Systematic exploration of prokaryotic form I rubisco maximal carboxylation rates

Benoit de Pins[1], Lior Greenspoon[1], Yinon M. Bar-On[1,†], Melina Shamshoum[1], Roee Ben-Nissan[1], Eliya Milshtein[1], Dan Davidi[2,‡], Itai Sharon[3], Oliver Mueller-Cajar[4], Elad Noor[1], Ron Milo[1,*].

[1] Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[2] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[3] Migal Galilee Research Institute, Kiryat Shmona 11016, Israel
[4] School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore

[†] Currently: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA
[‡] Currently: Aleph, Tel Aviv-Yafo 6688210, Israel
[*]Corresponding author: ron.milo@weizmann.ac.il

Abstract

Autotrophy is the basis for complex life on Earth. Its physiological limits are largely constrained by rubisco - the enzyme that catalyzes almost all carbon fixation on the planet. Yet, with only a small fraction of rubisco diversity kinetically characterized so far, the underlying biological factors driving the evolution of fast rubiscos in nature remain unclear. We conducted a high-throughput kinetic characterization of over 100 prokaryotic form I rubiscos, the most ubiquitous group of rubisco sequences in nature, to uncover the determinants of rubisco's carboxylation velocity. We show that phototrophy and the presence of a carboxysome CO_2 concentrating mechanism correlate with faster rubiscos (median rate difference of 3-5 fold). In contrast to previous assumptions, rubiscos originating from α-cyanobacteria exhibit the highest carboxylation rates among form I enzymes (≈10 s^{-1} median versus <7 s^{-1} in other groups). Our study ties biological and environmental insights associated with kinetic variation across rubiscos from nature.

Introduction

Biological carbon fixation is the keystone of food production and energy storage in the living world. Over 99% of global carbon fixation is catalyzed by rubisco[1], the most abundant enzyme in the biosphere[2]. Rubisco is mainly divided into four distinct forms (I, II, II/III, and III) and can be found in all domains of life, from plants to algae through autotrophic bacteria and archaea[3]. Within this diversity, form I is by far the most abundant of the four forms: it is used by all plants and cyanobacteria and is responsible for almost all CO_2 fixation in nature[4].

Paradoxically, while being the most abundant, rubisco is probably the slowest (i.e. low maximum carboxylation rate - $k_{cat,C}$) central metabolic enzyme[5,6]. In addition to its slow carboxylation rate, its active site can react with oxygen through a competitive oxygenase activity that ultimately leads to the release of 20-50% of the incorporated CO_2[7]. These properties are puzzling because of

rubisco's central role, which would lead us to expect a fast and efficient enzyme. Why has rubisco not evolved to be a more efficient enzyme? An extensive sampling of rubisco's genetic diversity is crucial to grasp the boundaries of its natural kinetic space.

Our group has recently developed a novel approach to systematically mine the kinetic space of natural rubiscos. We use computational methods to select representative rubiscos from the tremendous sequence diversity space. Gene synthesis is then used to generate expression constructs encoding many rubiscos for purification and kinetic characterization. In previous work we showed the feasibility of this approach by exploring form II and II/III rubisco variants[6]. Indeed, we found an uncharacterized rubisco that has a $k_{cat,C}$ two-fold higher than all previously-known rubiscos - demonstrating the potential of this approach to stretch the kinetic boundaries of this pivotal enzyme.

In this work, we expand our search to form I rubiscos, which represent 97% of known sequences[6]. Moreover, their immense ecological and sequence diversity[8–10], limited kinetic data, and higher carboxylation rates in comparison to plant rubiscos[5], make prokaryotic form I variants particularly interesting for exploring the evolution of fast carboxylating rubiscos. By conducting a first-of-its-kind large-scale study of uncharacterized prokaryotic from I rubiscos, and leveraging available meta-data on their sequences, we uncover the main biological and environmental determinants behind the evolution of fast-carboxylating rubiscos.

Results

Large scale survey of prokaryotic form I rubisco

To map the natural diversity of rubisco sequences, we performed an exhaustive search for rubisco homologs across the major genomic and metagenomic public databases. A total of 4300 unique sequences were identified as prokaryotic form I rubiscos (see Materials and Methods for more details).

By clustering at different sequence identity thresholds and across different rubisco subgroups (as detailed in the Materials and Methods and Supplementary Fig. 1), we selected 144 rubisco variants that represent the full spectrum of prokaryotic form I rubisco sequence diversity. Notably, prior work has so far characterized only 11 prokaryotic form I rubiscos using active site quantification, a method that allows for precise measurements of turnover rates while considering the enzyme's activation state (Fig. 1A).

In contrast to form II, II/III and III rubiscos which are composed of a ≈50 kDa large (L) subunit organized as a homodimer, form I rubiscos comprise an additional ≈15 kDa small subunit (S) in an L_8S_8 stoichiometry. Due to this complex oligomeric structure, recombinant expression of form I rubiscos is challenging[11]. To improve correct folding in *Escherichia coli*, we coexpressed rubiscos with the chaperone GroEL-GroES, which is known to help reconstitution of prokaryotic form I rubiscos[12–14]. In the case of variants originating from β-cyanobacteria, we coexpressed *rbcL* and *rbcS* together with their cognate chaperone *rbcX* whenever that gene was present in the operon. In addition, because these variants initially showed critical issues of solubility (with only a third of them soluble), we screened different homologs of rubisco accumulation factor 1 (Raf1),

a chaperone that mediates the assembly of β-cyanobacterial rubiscos[15–18]. We identified one homolog, from the bacteria *Euhalothece natronophila*, whose co-expression nearly doubled the number of solubly-expressed β-cyanobacterial rubiscos (Supplementary Fig. 2).

The carboxylation rates of each expressed rubisco were determined using a modified version of the spectroscopic coupled assay reported at Davidi *et al.*[6]. Here we directly assayed the crude cell lysates without purifying the enzyme in order to best preserve the quaternary structure of form I rubiscos. The method allows determining the specific carboxylation rate even without purification thanks to the use of CABP (see Methods and Supplementary Note 1). Since the assay is likely saturating the enzyme for CO_2, measured rates should approximate $k_{cat,C}$ values as suggested by the good match between the rates of five rubisco variants measured in our lab and their previously published values (Supplementary Fig. 3).

Out of 144 rubisco variants tested, 112 were successfully expressed and soluble. Of which, 98 exhibited significant catalytic activity (which we define as >0.5 reactions per active site per second). The median rate among active form I rubiscos was 5.4 s^{-1}, similar to the median $k_{cat,C}$ of plant rubisco literature values (4.7 s^{-1} when corrected to 30°C by assuming a Q10 value of 2.2[19]), with the fastest rate being 14 s^{-1} (Fig. 1B).

Altogether, our measurements complete a nearly tenfold increase in the number of kinetically characterized prokaryotic form I rubiscos. As described below, our results allow for the discovery of features associated with high-performing form I prokaryotic rubiscos.

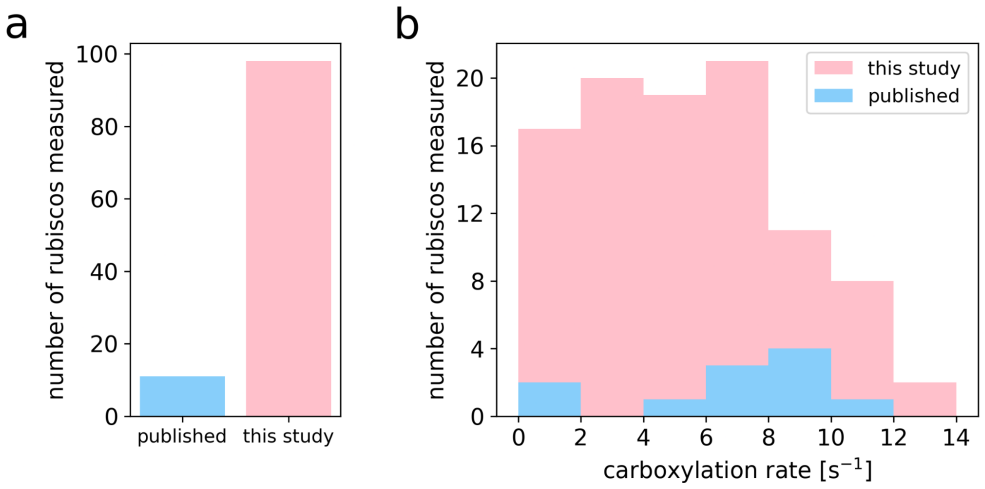

Fig. 1. Systematic exploration of the diversity of prokaryotic form I rubisco. *A) Number of prokaryotic form I rubisco variants with a carboxylation rate reported across the literature and in this study. B) Histogram of the carboxylation rates measured in this study and across the literature.*

Form I rubiscos from phototrophic bacteria are faster carboxylases than those from chemoautotrophs

Form I rubisco-expressing bacteria are facultative or obligate autotrophs: they can convert oxidized inorganic carbon (CO_2) into the reduced organic compounds forming their biomass. To fuel the energy-intensive reactions involved in carbon fixation, autotrophs can draw upon two different energy sources: light (photo-autotrophy) or chemical reactions (chemo-autotrophy). We collected information, when available, on the bacteria expressing the studied rubiscos and classified them into phototrophs and chemotrophs, to investigate whether one of these trophic modes was linked to higher carboxylation rates. To refrain from selection biases, we aimed to uniformly sample rubiscos from both classes (see Materials and Methods). We find that rubiscos originating from phototrophs have a carboxylation rate of 6.5 s^{-1} [4.4-7.9 s^{-1}] (median and interquartile range), three times faster than rubiscos derived from chemoautotrophs (2.0 s^{-1} [1.3-3.9 s^{-1}]) (Fig. 2A, Mann-Whitney U test, $p < 0.01$).

Potential confounding variables (e.g., phylogeny), limit our ability to indicate causality. Yet this result could serve as a basis to investigate if phototrophic bacteria may harbor faster carbon fixation as part of coping with the sun's high rate of energy delivery. In particular, phototrophic bacteria typically inhabit environments with fluctuating light intensity, which may require faster metabolic responses to adapt to changes in light availability. It is possible that rubisco adapted to this need.

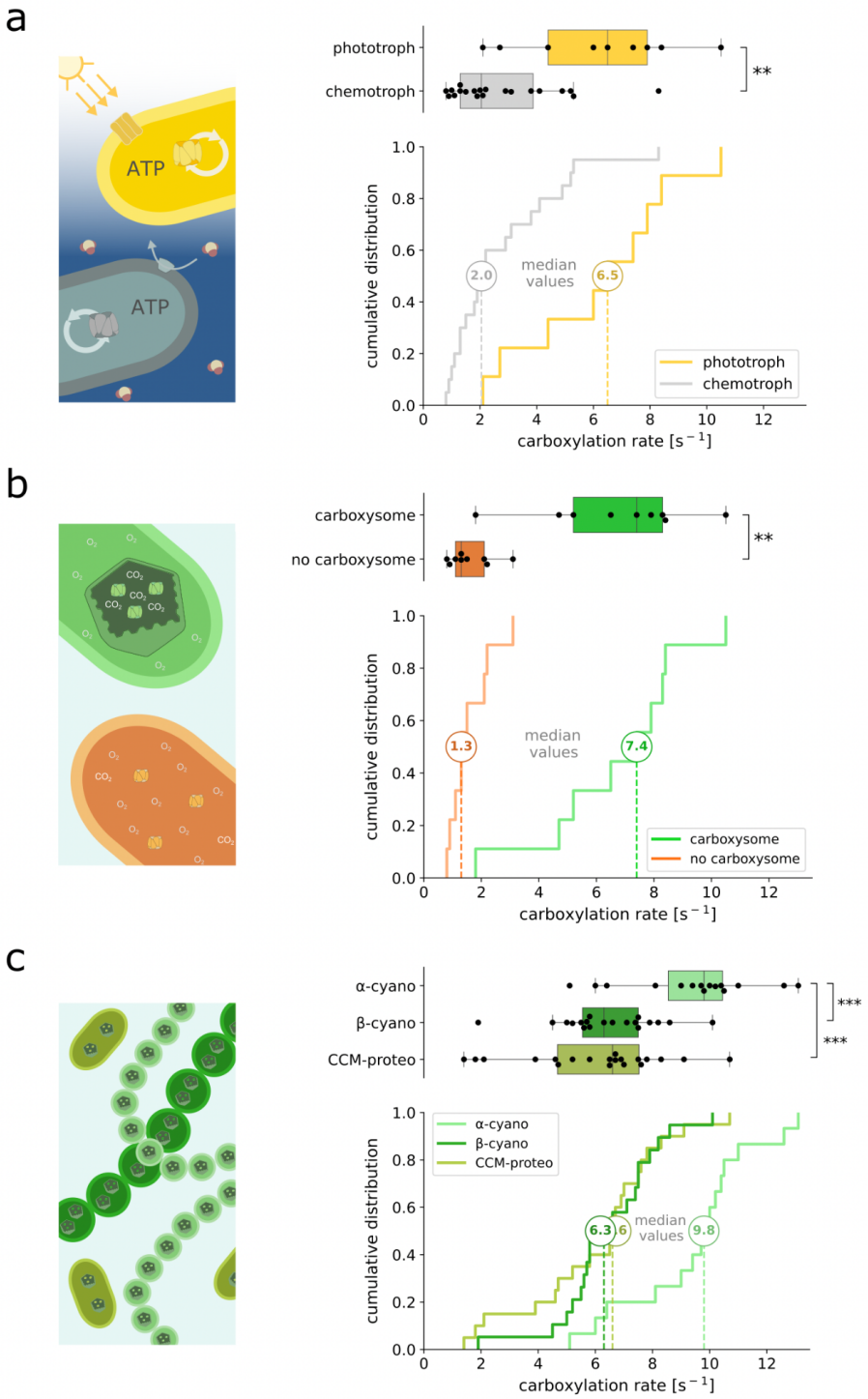

Fig. 2. Large-scale analysis of the biological parameters associated with fast carboxylating rubiscos *Box and cumulative distribution plots of rubisco carboxylation rates from different clusters: chemo- and phototrophic bacterial rubiscos (A), carboxysome-associated rubiscos and their counterparts (B), α- and β-cyanobacterial, and carboxysome-associated proteobacterial rubiscos (C). To ensure unbiased study of every group, we selected 20 and 9 class-representative chemo- and phototroph-associated rubiscos, 9 and 9 class-representative carboxysome-associated and non-associated rubiscos, and 15, 19 and 20 class-*

*representative α- and β- cyanobacterial, and carboxysome-associated proteobacterial rubiscos respectively (see Materials and Methods). Mann-Whitney U test (A and B) or Kruskal-Wallis followed by Dunn multiple comparison tests (C) were applied. **p < 0.01, ***p < 0.001. Legend abbreviations are as follows: α-cyano, α-cyanobacterial rubisco; β-cyano, β-cyanobacterial rubisco; CCM-proteo, carboxysome-associated proteobacterial rubisco.*

Carboxysome-associated form I rubiscos are significantly faster

Another biological parameter that could likely have influenced rubisco evolution is the presence of a carboxysome-based CO_2 concentrating mechanism (CCM). This cellular mechanism combines the active transport of inorganic carbon into the cell and the colocalization of carbonic anhydrase and rubisco inside subcellular proteinaceous microcompartments called carboxysomes, locally increasing CO_2 concentration around rubisco[20]. High CO_2 levels inhibit oxygenation by competitive inhibition, which might permit the use of faster, but less CO_2-specific, rubisco variants.

To compare carboxysomal and non-carboxysomal form I rubiscos, we uniformly sampled rubiscos from each class and compared their measured carboxylation kinetics (see Materials and Methods). We found that with a median catalytic rate of 7.4 s^{-1} [5.2-8.3 s^{-1}], carboxysome associated rubiscos are more than 5 times faster than their non-carboxysomal counterparts (1.3 s^{-1} [1.1-2.1 s^{-1}]) (Fig. 2B, Mann–Whitney U test, $p < 0.01$).

Carboxysomes likely evolved during the Proterozoic eon - in the context of the rise of oxygen and the decrease of carbon dioxide in our atmosphere[20] - to maintain carboxylase activity in this changing atmosphere. One evolutionary strategy is the emergence of carbon-concentrating mechanisms (CCMs), including carboxysomes, that locally concentrate CO_2 around rubisco and therefore maintain local gas concentrations favorable to carboxylation. Another strategy consists of the evolution of rubisco towards stronger affinity for CO_2. It has long been postulated that rubisco is a constrained enzyme, limited by a catalytic tradeoff between its carboxylation rate and affinity for CO_2 (illustrated by the positive correlation between $k_{cat,C}$ and $K_{M,CO2}$)[21,22]. Following this hypothesis[23], rubiscos with stronger affinity for CO_2 tend to exhibit lower carboxylation rates. In contrast, rubiscos which evolved in the context of a CCM with less pressure towards stronger CO_2 affinity, should present higher carboxylation rates. Our findings support this conjecture using the most comprehensive kinetic sampling to date.

Alpha-cyanobacteria express the fastest form I rubiscos across the tree of life

Carboxysomes are expressed by some proteobacteria (like the thiosulfate oxidizer *Halothiobacillus neapolitanus* or the purple sulfur bacteria *Thiorhodococcus drewsii*) and by all cyanobacteria. Cyanobacteria can be divided into two sub-clades, α- and β-cyanobacteria, based on the monophyletic grouping of their rubisco and carboxysome genes[24,25]. It has been posited that α- and β-cyanobacteria rubiscos had identical catalytic rates[26] or even, that β-cyanobacteria rubiscos are faster[22], as they include the fastest form I rubisco characterized up to now (from *Synechococcus elongatus* PCC 6301)[23]. However, such statements were made from scarce measurements, especially among α-cyanobacterial enzymes as only two $k_{cat,C}$ values are currently available[27,28]. We now reevaluate this hypothesis using a wider kinetic sampling of these

192 different rubisco subforms. We have class representative subsets of 15, 19 and 23 variants
193 uniformly covering α-, β-cyanobacteria, and proteobacteria carboxysome-associated rubiscos
194 diversity from our dataset (see Materials and Methods).
195
196 We find that, in contrast to previous statements, α-cyanobacterial rubiscos show the highest
197 carboxylation rates (9.8 s^{-1} [8.6-10.5 s^{-1}]) among all prokaryotic form I rubiscos, ≈50% higher than
198 β-cyanobacterial rubiscos and their proteobacterial counterparts (6.3 s^{-1} [5.6-7.5 s^{-1}] and 6.6 s^{-1}
199 [4.7-7.5 s^{-1}] respectively, Fig. 2C, Kruskal-Wallis test followed by Dunn multiple comparison test,
200 $p < 0.001$).
201
202 When oxygenic photosynthesis emerged in an ancestor of cyanobacteria ≈3 billion years ago[29,30],
203 it alleviated the dependence of these autotrophs on relatively scarce inorganic reductants and
204 probably contributed to the emergence of faster rubiscos. However, we do not find β-
205 cyanobacterial rubiscos to be faster than carboxysome-associated proteobacterial counterparts,
206 in contrast to α-cyanobacterial rubiscos. One possible difference between α- and β-cyanobacteria
207 could be their cell size. Indeed, α-cyanobacteria are known to encompass many members of the
208 so-called picocyanobacteria, the smallest cyanobacteria on Earth[31,26]. We therefore collected from
209 the literature the size dimensions of cyanobacteria, when available. With a median cell volume of
210 0.5 μm^3 (0.3-1.0 μm^3), α-cyanobacteria are more than 25 times smaller in volume than their β
211 counterparts (13.6 μm^3 [3.9-25 μm^3]) (Supplementary Fig. 4, Mann–Whitney U test, $p < 0.0001$).
212 Smaller cells offer higher surface-to-volume ratio and increased exchange with the medium[32],
213 which could support a higher supply of nutrients, and contribute to the evolution of faster rubiscos
214 in these specific cyanobacteria (see Supplementary Note 2 for more details).
215

216 ## Discussion

217 Rubisco is among the enzymes that have probably shaped Earth's biosphere and geosphere the
218 most. Its kinetic parameters result from billions of years of evolution, following (and causing)
219 changes in the atmosphere and climate. We present here a systematic large-scale survey of ≈140
220 rubisco variants covering the genetic diversity of prokaryotic form I rubiscos.
221
222 We note that the present study is limited by the risk of underestimating some rates if, for instance,
223 enzymes are partially denatured in the expression conditions used. Yet, a denatured enzyme
224 probably could not interact with RuBP or CABP, and more generally, we have no reason to
225 presume *a priori* a systematic bias affecting some of the studied groups over the others.
226
227 We find that on average, form I rubiscos from phototrophic bacteria are three times faster than
228 those from chemotrophs. A possible explanation could be that phototrophs tend to benefit from a
229 higher energy flux than chemotrophs, thus more easily driving the ATP-dependent reactions of
230 the Calvin-Benson-Bassham cycle. However, such a case mostly represents oxygenic
231 photosynthesis (cyanobacteria) which is further shown to be associated with the fastest rubiscos
232 in our work. The small number of rubiscos associated with anoxygenic photosynthesis in our
233 dataset prevents us from better studying this specific trophic mode. The latter is associated with
234 different situations and constraints, and most likely provides lower energy levels to cells. Our
235 previous survey[6] spanning the diversity of form II and form II/III rubiscos mostly identified one

group of fast carboxylating rubiscos, emerging from microaerophilic chemotrophs. The very small number of phototrophs in this previous study - absence of oxygenic phototrophs (exclusively composed of cyanobacteria) and low number of anoxygenic phototrophs - made this kind of comparison impossible with that data source.

In the context of significant shifts in oxygen and carbon dioxide levels since the Great Oxygenation Event, biophysical CCMs like carboxysomes evolved within some autotrophs, allowing for maintenance of high CO_2 levels around rubisco. We find that carboxysome-associated rubiscos are, on average, more than 5 times faster than others. Among the main parameters examined in this study, the presence of a carboxysome exhibits the strongest association with the occurrence of fast carboxylating rubiscos (Supplementary Fig. 5). This finding underscores the constraints imposed by CO_2 levels on rubisco carboxylation rate, as lowering this pressure appears to reliably facilitate the development of higher carboxylation rates. A compelling example within our dataset is the gammaproteobacteria *Hydrogenovibrio kuenenii*, which expresses two form I rubiscos. The carboxysomal rubisco has a rate of 8.3 s^{-1}, which is twice as fast as its non-carboxysomal counterpart (4.2 s^{-1}, Supplementary Data 1), exemplifying the major impact of carboxysome presence on rubisco carboxylation. Although we have not collected data regarding rubisco's Michaelis constant for CO_2 ($K_{M,CO2}$) in this study, the high CO_2 levels inside carboxysomes likely allow for fast but weaker-affinity rubiscos to evolve, in line with the hypothesis that a tradeoff exists between these two kinetic parameters[20].

We further analyzed the correlation between rubisco carboxylation rate and various biological and environmental parameters such as bacterial environmental source, rubisco subtype, bacterial halotolerance, pH, oxygen-sensitivity, or optimal growth temperature. As presented in Supplementary Fig. 6-11, these showed weaker or no correlations.

Overall, rubisco carboxylation rates seem to correlate with a select few environmental parameters, providing insights into autotrophic metabolism constraints. The fitness of autotrophic organisms largely depends on carbon fixation rates. This process relies on three necessary inputs: energy, water, and CO_2. When water and energy are highly abundant (aquatic phototrophs), the main limiting factor is the flow of carbon into the metabolic network. CO_2 needs to first reach the cell membrane, diffuse to the active site of rubisco, and successfully be involved in a catalytic event. Small cells (increased exchange) with a carboxysome (higher probability of catalysis) are therefore the ones most limited by the carboxylation rate of rubisco, as all other bottlenecks are relatively smaller in that case. The fact that the fastest rubiscos found in our study were from α-cyanobacteria, the smallest carboxysome-expressing oxygenic phototrophs in the microbial world, is consistent with this rationale and could be related to the finding that they account for the majority of cyanobacteria globally[26].

This study provides a systematic exploration of prokaryotic form I rubisco maximal rates and its relationship with various contextual factors that could have shaped the evolution of this most abundant enzyme on Earth. Gaining understanding of this dynamic process is crucial for predicting the evolution of autotrophs in a changing climate and the ability to engineer it for human needs.

Acknowledgments

We thank Yoav Peleg, Ron Sender, Noam Prywes, Ralf Steuer, Avi Flamholz and David Savage for important conversations and productive feedback on this manuscript. We thank Michelle Gehring for additional information about unpublished data from their laboratory. This research was supported by the Mary and Tom Beck Canadian Center for Alternative Energy Research, Miel de Botton, the Schwartz Reisman Collaborative Science Program, and the Charles and Louise Gartner Professorial Chair.

Materials and Methods

Rubisco sequence collection

Rubisco large subunit sequences were collected from (i) the NCBI's nr database[33] downloaded in December 2020 and searched following the method described in Davidi *et al.*[6]; (ii) in-house assemblies of the 244 samples from the Tara Oceans expedition[34]; (iii) assemblies and rubisco sequences published by [35–39]. Sequences outside the length range of 300-700 amino acids were removed. The remaining sequences were then clustered at an 80% sequence identity threshold using USEARCH algorithm[40]. Cluster representatives were aligned using MAFFT (v7.475, default parameters)[41], and columns with more than 95% gaps were removed using trimAl (v1.4.rev15, -gt 0.05)[42]. A phylogenetic tree was constructed using FastTree (v2.1.10, default parameters)[43]. To identify the different rubisco forms, we relied on annotated sequences from NCBI, Tabita *et al.*[44], and Banda *et al.*[39]. This process resulted in a total of 72,395 sequences, including 56,161 form I rubiscos, and most notably for this study, 4,302 non-eukaryotic form I rubiscos. The latter were further re-clustered at 90% identity using USEARCH algorithm (custom python script; see below), and a phylogenetic tree was constructed using RAxML[45].

Rubisco variants selection for characterization

Form I rubiscos are divided into 5 separate groups[9,38,46–48] (see Supplementary Fig. 7A): forms IA and IB (forming the “green” type, found in cyanobacteria and some proteobacteria); forms ICD and IE (the “red” type, found in proteobacteria); and the recently discovered form I “Anaero” (found in bacteria related to anaerobic, thermophilic Chloroflexaeota and Firmicutes)[49]. To comprehensively sample the sequence space of form I rubisco diversity, an iterative clustering approach of the large subunit gene was employed using USEARCH algorithm. The resulting representative sequences from each cluster were selected for characterization. Initially, 32 rubiscos were chosen to cover the entire diversity of form I rubisco at a threshold of 75% identity. Subsequently, further clustering was performed on smaller groups of rubiscos of particular interest, with increasing threshold percentages. Throughout the study, this successively included 38 rubiscos representing the diversity of form IA and B rubisco at 85% identity, 13 rubiscos representing the diversity of cyanobacterial rubisco at 88% identity, 31 rubiscos representing the diversity of IB rubisco at 91% identity, 20 rubiscos representing the diversity of cyanobacterial IA rubisco at 97.5% identity, and 23 rubiscos representing the diversity of proteobacteria carboxysome-associated rubisco at 90% identity (See Supplementary Fig. 1). These representative sequences could sometimes overlap, resulting in a total of 129 different rubisco sequences tested in this study. Additionally, 15 rubiscos were arbitrarily selected for setting-up the experimental pipeline. In total, 144 different rubisco variants were selected for characterization in this study.

Gene synthesis

For each chosen rubisco variant, the complete rubisco operon, encompassing rubisco large and small subunit genes, as well as the chaperone rbcX gene for IB rubiscos, was retrieved. The operons were then codon-optimized for expression in *E. coli* (Twist Codon Optimization tool) and synthesized by Twist Bioscience. Following synthesis, these operons were cloned into a pET-29b(+) overexpression vector (NdeI_XhoI insertion sites). Validation of gene synthesis and cloning was conducted through next-generation sequencing as part of the Twist bioscience service.

High-throughput rubisco expression

Chemocompetent BL21(DE3) cells, previously transformed with a pESL plasmid coding for the chaperone GroEL-GroES[13], were transformed with the rubisco library and incubated at 37°C, 250 rpm in 8 ml of LB media supplemented with 30 µg/ml chloramphenicol and 50 µg/ml kanamycin. Growth was performed in 24-deep-well plates. When cells reached an OD600 of 0.6, GroEL-GroES expression was induced by adding arabinose (0.2% final) and incubating at 23°C for 45 min. Rubisco expression was then induced by adding 0.2 mM IPTG (isopropyl β-d-thiogalactoside) and incubating at 23°C for 21 h. For protein extraction, cells were harvested by centrifugation (15 min; 4,000 g; 4°C) and pellets were lysed with BugBuster® ready mix (Millipore) for 25 min at room temperature; 70 µl BugBuster master mix (EMD Millipore) was added to each sample. Crude extracts were then centrifuged for 30 min at 4,000 g, 4°C to remove the insoluble fraction. For quality control, all samples were run on an SDS–PAGE gel (Supplementary Fig. 12).

Raf1-IB rubisco co-expression

For IB rubiscos (originating from β-cyanobacteria), a pilot study was performed to find an homologue of the chaperone Rubisco accumulation factor 1 (Raf1), that could help solubly-express these rubisco variants. 3 different Raf1 were first tested with their cognate rubisco, from 3 β-cyanobacteria (namely, *Trichormus variabilis*, *Pseudanabaena sp. PCC 6802*, and *Euhalothece natronophila*). Codon-optimized *raf1* genes were synthesized and cloned by Twist Bioscience in polycistron with their cognate rubisco genes into pET-29b(+) vectors. Proteins were expressed and tested *in vitro*, as described above, and Raf1 from *Euhalothece natronophila* was chosen, based on ability to solubilize its cognate rubisco. The gene was therefore cloned into a pESL plasmid, in the same operon as GroEL-ES genes, and transformed into chemocompetent BL21(DE3) cells. IB rubiscos were then transformed and expressed in these cells as described previously.

High-throughput determination of rubisco carboxylation rates

To determine rubisco carboxylation rates, we performed kinetic assays directly from the soluble fraction of prepared lysates since purifying both large and small rubisco subunits together was not feasible in a high-throughput manner. Soluble fractions were incubated with 4% CO_2 and 0.4% O_2 for rubisco activation (15 min; 4°C; plate shaker at 250 rpm). Following activation, rubisco carboxylase activities were tested as described in Davidi *et al.*[6]. Briefly, 10 µl of the activated rubisco sample was added to six aliquots of 80 µl of assay mix (a detailed list of all assay components and their sources is provided in Supplementary Table 1) containing different concentrations of CABP (0, 0, 10, 20, 30, and 90 nM). The mix was incubated for 15 minutes at

368 30°C in a plate reader (Infinite® 200 PRO; TECAN) connected to a gas control module (TECAN
369 pro200) ensuring atmospheric conditions of 4% CO_2 and 0.4% O_2. Rubisco carboxylation activity
370 was initiated by adding 10 µl RuBP to each sample (final concentration of 1 mM and a total volume
371 of 100 µl). The carboxylation rate was determined through a coupled reaction[50,51]. In brief, 3-
372 phosphoglycerate, the product of ribulose 1,5-bisphosphate carboxylation, was phosphorylated
373 and subsequently reduced into glyceraldehyde 3-phosphate, involving NADH oxidation. The latter
374 could be monitored through 340 nm absorbance for 15 minutes at 2-second intervals. To convert
375 from NADH to rubisco reactions per second, we assumed a stoichiometric ratio of 2:1 between
376 NADH and carboxylation reactions. The active-site concentration was determined by fitting a
377 linear regression model (custom python script; see below) to the measured reaction rates as a
378 function of the CABP inhibitor concentrations (See Supplementary Note 1 and Supplementary
379 Fig. 13 for more details). Due to the use of soluble fractions of total cell lysates, the initial
380 concentration of rubisco could not be determined beforehand, which often led to saturation of the
381 first assay with rubisco. To overcome this issue, we adjusted the concentration of lysates by
382 dilution, to obtain a rubisco concentration that enabled measurable inhibition by CABP, allowing
383 for an accurate quantification of the active-site concentration. We finally obtained the rate per
384 active site by dividing the activity with no CABP by the concentration of active sites. As was done
385 in Davidi *et al.*[6], and since not all variants were tested on the same day, the form-II rubisco from
386 *R. rubrum*, commonly used as a standard in rubisco kinetic assays, was consistently included as
387 an internal reference in every measurement.
388
389 **Rubisco variants selection for comparative analysis**
390 For comparing rubisco carboxylation rates between groups without sampling bias, we aimed to
391 select representative sets of variants uniformly covering the different groups studied. For the
392 comparison of rubiscos originating from phototrophic versus chemotrophic organisms, we
393 selected a set of 32 rubiscos, covering the entire diversity of form I rubiscos at a 75% similarity
394 threshold, enriched with a set of 36 rubiscos, covering the entire diversity of form IA and B
395 rubiscos at an 85% similarity threshold. Among these 68 variants, phototrophic rubiscos were
396 distinguished from chemotrophic ones based on their co-occurrence with photosystem genes in
397 the respective genome. This resulted in 53 chemotropic and 15 phototrophic rubiscos, of which
398 20 and 9 respectively were soluble and active *in vitro*.
399 For the comparison of carboxysome and non-carboxysome-associated rubiscos, we selected a
400 set of 14 rubiscos covering the entire diversity of carboxysome-associated rubiscos at an 85%
401 similarity threshold, and of 30 rubiscos covering the entire diversity of non-carboxysome-
402 associated rubiscos at a 75% similarity threshold. 9 and 9 of them were soluble and active *in vitro*.
403 For the comparison of α- and β- cyanobacterial, and carboxysome-associated proteobacterial
404 rubiscos, we selected 19 rubiscos covering the entire diversity of α-cyanobacteria rubiscos at a
405 97.5% similarity threshold, 29 rubiscos covering the entire diversity of β-cyanobacteria rubiscos
406 at a 91% similarity threshold, and 23 rubiscos covering the entire diversity of carboxysome-
407 associated proteobacterial rubiscos at a 90% similarity threshold. 15, 19 and 20 of them
408 respectively were soluble and active *in vitro*.
409

Data and Code Availability

All the data supporting the findings of this study as well as the codes used for generating our list of rubiscos and for analyzing the results is open source and can be found in the following link: https://gitlab.com/milo-lab-public/rubisco-F1.

References

1. Raven, J. A. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56**, 177–192 (2009).
2. Bar-On, Y. M. & Milo, R. The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4738–4743 (2019).
3. Prywes, N., Phillips, N. R., Tuck, O. T., Valentin-Alvarado, L. E. & Savage, D. F. Rubisco Function, Evolution, and Engineering. *Annu. Rev. Biochem.* (2023) doi:10.1146/annurev-biochem-040320-101244.
4. Wildman, S. G. Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate carboxylase-oxygenase). *Photosynth. Res.* **73**, 243–250 (2002).
5. Flamholz, A. I. *et al.* Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* vol. 58 3365–3376 Preprint at https://doi.org/10.1021/acs.biochem.9b00237 (2019).
6. Davidi, D. *et al.* Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.* **39**, e104081 (2020).
7. Mann, C. C. Genetic Engineers Aim to Soup Up Crop Photosynthesis. *Science* **283**, 314–316 (1999).
8. Tabita, F. R. Microbial ribulose 1, 5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth. Res.* **60**, 1–28 (1999).
9. Badger, M. R. & Bek, E. J. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *Journal of Experimental Botany* vol. 59 1525–1541 Preprint at https://doi.org/10.1093/jxb/erm297 (2008).
10. Witte, B. *et al.* Functional prokaryotic RubisCO from an oceanic metagenomic library. *Appl. Environ. Microbiol.* **76**, 2997–3003 (2010).
11. Aigner, H. *et al.* Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
12. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).

445 13. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. Reconstitution of active
446 dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two
447 chaperonin proteins and Mg-ATP. *Nature* **342**, 884–889 (1989).
448 14. Lin, Z. & Rye, H. S. Expansion and compression of a protein folding intermediate by GroEL.
449 *Mol. Cell* **16**, 23–34 (2004).
450 15. Huang, F. *et al.* Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating
451 Rubisco assembly and carboxysome biogenesis. *Proceedings of the National Academy of*
452 *Sciences* **117**, 17418–17428 (2020).
453 16. Xia, L.-Y. *et al.* Molecular basis for the assembly of RuBisCO assisted by the chaperone
454 Raf1. *Nature Plants* **6**, 708–717 (2020).
455 17. Hauser, T. *et al.* Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat.*
456 *Struct. Mol. Biol.* **22**, 720–728 (2015).
457 18. Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M. & Szczepaniak, A. Rubisco
458 Accumulation Factor 1 from Thermosynechococcus elongatus participates in the final
459 stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in Escherichia coli
460 cells and in vitro. *FEBS J.* **281**, 3920–3932 (2014).
461 19. Cen, Y.-P. & Sage, R. F. The regulation of Rubisco activity in response to variation in
462 temperature and atmospheric CO2 partial pressure in sweet potato. *Plant Physiol.* **139**,
463 979–990 (2005).
464 20. Flamholz, A. & Shih, P. M. Cell biology of photosynthesis over geologic time. *Curr. Biol.* **30**,
465 R490–R494 (2020).
466 21. Jordan, D. B. & Ogren, W. L. Species variation in the specificity of ribulose biphosphate
467 carboxylase/oxygenase. *Nature* **291**, 513–515 (1981).
468 22. Iñiguez, C. *et al.* Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2
469 concentrating mechanisms. *Plant J.* **101**, 897–918 (2020).
470 23. Savir, Y., Noor, E., Milo, R. & Tlusty, T. Cross-species analysis traces adaptation of
471 Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci. U. S. A.*
472 **107**, 3475–3480 (2010).
473 24. Badger, M. R. & Price, G. D. CO2 concentrating mechanisms in cyanobacteria: molecular
474 components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
475 25. Whitehead, L., Long, B. M., Price, G. D. & Badger, M. R. Comparing the in vivo function of
476 α-carboxysomes and β-carboxysomes in two model cyanobacteria. *Plant Physiol.* **165**,
477 398–411 (2014).
478 26. Cabello-Yeves, P. J. *et al.* α-cyanobacteria possessing form IA RuBisCO globally dominate

479 aquatic habitats. *ISME J.* (2022) doi:10.1038/s41396-022-01282-z.
480 27. Shih, P. M. *et al.* Biochemical characterization of predicted Precambrian RuBisCO. *Nat.*
481 *Commun.* **7**, 10382 (2016).
482 28. Long, B. M. *et al.* Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in
483 tobacco chloroplasts. *Nat. Commun.* **9**, 3570 (2018).
484 29. Fournier, G. P. *et al.* The Archean origin of oxygenic photosynthesis and extant
485 cyanobacterial lineages. *Proceedings of the Royal Society B: Biological Sciences* **288**,
486 20210675 (2021).
487 30. Sánchez-Baracaldo, P. & Cardona, T. On the origin of oxygenic photosynthesis and
488 Cyanobacteria. *New Phytol.* **225**, 1440–1446 (2020).
489 31. Jasser, I. & Callieri, C. Picocyanobacteria. in *Handbook of Cyanobacterial Monitoring and*
490 *Cyanotoxin Analysis* 19–27 (John Wiley & Sons, Ltd, 2017).
491 32. Young Kevin D. The Selective Value of Bacterial Shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–
492 703 (2006).
493 33. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–42 (2013).
494 34. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**,
495 1261359 (2015).
496 35. Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple
497 uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).
498 36. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain
499 Bacteria. *Nature* **523**, 208–211 (2015).
500 37. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected
501 biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 1–11 (2016).
502 38. Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III,
503 and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco
504 evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524 (2008).
505 39. Banda, D. M. *et al.* Novel bacterial clade reveals origin of form I Rubisco. *Nat Plants* **6**,
506 1158–1166 (2020).
507 40. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
508 **26**, 2460–2461 (2010).
509 41. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
510 Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
511 42. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
512 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973

513 (2009).
514 43. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood
515 trees for large alignments. *PLoS One* **5**, e9490 (2010).
516 44. Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H. & Kreel, N. E. Phylogenetic and
517 evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional
518 lessons provided by diverse molecular forms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**,
519 2629–2640 (2008).
520 45. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
521 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
522 46. Spreitzer, R. J. & Salvucci, M. E. Rubisco: structure, regulatory interactions, and
523 possibilities for a better enzyme. *Annu. Rev. Plant Biol.* **53**, 449–475 (2002).
524 47. Park, S. W. *et al.* Presence of duplicate genes encoding a phylogenetically new subgroup
525 of form I ribulose 1,5-bisphosphate carboxylase/oxygenase in Mycobacterium sp. strain
526 JC1 DSM 3803. *Res. Microbiol.* **160**, 159–165 (2009).
527 48. Grostern, A. & Alvarez-Cohen, L. RubisCO-based CO2 fixation and C1 metabolism in the
528 actinobacterium Pseudonocardia dioxanivorans CB1190. *Environ. Microbiol.* **15**, 3040–
529 3053 (2013).
530 49. Schulz, L. *et al.* Evolution of increased complexity and specificity at the dawn of form I
531 Rubiscos. *Science* **378**, 155–160 (2022).
532 50. Lilley, R. M. & Walker, D. A. The reduction of 3-phosphoglycerate by reconstituted
533 chloroplasts and by chloroplast extracts. *Biochimica et Biophysica Acta (BBA) -*
534 *Bioenergetics* **368**, 269–278 (1974).
535 51. Kubien, D. S., Brown, C. M. & Kane, H. J. Quantifying the amount and activity of Rubisco in
536 leaves. *Methods Mol. Biol.* **684**, 349–362 (2011).