Title

Deep-branching evolutionary intermediates reveal structural origins of form I rubisco

Journal

Current Biology, 33(24)

ISSN

0960-9822

Authors

Liu, Albert K

Kaeser, Benjamin

Chen, LinXing

et al.

Publication Date

2023-12-01

DOI

10.1016/j.cub.2023.10.053

Article

Current Biology

Deep-branching evolutionary intermediates reveal structural origins of form I rubisco

Highlights

- Form Iα and I″ rubisco help elucidate the evolution of form I oligomerization
- Form Iα rubisco adopts a dimeric assembly in solution
- Form I″ rubisco adopts an octameric assembly without small subunits
- A unique insertion primed form I″ enzymes for hetero-oligomerization

Authors

Albert K. Liu, Benjamin Kaeser, LinXing Chen, ..., Eva Nogales, Jillian F. Banfield, Patrick M. Shih

Correspondence

pmshih@berkeley.edu

In brief

Liu et al. structurally characterize the assemblies of deep-branching clades of rubisco, identifying representatives of evolutionary intermediates that help elucidate the evolution of the globally dominant form I enzyme.

Liu et al., 2023, Current Biology *33*, 5316–5325
December 18, 2023
https://doi.org/10.1016/j.cub.2023.10.053

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

Deep-branching evolutionary intermediates reveal structural origins of form I rubisco

Albert K. Liu,[1,2,3,4,18] Benjamin Kaeser,[5,18] LinXing Chen,[6,7] Jacob West-Roberts,[8] Leah J. Taylor-Kearney,[1,2,3] Adi Lavy,[9] Damian Günzing,[10] Wen-Jun Li,[11,12] Michal Hammel,[13] Eva Nogales,[5,13,14,15] Jillian F. Banfield,[7,8,9,16,17] and Patrick M. Shih[1,2,3,8,19,20,*]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA, USA
[4]Biochemistry, Molecular, Cellular and Developmental Biology Graduate Group, University of California, Davis, Davis, CA 95616, USA
[5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[6]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720, USA
[7]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA 94720, USA
[8]Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA 94720, USA
[9]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720, USA
[10]Department of Physics, University of Duisburg-Essen, 47057 Duisburg, Germany
[11]State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, P.R. China
[12]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, P.R. China
[13]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[14]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA 94720, USA
[15]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
[16]School of Geography, Earth and Atmospheric Sciences, University of Melbourne, Melbourne, VIC 3053, Australia
[17]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[18]These authors contributed equally
[19]X (formerly Twitter): @LabShih
[20]Lead contact
*Correspondence: pmshih@berkeley.edu
https://doi.org/10.1016/j.cub.2023.10.053

SUMMARY

The enzyme rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) catalyzes the majority of biological carbon fixation on Earth. Although the vast majority of rubiscos across the tree of life assemble as homo-oligomers, the globally predominant form I enzyme—found in plants, algae, and cyanobacteria—forms a unique hetero-oligomeric complex. The recent discovery of a homo-oligomeric sister group to form I rubisco (named form I′) has filled a key gap in our understanding of the enigmatic origins of the form I clade. However, to elucidate the series of molecular events leading to the evolution of form I rubisco, we must examine more distantly related sibling clades to contextualize the molecular features distinguishing form I and form I′ rubiscos. Here, we present a comparative structural study retracing the evolutionary history of rubisco that reveals a complex structural trajectory leading to the ultimate hetero-oligomerization of the form I clade. We structurally characterize the oligomeric states of deep-branching form Iα and I″ rubiscos recently discovered from metagenomes, which represent key evolutionary intermediates preceding the form I clade. We further solve the structure of form I″ rubisco, revealing the molecular determinants that likely primed the enzyme core for the transition from a homo-oligomer to a hetero-oligomer. Our findings yield new insight into the evolutionary trajectory underpinning the adoption and entrenchment of the prevalent assembly of form I rubisco, providing additional context when viewing the enzyme family through the broader lens of protein evolution.

INTRODUCTION

Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) serves as the entry point for nearly all inorganic carbon into the biosphere.[1,2] This enzyme fixes environmental carbon dioxide to its substrate, ribulose-1,5-bisphosphate (RuBP), to form a six-carbon intermediate, which is subsequently cleaved into two three-carbon molecules of 3-phosphoglycerate (3-PGA) for the downstream synthesis of organic compounds, most notably in the Calvin-Benson-Bassham cycle in oxygenic photosynthesis.[1,3–5] Though multiple forms of rubisco have been identified, form I enzymes are the predominant assembly, being involved in photosynthesis and representing over 90% of rubisco in nature.[5–7]

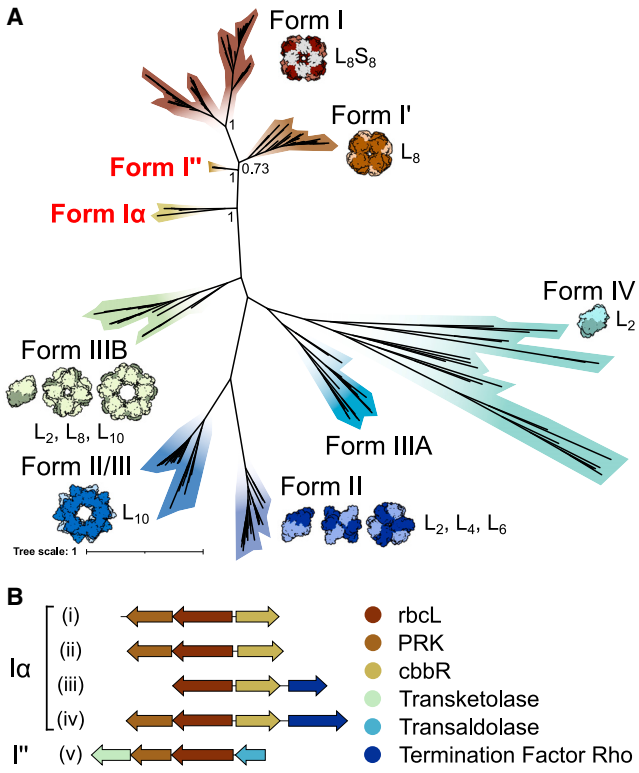

Figure 1. Metagenomic rubisco sequences cluster with novel deep-branching clades

(A) Unrooted maximum likelihood phylogenetic tree of rubisco forms. Felsenstein bootstrap values indicated at key nodes. Large subunits within rubisco dimers colored in alternating light and dark shades to illustrate dimeric interface. Form I PDB: 1RBL; form I' PDB: 6URA; form IV PDB: 2QYG; form II PDB: 5RUB, PDB: 7T1C, PDB: 5C2C; form II/III PDB: 5MAC; form IIIB PDB: 8DHT, PDB: 2CWX, PDB: 1GEH.

(B) Clinker diagram indicating genes surrounding rubisco large subunit (*rbcL*) gene. Form name indicated on left. Sequences in order: (i) PLM2_5_b1_jun17_scaffold_3874 (form Iα); (ii) GWC2_Chloroflexi_73_18_gwc2_scaffold_3548 (form Iα); (iii) Rifsed_csp1_10ft_1_scaffold_31868 (form Iα); (iv) RBG_16_CP_70_13_RBG_16_scaffold_22096 (form Iα); and (v) GXS_idba_scaffold_1654 (form I′′). All form Iα sequences correspond to the Limnocylindria class; the form I′′ sequence corresponds to the Firmicutes phylum.

See also Figures S1 and S2 and Table S1.

The basic structure of a functional rubisco is a dimer of two large subunits (RbcL, ~50 kDa) that assemble in a head-to-tail manner, resulting in the proper structure of the active site for catalysis.[3,6] While all other forms of rubisco (i.e., forms II, II/III, and III) are composed of these repeating dimers in various higher-order assemblies, form I enzymes are unique in their incorporation of an additional small subunit (RbcS, ~13–17 kDa) that cap the octameric RbcL core (L_8) at the junctions between dimers, resulting in a hexadecameric assembly (L_8S_8) (Figure 1A).[6,8–11] Form I rubisco has become structurally entrenched as a hexadecamer due to its strict dependence on the small subunits for catalysis, with enzymes losing activity and displaying decreased stability in their absence.[12–14] This entrenchment and the events leading up to it make the evolutionary trajectory of the form I clade anomalous among all rubiscos, as all other forms display multiple oligomeric states in each clade, highlighting the variance in higher-order structure outside of the form I clade.[6,7,9] This presents a unique opportunity to utilize rubisco as an interesting system for studying the evolution of oligomerization; patterns of both structural entrenchment and structural plasticity are observed within a single enzyme family, allowing for the broad sampling of sequences and assemblies to characterize the evolutionary trajectories giving rise to the diversity of extant rubisco.

Recent metagenomic studies have illuminated the evolutionary steps and missing lineages preceding the origin of the form I clade. We previously identified protein sequences that clustered between the form I clade and all other forms, named form I′.[14] Inspection of the metagenome-assembled genomes containing form I′ sequences revealed the absence of corresponding small subunit sequences, though other Calvin-Benson-Bassham cycle genes were identified.[14] The form I′ enzyme from *Candidatus* Promineofilum breve (*Ca.* P. breve) was found to adopt an octameric assembly similar to the octameric core found in form I enzymes, albeit without small subunits.[14] This discovery represents an evolutionary intermediate between other rubisco oligomers and the form I hexadecamer and, importantly, shows that octamerization occurred prior to small subunit acquisition.

Additional points of comparison are still needed to properly place the events that led to key evolutionary transitions. Specifically, the discovery of the form I′ clade alone does not reveal the origins of the form I clade. The divergence of the form I/I′ clades can be explained by two possible scenarios: (1) form I′ rubisco represent the homo-oligomeric state prior to the incorporation of the small subunit; (2) the form I′ clade lost the ability to bind the small subunit, either by losing the features driving hetero-oligomerization or by reverting back to a homo-oligomer. Deciphering the exact cause of this divergence is necessary to fully understand the possible structural entrenchment or plasticity that enabled the homo- to a hetero-oligomer transition.

As the breadth of metagenomic studies increases, rubisco sequences have been identified that do not cluster with any previously identified forms, with some revealing the existence of additional clades between form I and the remaining forms of rubisco.[15] Enhanced sampling across the rubisco phylogeny—especially in sparsely covered regions of the tree—is providing the requisite evolutionary reference points to begin to tease apart the origins of the form I clade. Two recent sister clades to form I rubisco, dubbed forms Iα and I′′, have been discovered in the lineage between forms I, I′, and IIIB, the next closest clade, providing much needed branches to resolve how this evolution unfolded.[15] However, our knowledge of the phylogenetic and structural properties of forms I′′ and Iα remains scant, as the phylogenetic order in which these two forms appear has proven challenging[7,13,15] (Figure 1A).

Schulz et al.'s recent study utilized the presence or absence of a single C-terminal insertion to constrain the topology of the form I′′ and Iα region in the rubisco phylogeny.[13] Due to the presence of this insertion in the form I and I′′ clades, but not in the form I′ or Iα clades, a parsimony-based argument was made to constrain the rubisco phylogeny, with their topology reflecting the divergence of the form I′′ and form I clades from an ancestral node and the form I′ clade diverging prior to both.[13] Using this constrained topology, ancestral sequence reconstruction (ASR) of

this node revealed an octameric enzyme containing the distinguishing C-terminal insertion, in agreement with their experimental understanding of the acquisition of hetero-oligomerization in form I rubisco.[13] However, the placement of the form I′, I″, and I clades on this phylogeny is poorly supported due the constraint imposed on this section of the phylogeny.[13] As a result, the poor support based on Schulz et al.'s makes the exact order in which these forms diverge unclear, and both additional taxa and unconstrained phylogenetic analyses are needed to better clarify this poorly resolved region in rubisco evolution.

Resolving this portion of the rubisco phylogeny is essential to understanding the key evolutionary steps in the transition from homo-oligomeric to hetero-oligomeric rubisco. Using molecular weight measurements, it has been posited that a form Iα enzyme adopts a dimeric assembly; however, no information is available regarding the assembly of form I″.[13,15] Without structural data from form I″, we cannot characterize the transition from a dimer to an octamer during the evolution of form I. As such, detailed structural characterization of new sequences in these intermediary clades will further our understanding of how the hetero-oligomeric rubisco evolved. Here, we investigate the assemblies of form Iα and I″ enzymes, further characterizing the trajectory and evolution of structural complexity leading up to the entrenchment of the form I clade.

RESULTS

Discovery of deep-branching rubisco sequences

Four rubisco large subunit protein sequences (RbcL) were identified in *Chloroflexota* genomes from aquifer sediment in Rifle, Colorado, via phylogeny and were designated as a new rubisco form, named Iα.[15] One of these sequences was truncated by approximately 100 residues, including known active site residues,[3] and was excluded from further analysis. Another form Iα sequence thought to originate from a member of the Limnocylindria class was found on an unbinned metagenomic contig, as identified by phylogeny utilizing reference sequences,[15] and samples obtained from the East River watershed, Gunnison County, Colorado.[16]

An additional RbcL protein sequence, clustering with the putative form I″ clade,[15] was also discovered in a sediment sample from the GongXiaoShe (GXS) hot spring in the Yunnan province of China, as collected in January 2016. The extracted genomic DNA was sequenced and analyzed to reveal a novel bacterial phylum ("Candidatus Kryptonia"),[17] and the contig containing the large subunit gene (*rbcL*) was taxonomically assigned to the bacteria domain. Manual curation and extension of the contig was attempted but failed as the sequencing coverage was insufficient, though taxonomic assignment of other genes indicated that it may have been isolated from a member of the Firmicutes phylum.

New metagenomic sequences bolster phylogenetic support of deep-branching intermediary clades

Utilizing the five metagenomic RbcL protein sequences, we generated a phylogeny of all forms of rubisco to query their positions across known forms (Figure 1A). The placement of two clades between form I′ and form IIIB rubisco corresponds to the previously identified form Iα, containing four of the sequences, while the remaining sequence clustered with the sparsely populated form I″.[13,15] In contrast with a previous study, our unconstrained maximum likelihood phylogenetic tree revealed that the form I″ clade is an intermediate between the form Iα and I′ clades (Figures 1A and S1); the alternative phylogeny, where the form I″ clade is instead found between the form I′ and I clades, was inferred from the presence of an insertion found in form I and I″ sequences that is absent in form I′ and Iα sequences.[13] At the branch point between form I″ and forms I′ and I, we report a Felsenstein bootstrap value of 1, compared with the alternative phylogeny from Schulz et al. reporting a value of 0.21 between form I″ and form I, and an approximate likelihood ratio of 0.[13] With the addition of four form Iα sequences and one form I″ sequence, support for the position of the form I″ clade between the form I′ and I clades increased in the absence of topological constraint, highlighting how the addition of new sequences can improve the robustness of phylogenetic support and overall confidence of the distribution of key rubisco clades.

Inspection of the region surrounding each *rbcL* gene sequence in the metagenomic contigs revealed a lack of corresponding small subunit (*rbcS*) genes (Figures 1B and S2). As bacterial rbcS is usually found within one or two genes of *rbcL*, this suggested that the rubisco encoded by these sequences would adopt a homomeric assembly.[18,19] Additionally, full genome scans searching for *rbcS* were performed on the genomes containing the four form Iα *rbcL* sequences, confirming the absence of the small subunit. However, the full genome of the form-I″-sequence-containing organism was not available for such an analysis, though it is likely that *rbcS* is not present in a cryptic site within its genome.[20,21] To structurally characterize the form Iα and I″ rubiscos, we synthesized the four form Iα genes in addition to the singular form I″ gene for heterologous protein expression and purification. Of the form Iα rubiscos, only one, a member of the Limnocylindria class (PLM2_5_b1_jun17_scaffold_3874), was soluble at a scale amenable for further analysis. Additionally, the form I″ rubisco, from the Firmicutes phylum (GXS_idba_scaffold_1654), was also soluble in quantities necessary for analysis. To validate the function of these metagenomic rubiscos, we conducted a spectroscopic kinetic assay to verify their catalytic activity. Indeed, the form Iα Limnocylindria sp. and form I″ Firmicutes sp. enzymes displayed carboxylase activity, albeit at a low rate, thus indicating that their active sites were indeed present and structured properly for catalysis (Table S1). With validation that our form Iα and I″ representatives were active, we proceeded with structural characterization of both.

Dimeric form Iα represents an oligomeric precursor

Previously, characterization of a form Iα rubisco by mass photometry yielded a molecular weight consistent with a dimeric assembly.[13] To query the solution-state assembly of our form Iα enzyme, we heterologously expressed and purified the aforementioned member of the Limnocylindria class, which shared 56.1% sequence identity with the previously characterized form I′ enzyme from *Ca.* P. breve.[14] As form Iα had previously been identified as adopting a dimeric assembly, we generated a homo-dimeric AlphaFold model using the form Iα Limnocylindria sp. protein sequence

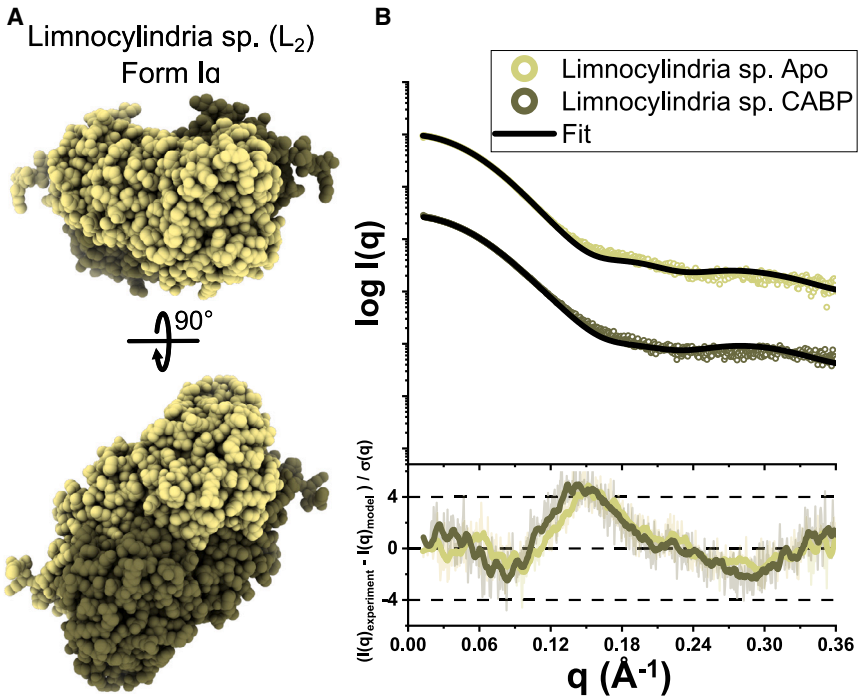

Figure 2. Form Iα enzyme forms dimer in solution

(A) AlphaFold model of the Limnocylindria sp. form Iα enzyme.

(B) SAXS curves for Limnocylindria sp. unbound (Apo) and bound to transition state analog (CABP) against AlphaFold model. Fit-residuals indicated below.

utilizing the ColabFold webtool for subsequent analyses (Figure 2A).[22]

To determine the solution-state assembly of this form Iα enzyme, we analyzed a purified sample by size exclusion chromatography coupled with small-angle X-ray scattering and multi-angle light scattering (SEC-SAXS-MALS), enabling the determination of the molecular weight of the enzyme as well as collection of a small-angle X-ray scattering (SAXS) profile (Figure 2B).[23–25] SEC-SAXS-MALS data were collected in both the absence and presence of the rubisco transition state analog 2-carboxyarabinitol 1,5-bisphosphate (2-CABP) in order to approximate the catalytic conformation of the enzyme active site and backbone.[14,26] During our previous characterization of the *Ca.* P. breve form I′ enzyme, analysis of the CABP-unbound (apo) and CABP-bound conditions on native polyacrylamide gel electrophoresis (native PAGE) showed a change in migration between the two conditions, though this was later revealed through SAXS analysis to be a phenomena unrelated to oligomeric state shift but rather a conformational change induced by CABP binding.[14] Thus, our analyses of the form Iα Limnocylindria sp. enzyme queried both conditions to address the potential for a shift in oligomeric state.[11,14,27] In both conditions, the SAXS data indicate a dimeric assembly in solution when using the predicted structure of the form Iα Limnocylindria sp. enzyme (Figure 2B). This observation is further supported by the MALS-derived molecular weights at 99.9 kDa for the unbound condition and 104.4 kDa for the 2-CABP-bound condition, in accordance with a homo-dimeric assembly of two large subunits at ~50 kDa each. This result is in agreement with a previous mass photometry measurement on a different form Iα enzyme, where the determined molecular weight was 98.2 kDa for a dimer.[13]

The dimeric assembly of form Iα is thus in agreement with the current model of the evolution of the form I enzyme octameric core. Considering that the minimal functional unit of rubisco is the dimeric assembly, the common ancestor of all rubiscos is widely considered to have been dimeric, which has also been experimentally demonstrated in form II enzymes.[1,9] Our structural characterization of the form Iα Limnocylindria sp. enzyme supports this concept, as the dimeric form Iα enzymes precede the adoption of the octameric state and acquisition of the small subunit that would give rise to form I hexadecameric assemblies. Thus, it is likely that, from a dimeric common ancestor, a radiation event gave rise to the clades of rubisco presently observed in nature, whereby a dimeric ancestor at the origin of each clade preceded the evolution of more complex multimeric assemblies.

Form I″ rubisco adopts an octameric assembly

As no form I″ enzymes had previously been structurally characterized, we lacked the information required to elucidate how this unique lineage places within the evolutionary transition to the form I clade. To address this knowledge gap, we synthesized, expressed, and purified the metagenomic form I″ rubisco from the Firmicutes phylum, which shares 66.3% sequence identity with the form I′ *Ca.* P. breve enzyme. We determined the structure of this form I″ enzyme using single-particle cryoelectron microscopy (cryo-EM) to an overall resolution of 2.2 Å, revealing an octameric homo-oligomer similar to that of form I′ (Figures 3A, S3, and S4). This assembly was also supported by SEC-SAXS-MALS analysis for both unbound and 2-CABP-bound complexes (Figure 3B), with MALS-determined molecular weights at 385.4 and 420.5 kDa, respectively. Prior to the characterization of the Firmicutes sp. enzyme, the oligomeric state of form I″ enzymes was unknown, with only four other sequences identified at present.[13] Our cryo-EM structure reveals the nuances in the stepwise evolution of octameric rubisco in early progenitors leading to the form I clade.

Form I″ interface conservation illuminates trends in rubisco oligomerization

Previous analysis of the interdimer interface of the *Ca.* P. breve form I′ rubisco identified key residues responsible for maintaining the octameric state.[14] Comparison of the homologous Firmicutes sp. form I″ interface residues with the ten key residues from the form I′ *Ca.* P. breve sequence showed the conservation of seven out of ten, with four of these seven also conserved in form I (Figure 4A). Superposition of the form I″ and I′ structures

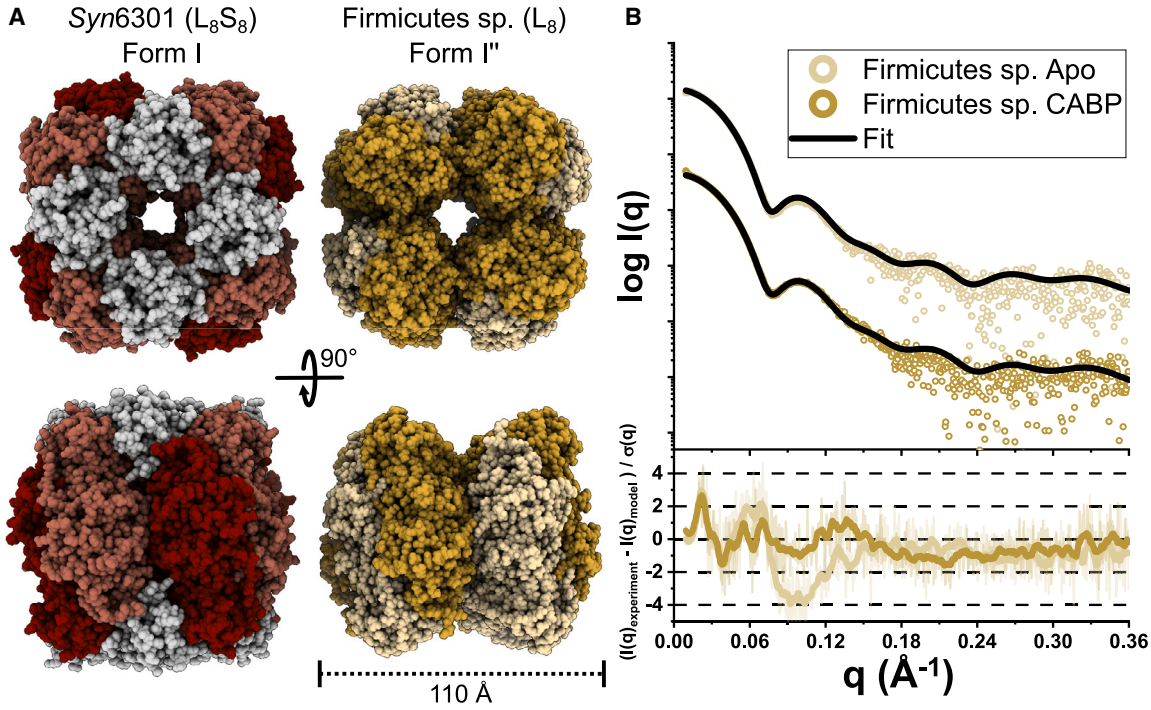

Figure 3. Form I′′ enzyme adopts octameric assembly

(A) Comparison of *Synechococcus elongatus* PCC6301 form I enzyme (PDB: 1RBL) and cryo-EM structure for Firmicutes sp. form I′′ enzyme PDB: 8U66.

(B) SAXS curves for Firmicutes sp. enzyme unbound (Apo) and bound to transition state analog (CABP) against the cryo-EM structure (continuous curves). Fit-residuals indicated below.

See also Figure S4.

further revealed the close proximity of these seven conserved residues at the interdimer interface, suggesting a similar role of these positions in maintenance of the form I′′ interface (Figure 4B). Comparison of form I′′ and form I′ non-conserved residues shows likely functional similarity that is lost in the presence of the small subunit in form I rubisco (Figure 4C). The previously reported loss of electrostatic interactions at the dimer-dimer interface in the hydrophobic Val 154 and Leu 158 residues in the form I Syn6301 structure (from Asp 161 and Trp 165 in *Ca* P. breve) is maintained in the homologous Gln 143 and Arg 147 residues in Firmicutes sp. form I′′, further illustrating the dependency of form I rubisco on the stabilizing mechanism conferred by small subunit binding.[14] These observations are in agreement with the evolutionary trajectory experienced by form I rubisco, where forms I′ and I′′ contain interdimeric interactions that are not present in form I, entrenching the octameric state.

Beyond the octameric core, the small subunit is another hallmark of form I rubisco assembly.[6,28] Though multiple roles have been suggested for the small subunit, one well-characterized function is the stabilization of the L_8S_8 holoenzyme.[13,14,29] In the absence of the small subunit, an isolated form I octameric core suffers from decreased thermal stability relative to both its native L_8S_8 assembly and the form I′ L_8, and ASR of form I enzymes has revealed that small subunit-less octamers are less soluble.[13,14] In accordance with these observations, it is possible that the stability conferred by the small subunit allowed for the exploration of large subunit sequence space, permitting the loss of stabilizing contacts at the interface between adjacent dimers in favor of mutations that result in beneficial catalytic performance.

Loss of C-terminal insertion precluded form I′ from evolving hetero-oligomeric assembly

The large subunits of form I rubiscos contain a unique C-terminal insertion that interacts with the small subunit, and previously characterized forms of rubisco lacking small subunits do not contain this insertion (Figure 5A).[13,14] This observation has been used to infer the evolutionary order of clades preceding form I. Thus, as form I′′ sequences contain this insertion, Schulz et al. place the form I′′ clade sibling to form I, rather than the form I′ clade as previously hypothesized.[13,14] In accordance with observations of this insertion, our presented form Iα sequences lack this insertion and the form I′′ Firmicutes sp. sequence contains it (Figures 5A and S5A). To query the nature of this insertion in form I′′, we performed a structural alignment of the *Syn*6301 form I and Firmicutes sp. form I′′ large subunits (Figure 5B). The C-terminal insertion in the form I *Syn*6301 large subunit contains an Arg-Asn-Glu motif that interacts with the small subunit. However, of the homologous residues in the form I′′ Firmicutes sp. enzyme, only the final glutamic acid residue is conserved (Figure 5B). Although the presence of this insertion appears to prime the octameric core for acquisition of the small subunit, the required residues for binding are not present, reflecting the intermediary nature of form I′′ in the form I rubisco evolutionary trajectory. Furthermore, the complete absence of this insertion from form I′ enzymes may preclude the small-subunit-binding

A

Form I′	Form I″	Form I
P. breve	Firmicutes sp.	*Syn*6301
Asp110	Asp92	Asp103
Glu114	Glu96	Glu107
Lys150	Lys132	Lys143
Asp161	Gln143	Val154
Trp165	Arg147	Leu158
Asp220	Asp202	Asp213
Tyr224	Tyr206	Phe217
Arg256	Arg238	Lys249
Asp292	Gly274	Gly285
Arg356	Glu338	Asp349
Tyr357	Phe339	His350
Lys376	Lys358	Pro369

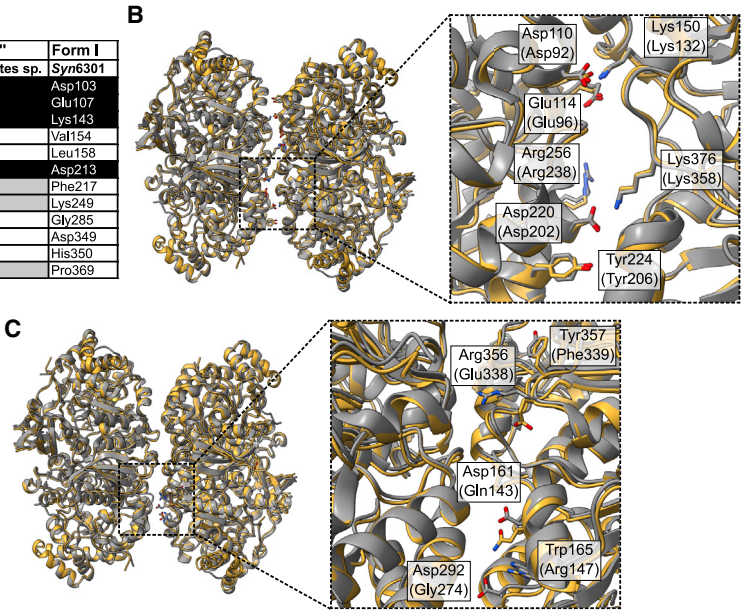

Figure 4. Form I′ and I″ interdimer interfaces are highly conserved

(A) Table of key residues in form I′ at the interdimer interface compared with homologous residues in form I″ and form I enzymes. Sample names indicated below each form in the second row. Black background indicates conserved in both; gray background indicates conserved in form I″.

(B and C) Overlay of *Ca.* P. breve form I′ (gray, PDB: 6URA) and Firmicutes sp. form I″ (orange, PDB: 8U66) dimer pairs. Conserved (B) and non-conserved (C) residues indicated in cutaway. Form I′ residue notation indicated, with form I″ in parentheses below.

event by reducing the number of stabilizing contacts formed between the large and small subunits, thus decreasing the likelihood that form I′ octamers would be capable of assembling a hexadecameric complex with an exogenous small subunit (Figure S5B).

Contextualizing the presence and identity of this insertion, it is possible to retrace the evolution of form I rubisco oligomerization in combination with the structural characterization of forms Iα and I″ conducted here (Figures 5C and 5D). From an ancestral intermediate immediately following form IIIB, a dimeric assembly was captured in the form Iα clade, representing the enzyme preceding the innovation of octamerization. Subsequently, octamerization was captured by the extant form I″ enzyme, which contained the C-terminal insertion that could interact with the small subunit (Figure 5D). Following form I″, loss of the insertion in form I′ sequences likely abolished the possibility for additional interface contacts between the large and small subunits, potentially precluding the observed octameric core from binding small subunits (Figure 5D). Finally, acquisition of the small subunit resulted in entrenchment of the hexadecameric form I assembly, with subsequent mutations at the interdimer interfaces resulting in dependence on the small subunits for stability and catalysis (Figure 5D).

DISCUSSION

The use of metagenomics has expanded our understanding of the distribution of rubisco sequences found in nature and revealed previously undiscovered forms.[7,30] In parallel, structural characterization of metagenomic rubisco forms has enabled the discovery of a diversity of oligomeric states, illustrating multiple evolutionary trajectories experienced by the enzyme.[9,14,31] For example, we recently demonstrated that form II rubiscos are capable of forming three distinct homo-oligomeric assemblies, in comparison with the single hetero-oligomeric hexadecamer found in the form I clade.[9] Although previous studies have characterized the molecular features governing homo-oligomeric rubisco, the continued combination of metagenomic and structural studies will permit the identification of distinct structural features in sibling clades and further our understanding of how rubisco evolved into the hetero-oligomeric assembly that has become the predominant form on Earth. Our characterization of representative members of forms I″ and Iα further resolves the means by which an ancestral rubisco evolved into the ubiquitous form I clade, i.e., that a dimeric ancestor first innovated the octameric state prior to the acquisition of the small subunit, represented by extant enzymes in forms Iα, I″/I′, and I, respectively.

There is a dearth of experimental characterization of both form I″ and Iα rubiscos, owing to their relatively recent discovery.[13,15] Here, we present the first structure of a form I″ octamer, as well as solution-state data supporting the dimeric assembly of form Iα, revealing evolutionary intermediates preceding the evolution of the form I hexadecamer. Though both the Limnocylindria sp. form Iα and the Firmicutes sp. form I″ showed minimal carboxylase activity, a more exhaustive characterization of all kinetic parameters could constitute future work to further investigate trends in rubisco activity. Comparison between the form I″ atomic structure and the previously determined form I′ structure shows that the form I″ enzyme displays both a partially conserved interdimeric interface as well as a distinctive C-terminal insertion present in all canonical form I enzymes that is absent in form I′ sequences. Furthermore, characterization of the insertion suggests that truncation of the insertion helix would eliminate stabilizing contacts with a small subunit, likely precluding form I′ octamers from ever adopting a hexadecameric assembly and supporting the hypothesis that form I′ rubisco lost the ability to bind the small subunit relative to the acquisition of the insertion at the branch point with the form I″ clade.

Our phylogeny places form I′ rubisco as the immediate sibling clade to form I rubisco. This is in contrast to the rubisco phylogeny proposed in Schulz et al., which argues that the form I″ clade is more closely related to form I rubisco.[13] Notably, Schulz et al. constrained their rubisco phylogeny to place the form I″ clade sibling to the form I clade based on the parsimonious assumption that there was a single gain of the C-terminal insertion that ultimately enabled the binding of the small subunit in the form I clade.[13] However, by constraining the topology based on this

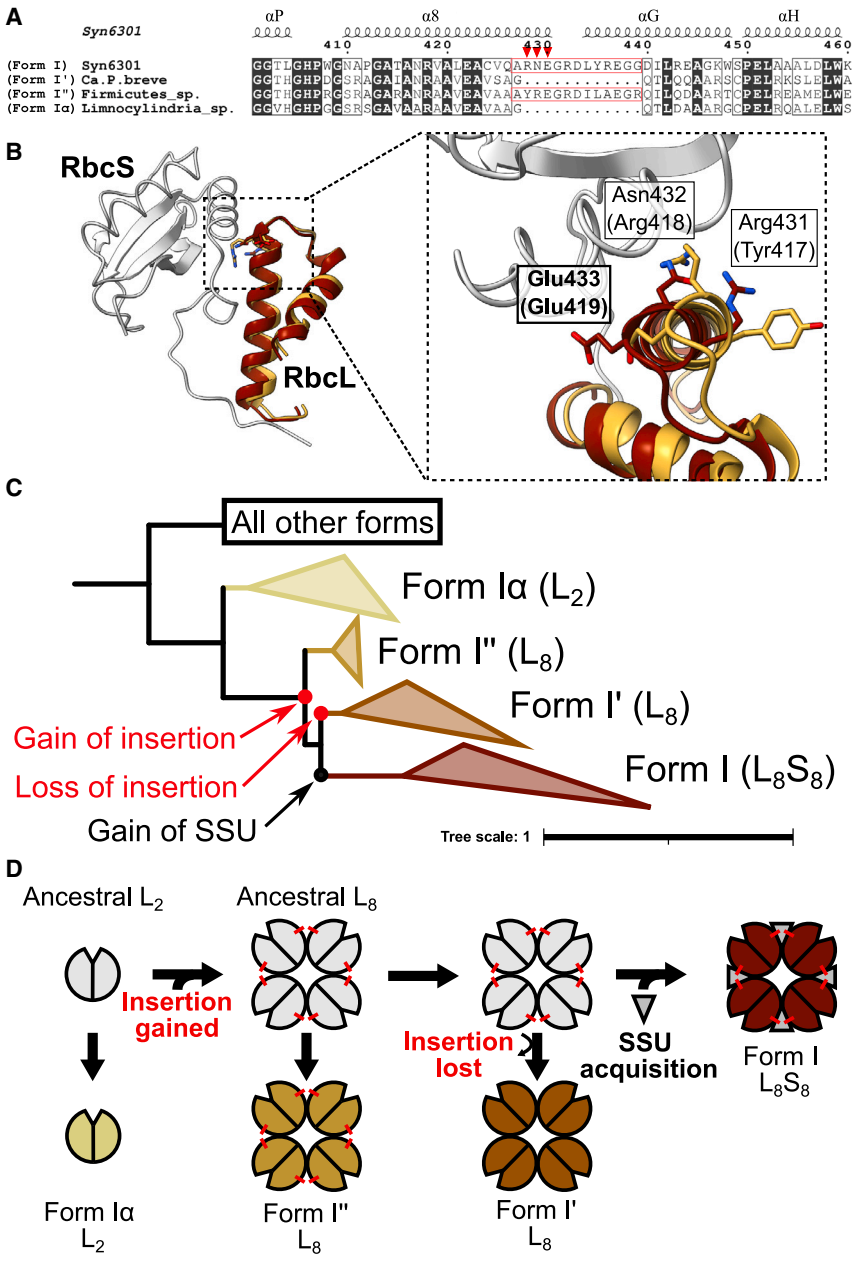

Figure 5. Form I″ C-terminal domain insertion shows minimal functional conservation

(A) Truncated sequence alignment of form I *Syn*6301, form I″ Firmicutes sp., form I′ *Ca.* P. breve, and form Iα Limnocylindria sp.; C-terminal insertion indicated in red box. Residues interacting with small subunit (SSU) in *Syn*6301 indicated by red arrowheads. Secondary structure labels assigned according to rubisco nomenclature from Knight et al.[8]

(B) Structural alignment of C-terminal insertion in *Syn*6301 (PDB: 1RBL) and Firmicutes sp. large subunits (PDB: 8U66). *Syn*6301 large subunit shown in red, Firmicutes sp. large subunit shown in orange. *Syn*6301 small subunit shown in silver. In cutaway, insertion residues interacting with SSU indicated in boxes; *Syn*6301 notation above, Firmicutes sp. notation in parentheses.

(C) Schematic of C-terminal insertion evolutionary trajectory leading up to the form I clade.

(D) Cartoon diagram of proposed form I evolutionary trajectory. C-terminal insertion represented in red.

See also Figure S5.

assumption, phylogenetic support for the divergence of the form I″ clade is substantially lower than that in our phylogeny at the node where the form I′ and I clades diverge. ASR of the form I″/I node from Schulz et al. revealed an octamer that contains the insertion[13]; however, this sequence was reconstructed based on their constrained phylogeny. Future studies that may compare ancestrally reconstructed sequences derived from both topologies will provide a deeper understanding of the nature and contributory role of C-terminal insertion in form I rubisco evolution. Importantly, ancestral sequences are limited to interpretation in the context of the particular phylogeny and dataset utilized. Although the use of an ASR-based approach to query rubisco evolution has yielded key enzymatic insights,[9,13,32,33] the extant sequences that are used to build the input alignment and phylogeny for such analyses ultimately dictate the output sequences. Thus, the future discovery and addition of new rubisco sequences that populate this portion of the tree will improve our sampling and help resolve this sparsely covered region of the rubisco phylogeny, which will ultimately improve the robustness of ASR studies.

Only a handful of sequences have been discovered in the clades nearest to the origin of form I. The sparse sampling in this pivotal region of the rubisco phylogeny highlights the importance of uncovering additional form I″ and Iα sequences. Further characterization of these enzymes will provide the added resolution required to further elucidate the molecular mechanisms governing the increase in complexity from a homo-oligomer to a hetero-oligomer. The presence of the small subunit in form I rubisco represents a mechanism of structural entrenchment, whereby the form I hexadecamer is structurally dependent on the presence of small subunits and their loss results in destabilization of the holoenzyme, serving as a selective force against mutations that would reverse small subunit binding.[13,34,35] However, in the broader context of the evolution of rubisco's oligomeric state, the selective pressure imposed by the dependency on small subunits precluded form I enzymes from innovating and adopting multiple assemblies, while other forms of rubisco (i.e., forms II, II/III, III) have been observed to adopt multiple homo-oligomeric states within their respective clades.[9,11,27] Thus, the comparison of form I and other forms of rubisco represents differences in lineages displaying protein structural entrenchment

or structural plasticity, respectively, providing a touchpoint in our understanding of how one enzyme family may evolve differing patterns of oligomerization. Our comparative analyses between forms I′, I″, and Iα rubisco have enabled the elucidation of the history of the form I′ clade as well as the ordering of the evolutionary events leading up to small subunit acquisition in form I enzymes. The presence of a small-subunit-binding insertion in form I″ rubisco indicates that the precursor structural features necessary to form a hetero-oligomeric complex were acquired early in the form I evolutionary trajectory; however, the actual binding event and entrenchment of the hexadecameric state occurred much later on, following the loss of the requisite insertion region in the form I′ clade.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL DETAILS
- METHOD DETAILS
  - Phylogenetic analysis
  - Protein modeling
  - Plasmids
  - Rubisco expression and purification
  - Spectroscopic kinetic parameter measurements
  - Cryo-electron microscopy sample preparation
  - Single-particle cryo-electron microscopy data collection, image processing, and model building
  - Size exclusion chromatography coupled small-angle X-ray scattering with in-line multiangle light scattering experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.10.053.

ACKNOWLEDGMENTS

We would like to thank Daniel J. Rosenberg for guidance in analyzing SEC-SAXS-MALS data, the Cal Cryo facility for support with cryo-EM data collection, and Kurt Stine for computational support for cryo-EM image analysis. P.M.S. and A.K.L. acknowledge support from a Packard Fellowship from the David Lucile Packard Foundation. A.K.L., L.J.T.-K., and P.M.S. acknowledge support from the DOE Joint BioEnergy Institute (http://www.jbei.org), supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. E.N. is a Howard Hughes Medical Institute Investigator. Support for generation and analysis of metagenome-derived sequences was provided by the Watershed Function Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award number DE-AC02-05CH11231. SEC-SAXS-MALS experiments were conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483.

AUTHOR CONTRIBUTIONS

A.K.L. and P.M.S. designed all experiments. L.X.C., J.W.-R., A.L., and J.F.B. conducted all metagenomic experiments and analyses. A.K.L. conducted all protein purifications and phylogenetic analyses. M.H. conducted all SEC-SAXS-MALS experiments. B.K. and E.N. conducted all cryo-EM experiments and data analysis. L.J.T.-K. conducted all kinetic parameter measurements. D.G. created code for the analysis of kinetic parameter measurement data. All authors contributed to the writing and review of this manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 6, 2023
Revised: September 26, 2023
Accepted: October 25, 2023
Published: November 17, 2023

REFERENCES

1. Erb, T.J., and Zarzycki, J. (2018). A short history of rubisco: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. Curr. Opin. Biotechnol. *49*, 100–107. https://doi.org/10.1016/j.copbio.2017.07.017.
2. Flamholz, A., and Shih, P.M. (2020). Cell biology of photosynthesis over geologic time. Curr. Biol. *30*, R490–R494. https://doi.org/10.1016/j.cub.2020.01.076.
3. Cleland, W.W., Andrews, T.J., Gutteridge, S., Hartman, F.C., and Lorimer, G.H. (1998). Mechanism of rubisco: the carbamate as general base. Chem. Rev. *98*, 549–562. https://doi.org/10.1021/cr970010r.
4. Nisbet, E.G., Grassineau, N.V., Howe, C.J., Abell, P.I., Regelous, M., and Nisbet, R.E.R. (2007). The age of rubisco: the evolution of oxygenic photosynthesis. Geobiology *5*, 311–335. https://doi.org/10.1111/j.1472-4669.2007.00127.x.
5. Bar-On, Y.M., and Milo, R. (2019). The global mass and average rate of rubisco. Proc. Natl. Acad. Sci. USA *116*, 4738–4743. https://doi.org/10.1073/pnas.1816654116.
6. Tabita, F.R., Satagopan, S., Hanson, T.E., Kreel, N.E., and Scott, S.S. (2008). 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. https://doi.org/10.1093/jxb/erm361.
7. Prywes, N., Phillips, N.R., Tuck, O.T., Valentin-Alvarado, L.E., and Savage, D.F. (2023). Rubisco function, evolution, and engineering. Annu. Rev. Biochem. *92*, 385–410. https://doi.org/10.1146/annurev-biochem-040320-101244.
8. Knight, S., Andersson, I., and Brändén, C.I. (1990). Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2·4 Å resolution: subunit interactions and active site. J. Mol. Biol. *215*, 113–160. https://doi.org/10.1016/S0022-2836(05)80100-7.
9. Liu, A.K., Pereira, J.H., Kehl, A.J., Rosenberg, D.J., Orr, D.J., Chu, S.K.S., Banda, D.M., Hammel, M., Adams, P.D., Siegel, J.B., et al. (2022). Structural plasticity enables evolution and innovation of RuBisCO assemblies. Sci. Adv. *8*, eadc9440. https://doi.org/10.1126/sciadv.adc9440.
10. Huang, Q., and Szebenyi, D.M.E. (2023). Crystal structure of a type III rubisco in complex with its product 3-phosphoglycerate. Proteins *91*, 330–337. https://doi.org/10.1002/prot.26431.
11. Gunn, L.H., Valegård, K., and Andersson, I. (2017). 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. https://doi.org/10.1074/jbc.M116.767145.

12. Andrews, T.J. (1988). Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. J. Biol. Chem. *263*, 12213–12219. https://doi.org/10.1016/S0021-9258(18)37741-X.
13. Schulz, L., Guo, Z., Zarzycki, J., Steinchen, W., Schuller, J.M., Heimerl, T., Prinz, S., Mueller-Cajar, O., Erb, T.J., and Hochberg, G.K.A. (2022). Evolution of increased complexity and specificity at the dawn of form I rubiscos. Science *378*, 155–160. https://doi.org/10.1126/science.abq1416.
14. Banda, D.M., Pereira, J.H., Liu, A.K., Orr, D.J., Hammel, M., He, C., Parry, M.A.J., Carmo-Silva, E., Adams, P.D., Banfield, J.F., et al. (2020). Novel bacterial clade reveals origin of form I RuBisCO. Nat. Plants *6*, 1158–1166. https://doi.org/10.1038/s41477-020-00762-4.
15. West-Roberts, J.A., Matheus-Carnevali, P.B., Schoelmerich, M.C., Al-Shayeb, B., Thomas, A.D., Sharrar, A., He, C., Chen, L.-X., Lavy, A., Keren, R., et al. (2021). The Chloroflexi supergroup is metabolically diverse and representatives have novel genes for non-photosynthesis based CO2 fixation. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.23.457424.
16. Lavy, A., Carnevali, P.B.M., Keren, R., Bill, M., Wan, J., Tokunaga, T.K., Williams, K.H., Hubbard, S.S., and Banfield, J.F. (2020). Taxonomically and metabolically distinct microbial communities with depth and across a hillslope to riparian zone transect. Preprint at bioRxiv. https://doi.org/10.1101/768572.
17. Eloe-Fadrosh, E.A., Paez-Espino, D., Jarett, J., Dunfield, P.F., Hedlund, B.P., Dekas, A.E., Grasby, S.E., Brady, A.L., Dong, H., Briggs, B.R., et al. (2016). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. Nat. Commun. *7*, 10476. https://doi.org/10.1038/ncomms10476.
18. Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. Photosynth. Res. *60*, 1–28. https://doi.org/10.1023/A:1006211417981.
19. Whitney, S.M., and Andrews, T.J. (2001). 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. https://doi.org/10.1105/tpc.13.1.193.
20. Badger, M.R., and Bek, E.J. (2008). Multiple rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. J. Exp. Bot. *59*, 1525–1541. https://doi.org/10.1093/jxb/erm297.
21. Mao, Y., Catherall, E., Díaz-Ramos, A., Greiff, G.R.L., Azinas, S., Gunn, L., and McCormick, A.J. (2023). The small subunit of rubisco and its potential as an engineering target. J. Exp. Bot. *74*, 543–561. https://doi.org/10.1093/jxb/erac309.
22. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.
23. Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., Tsutakawa, S.E., Jenney, F.E., Classen, S., Frankel, K.A., Hopkins, R.C., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat. Methods *6*, 606–612. https://doi.org/10.1038/nmeth.1353.
24. Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. *40*, 191–285. https://doi.org/10.1017/S0033583507004635.
25. Classen, S., Hura, G.L., Holton, J.M., Rambo, R.P., Rodic, I., McGuire, P.J., Dyer, K., Hammel, M., Meigs, G., Frankel, K.A., et al. (2013). Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. J. Appl. Crystallogr. *46*, 1–13. https://doi.org/10.1107/S0021889812048698.
26. Duff, A.P., Andrews, T.J., and Curmi, P.M.G. (2000). The transition between the open and closed states of is triggered by the inter-phosphate distance of the bound bisphosphate. J. Mol. Biol. *298*, 903–916. https://doi.org/10.1006/jmbi.2000.3724.
27. Alonso, H., Blayney, M.J., Beck, J.L., and Whitney, S.M. (2009). Substrate-induced assembly of Methanococcoides burtonii d-ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. J. Biol. Chem. *284*, 33876–33882. https://doi.org/10.1074/jbc.M109.050989.
28. Schneider, G., Lindqvist, Y., and Brändén, C.I. (1992). RuBisCO: structure and mechanism. Annu. Rev. Biophys. Biomol. Struct. *21*, 119–143. https://doi.org/10.1146/annurev.bb.21.060192.001003.
29. Spreitzer, R.J. (2003). Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. Arch. Biochem. Biophys. *414*, 141–149. https://doi.org/10.1016/S0003-9861(03)00171-1.
30. Jaffe, A.L., Castelle, C.J., Dupont, C.L., and Banfield, J.F. (2019). Lateral gene transfer shapes the distribution of RuBisCO among candidate phyla radiation bacteria and DPANN Archaea. Mol. Biol. Evol. *36*, 435–446. https://doi.org/10.1093/molbev/msy234.
31. Varaljay, V.A., Satagopan, S., North, J.A., Witte, B., Dourado, M.N., Anantharaman, K., Arbing, M.A., Hoeft McCann, S.H., Oremland, R.S., Banfield, J.F., et al. (2016). Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. Environ. Microbiol. *18*, 1187–1199. https://doi.org/10.1111/1462-2920.13138.
32. Lin, M.T., Salihovic, H., Clark, F.K., and Hanson, M.R. (2022). Improving the efficiency of RuBisCO by resurrecting its ancestors in the family Solanaceae. Sci. Adv. *8*, eabm6871. https://doi.org/10.1126/sciadv.abm6871.
33. Shih, P.M., Occhialini, A., Cameron, J.C., Andralojc, P.J., Parry, M.A.J., and Kerfeld, C.A. (2016). Biochemical characterization of predicted Precambrian RuBisCO. Nat. Commun. *7*, 10382. https://doi.org/10.1038/ncomms10382.
34. Hochberg, G.K.A., Liu, Y., Marklund, E.G., Metzger, B.P.H., Laganowsky, A., and Thornton, J.W. (2020). A hydrophobic ratchet entrenches molecular complexes. Nature *588*, 503–508. https://doi.org/10.1038/s41586-020-3021-2.
35. Gray, M.W., Lukeš, J., Archibald, J.M., Keeling, P.J., and Doolittle, W.F. (2010). Cell biology. Irremediable complexity? Science *330*, 920–921. https://doi.org/10.1126/science.1198594.
36. Frey, S., and Görlich, D. (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. J. Chromatogr. A *1337*, 95–105. https://doi.org/10.1016/j.chroma.2014.02.029.
37. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25. https://doi.org/10.1002/pro.3235.
38. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.
39. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. *20*, 1160–1166. https://doi.org/10.1093/bib/bbx108.
40. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165. https://doi.org/10.1093/bioinformatics/btr088.
41. Lemoine, F., Domelevo Entfellner, J.B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., and Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature *556*, 452–456. https://doi.org/10.1038/s41586-018-0043-0.
42. Kubien, D.S., Brown, C.M., and Kane, H.J. (2011). Quantifying the amount and activity of rubisco in leaves. In Photosynthesis Research Protocols Methods in Molecular Biology, R. Carpentier, ed. (Humana Press), pp. 349–362. https://doi.org/10.1007/978-1-60761-925-3_27.

43. Lilley, R.McC., and Walker, D.A. (1974). The reduction of 3-phosphoglycerate by reconstituted chloroplasts and by chloroplast extracts. Biochim. Biophys. Acta *368*, 269–278. https://doi.org/10.1016/0005-2728(74)90174-1.

44. Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y.M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D.G., Antonovsky, N., et al. (2020). Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. EMBO J. *39*, e104081, https://doi.org/10.15252/embj.2019104081.

45. Pierce, J., Tolbert, N.E., and Barker, R. (1980). Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogues. Biochemistry *19*, 934–942. https://doi.org/10.1021/bi00546a018.

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

47. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51. https://doi.org/10.1016/j.jsb.2005.07.007.

48. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.

49. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

50. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

51. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.

52. Rosenberg, D.J., Hura, G.L., and Hammel, M. (2022). Size exclusion chromatography coupled small angle X-ray scattering with tandem multiangle light scattering at the SIBYLS beamline. Chapter six. In Methods in Enzymology Small Angle Scattering Part A: Methods for Structural Investigation, J.A. Tainer, ed. (Academic Press), pp. 191–219. https://doi.org/10.1016/bs.mie.2022.08.031.

53. Hopkins, J.B., Gillilan, R.E., and Skou, S. (2017). BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J. Appl. Crystallogr. *50*, 1545–1553. https://doi.org/10.1107/S1600576717011438.

54. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. *38*, W540–W544. https://doi.org/10.1093/nar/gkq461.

55. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys. J. *105*, 962–974. https://doi.org/10.1016/j.bpj.2013.07.020.

56. Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320–W324. https://doi.org/10.1093/nar/gku316.

57. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

58. Meng, E.C., Pettersen, E.F., Couch, G.S., Huang, C.C., and Ferrin, T.E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics *7*, 339. https://doi.org/10.1186/1471-2105-7-339.

59. Crameri, F., Shephard, G.E., and Heron, P.J. (2020). The misuse of colour in science communication. Nat. Commun. *11*, 5444. https://doi.org/10.1038/s41467-020-19160-7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli BL21 DE3 Star	MacroLab, Berkeley, CA	N/A
Chemicals, peptides, and recombinant proteins		
bdSENP1 SUMO Protease	Frey and Görlich,[36] Banda et al.[14]	N/A
2-carboxyarabinitol 1,5-bisphosphate	Banda et al.[14]	N/A
Deposited data		
Firmicutes sp. cryo-EM structure	N/A	PDB ID 8U66
Firmicutes sp. and Limnocylindra sp. SEC-SAXS-MALS data	https://figshare.com/projects/Form_I_and_I_Rubisco/178857	N/A
Rubisco phylogeny data	https://figshare.com/projects/Form_I_and_I_Rubisco/178857	N/A
Recombinant DNA		
14xHis-bdSUMO-tagged form I'' and form Iα rbcL in pET-28-based plasmids	Twist Bioscience	This paper
Software and algorithms		
Python script for analyzing plate reader kinetic data	https://doi.org/10.5281/zenodo.7757660	This paper
UCSF ChimeraX	Goddard et al.[37]	https://www.cgl.ucsf.edu/chimerax/
Interactive Tree of Life	Letunic and Bork[38]	https://itol.embl.de/
MAFFT	Katoh et al.[39]	https://mafft.cbrc.jp/alignment/server/index.html
ProtTest 3.0	Darriba et al.[40]	https://github.com/ddarriba/prottest3
Other		
Emulsiflex C3	Avestin Inc.	N/A
ÄKTA pure	Cytiva Life Sciences	N/A
SIBYLS SEC-SAXS Beamline	Advanced Light Source, Berkeley, CA	https://bl1231.als.lbl.gov/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Patrick Shih (pmshih@berkeley.edu).

Materials availability

The Firmicutes sp. and Limnocylindria sp. 14xHis-bdSUMO-*rbcL* plasmids used in this study were synthesized by Twist Bioscience (South San Francisco, CA) and are available upon request.

Data and code availability

- All phylogenetic analysis and raw SAXS data are available at https://figshare.com/projects/Form_I_and_I_Rubisco/178857.
- The cryo-EM structure of the Firmicutes sp. rubisco has been deposited to the PDB (PDB ID: 8U66).
- The Python script for analyzing plate reader rubisco kinetics is available at https://doi.org/10.5281/zenodo.7757660.
- Any additional information for re-analysis of the data is available from the lead contact upon request.

EXPERIMENTAL MODEL DETAILS

All *E. coli* cultures for protein purification used *E. coli* BL21 DE3 Star cells (MacroLab, Berkeley, CA). Culture conditions used in this study are described in the method details section.

METHOD DETAILS

Phylogenetic analysis

Rubisco large subunit protein sequences were aligned with MAFFT using default parameters (https://mafft.cbrc.jp/alignment/server/).[39] The following number of sequences were used to represent each form within this phylogeny: form I, 80; form I', 32; form I'', 3; form Iα, 7; form II, 27; form II/III, 28; form IIIB, 19; form IIIA, 18; form IV; 21. The evolutionary model most appropriate for constructing a phylogenetic tree was determined using Prottest 3.0.[40] A maximum likelihood phylogenetic tree was constructed using RAxML-HPC BlackBox (v. 8.2.12) as implemented on cipres.org (default parameters with WAG model). Felsenstein bootstrap values were calculated using BOOSTER.[41]

Protein modeling

The Limnocylindria sp. dimer was modeled using the ColabFold webtool on Google Colaboratory.[22]

Plasmids

The form I'' and form Iα gene sequences were synthesized by Twist Biosciences and cloned into a modified pET28 vector containing an N-terminal His_{14}-bdSUMO tag.[36] pSF1389 and pBAD*ES*/*EL* were gifts.

Rubisco expression and purification

Form I'' and form Iα rubisco purification was performed as previously described.[14] Plasmids containing His_{14}-bdSUMO-tagged RbcL were cotransformed with pBAD*ES*/*EL* into BL21 DE3 Star *E. coli* competent cells (MacroLab, Berkeley, CA). Cells were grown in Luria-Bertani media at 37°C to an optical density at 600 nm of 0.6-0.8, at which point GroEL/ES overexpression was induced by the addition of 0.2% w/v arabinose for an additional two hours at 30°C. The cells were then resuspended in fresh media with arabinose, and rubisco expression was induced by the addition of 1 mM IPTG at 16°C for 16 hours. Pelleted cells were resuspended in a lysis buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 2 mM $MgCl_2$ hexahydrate, 5% glycerol), and were subject to a freeze-thaw cycle at -80°C. Thawed cells were then lysed using an Emulsiflex C3 (AVESTIN Inc, Ottawa, Canada), and cell lysate was clarified by centrifugation at 15,000 g. The soluble fraction was 0.44μm-filtered before application to HisPur Ni-NTA resin (Thermo Fisher) for batch binding. Columns were washed twice, first with a 25 mM imidazole wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole, and 10% glycerol), followed by a 50 mM imidazole wash buffer (20 mM sodium phosphate, 300 mM NaCl, 50 mM imidazole, and 10% glycerol). The column was then resuspended in SUMOlase buffer (20 mM HEPES pH 8.0, 100 mM NaCl, 1 mM dithiothreitol, 15 mM imidazole, and 20 mM MgCl2), and purified bdSENP1 was added and rocked overnight to facilitate tag cleavage.[9,14,36]

Spectroscopic kinetic parameter measurements

Enzyme kinetics were approximated using an NADH-coupled assay,[42,43] measured at 340 nm on a Spark TeCool (Tecan) plate reader using 96 well flat-bottom transparent plates (Corning, Costar). A complete list of assay reagents is included in Table S2. The assays were conducted in 100 mM HEPES or EPPS pH 8, at 25 ºC with orbital shaking at 1440 rpm. The enzymes were added to a mix of cofactors and NADH assay reagents and activated under 0.5% O_2 and 4% CO_2 for 20 minutes prior to initiating the reaction with ribulose-1,5-bisphosphate (RuBP). Active site concentration was estimated using the known rubisco inhibitor, CABP, as previously described.[44] Rubisco rates of activity were determined under a series of CABP concentrations (n=2). The resulting rubisco rates were plotted against CABP concentration, with the x-axis intercept proportional to rubsico active site concentration. V_{max} (n=3) was divided by this x-intercept ([E]) to generate the k_{cat}. The rubisco catalysis rates and CABP inhibition slopes were calculated using Python: 10.5281/zenodo.7757660. RuBP and CABP were synthesized and purified as previously described.[45,46] The rate of $A_{340\ nm}$ NADH oxidation was converted to molar concentrations of NADH using an experimentally determined conversion factor which accounted for both the extinction coefficient for NADH absorbance and the non-standard pathlength on the Tecan. This was achieved by plotting the $A_{340\ nm}$ absorption of serially diluted NADH mixes on the Tecan versus the NADH concentrations (Molar, Beer-Lambert Law, $\varepsilon_{340,\ NADH} = 6220\ M^{-1}cm^{-1}$) as determined spectroscopically on a machine of a known pathlength. The slope of the linear fit converted Tecan absorption values to NADH concentrations. The rate of NADH oxidation was converted to the rate of rubisco product formation, 3-phosphoglycerate (3-PGA), by dividing by two as a single rubisco cycle produces two molecules of 3-PGA and both 3-PGA molecules consume one molecule of NADH each in the forward reactions as part of the coupled assay.

Cryo-electron microscopy sample preparation

Ni-NTA–purified form I'' rubisco was further purified by anion exchange chromatography on a MonoQ 10/100 GL column and eluted by a linear NaCl gradient from 5 mM to 1 M. Fractions were analyzed by SDS-PAGE, followed by concentration and size exclusion chromatography on a Superose 6 Increase 10/300 GL column in a final buffer containing 20 mM HEPES pH 8.0, 100 mM NaCl, 25 mM $MgCl_2$, and 5 mM $NaHCO_3$. Samples were activated as previously described before incubation with a tenfold molar excess of previously synthesized 2-carboxyarabinitol 1,5-bisphosphate (CABP)[27]

Cryo-EM specimens were prepared on C-flat-1.2/1.3 400 mesh copper grids (Protochips) that were glow-discharged using a Tergeo-EM plasma cleaner (PIE Scientific). The cryo grids were produced using a FEI Mark IV Vitrobot. The chamber of the Vitrobot was kept at 4 °C and 100% relative humidity. 4 μl of sample was applied to the glow-discharged grid and blotted with filter paper for 5 seconds with the equipment-specific blotting force set at 4 after 30 seconds of incubation. After blotting, the grid was vitrified by plunging into liquid ethane.

Single-particle cryo-electron microscopy data collection, image processing, and model building

4266 movies were collected using a Titan Krios G3i microscope equipped with a Gatan Quantum energy filter (slit width 20 eV) and a K3 summit camera, using a defocus range of −0.5 to −2.0 μm. Automated image acquisition was carried out using SerialEM[47] with a nominal magnification of 81,000x, corresponding to a pixel size of 1.05 Å (0.525 Å super resolution). Image stacks of 50 movie frames were collected with a dose rate of 1.0 e-/Å^2/frame.

The 4,266 movies were imported into CryoSPARC, binned 2x and motion corrected using Patch Motion Correction. CTF was then estimated using the CTF Estimation job. 5,466,481 particles were picked using Blob Picker, with a particle diameter range of 90-150 A. After particle inspection, 3,903,212 particles were extracted using a 300x300 pixels box, and then Fourier cropped to a boxed size of 168 pixels. After 2D classification into 50 2D classes, 22 (containing 2,703,596 particles) were selected to generate an initial model using the Ab Initio Refinement job. Homogenous Refinement with a coloured noise model resulted in a 2.86 Å resolution volume. Subsequent Heterogenous Refinement produced 5 classes, 2 of which were selected (with a total of 1,557,723 particles) for another Homogenous refinement with a coloured noise model, yielding a 2.73 Å volume. After Global CTF Refinement and Local CTF Refinement, Non-Uniform Refinement yielded a volume with a resolution of 2.47 Å. The apparent D4 symmetry was then applied, and the volume used as a reference for another Non-Uniform Refinement with D4 symmetry, yielding a volume at 2.21 Å. This volume was sharpened using a supplied B Factor of 95.2. The data collection and image processing details are shown in Figure S3.

An initial model was generated using SWISS-MODEL with the Firmicutes sp. form I'' sequence and the previously reported structure of the form I' *Ca.* P. breve rubisco (PDB: 6URA). The resulting homology model was placed into the sharpened cryo-EM volume using UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM10331.[48] A rigid body fit was performed in Coot.[49] Several iterations of Real-Space Refinement in Phenix[50] followed by manual inspection in COOT were performed. The final model was validated using MolProbity.[51] Refinement statistics are available in Table S3.

Size exclusion chromatography coupled small-angle X-ray scattering with in-line multiangle light scattering experiments

Rubisco was purified as described above and concentrated to 2-5 mg/mL. Concentrated rubisco was then activated with an excess of $NaHCO_3$ before sample analysis. SEC-SAXS-MALS data were collected at the ALS beamline 12.3.1 at Lawrence Berkeley National Lab (Berkeley, CA, USA).[52] The X-ray wavelength was set at λ=1.24 Å and the sample-to-detector distance was 2075 mm resulting in scattering vectors (q) ranging from 0.01 Å^{-1} to 0.46 Å^{-1}. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. Data was collected using a Pilatus 3X 2M Detector (Dectris, Baden, Switzerland). Normalization and integration of each image was processed as previously described.[23] SEC was performed using a 1290 Infinity HPLC system (Agilent, Santa Clara, CA) coupled to a Shodex KW-803 column (Showa Denko, Tokyo, Japan). The column was equilibrated with a running buffer (20 mM HEPES-OH (pH 8.0), 300 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$) at a flow rate of 0.65 mL/min. 90-100 μL of sample was separated by SEC and the elution was monitored at 280 and 260 nm by an in-line Variable Wavelength Detector (VWD) (Agilent, Santa Clara, CA). MALS experiments were performed using an in-line 18-angle DAWN HELEOS II light scattering detector connected in tandem to an Optilab differential Refractive Index (dRI) detector (Wyatt Technology, Goleta, CA). System normalization and calibration was performed with bovine serum albumin using a 50 μL sample at 7 mg/mL in the same running buffer. The light scattering experiments were used to determine Molecular Weight (MW) across the principal peaks in the SEC analysis. UV, MALS, and dRI data was analyzed using Wyatt Astra 7 software to monitor the homogeneity of the sample across the elution peak complementary to the SEC-SAXS signal validation. A purpose-built SAXS flow cell was connected in-line immediately following the complementary spectroscopic techniques and two second X-ray exposures were collected continuously over the 25 min elution. The SAXS frames recorded prior to the protein elution peak were used to subtract 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 $qRg<1.5$. The elution peak was mapped by comparing integral ratios to background and Rg relative to the recorded frame using the program RAW.[53] Uniform Rg values across an elution peak represent a homogenous assembly and were merged to reduce noise in the curve. Final merged SAXS profiles (Figures 2B and 3B), were then compared to theoretical scattering curves generated from the ColabFold model of the Limnocylindria sp. dimer (Figure 2B) and the cryo-EM structure of the Firmicutes sp. octamer (Figure 3B) using FoXS.[54,55]

QUANTIFICATION AND STATISTICAL ANALYSIS

Multiple sequence alignments were generated using MAFFT and visualized with ESPript 3.0.[39,56] Phylogenetic trees were visualized using Interactive Tree of Life v5[38] and Felsenstein bootstrap values were calculated using BOOSTER.[41] UCSF ChimeraX was used for visualization of protein structures, structural alignment using the MatchMaker function, and preparation of manuscript figures.[37,57,58] The scientific color map ''roma'' was used in preparation of figures (http://doi.org/10.5281/zenodo.1243862)[59] and Inkscape software. All kinetic parameter checks were performed in technical triplicate (n=3) and active site quantification was performed in technical duplicate (n=2). The Python script for analyzing plate reader rubisco kinetics is available at 10.5281/zenodo.7757660. All SEC-SAXS-MALS experiments were performed once (n=1). SAXS data was processed using Wyatt Astra 7, RAW,[53] FoXS[54,55] and OriginLab software. All cryo-EM statistics and related information are available in Figure S3 and Table S3. SerialEM, CryoSPARC, SWISS-MODEL, UCSF Chimera, Coot,[49] and Phenix software were used in cryo-EM data collection and processing.