JBC Papers in Press. Published on October 6, 2017 as Manuscript M117.810861
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M117.810861

An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO_2-fixation kinetics.

Robert H. Wilson, Elena Martin-Avila, Carly Conlan, Spencer M. Whitney

From the Research School of Biology, The Australian National University, Acton, Australian Capital Territory 2601, Australia

Running Title: Novel solutions for improving Rubisco catalysis

To whom correspondence should be addressed: Spencer M. Whitney, Research School of Biology, Australian National University, Acton, Australian Capital Territory 2601, Australia; Tel: +61-2-6125-5073; E-mail: spencer.whitney@anu.edu.au

Keywords: Rubisco, carbon fixation, photosynthesis, chloroplast, protein engineering

An overarching goal of photosynthesis research is to identify how components of the process can be improved to benefit crop productivity, global food security and renewable energy storage. Improving carbon fixation has mostly focused on enhancing the CO_2 fixing enzyme Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). This grand challenge has mostly proved ineffective due to catalytic mechanism constraints and required chaperone complementarity that hinder Rubisco biogenesis in alternative hosts. Here we refashion *Escherichia coli* metabolism by expressing a phosphoribulokinase-neomycin phosphotrans-ferase fusion protein to produce a high fidelity, high throughput Rubisco directed evolution (RDE2) screen that negates false positive selection. Successive evolution rounds using the plant-like *Te*-Rubisco from the cyanobacterium *Thermosynechococcus elongatus* BP1 identified two large subunit and six small subunit mutations that improved carboxylation rate, efficiency and specificity. Structural analysis revealed the amino acids clustered in an unexplored subunit interface of the holoenzyme. To study its effect on plant growth the *Te*-Rubisco was transformed into tobacco by chloroplast transformation. As previously seen for *Synechocccus* PCC6301 Rubisco, the specialized folding and assembly requirements of *Te*-Rubisco hinder its heterologous expression in leaf chloroplasts. Our findings suggest that the ongoing efforts to improve crop photosynthesis by integrating components of a cyanobacteria CO_2-concentrating mechanism will necessitate co-introduction of the ancillary molecular components required for Rubisco biogenesis.

Improving carbon fixation in agriculture has mostly focused on enhancing the activity of the CO_2-fixing enzyme Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) by modifying the enzyme itself or increasing the CO_2-levels around it (1). Improving the kinetics of Rubisco has proved challenging due to its complex catalytic mechanism of fixing CO_2 to ribulose-1,5-bisphosphate (RuBP) and cleaving it into two 3-phosphoglycerate (3-PGA) molecules (2-4). The carboxylation rate is slow (k_{cat}^{C}, ~1 to 5 reactions per second in plants) and often confines flux through the Calvin-Benson-Bassham (CBB) cycle, and hence limits the rate of photosynthesis and plant growth. Because of this limitation large amounts of Rubisco are needed to support adequate CO_2-assimilation rates (5-7). This results in Rubisco being the Earth's most abundant protein (8). Further encumbering its performance, Rubisco also catalyzes RuBP oxygenation to produce 3-PGA and 2-phosphoglycolate (2-PG). The recycling of 2-PG back into 3-PGA via photorespiration is considered wasteful as it consumes energy and releases fixed CO_2 (9).

The potential for successfully improving the carboxylation properties of Rubisco is spurred on by findings that crop Rubisco is not the pinnacle of evolution, with kinetically more efficient Rubisco forms found in some red algae (1, 3, 6, 10). Evolutionary adaptation of Rubisco kinetics appears to have been constrained by the complexity of its catalytic chemistry and multiplexed subunit folding and assembly requirements (3, 11-13). The need to retain complementarity with ancillary proteins involved

in the biogenesis and metabolic repair of Rubisco appears particularly pertinent for Form I Rubisco that comprises eight RbcL and eight RbcS subunits to form a L_8S_8 complex. Examples of Rubisco dedicated components include the assembly chaperones BSD2, RbcX, Rubisco accumulation factors 1 and 2 (Raf1, Raf2) and the metabolic repair protein Rubisco activase (14). The protein folding chaperonins of bacteria (GroEL), chloroplasts (Cpn60) and their protein co-factors GroES (bacteria) and Cpn10/Cpn20 (chloroplasts) are also essential components in Rubisco biogenesis (15). The need for Rubisco to maintain complementarity with these varied chaperone and chaperonin components appears to limit the span of Rubisco isoforms that can be produced in *E. coli* as well as those that can be bioengineered in the chloroplasts of vascular plants. For example, the chloroplasts of the model C_3-plant tobacco can produce high levels of the RbcS-lacking bacterial Form II *R. rubrum* L_2 Rubisco and *Methanococcoides burtonii* archaeal L_{10} Rubisco (~10 to 25 µmol active sites.m^2) due to their simple biogenesis requirements (16, 17). Differences in the biogenesis requirements of L_8S_8 Rubisco from red algae and monocot plants however preclude their assembly in tobacco and *E. coli* (10, 18). By contrast the assembly requirements of *Synechococcus elongatus* PCC6301 L_8S_8 Rubisco (*Se*-Rubisco) are partially met in chloroplasts (~5 µmol active sites sites.m^2; (19)) and in *E. coli*, with ~98% of the *Se*-RbcL produced forming insoluble, misfolded aggregates in the bacterium (20).

To avoid the constraints a photosynthetic environment may pose on the catalytic evolution of Rubisco, modern laboratory evolution applications have made particular use of Rubisco dependent *E. coli* (RDE) screens (12, 21, 22). An elegant advance has been the extensive re-wiring of *E. coli* metabolism to incorporate a non-native CBB cycle where cell survival can be made dependent on CO_2-fixation by Rubisco (23). This contrasts with conventional RDE screens that simply express recombinant phosphoribulokinase (PRK) to produce ribulose-1,5-bisposphate (RuBP), the 5-carbon substrate of Rubisco. As summarized in Figure 1A, RuBP is unexplainably toxic to *E. coli* such that the rate of colony growth in RDE screens is dependent on the level of Rubisco activity. This type of selection has proved useful for evolving *M. burtonii* L_{10} Rubisco mutants with desired improvements in carboxylation rate (k_{cat}^c), CO_2-affinity (lower K_m for CO_2, K_c) and specificity for CO_2 over O_2 ($S_{c/o}$) (17). Other attempts to improve the catalysis of L_8S_8 Rubisco using RDE screens have only identified RbcL mutations that enhance the subunit's folding and assembly with RbcS into a functional L_8S_8 complex (*i.e.* improving Rubisco solubility) (22).

A feature of RDE screens that limit their throughput and reliability (*i.e.* selection fidelity) is the high proportion of false positives selected (24, 25). These arise from the inactivation of PRK function by transposon integration (Figure 1A, (22)). Here we present the development and implementation of a simple, faster, high throughput screen called RDE2 that specifically selects for improvements in Rubisco activity, not PRK inactivation mutants. By eliminating false positive selection we demonstrate the versatility of the RDE2 screen in selecting for mutants of *Thermosynechococcus elongatus* BP-1 (*Te*-) Rubisco that improve RbcL folding and assembly in *E. coli* (*i.e. Te*-Rubisco solubility), the enzyme's carboxylation kinetics, or both. We subsequently reveal, using chloroplast genome transformation, how the specialized folding and assembly requirements of *Te*-Rubisco hinder its translational testing in tobacco (*Nicotiana tabacum*). The outcomes demonstrate that Rubisco biogenesis in *E. coli* is not a reliable proxy for expression in chloroplasts. As a consequence, deciphering the crucial molecular partnerships required for Rubisco biogenesis is needed to optimize heterologous Rubisco expression in crop chloroplasts as well as adapting the RDE2 for future evolution of eukaryotic Form I Rubisco.

RESULTS

RDE2 screen development. Rubisco directed evolution studies that have utilized either RDE or photosynthetic mutant screens have almost exclusively selected for one or more amino acid substitutions in RbcL that improve Rubisco biogenesis (solubility), not catalysis (12, 22). Impeding success are the poor transformation efficiencies of photosynthetic hosts and the high proportion of false positives selected in RDE screens. As summarized in Figure 1A, false positive selection in RDE systems arise from

terminating RuBP production via transposon silencing of PRK expression (22). To circumvent this impediment a synthetic gene coding a fusion protein comprising *S. elongatus* PCC6301 PRK (*Se*-PRK) and C-terminal neomycin phosphotransferase (NPTII) was made (Figure 1B). The PRK-NPTII fusion protein was found to preserve the kanamycin resistance conferring properties of NPTII to *E. coli* in addition to retaining 70% the PRK activity (k_{cat} = 95 ± 3 s^{-1}) of unmodified *Se*-PRK (134 ± 6 s^{-1}).

The fidelity of the RDE and RDE2 screens were compared by plating XL-1Blue cells producing *Se*-PRK or the PRK-NPTII fusion, along with wild-type *Te*-Rubisco, on media containing 0.2% (w/v) arabinose (high PRK inducing) and no kanamycin. Under these conditions *Te*-Rubisco cannot support cell growth meaning any colonies formed are false positives due to transposon silencing of PRK activity (Figure 1A, (22)). More than 500 false positives colonies per 10^6 cells plated grew in the RDE screen while only 28 ± 17 colonies per 10^6 cells plated grew using the RDE2 screen. Inclusion of 100 μg/mL kanamycin in the growth media of replica RDE2 cell platings resulted in no colony growth. This confirmed the RDE2 screen is immune from generating false-positives as transposon silencing of PRK-NPTII activity also confers the cells sensitive to kanamycin (Figure 1A).

Identifying a suitable Rubisco to evolve. The inability of *E. coli* to meet the folding and assembly requirements of plant and algae Rubisco currently prevents their directed evolution using RDE screens (10, 12). By contrast the biogenesis requirements of cyanobacteria L_8S_8 Rubisco can be partially met by *E. coli*. For example, only a small proportion (<2%) of the 52 kDa *Se*-RbcL produced in *E. coli* can assemble with the more soluble 14 kDa *Se*-RbcS into functional *Se*-Rubisco (20). As a consequence *Se*-Rubisco produced in *E. coli* only accounts for 1 – 2 % (w/w) of the cell soluble protein (CSP). RDE screens have therefore primarily only identified substitutions in RbcL that enhance the biogenesis of *Se*-Rubisco (*i.e.* improved its solubility) (22). Among these are the F140L, V189A and F345I substitutions in RbcL (numbering relative to spinach and *Te*-Rubisco) that increase *Se*-Rubisco solubility between 3 to 14-fold in *E. coli* (Figure 2A,B; Table 1). Notably these mutations all impair k_{cat}^c, contrary to that proposed previously (11). By comparison the biogenesis requirements of *Te*-Rubisco, that already codes Ile-140, are more readily met in *E. coli* and expressed at ~6% w/w CSP (Figure 2A). Moreover, the apparent carboxylation efficiency under ambient O_2 ($k_{cat}^c/K_c^{21\%O2}$) and specificity for CO_2 over O_2 ($S_{c/o}$) of *Te*-Rubisco are approximately 40% and 18% higher, respectively, than *Se*-Rubisco (Table 1). Notably these favourable *Te*-Rubisco kinetics come at the expense of a slower carboxylation rate (k_{cat}^c) that is approximately 40% less than *Se*-Rubisco (Fig 2A,B). The high solubility of *Te*-Rubisco in *E. coli* and distinctive kinetics made it a superior target for evolution testing using RDE2.

First generation Te-Rubisco solubility mutant selection in RDE2. The efficacy of RDE2 to select for increased *Te*-Rubisco activity was initially tested using a library of 2.7 x 10^5 *Te-rbc*L mutants with an average mutation rate of 2.5 nucleotides (1.6 amino acid substitutions) per variant. After 5 days at 25°C eighteen faster growing colonies were identified on plates containing 0.1% (w/v) arabinose (defined as moderate PRK-NPTII induction) and found to code either V300A, F345I, F345L or P415A substitutions in RbcL. The kinetics of the slower growing V300A mutant resembled *Te*-Rubisco suggesting it improved RDE2 fitness possibly from the small (<10%) increase in the amount of soluble L_8S_8 made (Table 1). The faster growing F345I, F345L and P415A mutants produced significantly more *Te*-Rubisco, especially the F345I mutation that stimulated *Te*-Rubisco biogenesis to ~14% (w/w) of the CSP (Figure 2A,B). When coupled with the P415A mutation the solubility of the *Te*-F345I/P415A mutant increased to approximately 17% (w/w) CSP (Figure 2A,B). This improvement came at a cost to the kinetics of each mutant that all showed reductions in k_{cat}^c, $k_{cat}^c/K_c^{21\%O2}$ and $S_{c/o}$ (Table 1).

Sequence comparisons found both F345 and P415 are highly conserved among cyanobacteria and higher plant Form 1B RbcL but are located in separate, unconnected alpha helical regions (Fig 2C). Interestingly, these loci are modified in the Form 1A Rubisco lineage that has evolved independent of the assembly chaperones Raf1 and RbcX (11, 14).

Second generation Te-Rubisco catalytic mutant selection. Improving the kinetics of *Te*-Rubisco required a second round of RDE2 screening that targeted mutagenesis of both RbcL and RbcS (Figure 2D). A $3x10^5$ member mutagenic library of the full length *Te-rbc*LS operon coding the 1st generation P415A RbcL mutant (abbreviated *Te*-P415A) was used. Unlike the F345I and F345I/P415A mutants the catalytic properties of *Te*-P415A more closely matched wildtype *Te*-Rubisco (Figure 2A). Under strong PRK-NPTII selection (*i.e.* on media containing 0.2% (w/v) arabinose) the growth of cells producing wild-type (abbreviated *Te*-LS) or *Te*-P415A Rubisco were impeded (Figure 3A). After 5 days growth at 25°C in air supplemented with 2% (v/v) CO_2, fifteen colonies were isolated. The sequence and biochemistry of these 2nd generation mutants (abbreviated *Te*-2G Rubisco variants) were found to enhance RDE2 growth either through further improving *Te*-P415A Rubisco solubility or/and significantly enhancing its carboxylation properties (Figure 3A, Table 1). The fastest growing *Te*-2Ga mutant was independently selected 4 times and coded a V98M RbcS point mutation that improved Rubisco biogenesis, k_{cat}^{c}, $k_{cat}^{c}/K_c^{21\%O2}$ and $S_{c/o}$ by approximately 310, 28, 43 and 6% relative to *Te*-Rubisco (Figures 3B to E). Four additional *Te*-P415A Rubisco mutants with improvements in k_{cat}^{c} and $k_{cat}^{c}/K_c^{21\%O2}$ and unchanged solubility were also identified. These were found to code for point mutations in either RbcS (A48V, *Te-2*Gb; H37L, *Te-2*Gc or Y36N/G112D, *Te-2*Gd) or RbcL (L74M/D397N, *Te-2*Gg). The *Te-2*Gg mutant also showed a small, but significant, increase of approximately 4% in its $S_{c/o}$ relative to native *Te*-Rubisco (Table 1, Figure 3E).

Clustering of the catalytic mutations at a RbcL-RbcS interface in Te-Rubisco. Structural analysis revealed the H37L, Y36N, V98M, L74M and D397N substitutions all occur at residues that cluster on a surface exposed region of an RbcS-RbcL interface (Figure 4A,B). The A48V mutation occurs on the same interface but is oriented towards the inner enzyme surface (Figure 4A,C). The common locality of these mutations suggest this region to be a hot spot for *Te*-Rubisco catalytic modification, despite being located quite distant from the catalytic pockets in adjoining regions of paired RbcL subunits (Figure 4A). In support of this assertion, the RbcS residue Y36 is located directly adjacent to V98 in the *Te*-Rubisco holoenzyme (Fig 4B) and can convey comparable improvements in k_{cat}^{c} and $k_{cat}^{c}/K_c^{21\%O2}$ when mutated to Asn (Y36N, mutant *Te*-2Gd, Figure 3C,D). Curiously the selected RbcS mutants comprise residue changes adjacent to highly conserved regions of cyanobacterial and higher plant Form IB Rubisco (Figure S2). This suggests this RbcL-RbcS interface may pose a hot spot for manipulating the kinetics in related plant Rubisco lineages.

Second generation Te-Rubisco solubility mutant selection. Of the ten different *Te*-PA Rubisco derived mutants selected, five arose from improvements in the biogenesis of soluble *Te*-Rubisco, not catalysis (Figure 3A). The improved solubility derived from either a R51H substitution in RbcS (or A48V, see above) or from I393M, A398T or A414T mutations in RbcL (Table 1). Notably none of these RbcL substitutions were selected in the 1st generation RDE2 screen suggesting they complement the P415A mutation. Structural analyses reveal however that the residues associated with increasing *Te*-Rubisco solubility show no apparent relation to one another and are located quite distant from P415 within the quaternary Rubisco structure (Figure S3). Consistent with that postulated previously (11), this suggests these residues pose important determinants of RbcL subunit folding by GroEL or/and provide structural stability to avoid their mis-folding and aggregation in the absence of ancillary molecular assembly components.

Comparative analysis of Te-Rubisco biogenesis in chloroplasts. The varied biogenesis requirements within the Rubisco superfamily not only limit which isoforms that can be produced in *E. coli*, but also those that can be assembled within leaf chloroplasts (10, 14, 18). The availability of efficient plastome transformation capabilities in tobacco has made it the model plant for Rubisco bioengineering (13, 16). Consistent with leaf photosynthesis modelling (Figure 5A), the growth of tobacco producing *Se*-Rubisco necessitate high levels of CO_2 for growth in soil to compensate for the enzyme's poor kinetics and limited biogenesis potential in chloroplasts (19). Comparable modelling using *Te*-Rubisco kinetics show it would support higher rates of CO_2-assimilation in a C_3-plant over all intracellular CO_2

concentrations, more so in plants producing the *Te-2*Ga Rubisco mutant selected in this study using the RDE2 screen (Figure 5A).

To test the correlative potential between *Te*-Rubisco biogenesis in *E. coli* and leaf chloroplasts a synthetic Te-*rbc*LS operon coding for either *Te*-Rubisco or the high solubility F345I/P415A mutant was transformed in the tobacco chloroplast genome (plastome) in place of the native tobacco *rbc*L gene (Figure 5B and S4). To optimize translation, the codon use of the transformed genes matched the native tobacco *rbc*L and 12 amino acids of the native tobacco RbcL N-terminus were retained (Figure S4A). Independently transformed lines for each tobacco genotype (called tob^{TeLS} and tob$^{TeLS\text{-}FIPA}$, Figure 5B) were obtained and grown in tissue culture until homoplasmic (Figure 5C). Plastome sequencing confirmed correct integration of the trans-genes and RNA blots confirmed high Te-*rbc*L-*rbc*S and Te-*rbc*L-*rbc*S-*aad*A mRNA levels (Figure S4B). Nevertheless the tob^{TeLS} and tob$^{TeLS\text{-}FIPA}$ plants shared a pale green phenotype and could only survive in tissue culture (Figure 5C). CABP binding (Figure 5D) and native PAGE analysis (Figure 5E) confirmed the assembly of functional L_8S_8 *Te*-Rubisco in tob^{TeLS} and slightly more of the F345I/P415A mutated *Te*-Rubisco in tob$^{TeLS\text{-}FIPA}$. However the amount of Te-Rubisco produced were more than 1000-fold lower than Rubisco production in the wild-type tobacco controls (Figure 5D). Reciprocal mutagenic tests showed introducing the F345I, P415A or dual F345I/P415A mutations into the *N. tabacum* RbcL did not permit the assembly of tobacco Rubisco in *E. coli* (Fig. S5).

DISCUSSION

Following successive evolution rounds this study reveals a novel region in the subunit interface of the plant-like Rubisco from the *T elongatus* BP1 that offers multiple solutions for improving carboxylation rate, efficiency and specificity. Key to this success is the unwavering fidelity, faster screening and higher throughput of the new RDE2 screening platform. To date the RDE screen designed around a Δ*gapA E. coli* mutant (a strain called MM1, (26)) has proven the most effective in Rubisco directed applications due to the low number of false positives selected (12). Offsetting this improved fidelity is the slow growth rate of the MM1 strain and its need for specialized growth media that prolongs the screening timeframe to 9 to 12 days at the optimal growth temperature of 25°C (11, 17, 20, 26). Comparatively, the RDE2 screen is immune to false positive production and uses LB-growth media allowing the screens to be completed in 4-6 days at 25°C (Figure 2D). The feasibility of adapting the new CO_2-dependent sugar synthesizing CBB-*E. coli* strains (23) as an alternative RDE screen for selecting Rubisco catalytic mutants remains to be demonstrated (21).

There has been longstanding interest in modifying Rubisco catalysis via mutagenesis of RbcL (3, 5-7, 10). More recent RbcS mutagenic studies have demonstrated how it can pervasively influence plant Rubisco kinetics (27), a finding supported by structure-function surveys that indicate a role by the RbcS in the kinetic adaptation of Rubisco to environmental cues (18). X-ray crystallography comparisons show the quaternary RbcS structure remains highly conserved among the differing L_8S_8 lineages, especially among the plant and cyanobacteria Form 1B isoforms (Figure 6A). This suggests RbcS mediated changes to catalysis can be strongly influenced by subtle changes in electron density within key areas. One particular area, as demonstrated in this study, is located between the anti-parallel beta sheets around V98, Y36 and H37 in *Te*-RbcS that share structural similarity to C122, W38 and V39 in tobacco RbcS (Figure 6B). Generally, the M98, N36 or L37 substitutions differentially improved *Te*-Rubisco catalysis by modification towards more plant-like structures (Figure 6C). Notably these substitutions did not arise from limitations in codon redundancy at these residues in *Te*-RbcS. This suggests further improvements may be possible by additional rounds of laboratory evolution, or possibly even by rational design guided by the vast array of crystal structure information already available for plant and algae Rubisco (4).

An obvious extension of our study is to test how amino acid changes around this region in RbcS influence plant Rubisco catalysis. Such efforts are hindered by the inadequate transgenic methods for replacing or modifying all the multi-*RbcS* gene copies in plants. The low throughput of modern nucleus gene editing approaches limit their usefulness in mutagenic screening, despite

their capacity to introduce multiplexed nucleotide changes with relatively high accuracy (28). While our RDE2 screen poses a potential remedy in terms of screening throughput, our data highlights its versatility for evolving vascular plant Rubisco will require incorporation of requisite Rubisco assembly components from leaf chloroplasts. This observation accords with the underpinning need for ancillary protein complementarity for folding and assembly of Form I Rubisco in leaf chloroplasts (13). Understanding these requirements is critical to enabling heterologous Rubisco expression potential in cyanobacteria, *E coli* and chloroplasts. Despite the common prokaryotic connections of these hosts our work demonstrates that *E. coli* cannot be used as a proxy for chloroplast Rubisco assembly. This underpins the importance of understanding the specific chaperonin (Cpn60α, 60β), co-chaperonin (Cpn10, 20 and 21), and assembly chaperone requirements of plant and algal Rubisco, or strategies to circumvent their necessity for successful production in a non-native host. The last decade has seen significant advances into understanding the mechanisms for many of these components (14) but they have not yet been incorporated into a screen such as RDE2. Conceivably such an approach would provide a high throughput screen of plant (or algae) Rubisco activity that would be envisaged to function equivalently chloroplasts (29).

Efforts to introduce a cyanobacterial CO_2-concentration mechanism (CCM) into tobacco plastids are also likely to require the inclusion of Rubisco biogenesis components. In this study we show *Te*-Rubisco has the highest CO_2 affinity and CO_2/O_2 specificity of known cyanobacteria Rubisco. These kinetic features make *Te*-Rubisco better able to support C_3-photosynthesis, relative to the more commonly studied *Se*-Rubisco. The reverse would be the case if the extensive requirements for building a functional CCM could be met in plastids to support the higher k_{cat}^{C} of *Se*-Rubisco (30, 31). The limited biogenesis capacity of *Te*-Rubisco (Figure 5) and *Se*-Rubisco (19, 30, 32) in tobacco suggest that building a CCM in chloroplasts will not only require introducing the multiple carboxysome components, inorganic carbon membrane transporters and modulating CA levels, but also the incorporation of the necessary chaperonin/chaperone components for assembling cyanobacteria Rubisco.

Summary - Much of the effort to identify superior Rubiscos to date has focused on measuring the natural variation in Rubisco kinetics. This approach is slowed by the complexity of the catalytic assay methods and the vast spectrum of natural diversity available. Here we show how directed evolution using RDE2 poses a potentially faster alternative. The challenge is to extend past the prior success of RDE screens in investigating prokaryotic and Archaea Rubisco sources and adapt RDE2 for the directed evolution of plant and algae Rubisco. As with Rubisco engineering in plant chloroplasts, our data advocates that this will necessitate the co-expression of the complimentary ancillary proteins needed for the biogenesis and metabolic repair of the target Rubisco.

EXPERIMENTAL PROCEEDURES

Expression, purification and assay of PRK and PRK:NPTII-The *prkA* gene from *Synechococcus* PCC6301 was synthesized (Genscript) and cloned downstream of the arabinose inducible BAD promoter in pAC^{BAD} (26) to generate $pACYC^{PRK}$. Sequence coding a synthetic *nptII* gene was cloned in frame to the 3′ of *prkA* to produce the *prk-npt*II gene coding a PRK-NPTII fusion protein (GenBank accession MG000147). The *prk*A and *prk-npt*II genes were cloned in frame to a N-terminal 6xhistidine-ubiquitin (H_6Ub) fusion protein using the expression plasmid pHue, expressed in BL21(DE3) *E. coli*, purified by immobilised metal affinity chromatography and the 10kDa H_6Ub sequence removed as previously described (33). PRK activity (k_{cat}) was measured using a NADH enzyme-linked assay as described (34).

Comparing the selection stringency of RDE and RDE2- The *Te*-Rubisco expressing plasmid $pTrc^{TeLS}$ (Figure 1B) was produced by separately amplifying the *rbc*L and *rbc*S genes from genomic DNA isolated from *Thermosynechococcus elongatus* BP1 (gifted by Dr Warwick Hillier, ANU) and cloning into pTrcHisB as a *Nco*I-*Sal*I fragment (GenBank accession MG000148). *E. coli* XL1-Blue cells containing $pTrc^{TeLS}$ were transformed with either $pACYC^{PRK}$ (RDE) or $pACYC^{PRK:Kan}$ (RDE2, Figure 1B) and grown at 25°C for 6 days in air with 1% (v/v) CO_2 on LB-

agar containing 0.5 mM IPTG (*Te*-Rubisco induction), 0.2% (w/v) arabinose (high PRK/PRK-NPTII induction) and with/without 0.2 mg.ml^{-1} kanamycin.

Library construction and Rubisco selection using the RDE2 screen-The *rbc*LS operon in pTrc^{TeLS} (1st generation library) or pTrc^{TePA} (2nd generation library) were amplified from 10 ng plasmid DNA by error-prone PCR using the primers Trc55 (5'-GAGGTATATATTAATGTATCG-3') and Trc33 (5'-ATCTTCTCTCATCCGCCA-3') and the Genemorph II Random Mutagenesis kit (Agilent), as per the manufacturer's recommendations. Libraries were transformed into $CaCl_2$ competent RDE2 cells and plated onto LB media containing 32 µg/mL chloramphenicol, 200 µg/mL ampicillin, 100 µg/mL kanamycin, 0.5 mM IPTG and 0.05 to 0.25%[w/v] arabinose. After growing at 25°C for 3 – 7 days in air containing 2.5% [v/v] CO_2 the faster growing colonies were re-plated and then their pTrc^{TeLS} plasmid purified, sequenced, re-transformed into RDE2 and the screen repeated to confirm their selective advantage.

Rubisco content and catalysis-Cyanobacteria Rubisco expression was induced in XL-1Blue *E. coli* transformed with pTrc^{TeLS}, pTrcSynLS (20) or their mutant plasmid derivatives with 1 mM IPTG at 28°C. After 6 h the cells were harvested by centrifugation (5 min at 4°C, 6200 x *g*), and the cell pellets N_2-frozen and stored at -80°C. Their soluble protein was isolated following lysis using a French pressure cell (140 MPa) in ice cold extraction buffer (100 mM EPPES-NaOH, pH 8.05, 15 mM $MgCl_2$, 0.5 mM EDTA, 1 mM PMSF, 2.5 mM DTT). After centrifugation (3 min, 2°C, 15,000 x *g*) the supernatant was mixed with an equal volume of reaction buffer (100 mM EPPES-NaOH, pH 8.05, 15 mM $MgCl_2$, 0.5 mM EDTA) containing 40 mM $NaH^{14}CO_3$ and incubated at 25°C for 8 or 12 min (technical repeats) to activate Rubisco. The $^{14}CO_2$ fixation assays (0.5 mL total volume) were performed in 7.7 mL septum-capped scintillation vials in reaction buffer containing 10 µg ml^{-1} carbonic anhydrase and saturating (0.4 mM) RuBP synthesized and purified according to Kane et al., 1998. All assay components were equilibrated in CO_2-free air (*i.e.* 20.9% (v/v) O_2 in N_2) prior to adding a series of 5 differing amounts of $NaH^{14}CO_3$. The final $^{14}CO_2$ concentrations were 0 to 240 µM or 0 to 350 µM when assaying the wild-type and mutant variants of *Te*-Rubisco or *Se*-Rubisco, respectively. The assays were initiated by addition of 20µL of $^{14}CO_2$-activated *E. coli* soluble protein. RuBP-independent $^{14}CO_2$ fixation control assays were run for each protein sample and contained H_2O in place of RuBP. The assays were terminated after 1 min by rapid mixing with 0.2 mL 20% (v/v) formic acid then dried at 80°C before adding 0.5 mL H_2O and mixing with 1 mL Ultima-Gold scintillant (PerkinElmer). The fixed ^{14}C was measured in a Tri-carb 4910TR scintillation counter with the carboxylase activity between the technical repeats varying by <2% confirming the extracted Rubisco was fully activated and stable.

The specific activity of each $NaH^{14}CO_3$ stock was determined in assays (n=4) containing the highest $^{14}CO_2$ concentration and 5 nmol of pure RuBP. These assays were allowed to react for 30 or 90 min to ensure full RuBP fixation. After acid treatment and drying the scintillation counter ^{14}C values were divided by 5 to derive the specific activity value – which varied between 1500 to 1800 cpm per nmol CO_2 fixed.

The CO_2 levels in the assays were calculated using the Henderson-Hasselbalch derived equation

$$[CO_2] = \frac{(C_t)}{1 + \frac{V}{vqRT} + 10^{(pH-pK1)} + 10^{(2pH-pK1-pK2)}}$$

where *V/v* is the ratio of reaction vial headspace (*V*) to assay volume (*v*); C_t the total $NaHCO_3$ concentration (including the 0.4 µmol $NaH^{14}CO_3$ in the *E. coli* soluble protein); *q*, the CO_2 solubility at 1 atm (0.0329 Mol.L^{-1}.atm^{-1} at 25°C); *R*, the universal gas constant (0.082057 L.atm.K^{-1}.mol^{-1}); *T*, the assay temperature (298K) and dissociation constants of 6.251 (pK_1) and 10.329 (pK_2). The ^{14}C data were fitted to the Michaelis-Menten equation to derive the apparent Michaelis-Menton constant (K_m) for CO_2 under air levels of O_2 ($K_c^{21\%O2}$) and the maximal rate of carboxylation (V_c^{max}) from at least three independent *E. coli* soluble protein preparations. The caboxylation rate for each Rubisco (k_{cat}^{c}) was standardized by dividing V_c^{max} by the number of Rubisco-catalytic sites in the *E. coli* protein quantified by [^{14}C]-2-CABP binding native PAGE(20, 35). The soluble protein concentration in each *E. coli* extract was

quantified using a Coomassie dye binding assay against BSA.

Assays to measure the K_m for RuBP of *Te*-Rubisco (see legend to Table 1) were undertaken in reactions containing 20 mM $NaH^{14}CO_3$ and a series of six RuBP concentrations (0 to 150 µM).

Rubisco $S_{C/O}$ measurements were made as described in Kane *et al*, (1994) using recombinant Rubisco purified from *E. coli* by anion exchange chromatography then Superdex 200 (GE Life Sciences) size exclusion column chromatography (35). Each purified Rubisco was equilibrated with a gas mixture of 0.05 % (v/v) CO_2 and 99.6 - 99.95% (v/v) O_2 (accurately mixed using Wostoff gas mixing pumps) at 25°C in replica septum seal 20 mL glass vial assays comprising 1 mL specificity buffer (30 mM Triethanolamine pH 8.1, 10 mM $MgSO_4$, 10 µg.mL^{-1} carbonic anhydrase). After 1 hour the reactions were initiated by the addition of 1 nmol [1-^{3}H]-RuBP (10 MBq/nmol), then 10U alkaline phosphatase (Sigma) added 30 to 60 mins later. The resulting ^{3}H-glycerate and ^{3}H-glycolate products were then separated by HPLC and measured by scintillation counting as described (36). The $S_{C/O}$ factor was calculated using the following equation:

$$S_{c/o} = (R_{glycerate/glycolate}\ x\ \frac{M_{O_2}}{M_{CO_2}})\ x\ 0.037$$

Where from R is the ratio of ^{3}H-glycerate to ^{3}H-glycolate, M_{O_2} and M_{CO_2} the mole fractions of O_2 and CO_2, respectively, in the assay and 0.037 is the ratio between the aqueous solubility of O_2 and CO_2 at 25°C.

PAGE analyses-The leaf and *E. coli* soluble protein extracts were prepared and analyzed by SDS-PAGE, native PAGE and immunoblot analysis as described (35).

Tobacco chloroplast transformation-The plasmids pLEV*Te*LS (GenBank accession MG000149) and pLEV*Te*LS^{FIPA} were biolistically transformed into the plastome of the ^{cm}trL tobacco genotype as previously described (16). The resulting transplastomic genotypes tob^{TeLS} and tob$^{TeLS\text{-}FIPA}$ coded a synthetic *rbc*LS operon and the *aad*A marker gene (coding spectinomycin resistance) in place of the tobacco *rbc*L gene (Figure 5A). See Figure S4 for additional transformation and RNA blotting detail. Correct transgene insertion was confirmed by fully sequencing the PCR product amplified from total leaf genomic DNA isolated using the DNeasy® Plant Mini Kit and primers LSD (5'-CACGGAATTCGTGTCGAGTAG-3') and LSZ (5'-ATCCTTCTTTATTTCCTGC-3').

Leaf photosynthesis simulations-Photosynthetic CO_2 assimilation rates (A) in tobacco at 25°C under varying chloroplast CO_2 partial pressures (C_C) were simulated as the minimum of the Rubisco carboxylase-limited (A_c) and light-limited (A_j) CO_2-assimilation rates modelled according to (37):

$$A_c = \frac{m.k_{\text{cat}}{}^{c}\left(C_c.s_c - 0.5\,O_c/S_{c/o}\right)}{\left(C_c.s_c + K_c^{21\%O2}\right)} - R_d$$

$$A_j = \frac{\left(C_c.s_c - 0.5\,O_c/S_{c/o}\right)J_{\max}}{4\left(C_c.s_c + O_c/S_{c/o}\right)} - R_d$$

using the Rubisco kinetic values listed in Table S1 and a leaf Rubisco content (m) of 30 µmol active sites m^{-2} (or 5 µmol active sites m^{-2} for the assembly impaired *Se*-Rubisco producing tobacco line tob-SeLS, (19)); a maximal RuBP regeneration rate (J_{max}) of 160 µmol m^{-2} s^{-1}; a mitochondrial respiration rate (R_d) of 1 µmol m^{-2} s^{-1} and using the solubility constants 0.0334 M bar^{-1} (s_c) and 0.00126 M bar^{-1} (s_o) to calculate the CO_2 (C_c) and O_2 (O_c) concentrations in the chloroplast.

ACKNOWLEDGEMENTS:

Author contributions: RHW and SMW conceived the study. RHW designed RDE2, performed the directed evolution research and generated the tobacco transgenics that were maintained and analyzed by CC, EMA and SMW. RHW and SMW performed the Rubisco biochemical analyses and wrote the manuscript. We thank Andreas Bracher (Max-Planck Institute for Biochemistry, Germany) for constructive assistance in Rubisco structural modeling.

CONFLICT OF INTEREST:

The authors declare they have no conflicts of interest with the contents of this article.

REFERENCES

1. Long Stephen P, Marshall-Colon A, & Zhu X-G (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161(1):56-66.
2. Tcherkez G (2016) The mechanism of Rubisco-catalysed oxygenation. *Plant, cell & environment* 39(5):983-997.
3. Whitney SM, Houtz RL, & Alonso H (2011) Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant physiology* 155(1):27-35.
4. Andersson I & Backlund A (2008) Structure and function of Rubisco. *Plant Physiology and Biochemistry* 46(3):275-291.
5. Parry MAJ, *et al.* (2013) Rubisco activity and regulation as targets for crop improvement. *J Exp Botany* 64(3):717-730.
6. Sharwood RE (2017) Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phyt* 213(2):494-510.
7. Carmo-Silva E, Scales JC, Madgwick PJ, & Parry MAJ (2015) Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant, Cell & Environment* 38(9):1817-1832.
8. Ellis RJ (1979) The most abundant protein in the world. *Trends in biochemical sciences* 4(4):241-244.
9. Ort DR, *et al.* (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences of the United States of America* 112(28):8529-8536.
10. Whitney SM, Baldet P, Hudson GS, & Andrews TJ (2001) Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J* 26(5):535-547.
11. Durao P, *et al.* (2015) Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* 11(2):148-155.
12. Mueller-Cajar O & Whitney SM (2008) Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosyn Res* 98:667-675.
13. Whitney SM, Birch R, Kelso C, Beck JL, & Kapralov MV (2015) Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proceedings of the National Academy of Sciences of the United States of America* 112(11):3564-3569.
14. Bracher A, Whitney SM, Hartl FU, & Hayer-Hartl M (2017) Biogenesis and metabolic maintenance of Rubisco. *Annual Review of Plant Biology* 68:29-60.
15. Tsai Y-CC, Mueller-Cajar O, Saschenbrecker S, Hartl FU, & Hayer-Hartl M (2012) Chaperonin Cofactors, Cpn10 and Cpn20, of Green Algae and Plants Function as Hetero-oligomeric Ring Complexes. *J Biol Chem* 287(24):20471-20481.
16. Whitney SM & Sharwood RE (2008) Construction of a tobacco master line to improve Rubisco engineering in chloroplasts. *J Exp Bot* 59(7):1909-1921.
17. Wilson RH, Alonso H, & Whitney SM (2016) Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci Rep* 6:Article 22284.
18. Sharwood RE, Ghannoum O, Kapralov MV, Gunn LH, & Whitney SM (2016) Temperature responses of Rubisco from *Paniceae* grasses provide opportunities for improving C_3 photosynthesis. *Nature Plants* 2:16186.
19. Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, & Hanson MR (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature* 513(7519):547-550.
20. Mueller-Cajar O & Whitney SM (2008) Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. *The Biochemical journal* 414:205-214.
21. Antonovsky N, Gleizer S, & Milo R (2017) Engineering carbon fixation in *E. coli*: from heterologous Rubisco expression to the Calvin–Benson–Bassham cycle. *Current Opinion in Biotechnology* 47:83-91.
22. Wilson RH & Whitney SM (2017) Improving CO_2 fixation by enhancing Rubisco performance. *Directed Enzyme Evolution: Advances and Applications*, ed Alcalde M (Springer International Publishing, Cham), pp 101-126.
23. Antonovsky N, *et al.* (2016) Sugar Synthesis from CO_2 in *Escherichia col*i. *Cell* 166(1):115-125.
24. Cai Z, Liu G, Zhang J, & Li Y (2014) Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. *Protein Cell* 5(7):552-562.
25. Greene DN, Whitney SM, & Matsumura I (2007) Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. *The Biochemical journal* 404(3):517-524.

26. Mueller-Cajar O, Morell M, & Whitney SM (2007) Directed evolution of Rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochem.* 46(49):14067-14074.
27. Ishikawa C, Hatanaka T, Misoo S, Miyake C, & Fukayama H (2011) Functional incorporation of sorghum small subunit increases the catalytic turnover rate of Rubisco in transgenic rice *Plant Physiol* 156:1603-1611.
28. Yin K, Gao C, & Qiu J-L (2017) Progress and prospects in plant genome editing. *Nature Plants* 3:17107.
29. Whitney SM & Andrews TJ (2003) Photosynthesis and growth of tobacco with a substituted bacterial Rubisco mirror the properties of the introduced enzyme. *Plant Physiol* 133:287-294.
30. Hanson MR, Lin MT, Carmo-Silva AE, & Parry MAJ (2016) Towards engineering carboxysomes into C_3 plants. *Plant J* 87(1):38-50.
31. Price GD & Howitt SM (2014) Towards turbocharged photosynthesis. *Nature* 513(7519):497-498.
32. Occhialini A, Lin MT, Andralojc PJ, Hanson MR, & Parry MAJ (2016) Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO_2. *Plant J* 85(1):148-160.
33. Baker RT, *et al.* (2005) Using deubiquitylating enzymes as research tools. *Methods Enzymol.* 398:540-554.
34. He Z, *et al.* (1997) Ribulose-1,5-bisphosphate carboxylase/oxygenase activase deficiency delays senescence of ribulose-1,5-bisphosphate carboxylase/oxygenase but progressively impairs its catalysis during tobacco leaf development. *Plant Physiology* 115(4):1569-1580.
35. Sharwood RE, von Caemmerer S, Maliga P, & Whitney SM (2008) The catalytic properties of hybrid Rubisco comprising tobacco small and sunflower large subunits mirror the kinetically equivalent source Rubiscos and can support tobacco growth. *Plant Physiology* 146:83-96.
36. Kane HJ, *et al.* (1994) An improved method for measuring the CO_2/O_2 specificity of ribulosebisphosphate carboxylase-oxygenase. *Aust.J.Plant Physiol.* 21:449-461.
37. Farquhar GD, von Caemmerer S, & Berry JA (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* 149:78-90.
38. Whitfield JH, *et al.* (2015) Construction of a robust and sensitive arginine biosensor through ancestral protein reconstruction. *Protein Science* 24(9):1412-1422.

FOOTNOTES

This research was funded by the Australian Research Council grant CE140100015 awarded to SMW. Support for CC was provided as part of the Bill and Melinda Gates Foundation award RIPE: Realizing Increases in Photosynthetic Efficiency

The abbreviations used are: CABP, 2'-carboxyarabinitol-1,5-bisphosphate; carboxypentitol-P_2, isomeric mixture of carboxyarabinitol-P_2 and 2'-carboxyribitol-1,5-bisphosphate; RuBP, D-ribulose-1,5-bisphosphate; Rubisco, RuBP carboxylase/oxygenase.

TABLE 1: *E.coli* L_8S_8 Rubisco content and catalytic parameters at 25°C

Rubisco	Mutant library	Amino acid mutation(s)		Expression (% [w/w] sol *E. coli* protein)	k_{cat}^{c}	$K_c^{21\%\ O2}$	$k_{cat}^{c}/K_c^{21\%O2}$	$S_{C/O}$	T_m #
		RbcL	**RbcS**		(s^{-1})	(μM)	($mM.^{-1}.s^{-1}$)	($mol.mol^{-1}$)	(°C)
n = (**biological samples**, *technical replicates*)				(**3-5**, *2*)	(**3-9**, *2*)	(**3-5**, *2*)	(**3-5**)	(**3**, *3*)	(**1**,*3*)
Se-F345I	a,b	F345I	nil	14.4 ± 0.9**	10.3 ± 0.4**	285 ± 9**	36 ± 2**	41.8 ± 1.1**	n.m.
Se-F137I	a	F137I	nil	6.3 ± 1.1**	11.1 ± 0.5**	211 ± 5**	53 ± 3	46.0 ± 0.7	n.m.
Se-V186I	a	V189I	nil	6.2 ± 1.3**	10.6 ± 0.2**	186 ± 7**	57 ± 2**	39.4 ± 0.7**	n.m.
1st generation Te-rbcL – RDE2 mutants									
Te-F345I	1G	F345I	nil	14.6 ± 0.7**	4.8 ± 0.3**	101 ± 1	47 ± 2**	48.0 ± 0.5**	72.3 ± 0.3
Te-P451A	1G	P451A	nil	10.4 ± 0.8**	6.9 ± 0.2*	111 ± 2	62 ± 2*	52.6 ± 0.5	73.7 ± 0.1**
Te-V300A	1G	V300A	nil	7.1 ± 0.2	6.6 ± 0.2*	135 ± 9**	49 ± 6**	51.9 ± 0.6	71.4 ± 0.6
Te-FIPA	1G chimer	F345I, P451A	nil	17.5 ± 1.2*	4.8 ± 0.4**	102 ± 5	47 ± 6**	49.7 ± 0.4**	75.9 ± 0.1**
2nd generation Te-rbcLS – RDE2 mutants									
Te-2Ga (1)	2G	P451A	V98M	20.4 ± 3.3**	9.5 ± 0.3**	93 ± 3**	100 ± 3**	56.3 ± 0.3**	n.m.
Te-2Gb (2)	2G	P451A	A48V	9.7 ± 1.2	8.8 ± 0.2**	105 ± 3	84 ± 3**	48.6 ± 0.6**	n.m.
Te-2Gc (3)	2G	P451A	H37L	11.1 ± 0.2	8.4 ± 0.3**	100 ± 2*	84 ± 4**	49.7 ± 0.3**	n.m.
Te-2Gd (5)	2G	P451A	Y36N, G112D	9.7 ± 0.8	9.3 ± 0.2**	94 ± 0.4**	99 ± 2**	47.7 ± 0.1**	n.m.
Te-2Ge (8)	2G	P451A	Y36N, R51H	13.6 ± 1.6**	8.0 ± 0.4	120 ± 8	68 ± 2**	51.8 ± 0.2**	n.m.
Te-2Gf (10)	2G	A398T, P451A	A48V	13.3 ± 3.2	8.1 ± 0.1*	105 ± 3	75 ± 5	45.5 ± 0.1**	n.m.
Te-2Gg (11)	2G	L74M, D397N, P451A	nil	12.5 ± 0.8**	8.2 ± 0.5	92 ± 3**	90 ± 6**	55.4 ± 0.3**	n.m.
Te-2Gh (12)	2G	A414T, P451A	nil	13.9 ± 3.0*	8.0 ± 0.1	102 ± 2	78 ± 1	55.8 ± 0.5**	n.m.
Te-2Gi (13)	2G	I393M, P451A	nil	16.6 ± 2.4**	7.8 ± 0.4	90 ± 2**	87 ± 1**	48.1 ± 0.4**	n.m.
Te-2Gj (15)	2G	A398T, P451A	nil	16.6 ± 3.6**	6.4 ± 0.2**	110 ± 4	59 ± 3*	50.5 ± 0.5**	n.m.

Values are the mean ± SE of typically N≥3 biological samples assayed in duplicate or triplicate (details shown in top row). One-way ANOVA was undertaken with reference to either wild-type (wt) *Synechococcus* PCC 6301 Rubisco (*Se*-LS) or wild-type *Te*-Rubisco (*Te*-LS). Symbols show the statistical significance levels (* = $p < 0.05$; ** = $p < 0.01$) relative to wt enzymes. 1G, 1st generation RDE2 Te-LS Rubisco mutants selected on media containing 0.1% (w/v) arabinose (Figure 2D); 2G, 2nd gen mutants derived from *Te*-P415A under higher PRK selection on 0.25% (w/v) arabinose media (Figure 3). #Measurement of thermal stability (T_m) were made by circular dichroism spectroscopy as described (38). n.m., not measured. The K_m for RuBP of *Se*-LS and *Te*-LS Rubisco are 44±3 μM (20) and 36±2 μM (this study) respectively, ~10-fold less than the 0.4 mM RuBP concentration used in the $^{14}CO_2$-fixation assays.

FIGURE LEGENDS

FIGURE 1. An improved Rubisco dependent *E. coli* (RDE) screen A, RDE screens rely on the toxicity of RuBP (ribulose-1,5-P_2), a foreign metabolite produced from the pentose phosphate pathway (PPP) intermediate ribulose-5-phosphate (ribulose-5-P) via recombinant phosphoribulokinase (PRK) expression (regulated by arabinose addition). Cell viability is restored through either Rubisco expression (IPTG induced regulation), where growth rate is proportional to Rubisco activity level, or through transposon silencing of PRK expression that produces false positives at a high frequency (22). The new RDE2 screen utilizes a PRK-NPTII fusion comprising neomycin phosphotransferase (NPTII) tethered in frame to the C-terminus of PRK endowing kanamycin resistant to *E. coli* unless PRK-NPTII expression is transposon silenced. Toxic pathways to *E. coli* are highlighted in red. RDE2 screens are performed under high CO_2 (air + 1-2.5% [v/v] CO_2) with the ribulose-1,5-P_2 carboxylation reaction of Rubisco (positive sign) producing 3-phosphoglycerate (3-PGA), an intermediate of the glycolysis/gluconeogenesis pathways (shaded yellow)(reviewed in (12, 21)). B, Summary of the plasmids and *E. coli* growth conditions used in the RDE2 screen of an *rbcL* (±*rbcS*) mutant library. P_{trc}, P_{BAD}; IPTG and arabinose inducible promoters, respectively.

FIGURE 2. Directed evolution of Te-Rubisco using the RDE2 screen. A, Comparative L_8S_8 Rubisco biogenesis capacity (quantified by [^{14}C]-CABP binding) and B, confirmed by native PAGE. The CO_2-fixation rate (k_{cat}^{C}) of wildtype (wt) and prior RDE selected *Synechococcus elongatus* PCC 6301 (*Se*)-Rubisco solubility RbcL mutants (in green, F345I, F140I, V189I) (11, 20) relative to *Thermosynechococcus elongatus* BP1 (*Te*)-Rubisco (in blue) and the RbcL solubility F345I and P415A mutants identified using the RDE2 screen. Data are the mean ± S.E of >3 biological samples with significance relative to each wt Rubisco analyzed by one-way analysis of variance (ANOVA) with Turkey's post hoc analysis. **P < 0.01. See Table 1 for additional catalysis measurements. C, Partial RbcL alignment showing the conserved F345 and P415 loci in Form IB Rubisco compared to Y345 and A415 in the Form IA Rubisco examples shown. D, Example growth phenotypes of RDE2 cells producing *Te*-Rubisco and the P415A (*Te*-PA) solubility mutant under non-selective or increasingly selective conditions (*i.e.* intensifying arabinose induced PRK-NPTII expression) of the 1st and 2nd evolution rounds to enrich for selection of *Te*-Rubisco mutants with improved catalysis.

FIGURE 3. The catalysis and assembly of the 2nd generation *Te*-Rubisco mutants. A, Comparative growth after 6 days at 25°C of XL1-Blue *E. coli* producing wild-type, P415A and 2nd Gen *Te*-Rubisco (Te-2G) mutants on selective media (0.25% [w/v] arabinose 0.5 mM IPTG, 0.2 mg.ml^{-1} kanamycin, 0.1 mg.ml^{-1} ampicillin). Shown are the frequencies of selection for each mutant (from a mutagenic library of 3x10^5 colony forming units) and the amino acid substitutions in RbcL or/and RbcS identified. The effect of the mutations on B, L_8S_8 Rubisco biogenesis (quantified by [^{14}C]-2-CABP binding), C, CO_2-fixation rate (k_{cat}^{C}), D, apparent carboxylation efficiency (k_{cat}^{C} divided by the apparent K_m for CO_2 under ambient O_2; $K_C^{21\%O2}$) and E, specificity for CO_2 over O_2 ($S_{c/o}$) at 25°C, pH 8.0 were compared relative to the parental wild-type (*Te*-LS) and *Te*-P415A Rubiscos. Shown are the average data (± S.E of N >3samples, see Table 1). Significance of variation to *Te*-Rubisco determined by one-way ANOVA with Turkey's post hoc analysis *****P < 0.05, ******P < 0.01, *******P < 0.001.

FIGURE 4. Location of the catalysis enhancing residues in *Te*-Rubisco. A, Surface view from two orientations showing two *Te*-RbcL (purple and orange) and their antiparallel *Te*-RbcL partners (ribbon structures) bound by upper and lower RbcS monomers (green) with the loci of the 2nd Gen mutations shown in red. Bound CABP (an inhibitory structural analog of RuBP) in the RbcL catalytic sites is shown in magenta and the conserved K201 lysine shown in yellow. B,C, Structural detail of the clustered location and interactions of the catalysis influencing V98, A48, H37 and Y36 residues in *Te*-RbcS and the *Te*-RbcL amino acids L74 and D397 (with non-selected H-bond interacting residues shown in italics). Protein Data Bank code 3ZXW.

FIGURE 5. *Te*-Rubisco expression in tobacco. A, Simulated effect of *Se*-Rubisco (*Se*LS), *Te*-Rubisco (*Te*LS) and the catalytically improved *Te2*Ga-Rubisco (*Te*-2Ga) on CO_2 assimilation rate (***A***) in tobacco (a model C_3-plant***)*** at 25°C in response to varying chloroplast CO_2 partial pressures (***C***$_C$). ***A*** was calculated as the minimum of the Rubisco carboxylase-limited (A_c) and light-limited (A_j) rates modelled as described in Experimental Methods and assuming a leaf Rubisco content of 30 μmol active sites m^{-2} (or 5 μmol active sites m^{-2} as observed in tob-SeLS (19)). B, The tobacco plastome was transformed with plasmids pLEV-TeLS and pLEV-TeLS-FIPA containing homologous plastome flanking sequences [indicated by dashed lines; numbering relative to *N. tabacum* (wt) plastome; GenBank accession Z00044] that directed integration of codon optimized *rbcLS* operon (GenBank accession MG000149) coding wt or FIPA *Te*-Rubisco to produce the tobacco genotypes tob^{TeLS} and tob$^{TeLS\text{-}FIPA}$ (see Supplemental figure S4 for further detail). C, The tob^{TeLS} and tob$^{TeLS\text{-}FIPA}$ plants could be grown in tissue culture but not in soil as D, both produced finite levels of L_8S_8 *Te*-Rubisco as quantified by [^{14}C]-CABP binding and E, confirmed by native PAGE. Loading controls included tobacco Rubisco (tob), known amounts of *Te*-Rubisco (pTrc*Te*LS) expressed in *E. coli* and an empty vector (Trc) negative control. No L_8S_8 *Te*-Rubisco band was evident by Coomassie staining in the tob^{TeLS} and tob$^{TeLS\text{-}FIPA}$ leaf protein samples but detected by immunoblotting using a Se-Rubisco antibody. *, non-Rubisco tobacco protein that separates at the same location as *Te*-Rubisco by native PAGE.

FIGURE 6. Comparative structures of tobacco and *Te*-RbcS.
A, Structural overlay of *Nicotiana tabacum* (*Nt*RbcS, green, PDB code 4RUB) and *Thermosynechococcus elongatus* (*Te*RbcS, yellow, PDB code 3ZXW) RbcS structures B, Comparative positioning of the amino acid side chains along the anti-parallel beta sheet in wild-type *Nt*RbcS and *Te*RbcS at the loci selected in the 2nd generation (2G) *Te*-Rubisco mutants. C. Extension of panel B comparing the *Te*RbcS-2G amino acid substitutions with *Nt*RbcS.

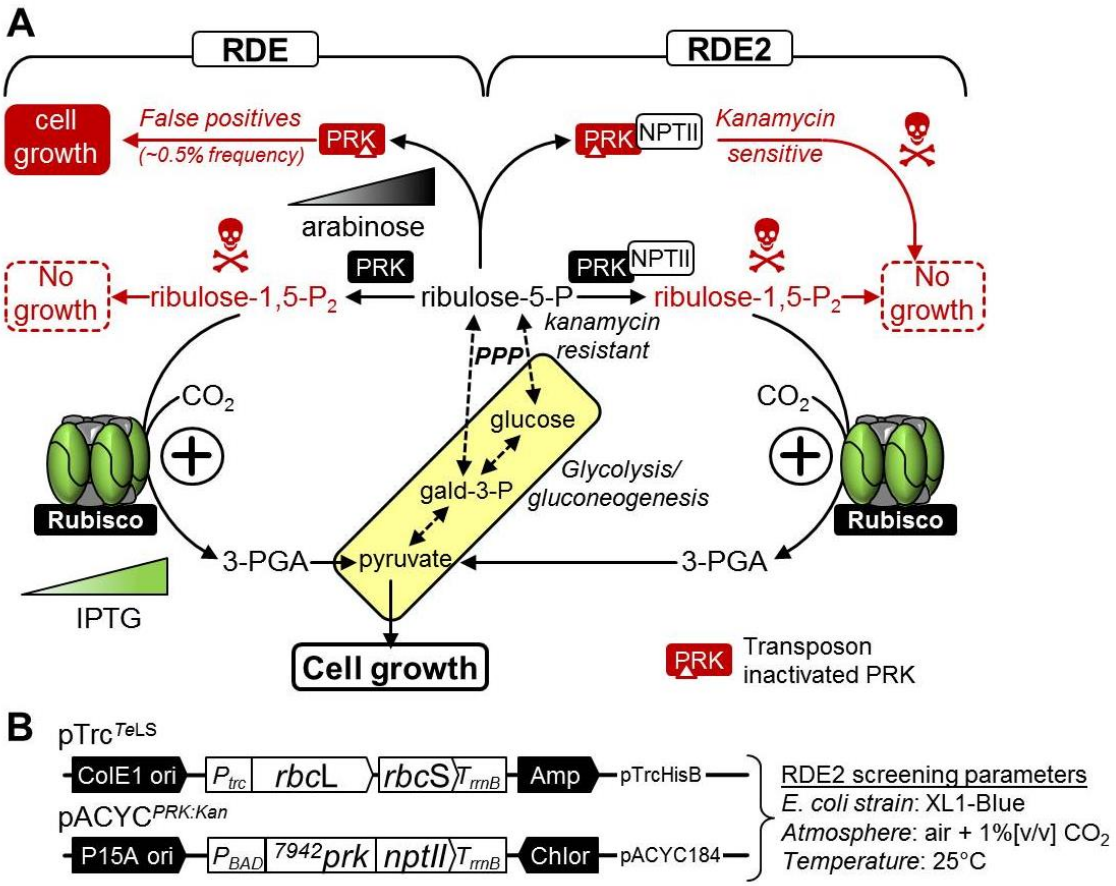

Figure 1

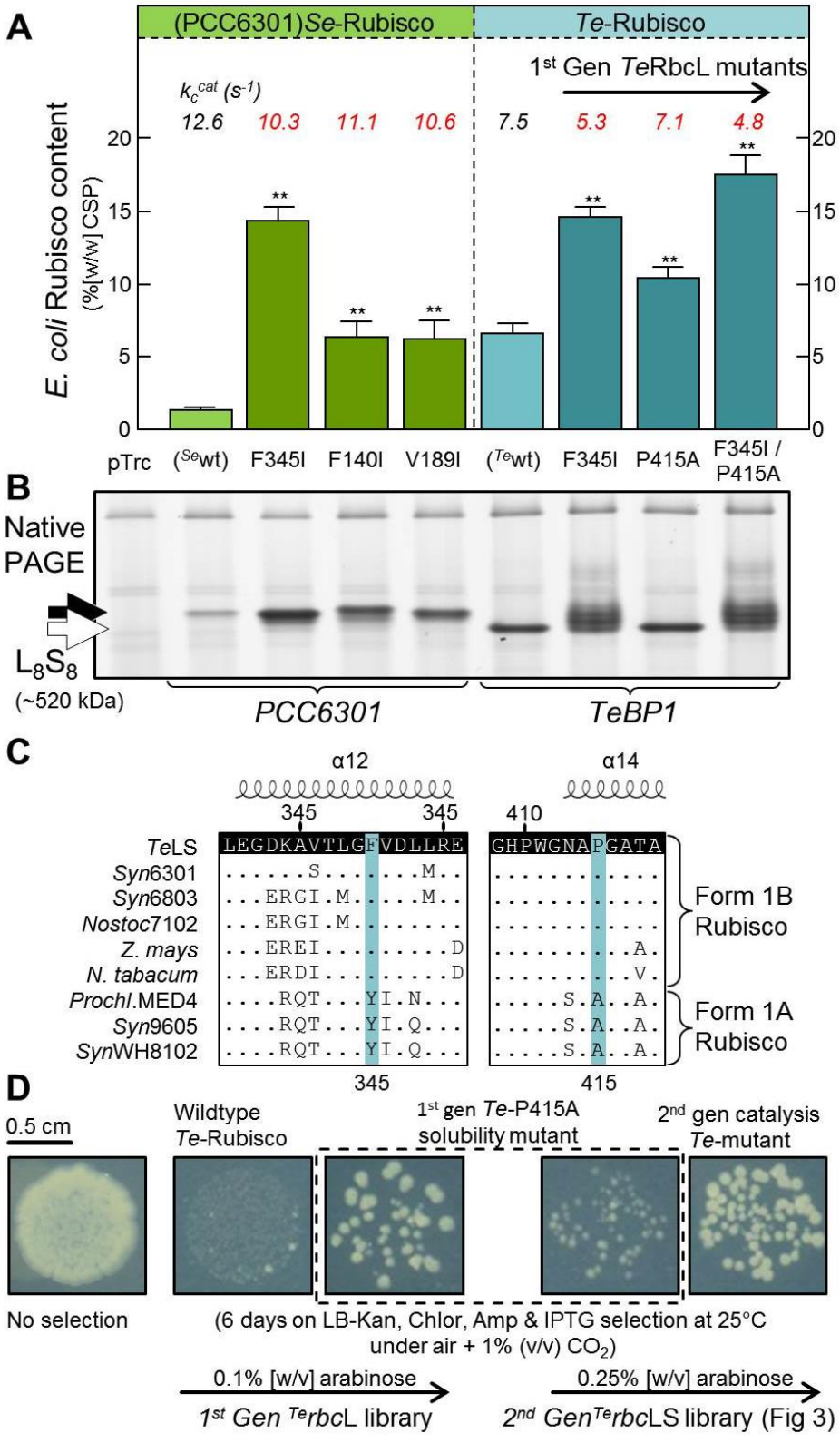

Figure 2

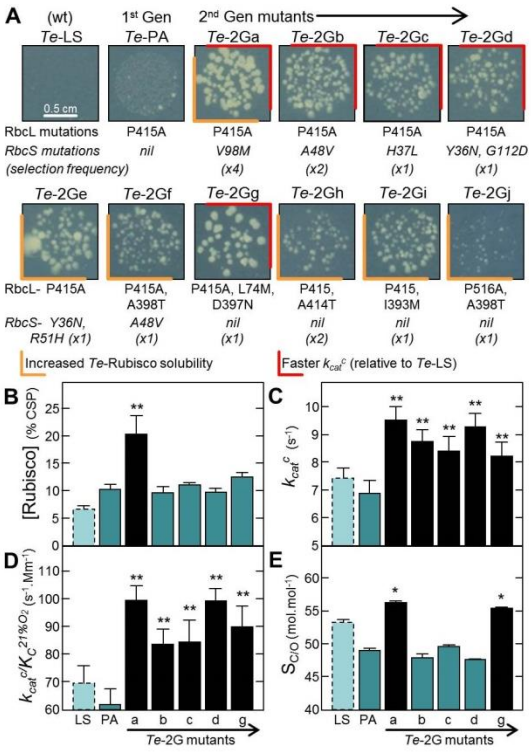

Figure 3

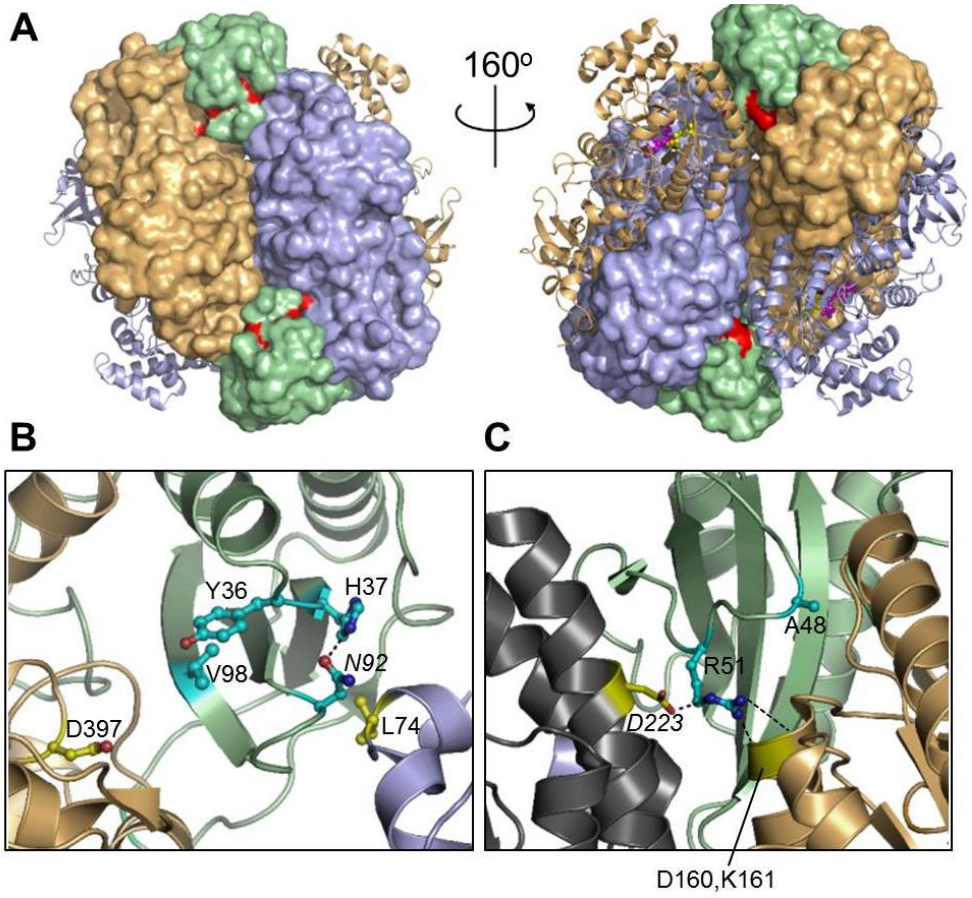

Figure 4

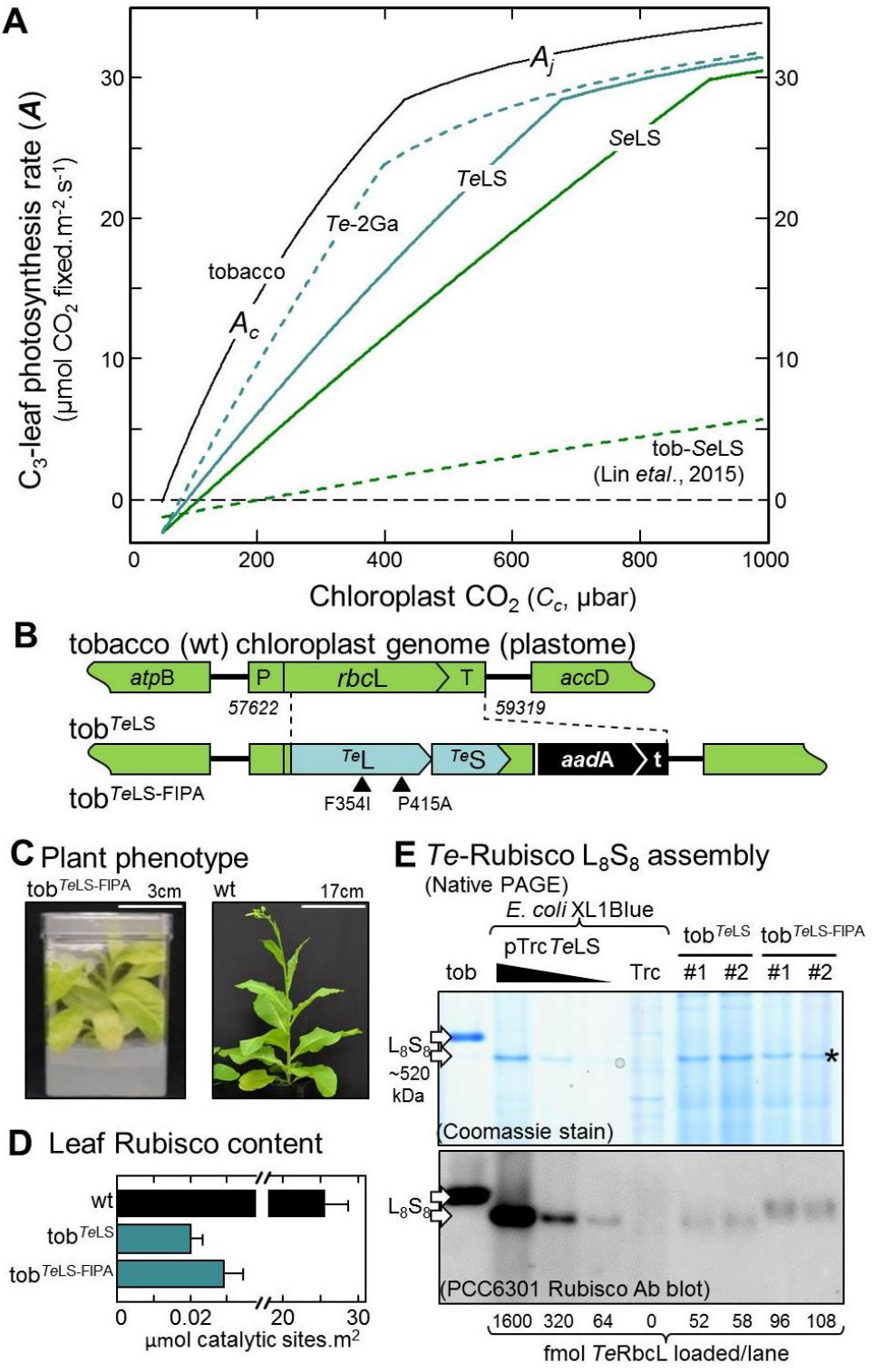

Figure 5

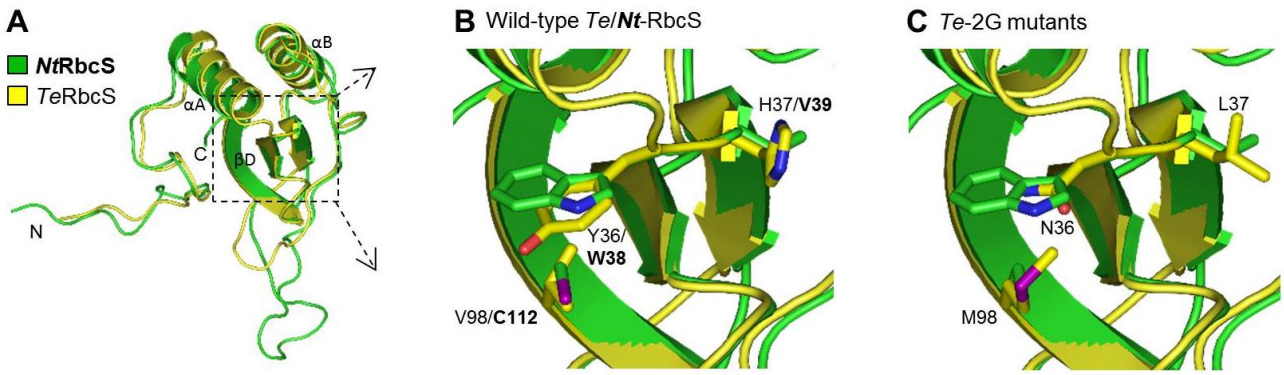

Figure 6

An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO_2-fixation kinetics.

Robert H. Wilson, Elena Martin-Avila, Carly Conlan and Spencer M. Whitney

J. Biol. Chem. *published online October 6, 2017*
