Molecular Biology and Evolution, 2025, **42**, 1–13
https://doi.org/10.1093/molbev/msae268
Advance access publication 30 December 2024
Discoveries

OXFORD MBE

Evolutionary Dynamics of RuBisCO: Emergence of the Small Subunit and its Impact Through Time

Kaustubh Amritkar [1,2], **Bruno Cuevas-Zuviría** [1,3], **Betül Kaçar** [1,*]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
[2]Biophysics Graduate Degree Program, University of Wisconsin-Madison, Madison, WI, USA
[3]Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Madrid, Spain

***Corresponding author:** E-mail: bkacar@wisc.edu.
Associate editor: Julian Echave

Abstract

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is an ancient protein critical for CO_2-fixation and global biogeochemistry. Form-I RuBisCO complexes uniquely harbor small subunits that form a hexadecameric complex together with their large subunits. The small subunit protein is thought to have significantly contributed to RuBisCO's response to the atmospheric rise of O_2 ~2.5 billion years ago, marking a pivotal point in the enzyme's evolutionary history. Here, we performed a comprehensive evolutionary analysis of extant and ancestral RuBisCO sequences and structures to explore the impact of the small subunit's earliest integration on the molecular dynamics of the overall complex. Our simulations suggest that the small subunit restricted the conformational flexibility of the large subunit early in its history, impacting the evolutionary trajectory of the Form-I RuBisCO complex. Molecular dynamics investigations of CO_2 and O_2 gas distribution around predicted ancient RuBisCO complexes suggest that a proposed "CO_2-reservoir" role for the small subunit is not conserved throughout the enzyme's evolutionary history. The evolutionary and biophysical response of RuBisCO to changing atmospheric conditions on ancient Earth showcase multi-level and trackable responses of enzymes to environmental shifts over long timescales.

Keywords: RuBisCO, small subunit, ancestral sequence reconstruction, structural dynamics.

Introduction

Life has been evolving on this planet for almost four billion years. Evolution allows for exploration of the vast sequence space comprised by polymers of the twenty standard amino acids—a space that exceeds the number of molecules in the universe (Wagner and Rosen 2014). From within this space, biology has discovered unique solutions to the challenges of growing and persisting in the ever-shifting diversity of Earth's environments. Undoubtedly, biogeochemically critical microbial enzymes are central to this dynamic, long-term interaction between life and the environment (Nealson and Conrad 1999; Falkowski et al. 2008; Kaçar 2024). One such key enzyme that evolved early in the history of life and persisted through planetary extremes is RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase).

RuBisCO is a globally critical, ancient enzyme with an intriguing history. It facilitates the rate-limiting step of the Calvin-Benson-Bassham cycle for carbon fixation, catalyzing the addition of atmospheric carbon dioxide (CO_2) with ribulose 1,5-bisphosphate (RuBP) (Andersson 2008). In the presence of oxygen (O_2), RuBisCO also catalyzes a competing oxygenation reaction in which RuBP combines with O_2, producing an autoinhibitory metabolite detrimental to the overall metabolic efficiency of carbon fixation (Fernie and Bauwe 2020). While the exact age of RuBisCO is not clearly known, paleobiological inferences suggest that it emerged prior to the rise of atmospheric O_2 ~2.5 billion years ago, a period known as the Great Oxidation Event (GOE) (Kacar et al. 2017; Ward and Shih 2019; Garcia et al. 2021; Kędzior et al. 2022). Accordingly, despite being infamously sensitive to O_2, RuBisCO was maintained by organisms through this drastic atmospheric upheaval that impacted the ecosystem and its constituent biomolecules (Ashida et al. 2005; Raymond and Segrè 2006; Wang et al. 2011; Young et al. 2012; Caetano-Anollés 2017; Kacar et al. 2017; Erb and Zarzycki 2018; Garcia et al. 2021).

Extant RuBisCO forms (I to IV) exhibit different dimeric and poly-dimeric assemblies. All forms have large subunits (RbcL), but Form-I RuBisCOs uniquely have an additional small subunit (RbcS) (Fig. 1a). Form-I RuBisCOs are also the most abundant (Tabita et al. 2008) and generally exhibit higher specificities for CO_2 (Flamholz et al. 2019). These attributes make RbcS a promising candidate to study how small accessory subunits can regulate the evolution of RuBisCO, though their precise evolutionary role remains unclear. RbcS is thought to be crucial for the assembly of the RbcL octameric complex made of RbcL homodimers (Liu et al. 2010; Esquivel et al. 2013; Joshi et al. 2015) and significantly impacts RuBisCO's catalytic parameters (Gatenby 1988; Spreitzer 2003; Genkov et al. 2010; Esquivel et al. 2013; Matsumura et al. 2020; Mao et al. 2023). For example, incorporation of ancestral RbcS in RuBisCO was shown to increase the enzyme's specificity for CO_2, indicating that emergence of RbcS played a critical role in the ancestor of Form-I RuBisCO (Schulz et al. 2022). Previous computational work

Received: July 17, 2024. **Revised:** November 25, 2024. **Accepted:** December 24, 2024

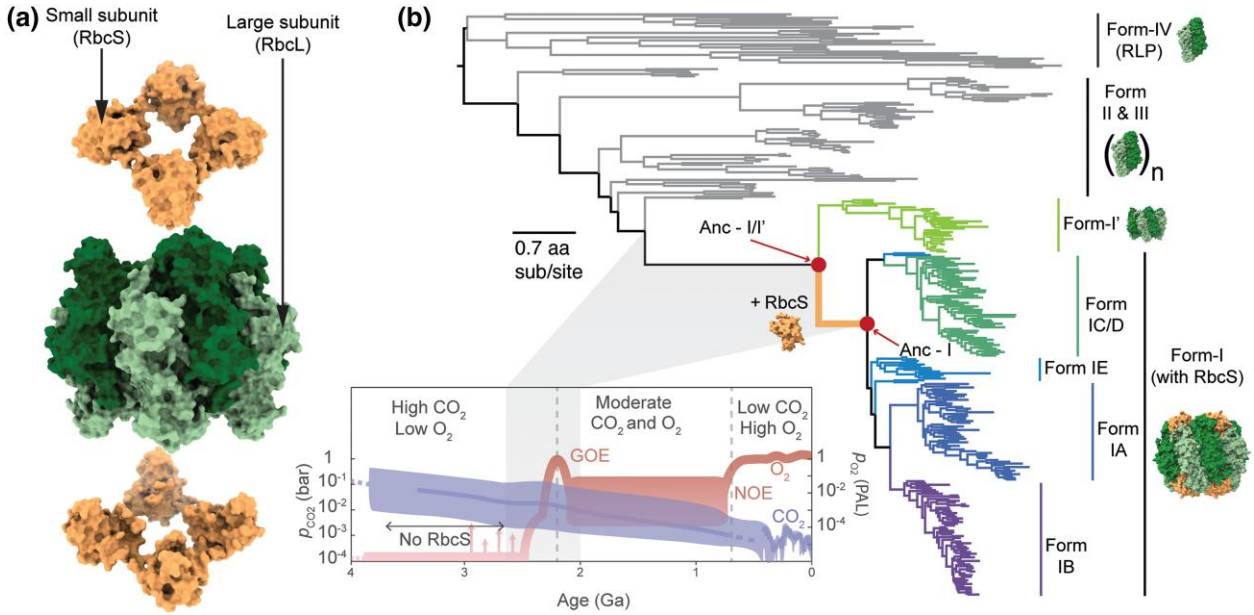

Fig. 1. The evolutionary history and diversification of RuBisCO. a) Architecture of the L_8S_8 structure of the Form-I RuBisCO complex, with RbcL at the centre (shown in green) and RbcS positioned at the top and bottom (in orange) (PDB:1BWV). b) RuBisCO phylogenetic tree highlighting the emergence of RbcS coinciding with the GOE (Kacar et al. 2017; Banda et al. 2020). Schematic on the right shows variation in the multimeric structure of the RuBisCO complex across different forms. Form-II and III RuBisCOs exhibit multiple homooligomeric states of the RbcL–dimer, such as RbcL–dimer, tetramer, or hexamer (Liu et al. 2022), represented here as $(RbcL–RbcL)_n$, where n is the number of RbcL dimers. Bottom inset schematic shows the change in atmospheric CO_2 and O_2 concentration through Earth's history (Rucker and Kaçar 2024).

also suggested that RbcS can act as a "CO_2-reservoir" to concentrate CO_2 molecules within the enzyme (Van Lun et al. 2014). However, the extent to which the small subunit has influenced the structural dynamics and evolutionary trajectory of the Form-I RuBisCO complex remains less understood.

Here, we used phylogenetic reconstructions, structural predictions as well as molecular dynamics (MD) simulations to track the evolution of the small subunit in the context of the RuBisCO complex over geologic time. Our exploration focused on the hypothesis that the molecular and structural innovations likely conferred a selective advantage to Form-I RuBisCO following the planetary rise of O_2 levels. We specifically focused on three areas: (i) the sequence and structural evolution of subunits within ancient RuBisCO complexes following integration of ancient RbcS, (ii) the impact of RbcS presence or absence on the structural dynamics of ancestral and extant RuBisCO complexes, and (iii) migration of CO_2 and O_2 gases in ancestral RuBisCO to assess a proposed CO_2-reservoir role for ancient small subunits (Van Lun et al. 2014).

Results and Discussion

Resurrection of Ancestral RuBisCOs

We built a maximum-likelihood phylogenetic tree from a concatenated alignment of RbcL and RbcS amino acid sequences representing RuBisCO Forms I to IV, including the recently described Form-I′ (Banda et al. 2020) (Fig. 1b). The phylogeny contains 194 sets of RbcL–RbcS homologs from Form-I and 135 RbcL homologs from all other forms (including Form-I′, II, II/III, III and IV). The tree contains sequences representative of known RuBisCO diversity and is rooted by Form-IV RuBisCO-like proteins in accordance with previous studies (Tabita et al. 2007; Kacar et al. 2017; Poudel et al. 2020). Form-I is categorized into four major subgroups: the "green-like" Form-IA and IB, prevalent in proteobacteria, cyanobacteria, green algae and plants, and the "red-like" Form-IC and ID, prevalent in proteobacteria and non-green algae (Spreitzer 2003; Tabita et al. 2008). Our phylogeny also resolves a Form-I subclade, Form-IE, that diverges before the last common ancestor of Form-IA and Form-IB (West-Roberts et al. 2021). Most sequences from the Form-IE clade are from metagenomic studies and belong to unclassified members of the *Chloroflexota* bacterial phylum. Form-I′ RuBisCOs cluster within a monophyletic clade sister to all other Form-I sequences. Form-I′ RuBisCOs are notable because their RbcL subunits are similar to those of Form-I in their multimeric assembly, but they lack the RbcS (Banda et al. 2020). Like Form-IE, the Form-I′ clade contains sequences from *Chloroflexota*. The topology of our concatenated phylogeny is in agreement with previously reported RbcL-based phylogenetic trees (Kacar et al. 2017; Banda et al. 2020).

We selected ancestral nodes situated along the evolutionary trajectory immediately before and after RbcS incorporation into RuBisCO for ancestral sequence reconstruction (Methods). Specifically, we inferred the common ancestor of both Form-I′ and Form-I RuBisCOs, as well as the ancestors of major Form-I clades as highlighted in Fig. 2a, inferring a total of seven ancestral nodes along the phylogenetic tree.

Sequence and Structural Diversity of Modern and Ancient RbcS Proteins

We analyzed all extant Form-I RbcS sequences from our database representative of known host taxonomic diversity (Fig. 2b, Methods). Extant Form-I RbcS homologs show remarkably low mean pairwise sequence identity (29%) compared with Form-I RbcL sequences (66%) (supplementary fig. S3, Supplementary Material online) as observed by others previously (Bracher et al. 2017; Mao et al. 2023; Bouvier et al. 2024).

We inferred sequences for 328 ancestral nodes within the phylogenetic tree, with 193 of these nodes belonging to the

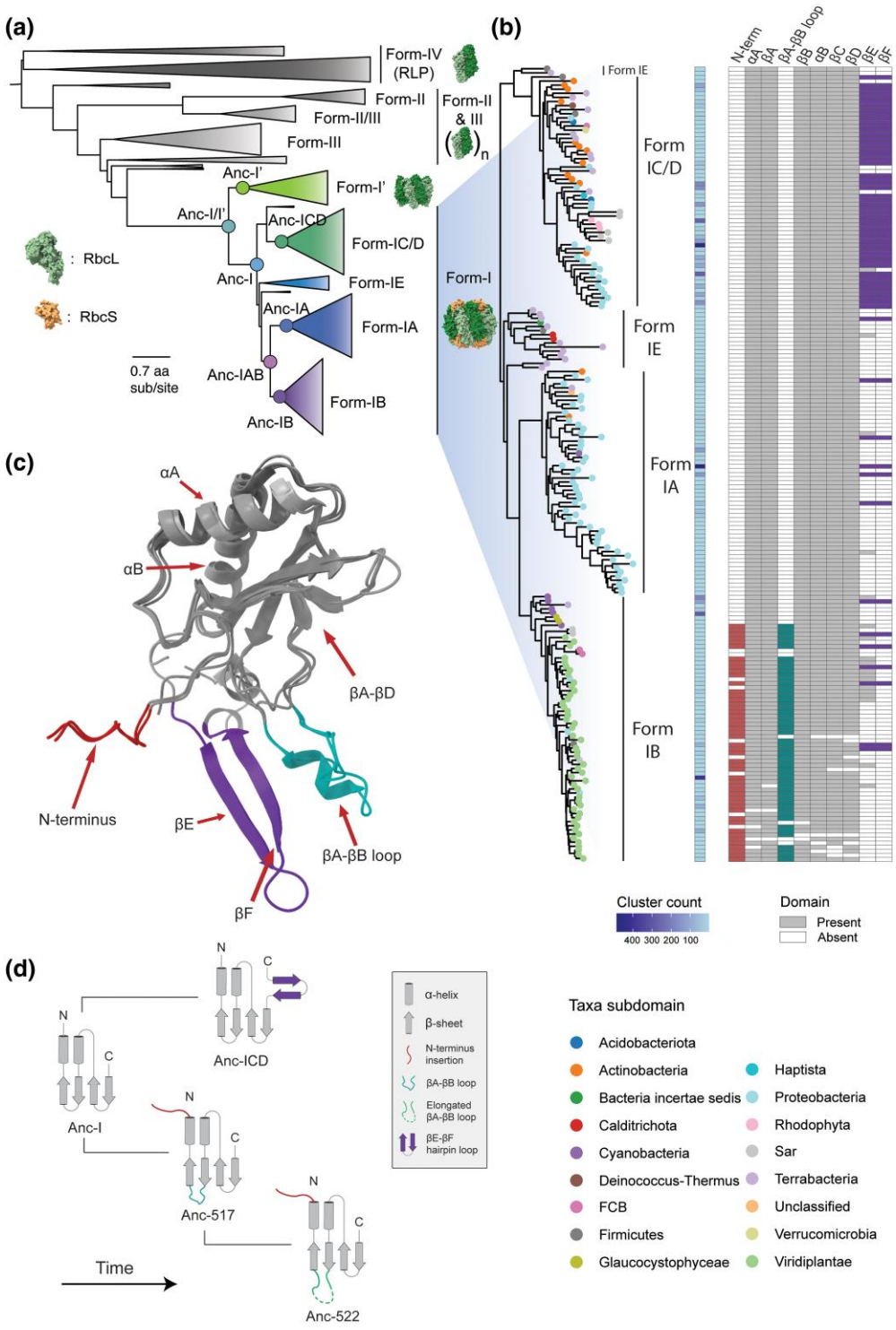

Fig. 2. Phylogenetic analysis of RbcS structural diversity. a) Collapsed maximum-likelihood phylogenetic tree of concatenated RbcL–RbcS sequences. Ancestral nodes used in this study are highlighted. b) Analysis of RbcS structural diversity, mapped to a clustered Form-I phylogeny, where each tip represents a cluster of RbcS sequences with >63% sequence identity. A heatmap for the number of sequences present in the RbcS cluster is represented in blue. Extant nodes are colored by microbial host taxonomic diversity. The presence and absence of different RbcS features in each extant sequence cluster is indicated by gray and white colors, and the color scheme for the presence of uncommon features is according to (c). c) Multiple RbcS structures across Form-I subgroups aligned to highlight the different RbcS structural features. d) Schematic representing the emergence of RbcS structural features across the Form-I RuBisCO phylogeny.

Form-I clade, including its common ancestor, Anc-I. Form-I RbcL ancestors are reconstructed with higher confidence than Form-I RbcS ancestors, with mean posterior probabilities of 0.98 and 0.90, respectively, across all positions (supplementary fig. S4, Supplementary Material online). Anc-I subunit sequences have a mean pairwise identity of 71.5% (RbcL) and 42.2% (RbcS) compared across all extant RuBisCOs in the phylogeny, indicating a higher conservation of RbcL compared with RbcS. The mean posterior probabilities for Anc-I RbcL and RbcS are 0.95 and 0.85, respectively.

We further explored historical, evolutionary variation of RbcS by identifying known, functionally significant sequence and structural features (Spreitzer 2003; Mao et al. 2023) (Fig. 2c) and mapping their presence across extant and reconstructed ancestral RbcS in our phylogenetic tree (Fig. 2b and d). We find that the RbcS structure contains novel features unique to different Form-I specific lineages. While all RuBisCO small subunits share a common core structure consisting of a β-sheet with four antiparallel strands (βA to βD) and two α-helices (αA and αB) (Knight et al. 1990), the Form-IB and Form-IC/D RbcS exhibit additional distinct characteristics.

Within the Form-IB clade, we find that most green-like eukaryotic RbcS homologs contain an additional N-terminal region as well as a βA–βB loop insertion. The N-terminal region is specifically known to be a signal peptide necessary for the entry of RbcS into the chloroplast, prior to assembly with RbcL (Schmidt and Mishkind 1986; Spreitzer 2003). The N-terminal extension feature is first observed in the Form-IB ancestry after the separation of cyanobacteria in the phylogeny (supplementary fig. S5, Supplementary Material online). Absence of the N-terminal extension in Anc-IB, the common ancestor of Form-IB, as well as older ancestors indicates that their ancient hosts did not translate RbcS and RbcL in separate organelles. Rather, expression of subunits in different organelles emerged as a trait following the bifurcation of cyanobacteria and Viridiplantae in Form-IB RuBisCO.

We find that the βA–βB loop insertion is similarly exclusive to green-like eukaryotic RbcS (Fig. 2b). Although the significance of the βA–βB loop in RbcS is not established, it plays an important part in regulating the size of the central solvent channel in RuBisCO complex (Esquivel et al. 2013). Like N-terminal extension, this loop first emerges after the splitting of cyanobacterial and eukaryotic lineages in Form-IB (supplementary fig. S5, Supplementary Material online). As observed with extant RbcS (Spreitzer 2003; Mao et al. 2023), the βA–βB loop insertion length also varies across Form-IB ancestors (10 to 28 aa). The loop has increased by ∼11 residues in all eukaryotic ancestors after the divergence from cyanobacterial ancestors and by ∼18 residues in the ancestors from green algae lineage within Form-IB (supplementary fig. S5, Supplementary Material online).

We also tracked the evolution of the C-terminal βE and βF hairpin loop that is unique to Form-ICD RbcS sequences (Fig. 2b). The common ancestor of Form-IC/D sequences, Anc-ICD, and all subsequent Form-IC/D ancestors have the βE and βF hairpin (Fig. 2d). By contrast, Anc-I RbcS does not have the βE and βF hairpin, suggesting this insertion happened after the divergence of Form-IC/D within the Form-I clade. The C-terminal β-hairpin is functionally significant for mediating the assembly of the oligomeric complex, allowing red-like RuBisCOs to assemble without specialized assembly chaperones (Joshi et al. 2015). The incorporation of the hairpin loop in Form-IC/D common ancestor, Anc-IC/D, suggests that independence from assembly chaperones may have first emerged in a red-like RuBisCO ancestor. Our analysis thus reveals that RbcS has undergone different clade-specific structural variations since its emergence in the Form-I ancestor.

Global Distribution of Residues Evolutionarily Linked to RbcS Incorporation

RbcL from Form-I and Form-I′ RuBisCOs exhibit the same L_8 oligomeric arrangement. Because the presence or absence of the small subunit is the only major structural distinction between the two forms (Banda et al. 2020), we performed comparative analyses to reveal the small subunit's impact on RbcL sequence and structure. We focused on the RbcL extant sequences from the Form-I and Form-I′ clades. We identified residues that are conserved within each clade, but differ between the two clades, which we refer to as "signature positions". These residues are more likely to contribute to the functional distinctions between Form-I and Form-I′, including interactions (or lack thereof) with RbcS. The interface region between the large and small subunit has been shown to be critical for the incorporation of RbcS into the Form-I RuBisCO complex (Knight et al. 1990; Van Lun et al. 2011; Ryan et al. 2019; Schulz et al. 2022). We therefore hypothesized that signature positions would cluster within the RbcL–RbcS interface.

Form-I and Form-I′ sequence analyses revealed eleven, discontinuous signature positions (Fig. 3a). These positions are clustered in three different regions on the RuBisCO structure. Two residues (Val 70 and Asp 74; position index from *Rhodobacter sphaeroides*) are located at the RbcL intradimer interface (highlighted in blue), three (Phe 201, Leu 401 and Gln 402) are buried in the RbcL monomer close to the active site (highlighted in purple) and six (Lys 260, Trp 284, Arg 286, Asn 288, Gly 159 and Glu162) are found at the RbcL dimer–dimer interface (highlighted in red) (Fig. 3a). Out of the six residues located at the RbcL dimer–dimer interface, Gly 159 and Glu 162 are the only two signature positions that are also present at the RbcL–RbcS interface. A similar comparison between the Form-IA/B/E and Form-IC/D RbcS sequences did not identify any signature positions (Methods).

We examined how these signature positions evolved through the emergence of RbcS by tracking their ancestral amino acid composition through the divergence of Form-I and Form-I′ clades. Residues at the signature positions in Anc-I and Anc-I′ most closely resemble those of their respective descendants (Fig. 3b). By contrast, the common ancestor of all Form-I and Form-I′ RuBisCOs, Anc-I/I′, has approximately equal proportions of Form-I- and Form-I′-like signature positions. Among these, signature positions at the Anc-I/I′ RbcL intradimer interface (labeled 70,74) are Form-I′-like, whereas those near the RbcL active site (labeled 201, 401, 402) are Form-I-like. Notably, the Anc-I/I′ RbcL dimer–dimer interface and RbcL–RbcS interface contains a mixture of Form-I- and Form-I′-like signature positions. These results identify the interface residues that are likely important for oligomerization of the RuBisCO complex and underwent distinct functional specialization in the presence (in Form-I) or absence (in Form-I′) of RbcS.

In sum, the signature positions are not restricted to the RbcL–RbcS interface but are scattered across the RbcL structure (Fig. 3). We suggest that RbcS integration resulted in distant mutations, which are known to influence protein function by altering intramolecular interactions and the global dynamics of protein complexes (Miton et al. 2021). We therefore submit that these positions reflect how the presence or absence of RbcS affected long-range motions and conformational dynamics of the RuBisCO complex, and therefore its evolutionary dynamics.

Molecular Dynamics Simulations of Modern and Ancient RuBisCOs

We studied the structural motions of RuBisCO complexes to investigate the influence of the presence or absence of RbcS.

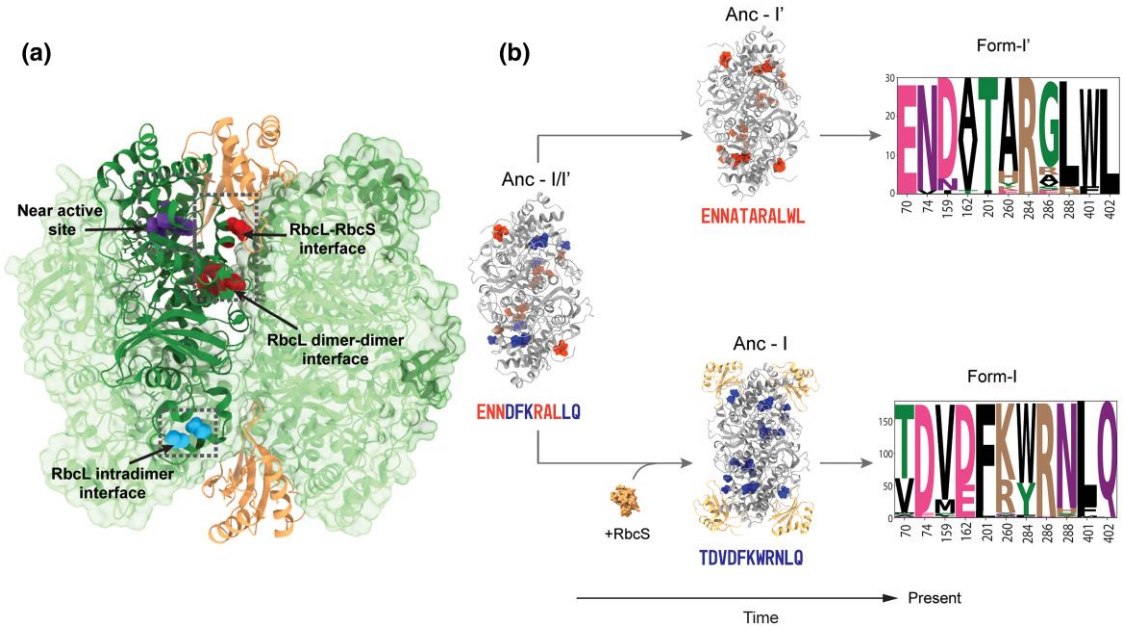

Fig. 3. Schematic representation of the structural distribution and evolution of the specialized protein "signature positions" between the RuBisCO Form-I and Form-I′. a) Signature residues are highlighted based on their structural proximity to each other along with the region it belongs to on the *Rhodobacter sphaeroides* L_4S_2 structure (PDB: 5NV3). Residues at the signature positions located at the RbcL intradimer interface, near active site and RbcL dimer–dimer interface are highlighted in blue, purple and red, respectively. b) Evolution of the signature residues corresponding to the divergence of Form-I and Form-I′ RuBisCOs. Amino acids present in the sequence motif are highlighted on the respective ancestor's L_2 or L_2S_4 structures. Sequence logos showing amino acid frequencies at each signature position across the Form-I and Form-I′ clades. Residue numbering according to the *R. sphaeroides*, L denotes RbcL and S denotes RbcS.

We built the ancestral RuBisCO large and small subunit structures (Methods). Modeling of the multimeric structural assembly of ancestral RuBisCO oligomers was guided by extant RuBisCO crystal structures (Methods). We simulated the molecular dynamics of seven ancestral and six extant RuBisCO complexes representing a range of host organisms and their associated environments (Fig. 4a, Table 1). We performed two simulation replicates for each RuBisCO variant, and data from both replicates are included in the analysis.

We calculated the root mean square deviation (RMSD) of backbone atoms for all residues in the RbcL and RbcS subunits relative to their average trajectory structure to evaluate conformational differences in the complex during the simulation. The mean RMSD of all the individual large and small subunits across all complexes is 0.928Å and 0.964Å, respectively. The average RbcL RMSD ranges from 0.78Å to 1.13Å, while the RbcS RMSD ranges from 0.76 to 1.26Å (Fig. 4b). The outliers in Anc-IAB small subunit RMSD values (Fig. 4b) reflect the N-terminal residues residue protrusion and unfolding observed across the simulation replicates. We performed pairwise comparisons of mean RMSD values among the different RuBisCOs. Ten out of all possible pairs show a significant pairwise RMSD difference for RbcL ($P < 0.01$, Tukey post-hoc test). And eight of these include the Anc-I′ or Anc-I/I′ complex, with a significantly higher RMSD in each case (supplementary table S3, Supplementary Material online). A higher RMSD for the Anc-I/I′ and Anc-I′ RbcL suggests a greater structural flexibility compared with the other complexes during the simulations.

We used the root mean square fluctuation (RMSF) of the different residues to identify region-specific differences in flexibility. Specifically, we studied the ancestral structures immediately before and after divergence of RbcS-less Form-I′ and RbcS-containing Form-I RuBisCOs. These ancestors include Anc-I/I′, Anc-I′, and Anc-I. The alteration in residue-wise fluctuations between the Anc-I′ and Anc-I/I′ complex during the MD simulation is close to zero (Fig. 4c, top). In contrast, the Anc-I RbcL exhibits less fluctuation compared with Anc-I/I′ during the simulations, where four distinct sections from the large subunit sequence show considerably higher fluctuations Anc-I/I′ compared with Anc-I (Fig. 4c, bottom). The sections with higher flexibility correspond to the top region of the large subunit that is positioned between the two small subunits in the RuBisCO complex (Fig. 4c, right). Out of the 45 RbcL residues present at the RbcS interface in Anc-I, 23 residues (~51%) show a significant difference in RMSF. The localization of RbcL residues near RbcS suggests that the presence of RbcS restricts the movement of these residues. This observation potentially accounts for the higher stability of the Anc-I large subunit compared with the RbcS-less Anc-I/I′ and Anc-I′. The differences in RMSF lend support to the hypothesis that RbcS plays a role in stabilizing the RbcL dimers in the octameric complex (Spreitzer 2003).

The extant Form-I′ RuBisCO (6URA) provides an additional basis for comparative analysis. Along with Anc-I′ and Anc-I/I′, RbcS is also absent in 6URA. Unlike its ancestor, however, the extant Form-I′ complex does not display a significantly higher RMSD during the pairwise comparison with the other RuBisCO systems (supplementary table S3, Supplementary Material online). Similarly, 6URA RbcL displays lower fluctuations per residue compared to the Anc-I/I′ and Anc-I′ RbcL (supplementary fig. S9, Supplementary Material online). This indicates that in the absence of RbcS, the modern Form-I′ RbcL has evolved a more stable L_8 without RbcS, potentially through other mutations. The L_8 assembly in the extant Form-I′ complex structure is maintained by a network of hydrogen bonds and salt bridges (Banda et al. 2020). Some of the key residues Asp161, Trp165 and Tyr224 involved in these

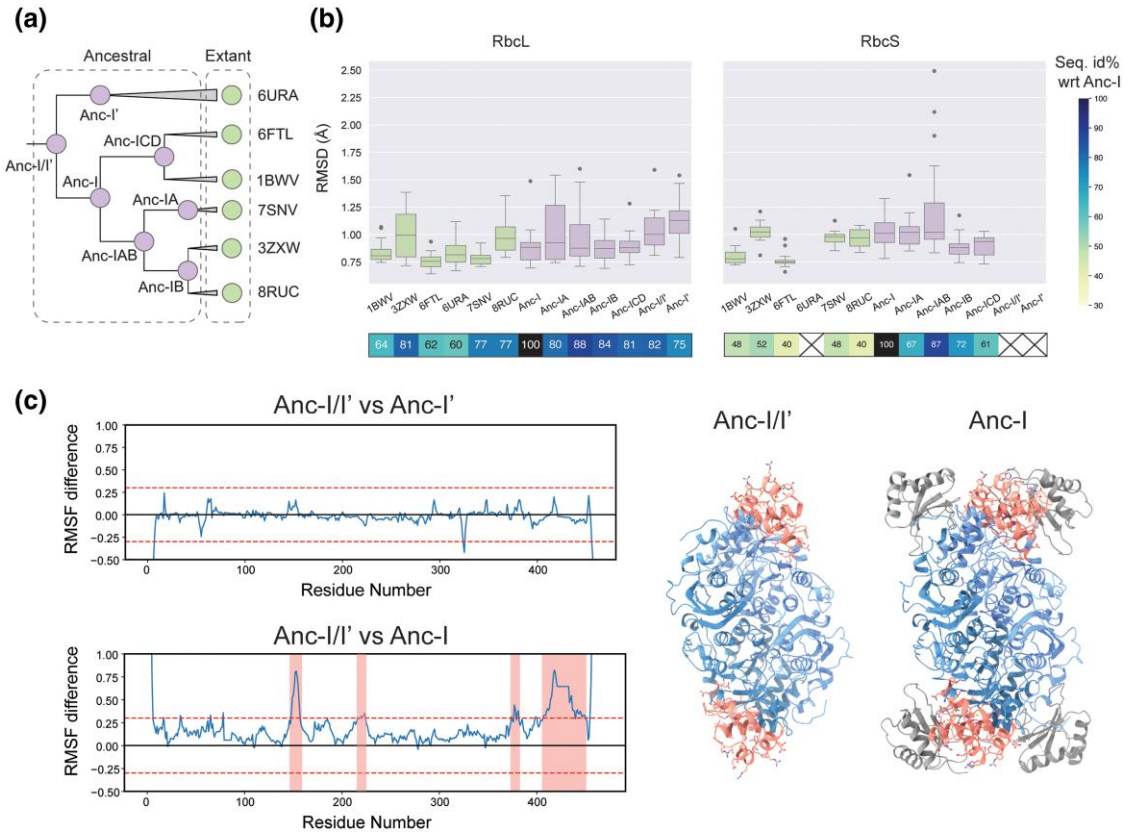

Fig. 4. Stability and fluctuation observed in the ancestral and extant RuBisCOs during MD simulations. a) Ancestral and extant RuBisCO homologs utilized in the study, distinct colors indicating the ancestral versus extant forms (Violet: ancestral RuBisCO, green: extant RuBisCO). b) Average root mean square deviation (RMSD) of all amino acids in the large and small subunits throughout the simulations. Outliers outside the quartile range are represented by circles. Heatmap on the *x*-axis represents the pairwise sequence identity for each sequence relative to the oldest RbcS ancestor, Anc-I. c) Difference in residue-wise root mean square fluctuation (RMSF) for RbcL between Anc-I/I′ versus Anc-I′ (top), and Anc-I/I′ versus Anc-I (bottom). Horizontal dashed red line represents the RMSF difference of 0.3 Å, used as threshold to classify the residue fluctuation as considerable, and the horizontal black line represents the baseline when there is a zero difference in RMSF between the residues from the two proteins. Residues with RMSF over 0.3 Å are highlighted in red on the Anc-I/I′ and Anc-I L_2 and L_2S_4 complex structures, respectively, on the right.

Table 1 Overview of extant RuBisCO structures selected for analysis

Structure	Form	Source organism	Environment[a]
1BWV (Sugawara et al. 1999)	I-CD	*Galdieria partita* (Red Algae)	High temperature, low pH
6FTL (Valegård et al. 2018)	I-CD	*Skeletonema marinoi* (Diatom)	Low temperature, arctic waters
7SNV (Blikstad et al. 2023)	I-A	*Halothiobacillus neapolitanus* (Gammaproteobacteria)	Aerobic, with alpha-carboxysome
8RUC (Andersson 1996)	I-B	*Spinacia Oleracea* (Spinach)	Mesophile, surface
3ZXW	I-B	*Thermosynechococcus vestitus* (Cyanobacteria)	Mesophile, tropics, with beta-carboxysome
6URA (Banda et al. 2020)	I′	*Candidatus Promineofilum breve* (Chloroflexi)	Anaerobic, marine sediments

[a]Environment specifies the known natural habitat or ecological niche and presence of CO_2-concentrating mechanism of the source organism. All the selected structures for extant RuBisCOs (except 7SNV) are complexed with 2-Carboxyarabinitol 1,5-Bisphosphate (CAP) intermediate, Magnesium ion (Mg^{2+}) and the carbamylated lysine (KCX) at the active site (Methods). The 8RUC structure does not include the N-terminal extension of the RbcS subunit.

interactions in the extant 6URA structure are not conserved in the analyzed RbcS-less ancestors, Anc-I′ and Anc-I/I′ (Asn140, Arg144 and Phe203, respectively) (supplementary fig. S7, Supplementary Material online). The absence of these key residue interactions may explain the higher fluctuations observed in the Anc-I′ and Anc-I/I′ RbcLs during the simulation.

While RbcS is indispensable for the activity of extant Form-I RuBisCOs (Andrews 1988; Lee and Tabita 1990), its present-day importance does not necessarily imply that it played an essential role at its initial emergence. Protein subunits can become entrenched in complexes via accumulations of neutral mutations that can be deleterious in monomers (Hochberg et al. 2020). However, our analysis further indicates the early importance of RbcS for enhanced stability of the RbcL octamer, along with prior experimental work demonstrating enhanced specificity toward CO_2 (Schulz et al. 2022). These findings collectively suggest an early advantage of RbcS in RuBisCO assembly at the time of its acquirement.

Impact of RbcS Integration on the Conformational Variation of Form-I RbcL

The global distribution of RbcL signature positions (Fig. 3), inferred to be functionally linked to the presence of RbcS,

suggests that RbcS integration affected the large subunit's overall dynamics and motion. We utilized MD simulations for RuBisCO complexes with and without RbcS to investigate the impact of RbcS on the enzyme's ability to explore conformational space. A protein's conformational space can be visualized as a set of different conformational states and the conformational variability of an enzyme is considered as an important factor for its functional diversity (Nobeli et al. 2009; Babtie et al. 2010).

To characterize the variation in conformations for the different RuBisCO simulations, we performed a principal component analysis (PCA) of the trajectory of the RbcL Cα atoms (Hayward and de Groot 2008; David and Jacobs 2014). This analysis allows us to reduce the dimensionality of the simulation trajectories into a few relevant dimensions, referred to as principal components (PC). To compare different RbcLs, we extracted the coordinates of conserved residues across all RbcLs and projected the trajectories onto a shared set of PCs (Methods). The dimensionality reduction shows the changes along the top two PCs (PC1 and PC2) for the RbcL's MD trajectories, capturing ~27% of all RbcL motions during the simulations (Fig. 5a). We observe that RbcL from each RuBisCO variant clusters closely together (Fig. 5b) suggesting that the dynamics of each individual RbcL are similar to those of its counterparts within the RuBisCO complex.

We observe that Form-I RuBisCOs (with RbcS) and non-Form-I RuBisCOs (without RbcS) occupy distinct regions along PC1 (Fig. 5c) with statistical significance (U-statistic = 2.09×10^7; $P = 0.0$; Mann–Whitney U test). In contrast, PC2, does not show a specific trend between RuBisCO variants. This suggests that the primary conformational variation (~19% of all RbcL motions) corresponds to the distinction between Form-I and non-Form-I RuBisCOs.

We simulated the dynamics of Anc-I, the common ancestor of Form-I RuBisCO, without the RbcS subunit and projected its trajectory onto the previously defined PCs for Form-I and Form-I′. The results show an intermediate positioning between Form-I and Form-I′ RuBisCOs (supplementary fig. S10a, Supplementary Material online). Removing RbcS from Form-I RuBisCO, shifts it dynamics closer to Form-I′, but does not fully replicate Form-I′ behavior. This indicates that RbcS plays a critical, though not exclusive, role in driving the distinct conformational variations between Form-I and Form-I′.

We assessed the impact of octameric oligomerization of RbcL on enzyme dynamics by including MD simulations of a Form-II RuBisCO (9RUB) in our PCA decomposition. Incorporating Form-II RbcL generates a distinct set of PCs (*PC′*). The first PC (PC1′) separates Form-II from Form-I and Form-I′; while the second PC (PC2′) distinguishes Form-I and Form-I′ (supplementary fig. S10b, Supplementary Material online). This separation suggests that the octamerization in Form-I/Form-I′ induces significant conformational changes compared with the non-octameric Form-II, with the Form-I and Form-I′ distinction becoming apparent only in PC2′.

Our results suggest that the presence or absence of RbcS in the RuBisCO complex impacts the major dynamics of RbcL. We hypothesize that the incorporation of RbcS induces a conformational shift in RbcL, a modification that has remained consistent across diverse Form-I and Form-I′ RuBisCO variants over evolutionary time. The functional diversity (e.g. promiscuity) of a protein is closely tied to its ability to explore diverse conformational states (Tokuriki and Tawfik 2009; Zou et al. 2015; Petrović et al. 2018; Jackson et al. 2022). In this context, our analysis shows that the RbcL explores distinct conformations in the presence of RbcS, suggesting that the shift in RuBisCO's conformational variability may be directly associated with the enzyme's increased CO_2-specificity following the emergence of RbcS (Schulz et al. 2022).

Heterogeneity of CO_2/O_2 Diffusion Patterns Across Ancestral and Extant RuBisCOs

Previous work suggested that RbcS acts as a CO_2-reservoir (Van Lun et al. 2014), increasing the concentration of CO_2 molecules near the enzyme and making CO_2 more accessible to the active site. This hypothesis also provides an additional functional justification for the integration of RbcS, given the increase in Earth's atmospheric O_2 concentration ~2.5 billion years ago, around the time of RbcS integration into the RuBisCO complex.

We performed MD simulations for the different RuBisCO complexes with CO_2 and O_2 gas molecules present in the

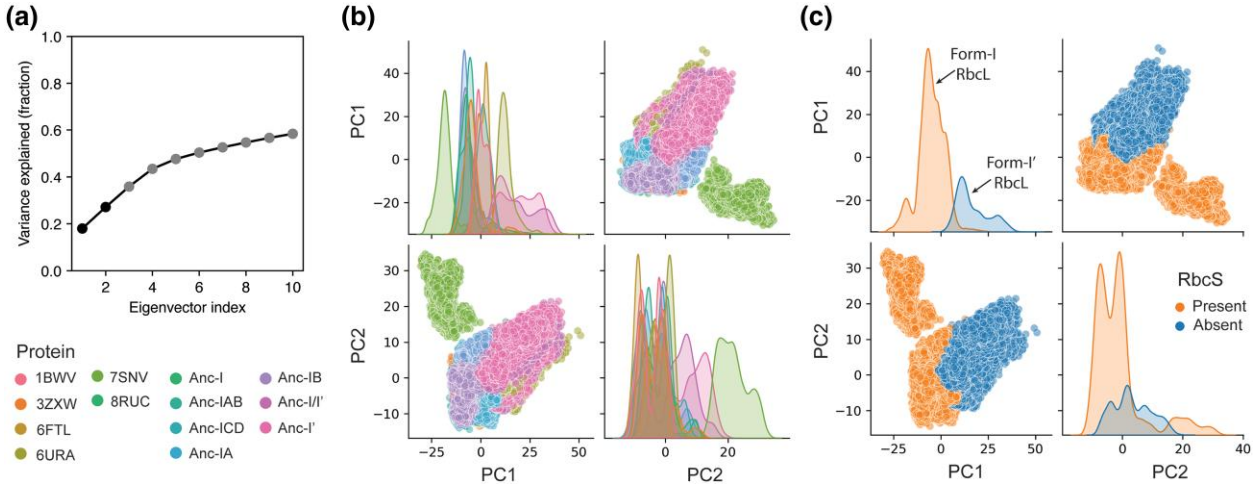

Fig. 5. Principal component analysis for the conserved RbcL residues of MD simulation trajectories across ancestral and extant Form-I and Form-I′ RuBisCO variants. a) Variance explained by each eigenvector in the PCA, showing the contribution of the top 10 principal components to overall fluctuation across all simulations. b) Pairwise representation of PC1 and PC2, highlighting the different Form-I and Form-I′ RuBisCO variants. c) Pairwise representation of PC1 and PC2, highlighting the RbcS presence and absence in the RuBisCO complex.

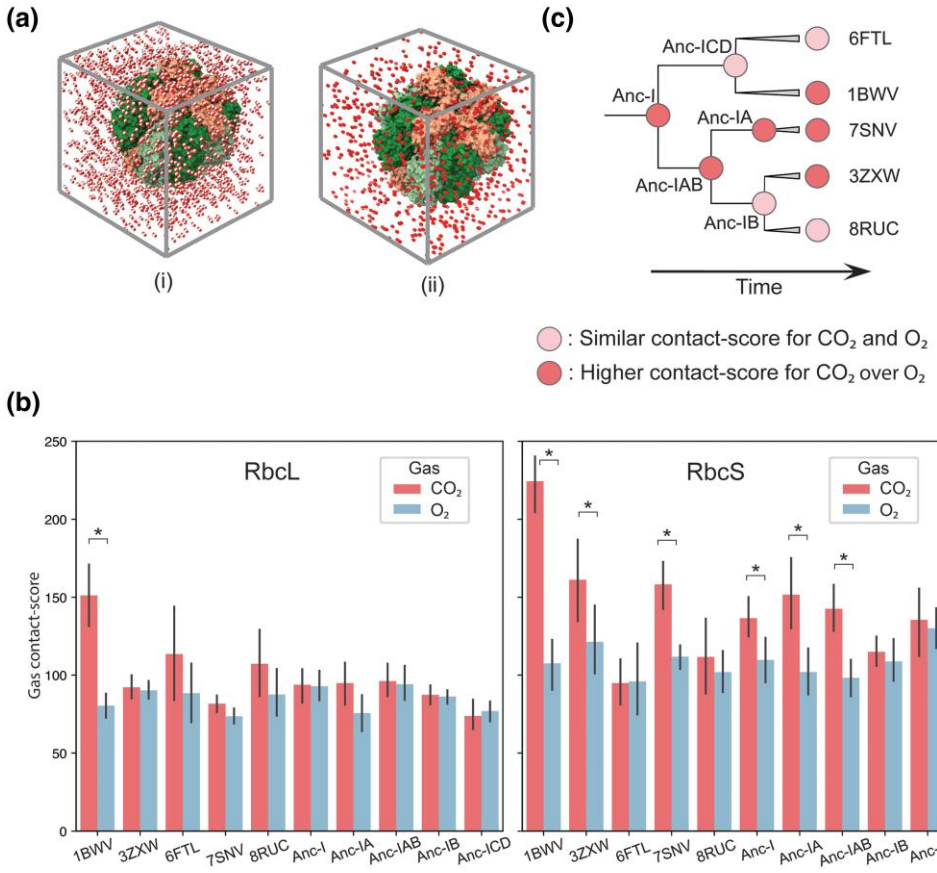

Fig. 6. CO_2 and O_2 gas diffusion molecular dynamics across different RuBisCO systems. a) A snapshot of the MD simulation box with CO_2 (i) and O_2 (ii) gas molecules in the medium at a concentration of 400 mM with the protein complex. b) Barplot representing the contact-score (number of gas contacts per residue) for RbcL and RbcS for each of RuBisCO for the two CO_2 and O_2 gas molecules. Error bars represent the standard deviation across the eight subunits in the RuBisCO complex. Proteins with a significant difference between the CO_2 and O_2 affinity are marked with an asterisk (*). c) Trend for the difference in contact-score for CO_2 and O_2 gas molecules for RbcS as observed in the barplot (B). Time is displayed along the *y*-axis.

medium (Fig. 6a) to assess the interaction of RbcS residues toward both gases. We estimated the relative contact-scores of the gas molecules by counting the number of interactions between the gases and the protein throughout the simulation within a distance threshold of 6Å. The MD simulations for O_2 and CO_2 gas molecules were performed separately (Fig. 6a). Five independent replicates for MD simulations of one extant (8RUC) and one ancestral (Anc-I) enzyme were conducted to assess the robustness and replicability of the gas dynamics modeling approach. To assess the impact of gas concentrations, we performed simulations with varying CO_2 and O_2 gas concentrations. The trends observed for affinity of the subunits across different gas concentrations remains relatively consistent between the two subunits (supplementary fig. S11, Supplementary Material online).

An independent two sample t-test was performed to compare the mean CO_2 and O_2 gas contacts across all simulations. For RbcL, we observe that only 1BWV has a significantly higher contact-score for CO_2 over O_2 ($t = 5.92$; $P = 3.7 \times 10^{-5}$, two-sided), whereas other complexes do not show a statistically significant ($P > 0.05$) difference (Fig. 6b). Alternatively, six RbcS (1BWV, 3ZXW, 7SNV, Anc-I, Anc-IA, Anc-IAB) out of the ten exhibit a significantly ($P < 0.05$) higher contact-score for CO_2 than for O_2 molecules in the surrounding medium (Fig. 6b, supplementary table S4, Supplementary Material online). RbcS that exhibit a higher affinity for CO_2 over O_2 are not confined to a specific clade in the phylogenetic tree (Fig. 6c). We find that these results do not vary significantly over five independent simulations (supplementary fig. S12, Supplementary Material online).

Conclusions

The co-evolution of the large and small subunits of RuBisCO presents an opportunity to analyze molecular evolutionary events within key macroevolutionary and geochemical transitions, independently documented by Earth's geologic record. We examined the historical impact of RuBisCO small subunit emergence on the structural motions and evolutionary trajectory of the enzyme, with specific focus on a possible adaptive role in concentrating CO_2. Both extant and ancestral Form-I RuBisCO complexes that contain RbcS show higher CO_2-specificity than other forms of RuBisCO lacking RbcS (Flamholz et al. 2019; Schulz et al. 2022). These observations have led to the hypothesis that the improved CO_2-specificity of RuBisCO after RbcS integration would have been advantageous after the GOE, given the subsequent rise in atmospheric O_2 levels and decrease in CO_2 concentrations. We illustrate that not all ancestral and extant RbcS proteins act as a reservoir to concentrate CO_2 around the RuBisCO complex. It is possible that certain RuBisCOs with RbcS might have adopted a CO_2-reservoir strategy based on other intercellular or possibly environmental factors as we find that this feature appears multiple times in the enzyme's history. Moreover, our analysis

highlights additional consequences of RbcS integration: increased stability of Form-I RbcL and decreased flexibility compared with Form-I′ RbcL. Thus, the small subunit integration shaped the enzyme's major conformational and functional variations in response to significant shifts in atmospheric composition.

The origin of the small subunit still remains unresolved, but these observations shed light on the circumstances that surrounded and likely facilitated the emergence of RbcS. Certain cyanobacteria have two other proteins that are part of the carbon fixation machinery: β-carboxysome structural protein (CcmM) and a RuBisCO activase-like protein (ALC). Both these proteins have one or multiple domains that are "RbcS-like" and are considered homologs to RbcS (Ryan et al. 2019; Wang et al. 2019; Lechno-Yossef et al. 2020). The presence of these protein domains is specific to cyanobacteria. Further studies should explore whether the RbcS-like protein domains could have emerged alongside or following the RbcS in response to the ancient shifts in atmospheric CO_2 and O_2 levels. In summary, we show that ancestral RuBisCO dynamically responded to a global environmental shift. The integration of the small subunit resulted in increased rigidity, allowing it to maintain substrate specificity despite decreased substrate availability. Ironically, this natural solution to an ancient challenge may have led to modern RuBisCO's notorious resistance to artificial improvements in CO_2 specificity.

Methods

Phylogenetic Reconstruction of RuBisCO

Homologous sequences for RbcS protein from the NCBI nonredundant database were identified using PSI-BLAST (Altschul et al. 1997; Altschul and Koonin 1998). The RbcS sequence from *Thermosynechococcus elongatus* (PDB id: 2YBV) (Gubernator et al. 2008) was used as the query sequence for five PSI-BLAST iterations with 50,000 sequences per iteration and an E-value cutoff of 0.005. The dataset was curated to remove partial sequences and sequences with incomplete annotations. The sequences in the dataset were dereplicated at 63% amino acid identity with CD-Hit (Li and Godzik 2006) and aligned by MAFFT (Katoh et al. 2002) (default parameters) with iteration refinement over 1,000 cycles. Any poorly aligned sequences were subsequently replaced with a different representative of their CD-HIT cluster.

A maximum-likelihood phylogenetic tree was constructed using IQ-Tree (Nguyen et al. 2015) with the LG + R6 evolutionary model (supplementary fig. S1, Supplementary Material online). ModelFinder Plus (Kalyaanamoorthy et al. 2017) was used to select the best-fit evolutionary model. Branch support values were calculated using the Shimodaira–Hasegawa–like approximate likelihood-ratio test (SH-aLRT) with 1,000 bootstrap replicates and 1,000 ultrafast bootstrap (UFBoot) replicates optimized by nearest neighbor interchange (NNI).

RbcL sequences from the same taxa represented in our RbcS sequence dataset were identified using the NCBI Identical Protein Groups database (IPG) (https://www.ncbi.nlm.nih.gov/ipg). For RbcS entries with missing IPG RbcL sequences, Taxonomy-restricted BLASTp was used to search for the corresponding RbcL sequence. Form-I RbcL sequences were combined with RbcL sequences for Form-I′, II, III and IV from Banda et al. (Banda et al. 2020) to build an RbcL dataset. An RbcL phylogenetic tree was constructed using IQ-Tree with the LG + R9 evolutionary model (all other parameters were the same as for the RbcS tree) (supplementary fig. S2, Supplementary Material online).

The sequences in the RbcL dataset were aligned by MAFFT, with the same parameters as those mentioned above, and were concatenated with the RbcS alignment. A maximum-likelihood phylogenetic tree was constructed for the concatenated alignment using IQ-Tree, with a partition model (Chernomor et al. 2016) and ModelFinder plus (Kalyaanamoorthy et al. 2017) to search and implement the best-fit evolutionary models corresponding to the RbcL (LG + R9) and RbcS (LG + R5) section of the concatenated alignment. In line with prior research, Form-IV "RuBisCO-like" protein sequences were used to root the phylogenetic tree (Kacar et al. 2017; Banda et al. 2020; Poudel et al. 2020; Camel and Zolla 2021; Schulz et al. 2022). Ancestral sequence inference for the concatenated RbcL–RbcS tree was performed using PamL4.9 (LG model) (Yang 2007). Reconstruction of the gaps in the ancestral sequences was performed using a binary likelihood model as described by Aadland et al. (Aadland et al. 2019).

RbcS Sequence and Structural Diversity Analysis

Candidate RbcS structures (listed in Table 1) from Form-I subclades were used to identify the different structural features in RbcS. The amino acid regions corresponding to these features were identified in the multiple sequence alignment of extant RbcS. The binary heatmap indicating the presence or absence of these features across all extant RbcS sequences in the phylogeny was generated based on the presence of at least half of the residues in the respective region of the RbcS alignment. Host taxonomy for the extant sequences was assigned using the REST API provided by Ensembl (https://rest.ensembl.org/).

RbcL Specialized Signature Position Analysis

Sequence specialization between the Form-I and Form-I′ extant RbcL sequences was analyzed using TwinCons (Penev et al. 2021), with blosum62 as substitution matrix for score calculation, and Zebra2 (Suplatov et al. 2020), using the webserver default parameters. Specialized residues identified by Zebra2 and TwinCons score were then filtered based on conservation in the multiple sequence alignment for Form-I and Form-I′ sequences. Same steps as above were followed to identify separately conserved amino acids between 131 RbcS sequences from Form-IA/B/E and 59 from Form-IC/D. None of the RbcS positions had a TwinCons score < -1, resulting in no signature positions.

Structural Modeling of RuBisCO

We predicted the structures of ancestral RuBisCOs using the deep-learning based Colabfold software (Mirdita et al. 2022). We built our models by combining predictions and structure alignments. We built the model for the RbcL–dimer with RbcS–dimer for each ancestral protein with templates from PDB, structure refinement using Amber and three prediction recycles. Low-confidence modeled terminal regions for the predicted structures were removed using a pLDDT threshold of 50. The extant RuBisCO structure from *Thermosynechococcus elongatus* (PDB: 2YBV) was used as a template to obtain a predicted hexadecameric complex using UCSF Chimera

(Pettersen et al. 2004). Anc-I′ and Anc-I/I′ did not have an RbcS sequence resulting in an octameric complex.

RuBisCO MD Simulations

We simulated the molecular dynamics of both extant and ancient RuBisCO complexes. We selected seven ancestral structures (nodes highlighted in Fig. 2a) corresponding to the last common ancestors of different Form-I and Form-I′ clades and six experimentally determined extant RuBisCO structures, representing the Form-IA, IB (two from this clade), IC/D (two from this clade), and I′ RuBisCO clades. PDB IDs for these entries are: 1BWV (red algae), 3ZXW (green algae), 6FTL (diatom), 6URA (chloroflexi), 7SNV (gammaproteobacteria) and 8RUC (plantae) (more details in Table 1). The 6URA entry represents the first case discovered of a Form-I-like RuBisCO without a small subunit. This selection represents the biological diversity of the RuBisCO enzyme while preserving the following attributes: high resolution (under 3.0 Å), carbamylation of the lysine at the active site (except for 7SNV) (Stec 2012), and CABP and magnesium at the active site. Sequence identity and alignments for ancestral and extant RuBisCOs analyzed in the study are presented in supplementary fig. S6 and S7, Supplementary Material online, respectively.

We obtained parameters for the CABP molecule through the standard Amber protocol for ligands (Wang et al. 2004). RuBisCO proteins include different post-translational modifications that required additional steps. We computed the CABP and the carbamylated lysine charges at HF/6-31*G level, as specified in the Amber protocol. The post-translational modifications found in 6FTL were modeled with semi-empirical charges. In the case of the modeled ancient structures, we manually added both ligands (CABP), ions (Mg^{2+}), and the post-translation modification at the carbamylated lysine (also for 7SNV).

We built the topology and parameters of the molecular dynamics systems using tleap (Maier et al. 2015) and Amber14 (Salomon-Ferrer et al. 2013). Periodic solvation boxes were constructed with 10 Å spacing and water molecules according to the TIP3P model (Jorgensen et al. 1983). Sodium and chloride ions counterbalanced the charge of the system. The particle-mesh Ewald summation method was used for long-range electrostatics and a 10 Å cutoff was set for short-range non-bonded interactions. Initial geometries in all systems were minimized at 5,000 conjugate-gradient steps after which water was equilibrated at 298 K and 1 atm for 100 ps at 2 fs time steps. Production runs were then performed for 250 ns in the NPT ensemble at $P = 1$ atm and $T = 298$ K. We used the hydrogen mass partitioning method to ensure the stability of the simulation under large integration steps, thereby increasing the simulation speed. Langevin dynamics for T control and Nosé-Hoover Langevin piston method for P control were used. We carried out the MD simulation on OpenMM 7.7 (Eastman et al. 2017) running in the Nvidia Tesla A100, L40 and H100 GPU nodes of the Center for High Throughput Computation at the University of Wisconsin-Madison. We performed two replicates of MD simulations. The equilibration of simulations across different complexes was evaluated by tracking the RMSD of the complex structure over time for both replicates (supplementary fig. S8, Supplementary Material online). The RMSD across all RuBisCO complexes stabilizes at a plateau, indicating that the complexes reached an equilibrated state.

Gas Diffusion MD Simulations

We obtained CO_2 and O_2 parameters through the standard Amber protocol. We computed the charges of these molecules at the HF/6-31*G level through the RESP protocol (Woods and Chappelle 2000). The embedding of proteins in solutions with gas molecules in dissolution consisted of replacing water molecules from a previously solvated system to avoid clashes among the gas molecules. We calculated the number of molecules corresponding to a given concentration using the Amber-recommended method for determining ion concentrations in a solvent. All other simulation aspects were consistent with those described in the previous section. These simulations were run for 75 ns, with the final 50 ns used for analysis following a 25 ns equilibration period.

Analysis of MD Simulations

We employed MDAnalysis (Michaud-Agrawal et al. 2011) and ProDy (Bakan et al. 2011) to analyze and process the outcome of the MD simulations. All simulations were conducted under identical temperature, pH, and solvent conditions. The initial 50 ns of the 250 ns simulations were designated for equilibration, and only the final 200 ns were used for analysis. Data from both MD simulation replicates were included, resulting in a combined simulation time exceeding 7 µs. The average system size was approximately 300,000 atoms, including solvent molecules.

RMSD and RMSF Calculations

We aligned all the simulation trajectories to their average structure and calculated the RMSD across all Cα backbone atoms for each RbcL and RbcS subunit using MDAnalysis. The boxplot in Fig. 4b presents the average RMSD value for each large and small subunit across the two simulation replicates. RMSF was calculated by aligning the MD trajectories to their reference structure first. The Cα atomic positions of each residue were mapped onto the respective sequence, and fluctuations were calculated as the average positional deviation from the mean structure over the simulation period across the eight RbcL subunits using MDAnalysis. Simulation data from replicates were included in the calculations of the RMSD and RMSF.

Principal Component Analysis

The principal component analysis helps reduce the complexity of MD simulation data by generating orthogonal eigenvectors (principal components, PCs) that represent the primary axes of motion. We conducted PC analyses using the scikit-learn implementation (Pedregosa et al. 2011). Specifically, we first extracted all conserved residues across select RbcL sequences shown in Fig. 4a. All the individual trajectories for each RbcLs were then aligned onto the reference extant RbcL structure (1BWV). The covariance matrix of positional fluctuations was constructed exclusively for the conserved RbcL Cα atoms. Eigenvalues and eigenvectors were derived from this covariance matrix to identify the PCs shared across all RbcL trajectories, including each simulation replicate. PCA, including the Anc-I without RbcS (supplementary fig. S10a, Supplementary Material online), was conducted by extracting trajectories for conserved residues and projecting them onto the PC-space defined by Form-I and Form-I′ RuBisCOs in Fig. 5. Additionally, a separate PCA was performed following the same procedure, incorporating RbcL simulation trajectories from Form-II

alongside Form-I and Form-I′, as shown in supplementary fig. S10b, Supplementary Material online.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Alessandro Senes, Phil Romero, Tina Wang and Srivatsan Raman for their feedback on the study, Aya Klos, Evrim Fer, Josh MacCready, Zach Adam and Amanda Garcia for critical reading and useful comments and suggestions on the manuscript. The Center for High Throughput Computing (CHTC) at the University of Wisconsin-Madison provided computing resources for this work.

Author Contribution

Conceptualization: KA, BCZ and BK; Methodology: KA, BCZ and BK; Data curation: KA and BCZ; Data analysis: KA, BCZ and BK; Figures: KA, BCZ and BK; Writing editing: KA (First draft) and BK; Resources and supervision: BK. All authors edited and approved the final draft.

Funding

This work was supported by the NASA Exobiology Program [NNH23ZDA001N] with additional support from the Human Frontier Science Program (HFSP) [RGY0072/2021], the NASA Interdisciplinary Consortia for Astrobiology Research (ICAR) Program [80NSSC17K0296] and the Hypothesis Fund. B.C.Z acknowledges the Margarita Salas Postdoctoral Fellowship, founded by the Unión Europea—Next Generation EU (B.C.Z.; UP2021-035).

Conflict of interest: The authors declare no competing interests.

Data Availability

The data underlying this article, along with the code used for analysis, can be found at: https://github.com/kacarlab/RuBisCO_evolution. Molecular dynamics simulation files for extant and ancestral RuBisCOs are available on Zenodo, at https://zenodo.org/records/14187581.

References

Aadland K, Pugh C, Kolaczkowski B. High-Throughput reconstruction of ancestral protein sequence, structure, and molecular function. In: Sikosek T, editors. *Computational methods in protein evolution*. New York, NY: Springer; 2019. p. 135–170.

Altschul SF, Koonin EV. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci*. 1998:23(11):444–447. https://doi.org/10.1016/S0968-0004(98)01298-5.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997:25(17):3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Andersson I. Large structures at high resolution: the 1.6 Å crystal structure of spinach ribulose-1,5- bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate. *J Mol Biol*. 1996:259(1):160–174. https://doi.org/10.1006/jmbi.1996.0310.

Andersson I. Catalysis and regulation in rubisco. *J Exp Bot*. 2008:59(7):1555–1568. https://doi.org/10.1093/jxb/ern091.

Andrews TJ. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J Biol Chem*. 1988:263(25):12213–12219. https://doi.org/10.1016/S0021-9258(18)37741-X.

Ashida H, Danchin A, Yokota A. Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism? *Res Microbiol*. 2005:156(5-6):611–618. https://doi.org/10.1016/j.resmic.2005.01.014.

Babtie A, Tokuriki N, Hollfelder F. What makes an enzyme promiscuous? *Curr Opin Chem Biol*. 2010:14(2):200–207. https://doi.org/10.1016/j.cbpa.2009.11.028.

Bakan A, Meireles LM, Bahar I. Prody: protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011:27(11):1575–1577. https://doi.org/10.1093/bioinformatics/btr168.

Banda DM, Pereira JH, Liu AK, Orr DJ, Hammel M, He C, Parry MAJ, Carmo-Silva E, Adams PD, Banfield JF, *et al*. Novel bacterial clade reveals origin of form I rubisco. *Nat Plants*. 2020:6(9):1158–1166. https://doi.org/10.1038/s41477-020-00762-4.

Blikstad C, Dugan EJ, Laughlin TG, Turnšek JB, Liu MD, Shoemaker SR, Vogiatzi N, Remis JP, Savage DF. Identification of a carbonic anhydrase–rubisco complex within the alpha-carboxysome. *Proc Natl Acad Sci U S A*. 2023:120(43):e2308600120. https://doi.org/10.1073/pnas.2308600120.

Bouvier JW, Emms DM, Kelly S. Rubisco is evolving for improved catalytic efficiency and CO2 assimilation in plants. *Proc Natl Acad Sci U S A*. 2024:121(11):e2321050121. https://doi.org/10.1073/pnas.2321050121.

Bracher A, Whitney SM, Hartl FU, Hayer-Hartl M. Biogenesis and metabolic maintenance of rubisco. *Annu Rev Plant Biol*. 2017:68(1):29–60. https://doi.org/10.1146/annurev-arplant-043015-111633.

Caetano-Anollés G. RubisCO and the search for biomolecular culprits of planetary change. *Bioessays*. 2017:39(11):201700174. https://doi.org/10.1002/bies.201700174.

Camel V, Zolla G. An insight of RuBisCO evolution through a multilevel approach. *Biomolecules*. 2021:11(12):1761–1761. https://doi.org/10.3390/biom11121761.

Chernomor O, Von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 2016:65(6):997–1008. https://doi.org/10.1093/sysbio/syw037.

David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol Biol*. 2014:1084:193–226. https://doi.org/10.1007/978-1-62703-658-0_11.

Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD, *et al*. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLOS Comput Biol*. 2017:13(7):e1005659. https://doi.org/10.1371/journal.pcbi.1005659.

Erb TJ, Zarzycki J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Curr Opin Biotechnol*. 2018:49:100–107. https://doi.org/10.1016/j.copbio.2017.07.017.

Esquivel MG, Genkov T, Nogueira AS, Salvucci ME, Spreitzer RJ. Substitutions at the opening of the rubisco central solvent channel affect holoenzyme stability and CO2/O2 specificity but not activation by rubisco Activase. *Photosynth Res*. 2013:118(3):209–218. https://doi.org/10.1007/s11120-013-9916-0.

Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science*. 2008:320(5879):1034–1039. https://doi.org/10.1126/science.1153213.

Fernie AR, Bauwe H. Wasteful, essential, evolutionary stepping stone? The multiple personalities of the photorespiratory pathway. *Plant J*. 2020:102(4):666–677. https://doi.org/10.1111/tpj.14669.

Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, Oltrogge LM, Alves R, Savage D, Milo R. Revisiting trade-offs between rubisco

kinetic parameters. *Biochemistry*. 2019:58(31):3365–3376. https://doi.org/10.1021/acs.biochem.9b00237.

Garcia AK, Cavanaugh CM, Kacar B. The curious consistency of carbon biosignatures over billions of years of earth-life coevolution. *ISME J*. 2021:15(8):2183–2194. https://doi.org/10.1038/s41396-021-00971-5.

Gatenby AA. Synthesis and assembly of bacterial and higher plant rubisco subunits in Escherichia coli. *Photosynth Res*. 1988:17(1-2):145–157. https://doi.org/10.1007/BF00047686.

Genkov T, Meyer M, Griffiths H, Spreitzer RJ. Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in Chlamydomonas. *J Biol Chem*. 2010:285(26):19833–19841. https://doi.org/10.1074/jbc.M110.124230.

Gubernator B, Bartoszewski R, Kroliczewski J, Wildner G, Szczepaniak A. Ribulose-1,5-bisphosphate carboxylase/oxygenase from thermophilic cyanobacterium thermosynechococcus elongatus. *Photosynth Res*. 2008:95(1):101–109. https://doi.org/10.1007/s11120-007-9240-7.

Hayward S, de Groot BL. Normal modes and essential dynamics. In: Kukol A, editors. *Molecular modeling of proteins. Methods molecular Biology*^{TM}. Totowa, NJ: Humana Press; 2008. p. 89–106.

Hochberg GKA, Liu Y, Marklund EG, Metzger BPH, Laganowsky A, Thornton JW. A hydrophobic ratchet entrenches molecular complexes. *Nature*. 2020:588(7838):503–508. https://doi.org/10.1038/s41586-020-3021-2.

Jackson C, Toth-Petroczy A, Kolodny R, Hollfelder F, Fuxreiter M, Kamerlin SCL, Tokuriki N. Adventures on the routes of protein evolution—in memoriam dan salah tawfik (1955–2021). *J Mol Biol*. 2022:434(7):167462. https://doi.org/10.1016/j.jmb.2022.167462.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983:79(2):926–935. https://doi.org/10.1063/1.445869.

Joshi J, Mueller-Cajar O, Tsai Y-CC, Hartl FU, Hayer-Hartl M. Role of small subunit in mediating assembly of red-type form I rubisco. *J Biol Chem*. 2015:290(2):1066–1074. https://doi.org/10.1074/jbc.M114.613091.

Kaçar B. Reconstructing early microbial life. *Annu Rev Microbiol*. 2024:78(1):463–492. https://doi.org/10.1146/annurev-micro-041522-103400.

Kacar B, Hanson-Smith V, Adam ZR, Boekelheide N. Constraining the timing of the great oxidation event within the rubisco phylogenetic tree. *Geobiology*. 2017:15(5):628–640. https://doi.org/10.1111/gbi.12243.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002:30(14):3059–3066. https://doi.org/10.1093/nar/gkf436.

Kędzior M, Garcia AK, Li M, Taton A, Adam ZR, Young JN, Kaçar B. Resurrected rubisco suggests uniform carbon isotope signatures over geologic time. *Cell Rep*. 2022:39(4):110726. https://doi.org/10.1016/j.celrep.2022.110726.

Knight S, Andersson I, Brändén CI. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 A resolution subunit interactions and active site. *J Mol Biol* 1990:215(1):113–160. https://doi.org/10.1016/S0022-2836(05)80100-7

Lechno-Yossef S, Rohnke BA, Belza ACO, Melnicki MR, Montgomery BL, Kerfeld CA. Cyanobacterial carboxysomes contain an unique rubisco-Activase-like protein. *New Phytol*. 2020:225(2):793–806. https://doi.org/10.1111/nph.16195.

Lee B, Tabita FR. Purification of recombinant ribulose-1,5-bisphosphate carboxylase/oxygenase large subunits suitable for reconstitution and assembly of active L8S8 enzyme. *Biochemistry*. 1990:29(40):9352–9357. https://doi.org/10.1021/bi00492a007.

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006:22(13):1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

Liu AK, Pereira JH, Kehl AJ, Rosenberg DJ, Orr DJ, Chu SKS, Banda DM, Hammel M, Adams PD, Siegel JB, *et al*. Structural plasticity enables evolution and innovation of RuBisCO assemblies. *Sci Adv*. 2022:8(34):eadc9440. https://doi.org/10.1126/sciadv.adc9440.

Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU, *et al*. Coupled chaperone action in folding and assembly of hexadecameric rubisco. *Nature*. 2010:463(7278):197–202. https://doi.org/10.1038/nature08651.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein Side chain and backbone parameters from ff99SB. *J Chem Theory Comput*. 2015:11(8):3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

Mao Y, Catherall E, Díaz-Ramos A, Greiff GRL, Azinas S, Gunn L, McCormick AJ. The small subunit of rubisco and its potential as an engineering target. *J Exp Bot*. 2023:74(2):543–561. https://doi.org/10.1093/jxb/erac309.

Matsumura H, Shiomi K, Yamamoto A, Taketani Y, Kobayashi N, Yoshizawa T, Tanaka SI, Yoshikawa H, Endo M, Fukayama H. Hybrid rubisco with complete replacement of rice rubisco small subunits by Sorghum counterparts confers C4 plant-like high catalytic activity. *Mol Plant*. 2020:13(11):1570–1581. https://doi.org/10.1016/j.molp.2020.08.012.

Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011:32(10):2319–2327. https://doi.org/10.1002/jcc.21787.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022:19(6):679–682. https://doi.org/10.1038/s41592-022-01488-1.

Miton CM, Buda K, Tokuriki N. Epistasis and intramolecular networks in protein evolution. *Curr Opin Struct Biol*. 2021:69:160–168. https://doi.org/10.1016/j.sbi.2021.04.007.

Nealson KH, Conrad PG. Life: past, present and future. *Philos Trans R Soc Lond B Biol Sci*. 1999:354(1392):1923–1939. https://doi.org/10.1098/rstb.1999.0532.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.

Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. 2009:27(2):157–167. https://doi.org/10.1038/nbt1519.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al*. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011:12:2825–2830.

Penev PI, Alvarez-Carreño C, Smith E, Petrov AS, Williams LD. TwinCons: conservation score for uncovering deep sequence similarity and divergence. *PLoS Comput Biol*. 2021:17(10):e1009541. https://doi.org/10.1371/journal.pcbi.1009541.

Petrović D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. *J R Soc Interface*. 2018:15(144):20180330. https://doi.org/10.1098/rsif.2018.0330.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004:25(13):1605–1612. https://doi.org/10.1002/jcc.20084.

Poudel S, Pike DH, Raanan H, Mancini JA, Nanda V, Rickaby REM, Falkowski PG. Biophysical analysis of the structural evolution of substrate specificity in RuBisCO. *Proc Natl Acad Sci U S A*. 2020:117(48):30451–30457. https://doi.org/10.1073/pnas.2018939117.

Raymond J, Segrè D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science*. 2006:311(5768):1764–1767. https://doi.org/10.1126/science.1118439.

Rucker HR, Kaçar B. Enigmatic evolution of microbial nitrogen fixation: insights from Earth's past. *Trends Microbiol*. 2024:32(6):554–564. https://doi.org/10.1016/j.tim.2023.03.011.

Ryan P, Forrester TJB, Wroblewski C, Kenney TMG, Kitova EN, Klassen JS, Kimber MS. The small RbcS-like domains of the β-carboxysome structural protein CcmM bind RubisCO at a site distinct from that binding the RbcS subunit. *J Biol Chem*. 2019:294(8):2593–5195. https://doi.org/10.1074/jbc.RA118.006330.

Salomon-Ferrer R, Case DA, Walker RC. An overview of the amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci*. 2013:3:(2):198–210. https://doi.org/10.1002/wcms.1121.

Schmidt GW, Mishkind ML. The transport of proteins into chloroplasts. 1986. Available from: www.annualreviews.org.

Schulz L, Guo Z, Zarzycki J, Steinchen W, Schuller JM, Heimerl T, Prinz S, Mueller-Cajar O, Erb TJ, Hochberg GKA. Evolution of increased complexity and specificity at the Dawn of form I rubiscos. *Science*. 2022:378(6616):155–160. https://doi.org/10.1126/science.abq1416.

Spreitzer RJ. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys*. 2003:414(2):141–149. https://doi.org/10.1016/S0003-9861(03)00171-1.

Stec B. Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proc Natl Acad Sci U S A*. 2012:109(46):18785–18790. https://doi.org/10.1073/pnas.1210754109.

Sugawara H, Yamamoto H, Shibata N, Inoue T, Okada S, Miyake C, Yokota A, Kai Y. Crystal structure of carboxylase reaction-oriented ribulose 1,5-bisphosphate carboxylase/oxygenase from a thermophilic red alga, *Galdieria partita*. *J Biol Chem*. 1999:274(22):15655–15661. https://doi.org/10.1074/jbc.274.22.15655.

Suplatov D, Sharapova Y, Geraseva E, Švedas V. Zebra2: advanced and easy-to-use web-server for bioinformatic analysis of subfamily-specific and conserved positions in diverse protein superfamilies. *Nucleic Acids Res*. 2020:48(W1):W65–W71. https://doi.org/10.1093/nar/gkaa276.

Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev*. 2007:71(4):576–599. https://doi.org/10.1128/MMBR.00015-07.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 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*. 2008:59(7):1515–1524. https://doi.org/10.1093/jxb/erm361.

Tokuriki N, Tawfik DS. Protein dynamism and evolvability. *Science*. 2009:324(5924):203–207. https://doi.org/10.1126/science.1169375.

Valegård K, Andralojc PJ, Haslam RP, Pearce FG, Eriksen GK, Madgwick PJ, Kristoffersen AK, van Lun M, Klein U, Eilertsen HC, *et al*. Structural and functional analyses of rubisco from Arctic diatom species reveal unusual posttranslational modifications. *J Biol Chem*. 2018:293(34):13033–13043. https://doi.org/10.1074/jbc.RA118.003518.

Van Lun M, Hub JS, Van Der Spoel D, Andersson I. CO2 and O2 distribution in rubisco suggests the small subunit functions as a CO2 reservoir. *J Am Chem Soc*. 2014:136(8):3165–3171. https://doi.org/10.1021/ja411579b.

Van Lun M, Van Der Spoel D, Andersson I. Subunit interface dynamics in hexadecameric rubisco. *J Mol Biol*. 2011:411(5):1083–1098. https://doi.org/10.1016/j.jmb.2011.06.052.

Wagner A, Rosen W. Spaces of the possible: universal darwinism and the wall between technological and biological innovation. *J R Soc Interface*. 2014:11(97):20131190. https://doi.org/10.1098/rsif.2013.1190.

Wang H, Yan X, Aigner H, Bracher A, Nguyen ND, Hee WY, Long BM, Price GD, Hartl FU, Hayer-Hartl M. Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature*. 2019:566(7742):131–135. https://doi.org/10.1038/s41586-019-0880-5.

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004:25(9):1157–1174. https://doi.org/10.1002/jcc.20035.

Wang M, Jiang YY, Kim KM, Qu G, Ji HF, Mittenthal JE, Zhang HY, Caetano-Anollés G. A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol Biol Evol*. 2011:28(1):567–582. https://doi.org/10.1093/molbev/msq232.

Ward LM, Shih PM. The evolution and productivity of carbon fixation pathways in response to changes in oxygen concentration over geological time. *Free Radic Biol Med*. 2019:140:188–199. https://doi.org/10.1016/j.freeradbiomed.2019.01.049.

West-Roberts JA, Matheus-Carnevali PB, Schoelmerich MC, Al-Shayeb B, Thomas AD, Sharrar A, He C, Chen L-X, Lavy A, Keren R, *et al*. The Chloroflexi supergroup is metabolically diverse and representatives have novel genes for non-photosynthesis based CO2 fixation. *bioRxiv*. https://doi.org/10.1101/2021.08.23.457424, 2021, preprint: not peer reviewed.

Woods RJ, Chappelle R. Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *J Mol Struct*. 2000:527(1-3):149–156. https://doi.org/10.1016/S0166-1280(00)00487-5.

Yang Z. PAML 4: phylogenetic analysis by Maximum likelihood. *Mol Biol Evol*. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Young JN, Rickaby REM, Kapralov MV, Filatov DA. Adaptive signals in algal rubisco reveal a history of ancient atmospheric carbon dioxide. *Philos Trans R Soc B Biol Sci*. 2012:367(1588):483–492. https://doi.org/10.1098/rstb.2011.0145.

Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol Biol Evol*. 2015:32(1):132–143. https://doi.org/10.1093/molbev/msu281.