto the complex of L_8S_8-SSUL1 (PDB: 6HBC). The RbcL structures are shown as surface, whereas the binding regions of Raf1 and SSUL1 are circled by the cyan and violet dashed lines, respectively.



Extended Data Fig. 10 | The C-terminus of RbcL interacts with Raf1. The RbcL and Raf1 structures are shown as electrostatic surface. The C-terminal residues of RbcL are shown as pink sticks, whereas Raf1 residues interacting with the C-terminus of RbcL are shown as cyan and orange sticks for the two subunits.

nature research

Corresponding author(s): Cong-Zhao Zhou

Last updated by author(s): Apr 6, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Bluice for X-ray data collection and SerialEM for cryo-EM data collection.
Data analysis	CCP4 7.0, Pymol 2.2.3, PHENIX 1.14, Coot 0.8.9, HKL2000 v712, UCSF Chimera 1.13, Relion 2.0, CTFFIND4, MotionCor2, GraphPad Prism 7.00, Adobe Photoshop CC 2017, ImageJ 1.8.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The structural factors and atomic coordinates of Raf1 and its complex with RbcL have been deposited in the Protein Data Bank (PDB; accession codes: 6KKN for Raf1 and 6KKM for L8F8). The cryo-EM structures of L8F8S8 and L8S4 have been deposited in the Protein Data Bank (accession codes: 6LRR for L8F8S8 and 6LRS for L8S4). The cryo-EM density maps of L8F8S8, L8S4, L8F8S4 and L8S8 have been deposited in the Electron Microscopy Data Bank (accession codes: EMDB-0959~EMDB-0962 for L8F8S8, L8S4, L8F8S4 and L8S8, respectively).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to determine the sample size. Not applicable.
Data exclusions	No data excluded.
Replication	Three replicates were performed in the activity assays (Fig. 4b). All replicates were successful.
Randomization	Animals and humans were not used in the study. The protein samples for the biochemical assays were randomly allocated into experimental groups.
Blinding	Animals and humans were not used in the study. Blinding was not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The antibodies, His-Tag Antibody Mouse McAb (Catalog number: 66005-1-Ig, CloneNo.: 1B7G5) and HRP-conjugated Affinipure Goat Anti-Mouse IgG(H+L) (Catalog number: SA00001-1), were used for detecting His-tagged RbcL proteins in Fig. 2e.
Validation	The antibodies were validated by western blot of purified His-tagged fusion proteins.