RESEARCH ARTICLE

ENZYME EVOLUTION

Evolution of increased complexity and specificity at the dawn of form I Rubiscos

Luca Schulz[1], Zhijun Guo[2], Jan Zarzycki[1], Wieland Steinchen[3,4], Jan M. Schuller[3,4], Thomas Heimerl[3,5], Simone Prinz[6], Oliver Mueller-Cajar[2], Tobias J. Erb[1,3]*, Georg K. A. Hochberg[3,4,7]*

The evolution of ribulose-1,5-bisphosphate carboxylase/oxygenases (Rubiscos) that discriminate strongly between their substrate carbon dioxide and the undesired side substrate dioxygen was an important event for photosynthetic organisms adapting to an oxygenated environment. We use ancestral sequence reconstruction to recapitulate this event. We show that Rubisco increased its specificity and carboxylation efficiency through the gain of an accessory subunit before atmospheric oxygen was present. Using structural and biochemical approaches, we retrace how this subunit was gained and became essential. Our work illuminates the emergence of an adaptation to rising ambient oxygen levels, provides a template for investigating the function of interactions that have remained elusive because of their essentiality, and sheds light on the determinants of specificity in Rubisco.

Ribulose-1,5-bisphosphate carboxylase/oxygenases (Rubiscos) that are paired with oxygenic photosynthesis are responsible for most inorganic carbon assimilation on Earth today (*1*, *2*). Rubisco ancestrally evolved in anaerobic environments, predating the emergence of oxygenic photosynthesis (*3*, *4*). As oxygenic photosynthesis evolved, Rubisco faced molecular oxygen, which acts as an undesired side substrate during catalysis. Reaction with O_2 produces 2-phosphoglycolate (2PG), a metabolite that inhibits carbon metabolism and causes a loss of carbon from metabolism (*5*, *6*). To reduce the buildup of 2PG and enable its conversion, several mitigation strategies evolved: recycling of 2PG through photorespiration (*6*, *7*), carbon-concentrating mechanisms that concentrate CO_2 around Rubisco (*8*), and the use of Rubiscos with higher specificity for CO_2 (*9*–*11*). Although the exact strategies used for photorespiration and carbon concentration vary, all aerobic phototrophs, in particular algae and plants, use high-specificity Rubiscos (*9*, *11*). Their evolution was an important ingredient in the rise of oxygenic photosynthesis. Yet, how and when Rubisco evolved high specificity remain unknown despite extensive research and speculation (*4*, *12*–*14*).

[1]Department of Biochemistry and Synthetic Metabolism, Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany. [2]School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore. [3]Center for Synthetic Microbiology (SYNMIKRO), Philipps University Marburg, 35043 Marburg, Germany. [4]Department of Chemistry, Philipps University Marburg, 35043 Marburg, Germany. [5]Department of Biology, Philipps University Marburg, 35043 Marburg, Germany. [6]Central Electron Microscopy Facility, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [7]Evolutionary Biochemistry Group, Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany.
*Corresponding author. Email: toerb@mpi-marburg.mpg.de (T.J.E.); georg.hochberg@mpi-marburg.mpg.de (G.K.A.H.)

The defining structural feature of all known high-specificity form I Rubiscos is their assembly into a complex of eight catalytic large subunits (LSUs) and eight noncatalytic small subunits (SSUs) (total stoichiometry: L8S8) (*1*). This stoichiometry evolved from simpler ancestors that did not interact with SSUs, as evidenced by modern-day SSU-independent form I′, II, and III Rubiscos (*3*, *13*). Because the SSU was previously shown to influence Rubisco catalysis (*15*–*20*) and because it is the most obvious structural difference between form I and other Rubiscos, it is thought to be at least partly responsible for increased specificity toward CO_2 (*10*, *17*). However, it has so far not been possible to directly test this hypothesis because the SSU is essential for both catalytic activity and solubility of form I Rubiscos (*15*, *21*, *22*). Studies on its effect have therefore been limited to in silico calculations (*23*), homolog shuffling experiments (*16*–*19*), and small-scale mutational perturbations (*17*, *20*), which have not been conclusive and have shown relatively small, mostly deleterious effects on specificity.

In this work, we overcome this challenge using ancestral sequence reconstruction (*24*) to recapitulate the evolution of form I Rubiscos. We identify the genetic and structural causes for the recruitment of the SSU, show that the SSU was indirectly responsible for an increase in specificity, and reveal how the solubility of form I Rubiscos became dependent on the SSU. Notably, our experiments suggest that form I Rubiscos increased their specificity for CO_2 over O_2 even before oxygen was abundant (*25*–*28*).

SSU gain coincided with rising specificity

To determine when Rubisco gained the SSU, we inferred maximum-likelihood phylogenies of Rubisco's LSU and SSU (Fig. 1A and figs. S1 and S2). On the LSU phylogeny, three clades of orthologs branch on the stem lineage toward form I Rubiscos (form I-α, form I′, and form I″) (*13*, *29*). These orthologs were identified from metagenome-assembled genomes (MAGs), which do not encode SSUs (fig. S3A). We also identified a fourth clade of true form I Rubiscos from MAGs that do encode SSUs (form I Anaero). This clade branches close to the last common ancestor (LCA) of known cyanobacterial and plant form I AB and proteobacterial and red algal form I CD Rubiscos (see the supplementary text).

These Rubisco-encoding MAGs were sampled in hot environments and belong to organisms related to anaerobic, thermophilic *Chloroflexaeota* (*30*) and *Firmicutes* (fig. S3, A, B, and C). For further characterization, we purified Rubiscos of clades branching close to the SSU's appearance and measured their oligomerization state using mass photometry (MP). Form I-α Rubiscos assembled into dimers, form I′ Rubiscos into octamers (*13*), and form I Anaero variants into L8S8 hexadecamers (Fig. 1B and fig. S3D). Furthermore, form I Anaero Rubiscos exhibited high thermal stability (>80°C), a high catalysis temperature optimum of 55° to 70°C (k_{catC} ~ 5 s^{-1}), and low specificity for CO_2 over O_2 for form I Rubiscos (specificity < 27; fig. S3, E, F, and G, and Table 1).

Together, the stoichiometry measurements and the absence of SSU genes in the MAGs of form I″–encoding organisms suggest that Rubisco began to interact with the SSU after branching of the form I″ clade. Furthermore, the MAG's origin in hot, anaerobic environments and the thermophilicity of form I Anaero Rubiscos suggest that the SSU first evolved in thermophilic anaerobes that existed before crown cyanobacteria diverged and thus before atmospheric oxygen levels rose substantially during the Great Oxygenation Event (GOE) (*25*–*28*). We cannot rule out that yet-unsampled organisms carry forms of Rubisco that conflict with these inferences. However, the phylogenetic interval in question is bracketed by three separate Rubisco clades with similar characteristics that derive from multiple sampling sites and sequencing projects. An anaerobic, thermophilic origin of form I Rubiscos is thus the most parsimonious scenario based on current data.

We next investigated how the evolution of Rubisco's L8S8 assembly influenced its function. To do this, we inferred sequences of ancestral Rubisco LSUs that existed before (AncL) and after (AncLS) the gain of the SSU (Fig. 1A and fig. S4, A and B). We also resurrected an ancestral small subunit (AncSSU). Because the SSU phylogeny has no outgroup, we reconstructed a deep node within the form I Anaero SSUs as a surrogate that existed close to the presumed root. MP of purified proteins showed that AncL formed a homo-octamer and that AncLS formed an L8S8 heterocomplex with

AncSSU (Fig. 1C). Both complexes were able to bind the active site inhibitor carboxyarabinitol-1,5-bisphosphate (CABP) and were stable to >75°C (fig. S4, C and D).

We quantified AncL and AncLS+AncSSU's carboxylation with ribulose-1,5-bisphosphate (RuBP) as well as their specificities for CO_2 over the side-reaction substrate O_2 ($S_{C/O}$) at 25°C. From AncL to AncLS+AncSSU, the $K_m(CO_2)$ (the Michaelis constant for CO_2) decreased from 568 to 69 μM, the K_m(RuBP) increased from 92 to 544 μM, and the maximal rate of carboxylation decreased from 0.61 to 0.29 (Table 1). Overall, this resulted in an increased catalytic efficiency for the carboxylation reaction (1.1×10^3 to 4.2×10^3) as well as a faster carboxylation at physiologically relevant dissolved CO_2 concentrations <500 μM and saturating RuBP concentrations in AncLS (fig. S4E). Lastly, the $S_{C/O}$ increased from 21.9 to 47.3, which is comparable to the change in specificity from a form II to a cyanobacterial form I Rubisco (Table 1). Thus, the carboxylation efficiency and specificity of Rubisco improved when the canonical form I L8S8 assembly first evolved. Subsequently, extant form I Anaero Rubiscos appear to have reverted to low specificities ($S_{C/O}$ values between 12 and 26), consistent with their anaerobic habitat. This suggests that binding of the SSU alone is not the sole determinant of specificity.

The SSU quickly became essential

Our next aim was to determine to what extent the SSU was responsible for the functional changes that we observed. We sought to identify a Rubisco that can bind the SSU but does not yet depend on it for solubility and activity. We first tested whether AncL was already able to bind AncSSU. We produced AncL with untagged AncSSU and vice versa. In either case, the untagged protein did not coelute (fig. S4,

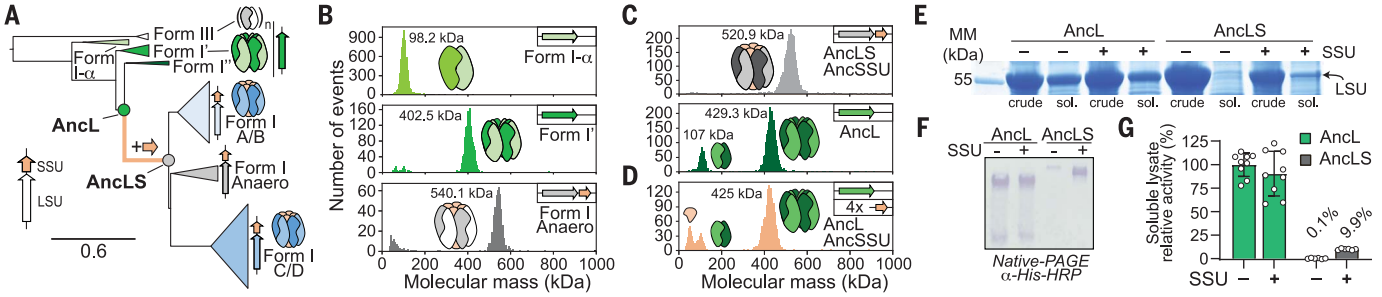

Fig. 1. The evolution of form I Rubiscos. (**A**) Reduced phylogeny of form I and related Rubiscos. Arrows depict operon structure. Complete phylogeny is shown in fig. S2. (**B**) MP spectra of metagenomic Rubiscos. Color scheme is according to (A). Inferred stoichiometries are depicted as cartoons. Representative spectra of two technical replicates. (**C**) MP spectra of ancestral Rubiscos AncL and AncLS+AncSSU. Representative spectra of more than five technical replicates. (**D**) MP spectrum of AncL after incubation with a fourfold molar excess of purified AncSSU. (**E**) Coomassie-stained SDS-PAGE of crude and soluble lysates of AncL and AncLS produced with or without AncSSU coproduction. MM, molecular mass; sol., soluble. Representative gel of more than five biological replicates. See fig. S5A for full gel. (**F**) Native PAGE and Western blot analyses of lysates from cultures producing AncL or AncLS with or without AncSSU coproduction. Loading control is shown in fig. S5B. HRP, horseradish peroxidase. (**G**) Relative Rubisco activities in soluble lysates of cultures producing AncL and AncLS with and without coproduction of AncSSU. Data are relative to measurements of AncL without AncSSU coproduction. For SDS-PAGE analysis of biological triplicates, see fig. S5C. $N = 9$ technical replicates for AncL, and $N = 6$ for AncLS. Error bars depict standard deviations (SDs).

Table 1. Kinetic characterization of ancestral and related extant Rubisco at 25°C. AncL+7 derives from AncL and contains seven substitutions at the LSU-SSU interface to enable SSU binding (Fig. 2). k_{catC} is the maximal rate of carboxylation under saturating substrate concentrations. $K_m(CO_2)$ and K_m(RuBP) are the Michaelis constants for CO_2 and RuBP, respectively. $S_{C/O} = [k_{catC}/K_m(CO_2)]/[k_{catO}/K_m(O_2)]$. *R. rubrum*, *Rhodospirillum rubrum*; *A. ferrooxidans*, *Acidithiobacillus ferrooxidans*; *Syn.*, *Synechococcus*; PCC, Pasteur Culture Collection. Values are means ± 95% confidence intervals with the number of technical replicates (*N*) indicated in parentheses. Extant form I Rubiscos were produced and measured with their native SSUs. Kinetic curves are shown in fig. S17. The $K_m(CO_2)$ measurement for RME08239 Rubisco was performed at 1 mM RuBP as opposed to 2.2 mM. n.d., not determined.

Rubisco name or identifier	k_{catC} (s^{-1})	$K_m(CO_2)$ (μM)	K_m(RuBP) (μM)	Specificity ($S_{C/O}$)	$k_{catC}/K_m(CO_2)$ (M^{-1} s^{-1})	$k_{catO}/K_m(O_2)$ (M^{-1} s^{-1})
AncL	0.61 ± 0.10 (2)	568 ± 177 (2)	92 (1)	21.9 ± 1.1 (9)	1.1×10^3	0.5×10^2
AncL+7	0.63 ± 0.11 (2)	499 ± 167 (2)	58 ± 18 (3)	25.3 ± 1.9 (5)	1.3×10^3	0.5×10^2
AncL+7 + 4x AncSSU	0.41 ± 0.02 (2)	148 ± 21 (2)	331 ± 76 (3)	29.6 ± 1.7 (4)	2.8×10^3	0.9×10^2
AncL+7 s437W	0.61 ± 0.11 (2)	460 ± 158 (2)	63 (1)	25.6 ± 1.4 (9)	1.3×10^3	0.5×10^2
AncL+7 s437W + 2x AncSSU	0.49 ± 0.02 (2)	162 ± 24 (2)	483 (1)	25.8 ± 2.2 (5)	3.0×10^3	1.2×10^2
AncL+7 e170N	0.55 ± 0.16 (2)	569 ± 300 (2)	28 (1)	25.6 ± 2.3 (6)	1.0×10^3	0.4×10^2
AncL+7 e170N + 4x AncSSU	0.47 ± 0.03 (4)	113 ± 29 (4)	203 (1)	38.7 ± 2.9 (6)	4.2×10^3	1.1×10^2
AncL+7 e170N s437W + 4x AncSSU	0.49 ± 0.02 (2)	119 ± 17 (2)	n.d.	40.7 ± 2.8 (4)	4.1×10^3	1.0×10^2
AncLS + AncSSU	0.29 ± 0.02 (2)	69 ± 16 (2)	544 (1)	47.3 ± 2.1 (9)	4.2×10^3	0.9×10^2
RME08239 + SSU (form I Anaero)	1.1 ± 0.12 (2)	200 ± 47 (2)	n.d.	12.2 ± 2.5 (2)	5.5×10^3	4.5×10^2
RMG64267 + SSU (form I Anaero)	1.94 ± 0.28 (2)	88 ± 53 (2)	67 (1)	26.4 (1)	2.2×10^4	8.4×10^2
R. rubrum (form II)	6.31 ± 0.56 (2)	116 ± 38 (2)	9 (1)	12.6 ± 0.5 (15)	5.4×10^4	4.3×10^3
A. ferrooxidans (form II)	n.d.	n.d.	n.d.	17.9 ± 0.5 (6)	n.d.	n.d.
Syn. PCC 6301 + SSU (form I)	n.d.	n.d.	n.d.	45.7 ± 1.7 (9)	n.d.	n.d.

F and G). Additionally, we purified AncL and AncSSU separately and used MP to probe the interaction, but we could not detect binding of AncSSU to AncL (Fig. 1D).

In parallel, we tested whether AncLS could still function without the SSU. We produced both ancestors with and without coproduction of AncSSU and quantified solubility using SDS–polyacrylamide gel electrophoresis (SDS-PAGE), Western blots, and Rubisco activities in clarified lysates. The yield of soluble AncLS was markedly reduced without AncSSU, whereas AncL titers were unchanged (Fig. 1, E, F, and G). The interaction with the SSU and a total dependence on it thus evolved in quick succession between AncL and AncLS. Interactions, dependencies, and catalytic properties of AncL and AncLS were qualitatively unchanged when validating their robustness to statistical uncertainty (*31*). Moreover, inferences of AncLS's stoichiometry and SSU dependence were unchanged when AncLS was reconstructed using an alternative phylogeny (*29*) (figs. S6, S7, and S8 and supplementary text).

Genetic basis of LSU-SSU interaction

Because AncL cannot yet interact with the SSU, and AncLS already cannot function without it, historical substitutions that separate the two constructs must be responsible for both creating the LSU-SSU interaction and making it essential. We reasoned that we might be able to construct a nonobligate LSU-SSU interaction by introducing only those substitutions into AncL that create the interaction.

To pinpoint relevant substitutions, we solved x-ray crystal structures of inhibitor-bound AncL and AncLS+AncSSU to 2.1- and 1.8-Å resolution, respectively (Fig. 2, A and B). AncL and AncLS differ by 95 substitutions, 14 of which are close to the SSU in the AncLS structure, and three insertions (Fig. 2, C and D, and fig. S4B). The 14 substituted sites make up only 22% of all interface sites and 25% of the total buried interface area in AncLS.

Next, we created AncL constructs containing single, double, or triple combinations of the 14 substitutions in proximity to the SSU (8, 28, and 10 variants, respectively). At least three substitutions were required to create a MP-detectable LSU-SSU interaction (fig. S9), although this interaction remained weak. To achieve tighter binding, we simultaneously introduced seven interface substitutions into AncL (Fig. 2D). This variant (AncL+7) was soluble without AncSSU and rapidly reconstituted an L8S8 heterocomplex upon mixing with purified AncSSU (Fig. 2E).

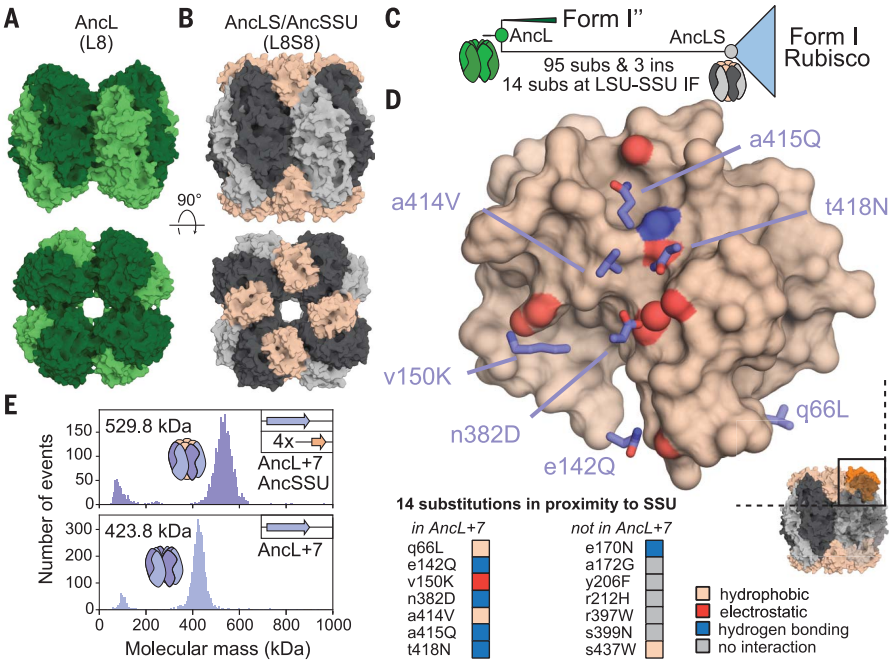

Fig. 2. A small set of substitutions can create the LSU-SSU interface. (**A**) Structure of CABP-bound AncL solved to 2.1 Å. (**B**) Structure of CABP-bound AncLS cocrystallized with AncSSU solved to 1.8 Å. LSU is shown in light and dark gray, and AncSSU is shown in beige. (**C**) Schematic phylogeny highlighting the differences between AncL and AncLS. subs, substitutions; ins, insertions; IF, interface. (**D**) Substitutions separating AncL from AncLS that occurred in proximity to the SSU. Small letters denote ancestral states, and capitalized letters denote derived states. Purple side chains depict substitutions that are transplanted into AncL+7. The SSU surface is depicted in beige. Heteroatoms within 3.8 Å of the substitutions are highlighted in red (oxygen) and blue (nitrogen). Water molecule is shown as a red sphere. (**E**) MP spectra of AncL+7 in isolation (bottom) or with a fourfold molar excess of purified AncSSU (top). Empirical masses and inferred stoichiometries are depicted in the histogram. Theoretical L8 mass: 428.0 kDa; theoretical L8S8 mass: 527.6 kDa. Data are from three technical replicates. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

SSU improved catalysis and specificity

We were now able to directly test how the interaction with the SSU affected Rubisco. To do this, we measured AncL+7's kinetic parameters in the presence or absence of AncSSU (the L8 and L8S8 forms of AncL+7, respectively). The carboxylation rate of AncL+7 decreased slightly when interacting with the SSU (k_{catC} values of 0.63 and 0.41, respectively), and the K_m(RuBP) increased sixfold. Notably, however, the $K_m(CO_2)$ was decreased threefold, which, in total, leads to a 2.2-fold improvement in carboxylation efficiency. Carboxylation is up to twofold faster at physiologically relevant dissolved CO_2 concentrations <500 μM in the L8S8 form at saturating RuBP concentrations (Fig. 3, A and B, and Table 1). We verified robustness of these inferences with an alternative SSU reconstruction, corresponding to a deeper but worse-reconstructed node on our SSU phylogeny (fig. S10).

Besides increasing carboxylation efficiency, the addition of AncSSU to AncL+7 markedly increased the temperature tolerance of catalysis from ~40°C up to 50° to 70°C, which results in an up to 12.5-fold improved carboxylation rate in the L8S8 form at 70°C (Fig. 3C). Yet, unexpectedly, AncSSU did not increase overall stability of AncL+7 toward denaturation: We found no difference in stability between the L8 and L8S8 forms upon exposure to elevated temperatures (Fig. 3D).

Next, we quantified how the SSU affected CO_2/O_2 specificity. The seven substitutions introduced at the LSU-SSU interface to create AncL+7 already caused a statistically significant increase in specificity from 21.9 to 25.3 (P = 0.0013, two-tailed t test; Fig. 3E and Table 1). This increase was amplified to 29.6 when the SSU was added. We could further improve specificity by introducing one more substitution at the LSU-SSU interface. Introducing e170N (where small letters denote ancestral states and capitalized letters denote derived states) into AncL+7 boosted specificity to 38.7 in the presence of AncSSU but did not improve AncL+7's specificity without AncSSU (Fig. 3E and Table 1). Together, eight substitutions and the presence of AncSSU almost doubled the specificity of AncL from 21.9 to 38.7, whereas no changes in or close to the active site were required. The SSU is thus directly and indirectly responsible for the key functional improvements associated with form I Rubiscos: The interaction directly improves

catalytic efficiency and catalysis at high temperatures. Its effect on specificity is (at least partly) indirect because it epistatically enhances the effect of substitution e170N, which has no influence on specificity without the SSU.

To elucidate the structural basis of the SSUs' effects on catalysis, we solved x-ray crystal structures of inhibitor-bound AncL+7 as well as AncL+7 e170N in their L8 and L8S8 forms (Fig. 3F and fig. S11A). In both complexes, individual monomers are virtually identical between the L8 and L8S8 forms [average root mean square deviation (RMSD) < 0.3 Å; fig. S11B]. The only observable structural difference is that the LSU octamer is slightly more compact in the L8S8 forms, but this does not translate into observable active site rearrangements (Fig. 3G and fig. S11, C and D). The high specificity–conferring e170N substitution at the LSU-SSU interface is far away from the active site, where it engages in a hydrogen-bonding interaction with Y83 and Q88 on the SSU (fig. S11E). The SSU thus improves catalysis through allosteric effects on either the open, non–inhibitor-bound state or protein dynamics.

A single substitution causes SSU dependence

All characterized form I Rubiscos depend on the SSU for solubility or catalysis (*15*, *21*, *22*), and our results indicate that this trait had already evolved by the time of AncLS. We therefore sought to discover how this dependence evolved. Drawing on prior work on how

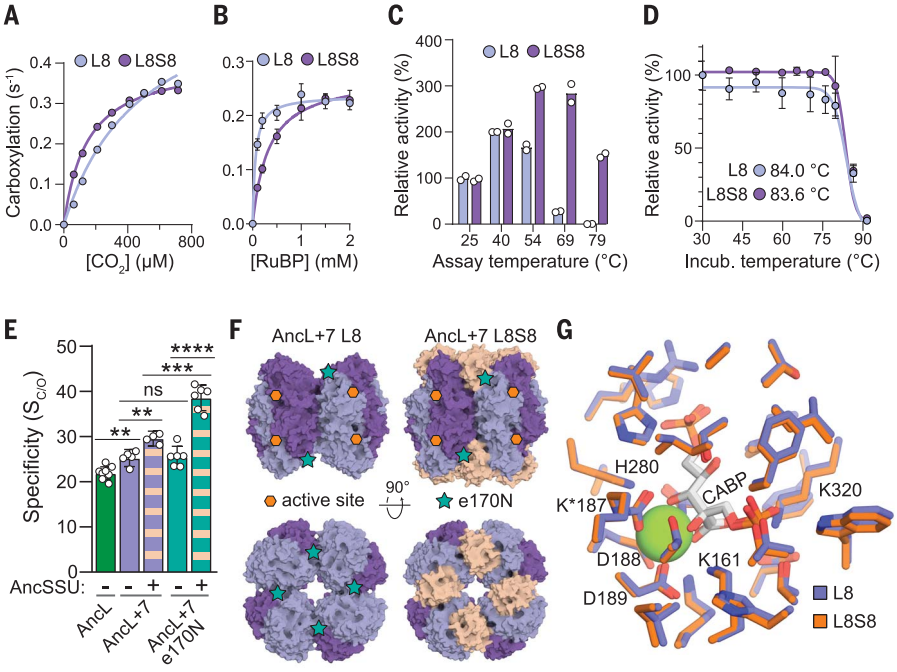

Fig. 3. The SSU allosterically enhances catalysis. (**A** and **B**) Kinetic characterization of AncL+7 with or without AncSSU (L8 and L8S8 forms, respectively). Data are the means of two (A) or three (B) technical replicates. Error bars indicate SDs. (**C**) Carboxylation rates of the L8 and L8S8 forms of AncL+7 at varying temperatures. Bars represent the means of two technical replicates. Activity is relative to the rate of catalysis at 25°C in L8 form. (**D**) Remaining activities of the L8 and L8S8 forms of AncL+7 after incubation at elevated temperatures, shown relative to incubation (Incub.) at 30°C. Curve is a sigmoidal fit (N = 3 technical replicates; error bars indicate SDs). (**E**) Specificity of AncL variants with and without AncSSU. Significance tested by two-tailed t test. ns, nonsignificant; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. N values are listed in Table 1, and error bars represent 95% confidence intervals. (**F**) Structures of CABP-bound AncL+7 crystallized as a homomer (left) or in complex with AncSSU (right) solved at 2.65 and 2.0 Å, respectively. LSU monomers are shown in light and dark purple, and AncSSU is shown in beige. (**G**) Active site residues of the L8 (purple) and L8S8 forms (orange) of AncL+7 with coordinated magnesium (green spheres) and bound CABPs (L8, light gray; L8S8, dark gray). K*187 marks the carbamylated active site lysine.

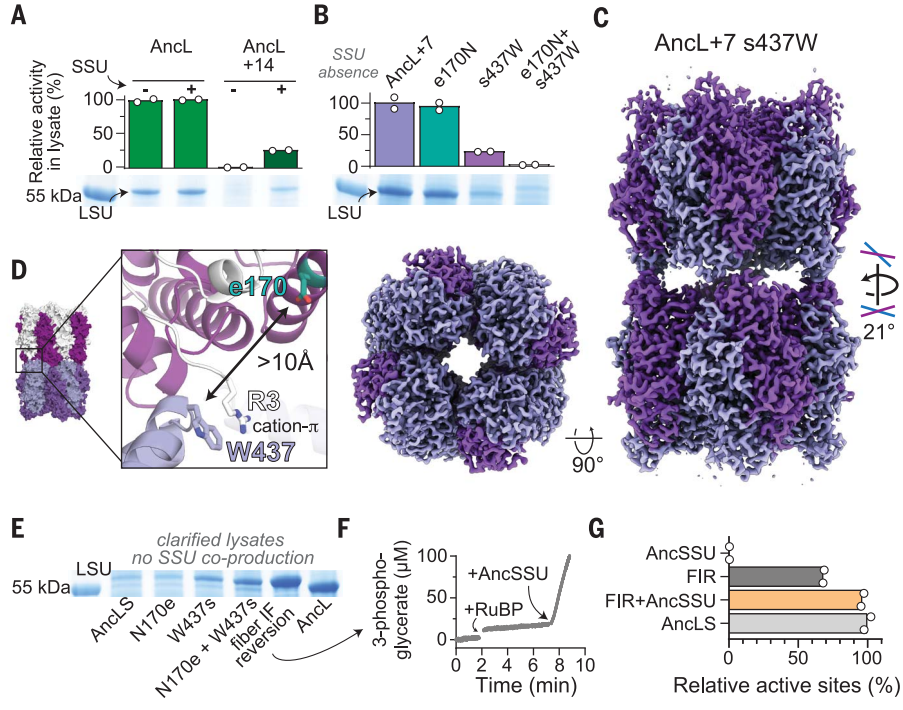

Fig. 4. Rubisco evolves to depend on the SSU for solubility and catalysis. (**A**) Relative Rubisco activities and SDS-PAGE analysis of soluble lysates from strains producing AncL and AncL+14 with or without AncSSU. LSU band is indicated. Full gel is shown in fig. S16A. N = 2 technical replicates. (**B**) Relative Rubisco activities and SDS-PAGE analysis of soluble lysates from strains producing AncL+7 and derivates thereof. Full gel is shown in fig. S16B, and crude lysate pictures are shown in fig. S16C. N = 2 technical replicates. (**C**) A 3.01-Å resolution cryo-EM density map of AncL+7 s437W fibrils. (**D**) Structural interactions across the fiber interface of sites that contain entrenching states. (**E**) Coomassie-stained SDS-PAGE of clarified lysates from cells producing AncLS and variants thereof. AncSSU was not coproduced. Genotype of the fiber IF reversion construct is listed in fig. S11H. Full gel is shown in fig. S16B. (**F**) 3-Phosphoglycerate production of fiber interface reversion is SSU-dependent, even though solubility is recovered. Arrows mark addition of RuBP and AncSSU (10-fold excess). (**G**) Relative active site contents of Rubisco LSU (0.2 nmol) and SSU (2 nmol) samples, as determined by ^{14}C-CABP binding assays. FIR, fiber interface reversion construct. Bars represent the means of two technical replicates.

interactions can become essential (*32*, *33*), we hypothesized that entrenching substitutions occurred near the SSU interface. To test this idea, we simultaneously transplanted all seven additional SSU-neighboring substitutions from AncLS into AncL+7. The resulting AncL+14 was only soluble when AncSSU was coproduced and formed an L8S8 heterocomplex (Fig. 4A and fig. S12A).

To pinpoint which changes cause SSU dependence, we introduced all 14 substitutions individually into AncL and tested their effect on solubility (fig. S12B and fig. S13). Substitution s437W reduced solubility of both AncL and AncL+7 by ~75%, as assessed by SDS-PAGE and relative Rubisco activities in clarified lysate (Fig. 4B). We examined the remaining six substitutions separating AncL+7 s437W from AncL+14 for their effect on solubility (fig. S12, C and D). Substitution e170N, which was also responsible for the specificity increase, had the most severe effect. When introduced into AncL+7 s437W, it caused a complete loss of solubility, which could be rescued by coproducing AncSSU. Notably, e170N alone had no negative effect on the solubility of AncL+7 (Fig. 4B). Thus, a single historical substitution (s437W) could make Rubisco become dependent on the SSU for solubility.

Introducing s437W had no measurable effect on most biochemical parameters of AncL+7 or on catalysis and specificity of AncL+7 e170N in the presence of AncSSU (fig. S14 and Table 1). Our results imply that Rubisco's dependence on the SSU for solubility is functionally neutral and is not required for the catalytic improvements associated with form I Rubiscos. This conclusion is further supported by the fact that most of the functional differences between AncL and AncLS are recapitulated in AncL+7 e170N+AncSSU, which does not depend on the SSU (Table 1).

Mechanisms of essentiality

To conclude our investigation, we sought to understand how these two substitutions make the LSU require the SSU for solubility. Rubiscos containing s437W formed higher-order oligomers in MP measurements and stacked into fibers in negative-stain electron microscopy images (fig. S12, E, F, and G). This led us to hypothesize that insolubility stems from LSU self-assembly into fibers (*34*), which can be prevented by the SSU.

To understand the molecular basis of fiber formation, we solved the structure of AncL+7 s437W fibers to 3.01 Å using cryo–electron microscopy (cryo-EM) (Fig. 4C and fig. S15). In fibers, LSU octamers interact at their apices, forming a helix with a twist of ~21° per octamer. The fiber-inducing W437 directly engages in a cation-π interaction with R3 of an adjacent LSU octamer at the fiber interface. N170, which exacerbates the effect of W437, seems to have a less direct mechanism of action, as it is >10 and >25 Å away from either of the fiber interfaces (Fig. 4D). The fiber interface partially overlaps with the LSU-SSU interface, which makes the cavity between interacting octamers too small to accommodate SSUs. Capping of LSU octamers by SSUs therefore makes fiber formation impossible (fig. S12, H and I).

We tested whether substitutions e170N and s437W remain the sole causes of insolubility in AncLS or whether additional entrenching substitutions accumulated at the SSU interface. Neither N170e nor W437s reversions, alone or as a pair, fully restored solubility without the SSU (Fig. 4E). Simultaneously reverting all 14 substitutions that occurred within 5 Å of the fiber interface fully recovered solubility in the absence of the SSU (Fig. 4E and fig. S12H). Hence, the region around the fiber interface remains the cause of insolubility, and additional entrenching substitutions beyond e170N and s437W accumulated after the gain of the SSU.

In the absence of SSUs, this so-called fiber interface reversion construct assembled into a conventional L8 complex but was only partially able to bind the active site inhibitor CABP and was barely active (~68% active site content; $k_{catC} < 0.05\ s^{-1}$). Addition of AncSSU quickly reconstituted an L8S8 complex, increased active site content to levels comparable to those of AncLS, and restored catalytic activity (Fig. 4, F and G, and fig. S12J). This indicated that even though we reverted the dependence on SSU for solubility, some of the remaining 81 substitutions separating AncL and the fiber interface reversion construct made Rubisco depend on the SSU for catalysis. This catalytic dependence is also present in the only form I Rubisco that is soluble without the SSU (*15*).

Conclusions

We show that the SSU modified the functional sequence space available to Rubisco. Consequently, certain substitutions became advantageous in the presence of the SSU that otherwise would not have been. However, it also enabled the accumulation of substitutions that created a dependence on the interaction. Rubisco is a slow-evolving (*35*), notoriously constrained enzyme (*36*, *37*), and the recruitment of an epistatic modifier like the SSU may have been the only way for it to access high specificity.

All features of modern form I Rubiscos—interaction with and dependence on the SSU, as well as high specificity—had already evolved by the time of AncLS. Notably, this timeline implies that high-specificity Rubiscos predate the GOE and may even predate the evolution of photosystem II, although the exact timing of this event is still contested (*25*–*28*, *38*). One possibility is that specificity first improved as a by-product of selection for decreasing $K_m(CO_2)$ alongside a drop in atmospheric CO_2 concentrations before the GOE (*39*). This would have fortuitously prepared Rubisco for the subsequent rise of atmospheric O_2 concentrations. However, there is also tentative evidence that O_2 (and, by implication, oxygenic photosynthesis) may have been present before the GOE (*28*, *38*, *40*–*42*), which could have directly selected for improved specificity. In either case, the evolution of the SSU facilitated Rubisco's coexistence with oxygenic photosynthesis, which helped set the stage for complex life in aerobic environments.

REFERENCES AND NOTES

1. I. Andersson, A. Backlund, *Plant Physiol. Biochem.* **46**, 275–291 (2008).
2. Y. M. Bar-On, R. Milo, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4738–4743 (2019).
3. F. R. Tabita, S. Satagopan, T. E. Hanson, N. E. Kreel, S. S. Scott, *J. Exp. Bot.* **59**, 1515–1524 (2008).
4. T. J. Erb, J. Zarzycki, *Curr. Opin. Biotechnol.* **49**, 100–107 (2018).
5. W. W. Cleland, T. J. Andrews, S. Gutteridge, F. C. Hartman, G. H. Lorimer, *Chem. Rev.* **98**, 549–562 (1998).
6. F. Flügel *et al.*, *Plant Cell* **29**, 2537–2551 (2017).
7. H. Bauwe, M. Hagemann, A. R. Fernie, *Trends Plant Sci.* **15**, 330–336 (2010).
8. J. A. Raven, C. S. Cockell, C. L. De La Rocha, *Phil. Trans. R. Soc. B* **363**, 2641–2650 (2008).
9. A. I. Flamholz *et al.*, *Biochemistry* **58**, 3365–3376 (2019).
10. R. J. Spreitzer, *Arch. Biochem. Biophys.* **414**, 141–149 (2003).
11. C. Iñiguez *et al.*, *Plant J.* **101**, 897–918 (2020).
12. B. Kacar, V. Hanson-Smith, Z. R. Adam, N. Boekelheide, *Geobiology* **15**, 628–640 (2017).
13. D. M. Banda *et al.*, *Nat. Plants* **6**, 1158–1166 (2020).
14. P. M. Shih *et al.*, *Nat. Commun.* **7**, 10382 (2016).
15. T. J. Andrews, *J. Biol. Chem.* **263**, 12213–12219 (1988).
16. T. J. Andrews, D. M. Greenwood, D. Yellowlees, *Arch. Biochem. Biophys.* **234**, 313–317 (1984).
17. S. Karkehabadi *et al.*, *Biochemistry* **44**, 9851–9861 (2005).
18. E. Martin-Avila *et al.*, *Plant Cell* **32**, 2898–2916 (2020).
19. M. T. Lin, W. D. Stone, V. Chaudhari, M. R. Hanson, *Nat. Plants* **6**, 1289–1299 (2020).
20. B. A. Read, F. R. Tabita, *Biochemistry* **31**, 519–525 (1992).
21. C. Liu *et al.*, *Nature* **463**, 197–202 (2010).
22. J. Joshi, O. Mueller-Cajar, Y.-C. C. Tsai, F. U. Hartl, M. Hayer-Hartl, *J. Biol. Chem.* **290**, 1066–1074 (2015).
23. M. van Lun, J. S. Hub, D. van der Spoel, I. Andersson, *J. Am. Chem. Soc.* **136**, 3165–3171 (2014).
24. G. K. A. Hochberg, J. W. Thornton, *Annu. Rev. Biophys.* **46**, 247–269 (2017).
25. P. M. Shih, J. Hemp, L. M. Ward, N. J. Matzke, W. W. Fischer, *Geobiology* **15**, 19–29 (2017).
26. B. E. Schirrmeister, J. M. de Vos, A. Antonelli, H. C. Bagheri, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1791–1796 (2013).
27. B. E. Schirrmeister, M. Gugger, P. C. J. Donoghue, *Palaeontology* **58**, 769–785 (2015).
28. N. J. Planavsky *et al.*, *Nat. Geosci.* **7**, 283–286 (2014).
29. J. A. West-Roberts *et al.*, bioRxiv 2021.08.23.457424 [Preprint] (2021). https://doi.org/10.1101/2021.08.23.457424.
30. Y. Sekiguchi *et al.*, *Int. J. Syst. Evol. Microbiol.* **53**, 1843–1851 (2003).
31. G. N. Eick, J. T. Bridgham, D. P. Anderson, M. J. Harms, J. W. Thornton, *Mol. Biol. Evol.* **34**, 247–261 (2017).
32. G. K. A. Hochberg *et al.*, *Nature* **588**, 503–508 (2020).
33. A. Stoltzfus, *J. Mol. Evol.* **49**, 169–181 (1999).
34. H. Garcia-Seisdedos, C. Empereur-Mot, N. Elad, E. D. Levy, *Nature* **548**, 244–247 (2017).
35. J. W. Bouvier, D. M. Emms, S. Kelly, bioRxiv 2022.07.06.498985 [Preprint] (2022). https://doi.org/10.1101/2022.07.06.498985.
36. G. G. B. Tcherkez, G. D. Farquhar, T. J. Andrews, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7246–7251 (2006).
37. J. W. Bouvier *et al.*, *Mol. Biol. Evol.* **38**, 2880–2896 (2021).
38. T. Cardona, J. W. Murray, A. W. Rutherford, *Mol. Biol. Evol.* **32**, 1310–1328 (2015).
39. A. M. Hessler, D. R. Lowe, R. L. Jones, D. K. Bird, *Nature* **428**, 736–738 (2004).
40. A. M. Satkoski, N. J. Beukes, W. Li, B. L. Beard, C. M. Johnson, *Earth Planet. Sci. Lett.* **430**, 43–53 (2015).
41. J. Jabłońska, D. S. Tawfik, *Nat. Ecol. Evol.* **5**, 442–448 (2021).
42. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
43. L. Schulz, G. Hochberg, T. Erb, Evolution of increased complexity and specificity at the dawn of Form I Rubiscos,

version 1.0, Edmond (Max Planck Digital Library, 2022); https://doi.org/10.17617/3.F9BWOU.

ACKNOWLEDGMENTS

The authors thank the Central Electron Microscopy Facility at the Max Planck Institute of Biophysics for expertise and access to their instruments. We acknowledge support from the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and the EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). **Funding:** O.M.-C. is thankful for support from a Ministry of Education (MOE) of Singapore Tier 2 grant (MOE-T2EP30120-0005). J.M.S. is grateful for an Emmy Noether grant (SCHU 3364/1-1) from the Deutsche Forschungsgemeinschaft (DFG). L.S., J.Z., S.P., T.J.E., and G.K.A.H. are grateful for generous support from the Max Planck Society. L.S. thanks Refeyn Ltd. for a travel grant that funded the presentation of this work.

Author contributions: L.S., T.J.E., and G.K.A.H. conceived the project, analyzed data, and planned experiments. L.S. performed molecular work, phylogenetics, ancestral sequence reconstruction, protein purification, cryo-EM and x-ray analysis, comparative in vitro biochemistry, and MP measurements. Z.G. performed enzyme kinetic analysis and specificity constant measurements. O.M.-C. supervised enzyme kinetic analysis and specificity constant measurements. W.S. analyzed x-ray structures. J.Z. collected, solved, refined, and analyzed x-ray structures and refined the cryo-EM structure. T.H. performed negative-stain EM. S.P. performed cryo-EM grid preparation and collected cryo-EM datasets, and J.M.S. processed the cryo-EM datasets. O.M.-C., T.J.E., and G.K.A.H. supervised the project. L.S., T.J.E., and G.K.A.H. wrote the manuscript with contributions and comments from all authors. **Competing interests:** O.M.-C. has consulted for FL79 Inc. This company did not fund and had no role in this research project. The authors declare no other competing interests. **Data and materials availability:** All raw data for MP spectra and kinetic traces as well as phylogenetic trees, alignments, and ancestral sequences are deposited on Edmond, the Open Research Data Repository of the Max Planck Society for public access (*43*). Atomic structures reported in this paper are deposited to the Protein Data Bank under accession codes 7QSV, 7QSW, 7QSX, 7QSY, 7QSZ, 7QT1, and 7QVI. The cryo-EM data were deposited to the Electron Microscopy Data Bank under EMD-14178.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq1416
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S7
References (*44*–*73*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 21 March 2022; accepted 26 August 2022
10.1126/science.abq1416