In vivo directed evolution of an ultra-fast RuBisCO from a semi-anaerobic environment imparts oxygen resistance

Julie L. McDonald[1,2], Nathan P. Shapiro[1], Spencer M. Whitney[3], Robert H. Wilson[1,*], Matthew D. Shoulders[1,4,5,*]

1. Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA
2. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
3. Plant Science Division, Research School of Biology, The Australian National University, Canberra, Australian Capital Territory, Australia
4. Broad Institute of MIT and Harvard, Cambridge, MA, USA
5. Koch Institute for Integrative Cancer Research at MIT, Cambridge, MA, USA

Email: rhwilson@mit.edu; mshoulde@mit.edu

Author Contributions: J.L.M., R.H.W., and M.D.S. conceived this work. All authors designed experiments and analyzed data. J.L.M., N.P.S., R.H.W., and S.M.W. performed experiments. J.L.M., R.H.W, M.D.S, and S.M.W. wrote the manuscript. All authors edited the manuscript.

Abstract

Carbon dioxide assimilation by the enzyme Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (RuBisCO) underpins biomass accumulation in photosynthetic bacteria and eukaryotes. Despite its pivotal role, RuBisCO has a slow carboxylation rate ($k_{cat}^{CO_2}$) and is competitively inhibited by oxygen. These traits impose limitations on photosynthetic efficiency, making RuBisCO a compelling target for improvement. Interest in RuBisCO from *Gallionellaceae* bacteria, which comprise a dimer or hexamer of large subunits, arises from their nearly 5-fold higher $k_{cat}^{CO_2}$ than the average RuBisCO enzyme. Here, a comprehensive kinetic analysis of a *Gallionellaceae* RuBisCO (GWS1B) reveals the fastest CO_2 fixation rate measured for this enzyme (25.8 s^{-1}), and exposes an extreme sensitivity to oxygen inhibition, consistent with its evolution under semi-anaerobic environments. We used a novel *in vivo* mutagenesis-mediated screening pipeline to evolve GWS1B over six rounds under oxygenic selection, identifying three catalytic point mutants with improved ambient carboxylation efficiency; Thr-29-Ala (T29A), Glu-40-Lys (E40K) and Arg-337-Cys (R337C). Full kinetic characterization showed that each substitution enhanced the CO_2 affinity of GWS1B under oxygenic conditions while simultaneously subduing oxygen affinity, leading to 24% (E40K), 9% (T29A) and 7% (R337C) enhancements in carboxylation efficiency under ambient O_2. By contrast, under the near anaerobic natural environment of *Gallionellaceae*, the carboxylation efficiency of each mutant was impaired ~16%. These findings demonstrate the efficacy of artificial directed evolution to access novel regions of catalytic space in RuBisCO, and a capacity for rapid evolution of kinetic traits in response to environmental change.

Significance

Given RuBisCO's crucial role in carbon dioxide assimilation, addressing its slow carboxylation rate and oxygen inhibition is a significant challenge. Utilizing the fastest known, yet highly oxygen-sensitive, RuBisCO from the bacteria *Gallionellaceae*, we applied a novel *in vivo* directed evolution pipeline in *Escherichia coli* to discover mutations that specifically enhance carboxylation efficiency under ambient oxygen, a condition distinct from *Gallionellaceae's* natural semi-anaerobic environment. Our findings underscore the potential of directed evolution to unlock new catalytic capabilities for RuBisCO, with implications for both fundamental research and practical agricultural applications.

Introduction

The enzyme Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (RuBisCO) is essential for life on Earth, fixing atmospheric carbon dioxide (CO_2) to organic carbon during photosynthesis. The RuBisCO enzyme family has diverse isoforms across phylogeny. Plants, algae, and many mixotrophic and autotrophic bacteria utilize Form I RuBisCO, where the enzyme is composed of a large (RbcL) and small (RbcS) subunit as an $RbcL_8RbcS_8$ hexadecamer. Form II and Form III RuBisCO, found in bacteria and archaea, lack RbcS and present as arrangements of RbcL dimers ($RbcL_2$–$RbcL_{10}$). Form II/III RuBisCO display lower substrate specificity for CO_2 over oxygen (O_2; $S_{C/O}$) than Form I enzymes, attributed to their lack of a small subunit. RbcS evolved in Form I RuBisCO to support oxygenic photosynthesis (1).

Despite their widespread prevalence in Nature and essentiality in carbon fixation, RuBisCO enzymes display striking inefficiencies. The rate of RuBisCO-mediated carboxylation is most often in the range of 1–10 reactions per second (2) – much slower than most other enzymes involved in carbon metabolism. The enzyme also reacts promiscuously with atmospheric O_2, producing glycolate that must be recycled via energy-consuming pathways, including photorespiration in plants (3) and phosphoglycolate salvage in bacteria (4). RuBisCO's weak specificity for CO_2 may be related to its evolutionary emergence in an anoxic atmosphere, prior to the rise of oxygenic photosynthesis (5, 6). Development of the current oxygen-rich environment forced RuBisCO to gain selectivity for CO_2 over O_2, though there is significant evidence that this increase in CO_2 specificity was met with a decrease in carboxylation rate ($k_{cat}^{CO_2}$) (2, 7, 8). Form II RuBisCO featuring particularly high $k_{cat}^{CO_2}$ values (>10 s^{-1}) are associated with environments high in dissolved CO_2 and/or low in O_2, such as acidic iron-rich soils (9) and the deep sea (10).

A large body of both theoretical and experimental research indicates that improvements to RuBisCO carboxylation efficiency can enhance photosynthetic efficiency (11-17), making RuBisCO one of the most promising engineering targets to enhance biological carbon capture and improve crop yields. A major goal of RuBisCO engineering for enhanced photosynthetic efficiency has been to increase the speed of carboxylation while avoiding a concomitant increase in oxygenation. Several directed evolution attempts have been successful in improving $k_{cat}^{CO_2}$ (18-21), affinity for CO_2 under anaerobic conditions [$K_M^{CO_2}(0\%\ O_2)$] (18, 22-25), and $S_{C/O}$ (18, 19) of RuBisCO. However, these studies either did not measure or failed to innovate on the biologically-relevant parameter of O_2 resistance (18, 21-28).

In addition to RuBisCO engineering efforts, several studies have sought to profile the natural catalytic landscape of the RuBisCO family (29-31). Form II RuBisCO enzymes from the

semi-anaerobic bacterial family *Gallionellaceae* have garnered attention from such profiling efforts and from metagenomic analyses (30, 32), as they display some of the fastest RuBisCO-mediated carboxylation rates ever measured (15–22 s^{-1}) (30, 33). Interestingly, *Gallionellaceae* RuBisCO, while very fast, also have high substrate affinity for O_2 (33). The combination of these extreme and unique properties (fast carboxylation kinetics and high sensitivity to O_2) position *Gallionellaceae* RuBisCO as a novel starting point for directed evolution, with the potential to uncover distinct loci important for mitigating oxygenation.

Here, we describe an enhanced *in vivo* laboratory directed evolution pipeline for RuBisCO enzymes that utilizes the mutagenesis tool MutaT7 (34-39) to circumvent low-throughput and shallow *in vitro* mutagenesis that have hampered previous RuBisCO directed evolution campaigns (18-28). MutaT7 is a targeted, genomically-encoded mutagen composed of a deaminase fused to a T7 RNA polymerase. MutaT7 processively and selectively mutates any DNA region that is flanked by a T7 promoter and terminator. The high mutagenesis frequency (34-39) of MutaT7 positions it as a useful tool for deep sampling of the RuBisCO sequence landscape. We applied MutaT7 (36) in tandem with RuBisCO-dependent *Escherichia coli* (RDE/RDE2) (20) to evolve a hexameric RuBisCO from *Gallionellaceae* (termed GWS1B) that was previously identified in a groundwater sample (32). This campaign converged on multiple "winner" mutants encoding amino acid substitutions positioned near the active site and oligomeric interfaces. Kinetic analysis revealed that these mutants were less sensitive to O_2 than wild-type GWS1B, displaying consistently decreased O_2 affinity ($K_M^{O_2}$) alongside improved carboxylation efficiency in air.

In sum, by beginning our evolution campaign with an exceptionally fast but also remarkably O_2-sensitive natural RuBisCO, we were able to uncover RuBisCO variants that exist in a region of catalytic space not previously observed for RuBisCO. These results strongly motivate continued RuBisCO engineering efforts that can improve the enzyme beyond its natural capabilities.

Results

Biochemical characterization of GWS1B RuBisCO

We began by characterizing the activity of wild-type GWS1B. GWS1B, here expressed with a C-terminal polyhistidine tag for purification, displayed the highest carboxylation rate reported for any RuBisCO to-date (25.8 s^{-1}; **Figure 1A**). This high rate is consistent with other members of the *Gallionellaceae* family (30), and far surpasses other well-studied Form II enzymes such as that from *Rhodospirillum rubrum* (22, 40). Notably, the carboxylation rate of the C-terminally-tagged GWS1B exceeded previous measurements of an N-terminally tagged isoform of the enzyme and was consistent with measurement of the untagged GWS1B (**Supplementary Table 1**) (32, 33). Thus, we used C-terminally tagged GWS1B for further work, enabling straightforward purification of the protein for spectrophotometric determination of carboxylation rate and to measure $S_{C/O}$.

For clarity, we note that GWS1B is distinct from the similarly fast RuBisCO enzyme from *Gallionella* sp. ($k_{cat}^{CO_2} = 22 \pm 1$ s^{-1}) (30). GWS1B assembles as a hexamer rather than as a dimer, and shares only 68% primary sequence identity to *Gallionella sp.* For comparison, RbcL from β-cyanobacteria and vascular plants share ~80% primary sequence identity but differ substantially in virtually every catalytic parameter. Accessing a high rate of CO_2 fixation is therefore not limited to an isolated region of sequence space but can be accessed by multiple evolutionary routes.

Despite its high rate of carboxylation, GWS1B's overall carboxylation efficiency is severely reduced by the presence of O_2. GWS1B experiences a 73% decrease in carboxylation efficiency when O_2 is present at ambient air levels (21% O_2), compared to 62% for *R. rubrum* RuBisCO – which is itself already considered a highly O_2-sensitive Form II enzyme (**Figure 1B**).

Alongside GWS1B's extreme CO_2 fixation rate and O_2 sensitivity, the remarkably high solubility of GWS1B upon expression in *E. coli* (~77% of the total expressed RbcL; **Figure 1C**) presents another compelling attribute for directed evolution in RDE. Weakly soluble RuBisCO typically circumvent intended selection for improved RuBisCO kinetic properties, as solubility-improving variants can yield a fold-increase in total RuBisCO active site abundance, conferring a higher fitness advantage compared to kinetic variants that display improvement in carboxylation rate (20, 22). Here, the use of an already highly soluble RuBisCO, such as GWS1B, may be expected to reduce or even eliminate this path to selection in RDE that can evade kinetic improvements. Moreover, the high solubility of GWS1B enables expression of large amounts of soluble RuBisCO from the single-copy number bacterial artificial chromosome

(BAC) vectors that are preferentially used in concert with MutaT7-driven *in vivo* directed evolution campaigns (41).

In vivo directed evolution of GWS1B RuBisCO using MutaT7-RDE

Past RuBisCO directed evolution campaigns have relied on gene diversification by error-prone PCR, transformation of the resulting gene library into *E. coli*, and subsequent selection using a RuBisCO-dependent chassis (**Figure 1D**) (18-28). These campaigns, while moderately successful in improving some RuBisCO catalytic parameters, have suffered from low-throughput mutagenesis – often only proceeding through a single round owing to the labor-intensive process of error-prone PCR-mediated library creation and analysis. We sought to combine MutaT7 with RDE to create a new type of *in vivo* mutagenesis and selection system for RuBisCO evolution and enable efficient repetition of directed evolution rounds.

To enable this goal, we first transformed the selection plasmid pACYC^{BAD}PRK-NptII into MutaT7-Dual8 *E. coli*. MutaT7-Dual8 is a DH10B-derived *E. coli* strain that contains two genomically-encoded mutagenic T7 RNA polymerase (T7 RNAP) fusions that install transition mutations at loci between a T7 RNAP promoter and terminator sequence (36) (**Figure 1D**). pACYC^{BAD}PRK-NptII is a selection plasmid that enables arabinose-tunable expression of phosphoribulokinase (PRK), an enzyme that phosphorylates the endogenous metabolite ribulose-5-phosphate (R5P) to yield ribulose-1,5-bisphosphate (RuBP), the sugar substrate of RuBisCO. RuBP accumulation is toxic to *E. coli*, thus rendering cell survival dependent on RuBisCO-mediated RuBP clearance (20, 22, 27) (**Figure 1D**). Previous work has shown that fusion of PRK to the kanamycin resistance protein NptII, as in pACYC^{BAD}PRK-NptII, reduces false positives by linking PRK elimination to kanamycin sensitivity (20). When PRK-NptII is present as a fusion protein, selection escape via elimination or early termination of the PRK protein also eliminates kanamycin resistance, meaning that inclusion of kanamycin in the growth media can remove these otherwise common “cheaters”.

To test GWS1B expression in non-mutagenic DH10B-RDE to ensure that RDE was suitable for variant selection, we spotted bacteria expressing GWS1B or an empty vector control onto agar plates under three conditions; one where no protein expression was induced (permissive growth), one where GWS1B was expressed (induced growth), and one where both GWS1B and PRK-NptII were expressed (selective growth) (**Figure 1E**).Though GWS1B expression appeared to have a somewhat toxic effect on *E. coli* cells under induced growth, *E. coli* expressing the wild-type enzyme grew slowly while the empty vector control completely failed to grow under selection, as expected. This difference in growth rate suggested to us that

RDE is suitable for GWS1B evolution, as catalytically-dead GWS1B should fail to grow and catalytically-improved variants can easily establish a faster, selectable growth rate relative to the wild-type enzyme. We thus transformed MutaT7-Dual8 containing pACYC^{BAD}PRK-NptII with a MutaT7-targeting BAC vector encoding polyhistidine-tagged GWS1B (BAC-GWS1B). We named this mutagenesis-selection system, containing all the necessary components for *in vivo* directed evolution of RuBisCO, MutaT7-RDE.

MutaT7-RDE containing BAC-GWS1B was subjected to six rounds of evolution under increasing selection pressure from PRK induction (**Figure 2A**). MutaT7-driven *in vivo* mutagenesis and selection allowed us to conduct all six rounds of evolution in just 30 working days, in contrast to only 1–2 rounds conducted in past RDE campaigns over a similar timeframe (42). After a 24-hour continuous mutagenesis period in liquid culture initiated by expression of MutaT7, the bacteria were plated on solid selective agar and incubated under an atmosphere of 10% CO_2-enriched air (21% O_2) to, ideally, select for variants that would influence GWS1B's specificity and/or carboxylation rate under an inhibitory O_2 concentration (**Figure 1B**). After three days of selective growth, plasmid pools containing both BAC-GWS1B and the PRK-expressing plasmid pACYC^{BAD}PRK-NptII were extracted. To remove any PRK-NptII fusions that may have been mutated to weaken selection, pACYC^{BAD}PRK-NptII was removed from this pool by digestion with the restriction enzyme SfiI. This digestion leaves a purified "winning" pool of mutated GWS1B plasmids, signifying one complete round of evolution. The winner pool was retransformed into MutaT7-RDE and the process was repeated for six rounds with incremental increases in selection pressure via increasing PRK expression.

We collected purified GWS1B mutant libraries from each evolution round and used them to further enforce selection. We transformed the selected libraries into *E. coli* that contained a clean copy of pACYC^{BAD}PRK-NptII and did not contain MutaT7 in the genome. Selection was then repeated at 10% CO_2 and 21% O_2 and single-colony isolates were extracted. Isolates from this enrichment step were characterized using whole-plasmid sequencing. The winning populations, which grew under PRK-NptII-expressing conditions were also sequenced using next-generation sequencing (NGS) after rounds 1, 2, 4, and 6, which we first analyzed per round to determine the extent of targeted mutagenesis.

Short-read NGS on the winner populations revealed the overall frequency and distribution of mutations at each position on BAC-GWS1B. Mutations were highly enriched between the T7 promoter and terminator regions (**Figure 2B** and **Supplementary Figure 1**) (36), including in the tac promoter sequence driving GWS1B expression (**Supplementary Table 2**), which lies immediately downstream of pT7. This pattern indicates that MutaT7 provided

efficient and targeted *gws1b* mutagenesis. As expected, transition mutations accounted for all non-synonymous mutations in the *gws1b* gene identified at ≥0.10% frequency, which was the limit of sequencing fidelity. Mutational frequencies at over 20 specific sites in *gws1b* rose substantially within the total population between rounds two and four of selection and generally continued to enrich with increasing selection pressure into round six, suggesting that these sites were key to selective advantages within MutaT7-RDE (**Supplementary Table 2**). Sequence alignment of GWS1B with diverse RuBisCO isoforms revealed that most selected mutations occurred at poorly conserved sites within the protein (**Supplementary Figure 2**), consistent with mutational tolerance in these regions that was exploited by MutaT7-RDE.

After establishing that targeted mutagenesis on *gws1b* and selection had occurred, we sought to characterize the most commonly occurring enzyme variants. Owing to the high rate of mutagenesis by MutaT7, we speculated that many GWS1B variants may have multiple epistatic mutations that could not be identified as coupled to each other using short-read NGS. We therefore turned to whole-plasmid sequencing results for clonal isolates from different rounds of evolution. Many isolates contained multiple non-synonymous mutations in the *gws1b* gene (**Figure 2C**), which accumulated as early as the first round of evolution. Many isolates also had synonymous mutations in *gws1b* as well as mutations in the promoter and plasmid replication elements (**Supplementary Table 3**). To remove selection artefacts from expression and copy number changes in the BAC plasmid, clonal isolates were subcloned into a clean pTrc backbone that is commonly used when screening error prone PCR libraries of RuBisCO in RDE (19, 20, 22, 28). In addition, a library of 'winner' alleles was subcloned into pTrc from the round 4 and 6 BAC winner plasmid pool and cloned into pTrc. The pTrc libraries were screened in RDE as previously described but under maximum selective pressure (0.2% arabinose) as the 'final' round of selection.

Evaluating the solubility, speed, and stability of GWS1B RuBisCO variants

We purified the 11 isolated GWS1B variants from the pTrc plasmid backbone for initial testing to identify those that should be pursued further for in-depth analysis. As a rapid checkpoint for this variant group, we used the purified enzymes for spectrophotometric determination of carboxylation activity (30, 43). Of the 11 variants isolated from MutaT7-RDE clonal screening, six had similar carboxylation rates to wild-type GWS1B, while five displayed significantly less carboxylation activity (**Supplementary Figure 3**). The inactivity of some selected variants indicates that there is an accessible selective advantage in RDE other than enhancing carboxylation kinetics. Directed evolution campaigns, including RDE-based selection

campaigns, often encounter false positives (21, 27, 42, 44) – for example, citing plasmid instability (27) or transposon insertion in *prk* (19, 20, 42). Fusion of PRK to NptII should eliminate false positives owing to transposon insertion or plasmid instability, but does not protect against ablation of PRK function via a point mutation. Indeed, several *prk* point mutations were identified in the sequenced pACYC^{BAD}PRK-NptII plasmid pools (**Supplementary Figure 4**). We speculate that these mutations reduce or destroy PRK phosphorylation activity while maintaining kanamycin resistance, meaning that escape mutants leading to loss of RuBisCO function can sometimes survive.

We proceeded with seven GWS1B variants; six that maintained carboxylation activity as well as the triple mutant GWS1B R172C/H237R/K446E, as precipitation of the triple mutant during purification led us to hypothesize that the purification process may have interfered with spectrophotometric measurement of its carboxylation activity. GWS1B E40K also contained the substitution H464R in the polyhistidine tag (**Supplementary Table 3**), potentially suggesting a selective advantage from tag modification. However, when we compared GWS1B E40K and GWS1B E40K/H464R using RDE spot testing (**Supplementary Figure 5**), we found that the two variants performed nearly identically, suggesting that the histidine tag substitution has a minimal effect on selection and activity.

We first evaluated whether these variants were selected in MutaT7-RDE due to increased production of soluble GWS1B. To probe GWS1B expression, we measured the percentage of soluble GWS1B relative to total cellular protein via ^{14}C-2-carboxyarabinitol-1,5-bisphosphate (^{14}C-CABP) binding (43) (**Figure 2D**). With the exception of GWS1B E40K, all variants had similar or reduced solubility relative to the wild-type enzyme in *E. coli*. These observations contrast with prior RuBisCO directed evolution campaigns that largely identified RuBisCO variants with improved solubility (20, 27, 28), thereby providing confidence in selection of active GWS1B variants based on enhanced kinetics rather than increased solubility. We also measured the temperature of aggregation (T_{agg}) of each variant as a proxy for stability (**Figure 2E**). Most of the purified variants displayed reduced stability relative to wild-type GWS1B, especially the triple mutant GWS1B R172C/H237R/K446E – consistent with its precipitation *in vitro*. Strikingly, there was no correlation between GWS1B variant stability and spectrophotometric activity (**Supplementary Figure 3**) (45). Some variants lost activity but had a similar T_{agg} to wild-type GWS1B, whereas other variants maintained activity at similar or greater levels to wild-type GWS1B while displaying a significantly lower T_{agg}. Finally, we probed the solubility of each variant in *E. coli* relative to the total cellular protein using SDS-PAGE,

revealing that none of the GWS1B variants meaningfully accumulated in the insoluble protein fraction (**Figure 2F**).

Comparing the kinetic properties of the evolved GWS1B RuBisCO mutants

Because GWS1B directed evolution was conducted under ambient O_2, a foreign environment for this RuBisCO, we hypothesized that the selective advantage of GWS1B variants in RDE could be attributable to maintaining enhanced carboxylation and reduced oxygenation specifically at ambient O_2 concentration. To test this hypothesis, we performed a full kinetic analysis of the GWS1B variants I45V/P153S, V294A, E40K, P86S, T29A, and R337C, as well as the triple mutant R172C/H237R/K446E (**Figures 3A**–**F** and **Supplementary Table 4**). The carboxylation rate of all variants remained similar to or decreased relative to wild-type GWS1B (**Figure 3A**), perhaps as expected owing to the already extremely high carboxylation rate of the wild-type enzyme that limits selection pressure on that parameter. $S_{C/O}$ (**Figure 3B**) also remained comparable or slightly decreased relative to wild-type GWS1B. Critically, however, the $K_{\mathrm{M}}^{\mathrm{CO_2}}$ $(21\%\ \mathrm{O_2})$ of nearly all variants was improved (**Figure 3C** and **Supplementary Figure 6**), likely owing to selection at ambient O_2.

We also derived $K_{\mathrm{M}}^{\mathrm{O_2}}$ (**Figure 3D** and **Supplementary Figure 6**) from our measurements, revealing that nearly all the selected GWS1B variants had a weaker affinity for oxygen than wild-type GWS1B, although this enhancement was usually accompanied by an increase in oxygenation rate. While $K_{\mathrm{M}}^{\mathrm{CO_2}}$ $(21\%\ \mathrm{O_2})$ was improved for most variants, $K_{\mathrm{M}}^{\mathrm{CO_2}}$ $(0\%\ \mathrm{O_2})$ was generally worse (**Supplementary Table 4**). As a result, the carboxylation efficiency of the measured variants was decreased in the absence of oxygen (**Figure 3E**). In contrast, substantial improvements to $K_{\mathrm{M}}^{\mathrm{CO_2}}$ $(21\%\ \mathrm{O_2})$ and $K_{\mathrm{M}}^{\mathrm{O_2}}$ drove a significant net improvement in the ambient carboxylation efficiency of GWS1B E40K (24% increase) and modest increases for GWS1B T29A and GWS1B R337C (**Figure 3F**).

To better understand the selection of improved variants *in vivo*, we calculated the carboxylation efficiency of GWS1B E40K, GWS1B T29A, and GWS1B R337C relative to the wild-type enzyme over a range of CO_2 concentrations at 0% and 21% O_2 (**Figure 3G**). As expected, these three variants outperformed the wild-type enzyme at ambient O_2, especially at low local CO_2. Meanwhile, wild-type GWS1B outperforms the variants over the entire CO_2 range at 0% O_2, suggesting that directed evolution in an oxygenic environment provides resistance to O_2 at the expense of anaerobic kinetics.

GWS1B I45V/P153S and GWS1B V294A had a slightly worse carboxylation efficiency in air relative to wild-type GWS1B. Since wild-type GWS1B expression hampered *E. coli* growth (**Figure 1E**), we hypothesized that selected variants with a decreased carboxylation efficiency may be less toxic to *E. coli* cells, leading to faster growth and selection. We performed overnight growth analyses of *E. coli* cultures to determine if expression of these variants affected cellular growth rate. Indeed, *E. coli* expressing GWS1B I45V/P153S and GWS1B V294A grew much more quickly than *E. coli* expressing wild-type GWS1B when IPTG was added (**Supplementary Figure 7**), supporting that their selection was due to reduction of toxicity caused by GWS1B expression. Interestingly, the unstable triple mutant GWS1B R172C, H237R, K446E experienced the largest improvement in both $K_{\mathrm{M}}^{\mathrm{CO_2}}$ $(21\%\ \mathrm{O_2})$ and $K_{\mathrm{M}}^{\mathrm{O_2}}$. Nonetheless, these changes were met with a marked decrease in $k_{cat}^{CO_2}$, leading to a decrease in ambient carboxylation efficiency. Its selection in RDE does not seem to be attributable to decreased toxicity (**Supplementary Figure 7**), but perhaps may be due to decreased glycolate production owing to its weak $K_{\mathrm{M}}^{\mathrm{O_2}}$ (40, 46).

Selected substitutions in GWS1B cluster near the active site and at protein–protein interfaces

The selected GWS1B variants all have amino acid substitutions either in the N-terminal portion of the protein near the active site (T29A, I45V; **Figure 4A**) or at protein–protein interfaces in GWS1B (E40K, P86S, P153S, R172C, H237R, V294A, K446E; **Figures 4B** and **4C**). The R337C substitution is the one exception to this trend, as it is positioned on the surface of the protein distal to both dimer and hexamer interfaces (**Figure 4B**). Notably, there was one GWS1B variant containing substitutions in both regions – the I45V/P153S variant. The prevalent selection of substitutions at protein–protein interfaces suggests a possible connection between GWS1B oligomerization state and O_2 sensitivity. To test this hypothesis, we performed native gel electrophoresis on purified variant enzymes. For both wild-type GWS1B and all the variants, we observed single bands at approximately 308 kDa, consistent with a hexameric state (**Supplementary Figure 8**). Thus, changes in catalytic parameters are not likely to be attributable to a substantial change to oligomeric state. In addition to the substitutions at protein–protein interfaces, the T29A substitution is buried in the protein, separated from the active site by an alpha-helix that positions it approximately 20 Å away (**Figure 4A**). It is possible that this amino acid change creates subtle differences in the active site architecture that lead to decreased availability to O_2 (16, 47).

Finally, to understand the selected substitutions across the RuBisCO family, we mapped them onto RuBisCO from *Nicotiana tabacum* and *R. rubrum*. Most amino acid loci aligned well to their respective functional region (active site or interface) in both *R. rubrum* and *N. tabacum* RuBisCO (**Supplementary Table 5**), suggesting possible transferability of kinetic properties or a generalizable selection mechanism. Additionally, we compared the variants with the *R. rubrum* RuBisCO deep mutational scanning analysis of Prywes et al. (2025), finding that 50% of the selected mutations in GWS1B occurred at loci where fitness was improved in *R. rubrum* (**Supplementary Table 5**), which is striking considering that only 0.14% of substitutions in *R. rubrum* were found to be fitness-improving (40). Taken together, these improvements to RuBisCO catalysis in GWS1B reveal multiple pathways to O_2 resistance across RuBisCO phylogeny and support the notion that laboratory-based *in vivo* directed evolution can push RuBisCO beyond perceived natural catalytic boundaries.

Discussion

In this work, we introduce an *in vivo* directed evolution platform for RuBisCO enzymes and use it to evolve RuBisCO from *Gallionellaceae*, yielding variants with improved carboxylation efficiency in air. This system builds upon years of foundational work on RuBisCO evolution by accelerating the duration of evolution rounds, reducing researcher intervention, and deeply sampling RuBisCO sequence space. The success of this evolution campaign is both a proof-of-concept for *in vivo* RuBisCO directed evolution and a demonstration of exceeding the known limits of RuBisCO catalysis.

Exerting an oxygenic selection pressure on a RuBisCO isoform isolated from a low-oxygen environment yielded improvements to $K_M^{CO_2}(21\%\ O_2)$ and $K_M^{O_2}$, leading to a substantial net increase of carboxylation efficiency in air for this normally highly O_2-sensitive RuBisCO. The specific substitutions found to provide these catalytic benefits have not been previously characterized and may be transferable across other RuBisCO isoforms to attain catalytic improvements. Indeed, photorespiration, caused by oxygenation of RuBP by RuBisCO, accounts for some 30% productivity loss in crops (3). Reducing photorespiration has already proven to be beneficial to plant growth (14, 48), meaning that amino acid substitutions in RuBisCO that reduce oxygenation efficiency without a concomitant loss in carboxylation efficiency can provide similar effects. This significance to crop growth, as well as the identified GWS1B variants that demonstrate that anaerobic and ambient carboxylation efficiencies are not intrinsically linked, emphasize that, going forward, the effects of O_2 on RuBisCO activity must be considered when undertaking any RuBisCO engineering campaign.

The mechanism by which the selected mutations provide O_2 resistance is not yet clear. Seven of the ten selected substitutions lie at either the dimer or hexamer interface of GWS1B, which is not the first instance of mutations at RuBisCO protein–protein interfaces providing O_2 resistance. In fact, mutations clustered at the interface between the large and small subunits during directed evolution of the Form I *Thermosynechococcus elongatus* RuBisCO caused an increase in carboxylation efficiency in air (20). Designed mutations that transition GWS1B from hexamer to dimer caused an increase in $S_{C/O}$ and a large increase in $K_M^{O_2}$, although these particular designed mutations were met with a decrease in $k_{cat}^{CO_2}$ that was detrimental to carboxylation efficiency (33). A recent deep mutational scan of Form II *R. rubrum* RuBisCO identified only a handful of single amino acid substitutions that could meaningfully improve enzyme affinity for CO_2, and the majority of those substitutions were at the dimer interface (40). These identified *R. rubrum* variants shared the same fate of a catastrophic decrease in $k_{cat}^{CO_2}$.

Therefore, there is accumulating evidence, including this work, that the protein–protein interfaces of RuBisCO enzymes across all forms are involved in modulating the balance between carboxylation and oxygenation. Importantly, the GWS1B variants identified here subvert the trade-off that aforementioned mutants have so far proven prey to, in that improvements to their substrate affinities are *not* accompanied by a decrease in carboxylation rate.

In conclusion, we deployed a novel *in vivo* directed evolution system to expand the catalytic limits of RuBisCO beyond what has been observed in nature, focusing on minimizing oxygenation while maintaining efficient carboxylation. Future use of MutaT7-RDE, as well as orthogonal protein engineering approaches, will continue to elucidate the RuBisCO sequence-structure-function landscape. Efficient RuBisCO enzymes engineered in the laboratory will provide key insights into the fundamental mechanisms of carboxylation and oxygenation and can be utilized as essential starting points to improve photosynthesis.

Materials and Methods

MutaT7-RDE workflow

PACYC^{BAD}PRK-NptII was transformed into Dual8 *E. coli* (36) to generate MutaT7-RDE. These cells were rendered chemically competent by treatment with ice cold 0.1 M $CaCl_2$. BAC-GWS1B was transformed by heat shock at 42 °C. A single colony of this strain was picked and resuspended in 5 mL Luria broth (LB) containing 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to induce mutagenesis via MutaT7. Mutagenesis cultures were grown for 24 h at 37 °C/180 rpm. 0.1 µL of this culture was plated on LB agar containing 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, and 50 µg/mL triphenyl tetrazolium chloride (TTC; Sigma-Aldrich) to confirm viability of the culture and colony forming units (cfu) per volume of culture. 250 µL aliquots of the remaining culture were plated on LB agar plates containing 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, 400 µg/mL kanamycin, 0.5 mM IPTG, 50 µg/mL TTC, and 0.025%–0.2% [w/v] of L-arabinose (Sigma-Aldrich) for selection via PRK induction. The concentration of arabinose was increased stepwise by evolutionary round in increments of 0.025% from round one to round four before increasing in increments of 0.05% from round five (0.15%) to six (0.2%). Selective plates were incubated at RT and 10% [v/v] CO_2 (g) in synthetic air (21% [v/v] O_2 (g) balanced with N_2 (g); Linde) for approximately 72 h or until colony growth was observed. Total bacterial mass on each plate was resuspended into 5–10 mL LB, pelleted at 5,000 x g for 10 min, and total plasmid DNA was extracted using a Wizard® *Plus* SV Miniprep kit (Promega). These extracts were submitted for next-generation sequencing as described below.

1 µg of remaining total plasmid DNA was digested with the restriction enzyme SfiI (New England Biolabs) for 2 h at 50 °C then cleaned using a Wizard® SV PCR clean up kit (Promega) and eluted into 15 µL of pure H_2O. 5 µL of purified DNA was electroporated into electrocompetent MutaT7-RDE and recovered for 1 h in 1 mL LB at 37 °C/180 rpm. 1 µL of recovery culture was plated on LB agar containing 100 µg/mL ampicillin to assess transformation efficiency and the remainder was diluted to 5 mL in LB to initiate the second round of mutagenesis as described above.

For non-mutagenic rescreening, the digested plasmid pool as prepared above was electroporated into DH10B (New England Biolabs) cells containing a fresh copy of pACYC^{BAD}PRK-NptII (DH10B-RDE). 50 µL of the recovered transformation was plated to the identical selective condition as when isolated from MutaT7-RDE. The fastest growing colonies were picked from each plate and were submitted for whole-plasmid sequencing.

To eliminate any confounding effects from promoter or plasmid replication mutations, we subcloned *rbcL* from the BAC Round 4 (0.1% arabinose selection) and Round 6 (0.15% arabinose selection) winner pools using primers 1 and 2 (**Supplementary Table 7**). The resulting library of PCR amplicons was digested using NcoI and SacI and ligated into pTrc to generate two libraries in the pTrc backbone. The libraries were screened for RDE growth at 0.2% arabinose in 10% [v/v] CO_2 in synthetic air (21% [v/v] O_2 balanced with N_2). The fastest-growing colonies were directly picked and sequenced. Isolates containing the nonsense and (sense) mutations encoding I45V, P153S (I18I, F251F), V294A (S50S, G359G) were isolated from the round 4 pTrc library. E40K (H464R, F176F) was isolated from the 0.15% library.

Whole-plasmid sequencing

Fast-growing colonies were picked from agar plates, resuspended in 5 mL of LB, and grown at 37 °C/180 rpm overnight. These cultures were pelleted, miniprepped, and submitted to Plasmidsaurus Inc. for sequencing. Output fasta files were aligned to BAC-GWS1B.

Next-generation sequencing

Mutated cultures after each round of evolution were plated on LB agar containing 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, and 50 µg/mL TTC and then grown at RT and 10% CO_2 for three d. The agar was scraped to collect all colonies and plasmids were isolated as described above. Concentrations of the plasmid pools were normalized to 20 ng/µL and samples were prepared using the Nextera DNA Flex kit (Illumina). Samples were sequenced in duplicate using a 150 paired-end kit on the Element Biosciences AVITI sequencer. Quality scores for each sample ranged between 30–45.

Output fastQ files were aligned to BAC-GWS1B and pACYC^{BAD}PRK-NptII using the Burrows-Wheeler Aligner (BWA) (49). Alignments were down-sampled to 10% to reduce file size, then were indexed, sorted, and a pileup file was generated using Samtools (50). Non-reference base frequencies at each position in the reference plasmid were calculated using Samtools and a custom Perl script. Mutational scatter plots were generated from these tables using Prism. Further analysis of high-percentage SNPs was completed with Geneious Prime. Mutations at 0.10% or higher frequency (determined using a read quality score of 30) in BAC-GWS1B that were present in duplicate sequencing of at least one round of evolution are listed in **Supplementary Table 2**. Full scripts for NGS processing are detailed at https://github.com/julielmcdonald/GWS1B.

Protein purification

GWS1B isolates were cloned into a pTrc high-copy plasmid backbone using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) and transformed into DH10B chemically competent cells (New England Biolabs). A colony from this transformation was picked and grown in 50 mL of LB + 100 µg/mL ampicillin for 16 h at 37 °C/180 rpm. 25 mL of this culture was used to inoculate 1 L of terrific broth (TB) + 100 µg/mL ampicillin in a baffled 3 L flask. The 1 L culture was grown at 37 °C/180 rpm until an optical density (OD) of 0.6–0.8 was reached, approximately 2.5 h. The temperature was reduced to 23 °C and GWS1B expression was induced by addition of 0.5 mM IPTG to the culture. Induction continued for 16 h, then the culture was transferred to 4 °C. The cells were pelleted at 4 °C, the supernatant was discarded, and the pellet was scraped and frozen at -80 °C for future use.

The bacterial pellet was thawed on ice, then resuspended in approximately 20 mL of 20 mM HEPES at pH 8.0 containing 50 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF; Amersco), 5 mM DL-Dithiothreitol (DTT; Sigma-Aldrich), and 1 µL of DNase (Sigma-Aldrich). The resuspension was lysed using a needle-tipped sonicator at 50% amplitude, 20 s on, 20 s off for three cycles. The insoluble fraction was separated by centrifugation at 20,000 x g for 20 min. The pellet was discarded and the soluble lysate (supernatant) was filtered through a 0.22 µm filter. The soluble lysate was loaded onto a column packed with 1 mL HisPur™ Cobalt Resin (Thermo Fisher Scientific) and equilibrated with 20 mM HEPES at pH 8.0 containing 50 mM NaCl. The column was washed with 5 column volumes (CV) of 20 mM HEPES at pH 8.0 containing 50 mM NaCl and 20 mM imidazole (Alfa Aesar). GWS1B proteins were eluted with 2 CV of 20 mM HEPES at pH 8.0 containing 50 mM NaCl and 500 mM imidazole. The eluted protein was dialyzed overnight against 20 mM HEPES at pH 8.0 containing 50 mM NaCl, concentrated using a Amicon® Ultra Centrifugal Filter (Sigma-Aldrich) with a 100 kDa molecular weight cutoff, and flash frozen for future use. Purity and oligomeric state were evaluated using Native PAGE (**Supplementary Figure 8**).

SDS-PAGE, Western Blotting, and Native PAGE

For SDS-PAGE and western blotting of protein lysates, 10 mL of culture grown as in the "Protein purification" section above was resuspended in 1 mL of 20 mM HEPES at pH 8.0 containing 50 mM NaCl, 1 mM PMSF, 5 mM DTT, and 1 µL of DNase, and was lysed by sonication using a needle-tipped sonicator at 20% amplitude, 20 s on, 20 s off for two cycles. 100 µL of total protein fraction was saved, and the insoluble fraction was then separated by centrifugation at 20,000 x g for 20 min. The supernatant was taken as the soluble protein

fraction. Total protein content of both fractions was determined using the Pierce™ Bradford Protein Assay Kit (Thermo Fisher Scientific). 10 µg of total or soluble protein was loaded onto a NuPAGE™ Bis-Tris Mini Protein Gels, 4–12% (Thermo Fisher Scientific). The gel was run in NuPAGE™ MES SDS Running Buffer (Thermo Fisher Scientific) for 30 min at 200 V.

For staining via Western blotting, the protein was transferred onto nitrocellulose membranes (Hybond C, APBiotech) using an Xcell transfer cell (Novex) according to the manufacturer's specifications. Immunoblotting was performed as previously described (51), using a *R. rubrum* RuBisCO-reactive antibody. Immunoreactive proteins were visualized using AttoPhos reagent (Promega) with a Vistra FluorImager.

For native PAGE, purified GWS1B proteins were thawed on ice and diluted to 1 µg/mL in 20 mM HEPES at pH 8.0 containing 50 mM NaCl. 10 µg of protein was loaded onto a Novex™ Tris-Glycine 4–12% Mini Protein Gel (Thermo Fisher Scientific). The gel was run in NativePAGE™ Running Buffer (Thermo Fisher Scientific) at 120 V for 2.5 h at 4 °C, then was stained with Coomassie InstantBlue® Protein Stain (Abcam). 10 µg of bovine thyroglobulin (Sigma-Aldrich) and 10 µg of bovine serum albumin (Sigma-Aldrich) were used as molecular weight markers.

Growth curves

pTrc-GWS1B plasmids were transformed into DH10B chemically competent cells (New England Biolabs). A single colony was used to inoculate 5 mL of LB containing 100 µg/mL ampicillin and was incubated overnight at 37 °C/180 rpm. This culture was used to inoculate 200 µL of LB to an OD of 0.005 in a 96-well plate. OD was measured at 600 nm for 16 h in an Agilent BioTek Synergy HTX plate reader. Data reported are means of normalized triplicate growth curves.

RDE spot testing

DH10B cells containing pACYC^{BAD}PRK-NptII and pTrc-GWS1B isolates were streaked on LB agar containing 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, and 50 µg/mL TTC and grown at 37 °C until colonies appeared. A colony was picked and resuspended into 1 mL of LB containing 100 µg/mL ampicillin and 35 µg/mL chloramphenicol. This culture was incubated at 37 °C/180 rpm until slightly cloudy, to an OD of ~0.2. OD was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific). The culture was diluted to an OD of 0.1 by adding an appropriate volume of LB and then was diluted once more 1:1000 to a final OD of .0001. 20 µL of the final diluted culture was spotted onto LB agar containing 100 µg/mL

ampicillin, 35 µg/mL chloramphenicol, 50 µg/mL TTC, 0.5 mM IPTG, 400 µg/mL kanamycin, and increasing concentrations (0% to 0.3%) of arabinose for selection. The plates were incubated at RT and 2% CO_2 (g). Growth was monitored every other day for 10 d.

^{14}C-CABP binding analysis

pTrc-GWS1B plasmids were transformed into DH10B chemically competent cells (New England Biolabs). A single colony was used to inoculate 5 mL of LB containing 100 µg/mL ampicillin and was incubated overnight at 37 °C/180 rpm. 300 µL of saturated overnight culture was used to inoculate 15 mL of LB containing 100 µg/mL ampicillin in a 50 mL baffled flask. This culture was grown to an OD of 0.6–0.8 at 37 °C/180 rpm (~2.5 h). GWS1B expression was induced by adding IPTG to a final concentration of 0.5 mM. Incubation temperature was reduced to 23 °C and cells were grown for 16 h. The culture was then pelleted and used immediately or stored at -80 °C.

Bacterial pellets were thawed on ice, then resuspended in 500 µL of 50 mM EPPS-NaOH at pH 8.0 containing 15 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA, Sigma-Aldrich), 1 mM PMSF, 5 mM DTT, 1 µL of DNase, and 10 mM $NaHCO_3$ (Sigma-Aldrich) for RuBisCO activation. The resuspension was lysed using a needle-tipped sonicator at 50% amplitude for 20 s. The insoluble fraction was pelleted by centrifugation at 4 °C, 20,000 x g for 15 min. Total protein concentration of the soluble fraction was determined using the Pierce™ Bradford Protein Assay Kit. RuBisCO active site content was quantified using ^{14}C-CABP binding, as previously described (43, 52).

Spectrophotometric determination of RuBisCO activity

Spectrophotometric determination of RuBisCO enzymatic activity was conducted using a nicotinamide adenine dinucleotide (NAD^+/NADH) reduction assay that has been previously described (30, 43). Briefly, assay buffer was prepared with 100 mM EPPS-NaOH at pH 7.6 containing 25 mM $MgCl_2$ and 1 mM EDTA. A mixture of coupling enzymes containing creatine phosphokinase (Sigma-Aldrich), carbonic anhydrase (Sigma-Aldrich), 3-phosphoglycerate kinase (PGK; Sigma-Aldrich), triose-phosphate isomerase (TIM; Sigma-Aldrich), Glycerol-3-phosphate dehydrogenase (GAPDH; Sigma-Aldrich), and glyceraldehyde-3-phosphate dehydrogenase (GlyPDH; Sigma-Aldrich) to perform the reduction reaction was prepared as described (43).

Purified GWS1B proteins were diluted to approximately 1 mg/mL in assay buffer and their concentration was measured in triplicate by absorbance at 280 nm using a NanoDrop

spectrophotomer. The sample was then diluted 1:500 into assay buffer containing 10 mM $NaHCO_3$, 1 mM ATP/5 mM P-creatine (Sigma-Aldrich), 0.2 mM NADH (Sigma-Aldrich), and 1:50 coupling enzyme mix. 195 µL of this mixture was added in triplicate to a 96-well plate. 5 µL of 20 mM purified RuBP (prepared as below) was added to each well using a multichannel pipette to start the reaction. Absorbance at 340 nm of each well was monitored for 20 min at 53 s intervals at 30 °C. The slope of A340 over time was used to calculate the nanomoles of RuBP consumed (s^{-1}) and the corresponding RuBisCO carboxylation rate (s^{-1}) (43).

Kinetic characterization of GWS1B variants

RuBP was synthesized and purified following established methods (53). Cell cultures that were lysed for ^{14}C-CABP binding analysis were used to measure $k_{cat}^{CO_2}$ and $K_M^{CO_2}$ with described $^{14}CO_2$-fixation assays (20) with the following adaptations for GW1SB: RuBisCO enzymes were activated with 10 mM $NaH^{14}CO_3$ and a range of 0–300 µM $^{14}CO_2$ was used to determine $K_M^{CO_2}$. $K_M^{CO_2}$ was measured in buffer equilibrated with 0%, 10%, or 20% O_2 (v/v) in N_2. The slope of $K_M^{CO_2}$ vs. [O_2] was used to determine $K_M^{O_2}$, following the equation:

$$K_{M\ app}^{CO_2} = K_M^{CO_2}\ (1 + \frac{[O_2]}{K_M^{O_2}})$$

where $K_{M\ app}^{CO_2}$ is equal to $K_M^{CO_2}$ at varying concentrations of O_2 (22). $S_{C/O}$ was quantified using a previously described [1-^{3}H]-RuBP consumption assay (53) using purified GWS1B. $k_{cat}^{O_2}$ was calculated using the following equation for $S_{C/O}$:

$$S_{\mathrm{C/O}} = \frac{k_{cat}^{CO_2}\ x\ K_M^{O_2}}{k_{cat}^{O_2}\ x\ K_M^{CO_2}}$$

Acknowledgments

This work was supported by the National Science Foundation's Division of Molecular and Cellular Biosciences (EAGER Grant 2244770) to M.D.S., the Abdul Latif Jameel Water and Food Systems Lab Grand Challenge Grant to M.D.S., R.H.W., and S.W. and the Martin Family Society Fellowship for Sustainability to J.L.M. We extend our gratitude to the MIT BioMicroCenter staff for their assistance with high-throughput sequencing experiments. We thank Dr. Jiacheng Lin for his helpful comments and editing of the manuscript and Sarah Kaines for technical assistance.

References

1. L. Schulz *et al.*, Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155-160 (2022).
2. A. I. Flamholz *et al.*, Revisiting trade-offs between rubisco kinetic parameters. *Biochem.* **58**, 3365-3376 (2019).
3. B. J. Walker, A. VanLoocke, C. J. Bernacchi, D. R. Ort, The costs of photorespiration to food production now and in the future. *Annu. Rev. Plant Biol.* **67**, 107-129 (2016).
4. N. J. Claassens *et al.*, Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. *Proc. Natl. Acad. Sci.* **117**, 22452-22461 (2020).
5. E. G. Nisbet *et al.*, The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiol.* **5**, 311-335 (2007).
6. T. J. Erb, J. Zarzycki, A short history of RubisCO: The rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr. Opin. Biotech.* **49**, 100-107 (2018).
7. Y. Savir, E. Noor, R. Milo, T. Tlusty, Cross-species analysis traces adaptation of rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci.* **107**, 3475-3480 (2010).
8. P. M. Shih *et al.*, Biochemical characterization of predicted Precambrian RuBisCO. *Nat. Commun.* **7**, 10382 (2016).
9. M. Esparza, E. Jedlicki, C. González, M. Dopson, D. S. Holmes, Effect of CO_2 Concentration on Uptake and Assimilation of Inorganic Carbon in the Extreme Acidophile *Acidithiobacillus ferrooxidans*. *Front. microbiol.* **10** (2019).
10. J. Zhang *et al.*, Discovery of a readily heterologously expressed Rubisco from the deep sea with potential for CO_2 capture. *Bioresour. Bioprocessing* **8**, 86 (2021).
11. R. E. Sharwood, O. Ghannoum, S. M. Whitney, Prospects for improving CO_2 fixation in C_3-crops through understanding C_4-Rubisco biogenesis and catalytic diversity. *Curr. Opin. Plant Biol.* **31**, 135-142 (2016).
12. R. E. Sharwood, Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phytol.* **213**, 494-510 (2017).
13. X.-G. Zhu, D. R. Ort, M. A. Parry, S. von Caemmerer, A wish list for synthetic biology in photosynthesis research. *J. Exp. Bot.* **71**, 2219-2225 (2020).
14. A. P. Cavanagh, P. F. South, C. J. Bernacchi, D. R. Ort, Alternative pathway to photorespiration protects growth and productivity at elevated temperatures in a model crop. *Plant Biotech. J.* **20**, 711-721 (2022).
15. M. Gionfriddo, T. Rhodes, S. M. Whitney, Perspectives on improving crop Rubisco by directed evolution. *Semin. Cell Dev. Biol.* **155**, 37-47 (2023).
16. N. Prywes, N. R. Phillips, O. T. Tuck, L. E. Valentin-Alvarado, D. F. Savage, Rubisco function, evolution, and engineering. *Annu. Rev. Biochem.* **92**, 385-410 (2023).
17. K. D. Cocon, P. Luis, The potential of RuBisCO in CO_2 capture and utilization. *Prog. Energy Combust. Sci.* **105**, 101184 (2024).
18. P. Durão *et al.*, Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* **11**, 148-155 (2015).
19. R. H. Wilson, H. Alonso, S. M. Whitney, Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci Rep* **6**, 22284 (2016).
20. R. H. Wilson, E. Martin-Avila, C. Conlan, S. M. Whitney, An improved *Escherichia coli* screen for rubisco identifies a protein–protein interface that can enhance CO_2-fixation kinetics. *J. Biol. Chem.* **293**, 18-27 (2018).
21. Y. Zhou, S. Whitney, Directed evolution of an improved rubisco; *in vitro* analyses to decipher fact from fiction. *Int. J. Mol. Sci.* **20**, e5019 (2019).

22. O. Mueller-Cajar, M. Morell, S. M. Whitney, Directed evolution of rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochem.* **46**, 14067-14074 (2007).
23. S. Satagopan, S. S. Scott, T. G. Smith, F. R. Tabita, A Rubisco mutant that confers growth under a normally "inhibitory" oxygen concentration. *Biochem.* **48**, 9076-9083 (2009).
24. S. Satagopan, F. R. Tabita, RubisCO selection using the vigorously aerobic and metabolically versatile bacterium *Ralstonia eutropha*. *FEBS J.* **283**, 2869-2880 (2016).
25. S. Satagopan, K. A. Huening, R. F. Tabita, Selection of cyanobacterial (*Synechococcus* sp. Strain PCC 6301) RubisCO variants with improved functional properties that confer enhanced CO_2-dependent growth of *Rhodobacter capsulatus*, a photosynthetic bacterium. *mBio* **10** (2019).
26. S. A. Smith, F. R. Tabita, Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) Ribulose-1,5-bisphosphate Carboxylase/Oxygenase. *J. Mol. Biol.* **331**, 557-569 (2003).
27. M. R. Parikh, D. N. Greene, K. K. Woods, I. Matsumura, Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E. coli*. *Protein Eng. Des. Sel.* **19**, 113-119 (2006).
28. O. Mueller-Cajar, S. M. Whitney, Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. *Biochem.* **414**, 205-214 (2008).
29. J. Galmés *et al.*, Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ.* **37**, 1989-2001 (2014).
30. D. Davidi *et al.*, Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.* **39**, e104081 (2020).
31. B. de Pins *et al.*, A systematic exploration of bacterial form I rubisco maximal carboxylation rates. *EMBO J.* **43**, 3072-3083 (2024).
32. V. A. Varaljay *et al.*, Functional metagenomic selection of ribulose 1,5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environ. Microbiol.* **18**, 1187-1199 (2016).
33. A. K. Liu *et al.*, Structural plasticity enables evolution and innovation of RuBisCO assemblies. *Sci. Adv.* **8**, eadc9440 (2022).
34. B. Álvarez, M. Mencía, V. de Lorenzo, L. Á. Fernández, *In vivo* diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.* **11**, 6436 (2020).
35. H. Park, S. Kim, Gene-specific mutagenesis enables rapid continuous evolution of enzymes *in vivo*. *Nucleic Acids Res.* **49**, e32 (2021).
36. A. A. Mengiste *et al.*, Expanded MutaT7 toolkit efficiently and simultaneously accesses all possible transition mutations in bacteria. *Nucleic Acids Res.* **51**, e31-e31 (2023).
37. D. Seo, B. Koh, G.-e. Eom, H. W. Kim, S. Kim, A dual gene-specific mutator system installs all transition mutations at similar frequencies *in vivo*. *Nucleic Acids Res.* **51**, e59 (2023).
38. A. A. Mengiste *et al.*, MutaT7^{GDE}: A single chimera for the targeted, balanced, efficient, and processive installation of all possible transition mutations *in vivo*. *ACS Synth. Biol.* **13**, 2693-2701 (2024).
39. C. L. Moore, L. J. Papa, M. D. Shoulders, A Processive protein chimera introduces mutations across defined DNA regions *in vivo*. *J. Am. Chem. Soc.* **140**, 11560-11564 (2018).
40. N. Prywes *et al.*, A map of the rubisco biochemical landscape. *Nature*, 1-6 (2025).
41. R. S. Molina *et al.*, *In vivo* hypermutation and continuous evolution. *Nat. Rev. Methods Primers* **2**, e36 (2022).

42. R. H. Wilson, S. M. Whitney, "Improving CO_2 Fixation by Enhancing Rubisco Performance" in Directed Enzyme Evolution: Advances and Applications, M. Alcalde, Ed. (Springer International Publishing, 2017), pp. 101-126.
43. D. S. Kubien, C. M. Brown, H. J. Kane, Quantifying the amount and activity of Rubisco in leaves. *Methods Mol. Biol.* **684**, 349-362 (2011).
44. Z. Cai, G. Liu, J. Zhang, Y. Li, Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of rubisco. *Protein & Cell* **5**, 552-562 (2014).
45. R. A. Studer, P. A. Christin, M. A. Williams, C. A. Orengo, Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc. Natl. Acad. Sci.* **111**, 2223-2228 (2014).
46. M. T. Pellicer, M. F. Nuñez, J. Aguilar, J. Badia, L. Baldoma, Role of 2-Phosphoglycolate phosphatase of *Escherichia coli* in metabolism of the 2-phosphoglycolate formed in DNA repair. *J. Bacteriol.* **185**, 5815-5821 (2003).
47. C. Bathellier *et al.*, Ribulose 1,5-bisphosphate carboxylase/oxygenase activates O_2 by electron transfer. *Proc. Natl. Acad. Sci.* **117**, 24234-24242 (2020).
48. P. F. South, A. P. Cavanagh, H. W. Liu, D. R. Ort, Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* **363**, eaat9077 (2019).
49. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
50. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
51. S. M. Whitney, P. Baldet, G. S. Hudson, T. J. Andrews, Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* **26**, 535-547 (2001).
52. S. M. Whitney, R. E. Sharwood, "Plastid Transformation for Rubisco Engineering and Protocols for Assessing Expression" in Chloroplast Biotechnology: Methods and Protocols. (Humana Press, 2014), pp. 245-262.
53. H. Kane *et al.*, An improved method for measuring the CO_2/O_2 specificity of Ribulose bisphosphate Carboxylase-Oxygenase. *Funct. Plant Biol.* **21**, 449 (1994).

FIGURE 1

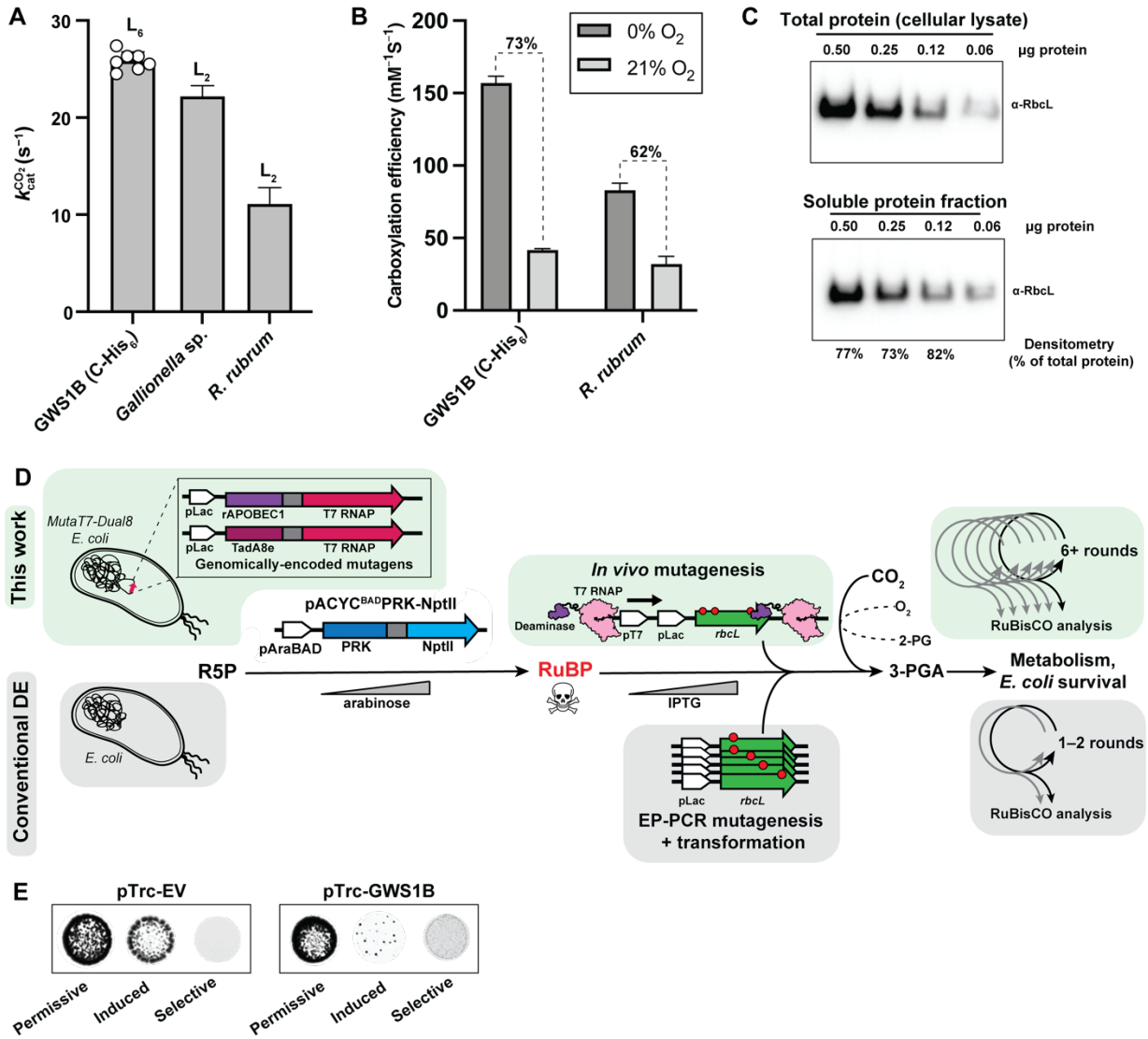

Figure 1. Biochemistry of GWS1B RuBisCO and potential for *in vivo* directed evolution in *E. coli*. A. Carboxylation rates for C-terminally-tagged GWS1B RuBisCO (this work), as well as for *Gallionella* sp. RuBisCO (30) and *R. rubrum* RuBisCO (22, 40). **B.** Carboxylation efficiencies at 0% O_2 and 21% O_2 for GWS1B and *R. rubrum* RuBisCO (22). **C.** Western blot analysis of RuBisCO levels detected in total protein lysate (top) compared to the soluble protein fraction (bottom) when GWS1B is expressed from the pTrc-GWS1B plasmid, revealing the very high soluble expression of GWS1B. **D.** Traditional RuBisCO directed evolution workflow versus targeted mutagenesis-mediated *in vivo* directed evolution workflow. *In vivo* directed evolution is enabled by in-cell, MutaT7-based mutagenesis rather than by extracellular, error-prone PCR (EP-PCR) mutagenesis followed by transformation into *E. coli* cells. MutaT7-based mutagenesis, enabled by a deaminase-T7 RNA polymerase (T7 RNAP) fusion binding to and processing from the T7 RNAP promoter site (pT7) on *rbcL*, generates high levels of downstream mutations in the plasmid DNA and avoids the need for step-wise generation of mutations and transformation, enabling deeper mutagenesis and more rapid progress between rounds of directed evolution. In both workflows, an exogenous *prk* gene is expressed to create a circuit where *E. coli* survival is dependent on RuBisCO-catalyzed carboxylation and detoxification of RuBP. R5P = Ribulose-5-phosphate; RuBP = Ribulose-1,5-bisphosphate; 2-PG = 2-phosphoglycolate; 3-PGA = 3-

phosphoglycerate. **E.** RDE testing of GWS1B against an empty vector (pTrc-EV) control. Permissive growth media contains no additives, induced growth media contains 0.5 mM IPTG for expression of RuBisCO, and selective growth media contains 0.5 mM IPTG, kanamycin at 400 mg/mL, and 0.15% arabinose for expression of PRK-NptII.

FIGURE 2

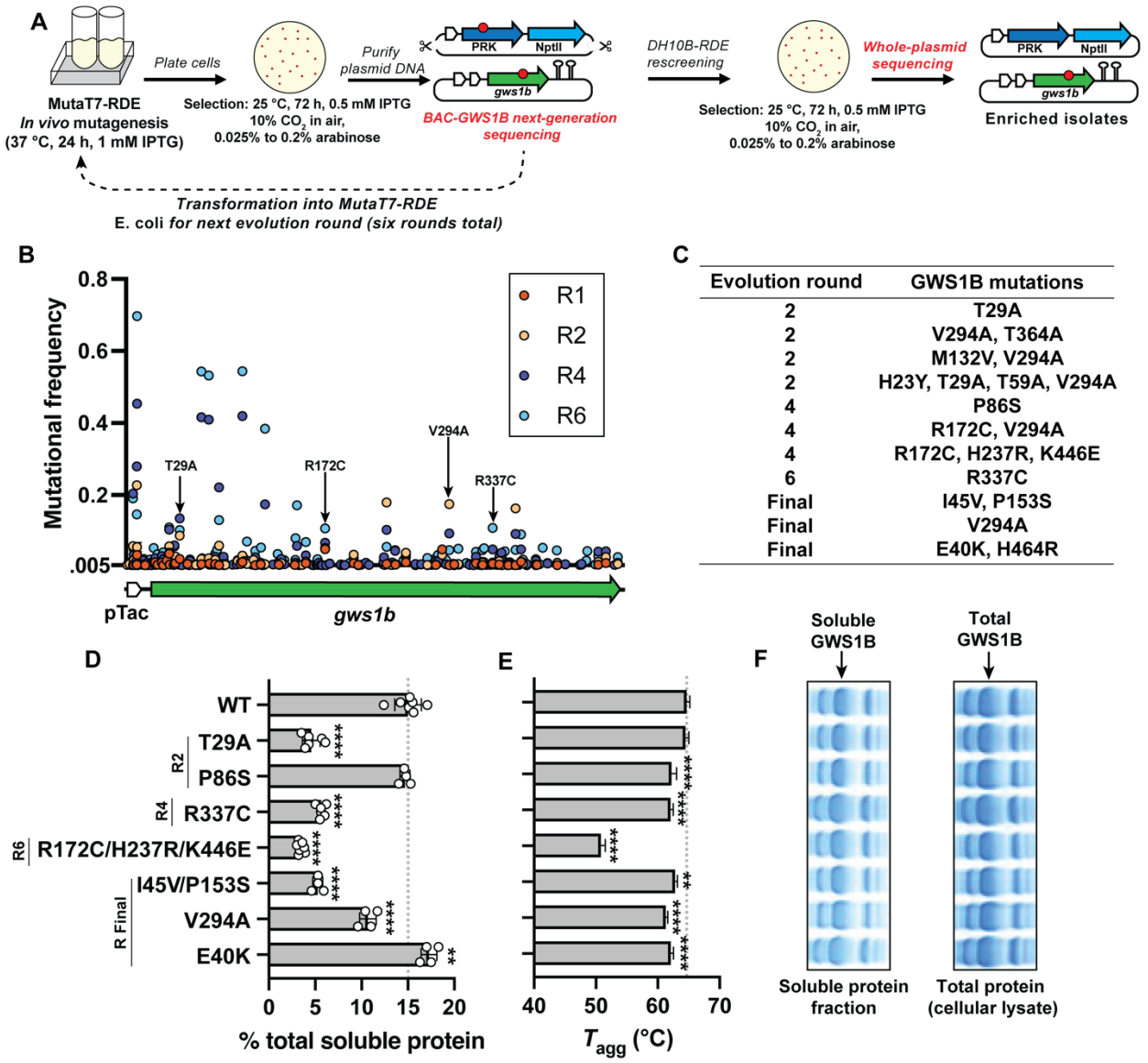

Figure 2. Identification and biochemical characterization of GWS1B variants: A. Schematic of MutaT7-RDE workflow. **B.** Map of the promoter and GWS1B coding region of the BAC evolution plasmid showing mutational frequency at the DNA base level during sequenced rounds of evolution. Non-synonymous mutations from clonal isolates are indicated with arrows **C.** Non-synonymous mutations found in clonal isolates of GWS1B in rounds 1, 2, 4, 6, and final round of evolution. **D.** Soluble, folded RuBisCO content of selected variants in *E. coli* quantified by [^{14}C]-CABP binding (43). Data are the means and standard deviations of three replicates with significance shown relative to wild-type GWS1B as analyzed by one-way ANOVA. **, $p \leq .005$; ***, $p \leq .0005$; ****, $p \leq .00005$. **E.** Aggregation temperature (T_{agg}) of selected variants. Significance testing and labels are consistent with (**D**). **F.** SDS-PAGE gels showing soluble and total protein fractions upon expression of GWS1B variants in *E. coli*.

FIGURE 3

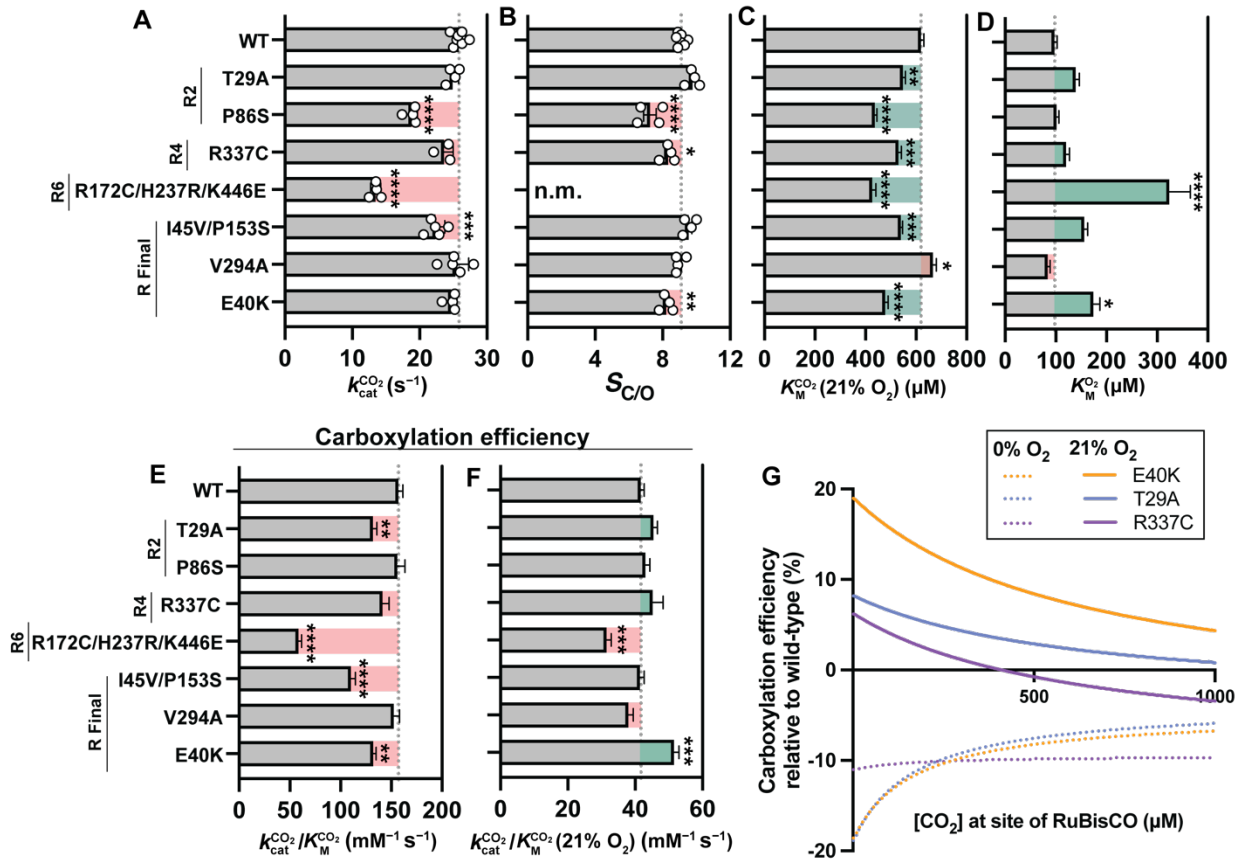

Figure 3. Kinetic characterization of GWS1B variants: A. Carboxylation rate. **B.** Specificity factor. **C.** Affinity for CO_2 in air, determined through extrapolation of O_2 vs. $K_M^{CO_2}$. **D.** Affinity for O_2. **E.** Carboxylation efficiency at 0% O_2. **F.** Carboxylation efficiency at 21% O_2. **G.** Carboxylation efficiency of GWS1B E40K, GWS1B T29A, and GWS1B R337C relative to wild-type GWS1B at 0% or 21% O_2 as a function of CO_2 concentration.

FIGURE 4

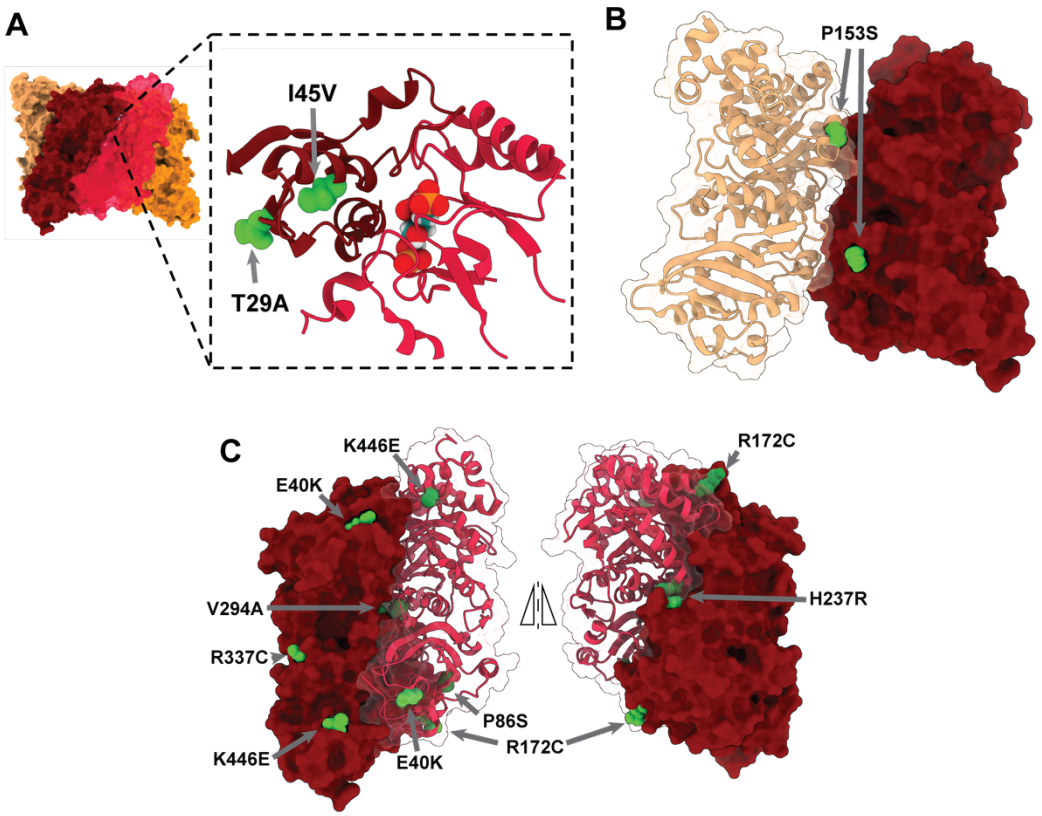

Figure 4. Mapping evolved variants onto GWS1B: A. Substitutions T29A and I45V are positioned near the GWS1B active site. PDBID: 5C2G (32). **B.** P153S occurs at the hexamer interface of GWS1B. **C.** E40K, P86S, R172C, H237R, V294A, and K446E occur at the dimer interface. R337C is surface-exposed.