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	**Distribution of Form I' Chloroflexi genomes.**	Shih_EDFig1.eps	Maximum likelihood phylogenetic tree of Chloroflexi using ribosomal protein S3 (rpS3) as a marker gene. To map the distribution of Form I' Rubisco genes onto genomes, all MAGs were scanned for presence of both rpS3 and Form I' Rubisco. MAGs containing Form I' Rubisco are highlighted in orange. The scaffolds that encode RbcL vary in size substantially, ranging up to ~106 kbp in length (available as Supplementary Data). At least partial genomic context could be determined in most cases and the gene for phosphoribulokinase was adjacent. In some cases, additional CBB Cycle pathway genes were present in an operon with Rubisco, strongly supporting the function of Rubisco in this pathway. In a subset of cases, other pentose phosphate pathway genes were co-encoded. In no case was there evidence for RbcS, either on the scaffold or in the draft genome bin (where a bin was available). Gene predictions were established via a standard annotation pipeline[50,51] and augmented by HMM-based profiling and domain analysis.
Extended Data Fig. 2	**In Form I-containing Chloroflexi operons, *rbcL* and *rbcS* are always found next to each**	Shih_EDFig2.eps	Fragment operons from an example set of 10 Form I Rubisco-containing Chloroflexi genomes shows that *rbcS* is always found next to *rbcL*, similar to Form I Rubisco found in Cyanobacteria and Proteobacteria[11]. Form I' Rubisco-

	other, unlike Form I'-containing Chloroflexi operons that lack *rbcS*.		containing Chloroflexi genomes do not contain small subunit *rbcS* (Fig. 1b). Scaffold names are shown to the right of their corresponding genome fragments.
Extended Data Fig. 3	**PAGE analyses.**	Shih_EDFig3.eps	**a**, Non-denaturing PAGE gel with a molecular weight marker (M, lane 1), and purified proteins of all three candidate Form I' Rubisco (P. breve, 241187, and 170907) with (+) or without (-) prior activation and incubation with 10-fold molar excess of 2CABP. 241187 and 170907 denotes scaffolds B_1_S1_170907_scaffold_241187_5_Tax=RBG_16_Chloroflexi_63_12 and S_p2_S4_170907_scaffold_85440 Rubisco, respectively. **b**, SDS-PAGE analysis of crude cell lysate from 1) over-expression of untagged P. breve Rubisco with co-expression of GroEL/ES from pBAD33*EL/ES*, 2) over-expression of His_{14}-bdSUMO-tagged P. breve Rubisco with co-expression of GroEL/ES from pBAD33*EL/ES*, and 3) over-expression of His_{14}-bdSUMO-tagged P. breve Rubisco without overexpression of GroEL/ES (background GroEL/ES expression from *E. coli*). Without GroEL/ES overexpression, untagged RbcL comprises 8 ± 1.0 (*n* = 3) of the total soluble protein, which improves to 14 ± 0.5 (*n* = 3) when GroEL/ES overexpression in induced (see Methods). When the His_{14}-bdSUMO tag is included on the N-terminal end of RbcL, soluble expression is 7 ± 0.8 (*n* = 3) and 14 ± 0.8 (*n* = 3) of the total soluble protein, without and with GroEL/ES overexpression, respectively. Reported values collected from *n* separate experiments (separately grown *E. coli* cultures) reflect the mean ± standard deviation.
Extended Data Fig. 4	**Form I Rubisco possess a unique RbcL C-terminal extension that interacts with RbcS, which is not found in Form I' Rubisco.**	Shih_EDFig4.tiff	**a**, Sequence alignment of representative Rubisco RbcL sequences from Forms I, I', II, II/III, IIIA and IIIb. Strictly conserved residues have a red background, residues well conserved within a group are indicated by red letters, and the remaining residues are in black letters. Gaps are represented by dots. Residue numbering along the top refers to P. breve RbcL. Symbols above blocks of sequences correspond to the

			secondary structure of P. breve RbcL: α, α-helix; β, β-strand; η, 310-helix. The secondary structure elements were named according to Knight et al., 1990[52]. The positions of loop 6 (black dotted lined), the Form II/III-specific Rubisco assembly domain (cyan line), and the Form I-specific C-terminal extension (purple line) are indicated. The RbcX binding domain-specific to Form IB Rubisco is boxed in pink. The sequence alignment was created using the UniProt RbcL sequences P22859 (*Allochromatium vinosum*), O85040 (*Halothiobacillus neapolitanus*), A0A4D4IZ26 (*Zea mays*), P00880 (*Syn*6301), Q1QH22 (*Nitrobacter hamburgensis*), Q3IYC2 (*Rhodobacter sphaeroides*), P51226 (*Porphyra purpurea*), Q9GGQ2 (*Vaucheria litorea*), E1IGS1 (*Oscillochloris trichoides*), A0A0P9FAF0 (*Kouleothrix aurantiaca*), A4WW35 (*Rhodobacter sphaeroides*), P04718 (*Rhodospirillum rubrum*), Q12TQ0 (*Methanococcoides burtonii*), A0A1L3Q3Y6 (*Methanohalophilus halophilus*), B5IH56 (*Aciduliprofundum boonei*), O93627 (*Thermococcus kodakarensis*), J1ANE7 (*Methanofollis liminatans*), and Q2FSY4 (*Methanospirillum hungatei*). The sequences for representative Form I' homologs are presented in this study (Supplementary Data 1). **b**, Overlay of amino acid residues 408-458 of *Syn*6301 Rubisco (tan) with residues 415-453 of P. breve Rubisco (blue) depicting the unique RbcL C-terminal extension found in Form I enzymes, but not in Rubisco homologs that do not possess RbcS. Residues R428, N429, and E430 of *Syn*6301 RbcL contact residues N29 and Y32 at the interface of *Syn*6301 RbcS (purple).
Extended Data Fig. 5	**Negative-staining electron microscopy 2D images of P. breve Rubisco.**	Shih_EDFig5.eps	Images reflect the highest resolution data collected with activated P. breve Rubisco in phosphate buffer. The experiment was performed once (*n* = 1).
Extended Data Fig. 6	**Extended SEC-SAXS-MALS data.**	Shih_EDFig6.eps	Experimental SAXS profiles (black) of P. breve Rubisco in the absence (purple) or presence (blue) of bound 2CABP is

			displayed with the calculated scattering from the atomistic models shown in Fig. 3c. Inset shows the Guinier plot of experimental SAXS profiles with the linear fit in the q×Rg < 1.6 limits.
Extended Data Fig. 7	**Amino acid sequence alignment of *Syn*6301 RbcL and P. breve RbcL.**	Shih_EDFig7.eps	**a**, Structure-based sequence alignment was originally made using PROMALS3D[53] using 1RBL and 6URA structures, then aligned with the complete RbcL sequences using MAFFT[54]. Darker shades indicate higher sequence conservation between amino acids. *Syn*6301 and P. breve RbcL residues involved in dimer-dimer interactions are highlighted in green and blue, respectively. *Syn*6301 RbcL residues involved in RbcS contacts are annotated with red stars. All contact residues were identified using CCP4 CONTACTS[55]. **b**-**c**, Cross-section depictions of 1RBL, without RbcS, and P. breve Rubisco highlighting dimer-dimer interactions as in panel a. **d**, Map of *Syn*6301 RbcL residues involved in RbcS interactions, highlighted in red as in panel a.
Extended Data Fig. 8	**Mutating key amino acid residues at the dimer-dimer interface of P. breve Rubisco disrupts octameric oligomeric assembly.**	Shih_EDFig8.eps	Native PAGE gel of recombinant WT, K150A, D161A, W165A, D220A, and Y224A P. breve Rubisco. Native Mark protein ladder denoted by "M". Site directed mutants destabilize the interface between RbcL dimers leading to break down of higher-order (i.e., L_8) oligomers into Rubisco species with variable oligomeric state and conformations, which results in a variety of lower molecular weight migration patterns within the Native PAGE gel. Experiment was performed once (n = 1).
Extended Data Fig. 9	**Site directed mutagenesis of *Syn*6301 dimer-dimer interface residues imparts marginal stability in the absence of RbcS.**	Shih_EDFig9.eps	**a**, Protein thermal shift data displaying the mean fluorescent signal collected from four separate trials for WT *Syn*6301 RbcL, three separate mutant proteins, L158W, V154D, D349R and a combined four mutant protein, 4SDM (L158W, V154D, F217Y, and D349R). Mutations were designed to reflect homologous dimer-dimer interface residues present in P. breve Rubisco. The peaks corresponding to thermal denaturation of L_8 quaternary structure are boxed, and analysis statistics are presented in the below table. Tm values represent the mean

			and standard deviation of *n* number of experiments conducted with the same protein sample. Two-tailed P-values for unpaired t test with Welch's corrections are reported in the last column using WT *Syn*6301 RbcL as the reference comparison. *n* = number of technical replicates conducted in experiment. ns = not significant. ** P <0.005, *** P <0.0005. **b**, Native gel of purified recombinant WT and mutant *Syn*6301 proteins used in experiment.

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 Information.pdf	Supplementary Note, Supplementary Tables 1-3.
Reporting Summary	Yes	nr-reporting-summary_PMS_20200710.pdf	

B. Additional Supplementary Files

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 Data	1	Supp. Data 1.txt	Fasta file containing protein amino acid sequences for Form I' enzymes identified from MAGs.
Supplementary Data	2	Supp. Data 2.xlsx	Representative MAG genbank scaffolds.
Supplementary Data	3	Supp. Data 3.xlsx	Site-directed mutagenesis primers and synthesized candidate Form I' *rbcL* gene sequences.

Novel Bacterial Clade Reveals Origin of Form I Rubisco

Douglas M. Banda[1,2], Jose H. Pereira[3,4,†], Albert K. Liu[1,2,†], Douglas J. Orr[5], Michal Hammel[4], Christine He[6], Martin A.J. Parry[5], Elizabete Carmo-Silva[5], Paul D. Adams[3,4], Jillian F. Banfield[6,7,8,9,*] & Patrick M. Shih[1,2,10,11,*]

[1] Department of Plant Biology, University of California, Davis, CA, USA.
[2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[3] Technology Division, Joint BioEnergy Institute, Emeryville, CA, USA.
[4] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.
[5] Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK.
[6] Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA.
[7] Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA.
[8] Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA.
[9] Chan Zuckerberg Biohub, San Francisco, CA, USA.
[10] Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA, USA.
[11] Genome Center, University of California, Davis, Davis, CA, USA.

[†] Contributed equally
*Correspondence to: pmshih@ucdavis.edu or jbanfield@berkeley.edu

Abstract

Rubisco sustains the biosphere through the fixation of CO_2 into biomass. In plants and cyanobacteria, Form I Rubisco is structurally comprised of large and small subunits, whereas all other Rubisco Forms lack small subunits. Thus, the rise of the Form I complex through the innovation of small subunits represents a key, yet poorly understood, transition in Rubisco's evolution. Through metagenomic analyses, we discovered a previously uncharacterized clade sister to Form I Rubisco that evolved without small subunits. This clade diverged prior to the evolution of cyanobacteria and the origin of the small subunit; thus, it provides a unique reference point to advance our understanding of Form I Rubisco evolution. Structural and kinetic data presented here reveal how a proto-Form I Rubisco assembled and functioned without the

structural stability imparted from small subunits. Our findings provide insight into a key evolutionary transition of the most abundant enzyme on Earth and the predominant entry point for nearly all global organic carbon.

Main Text:

Of all known enzymes, few have been more integral in linking the evolution of life with the geochemical cycles of our planet than Rubisco (D-ribulose 1,5-bisphosphate carboxylase/oxygenase)[1]. Rubisco sources nearly all organic carbon to the biosphere through the fixation of atmospheric CO_2 with ribulose 1,5-bisphosphate (RuBP) into biomass, thus sustaining our entire food supply. Rubisco also possesses competing oxygenase activity, which is thought to be a vestige of its evolution in a young, oxygen-depleted atmosphere; yet it has co-evolved with Rubisco's carboxylase activity over billions of years. Although there are several distinct Forms of Rubisco found across all three domains of life[2,3], the vast majority of carbon fixation on Earth is driven specifically by Form I Rubisco (found in plants, cyanobacteria, algae, and select bacteria phyla); thus, the evolution of this unique Form of Rubisco has profoundly shaped the trajectory of our planet.

Structurally, all Forms of Rubisco are composed of at least two large subunits (RbcL, ~50 kDa) which assemble head-to-tail as catalytically active dimers. From this rudimentary dimeric scaffold (found in Form II and III homologs), Rubisco has evolved to function in higher-order structures of large subunits including hexamers (Form II), octamers (Form I), and decamers (Form III). Form I homologs, however, are structurally unique from their divergent Form II and Form III counterparts due to the presence of additional small subunits (RbcS, ~13-17 kDa), which cap either end of a central octameric RbcL assembly to form a hexadecameric (L_8S_8) holoenzyme. Thus, understanding the origins of RbcS is part and parcel to investigating the evolution of Form I Rubisco.

Although not in direct participation with the active site, RbcS is accepted as an indispensable structural component of Form I Rubisco[4–6]. For example, cyanobacterial Rubisco from *Synechococcus* sp. strain PCC 6301 (*Syn*6301) retains approximately 1% of its carboxylase activity in the absence of RbcS[4], suggesting that active site structural integrity is compromised. Furthermore, Form I Rubisco from *Rhodobacter sphaeroides* relies on RbcS to correctly arrange RbcL geometry for proper activity[7], and plant Rubisco RbcL form insoluble aggregates when expressed without RbcS *in planta*[8,9]. Despite its demonstrated significance in Rubisco catalysis, the structural role RbcS has played in the evolution of Form I Rubisco has long been debated[6]. This quandary, in part, stems from the fact that we have not identified Form I Rubisco that function without small subunits. Thus, the identification and characterization of a small subunit-less Form I Rubisco would provide the necessary reference point from which to better examine the evolutionary role of RbcS. Towards this end, we searched metagenomic datasets for a "missing link" between the evolution of the Form I clade and all other Forms of Rubisco. Here, we report the discovery of a Form I Rubisco with octameric oligomeric assembly that evolved without RbcS, thus challenging our understanding of the structural properties that govern the activity of the most prominent Form of Rubisco.

Discovery of Form I Rubisco that lack small subunits

To determine whether Form I Rubisco lacking small subunits occur in nature, we analyzed a diverse set of metagenomic datasets derived from environmental communities of largely uncultivated bacteria. Our analyses specifically targeted the identification of uncharacterized bacterial *rbcL* genes, which are usually found within operons encoding other key Calvin-Benson-Bassham (CBB) cycle genes[10]. Through this process, we identified 24 *rbcL* genes with gene products that share high sequence homology (52-61%) to known Form I Rubisco.

Notably, the average amino acid sequence identity between different Forms of Rubisco is approximately 30%, thus it is possible that the identified *rbcL* genes were either within the Form I clade, or within a close sister clade[2]. Further phylogenetic analyses confirmed that the newly discovered *rbcL* sequences indeed form a monophyletic clade sister to Form I Rubisco. Given the unique phylogenetic proximity to Form I, we named this new clade Form I' to distinguish it from all other *bona fide* Forms of Rubisco (**Fig. 1a**).

Where metagenome-assembled contigs were of sufficient length to reveal the genomic context surrounding Form I' *rbcL* genes, all identified operons encoded other CBB cycle genes, including the only other CBB cycle-specific gene, phosphoribulokinase (PRK) (**Fig. 1b**). Closer inspection of metagenome-assembled genomes (MAGs) containing Form I' *rbcL* genes indicated the absence of *rbcS* upstream or downstream of *rbcL*. Notably, bacterial Form I *rbcL* and *rbcS* genes are always found within one or two genes of another in operons[11,12]. Given that Form I' Rubisco lacks RbcS similar to all other non-Form I Rubisco found in various bacteria and archaea, this suggests that the Form I' clade represents a distinct Form of Rubisco that likely diverged from the Form I clade prior to the origin of RbcS.

Surprisingly, all Form I' genes identified from MAGs were found exclusively in a single order of the Chloroflexi phylum, Anaerolineales (**Extended Data Figs. 1 and 2**). Although Chloroflexi are commonly known for their phototrophic members in the order Chloroflexales, the majority of the phylum is composed of phenotypically-diverse filamentous bacteria that are non-phototrophic, such as the Anaerolineales[13]. Of the known phototrophic examples of Chloroflexi within the order Chloroflexales, most perform carbon fixation via the 3-hydroxypropionate bicycle (*e.g. Chloroflexus sp.*), or with Form I Rubisco via the CBB cycle (*e.g. Oscillochloris trichoides*, *Chlorothrix halophila*, and *Kouleothrix aurantiaca*)[14]. Form I'-

containing MAGs were not found to contain characteristic 3HP bicycle genes such as propionyl-CoA synthetase, malonyl-CoA reductase/3-hydroxypropionate dehydrogenase, and malonyl-CoA/succinyl-CoA reductase, suggesting that the bacteria consortium from which MAGs were derived use the CBB cycle for autotrophy. Though some examples of phototrophic Chloroflexi have recently been described in clades sister to the Anaerolinea (e.g. the class-level clade *Candidatus* Thermofonsia)[15], none possessed carbon fixation pathway genes and were presumed to be photoheterotrophic. Studies have demonstrated that phototrophy within Chloroflexi may be driven by horizontal gene transfer[15,16]; however, the tight phylogenetic distribution of Form I' genes within the order Anaerolineales suggests otherwise, albeit future studies may reveal genomes outside of Anaerolineales that possess Form I' genes.

Form I' Rubisco is functional despite lack of small subunits

To characterize genes discovered from MAGs, representative Form I' Rubisco homologs were recombinantly expressed and purified (**Extended Data Fig. 3a**) from *E. coli* overexpressing the bacterial chaperonin system GroEL-GroES (homologous to Cpn60-Cpn10/20 in plants), a necessary component of Rubisco biogenesis[17,18]. The assembly of hexadecameric Form I homologs in cyanobacteria and plants require auxiliary chaperones such as RbcX and Raf1, which aid in the stabilization of the octameric RbcL core before the addition of small subunits[19,20]. Other Form I homologs, however, do not require homologous assembly factors but instead rely on RbcS for efficient assembly, which has been demonstrated for Rubisco from the photosynthetic proteobacterium *Rhodobacter sphaeroides*[7]. RbcX was not found in Form I' containing MAGs (**Fig. 1b**). Consistent with this finding, all Form I' sequences do not possess the C-terminal binding domain for RbcX[8,21] (**Extended Data Fig. 4a**). Furthermore, Form I' homologs identified to date do not possess small subunits, precluding the necessity of chaperones

involved in the assembly of hexadecameric Rubisco[19]. Some archaeal Rubisco possess an extra C-terminal domain that is proposed to aid in RbcL core assembly[22], but this unique insertion is not found within the described representative homologs of the Form I' clade (**Extended Data Fig. 4a**). Notably, *Syn*6301 Rubisco expressed in *E. coli* makes up ~1-2% of the total soluble protein, but this number improves to ~6% with the associated overexpression of GroEL/ES[23]. In comparison, Rubisco from *R. sphaeroides* comprises ~16% of the total soluble protein when heterologously expressed in *E. coli*, which jumps to 33% with the overexpression of GroEL/ES[7]. With the system outlined in this work, Form I' Rubisco was found to express at ~7-8% of the total soluble protein in BL21(DE3) *E. coli*, which improves to ~14-15% when overexpressed with GroEL/ES (**Extended Data Fig. 3b**). Currently, it is unknown whether the expression levels of Form I' Rubisco in *E. coli* are intrinsic to its amino acid sequence alone, or if auxiliary chaperone factors are necessary for higher expression. Though the Chloroflexi from which these sequences are derived may possess a unique assembly factor that aids in Rubisco biogenesis, no such protein was identified from the metagenomic datasets presented in this work.

To assess the catalytic activity of a representative Form I' homolog, we performed detailed enzyme kinetic measurements on Form I' Rubisco from the mesophilic Chloroflexi species "*Candidatus* Promineofilum breve" (P. breve) using the method of Parry *et al.*[24]. At saturating substrate concentrations, Rubisco proteins exhibit maximal rates of catalysis (V_C and V_O for carboxylation and oxygenation, respectively), generally at the expense of the concentration of substrate necessary to achieve a maximal rate (represented by the Michaelis constants K_C and K_O for carboxylation and oxygenation, respectively, which can be considered conceptually as pseudo-dissociation constants for the binding of either CO_2 or O_2)[25,26].

P. breve Rubisco demonstrated relatively slow V_C and about average K_C when compared to the reported measurements of Form I enzymes at 25 °C[26] (**Table 1**, **Fig. 2**). Conversely, the enzyme demonstrated slightly above average V_O and below average K_O. This is consistent with the discovery of the Form I' clade within the order Anaerolineales, which is typically comprised of obligate anaerobes[27], although genomic signatures of aerobic respiration have recently been discovered in some examples of Anaerolineae[28,29]. Together, these kinetic parameters culminated in a specificity for CO_2 over O_2 (represented by the specificity factor ($S_{C/O}$), a measure of the catalytic efficiency of the carboxylation reaction over the oxygenation reaction) that is lower relative to values reported for Form I enzymes, but higher than Form II and Form III homologs (**Supplementary Table 1**). It is unclear at this time whether the high oxygenase specificity of P. breve Rubisco is linked to the absence of RbcS. Notably, Form I' and Form I Rubisco lineages diverged before the evolution of cyanobacteria suggesting that Form I' enzymes may have evolved in anaerobic conditions.

Form I' Rubisco is octameric, reminiscent of Form I Rubisco

The Form I clade is structurally characterized by two features distinct from other Forms of Rubisco: 1) the presence of RbcS, and 2) the oligomeric assembly of RbcL into octamers. Given the close phylogenetic placement to the Form I clade, we hypothesized that Form I' homologs may possess octameric oligomeric assembly of RbcL, which has not been previously observed for Rubisco in nature. Size exclusion chromatography (SEC) and non-denaturing PAGE analyses revealed that recombinant P. breve RbcL dimers (~100-110 kDa) oligomerized into a higher-order structure (**Fig. 3d**). Previous studies have demonstrated that the addition of the Rubisco-specific transition-state analog, 2-carboxyarabinitol 1,5-bisphosphate (2CABP), may influence the oligomeric state of the enzyme[30]. Incubation of magnesium-bound and CO_2-

191 activated P. breve Rubisco with 2CABP resulted in an observed structural compaction, evident
192 from both later elution in SEC traces, as well as slower migration in non-denaturing gels (**Fig. 3**).

193 To more rigorously characterize the solution-state oligomeric assembly of P. breve
194 Rubisco, we performed SEC coupled to small-angle X-ray scattering (SAXS) and multiangle
195 light scattering (MALS) (SEC-SAXS-MALS) experiments[31] with activated P. breve Rubisco in
196 the presence or absence of 2CABP. Protein molecular weights determined by MALS (~400-440
197 kDa) supported the oligomerization of P. breve Rubisco as an L_8 complex (theoretical octamer
198 M.W. ~409 kDa), similar to the octameric assembly of RbcL in related Form I enzymes (**Fig. 3).**
199 These observations were corroborated by negative-staining electron microscopy images
200 (**Extended Data Fig. 5**). Experimentally determined pair-distribution, or P(r), functions
201 displayed significant broadening and elongation of P. breve Rubisco in the absence of 2CABP
202 relative to the 2CABP-bound protein (**Fig. 3b**). This observation agrees well with the larger
203 radius of gyration (Rg) values of the 2CABP-bound (Rg ~ 46.8 ± 0.4 Å) versus unbound (Rg ~
204 45.0 ± 0.5 Å) protein.

205 In the absence of substrate, Form I Rubisco proteins exist in an "open" conformation that
206 is structurally characterized, in part, by an extended C-terminal domain that is disordered and
207 positioned away from the active site[32]. Upon active site binding of RuBP, the extended C-
208 terminal domain flips down over the active site with Loop 6 to produce a compact "closed"
209 conformation primed for catalysis. In order to account for observed differences in the radius of
210 gyration between 2CABP-bound and unbound structures, we generated theoretical SAXS data
211 from computational models of octameric P. breve Rubisco either in a compact "closed" state
212 (i.e., bound to 2CABP) or an "open" state with disordered C-terminal domains (**Fig. 3c**). Indeed,
213 theoretical SAXS data produced from these models matched well with the experimentally

214 determined P(r) functions (**Fig. 3b**) and SAXS profiles (**Extended Data Fig. 6**, **Supplementary**
215 **Table 2**, χ^2 = 1.8 and 1.4 for closed and open models, respectively).

216 Overall, the combination of SEC-SAXS-MALS and electron microscopy experiments
217 support an L_8 oligomerization of Form I' Rubisco reminiscent of the L_8S_8 Form I Rubisco.
218 Because no other Form of Rubisco has been convincingly demonstrated to express as octamers in
219 nature (see **Supplementary Note**), the most parsimonious history consistent with our data
220 suggests that the common ancestor of Form I and Form I' clades evolved an octameric core
221 assembly prior to the evolution of RbcS.

222 **Form I' Rubisco structure yields insight into Form I Rubisco evolution**

223 To obtain higher molecular resolution of P. breve Rubisco, we solved a 2.2 Å crystal
224 structure of the activated enzyme in complex with 2CABP (**Fig. 4, Supplementary Table 3**).
225 Superposition of P. breve RbcL onto the structure of *Syn*6301 L_8S_8 Rubisco (PDB ID: 1RBL)[33]
226 resulted in a Cα RMSD of 0.68 Å between 435 pruned atom pairs (97.5% of P. breve RbcL
227 amino acid sequence), with a Q-score of 0.87[34]. As with all other *bona fide* Rubisco, all key
228 active site residues[35,36] were positioned in an αβ-barrel (TIM-barrel) domain (residues 158-405).

229 Many of the characteristic Form I hydrophobic RbcL residues at the interface of large and
230 small subunits[37] were either functionally substituted on the surface of P. breve Rubisco (~31%)
231 or completely absent (~4%), based on sequence homology to *Syn*6301 RbcL (**Extended data**
232 **Fig. 7**). RbcL surface residues between the two structures displayed strikingly similar
233 electrostatic characteristics (**Fig. 4**), which was unexpected given that P. breve Rubisco had not
234 evolved to interact with RbcS, unlike its closely related *Syn*6301 homolog. Because of this
235 observation and the close phylogenetic relationship between the Form I and Form I' clades, a

competing hypothesis is that Form I' evolved *from* Form I homologs and subsequently lost RbcS, as opposed to the hypothesis that Form I' and Form I Rubisco diverged from a common ancestor. To explore this further, we investigated the observation that Form I homologs possess an RbcL "C-terminal extension" (residues 430-442 of *Syn*6301 Rubisco, **Extended Data Fig. 4a**) not found in Rubisco that lack RbcS (*i.e.*, all other Forms of Rubisco). This unique C-terminal extension has evolved in Form I lineages to stabilize key RbcL interactions with RbcS[38] (**Extended Data Fig. 4b**). The Form I' enzymes identified in this study do not possess this unique C-terminal extension important for RbcS interactions, supporting the hypothesis that Form I' and Form I Rubisco diverged from a common ancestor. This is in accordance with the parsimonious observation that all non-Form I Rubisco lack RbcS, suggesting that the common ancestor to both Form I and Form 1' clades most likely lacked RbcS.

In the absence of RbcS, we hypothesized that P. breve Rubisco must possess fortified interactions at the RbcL dimer-dimer interface to support octameric assembly. Indeed, P. breve Rubisco possesses an extensive network of hydrogen bonds and salt bridges at the interdimer interface that is not present in *Syn*6301 Rubisco (**Fig. 5a**). Site-directed mutagenesis of key amino acid residues within this network (Lys150, Asp161, Trp165, Asp220, and Tyr224) to alanine abolished P. breve Rubisco's octameric assembly (**Extended Data Fig. 8**), demonstrating their importance in maintaining holoenzyme stability in the absence of RbcS. Notably, homologous amino acid positions to Asp161, Trp165, and Tyr224 within *Syn*6301 (Val154, Leu158, and Phe217, respectively) are incapable of forming a similar electrostatic network due to their side-chain physicochemical properties, necessitating interactions with RbcS for complex stability (**Extended Data Fig. 7**).

To quantitatively evaluate how subunit interactions within *Syn*6301 and P. breve Rubisco affect the thermal stability of the complex quaternary structure, we employed a protein thermal shift assay[39] (**Fig. 5b**). In the absence of RbcS, *Syn*6301 Rubisco displayed a two-phase melting profile; the first phase (Tm = 58.6 ± 0.2 °C) resulting from quaternary structure disassembly (*i.e.*, the dissociation of octamers into dimers), and the second phase (Tm = 70.6 ± 0.2 °C) corresponding to the simultaneous denaturation of RbcL dimers and RbcL secondary structure[40]. In the presence of RbcS, *Syn*6301 Rubisco was significantly stabilized such that L_8S_8 disassembly was shifted by more than 15 °C relative to *Syn*6301 L_8 (Tm = 75.5 ± 0.1 °C). Interestingly, P. breve Rubisco disassembly displayed a modest increase in Tm (82.6 ± 0.1 °C) relative to *Syn*6301 L_8S_8, but a significant increase when compared to the Tm measured for *Syn*6301 in the absence of RbcS, consistent with the predicted added stability due to interdimer interface interactions. To stabilize *Syn*6301 in the absence of RbcS, we mutated RbcL residues known to interact with RbcS to mimic part of the electrostatic network stabilizing P. breve oligomeric assembly (**Extended Data Fig. 9**). This effort yielded modest improvement in stability, highlighting the complexity of forming octamers in the absence of RbcS.

Discussion

Accrued evidence from investigations into the evolutionary adaptability of proteins supports a common trend: the catalytic promiscuity of an enzyme is inversely proportional to its conformational stability[41–43]. In line with previous observations[6], the data presented in this work suggests that the innovation of a distinct structural subunit (i.e., RbcS) imparted structural stability to Rubisco during the evolution of its carboxylase and oxygenase activities towards "Pareto optimality"[44]. Form I' Rubisco from *Ca.* P. breve demonstrated high oxygenase activity

and lower specificity when compared to Form I homologs (**Fig. 2**, **Table 1**), likely stemming from the anaerobic lifestyle of the Anaerolineales order of Chloroflexi from which sequences were discovered. Furthermore, the divergence of Form I' and Form I Rubisco from a common ancestor predates the origin of cyanobacteria; thus it is likely that Form I' Rubisco originated during the Archean Eon when atmospheric oxygen was scarce. Collectively, these observations suggest that the appearance of RbcS and the evolutionary transition from L_8 to L_8S_8 may have been an evolutionary response to the rise of oxygen ~2.4 Ga. This environmental transition may have provided a strong selective pressure to L_8-containing autotrophs (e.g., stem-group cyanobacteria) that necessitated a tradeoff between conformational rigidity (i.e., enhanced interactions at the dimer-dimer interface of octameric Rubisco) and active site plasticity. The selective pressure driving this tradeoff likely stemmed from an increased demand for improved carboxylation activity to drive flux through carbon metabolism during a rapidly changing paleoatmosphere[45,46]. To evolve this conformational dynamism while maintaining an optimized oligomeric state (i.e., L_8), we posit that RbcS evolved to facilitate the adaptive evolution of Rubisco's catalytic activity, effectively buffering the cost of destabilizing mutations and allowing the sampling of higher genetic diversity during the random walk through sequence space.

In addition to the evolutionary insight gleaned from this work, the discovery of the Form I' clade from MAG's may offer alternative means to explore Rubisco engineering efforts in plants. Notably, Form I Rubisco has long been recalcitrant to directed evolution experiments for improved carbon fixation, with notable exceptions[47], in part due to challenges associated with effectively exploring the sequence space of two genes (i.e., RbcL and RbcS) simultaneously; thus, the absence of RbcS in Form I' enzymes may streamline such future efforts. Overall,

performing directed evolution experiments[47,48] with P. breve Rubisco in conjunction with the continued characterization of the Form I' clade will offer novel opportunities to advance our understanding of Rubisco evolution.

Methods

Metagenomic and phylogenetic analysis. All metagenomes were sequenced using 150 bp, paired-end Illumina reads and assembled into scaffolds using either IDBA-UD or Megahit. Scaffolds were binned based on GC content, coverage, presence of ribosomal proteins, presence/copies of single copy genes, tetranucleotide frequency, and patterns of coverage across samples. Bins were manually curated, dereplicated, and filtered for completeness and contamination. Genes were predicted using hidden Markov models (HMMs) based on Pfam, TIGRfams, KEGG, and custom databases. Phylogeny of bins containing Rubisco genes was identified using overall scaffold gene content as well as maximum likelihood phylogenetic trees of 16 concatenated ribosomal protein sequences. Rubisco gene sequences were dereplicated at 97% amino acid identity using CD-Hit, aligned using MAFFT (default parameters), and columns with >95% gaps were removed using TrimAI. A maximum-likelihood phylogenetic tree was constructed using RAxML-HPC BlackBox (v. 8.2.10) as implemented on cipres.org (default parameters with LG model). To construct Figure 1A, branches with bootstrap values of <0.65 were collapsed. Both the alignment file and the tree file with bootstrap values are available on figshare (DOI: 10.6084/m9.figshare.9980630).

Plasmids, cloning, and site-directed mutagenesis. Representative Form I' *rbcL* genes were synthesized by Twist Biosciences (San Francisco, CA) (sequences available as supplementary data) and cloned into a pET28 vector with an N-terminal His_{14}-bdSUMO tag[59]. Plasmids pSF1389[59], pET11a-*Syn*6301-*rbcLS*, pET11A-*Syn*6301-*rbcL*, pBAD*ES/EL*, and pG-KJE8[21] were gifts. Site-directed mutagenesis (SDM) was conducted using an Agilent QuikChange SDM kit and standard procedures. Primers were designed using the Agilent QuikChange Primer Design tool (available as supplementary data).

Expression and purification of recombinant proteins. *Brachypodium distachyon* SUMO-specific protease (bdSENP1) was prepared by transforming pSF1389 into chemically competent

BL21 DE3 Star *E. coli* cells (Macrolab, QB3-Berkeley, CA). Cells were grown to mid-log phase at 37 °C (OD_{600} ~ 0.6) and induced with 0.3 mM IPTG for 3 hours. Cells were resuspended in pH 7.0 Lysis Buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 2 mM $MgCl_2$) with ~5 mM PMSF and subject to a freeze-thaw cycle before lysis by use of a Microfluidizer high pressure homogenizer (Microfluidics, Westwood, MA), and centrifugation (15,000 RCF, 20 min). Soluble protein was 0.2/0.8 μm filtered and applied to Ni-NTA Resin (Thermo Fisher, Waltham, MA) and batch bound according to the manufacturer's protocols. Columns were washed thoroughly before elution. TEV protease (MilliporeSigma, Burlington, MA) was added to the eluted fraction according to the manufacturer's suggestion and rocked gently overnight at 4 °C to facilitate His tag cleavage. The flow-through from TEV protease reactions was buffer exchanged into pH 7.0 Ni Equilibration buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 10% glycerol) and passed over Ni-NTA resin again to separate cleaved His tag from the target protein. bdSENP1-containing flow-through was analyzed by SDS-PAGE for purity and stored at -80 °C in storage buffer (20 mM sodium phosphate pH 7.0, 300 mM NaCl, 1 mM DTT, 10% glycerol).

P. breve Rubisco was prepared by co-transforming plasmids containing His_{14}-bdSUMO-tagged P. breve RbcL into chemically competent BL21 DE3 Star *E. coli* with pBAD*ES*/*EL* plasmid. Cells were grown to mid-log phase at 30 °C (OD_{600} ~ 0.6) and overexpression of GroEL/ES was induced by the addition of 0.2% w/v arabinose, and further incubation for 2 hours. Cells were resuspended in fresh LB media (without arabinose) with 300 mM NaCl and 20 mM L-proline and shaken for 16 hours at 16 °C. Pelleted cells were resuspended in pH 8.0 Lysis Buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 2 mM $MgCl_2$) with ~5 mM PMSF and subject to a freeze-thaw cycle at -80 °C before lysis by use of a Microfluidizer high pressure homogenizer. The soluble fraction was collected by centrifugation (15,000 RCF, 20 min) and 0.2/0.8 μm filtered. Clarified cell lysate was batch-bound to pre-equilibrated Ni-NTA resin as described above. Columns were washed thoroughly before resuspension in bdSENP1 Reaction Buffer (20 mM sodium phosphate pH 8.0, 300 mM NaCl, 1 mM DTT, 10% glycerol). Purified bdSENP1 was added to resuspended columns and rocked gently overnight at 4 °C to facilitate cleavage of the His_{14}-bdSUMO tag from the target protein. Flow-through from the bdSENP1 reaction was applied to a 5 mL HiTrap Q FF column equilibrated in Q Buffer A (100 mM HEPES pH 8.0). Protein was eluted off the column over a linear NaCl gradient from 5 mM to 1 M. Eluted fractions were analyzed by SDS-PAGE prior to concentration and separation by size exclusion chromatography using a Superose 6 Increase 10/300 GL column (GE Healthcare Life Sciences, Marlborough, MA) equilibrated in SEC Buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 25 mM $MgCl_2$, 1 mM DTT, 5 mM $NaHCO_3$). Eluted SEC fractions were analyzed by SDS-PAGE and Native PAGE for Rubisco content and purity. Samples were stored in 20 mM sodium phosphate pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$ at -80 °C.

*Syn*6301 RbcLS was prepared in a similar fashion to previous reports[21,40]. Plasmids *Syn*6301-*rbcLS*-pET11A and pBAD*ES/EL* were co-transformed into BL21 DE3 Star *E. coli* cells. Cells were grown to mid-log phase at 30 °C (OD_{600} ~ 0.6) and overexpression of GroEL/ES was induced by 0.4% w/v arabinose for 1.5 hours. Cells were resuspended in fresh media (without arabinose) and induced with 1 mM IPTG for 16 hours at 16 °C. Cells were lysed by using a Microfluidizer high pressure homogenizer and centrifuged (15,000 RCF, 20 minutes). Soluble protein from whole-cell lysate was 0.2/0.8 μm filtered and subject to ammonium sulfate precipitation at the 30-40% cut (where the protein is soluble at 30% w/v ammonium sulfate, but precipitates at 40% saturation. Precipitated protein was resuspended in pH 8.0 Lysis Buffer, desalted, and applied to a MonoQ 10/100 GL column (GE Healthcare Life Sciences, Marlborough, MA) equilibrated in Q Buffer A. Protein was eluted off the column over a linear NaCl gradient from 5 mM to 1 M. Eluted fractions were analyzed by SDS-PAGE prior to concentration and size exclusion chromatography as described for P. breve Rubisco. Samples were stored in 20 mM sodium phosphate pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$ at -80 °C.

*Syn*6301 RbcL expressed without RbcS was prepared in a similar fashion to previous reports[21,40]. Plasmids *Syn*6301-*rbcL*-pET11A and pG-*KJE8* were co-transformed into BL21 DE3 Star *E. coli* cells. Cells were grown to mid-log phase at 30 °C (OD_{600} ~ 0.6) and overexpression of dnaK/dnaJ/grpE was induced by 0.4% w/v arabinose for 2 hours. Cells were resuspended in fresh media (without arabinose) and induced with 1 mM IPTG for 16 hours at 16 °C. Cells were lysed and centrifuged as described for *Syn*6301 RbcLS. Soluble protein from whole-cell lysate was subject to ammonium sulfate precipitation at the 50-60% cut. Precipitated protein at 60% saturation was resuspended in lysis buffer and purified via anion exchange and size exclusion chromatography, then stored at -80 °C as described for *Syn*6301 RbcLS.

PAGE analyses. Rubisco samples were activated with excess $NaHCO_3$ and incubated with 10-fold molar excess 2-carboxyarabinitol 1,5-bisphosphate (2CABP) as described previously[30]. 2CABP was synthesized according to previously described methods[60,61]. SDS-PAGE samples were prepared according to standard procedures in Laemmli Sample Buffer (Bio-rad, Hercules, CA) with 2-mercaptoethanol, and heated at 98 °C for 5 minutes, followed by centrifugation in a benchtop centrifuge at maximum speed for 1 minute. Samples were resolved on 12% Mini-PROTEAN® TGX™ precast protein gels (Bio-rad) in 1x Tris/Glycine/SDS buffer (Bio-Rad) and stained in AcquaStain (Bulldog Bio, Portsmouth, NH). Non-denaturing PAGE samples were prepared by mixing protein with Native Sample Buffer (Bio-Rad) at 4 °C. Samples were resolved at 4 °C on 4-15% Mini PROTEAN® TGX™ precast protein gels (Bio-rad) in 1x Tris/Glycine buffer (Bio-Rad) and visualized by staining with AcquaStain.

Crystallization, X-ray data collection, and structure determination. For crystallography, P. breve Rubisco was prepared as described above, but with a final buffer composition of 100 mM HEPES-OH pH 8.0, 100 mM NaCl, 25 mM $MgCl_2$, 1 mM DTT, 5 mM $NaHCO_3$. Samples at 10-

15 mg/mL were activated as described above. Samples crystallized in the presence of 2CABP were incubated for 1 hr at ambient temperature in the presence of a 10-fold molar excess of 2CABP before setting up crystal trays. P. breve Rubisco protein was screened using the crystallization screens: Berkeley Screen[62], Crystal Screen, SaltRx, PEG/Ion, Index and PEGRx (Hampton Research, Aliso Viejo, CA). The crystals of P. breve Rubisco were found in 0.1 M Tris pH 8.0 and 30 % Polyethylene glycol monomethyl ether 2,000 obtained by the sitting-drop vapor-diffusion method with drops consisting of a mixture of 0.2 µL of protein solution and 0.2 µL of reservoir solution.

A crystal of P. breve Rubisco was placed in a reservoir solution containing 20% (v/v) glycerol, then flash-cooled in liquid nitrogen. The X-ray data sets for P. breve Rubisco were collected at the Berkeley Center for Structural Biology beamline 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory (LBNL). The diffraction data were recorded using an ADSC-Q315r detector. The data sets were processed using the program Xia2[63].

The P. breve Rubisco crystal structure was determined by the molecular-replacement method with the program PHASER[64] within the Phenix suite[65,66], using as a search model the structure of a Rubisco from *Thermosynechococcus elongatus* (PDB code 2YBV), which shows 57 % sequence identity to the target. The atomic positions obtained from molecular replacement and the resulting electron density maps were used to build the P. breve Rubisco structure and initiate crystallographic refinement and model rebuilding. Structure refinement was performed using the phenix.refine program[66]. Translation-libration-screw (TLS) refinement was used, with each protein chain assigned to a separate TLS group. Manual rebuilding using COOT[67] and the addition of water molecules allowed construction of the final model. The final model of P. breve Rubisco has an R factor of 18.8 % and an R_{free} of 22.5 %. Root-mean-square deviation differences from ideal geometries for bond lengths, angles and dihedrals were calculated with Phenix. The stereochemical quality of the final model of P. breve Rubisco was assessed by the program MOLPROBITY[68].

Small-angle X-ray–scattering (SAXS) data collection and analysis. Small-angle X-ray scattering (SAXS) coupled with multi-angle light scattering (MALS) in line with size-exclusion chromatography (SEC) experiments were performed with 50 µL samples containing 4.6 mg/mL of P. breve Rubisco incubated with or without 2CABP prepared in 20 mM HEPES-OH (pH 8.0), 300 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$. SEC-SAXS-MALS data were collected at the ALS beamline 12.3.1 at Lawrence-Berkeley National Lab[69]. The X-ray wavelength was set at λ=1.127 Å and the sample-to-detector distance was 2100 mm resulting in scattering vectors (q) ranging from 0.01 $Å^{-1}$ to 0.4 $Å^{-1}$. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20 °C and the data was processed as described[70]. Briefly, a SAXS flow cell was directly coupled with an online 1260 Infinity HPLC system (Agilent, Santa Clara, CA) using a Shodex KW804 column (Showa Denko, Tokyo,

Japan). The column was equilibrated with running buffer (20 mM HEPES-OH (pH 8.0), 300 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$) with a flow rate of 0.5 mL/min. 90 µL of sample was separated by SEC, and three second X-ray exposures were collected continuously during a 30 min elution. The SAXS frames recorded prior to sample analysis were subtracted from all other frames. The subtracted frames were investigated by radius of gyration (Rg) derived by the Guinier approximation, $I(q) = I(0) \exp(-q^2Rg2/3)$ with the limits $qRg<1.6$. The elution peak was mapped by comparing integral of ratios to background and Rg relative to the recorded frame using the program SCÅTTER. Uniform Rg values across an elution peak represent a homogenous assembly. Final merged SAXS profiles, derived by integrating multiple frames across the elution peak, were used for further analysis including Guinier plot which determined aggregation free state. The program SCÅTTER was used to compute the pair distribution, or P(r), functions presented in Figure 3B. P(r) functions were normalized based on the molecular weight determined by SCÅTTER using volume of correlation Vc[49] (Supplementary Table 2). Eluent was subsequently split 3:1 between the SAXS line and a series of UV detectors at 280 and 260 nm, a MALS detector, a quasi-elastic light scattering (QELS) detector, and a refractometer detector. MALS experiments were performed using an 18-angle DAWN HELEOS II light scattering detector connected in tandem to an Optilab refractive index concentration detector (Wyatt Technology, Goleta, CA). System normalization and calibration was performed with bovine serum albumin using a 45 µL sample at 10 mg/mL in SEC Buffer and a dn/dc value of 0.19. The light scattering experiments were used to perform analytical scale chromatographic separations for M.W. determination of the principal peaks in the SEC analysis. UV, MALS, and differential refractive index data was analyzed using Wyatt ASTRA 7 software to monitor the homogeneity of the sample across the elution peak complementary to the above-mentioned SEC-SAXS signal validation.

SAXS modeling. The atomistic model of P. breve Rubisco in the open conformation was prepared based on the crystal structure of the closed conformation presented in this study by including missing N- and C-terminal residues using the program MODELLER[71]. Different extensions and compactions of the unfolded tails were built to screen conformational variability. The experimental SAXS profiles were then compared to theoretical scattering curves generated from these atomistic models using FoXS[57,58]. Theoretical scattering profiles were used to calculate P(r) functions and further compared to experimental P(r) functions to validate solution state conformations of P. breve Rubisco.

Negative-staining electron microscopy. 3 µL of 1 mg/mL P. breve Rubisco in SEC Buffer were applied to a glow-discharged carbon grid (30 mA, 30 sec) and incubated for 1 min at room temperature. Five drops of 2% uranyl acetate were then sequentially applied and blotted off for negative staining. 50 images were taken on a JEOL 2100F at x40,000 nominal magnification, 200 kV, with 1.48 Å/pixel sampling on a DE-20 detector. 4062 particles were selected and 2-D classified using cisTEM.

Rubisco activity assays. Rubisco specificity was determined using the method of Parry *et al*.[24], with the exception that the activation buffer included 250 mM NaCl to enhance the solubility of P. breve Form I' Rubisco, and pKa of 6.11 was used for calculations. Measurements using *T. aestivum* (bread wheat) Rubisco were used for normalization as previously described[24], and results from testing with *T. aestivum* Rubisco showed no effect of NaCl in the activation buffer. Purified Rubisco was used to determine catalytic properties as described previously[72], with the following alterations to protein desalting and activation: an aliquot of concentrated Rubisco was diluted with an activation mix containing 100 mM Bicine-NaOH pH 8.0, 20 mM $MgCl_2$, 250 mM NaCl, 10 mM $NaHCO_3$, and 1 % (v/v) Plant Protease Inhibitor cocktail (Sigma-Aldrich, UK). This was then incubated on ice for 20 min before used to assay at CO_2 concentrations of 20, 40, 60, 120, 280, and 400 μM. These were combined with O_2 concentrations of either 0, 21, 40, or 70 % (v/v) to determine K_O. V_O was calculated from measured parameters using the equation $S_{C/O} = (V_C/K_C)/(V_O/K_O)$. V_C was determined using measurements with 0% O_2. An aliquot of the activated protein was used for determination of Rubisco active sites via ^{14}C-CABP binding using the method of Sharwood *et al*.[73] with 250 mM NaCl, instead of the typical 75 mM, in the activation buffer.

Protein thermal shift (PTS) assay. The PTS assay was conducted using a Protein Thermal Shift™ kit (Thermo Fisher, Waltham, MA). Samples were prepared with 1 mg/mL protein in 1x PTS phosphate buffer, and 4x PTS dye in Thermo Fisher MicroAmp Optical 8-Tube Strips. Assay was conducted on an Applied Biosciences QuantStudio 3 RT-PCR machine. The assay consisted of initial cooling and hold at 16 °C for 1 minute, followed by an 0.05°C/s increase to 95 °C, and a final hold at 95 °C for 1 minute. Data was analyzed in Protein Thermal Shift™ Software.

Other software. Structure-based sequence alignments were conducted using PROMALS3D[53] and MAFFT[54]. Analyses of protein amino acid contacts and subunit interface thermodynamics were performed using CCP4 CONTACTS[55], and PISA[74,75], respectively. UCSF Chimera[76] was utilized for the visualization of protein models, generating electrostatic potential maps, and the preparation of manuscript figures.

Data availability. Form I' RbcL amino acid sequences are included as a supplementary file (Supplementary data 1). Sequences used to generate Fig. 1a were uploaded to figshare (DOI: 10.6084/m9.figshare.9980630) along with the associated phylogenetic tree. Representative MAG genbank scaffolds are included as a supplementary file (Supplementary data 2). Site-directed mutagenesis primers and synthesized candidate Form I' *rbcL* genes are included as a supplementary file (Supplementary data 3). The structural coordinates of 2CABP-bound P. breve Rubisco have been deposited in the PDB under the accession ID 6URA. The crystal structure of Syn6301 Rubisco can be found on the PDB under the accession ID 1RBL. Publicly available databases used in this study include: PDB (www.rcsb.org), pfam (www.pfam.xfam.org), TIGRfams (www.tigrfams.jcvi.org), and KEGG database (www.genome.jp/kegg.html).Two Chloroflexi genomes identified in this study are available at:

https://ggkbase.berkeley.edu/Chloroflexi_Rubisco_PatrickShih/organisms.

Materials & correspondence. Correspondence and material requests should be addressed to P.M. Shih and J.F. Banfield

Author information

Author contributions. D.M.B, A.K.L., and P.M.S. designed experiments. D.M.B and A.K.L. prepared all protein samples, performed all PAGE analyses, and protein thermal shift experiments. M.H. performed all SEC-SAXS-MALS experiments and data analysis. J.H.P. performed X-ray crystallography data acquisition, image processing, and structure determination. D.M.B. performed all structural analyses. A.K.L. performed all site-directed mutagenesis experiments. D.J.O. performed all Rubisco activity and kinetic measurements. C. H. and J.F.B performed all metagenomic and phylogenetic analyses. All authors participated in writing and manuscript preparation.

Acknowledgments. D.M.B, A.K.L., and P.M.S. acknowledge support from a Society in Science–Branco Weiss fellowship from ETH Zurich. J.H.P., P.D.A., and P.M.S. acknowledge support from the Joint BioEnergy Institute which is supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research under Contract No. DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. C.H. and J.F.B. thank Adi Lavy and Allison Sharrar for providing unpublished Rubisco sequences, Jacob West-Roberts for assistance, the Rifle IFRC/SFA 2.0 Metagenomics and Proteomics Data Analysis Project, the Allen Foundation, the Chan Zuckerberg Biohub, and the Innovative Genomics Institute for support. C.H. acknowledges the Camille and Henry Dreyfus Foundation for a postdoctoral fellowship, and the Joint Genome Institute CSP for sequencing. M.H. acknowledge support from the Department of Energy BER Integrated Diffraction Analysis Technologies (IDAT) program, NIGMS grant P30 GM124169-01, and ALS-ENABLE for SAXS

data collection at SIBYLS. D.J.O., M.A.J.P., and E.C.S. acknowledge support from the UK Biotechnology and Biological Sciences Research Council (BBSRC; grant number BB/I024488/1). We would like to thank Manajit Hayer-Hartl (Max Planck Institute of Biochemistry, Martinsried, Germany) for the kind donation of the *Syn*6301-*rbcL*-pET11a, *Syn*6301-*rbcLS*-pET11a, pG-*KJE8*, and pBAD33*ES/EL* plasmids used in this study. Additionally, we would like to thank Noam Prywes for the kind donation of the pET28-His_{14}-bdSUMO and pSF1389 plasmids. We would also like to thank Fei Guo and the UC Davis BioEM core facility for EM images, and the laboratory of Justin Siegel (UC Davis Genome Center) for use of their qPCR machine for protein thermal shift experiments. We are grateful to Avi Flamholz for collecting publicly available Form I Rubisco kinetic data used in this study, and to Alyssa Marinas and Rick Vermon Callado for assisting with enzyme purifications. We thank Kasey Markel for his edits and suggestions on the manuscript.

Declarations of interest

The authors declare no competing interests.

References

1. Nisbet, E. G. *et al.* The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology* **5**, 311–335 (2007).
2. Tabita, F. R. *et al.* Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiol. Mol. Biol. Rev.* **71**, 576–599 (2007).
3. Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524 (2007).
4. 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).
5. Morell, M. K., Wilkin, J. M., Kane, H. J. & Andrews, T. J. Side reactions catalyzed by ribulose-bisphosphate carboxylase in the presence and absence of small subunits. *J. Biol. Chem.* **272**, 5445–5451 (1997).
6. Spreitzer, R. J. Role of the small subunit in ribulose-1,5-bisphosphate

carboxylase/oxygenase. *Archives of Biochemistry and Biophysics* vol. 414 141–149 (2003).

7. Joshi, J., Mueller-Cajar, O., Tsai, Y.-C. 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).

8. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).

9. Grabsztunowicz, M., Górski, Z., Luciński, R. & Jackowski, G. A reversible decrease in ribulose 1,5-bisphosphate carboxylase/oxygenase carboxylation activity caused by the aggregation of the enzyme's large subunit is triggered in response to the exposure of moderate irradiance-grown plants to low irradiance. *Physiol. Plant.* **154**, 591–608 (2015).

10. Kusian, B. & Bowien, B. Organization and regulation of cbb CO2 assimilation genes in autotrophic bacteria. *FEMS Microbiol. Rev.* **21**, 135–155 (1997).

11. Tabita, F. R. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynth. Res.* **60**, 1–28 (1999).

12. Whitney, S. M. & Andrews, T. J. The gene for the ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) small subunit relocated to the plastid genome of tobacco directs the synthesis of small subunits that assemble into Rubisco. *Plant Cell* **13**, 193–205 (2001).

13. Bryant, D. A. & Liu, Z. Chapter Four - Green Bacteria: Insights into Green Bacterial Evolution through Genomic Analyses. in *Advances in Botanical Research* (ed. Beatty, J. T.) vol. 66 99–150 (Academic Press, 2013).

14. Shih, P. M., Ward, L. M. & Fischer, W. W. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. *Proc. Natl. Acad. Sci.*

U. S. A. **114**, 10749–10754 (2017).

15. Ward, L. M., Hemp, J., Shih, P. M., McGlynn, S. E. & Fischer, W. W. Evolution of Phototrophy in the Chloroflexi Phylum Driven by Horizontal Gene Transfer. (2018) doi:10.3389/fmicb.2018.00260.

16. Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of Oxygenic Photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016).

17. Roy, H. Rubisco assembly: a model system for studying the mechanism of chaperonin action. *Plant Cell* **1**, 1035–1042 (1989).

18. Hayer-Hartl, M. From chaperonins to Rubisco assembly and metabolic repair. *Protein Sci.* **26**, 2324–2333 (2017).

19. Aigner, H. *et al.* Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).

20. Wilson, R. H. & Hayer-Hartl, M. Complex Chaperone Dependence of Rubisco Biogenesis. *Biochemistry* **57**, 3210–3216 (2018).

21. Saschenbrecker, S. *et al.* Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco. *Cell* **129**, 1189–1200 (06/2007).

22. Gunn, L. H., Valegård, K. & Andersson, I. A unique structural domain in *Methanococcoides burtonii* ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) acts as a small subunit mimic. *J. Biol. Chem.* **292**, 6838–6850 (2017).

23. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342**, 884–889 (1989).

24. Parry, M. A. J., Keys, A. J. & Gutteridge, S. Variation in the Specificity Factor of C3

Higher Plant Rubiscos Determined by the Total Consumption of Ribulose-P2. *J. Exp. Bot.* **40**, 317–320 (1989).

25. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences* **103**, 7246–7251 (2006).
26. Flamholz, A. I. *et al.* Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* **58**, 3365–3376 (2019).
27. Yamada, T. & Sekiguchi, Y. Cultivation of uncultured chloroflexi subphyla: significance and ecophysiology of formerly uncultured chloroflexi 'subphylum i' with natural and biotechnological relevance. *Microbes Environ.* **24**, 205–216 (2009).
28. Hemp, J., Ward, L. M., Pace, L. A. & Fischer, W. W. Draft Genome Sequence of Ornatilinea apprima P3M-1, an Anaerobic Member of the Chloroflexi Class Anaerolineae. *Genome Announc.* **3**, (2015).
29. Ward, L. M., Hemp, J., Pace, L. A. & Fischer, W. W. Draft Genome Sequence of Leptolinea tardivitalis YMTK-2, a Mesophilic Anaerobe from the Chloroflexi Class Anaerolineae. *Genome Announc.* **3**, (2015).
30. 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).
31. Knott, G. J. *et al.* Structural basis for AcrVA4 inhibition of specific CRISPR-Cas12a. *Elife* **8**, (2019).
32. 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).

33. Newman, J., Branden, C. I. & Jones, T. A. Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from Synechococcus PCC6301. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 548–560 (1993).

34. Lu, Z., Zhao, Z. & Fu, B. Efficient protein alignment algorithm for protein search. *BMC Bioinformatics* **11 Suppl 1**, S34 (2010).

35. 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).

36. Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008).

37. van Lun, M., van der Spoel, D. & Andersson, I. Subunit interface dynamics in hexadecameric rubisco. *J. Mol. Biol.* **411**, 1083–1098 (2011).

38. Schneider, G. *et al.* Comparison of the crystal structures of L2 and L8S8 Rubisco suggests a functional role for the small subunit. *EMBO J.* **9**, 2045–2050 (1990).

39. Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. Protein Sci.* **79**, 28.9.1–14 (2015).

40. Greene, D. N., Whitney, S. M. & Matsumura, I. Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. *Biochem. J* **404**, 517–524 (2007).

41. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678–687 (2005).

42. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002 (2008).

43. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
44. Erb, T. J. & Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Curr. Opin. Biotechnol.* **49**, 100–107 (02/2018).
45. Badger, M. R., Hanson, D. & Dean Price, G. Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* **29**, 161–173 (2002).
46. Studer, R. A., Christin, P.-A., Williams, M. A. & Orengo, C. A. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proceedings of the National Academy of Sciences* **111**, 2223–2228 (2014).
47. Zhou, Y. & Whitney, S. Directed Evolution of an Improved Rubisco; In Vitro Analyses to Decipher Fact from Fiction. *Int. J. Mol. Sci.* **20**, (2019).
48. Wilson, R. H., Alonso, H. & Whitney, S. M. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci. Rep.* **6**, 22284 (2016).
49. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).
50. Diamond, S. *et al.* Mediterranean grassland soil C-N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat Microbiol* **4**, 1356–1367 (2019).
51. Lavy, A. *et al.* Microbial communities across a hillslope-riparian transect shaped by proximity to the stream, groundwater table, and weathered bedrock. *Ecol. Evol.* **9**, 6869–6900 (2019).
52. Knight, S., Andersson, I. & Brändén, C. I. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 A resolution. Subunit interactions and active

site. *J. Mol. Biol.* **215**, 113–160 (1990).

53. Pei, J., Kim, B.-H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).

54. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* (2017) doi:10.1093/bib/bbx108.

55. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137 (2003).

56. Mueller-Cajar, O., Morell, M. & Whitney, S. M. Directed Evolution of Rubisco in Escherichia coli Reveals a Specificity-Determining Hydrogen Bond in the Form II Enzyme. *Biochemistry* **46**, 14067–14074 (2007).

57. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–4 (2010).

58. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).

59. Frey, S. & Görlich, D. A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J. Chromatogr. A* **1337**, 95–105 (2014).

60. Kane, H. J., Wilkin, J. M., Portis, A. R. & John Andrews T. Potent inhibition of ribulose-bisphosphate carboxylase by an oxidized impurity in ribulose-1,5-bisphosphate. *Plant Physiol.* **117**, 1059–1069 (1998).

61. Pierce, J., Tolbert, N. E. & Barker, R. Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogues. *Biochemistry* **19**, 934–942 (1980).

62. Pereira, J. H., McAndrew, R. P., Tomaleri, G. P. & Adams, P. D. Berkeley Screen: a set of 96 solutions for general macromolecular crystallization. *J. Appl. Crystallogr.* **50**, 1352–1358 (2017).

63. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).

64. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

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

66. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).

67. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

68. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–83 (2007).

69. Dyer, K. N. *et al.* High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol.* **1091**, 245–258 (2014).

70. Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612 (2009).

71. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

72. Prins, A. *et al.* Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. *J. Exp. Bot.* **67**, 1827–1838 (03/2016).

73. Sharwood, R. E., Ghannoum, O. & Whitney, S. M. Prospects for improving CO2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity. *Curr. Opin. Plant Biol.* **31**, 135–142 (2016).

74. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

75. Krissinel, E. Crystal contacts as nature's docking solutions. *J. Comput. Chem.* **31**, 133–143 (2010).

76. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Figure Legends

Fig. 1. Metagenomics-enabled identification of a novel clade of Form I Rubisco that lack small subunits. a, Maximum likelihood phylogeny of Rubisco RbcL. By including recently discovered metagenome-assembled genomes (MAGs) from Chloroflexi, the emergence of a *bona fide*, well-supported clade of Rubisco was identified (Form I'). Black circles indicate bootstrap values of 100 and white circles indicate bootstrap values >90. **b**, Example Chloroflexi operons with Form I' Rubisco (dark blue) reveal no presence of a *rbcS*, a defining feature of Form I Rubisco, which are almost always found immediately neighboring *rbcL* in bacteria; however, other CBB cycle-related genes are found in the operon (light blue). White, other enzymes; gray, hypothetical protein. Annotated loci (i-v) represent Scaffolds 211530, 92, 509483, 467972, and 172446, respectively. For the full annotation information see Supplementary Data 2. GAPDH, glyceraldehyde-3-phosphate dehydrogenase; cbbT,

transketolase; PRK, phosphoribulokinase; FBP, fructose bisphosphate; TBP, tagatose bisphosphate; cbbF, fructose 1,6-bisphosphatase.

Fig. 2. Comparison of P. breve kinetic data to reported values of Form I Rubisco. Scatter plots of reported Form I Rubisco kinetic data (black circles) collected at 25 °C[26] against P. breve Form I' Rubisco (green dots), including maximum rates of carboxylation and oxygenation of RuBP (V_C and V_O, respectively), the catalytic efficiency of carboxylation over oxygenation ($S_{C/O}$), and Michaelis constants for carboxylation and oxygenation of RuBP (K_C and K_O, respectively). Gray dotted lines represent the median for collected Form I Rubisco kinetic data.

Fig. 3. Solution-state characterization of Form I' oligomerization reveals an octameric holoenzyme reminiscent of canonical Form I Rubisco. a, SEC-SAXS-MALS chromatograms of the separation of activated P. breve Rubisco in the absence (top) or presence (bottom) of bound 2CABP. Solid gray lines represent the UV absorbance reading at 280 nm, dashed black lines represent the integrated SAXS signal, while circles represent molecular mass (light blue) data collected from MALS, and Rg values for each SAXS frame (dark blue) versus elution time. **b**, Experimental P(r) functions determined from SAXS profiles (black dashes) of P. breve Rubisco in the open conformation (light blue) or bound to 2CABP (dark blue). The area under the P(r) function is normalized relative to the molecular weight estimated by SAXS[49] and is listed in Supplementary Table 2. Theoretical P(r) functions are calculated from the theoretical SAXS curves of the corresponding models shown in panel C. The radius where P(r) approaches zero intensity identifies the maximal dimension of the macromolecule (dashed arrows). **c**, Surface representation models of P. breve Rubisco with extended (open conformation) or compact (closed conformation) C-terminal regions. **d**, A representative non-denaturing PAGE gel demonstrating the migration of P. breve Rubisco in the absence (-) or presence (+) of 2CABP. M = molecular weight marker. Native gel electrophoresis experiment was performed at $n > 10$.

Fig. 4. Crystal structure of Form I' Rubisco compared to cyanobacterial Form I Rubisco. Comparison of the structural models of **a**, Form I Rubisco from *Synechococcus sp*. strain PCC 6301 (PDB ID: 1RBL) RbcL (green) with RbcS (tan), and **b**, Form I' Rubisco from P. breve (PDB ID: 6URA, blue) which lacks RbcS. Coulombic electrostatic potential maps of 1RBL (RbcS removed) and P. breve Rubisco are illustrated by the charge distributions (negative, red; neutral, white; positive, blue) of the surface residues of either structure.

Fig. 5. Structure of Form I' Rubisco suggests that a tradeoff between stability and catalytic activity spurred the evolution of the small subunit. a, Salt bridge and hydrogen bond networks present at the dimer-dimer interface of P. breve Rubisco mediate holoenzyme stability in the absence of small subunits. Separate RbcL dimers at the dimer-dimer pair are distinguished by

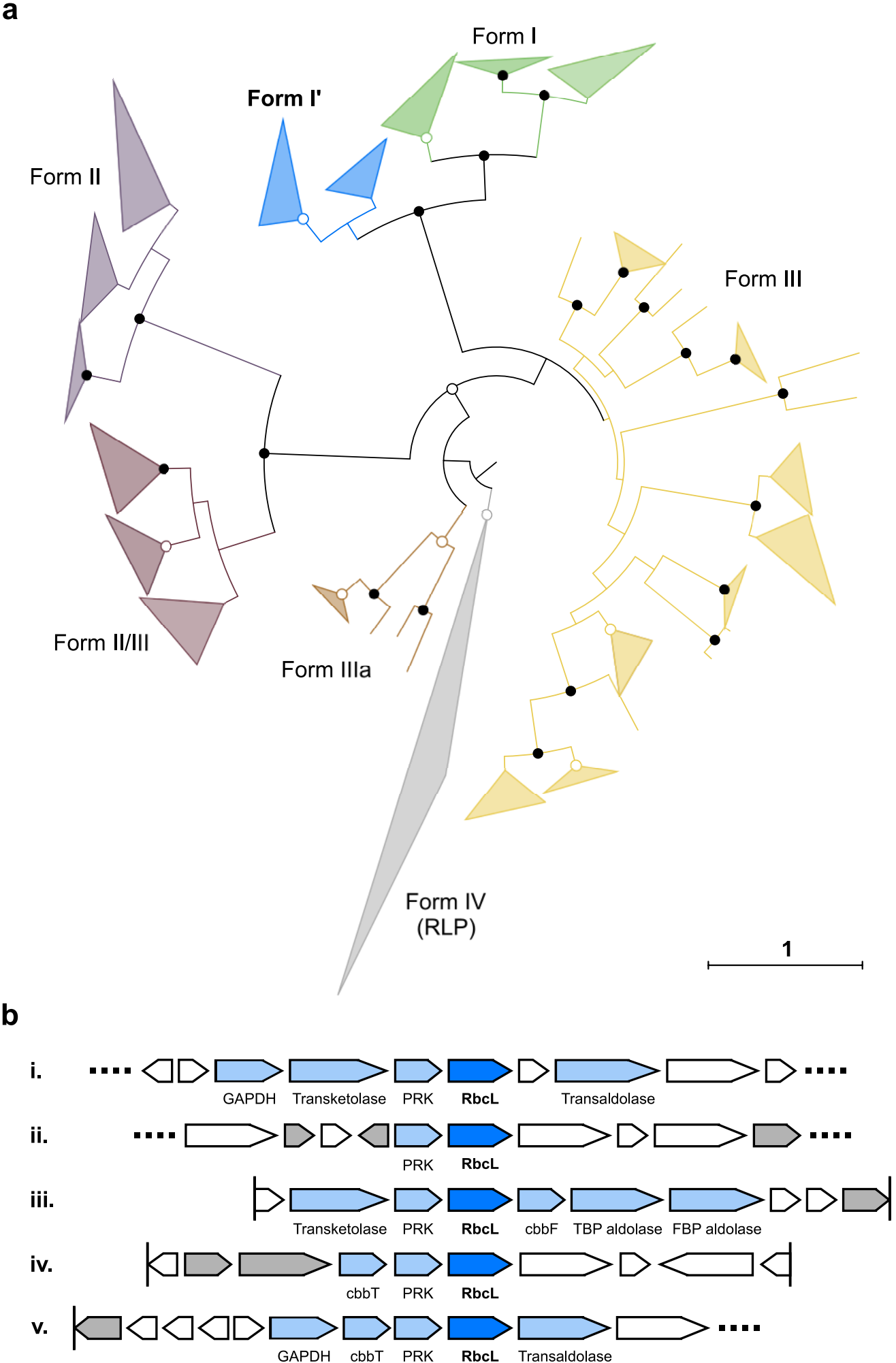
two separate shades of blue. **b**, Protein thermal shift assay data with annotated melting temperatures (Tm) for the disassembly of RbcL dimer quaternary structure from wild-type *Syn*6301 RbcL, *Syn*6301 RbcLS, and P. breve RbcL. Reported Tm values represent the average measured from a total of four experiments.

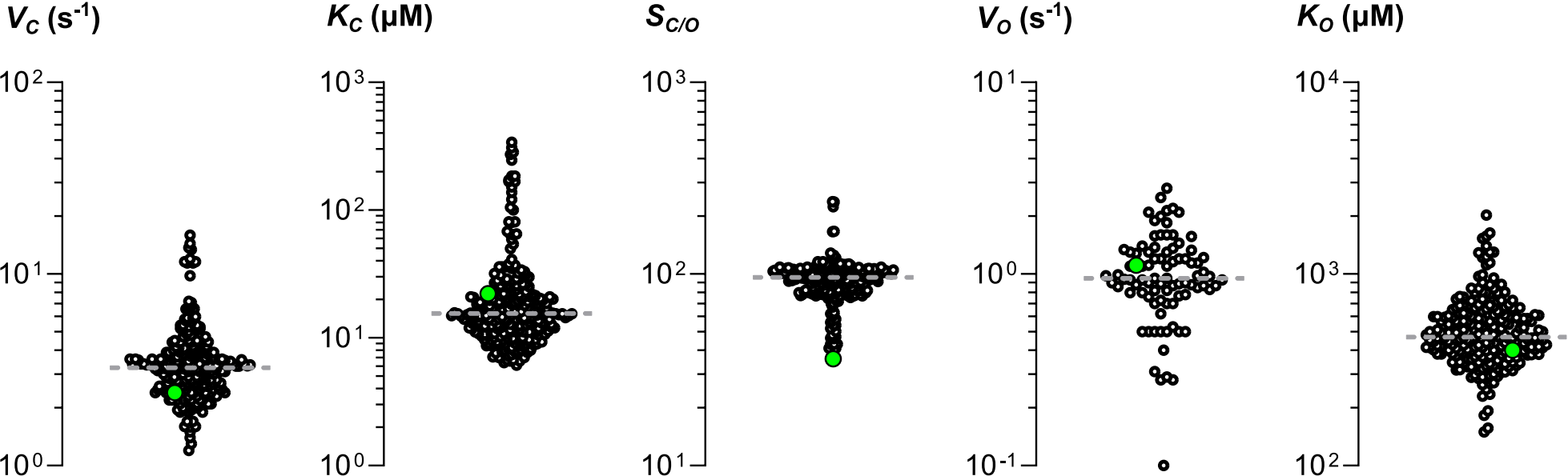
Tables

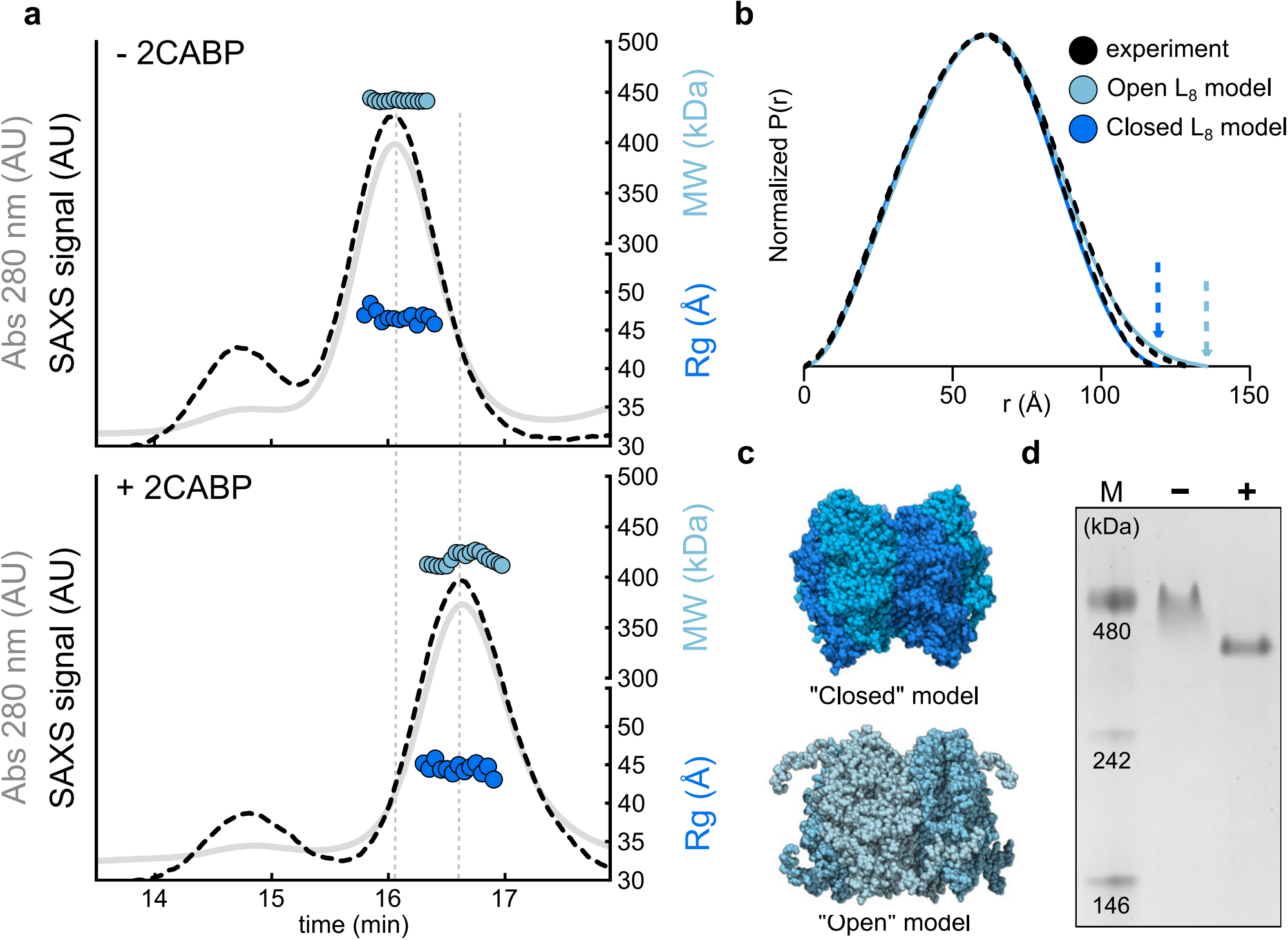
Table 1 | Kinetic characterization of Form I' Rubisco at 25 °C.

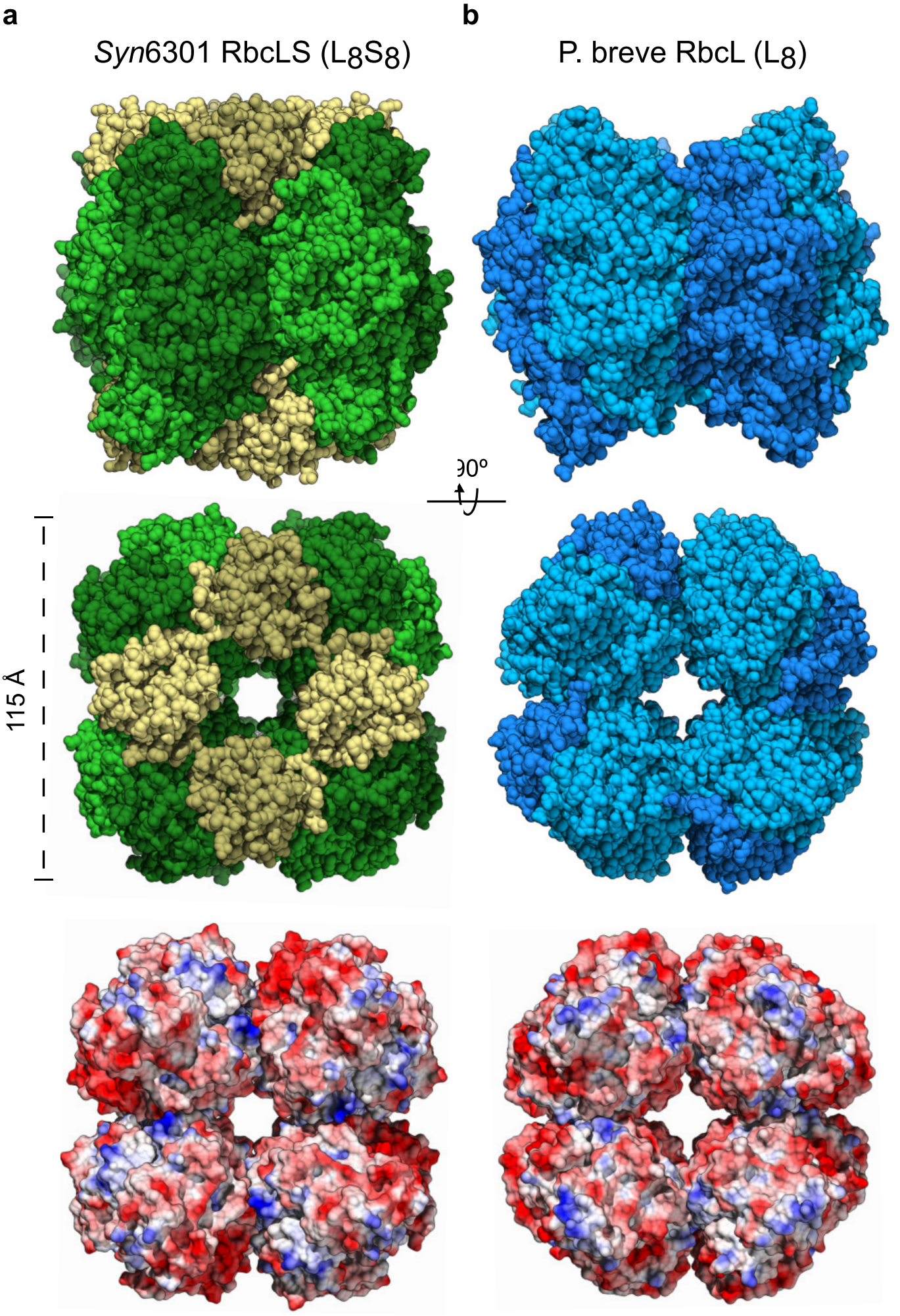
Rubisco	V_C **(s^{-1})**	K_C **(μM)**	$S_{C/O}$	V_O **(s^{-1})**	K_O **(μM)**
Form I' P. breve	2.23 ± 0.04 (*5*)	22.2 ± 9.7 (*5*)	36.1 ± 0.9 (*10*)	1.11(5)	401 ± 115 (*5*)
Form I *Synechococcus sp.* strain PCC 6301	14.3 ± 0.71 (*4*)	235 ± 20.0 (*4*)	56.1 ± 1.3 (*4*)	1.10 (4)	983 ± 81 (*4*)

V_C and V_O correspond to the maximal rates of the carboxylation and oxygenation reactions, respectively, under saturating substrate concentrations. K_C and K_O are the Michaelis constants (K_M) for the carboxylation and oxygenation reactions, respectively. $S_{C/O} = (V_C/K_C)/(V_O/K_O)$. Values represent the mean ± S.E. with *n* indicated in parentheses, where *n* reflects the number of experiments conducted with the same protein sample.


a
Form I
Form I'
Form II
Form III
Form II/III
Form IIIa
Form IV
(RLP)
1
b
i.
GAPDH
Transketolase
PRK
RbcL
Transaldolase
ii.
PRK
RbcL
iii.
Transketolase
PRK
RbcL
cbbF
TBP aldolase
FBP aldolase
iv.
cbbT
PRK
RbcL
v.
GAPDH
cbbT
PRK
RbcL
Transaldolase


V_C (s^{-1})
K_C (μM)
$S_{C/O}$
V_O (s^{-1})
K_O (μM)
10^2
10^1
10^0
10^3
10^2
10^1
10^0
10^3
10^2
10^1
10^1
10^0
10^{-1}
10^4
10^3
10^2


a
- 2CABP
+ 2CABP
Abs 280 nm (AU)
SAXS signal (AU)
MW (kDa)
Rg (Å)
500
450
400
350
300
50
45
40
35
30
14
15
16
17
time (min)
b
experiment
Open L8 model
Closed L8 model
Normalized P(r)
0
50
r (Å)
100
150
c
"Closed" model
"Open" model
d
M
−
+
(kDa)
480
242
146


a
b
Syn6301 RbcLS (L8S8)
P. breve RbcL (L8)
90°
115 Å

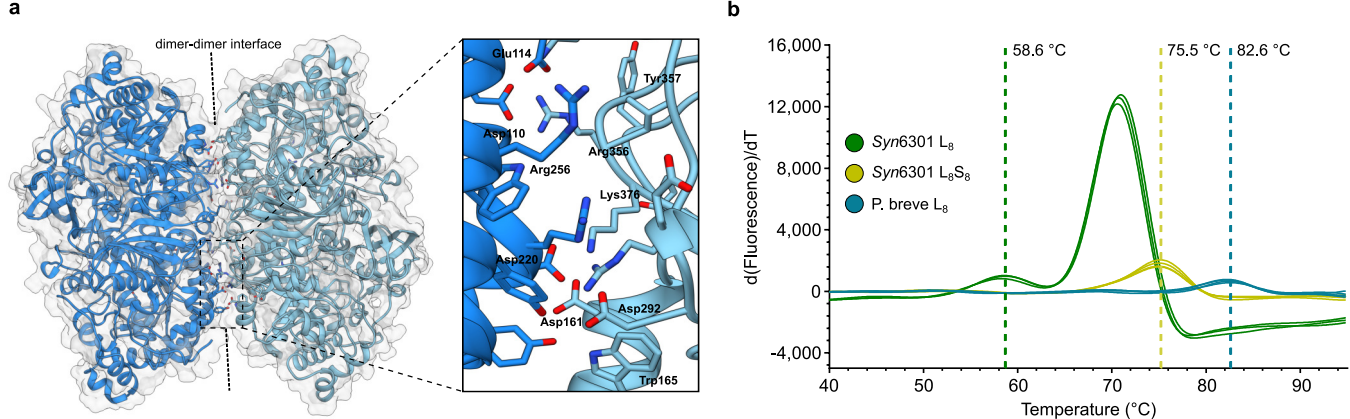
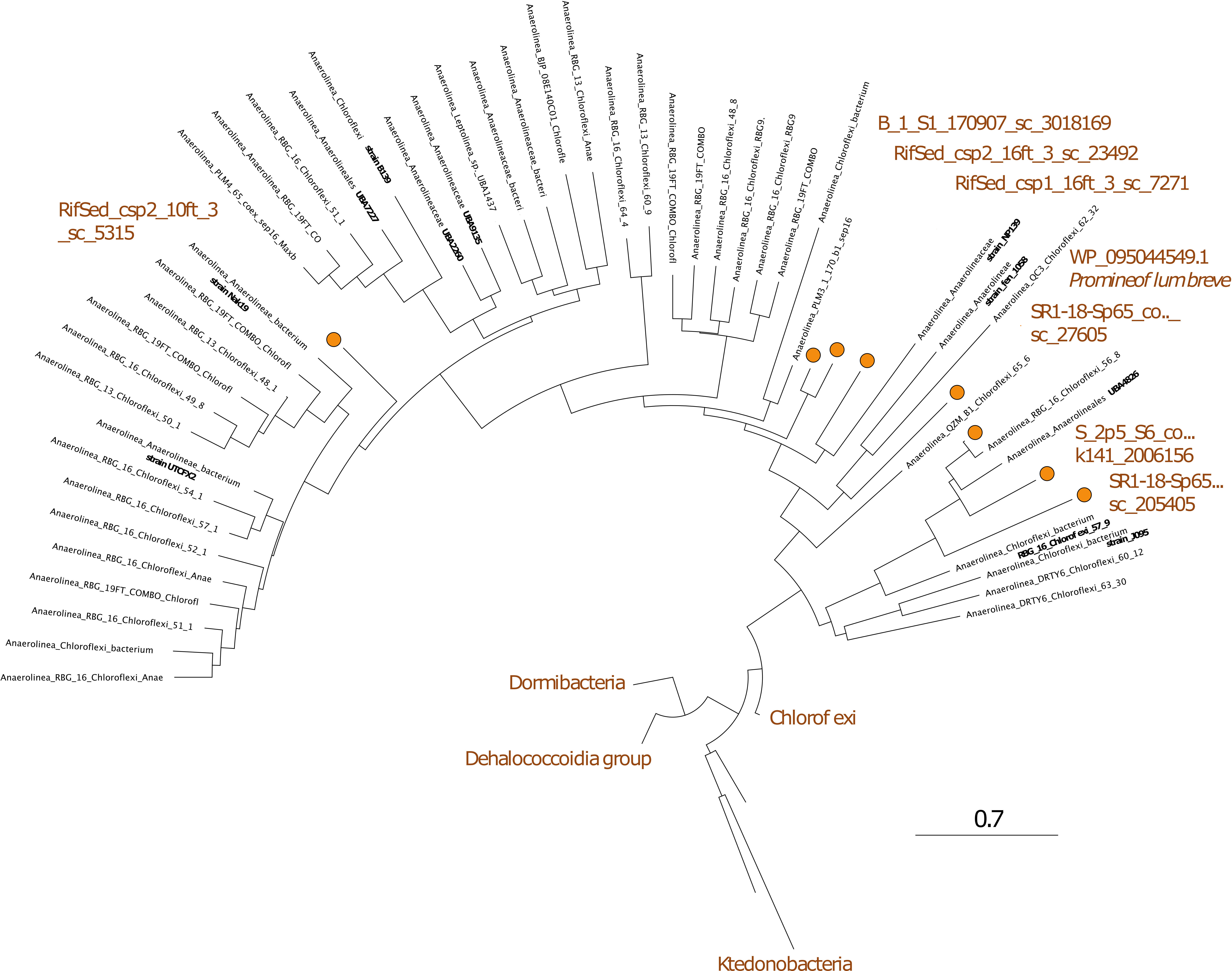
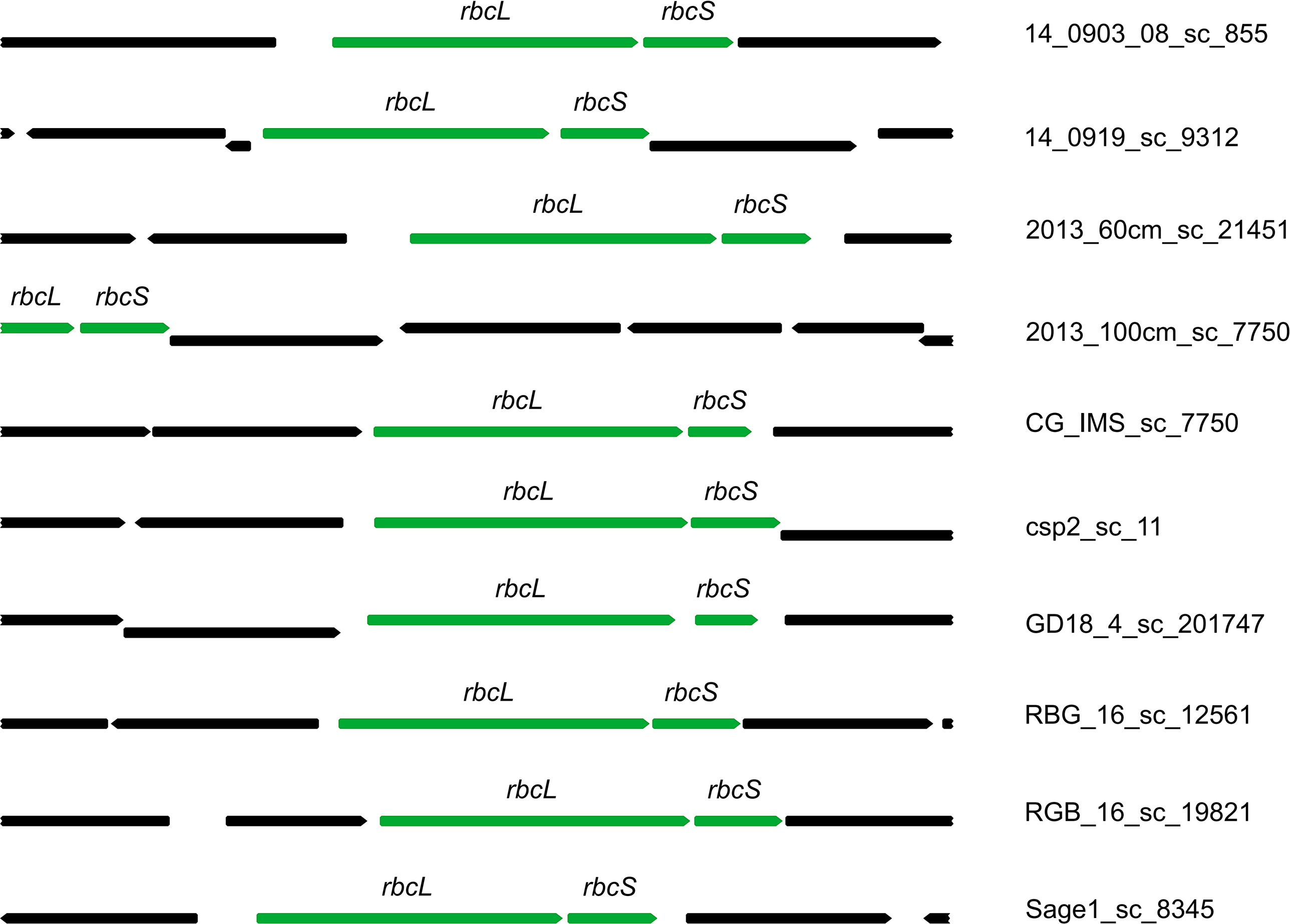

a
dimer-dimer interface
Glu114
Tyr357
Asp110
Arg356
Arg256
Lys376
Asp220
Asp292
Asp161
Trp165
b
16,000
12,000
8,000
4,000
0
-4,000
d(Fluorescence)/dT
58.6 °C
75.5 °C
82.6 °C
Syn6301 L8
Syn6301 L8S8
P. breve L8
40
50
60
70
80
90
Temperature (°C)

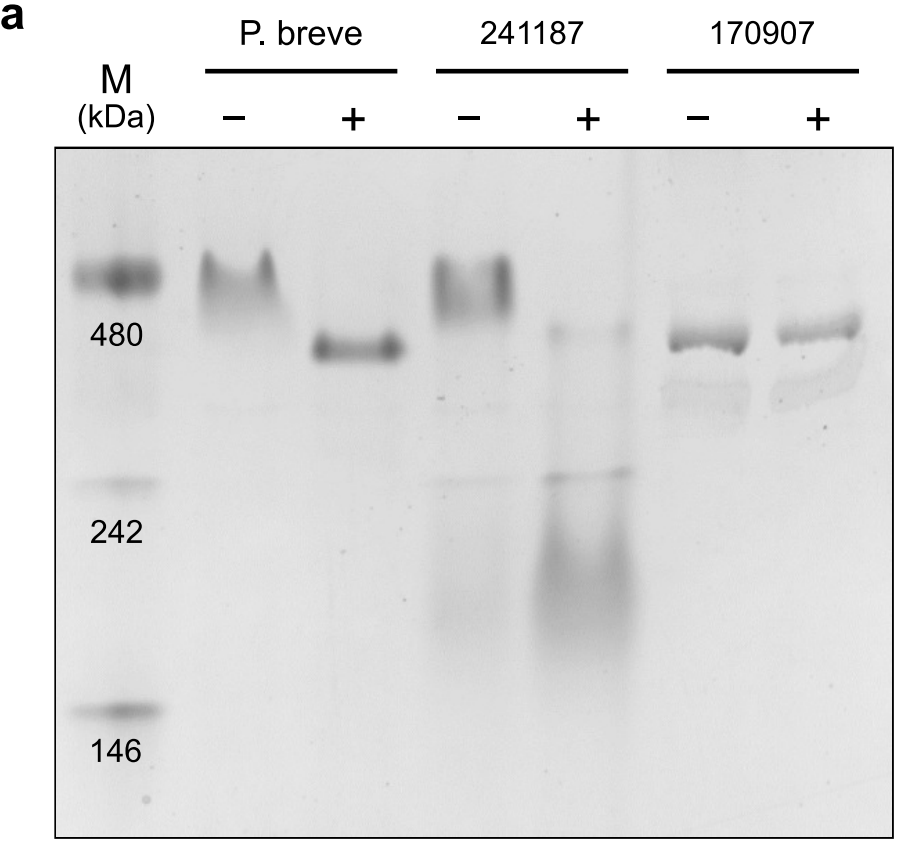
Table 1 | Kinetic characterization of Form I' Rubisco at 25 °C.

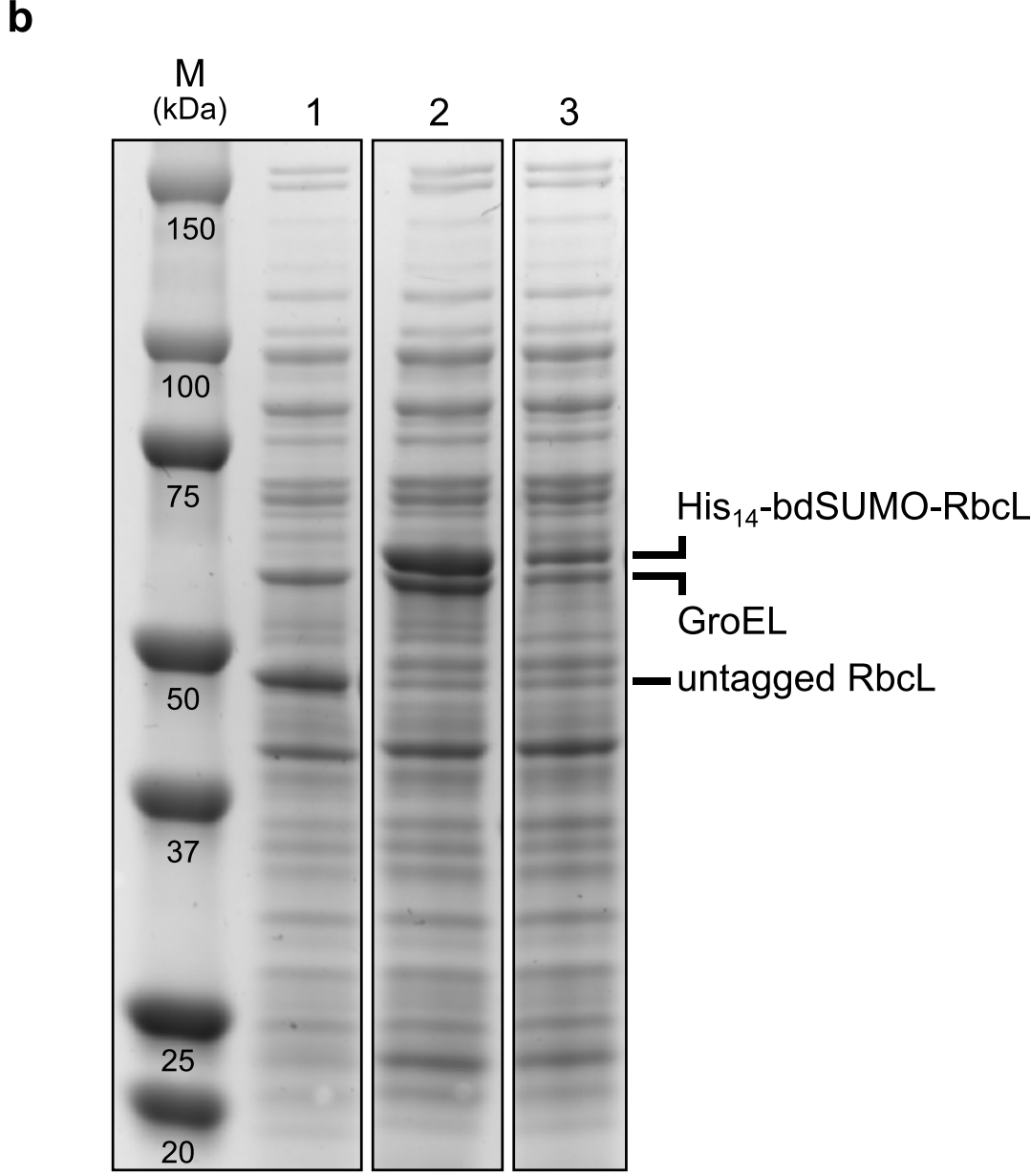
Rubisco	V_C **(s^{-1})**	K_C **(μM)**	$S_{C/O}$	V_O **(s^{-1})**	K_O **(μM)**
Form I' P. breve	2.23 ± 0.04 (*5*)	22.2 ± 9.7 (*5*)	36.1 ± 0.9 (*10*)	1.11(5)	401 ± 115 (*5*)
Form I *Synechococcus sp.* strain PCC 6301	14.3 ± 0.71 (*4*)	235 ± 20.0 (*4*)	56.1 ± 1.3 (*4*)	1.10 (4)	983 ± 81 (*4*)

V_C and V_O correspond to the maximal rates of the carboxylation and oxygenation reactions, respectively, under saturating substrate concentrations. K_C and K_O are the Michaelis constants (K_M) for the carboxylation and oxygenation reactions, respectively. $S_{C/O} = (V_C/K_C)/(V_O/K_O)$. Values represent the mean ± S.E. with *n* indicated in parentheses, where *n* reflects the number of experiments conducted with the same protein sample.

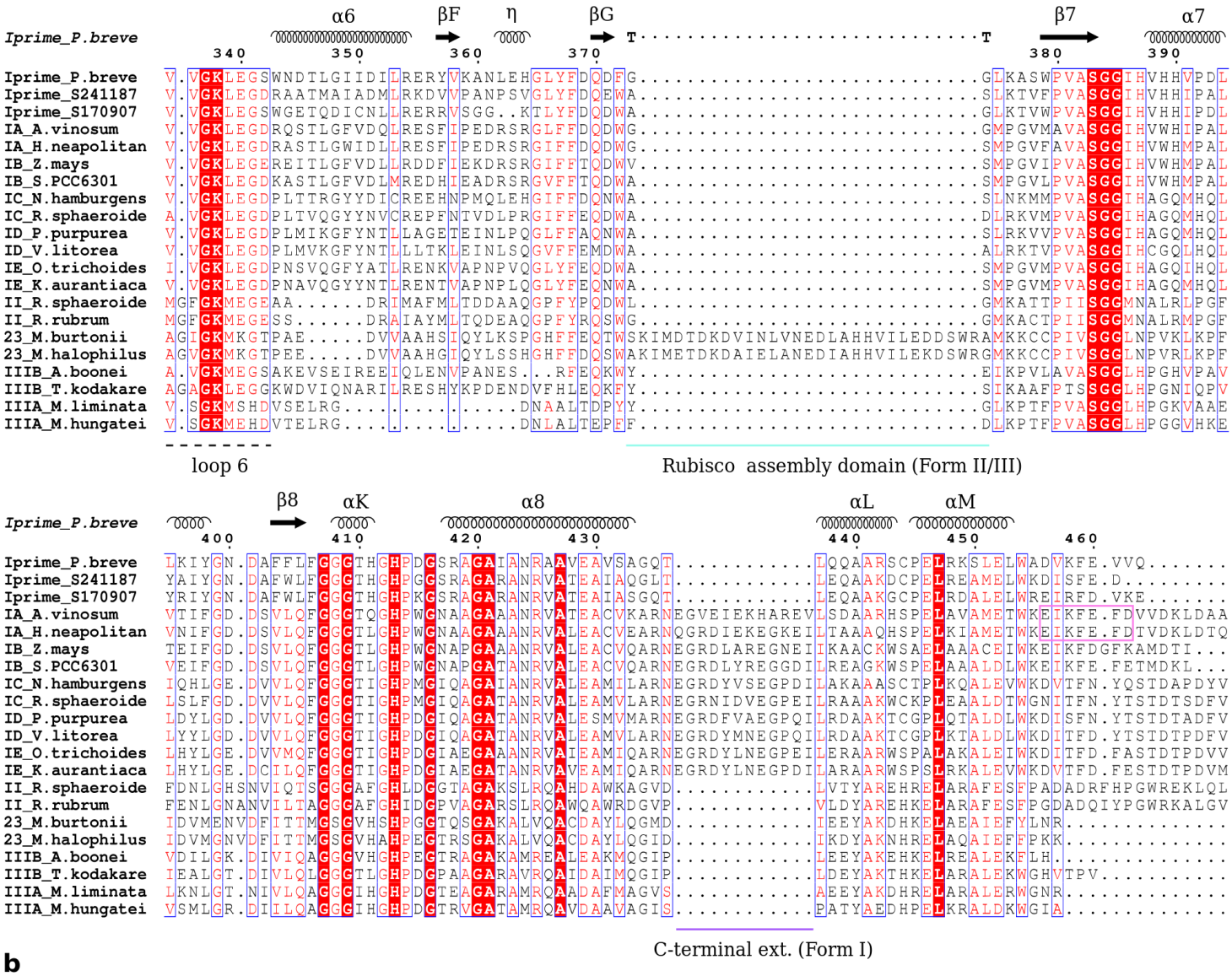

B_1_S1_170907_sc_3018169
RifSed_csp2_16ft_3_sc_23492
RifSed_csp1_16ft_3_sc_7271
RifSed_csp2_10ft_3_sc_5315
WP_095044549.1 Promineofilum breve
SR1-18-Sp65_co..._sc_27605
S_2p5_S6_co... k141_2006156
SR1-18-Sp65... sc_205405
Dormibacteria
Dehalococcoidia group
Chloroflexi
Ktedonobacteria
0.7
Anaerolinea_Chloroflexi_bacterium
Anaerolinea_RBG_16_Chloroflexi_Anae
Anaerolinea_RBG_16_Chloroflexi_51_1
Anaerolinea_RBG_19FT_COMBO_Chlorofl
Anaerolinea_RBG_16_Chloroflexi_52_1
Anaerolinea_RBG_16_Chloroflexi_57_1
Anaerolinea_RBG_16_Chloroflexi_54_1
Anaerolinea_Anaerolineae_bacterium strain UTCFX2
Anaerolinea_RBG_13_Chloroflexi_50_1
Anaerolinea_RBG_16_Chloroflexi_49_8
Anaerolinea_RBG_19FT_COMBO_Chlorofl
Anaerolinea_RBG_13_Chloroflexi_48_1
Anaerolinea_RBG_19FT_COMBO_Chlorofl
Anaerolinea_Anaerolineae_bacterium strain Nak19
Anaerolinea_PLM4_65_coex_sep16_Maxb
Anaerolinea_Anaerolinea_RBG_19FT_CO
Anaerolinea_RBG_16_Chloroflexi_51_1
Anaerolinea_Anaerolineales UBA7227
Anaerolinea_Chloroflexi strain B139
Anaerolinea_Anaerolineaceae UBA2260
Anaerolinea_Anaerolineaceae UBA9135
Anaerolinea_Leptolinea_sp._UBA1437
Anaerolinea_Anaerolineaceae_bacteri
Anaerolinea_BJP_08E140C01_Chlorofle
Anaerolinea_RBG_13_Chloroflexi_Anae
Anaerolinea_RBG_16_Chloroflexi_64_4
Anaerolinea_RBG_13_Chloroflexi_60_9
Anaerolinea_RBG_19FT_COMBO_Chlorofl
Anaerolinea_RBG_19FT_COMBO
Anaerolinea_RBG_16_Chloroflexi_48_8
Anaerolinea_RBG_16_Chloroflexi_RBG9.
Anaerolinea_RBG_16_Chloroflexi_RBG9
Anaerolinea_RBG_19FT_COMBO
Anaerolinea_Chloroflexi_bacterium
Anaerolinea_PLM3_1_170_b1_sep16
Anaerolinea_Anaerolineaceae strain NP139
Anaerolinea_Anaerolineae strain fen_1058
Anaerolinea_QC3_Chloroflexi_62_32
Anaerolinea_QZM_B1_Chloroflexi_65_6
Anaerolinea_RBG_16_Chloroflexi_56_8
Anaerolinea_Anaerolineales UBA4826
Anaerolinea_Chloroflexi_bacterium RBG_16_Chloroflexi_57_9
Anaerolinea_Chloroflexi_bacterium strain J095
Anaerolinea_DRTY6_Chloroflexi_60_12
Anaerolinea_DRTY6_Chloroflexi_63_30


rbcL
rbcS
14_0903_08_sc_855
rbcL
rbcS
14_0919_sc_9312
rbcL
rbcS
2013_60cm_sc_21451
rbcL
rbcS
2013_100cm_sc_7750
rbcL
rbcS
CG_IMS_sc_7750
rbcL
rbcS
csp2_sc_11
rbcL
rbcS
GD18_4_sc_201747
rbcL
rbcS
RBG_16_sc_12561
rbcL
rbcS
RGB_16_sc_19821
rbcL
rbcS
Sage1_sc_8345

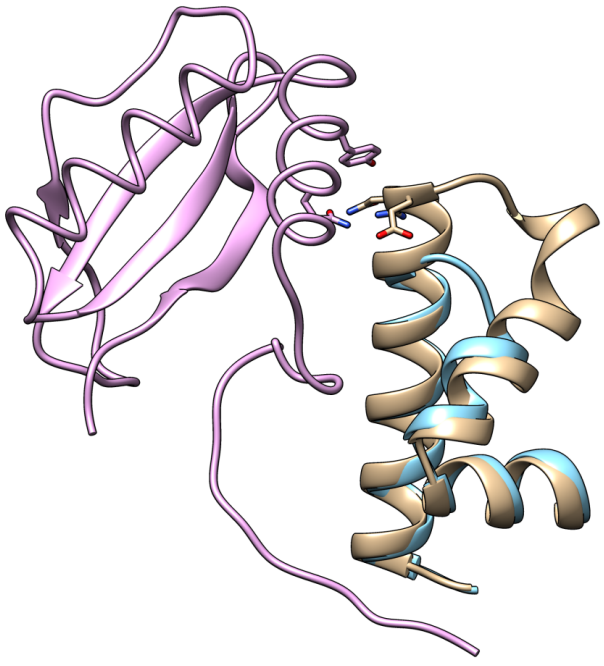

a
P. breve
241187
170907
M
(kDa)
–
+
–
+
–
+
480
242
146

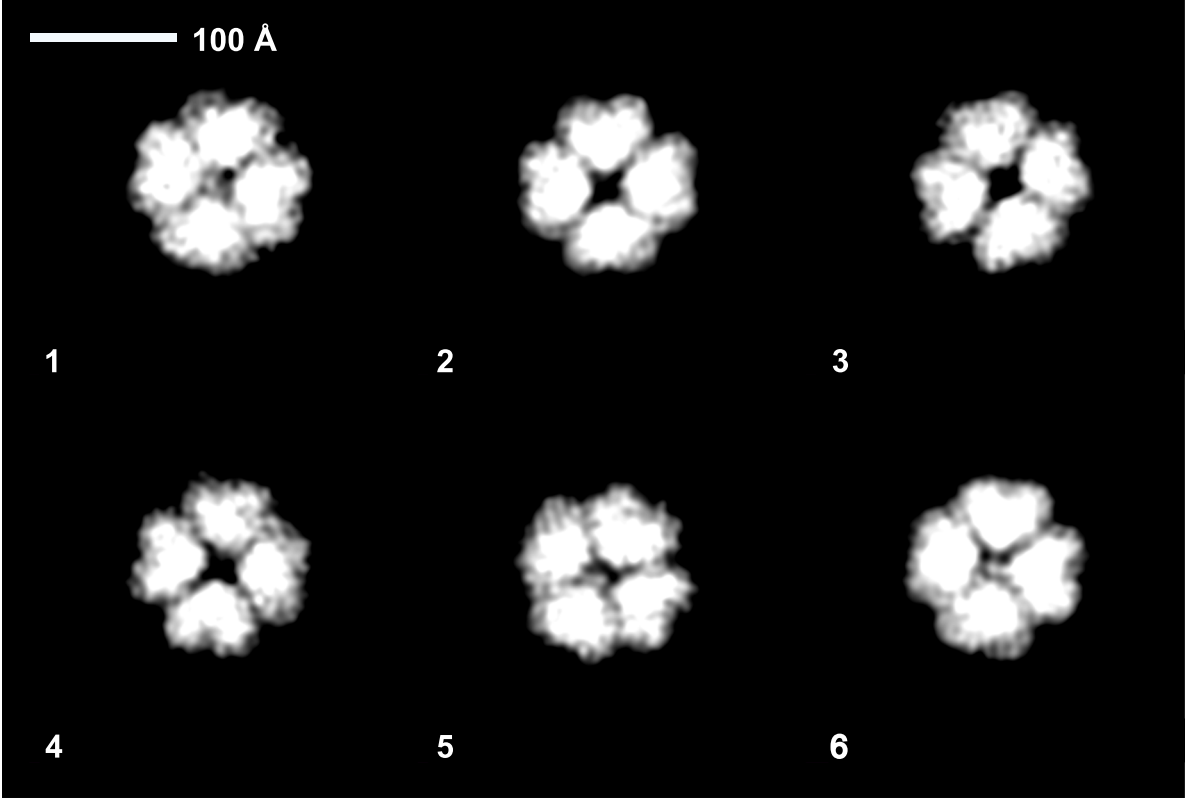

b
M
(kDa)
1
2
3
150
100
75
50
37
25
20
His14-bdSUMO-RbcL
GroEL
untagged RbcL

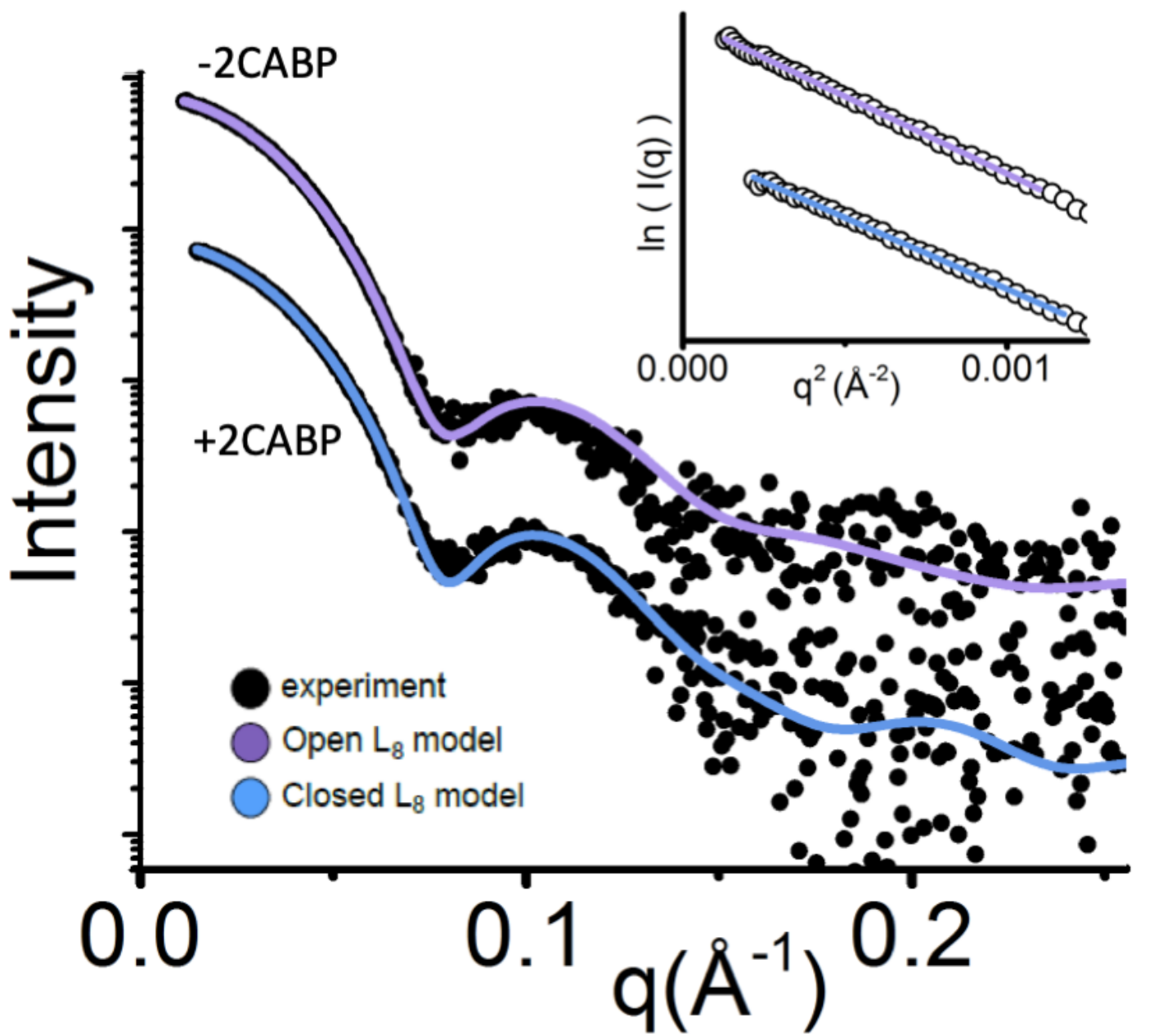
a

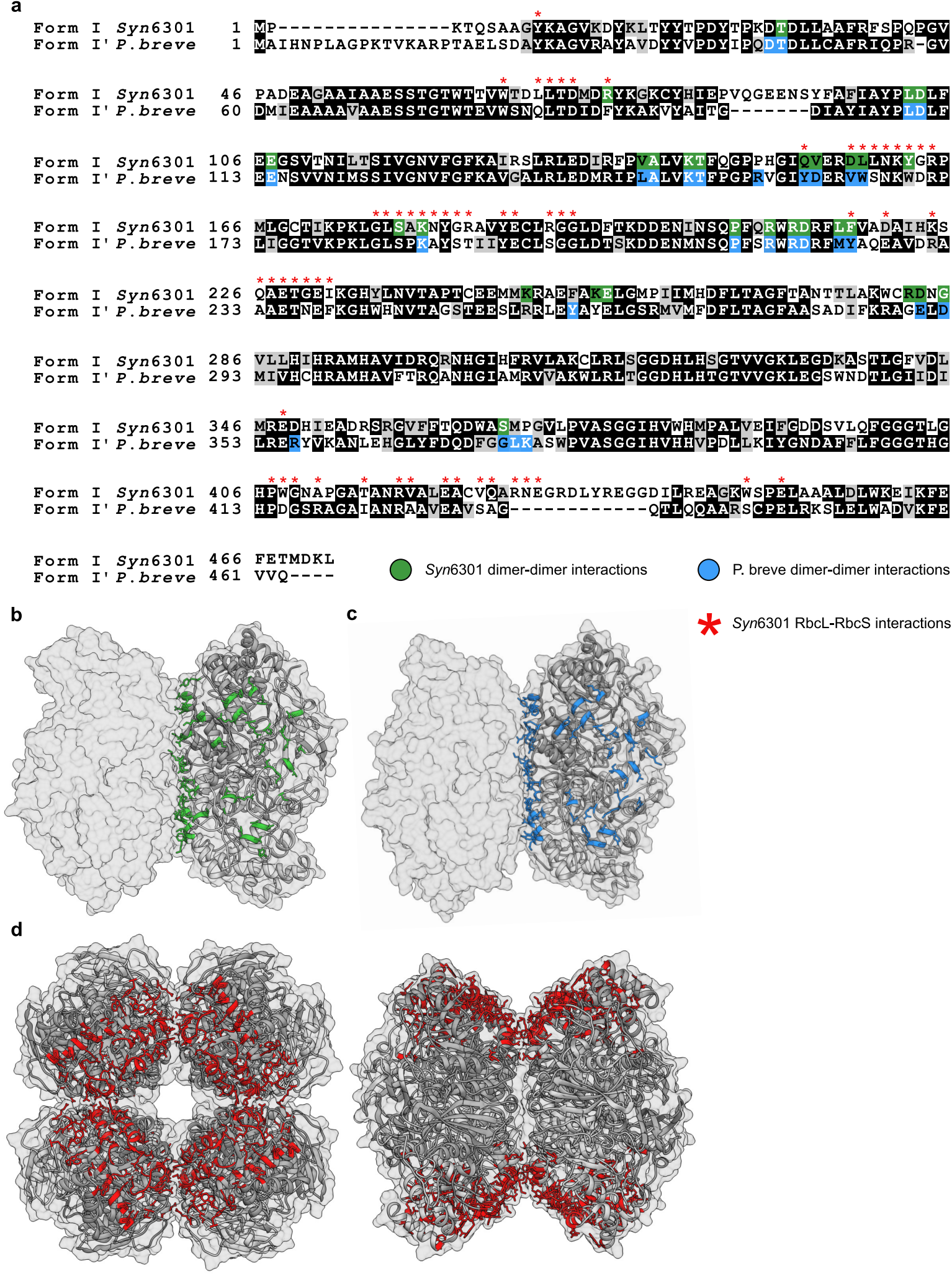

α6
βF
η
βG
β7
α7
Iprime_P.breve
340
350
360
370
380
390
Iprime_P.breve
Iprime_S241187
Iprime_S170907
IA_A.vinosum
IA_H.neapolitan
IB_Z.mays
IB_S.PCC6301
IC_N.hamburgens
IC_R.sphaeroide
ID_P.purpurea
ID_V.litorea
IE_O.trichoides
IE_K.aurantiaca
II_R.sphaeroide
II_R.rubrum
23_M.burtonii
23_M.halophilus
IIIB_A.boonei
IIIB_T.kodakare
IIIA_M.liminata
IIIA_M.hungatei
loop 6
Rubisco assembly domain (Form II/III)
β8
αK
α8
αL
αM
Iprime_P.breve
400
410
420
430
440
450
460
Iprime_P.breve
Iprime_S241187
Iprime_S170907
IA_A.vinosum
IA_H.neapolitan
IB_Z.mays
IB_S.PCC6301
IC_N.hamburgens
IC_R.sphaeroide
ID_P.purpurea
ID_V.litorea
IE_O.trichoides
IE_K.aurantiaca
II_R.sphaeroide
II_R.rubrum
23_M.burtonii
23_M.halophilus
IIIB_A.boonei
IIIB_T.kodakare
IIIA_M.liminata
IIIA_M.hungatei
C-terminal ext. (Form I)

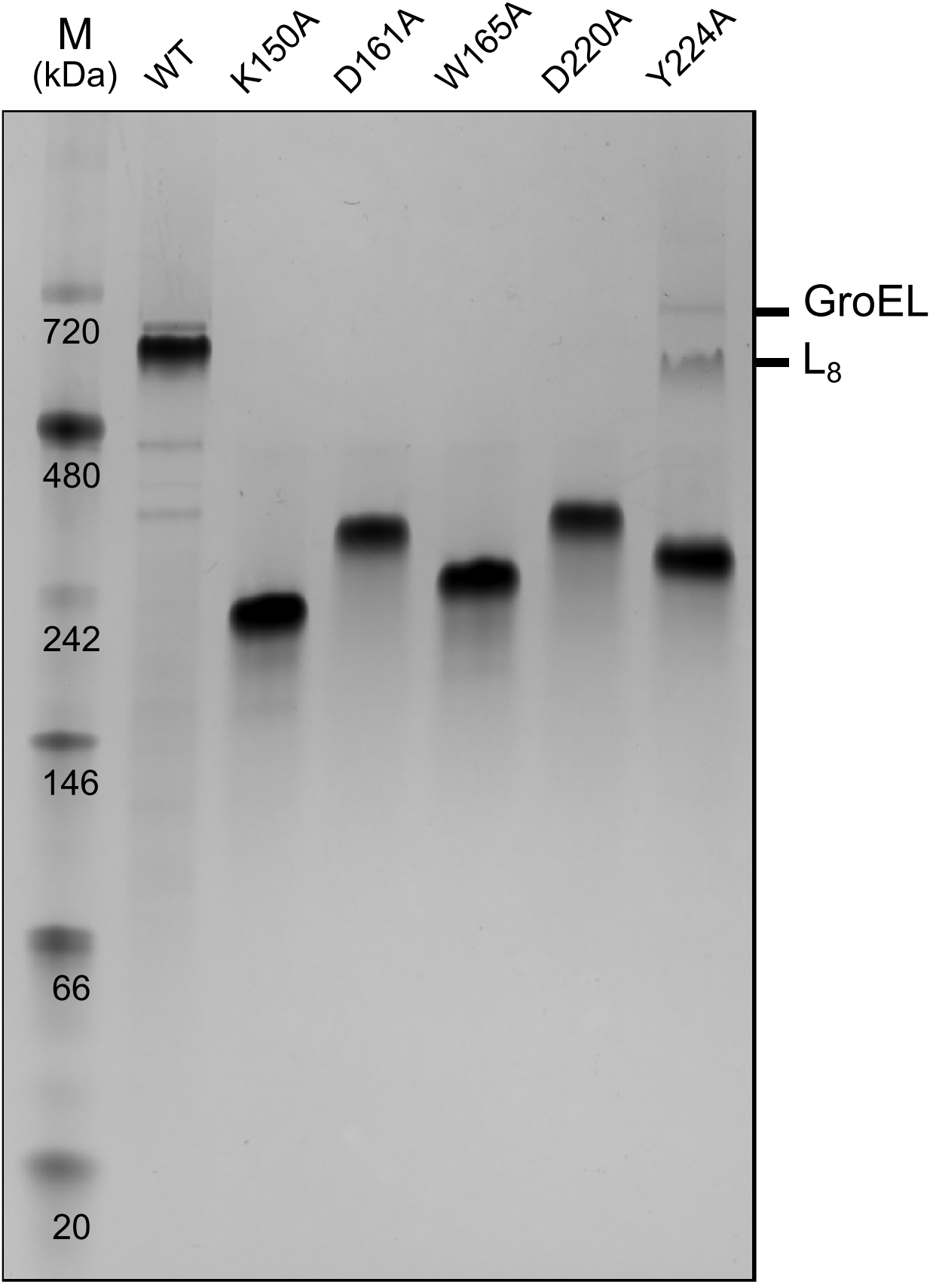
b

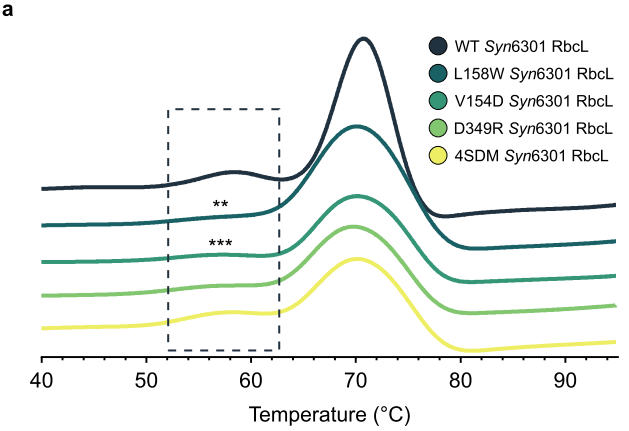


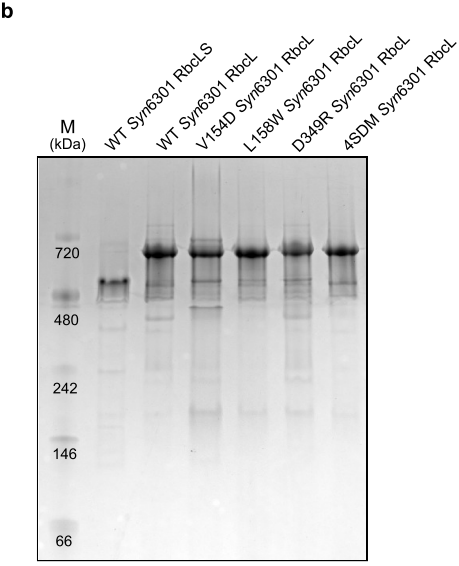

100 Å
1
2
3
4
5
6


-2CABP
+2CABP
Intensity
experiment
Open L8 model
Closed L8 model
0.0
0.1
0.2
q(Å-1)
ln (I(q))
0.000
0.001
q2 (Å-2)


a
Form I Syn6301 1 MP----------KTQSAAGYKAGVKDYKLTYYTPDYTPKDTDLLAAFRFSPQPGV
Form I' P.breve 1 MAIHNPLAGPKTVKARPTAELSDAYKAGVRAYAVDYYVPDYIPQDTDLLCAFRIQPR-GV
Form I Syn6301 46 PADEAGAAIAAESSTGTWTTVWTDLLTDMDRYKGKCYHIEPVQGEENSYFAFIAYPLDLF
Form I' P.breve 60 DMIEAAAAAVAAESSTGTWTEVWSNQLTDIDFYKAKVYAITG-------DIAYIAYPLDLF
Form I Syn6301 106 EEGSVTNILTSIVGNVFGFKAIRSLRLEDIRFPVALVKTFQGPPHGIQVERDLLNKYGRP
Form I' P.breve 113 EENSVVNIMSSIVGNVFGFKAVGALRLEDMRIPLALVKTFPGPRVGIYDERVWSNKWDRP
Form I Syn6301 166 MLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENINSQPFQRWRDRFLFVADAIHKS
Form I' P.breve 173 LIGGTVKPKLGLSPKAYSTIIYECLSGGLDTSKDDENMNSQPFSRWRDRFMYAQEAVDRA
Form I Syn6301 226 QAETGEIKGHYLNVTAPTCEEMMKRAEFAKELGMPIIMHDFLTAGFTANTTLAKWCRDNG
Form I' P.breve 233 AAETNEFKGHWHNVTAGSTEESLRRLEYAYELGSRMVMFDFLTAGFAASADIFKRAGELD
Form I Syn6301 286 VLLHIHRAMHAVIDRQRNHGIHFRVLAKCLRLSGGDHLHSGTVVGKLEGDKASTLGFVDL
Form I' P.breve 293 MIVHCHRAMHAVFTRQANHGIAMRVVAKWLRLTGGDHLHTGTVVGKLEGSWNDTLGIIDI
Form I Syn6301 346 MREDHIEADRSRGVFFTQDWASMPGVLPVASGGIHVWHMPALVEIFGDDSVLQFGGGTLG
Form I' P.breve 353 LRERYVKANLEHGLYFDQDFGGLKASWPVASGGIHVHHVPDLLKIYGNDAFFLFGGGTHG
Form I Syn6301 406 HPWGNAPGATANRVALEACVQARNEGRDLYREGGDILREAGKWSPELAAALDLWKEIKFE
Form I' P.breve 413 HPDGSRAGAIANRAAVEAVSAG----------QTLQQAARSCPELRKSLELWADVKFE
Form I Syn6301 466 FETMDKL
Form I' P.breve 461 VVQ----
Syn6301 dimer-dimer interactions
P. breve dimer-dimer interactions
Syn6301 RbcL-RbcS interactions
b
c
d


M
(kDa)
WT
K150A
D161A
W165A
D220A
Y224A
720
480
242
146
66
20
GroEL
L_8





Rubsico	Tm (°C)	*n*	p-value
WT *Syn*6301 RbcL	58.6 ± 0.2	4	ref.
L158W *Syn*6301 RbcL	57.7 ± 0.3	4	0.0037
V154D *Syn*6301 RbcL	57.3 ± 0.3	4	0.0007
D349R *Syn*6301 RbcL	57.9 ± 0.9	4	ns
4SDM *Syn*6301 RbcL	58.6 ± 0.1	4	ns