Small subunits can determine enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli*

Myat T. Lin[1], William D. Stone[1,2], Vishalsingh Chaudhari[1], Maureen R. Hanson[1]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, 14853
[2]Present address: Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, 20723

Abstract

Rubisco catalyses the first step in carbon fixation and is a strategic target to improve photosynthetic efficiency. In plants, Rubisco is composed of eight large and eight small subunits and its biogenesis requires multiple chaperones. We optimised a system to produce tobacco Rubisco in *Escherichia coli* by co-expressing chaperones in auto-induction medium. We successfully assembled tobacco Rubisco in *E. coli* with each small subunit that is normally encoded by the nuclear genome. Even though each enzyme carries only a single type of small subunit in *E. coli*, the enzymes exhibit carboxylation kinetics very similar to that of the native Rubisco. Tobacco Rubisco assembled with a recently discovered trichome small subunit has a higher catalytic rate and a lower CO_2 affinity than those assembled with other small subunits. Our *E. coli* expression system will allow probing of features of both subunits of Rubisco that affect its kinetic properties.

Introduction

In photosynthetic organisms including plants, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco, EC 4.1.1.39) catalyses the fixation of CO_2 from air to ribulose-1,5-bisphosphate (RuBP). Thus, Rubisco represents a major gateway for inorganic carbon to enter the biosphere and has far-reaching impacts on a global scale. The carboxylation of RuBP by Rubisco is a well-known bottleneck in photosynthesis of C_3 plants because of its slow catalytic turnover rate ($k_{cat} \approx$ 2-3 $^{-1}$)[1]. Rubisco also catalyses a competing RuBP oxygenation process that generates a toxic byproduct, 2-phosphoglycolate, which has to be recycled through the energy-intensive photorespiration pathway[2]. The efficiency loss due to the RuBP oxygenation depends on temperature, ambient CO_2 and other environmental conditions and has been estimated to lower the yield of C_3 crops by about 20-36%[3].

The minimal catalytic unit in Rubisco is composed of two large subunits (RbcLs) arranged in antiparallel orientation where two active sites are formed at the dimeric interface[1]. Plants, cyanobacteria and many autotrophic prokaryotes possess form I Rubisco, which is usually represented by L_8S_8 (8 large and 8 small subunits) as it is composed of four RbcL dimers or $(L_2)_4$ encoded by a single chloroplast gene and eight small subunits (RbcSs) expressed from a family of nuclear genes in plants[1]. Dinoflagellates, many proteobacteria and archaea possess forms II, III or II/III Rubisco without RbcS, and a recent survey highlighted their wide range of kinetic capabilities[4].

An important kinetic parameter of Rubisco is its CO_2/O_2 specificity factor ($S_{C/O}$), which is the ratio of the catalytic efficiencies of its carboxylation reaction (k_{cat}^{c}/K_C) and oxygenation reaction (k_{cat}^{o}/K_O), where k_{cat}^{c} and k_{cat}^{o} are the catalytic rates and K_C and K_O are Michaelis-Menten constants for CO_2 and O_2 respectively[5]. Based on data from form I green Rubisco lineage, a Rubisco that binds its carboxylated intermediates tighter is associated with delayed formation and release of products from its catalytic site leading to a higher $S_{C/O}$, a lower K_C and a lower k_{cat}^{c}[6,7]. Due to the lack of a carbon-concentrating mechanism (CCM), C_3 Rubisco enzymes typically have higher $S_{C/O}$ and lower k_{cat}^{c} than C_4 enzymes[8]. To overcome the low catalytic rates, C_3 plants have evolved to produce much more Rubisco and invest as much as 25% of leaf nitrogen in just one enzyme[9]. The enzymes in diatoms, which belong to non-green or red form I Rubisco, were found to possess kinetic parameters that indicate the presence of varying strength of CCM and deviate from the established relatioship[10].

RbcLs are prone to aggregation and require GroEL/ES-type chaperonin for proper folding followed by assistance from additional chaperones for dimerization and, in the case of form I Rubisco, subsequent assembly with RbcSs to form the L_8S_8 holoenzyme[11]. Form I Rubisco from certain cyanobacteria can be assembled in *E. coli* and *Nicotiana tabacum* (tobacco) chloroplasts without additional chaperones[12-15]. On the other hand, plant Rubisco has strict assembly requirements that can only be met by closely related species[16]. The Rubisco subunits from red algae and RbcLs from monocot Paniceae grasses failed to assemble into active enzyme in tobacco due to lack of compatible assembly factors[17-19]. Previous studies on maize and Arabidopsis photosynthesis mutants lacking Rubisco identified three essential chaperones for Rubisco assembly: BSD2 (bundle-sheath defective 2), RAF1 and RAF2 (Rubisco accumulation factors 1 and 2)[20-23]. Subsequent structural studies have revealed the mechanism of RAF1, RBCX and BSD2 in assisting the dimerization of RbcL and subsequent stabilisation of the L_8 core before the formation of the final holoenzyme with RbcSs[24-27].

In a recent breakthrough, Aigner and co-workers were able to produce functional Arabidopsis Rubisco in *E. coli* by co-expressing five additional chaperones: RBCX, RAF1, RAF2, BSD2 and the chaperonin complex made up of Cpn60α, Cpn60β and Cpn20[24]. They also demonstrated the assembly of a small amount of tobacco Rubisco by replacing the Arabidopsis RAF1 with a tobacco homolog. Needless to say, this advance will greatly facilitate functional studies of plant Rubisco. While both nuclear and chloroplast transformation can be carried out in model plants Arabidopsis and tobacco, tobacco offers an advantage in Rubisco engineering because transgenic tobacco plants with modified Rubisco can be investigated in field trials to access their photosynthetic performance when standard agricultural practices are employed[28,29]. Thus, an *E. coli* expression system with improved assembly of tobacco Rubisco will be

useful to identify enzymes with enhanced catalytic properties whose performance can be tested in the field.

In the current study, we modified the expression vectors used by Aigner and co-workers[24] to co-express chaperonin and chaperones from Arabidopsis and tobacco, utilised an auto-induction culture system, and have considerably increased expression of tobacco Rubisco. While the previous study expressed only a single Arabidopsis or a single tobacco RbcS along with the chloroplast-encoded RbcL in *E. coli*, we individually expressed all eight non-identical members of the tobacco RbcS family to assess their abilities to assemble properly into active enzyme in *E. coli*. We then analysed the tobacco Rubisco enzymes produced with each individual RbcS in *E. coli*, observed whether the expected molecular mass was achieved, and measured their carboxylation kinetics. We found that the Rubisco produced in *E. coli* with a recently identified RbcS present in the secretory glands of tobacco trichomes has a considerably different kinetic profile[30].

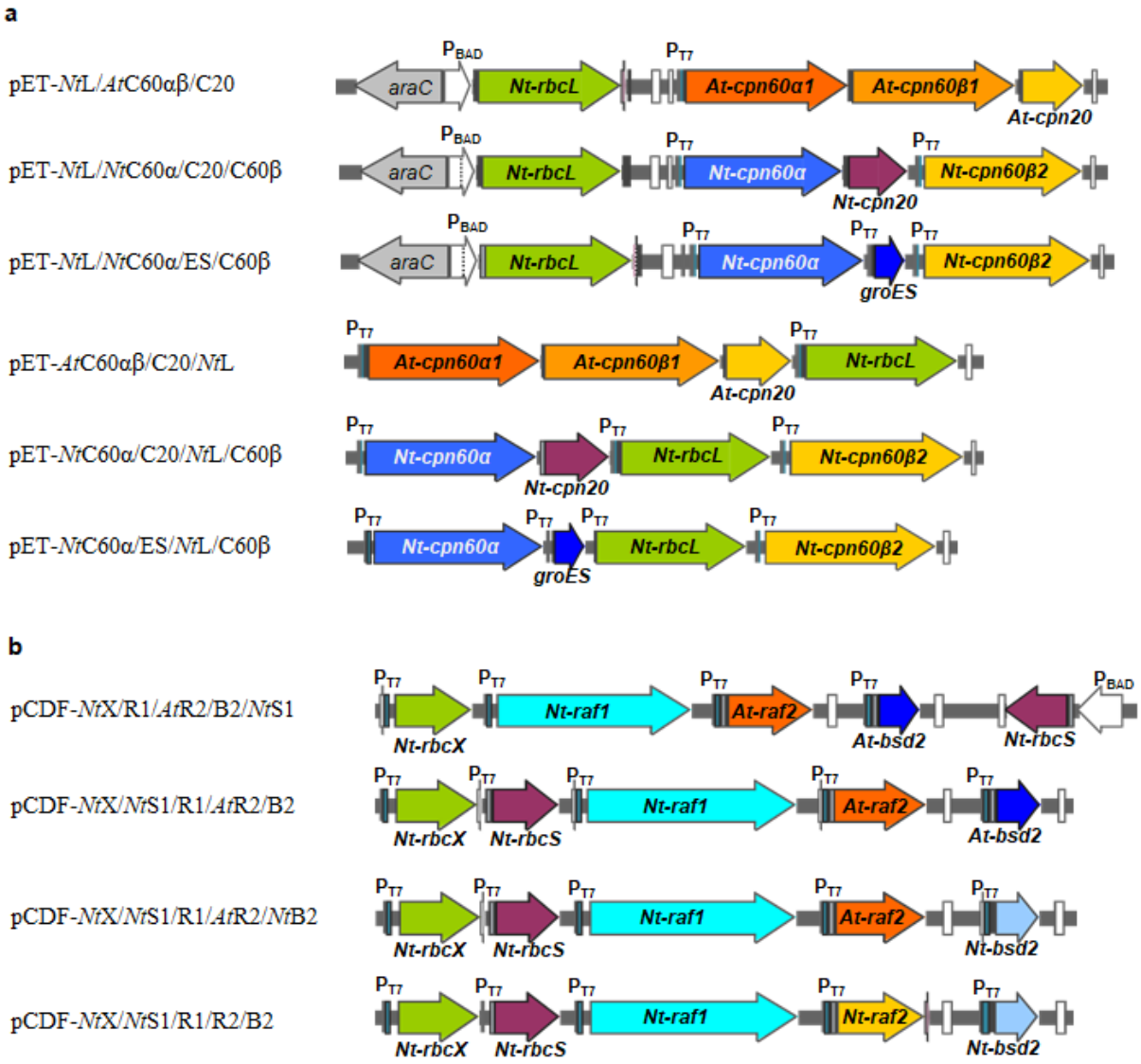


Fig. 1 | Gene arrangements in pET and pCDF *E. coli* expression vectors created in this study. a, pET vectors were used to express *Nt-rbcL* from either P_{BAD} or P_{T7} and *cpn60α*, *cpn60β* and either *cpn20* or *groES* from P_{T7}. **b**, pCDF vectors were used to express *Nt-rbcS* from either P_{BAD} or P_{T7} and *rbcX*, *raf1*, *raf2* and *bsd2* from P_{T7}.

Results

A modified two-vector system expresses tobacco Rubisco in *E. coli*. The pET and pCDF expression vectors we used in this study are based on the respective p*At*C60αβ/C20 and p*Nt*R1/*At*R2/B2 vectors created in a previous study[24]. Instead of expressing the two tobacco Rubisco subunits from a third vector, we introduced the tobacco *rbcL* and *rbcS-S1* genes into the pET and pCDF vectors, respectively. The pET vectors consist of three combinations of *cpn60α*, *cpn60β* and *cpn20* genes from either Arabidopsis (*At*) or tobacco (*Nt*) or *groES* from *E. coli* under T7 promoters. One set of pET vectors have the tobacco *rbcL* gene under an arabinose-inducible P_{BAD} promoter, while a T7 promoter controls the tobacco *rbcL* gene in the other set of pET vectors (Fig. 1a). The final gene arrangements in the vectors expressing Arabidopsis chaperonin are different from those expressing tobacco chaperonin. The four pCDF vectors contain tobacco *rbcX* and *raf1* genes as well as *raf2* and *bsd2* genes from either Arabidopsis or tobacco under T7 promoters (Fig. 1b). The first pCDF vector has the tobacco *rbcS-S1* gene controlled by a P_{BAD} promoter, while the other three have the *rbcS-S1* gene under a T7 promoter. Our modified two-vector system is compatible with *E. coli* expression strains that have extra tRNAs for rare *E. coli* codons such as Rosetta (DE3).

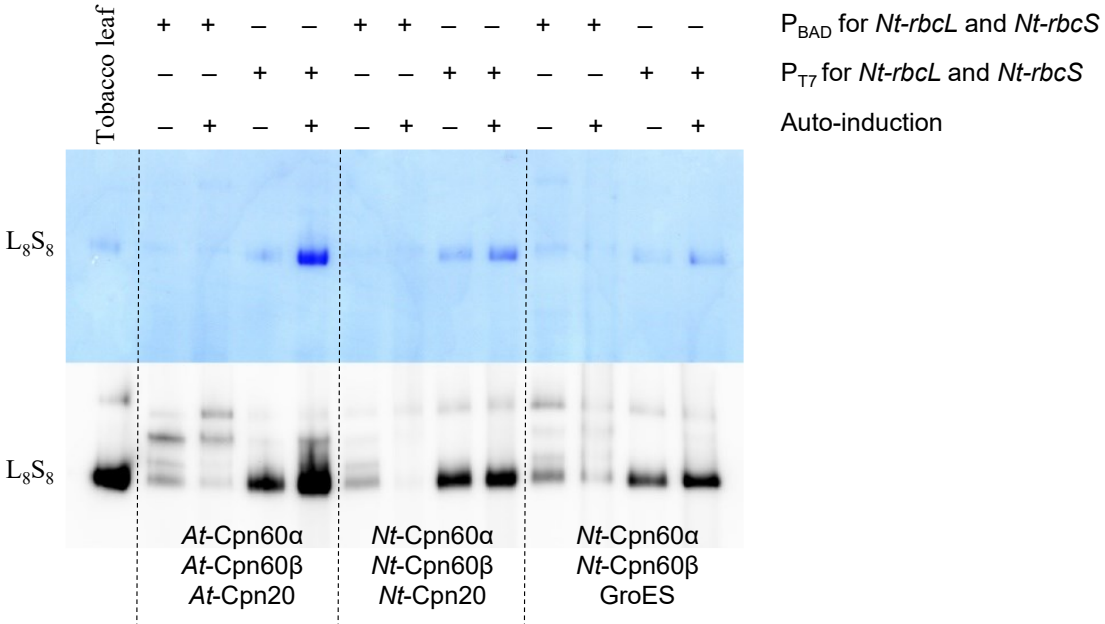


Fig. 2 | The native PAGE analysis of BL21 Star™ (DE3) *E. coli* soluble extracts with tobacco Rubisco subunits expressed from either P_{BAD} or P_{T7} promoter. The top panel shows Coomassie blue staining, and the bottom panel shows the immunoblot with the antibody against Rubisco. Expression was induced at 23 °C with either IPTG and/or arabinose in LB medium or auto-induced in ZYP-5052 medium for 18-20 h. Cpn60α, Cpn60β and Cpn20 from either Arabidopsis (At) or tobacco (Nt) or GroES from *E. coli* were co-expressed. Additional chaperones co-expressed were Nt-RBCX, Nt-RAF1, At-RAF2 and At-BSD2. *E. coli* pellets from 6 mL cultures were each resuspended in 500 µL lysis buffer and 7.5 µL of each soluble extract was loaded onto the gel. The native PAGE was performed once for the same set of samples. These same samples were also included in quantification Rubisco active sites (Fig. 3a,b).

T7 promoters under auto-induction conditions. For the Rubisco genes controlled by P_{BAD} promoters, we followed the previously described two-step expression protocol to express the chaperones with induction by isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours at 23 °C followed by the subsequent expression of the Rubisco subunits in a fresh LB medium containing arabinose[24]. We also

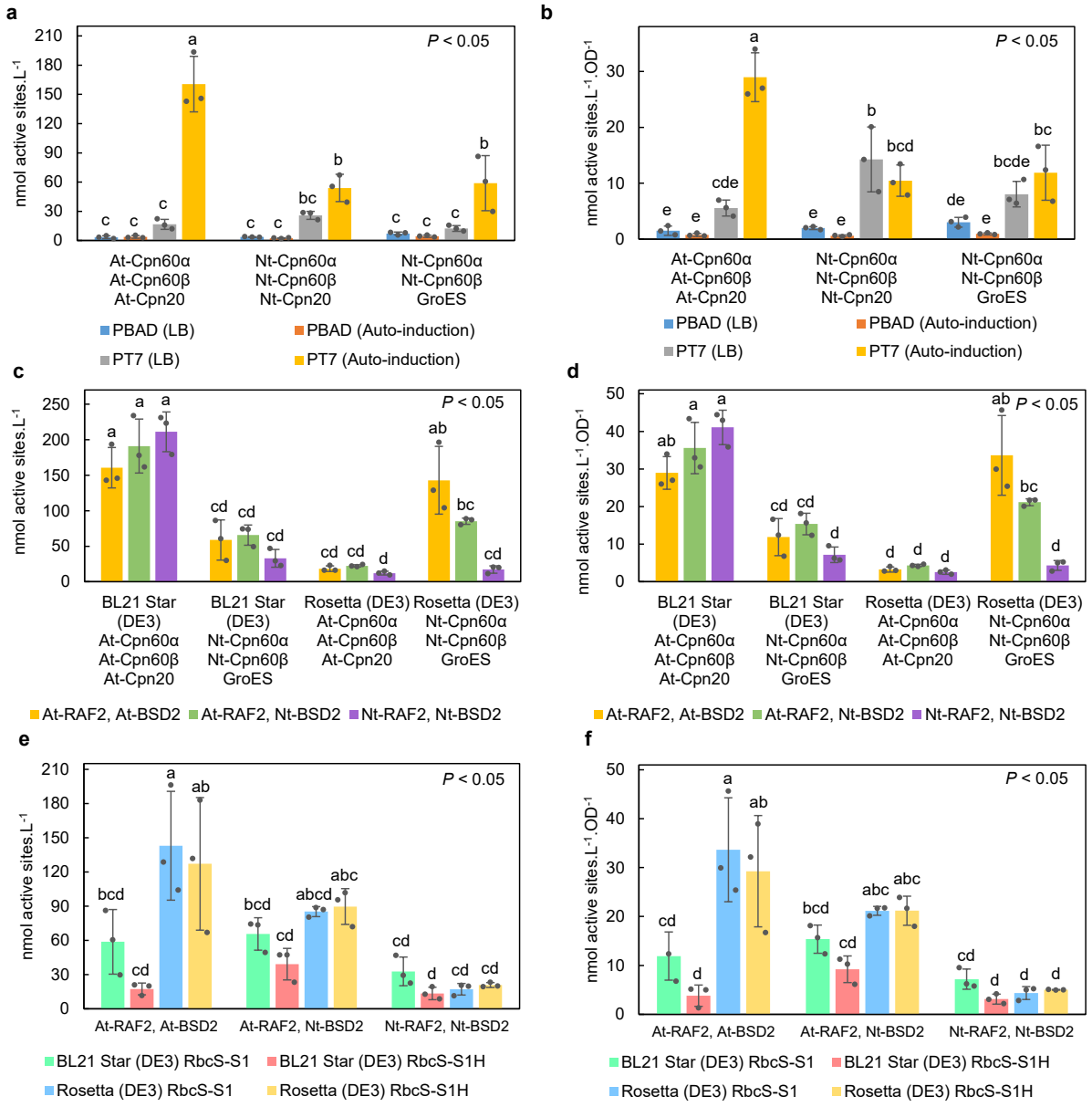


Fig. 3 | Comparison of Rubisco yields from *E. coli* under different expression conditions. a,**b**, *Nt-rbcL* and *Nt-rbcS-S1* were expressed in BL21 Star™ (DE3) *E. coli* from either P_{BAD} or P_{T7} promoter and induced with either IPTG and/or arabinose in LB medium or auto-induced with lactose-containing medium (ZYP-5052) supplied with arabinose. **c**,**d**, *Nt-rbcL* and *Nt-rbcS-S1* were expressed in either BL21 Star™ (DE3) or Rosetta (DE3) *E. coli* from P_{T7} promoter with ZYP-5052 medium along with different combinations of Arabidopsis and tobacco chaperones. **e**,**f**, *Nt-rbcL* and either *Nt-rbcS-S1* or *Nt-rbcS-S1H*, which encodes RbcS-S1 with a 6xHis tag, were expressed in either BL21 Star™ (DE3) or Rosetta (DE3) *E. coli* from P_{T7} promoter with ZYP-5052 medium. Cpn60α, Cpn60β, Cpn20, RAF2 and BSD2 were from either Arabidopsis (At) or tobacco (Nt) as indicated in individual panels. In some *E. coli* strains, GroES from *E. coli* were co-expressed instead of Nt-Cpn20. Additional chaperones co-expressed were Nt-RBCX and Nt-RAF1. Nt-Cpn60α, Nt-Cpn60β and GroES were co-expressed in **e** and **f**. Expression was performed at 23 °C. Rubisco active sites were quantified from bound ^{14}C-CABP separated in size-exclusion chromatography and normalized with culture volume (L) in **a**, **c**, **e** or with *E. coli* cell density per L in **b**, **d**, **f**. The error bars indicate mean values and standard deviations from measurements with three *E. coli* growth experiments for each condition in 6 mL culture media. In each chart, data were analyzed with one-way ANOVA, and the P-value was obtained from F-statistics. Tukey's honest significance test was then carried out in each chart, and samples with *P* value > 0.05 are indicated with the same letter.

expressed both tobacco RbcL and RbcS from T7 promoters (P_{T7}) because a recent study achieved a six-fold increase in Rubisco production with simultaneous expression of the Arabidopsis Rubisco subunits and chaperones from T7 promoters[31]. In addition, we tested auto-induction of gene expression from T7 promoters using a rich medium called ZYP-5052 that contains phosphate buffer along with glycerol, glucose and lactose as carbon sources[32]. The *E. coli* cultures grown in the ZYP-5052 medium automatically induce gene expression as the cells begin to metabolize lactose when the medium runs out of glucose. Minimal leaky expression prior to induction and relatively stable pH typically allowed the auto-induction medium to achieve higher cell density[32]. Our BL21 Star™ (DE3) cultures expressing the tobacco Rubisco in the auto-induction medium typically had final cell densities of ~ 5-6 OD_{600} (optical density measured as absorbance at 600 nm) compared to ~2 OD_{600} in LB medium.

We analysed the soluble extracts of the *E. coli* cultures on native PAGE (Polyacrylamide Gel Electrophoresis) and immunoblotting with an antibody against Rubisco. Consistent with the recent study, we found that using T7 promoters for the Rubisco genes gave higher levels of L_8S_8 Rubisco complex than using arabinose-inducible P_{BAD} promoters[31]. Using T7 promoters with auto-induction medium resulted in the highest yield of tobacco Rubisco (Fig. 2). Interestingly, co-expressing Arabidopsis chaperonin, Cpn60α, Cpn60β and Cpn20, provided better production of tobacco Rubisco than co-expressing tobacco chaperonin with either Cpn20 or GroES under auto-induction conditions.

We quantified the Rubisco active sites in the *E. coli* extracts (n = 3 or measurements on three cultures for each condition) with Rubisco-bound ^{14}C-carboxyarabinitol bisphosphate (CABP). The results were generally consistent with the observation from immunoblotting of the native PAGE, with the T7 promoter providing higher Rubisco levels than the P_{BAD} promoter (Fig. 3a,b). When induced with IPTG in LB medium, the tobacco chaperonin system with Cpn60α, Cpn60β and Cpn20 gave better Rubisco yields than either the Arabidopsis chaperonin or tobacco Cpn60 subunits with GroES. The ZYP-5052 auto-induction medium further improved the Rubisco expression in all cases that used T7 promoters to express the Rubisco subunits. However, for the two *E. coli* strains with tobacco Cpn60 subunits, the improved Rubisco yields were entirely due to the higher final cell densities attained in the auto-induction medium (Fig. 3a,b). Remarkably, the *E. coli* strain with the Arabidoposis chaperonin grown in the auto-induction medium produced much higher levels of tobacco Rubisco from the T7 promoter even when the yields were normalised with the final cell densities (Fig. 3b).

BL21 Star™ (DE3) vs. Rosetta (DE3) with different chaperones. Since our two-vector expression system is compatible with Rosetta (DE3) that has extra tRNAs for rare *E. coli* codons, we compared the performance of BL21 Star™ (DE3) and Rosetta (DE3) with different combinations of chaperones. Arabidopsis chaperonin, Cpn60α, Cpn60β and Cpn20, again gave the highest levels of tobacco Rubisco in BL21 Star™ (DE3) under auto-induction condition regardless of the presence of Arabidopsis or tobacco RAF2 and BSD2, while the same sets of vectors in Rosetta (DE3) produced about an order of magnitude lower amounts of Rubisco (n = 3) (Fig. 3c,d). In contrast, Rosetta (DE3) co-expressing tobacco Cpn60α, Cpn60β, *E. coli* GroES, and Arabidopsis RAF2 and BSD2 also produced Rubisco levels comparable to that in BL21 Star™ (DE3) with Arabidopsis chaperonin proteins. The yields normalised with culture volume generally had the same profile as those produced per unit cell, as the final cell densities under auto-induction conditions were similar across all samples (Fig. 3c,d).

6xHis tag on RbcS did not increase the yield. We did not observe any improvement in the yield of tobacco Rubisco when RbcS-S1H, which was fused with a 6xHis tag, was co-expressed with tobacco Cpn60α, Cpn60β and *E. coli* GroES in either BL21 Star™ (DE3) or Rosetta (DE3) under auto-induction conditions regardless of the sources of RAF2 and BSD2 (Fig. 3e,f). In fact, RbcS-S1H reduced the Rubisco yield in BL21 Star™ (DE3). Since the tobacco RAF2 performed more poorly than the Arabidopsis RAF2 in most cases, we did not co-express the tobacco RAF2 with the Arabidopsis BSD2 in our study.

Tobacco Rubisco with different RbcSs. In higher plants, the Rubisco RbcSs are expressed from multiple nuclear genes, usually encoding slightly different amino acid sequences. Being an allotetraploid, tobacco inherited two sets of *rbcS* genes from its parents *N. sylvestris* (“S” lines) and *N. tomentosiformis* (“T” lines)[33]. We analysed their expression levels using publicly available RNA-Seq data generated from tobacco leaf samples[34] and found that eleven of them are being transcribed, with *Nt-rbcS-T1* having the highest levels of transcripts for both young and mature leaves and *Nt-rbcS-S1a*, *Nt-rbcS-S1b*, *Nt-rbcS-S2* and *Nt-rbcS-S5* making up most of the remaining transcripts (Fig. 4a). Several of these genes encode identical mature small subunits, resulting in seven unique RbcSs: RbcS-S1 (from *Nt-rbcS-S1a* and *Nt-rbcS-S1b*), RbcS -S2, RbcS-S5, RbcS -T1, RbcS -T2 (from *Nt-rbcS-S3*, *Nt-rbcS-S4*, *Nt-rbcS-T2* and *Nt-rbcS-T3a*), RbcS-T4a and RbcS-T5 (Fig. 4b). In addition to these canonical RbcSs, recent studies discovered a phylogenetically diverse line of RbcS, denoted hereafter as RbcS-Tc, that is mainly expressed in non-photosynthetic tissues of plants including tobacco trichomes[30,35,36].

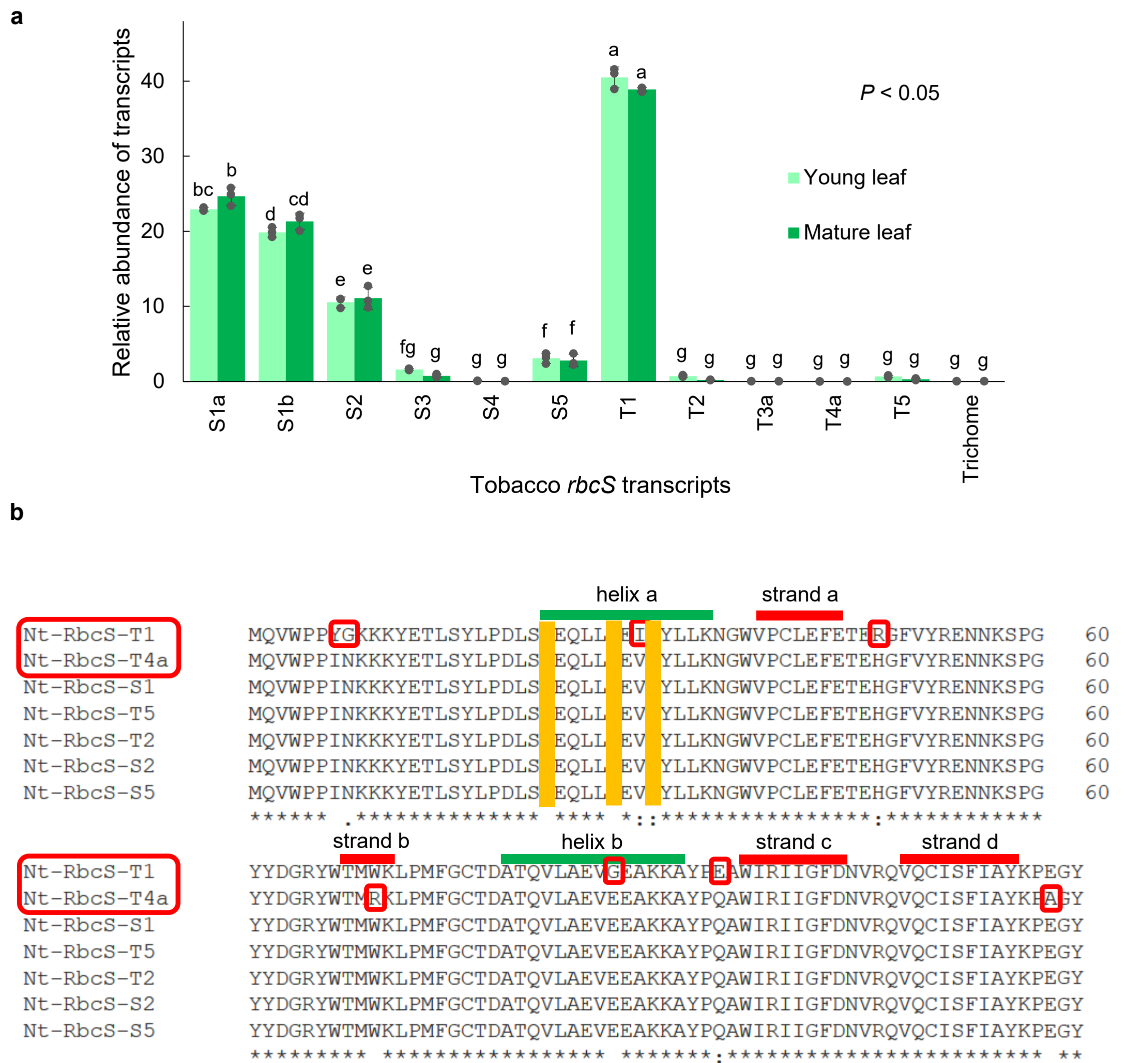


Fig. 4 | Survey of the Rubisco small subunits in tobacco. a, Comparison of transcript abundances estimated from an RNA-Seq experiment of tobacco leaf tissue[34]. The error bars represent mean values and standard deviations obtained from three SRA files. Data were analyzed with one-way ANOVA, and P-value was obtained from F-statistics. Tukey’s honest significance test was then carried out, and samples with P value > 0.05 are indicated with the same letter. **b**, Multiple sequence alignment of seven unique tobacco small subunits with Clustal Omega. The chloroplast transit peptides are not included. The three variable residues in helix a are highlighted in yellow. The unique residues found only in Nt-RbcS-T1 and Nt-RbcS-T4a are indicated with red rectangles.

We co-expressed the tobacco RbcL and each of these eight tobacco RbcS and RbcS-S1H, which has a 6xHis tag fused at the C-terminus of RbcS-S1, in BL21 Star™ (DE3) with Arabidopsis Cpn60α, Cpn60β, Cpn20 and RAF2 and tobacco RAF1, RBCX and BSD2 in either LB medium or ZYP-5052 auto-induction medium. For samples obtained with LB medium, we were able to detect an L_8S_8 Rubisco band at a slightly higher molecular weight (MW) than that from a tobacco leaf extract on a native PAGE immunoblot for six tobacco RbcSs (S1, S2, S5, T2, T5 and Tc) (Fig. 5a). A second band at around 720 kDa that corresponds to the RbcL-chaperonin complex was also present in most *E. coli* extracts. In samples with RbcS-T1, RbcS-T4a and RbcS-S1H, only a very strong 720-kDa band was observed on the immunoblot indicating Rubisco with those three RbcS failed to assemble properly in *E. coli* grown in LB medium. When ZYP-5052 auto-induction medium was used to express the Rubisco from the same BL21 Star™ (DE3) strains, all samples gave a 540-kDa L_8S_8 band very similar to that from a tobacco leaf extract, as well as a less pronounced RbcL-chaperonin complex, except for the sample with RbcS-T4a (Fig. 5b). Interestingly, when Rosetta (DE3) strains were used to co-express tobacco Cpn60α, Cpn60β and *E. coli* GroES in the auto-induction medium, the sample with RbcS-T1 failed to produce an L_8S_8 Rubisco complex (Extended Data Fig. 1). These results suggest that growth media, the *E. coli* expression strain and the chaperonin being co-expressed may influence Rubisco assembly in *E. coli*. For our *E. coli* expression system, the tobacco Rubisco expressed in ZYP-5052 auto-induction medium from BL21 Star™ (DE3) with Arabidopsis chaperonin resembled the native enzyme more closely than those obtained in LB medium or Rosetta (DE3) with tobacco chaperonin.

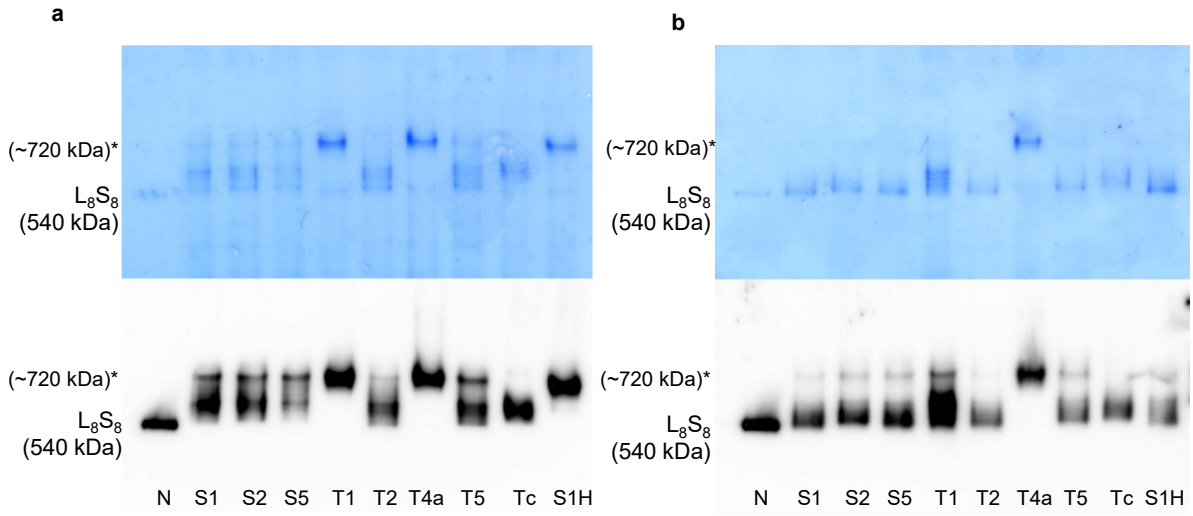


Fig. 5 | Native PAGE immunoblots of tobacco Rubisco expressed in BL21 Star™ (DE3) *E. coli* with different small subunits. a, Expression was induced with IPTG in LB medium at 23 °C for 18-20 h. **b**, Expression was auto-induced in ZYP-5052 medium at 23 °C for 18-20 h. For **a** and **b**, At-Cpn60α, At-Cpn60β, At-Cpn20, Nt-RBCX, Nt-RAF1, At-RAF2 and Nt-BSD2 were co-expressed in all samples. Approximately 5-10 µg of total soluble extract was loaded for each sample. Approximately 0.2 µg of total soluble extract from a young tobacco leaf (N) was also included as a control. The top panels show Coomassie blue staining, and the bottom panels show the immunoblots using the antibody against Rubisco. The bands for L_8S_8 Rubisco (540 kDa) and chaperonin-RbcL complex (~720 kDa, indicated with asterisks) are marked next to each panel. The native PAGE was performed once for each set of samples and at least twice for samples with RbcS-S1, RbcS-T1, RbcS-T4a, RbcS-Tc and RbcS-S1H with similar results.

Next, we compared the RuBP carboxylation activities of these *E. coli* extracts with different RbcSs. For RbcS-S5, T2, T5 and Tc, the samples from either LB or auto-induction medium had similar carboxylation rates as the native enzyme from a tobacco leaf (Fig. 6). For RbcS-S1 and RbcS-S5, the samples produced from LB medium had about 15-20% lower carboxylation rates, while those with RbcS-T1 and RbcS-S1H from LB medium had about 50-60% lower carboxylation rates than their respective samples from the auto-induction medium. In contrast, we could not detect any carboxylation activity from

both samples with RbcS-T4a, which lacked the L_8S_8 Rubisco band on the native PAGE. RbcS-T4a has unique residue substitutions, which may explain why it was unable to assemble active Rubisco holoenzyme in *E. coli* (Fig. 4b,5,6). Since the transcript level of *rbcS-T4a* is negligible, it is likely that RbcS-T4a is nonfunctional and physiologically irrelevant in tobacco. In the samples from LB medium with RbcS-T1 and RbcS-S1H, the 540 kDa L_8S_8 band was missing, and instead, a 720 kDa band with similar migration as the RbcL-chaperonin complex was greatly enhanced (Fig. 5a). The RbcL-chaperonin complex was previously shown to be mostly inactive[24], while these samples exhibit ~50% RuBP carboxylation activities. At present, the nature of the complex in the samples with RbcS-T1 and RbcS-S1H is not known, although it seems likely to be more advanced in assembly than the RbcL-chaperonin complex, perhaps trapped in a misfolded configuration with respective RbcS and/or possibly one or more chaperones, such as BSD2 that are involved in a late assembly stage.

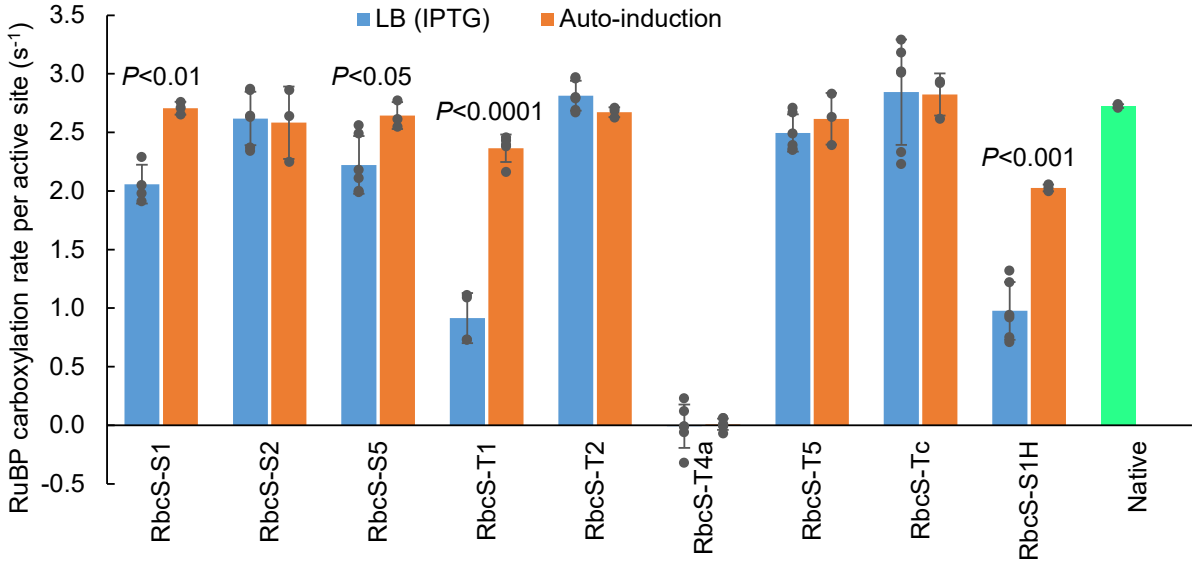


Fig. 6 | RuBP carboxylation rates at 42 µM [CO_2] for tobacco Rubisco with individual small subunits expressed in *E. coli* and induced with either IPTG in LB medium or auto-induced in ZYP-5052 medium. At-Cpn60α, At-Cpn60β, At-Cpn20, Nt-RBCX, Nt-RAF1, At-RAF2 and Nt-BSD2 were co-expressed. The RuBP carboxylation rate of native Rubisco from tobacco leaf tissue measured at 37 µM [CO_2] was also included for comparison. Carboxylation rates were measured from incorporation of ^{14}C into RuBP in the absence of O_2 at 25 °C and normalized with Rubisco active sites quantified with bound ^{14}C-CABP separated in size-exclusion chromatography. The error bars represent mean values and standard deviations of measurements from three to six *E. coli* growth experiments for each condition. P-values < 0.05 from two-tailed homoscedastic t-tests between the samples from LB medium and ZYP-5052 medium are indicated.

Compared to the ZYP-5052 medium, LB medium did not contain phosphate buffer for pH stability, nor three carbon sources (glycerol, glucose and lactose) to automatically switch carbon metabolism, nor Mg and trace metals. Since BL21 Star™ (DE3) expressing the Rubisco with RbcS-T1 in LB medium had greatly compromised enzymatic activities and complex configuration, we explored the expression of the Rubisco with RbcS-S1 and RbcS-T1 in three modified media: (1) ZYP, which lacks the three carbon sources and requires IPTG for induction, (2) ZYP-5052ΔMg, which lacks Mg, and (3) ZYP-5052Δmetals, which lacks trace metals. The correct band for the Rubisco complex was missing on the native PAGE for the RbcS-T1 Rubisco samples obtained with ZYP and ZYP-5052ΔMg media, similar to that from LB medium (Extended Data Fig. 2). The same samples also displayed markedly lower carboxylation rates, indicating that both auto-induction conditions provided by the carbon sources and the addition of Mg in the ZYP-5052 medium were responsible for enhanced assembly of tobacco Rubisco

with RbcS-T1. In contrast, the tobacco Rubisco with RbcS-S1 produced from all three modified media had the same carboxylation activities and displayed a proper band for the L_8S_8 Rubisco complex on the native PAGE (Extended Data Fig. 2).

We also measured the RuBP carboxylation rates at six different dissolved CO_2 levels for each BL21 Star™ (DE3) *E. coli* extract with different RbcS and fitted the data to the standard Michaelis-Menten model to estimate their K_C and k_{cat} values (Fig. 7a). We found that the enzymes produced in the auto-induction medium with each of the five RbcSs (S1, S2, S5, T2 and T5) had an overall carboxylation kinetic profile that is statistically similar to the native Rubisco from a tobacco leaf although their average K_C values were slightly higher (Fig. 7b, Table 1). In contrast, the enzyme with RbcS-T1 produced in the auto-induction medium and that with RbcS-S1 produced in LB medium had lower k_{cat} values. Expressing the enzyme with RbcS-T1 in LB medium further lowered its k_{cat} by ~50%. We also found that the enzyme with RbcS-S1H that has a 6xHis tag fused to the C-terminus of RbcS-S1 had an average k_{cat} that was ~30% lower than the enzyme with untagged RbcS-S1. Expressing the enzyme with RbcS-S1H in LB medium further lowered its k_{cat}, which was similar to that of the enzyme with RbcS-T1 produced in LB medium. This is consistent with the significantly lower carboxylation activities of both the enzymes with RbcS-T1 and RbcS-S1H produced in LB medium (Fig. 6).

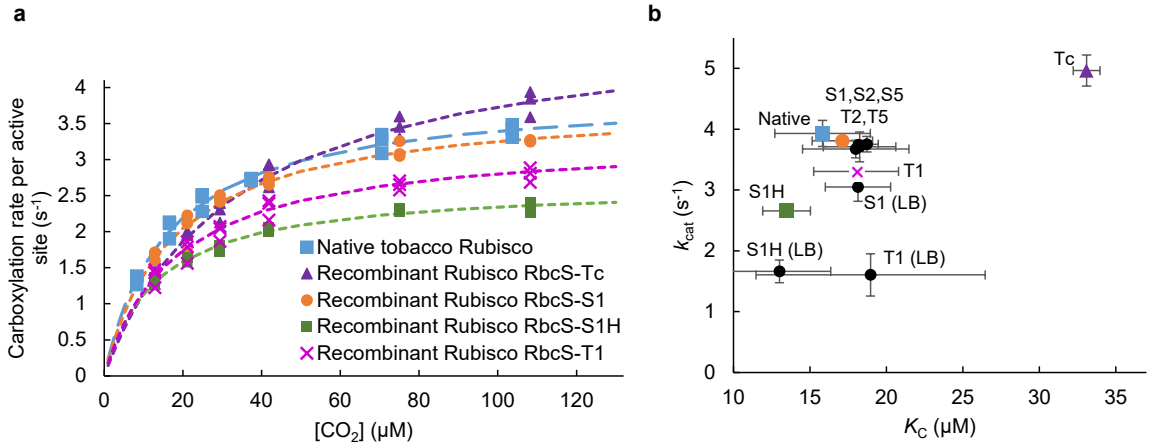


Fig. 7 | The enzyme kinetics of tobacco Rubisco expressed in *E. coli* with different small subunits compared to the native tobacco Rubisco. a, At-Cpn60α, At-Cpn60β, At-Cpn20, Nt-RBCX, Nt-RAF1, At-RAF2 and Nt-BSD2 were co-expressed in BL21 Star™ (DE3) either in LB or ZYP-5052 auto-induction medium at 23 °C for 18-20 h. RuBP carboxylation rates were measured from incorporation of ^{14}C into RuBP in the absence of O_2 at 25 °C. [CO_2] in the reaction mixtures ranged from 12.9 to 108.3 μM for the *E. coli* extracts and 8.3 to 103.7 μM for tobacco leaf extracts. Rubisco active sites were quantified with bound ^{14}C-CABP separated in size-exclusion chromatography. The data were fitted to the Michaelis-Menten equation with nonlinear regression. The fitted models with mean parameters obtained from three *E. coli* growth experiments are shown as dashed lines. Data with additional *E. coli* samples are included in Extended Data Fig. 3. **b**, The Michaelis-Menten constants for CO_2, K_C, and turnover numbers, k_{cat}, were obtained from nonlinear regression with the error bars representing the mean values and standard deviations of fitted models from three *E. coli* growth experiments. Their values are summarized in Table 1.

RbcS-Tc assembled Rubisco with different kinetic properties. We co-expressed the tobacco trichome small subunit, RbcS-Tc, in BL21 Star™ (DE3) with the same set of chaperonin and chaperones as described above. The corresponding *E. coli* extract had a normal L_8S_8 Rubisco band on the native PAGE immunoblot and displayed robust RuBP carboxylation activities (Fig. 5,6). The tobacco Rubisco with RbcS-Tc produced in *E. coli* had carboxylation kinetics that is markedly different from the native enzyme found in mesophyll tissue as well as those assembled in *E. coli* with other RbcS (Fig. 7). Its higher turnover number and lower affinity for CO_2 are consistent with the recent findings concerning tobacco trichome Rubisco[30].

Table 1. The enzyme kinetics of tobacco Rubisco produced in *E. coli* with different RbcS. The averages and standard deviations were obtained from three biological repeats, and *P*-values were from two-tailed homoscedastic t-tests in comparison to the native enzyme. The growth conditions were the same as those in Fig. 7.

RbcS Present	K_C (μM)	P-value	k_{cat} (s^{-1})	*P*-value
Native	15.8 ± 3.1		3.9 ± 0.2	
S1	17.1 ± 2.0	0.5745	3.8 ± 0.0	0.3790
S2	18.3 ± 2.4	0.3445	3.7 ± 0.2	0.3076
S5	18.7 ± 0.7	0.1921	3.8 ± 0.1	0.2872
T1	18.0 ± 2.8	0.3704	3.3 ± 0.0	0.0020
T2	18.2 ± 1.7	0.3161	3.7 ± 0.1	0.2267
T5	18.0 ± 3.5	0.4657	3.7 ± 0.1	0.1604
Tc	33.1 ± 0.9	0.0008	5.0 ± 0.3	0.0058
S1H	13.5 ± 1.6	0.3068	2.7 ± 0.1	0.0006
S1 (LB)	18.1 ± 2.1	0.3485	3.1 ± 0.3	0.0088
T1 (LB)	19.0 ± 7.5	0.5428	1.6 ± 0.3	0.0024
S1H (LB)	13.0 ± 3.4	0.3467	1.7 ± 0.2	0.0002

Discussion

Aigner and co-workers recently overcame a major obstacle in functional studies of plant Rubisco by successfully expressing active Arabidopsis Rubisco in *E. coli*[24]. We modified two of the expression vectors used in their studies such that the tobacco RbcL and RbcS were co-expressed with five chaperones. Consistent with a recent study[31], expressing the Rubisco subunits from T7 promoters produced more enzymes than expressing from P_{BAD} promoters (Fig. 3a,b). Using an auto-induction medium with the T7 promoters allowed the cultures to achieve approximately three-fold higher cell densities and further increased the production of Rubisco compared to LB medium. Surprisingly, we achieved the highest yields of tobacco Rubisco when Arabidopsis chaperonin, namely Cpn60α, Cpn60β and Cpn20, was co-expressed as opposed to the tobacco chaperonin. The operon structure in the pET vector that expresses the Arabidopsis chaperonin is different from that for the tobacco chaperonin (Fig. 1a). Thus, the poor performance of the tobacco chaperonin in the auto-induction medium could be due to suboptimal expression of those proteins, and it will be useful to quantify each protein in *E. coli* extracts in the future by either mass spectrometry or immunoblotting.

We also compared RAF2 and BSD2 from Arabidopsis and tobacco in our *E. coli* system. In many cases, the production of tobacco Rubisco was not hindered when Arabidopsis chaperones RAF2 and BSD2 were co-expressed instead of the tobacco versions (Fig. 3c,d). This is not surprising for BSD2 since a recent study showed that BSD2 from Arabidopsis or maize was able to fully restore the Rubisco assembly in RNAi-bsd2 tobacco plants[37]. Overall, our results indicate that it sometimes may not be necessary to co-express the Rubisco subunits with RAF2 and BSD2 chaperones from the same plant species in order to achieve robust production of Rubisco in *E. coli*. Previously, replacing tobacco RbcL with the Arabidopsis RbcL in tobacco chloroplasts achieved higher Rubisco levels when Arabidopsis RAF1 was co-expressed[17]. When tobacco Rubisco was assembled for the first time in *E. coli*, tobacco RAF1 was shown to improve its yield[24]. In our study, incorporation of tobacco RBCX may have further improved the expression of tobacco Rubisco in *E. coli*. We cannot explain why the auto-induction medium selectively boosted the expression of tobacco Rubisco in the presence of Arabidopsis chaperonin only in BL21 Star™ (DE3) and why the Arabidopsis RAF2 outperformed the tobacco RAF2 in most of our experiments. It will be important to further investigate the interchangeability of chaperones from different plant species when the *E. coli* expression system is expanded to more plant species in the future.

The composition of small subunits in the native Rubisco from a tobacco leaf may be heterogeneous, unlike the enzymes expressed in *E. coli*. In addition, native Rubisco subunits in higher

plants are known to undergo multiple conserved post-translational modifications (PTMs), and their exact roles are not well understood[38]. The conserved N-terminal acetylation of RbcL and possibly other PTMs were found to be missing in the Arabidopsis Rubisco expressed in *E. coli*[24]. Different PTM patterns could also partly explain slower migration of some tobacco Rubisco expressed in *E. coli*, and further analyses are needed to understand the roles of different PTMs on the Rubisco subunits. Nevertheless, we found that most tobacco Rubisco expressed in *E. coli* and the native Rubisco had very similar carboxylation kinetics except for the enzymes with RbcS-T1 and RbcS-S1H, which had on average ~15% and 30% lower k_{cat} respectively (Table 1). Fusing a 6xHis tag to an Arabidopsis RbcS was previously shown to facilitate purification of Arabidopsis Rubisco assembled in *E. coli* without adverse side-effect on its kinetics[31]. However, our findings suggest that a 6xHis tag on the tobacco RbcS-S1 can negatively affect the Rubisco kinetics.

The LB medium not only gave low yield of Rubisco in our expression system, it also failed to properly assemble the enzymes with RbcS-T1 and RbcS-S1H, resulting in dramatically compromised catalytic rates (Fig. 5,6). The stress subjected to *E. coli* cells grown in LB medium upon IPTG induction appears to be responsible for much lower final cell densities and adversely affected protein synthesis. Our modified expression vectors produced robust amounts of tobacco Rubisco in BL21 Star™ (DE3) under auto-induction conditions. Although Rosetta (DE3) was found to produce higher amounts of Rubisco when tobacco chaperonin was co-expressed (Fig. 3e,f), it was not able to properly assemble Rubisco with RbcS-T1 even under auto-induction conditions (Extended Data Fig. 1). Thus, it is important to explore multiple *E. coli* expression strains as well as induction conditions when the current *E. coli* expression systems are expanded for Rubisco from other plant species in the future. Overall, our results demonstrate that our *E. coli* system produced tobacco Rubisco enzymes that are functionally similar to the native enzyme, and can be useful in future studies to identify features of the protein sequence that alter enzyme kinetics, assembly or affinity for CO_2 or O_2.

RbcS-Tc found in tobacco trichomes was also able to assemble into functional Rubisco in *E. coli* with higher k_{cat} and K_C values than the native enzyme expressed in mesophyll tissue. RbcS-Tc belongs to a phylogenetically distinct line of small subunits named T-type and usually expressed in non-photosynthetic tissues with only about 65% sequence identity to the canonical small subunits recently named M-type[35,36]. Previous studies with a rice T-type homolog as well as the same tobacco trichome subunit indicated that these Rubisco enzymes exhibit higher k_{cat} and K_C values than those in mesophyll tissue[30,39]. Although the active sites are located within dimers of large subunits in the holoenzyme, it has long been known that RbcS can influence the carboxylation kinetics likely by modifying the conformation of the enzyme[40]. For example, studies in Chlamydomonas Rubisco indicated that changes along the interface between the RbcL and RbcS resulted in an altered CO_2/O_2 specificity factor and other kinetic parameters[41,42]. In another study, a transgenic rice expressing a small subunit from sorghum, a C_4 plant, had Rubisco with higher catalytic rates than the rice enzyme[43].

Phylogenetic analysis of Rubisco has revealed that the substitution rates in RbcS are much higher than those in RbcL in higher plants[44]. It is possible that diversification of RbcS represents one strategy that plants have employed to adjust the kinetics of Rubisco as they adapt to different environmental conditions. Some information has already become available from analysis of *Flaveria* species, in which both C_3 and C_4 carbon fixation exists[45] and a recent evolutionary study on a wider range of angiosperms[46]. Expression of variant and mutant RbcSs in *E. coli* in the future should provide a wealth of information concerning their roles in Rubisco enzyme kinetics and assembly.

One potential approach to identify Rubisco with improved kinetics is directed evolution using Rubisco-dependent *E. coli* strains. This approach has been applied for Rubisco from cyanobacteria, *Rhodospirillum rubrum*, *Methanococcoides burtonii* and *Rhodobacter sphaeroides* and can be extended to plant Rubisco by co-expressing the required chaperonin and chaperones in *E. coli* [47-50]. Our current study demonstrates that extension of the *E. coli* system to Rubisco enzymes from important crops should be feasible by incorporating appropriate chaperones and using suitable induction conditions.

Methods

Construction of *E. coli* expression plasmids. All the genes expressed in this study are summarized in Extended Data Fig. 4, and the *E. coli* expression vectors in Extended Data Fig. 5. All primers used were obtained from Integrated DNA Technologies and listed in Supplementary Table 1. Phusion™ high-fidelity DNA polymerase, FastDigest restriction enzymes and T4 DNA ligase from Thermo Scientific were used to generate amplicons, restriction digests and ligation products respectively. The ligation products were transformed into chemically competent DH5α *E. coli* and selected on LB agar medium with 100 μg/mL ampicillin or spectinomycin. Two *E. coli* colonies were normally selected for growth in liquid culture and 2-3 mL each was then used to isolate the plasmids using Nucleospin® Plasmid miniprep kit from Macherey-Nagel. Sanger sequencing of the ligated regions in new plasmids were performed by Biotechnology Resource Center at Cornell University. PureLink™ RNA Mini kit from Thermo Fisher Scientific was used to prepare total RNA from tobacco leaf tissue, and cDNA was subsequently prepared with SuperScript™ III First-Strand Synthesis System from Thermo Fisher Scientific.

A template vector to hold each DNA piece was constructed as follows. The *aadA* operon from BJF-070 vector[51] was removed by self-ligation of the NsiI digest. An amplicon was generated from the resulting vector using NsiI-BJF3 and BamHI-BJF5 primers and ligated into the BamHI and NsiI sites of the vector to introduce SbfI and NotI sites upstream of the NsiI locus to obtain BJFE-BB vector. Next, an amplicon generated with BB-SD-f and BJF5623r primers was digested with SbfI and AgeI and ligated into the SbfI and AgeI sites of BJFE-BB vector to introduce MluI and MauBI sites upstream of the NotI locus to obtain BJFE-BB-SD vector. Then, MauBI+NdeI digest of an amplicon generated with MauBI-T7P-f and NdeI-RBS-r primers and NdeI+NotI digest of the Nt_cpn60α gene amplified from tobacco cDNA with NdeI-Nt_c60a-f and NotI-Nt_c60a-r primers were ligated into the MauBI and NotI sites of BJFE-BB-SD vector to obtain BJFE-T7P-RBS-Nt_cpn60α vector. The NdeI and NotI digest of the BJFE-T7P-RBS-Nt_cpn60α vector was used as a template for subsequent ligations of different genes into BJFE-T7P-RBS template vector. In this work, such template vectors were created for groES, Nt_rbcX, Nt_rbcS-S1, Nt_rbcS-S1H, Nt_rbcS-S2, Nt_rbcS-S5, Nt_rbcS-T1, Nt_rbcS-T2, Nt_rbcS-T4a, Nt_rbcS-T5 and Nt_rbcS_Tc genes. For Nt_rbcL gene, MauBI-T7P-lacO-f and lacO-RBS-for primers were used to insert the lacO sequence downstream of the T7 promoter to obtain BJFE-T7P-lacO-RBS-Nt_rbcL vector. MauBI-extRBS-F, RBS-Nt_c20-f and NotI-Nt_c20-r primers were used to amplify Nt_cpn20 gene and ligated into the MauBI and NotI sites of BJFE-BB-SD vector to obtain BJFE-exRBS-Nt_cpn20 vector.

Nt_rbcL gene was amplified with NcoI-NtrbcLF and NtrbcLR-S-H primers and ligated into NcoI and HindIII sites of pBAD-Dest49 vector (Thermo Fisher Scientific) to obtain pBAD-Nt_rbcL_S2A vector, which encodes Ala instead of Ser as the second residue. To obtain the wild-type Nt_rbcL gene, the two amplicons generated from pBAD_138F+NtrbcLR_A2S and NtrbcLF_A2S+NtrbcL_518R primers were joined with overlap extension pcr and ligated into the two BamHI sites of pBAD-Nt_rbcL_S2A vector to obtain pBAD-Nt_rbcL vector. The Nt_rbcL gene with P_{BAD} promoter was amplified from pBAD_Nt_rbcL vector with NaeI-araC-F and NaeI-rrnBT2-R primers, digested with NaeI and ligated into the two NaeI sites of pATC60αβ/C20 vector[24] to obtain pET-NtL/AtC60αβ/C20 vector (Fig. 1a).

To replace At_cpn60α and At_cpn60β with tobacco genes, Nt_cpn60β and Nt_cpn60α were first amplified from tobacco cDNA with NdeI-NtC60bF+EcoRV-NtC60bR and NcoI-NtC60aF+AscI-NtC60aR primers respectively and ligated into NdeI+EcoRV and NcoI+AscI sites of pRSB-Duet1 vector (Novagen) to obtain pRSF-NtC60αβ vector. Next, the Nt_rbcL fragment between the two MluI sites in pET-NtL/AtC60αβ/C20 vector was removed by self-ligation of the MluI-digested vector to obtain pET-AtC60αβ/C20-MluI vector. Nt_cpn60αβ was amplified from pRSF-NtC60αβ vector with NheI-NtC60aF and BsiWI-NtC60bR primers, digested with NheI and BsiWI and ligated into the XbaI and Acc65I sites of pET-AtC60αβ/C20-MluI vector to obtain pET-NtC60αβ-2MluI vector. Then, the 3' end of Nt_cpn60β was amplified with NtC60b_dMluI_F and pET148_R primers, digested with XbaI and AatII and ligated into the XbaI and AatII sites of pET-NtC60αβ-2MluI vector to obtain pET-NtC60αβ-1MluI vector. Then, the Nt_rbcL fragment between the two MluI sites previously removed was ligated back into the MluI site

to obtain pET-NtL-NtC60αβ vector. Finally, RBS-Nt_cpn20 and T7P-RBS-groES modules were digested out of their respective BJFE vectors with MauBI and NotI and ligated into the AscI and NotI sites of pET-NtL/NtC60αβ vector to obtain pET-NtL/NtC60α/C20/C60β and pET-NtL/NtC60α/ES/C60β vectors respectively (Fig. 1a).

To drive Nt_rbcL with a T7 promoter, T7P-lacO-RBS-Nt_rbcL was amplified from BJFE-T7P-lacO-RBS-Nt_rbcL with AgeI-T7P-for and AgeI-NotI-rev primers, digested with AgeI and ligated into the AgeI site of pATC60αβ/C20 vector to obtain pET-AtC60αβ/C20/NtL vector (Fig. 1a). Next, the Nt_rbcL gene with P_{BAD} promoter in pET-NtL/NtC60α/C20/C60β and pET-NtL/NtC60α/ES/C60β vectors was removed by self-ligation of NaeI digests to obtain pET-NtC60α/C20/C60β and pET-NtC60α/ES/C60β vectors respectively. Finally, T7P-lacO-RBS-Nt-rbcL was digested out of BJFE-T7P-lacO-RBS-Nt_rbcL vector with MauBI and NotI and ligated into the AscI and NotI sites of pET-NtC60α/C20/C60β and pET-NtC60α/ES/C60β vectors to obtain pET-NtC60α/C20/NtL/C60β and pET-NtC60α/ES/NtL/C60β vectors respectively (Fig. 1a).

Nt_rbcS1 with P_{BAD} promoter was assembled with overlap extension PCR of ParaBAD amplified with ParaF2 and ParaR primers and Nt_rbcS1 amplified from Para-S1aF and B1002-S1aR primers, digested with Kpn2I and ligated into the AgeI site of pNtR1/AtR2/B2 vector[24] to obtain pCDF-NtR1/AtR2/B2/NtS1 vector. Nt_rbcX was amplified from BJFE-T7P-RBS-Nt_rbcX with NcoI-NtRbcXF and NotI-R primers, digested with NcoI and NotI and ligated into NcoI and NotI sites of pCDF-NtR1/AtR2/B2/NtS1 vector to obtain pCDF-NtX/R1/AtR2/B2/NtS1 (Fig. 1b). The NcoI-XhoI digest with NtX/R1/AtR2/B2 from pCDF-NtX/R1/AtR2/B2/NtS1 was ligated into NcoI and XhoI sites of pNtR1/AtR2/B2 vector to obtain pCDF-NtX/R1/AtR2/B2 vector. MauBI and NotI digest from BJFE-T7P-RBS-Nt_rbcS1 was ligated into AscI and NotI sites of pCDF-NtX/R1/AtR2/B2 vector to obtain pCDF-NtX/NtS1/R1/AtR2/B2 vector (Fig. 1b).

To replace At_raf2 and At_bsd2 genes with Nt_raf2 and Nt_bsd2 genes, intergenic regions from the pCDF vector were joined with Nt_raf2 and Nt_bsd2 genes using PstI-pCDF-F, pCDF-IG2-R, pCDF-IG2-RN, pCDF-IG2-F, XhoI-pCDF-R, IG2-Ntraf2FN, IG2-Ntraf2R, IG2-Ntbsd2F and IG2-Ntbsd2R primers in overlap extension PCRs. Nt_bsd2 module was ligated into SpeI and XhoI sites of pCDF-NtX/R1/AtR2/B2 vector to obtain pCDF-NtX/R1/AtR2/NtB2 vector. Similarly, Nt_raf2 module was ligated into PstI and SpeI sites of pCDF-NtX/R1/AtR2/NtB2 vector to obtain pCDF-NtX/R1/R2/B2 vector. Finally, MauBI and NotI digest from BJFE-T7P-RBS-Nt_rbcS1 was ligated into AscI and NotI sites of pCDF-NtX/R1/AtR2/NtB2 and pCDF-NtX/R1/R2/B2 vectors to obtain pCDF-NtX/NtS1/R1/AtR2/NtB2 and pCDF-NtX/NtS1/R1/R2/B2 vectors respectively (Fig. 1b). Similarly, the other tobacco *rbcS* genes were transferred from their respective BJFE vectors to the final pCDF vectors.

Expression of tobacco Rubisco in *E. coli*. Chemically competent BL21 Star™ (DE3) or Rosetta (DE3) *E. coli* strains were simultaneously transformed with two expression vectors: a pET vector with Nt_rbcL, either Nt_cpn60α or At_cpn60α, either Nt_cpn60β or At_cpn60β and either Nt_cpn20, At_cpn20 or groES genes and a pCDF vector with Nt_rbcS, Nt_raf1, Nt_rbcX, either Nt_raf2 or At_raf2 and either Nt_bsd2 or At_bsd2 genes. Transformed cells were selected on LB agar plates with 100 μg/mL ampicillin and 100 μg/mL spectinomycin. For Rosetta (DE3), 25 μg/mL chloramphenicol was also added to the medium. *E. coli* cultures were grown at 37°C with the exception of those having the Nt_cpn20 gene, which required growth at room temperature (~ 25°C) for unknown reasons.

For genes driven by T7 promoters, two induction approaches were tested: IPTG-induction or auto-induction. For IPTG-induction, overnight *E. coli* cultures in LB medium were diluted 500x into 6 mL LB medium with antibiotics, which were grown at 37°C 250 rpm for 3-4 hr before they were induced with 0.3 mM IPTG and grown at room temperature 250 rpm for additional 18-20 hr. For cultures having the Rubisco genes under P_{BAD} promoters, three hours after the addition of IPTG, the cultures were pelleted at 2500 rcf for 5 min and resuspended in equal volumes of LB medium with 0.4% w/v L-arabinose. They were then grown at room temperature at 250 rpm for additional 17-18 hr. For auto-induction, overnight *E. coli* cultures grown in ZYPG medium were diluted 50x into 6 mL ZYP-5052 medium with antibiotics and grown at 37°C 250 rpm for 5-6 hr. The cultures were then transferred to

room temperature and grown at 250 rpm for additional 17-19 hours. L-arabinose was added to cultures having the Rubisco genes under P_{BAD} promoters to a final concentration of 0.4% w/v three hours after they were transferred to room temperature.

ZYPG non-inducing medium is made up of 1% tryptone, 0.5% yeast extract technical, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 0.5% glucose, 2 mM $MgSO_4$ and trace metals (50 µM $FeCl_3$, 20 µM $CaCl_2$, 10 µM $MnCl_2$, 10 µM $ZnSO_4$, 2 µM $CoCl_2$, 2 µM $CuCl_2$, 2 µM $NiCl_2$, 2 µM Na_2MoO_4 and 2 µM H_3BO_3 in 60 µM HCl). ZYP-5052 auto-inducing medium is composed of 1% tryptone, 0.5% yeast extract technical, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 2 mM $MgSO_4$ and the same trace metals. Both ZYPG and ZYP-5052 media were prepared according to a previous report[32].

The *E. coli* cultures typically reached final OD600 values ~2-3 under IPTG-induction and ~4-5 under auto-induction. They were pelleted at 2500 rcf for 5 min and stored at -80°C before use. Each pellet was resuspended in 400-500 µL of lysis buffer made up of 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 1 mM EDTA, 30 mM $NaHCO_3$, 2 mM DTT and Pierce™ protease inhibitor mini-tablet from Thermo Scientific. The cells were lysed under ice-cold condition for 1 sec with Heat Systems – Ultrasonics Inc. W380 sonicator under 40% duty cycle and output control set to 4. Insoluble materials were removed by centrifugation at 16,000 rcf and 4°C for 5 min. The soluble protein extracts in buffer containing 30 mM $NaHCO_3$ and 10 mM $MgCl_2$ were then incubated on ice for 45-60 minutes to fully activate the Rubisco active sites before they were used in ^{14}C-CABP binding experiments or RuBP carboxylation experiments as described below.

Native PAGE of *E. coli* extracts. The *E. coli* samples were prepared as described above. Tobacco leaf samples were prepared based on a previous protocol[52] as follows. About 5 cm^2 each of leaf tissues was ground in a Wheaton homogenizer in 100 mM Bicine/NaOH pH 7.9, 5 mM $MgCl_2$, 1 mM EDTA, 5 mM ε–aminocaproic acid, 2 mM benzamidine, 50 mM 2-mercaptoethanol, protease inhibitor cocktail, 1 mM phenylmethanesulfonyl fluoride, 5% w/v poly(ethylene glycol) 4000, 10 mM $NaHCO_3$ and 10 mM DTT. After the cell debris were removed by centrifugation at 16,000 rcf and 4°C for 5 minutes, the supernatant was applied to a 2 mL Zeba™ spin desalting column equilibrated with 100 mM Bicine-NaOH pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM benzamidine, 1 mM ε-aminocaproic acid, 1 mM KH_2PO_4, 2% w/v poly(ethylene glycol) 4000, 10 mM $NaHCO_3$ and 10 mM DTT and eluted at 2,000 x g for 2 min at 4°C. The protein concentrations were estimated by a standard Bradford assay. Approximately 10-20 µg of total soluble proteins from each sample was mixed with 1/3 volume of sample buffer containing 200 mM Bis-Tris pH 7.2, 200 mM NaCl, 0.004% PonceauS and 40% glycerol and loaded to an Invitrogen™ 15-well 3-12% Bis-Tris protein gel from Thermo Fisher Scientific. The anode buffer was 50 mM Bis-Tris and 50 mM Tricine pH 6.8, while the cathode buffer contained an additional 0.002% Coomassie Brilliant Blue G250. The gels were run at 150 V for 30 min and then at 250 V for 60-70 min at 4°C before being transferred to PVDF membranes with 0.45 µm removal rating at 100 V for 1 hour. The membranes were then blocked with 5% milk in TBST buffer at room temperature for 1 hour and incubated with the antibody against Rubisco (kindly provided by P. John Andralojc from Rothamsted Research) in 5% milk in TBST buffer at 4 °C overnight. The primary antibody was detected with an HRP-conjugated secondary antibody in 2.5% milk in TBST buffer at room temperature and chemiluminescence produced with ECL substrate was recorded with a ChemiDoc™ MP Imaging System from Bio-Rad.

Quantification of Rubisco active sites. We synthesized ^{14}C-CABP from RuBP (Sigma-Aldrich part number 83895) and ^{14}C-KCN (specific activity of 5 mCi/mmol from American Radiolabeled Chemicals Inc part # 0136-1 mCi) based on a previously described protocol[53] as follows. First, 200 µL of 0.1 M ^{14}C KCN in 0.1 M Tris-acetate buffer pH 8.5 was mixed with approximately 16 µmol of RuBP (600 µL of 26.7 mM) in a glass vial and sealed for 48-72 hours at 23 °C. The mixture was then applied to 0.9 ml of Dowex 50WX8 200-400 mesh resin (Sigma part # 217514) that had been equilibrated with H_2O after being treated with 5 mL of 0.1 N HCl for 1 min, and the filtrate was collected. The resin was washed with 2 mL H_2O, and the filtrates were combined and dried under N_2 gas for 5 hours. The dried ^{14}C CPBP was

resuspended in 1.225 mL of 50 mM Bicine NaOH pH 9.3 at a final concentration of ~12 mM and stored at -20C in 0.125 mL aliquots. Before use, each aliquot was diluted with 0.5 mL of 0.11 M Bicine 22 mM $MgCl_2$ pH 8 and left overnight at 23 °C for complete delactonisation.

For size exclusion chromatography, 100 µL of each sample incubated with 7.2 nmol ^{14}C-CABP for at least 20 minutes at 23 °C was applied to 10 mL of Sephadex G50 Fine (Santa Cruz Biotechnology) equilibrated with 20 mM Bicine NaOH, 75 mM NaCl pH 8.0 in a 0.7 x 30 cm glass chromatography column (Bio-Rad part number 7374731). The resin was carefully washed with 200 µL buffer and three times with 750 µL of the same buffer. After that, the eluent was collected in 8.3 mL glass vials for the next eight 750 µL applications of the buffer as outlined in a previous protocol[54], and each vial was mixed thoroughly with 2 mL of Ultima Gold liquid scintillation cocktail from PerkinElmer. The ^{14}C radioactivity was then measured in a Beckman LS 6000IC scintillation counter, and the number of Rubisco active sites in nmol in each sample was calculated with the following formula:

$$\frac{\text{radioactivity from Rubisco fractions in cpm} \times \frac{10^6\ \text{nmol}}{\text{mmol}}}{\text{specific radioactivity of CABP in} \frac{\text{mCi}}{\text{mmol}} \times 2.22 \times 10^9 \frac{\text{dpm}}{\text{mCi}} \times \text{efficiency in} \frac{\text{cpm}}{\text{dpm}}}$$

We confirmed that in all our experiments to quantify Rubisco active sites, the amount of ^{14}C-CABP incubated with each sample was at least 20-fold molar excess of the Rubisco active sites in the sample.

Determination of RuBP carboxylation kinetics. RuBP carboxylation rates were measured from the incorporation of ^{14}C into RuBP as follows. About 25 mL of an assay buffer consisting of 110 mM Bicine-NaOH and 22 mM $MgCl_2$ at pH 8.0 was equilibrated with CO_2-free N_2 gas for at least 1 hour and mixed with about 1 mg of carbonic anhydrase (Sigma-Aldrich). Each glass vial with 8.3 mL internal volume containing 918 µL of the assay buffer was then sealed with open-top caps with bonded PTFE faced silicone liners (Wheaton part number W240842), and further equilibrated with CO_2-free N_2 gas for at least 30 minutes at 23 °C followed by the addition of 50 µL of different ^{14}C bicarbonate stock solutions with known specific activities and 11.5 µL of 35 mM RuBP (a generous gift from Dr. Douglas Orr at Lancaster University). The final concentrations of dissolved CO_2 ranged from 12.9-108.3 µM for the *E. coli* samples and 8.3-103.7 µM for the tobacco extracts. After equilibration at 25 °C for at least 15 minutes, the reaction was then initiated by the addition of 20 µL of each sample containing Rubisco to the vials and stopped exactly one minute later by the addition of 200 µL of 20% (v/v) formic acid. The caps were then removed from the vials, which were then left on a heating block set at 100 °C. Once almost all solutions in the vials were evaporated, the leftover residue in each vial was dissolved in 0.5 mL of ddH_2O and thoroughly mixed with 3.5 mL of Ultima Gold liquid scintillation cocktail, and the acid-stable ^{14}C compounds in the vials were counted with a Beckman LS 6000IC scintillation counter. Each sample incubated with 200 µM CABP for 30 min was also run to obtain the background signals. The RuBP carboxylation rates in nmol/s were calculated with the following formula:

$$\frac{\text{radioactivity of leftover residue in} \frac{\text{cpm}}{\text{min}} \times \frac{10^6\ \text{nmol}}{\text{mmol}}}{\text{specific radioactivity of bicarbonate in} \frac{\text{mCi}}{\text{mmol}} \times 2.22 \times 10^9 \frac{\text{dpm}}{\text{mCi}} \times \text{efficiency in} \frac{\text{cpm}}{\text{dpm}} \times \frac{60\ \text{s}}{\text{min}}}$$

We used the Henderson-Hasselbalch equation with a pKa value of 6.11 to calculate $[CO_2]$ and adjusted it for partition between the aqueous buffer and the air space in each sealed reaction vial. The catalytic rates at six different $[CO_2]$ concentrations were then fitted to the standard Michaelis-Menten equation with nonlinear least-squares regression in RStudio (Version 1.2.1335) to obtain the K_C and V_{max}. The value of k_{cat} was then obtained by dividing V_{max} in nmol/s with the Rubisco active sites in each sample in nmol.

Relative expression of *rbcS* genes in Tobacco. Publicly available SRA files used for tobacco genome sequencing (NCBI Bioproject -PRJNA208209)[34] were utilised to quantify relative abundance of genes. Three transcriptomic SRA files each for young leaf and mature leaf were used to quantitate relative abundance of *rbcS* genes using Kallisto[55] with standard parameters.

Statistical analyses

RStudio (Version 1.2.1335) was used to perform ANOVA followed by Tukey's HSD test using agricolae package with standard parameters. Two-tailed t-tests were performed with Microsoft Excel 2016.

Data availability

The accession numbers of proteins expressed in this study are listed in Extended Data Fig. 4 and publicly available at https://www.ncbi.nlm.nih.gov or https://solgenomics.net. The tobacco transcriptomic data are available at https://www.ncbi.nlm.nih.gov under Bioproject accession PRJNA208209.

References

1 Whitney, S. M., Houtz, R. L. & Alonso, H. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol.* **155**, 27-35, doi:10.1104/pp.110.164814 (2011).

2 Bauwe, H., Hagemann, M., Kern, R. & Timm, S. Photorespiration has a dual origin and manifold links to central metabolism. *Curr. Opin. Plant Biol.* **15**, 269-275, doi:10.1016/j.pbi.2012.01.008 (2012).

3 Walker, B. J., VanLoocke, A., Bernacchi, C. J. & Ort, D. R. The costs of photorespiration to food production now and in the future. *Annu. Rev. Plant. Biol.* **67**, 107-129, doi:10.1146/annurev-arplant-043015-111709 (2016).

4 Davidi, D. *et al.* Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.* e104081. doi:10.15252/embj.2019104081 (2020).

5 Jordan, D. B. & Ogren, W. L. A sensitive assay procedure for simultaneous determination of ribulose-1,5-bisphosphate carboxylase and oxygenase activities. *Plant Physiol.* **67**, 237-245 (1981).

6 Flamholz, A. I. *et al.* Revisiting trade-offs between Rubisco kinetic parameters. *Biochemistry* **58**, 3365-3376, doi:10.1021/acs.biochem.9b00237 (2019).

7 Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7246-7251, doi:10.1073/pnas.0600605103 (2006).

8 Iniguez, C. *et al.* Evolutionary trends in RuBisCO kinetics and their co-evolution with CO_2 concentrating mechanisms. *Plant J.* **101**, 897-918, doi:10.1111/tpj.14643 (2020).

9 Evans, J. R. Photosynthesis and nitrogen relationships in leaves of C_3 plants. *Oecologia* **78**, 9-19, doi:10.1007/BF00377192 (1989).

10 Young, J. N. *et al*. Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *J. Exp. Bot.* **67**, 3445-3456, doi:10.1093/jxb/erw163 (2016).

11 Wilson, R. H. & Hayer-Hartl, M. Complex chaperone dependence of Rubisco biogenesis. *Biochemistry* **57**, 3210-3216, doi:10.1021/acs.biochem.8b00132 (2018).

12 Gatenby, A. A., van der Vies, S. M. & Bradley, D. Assembly in *E. coli* of a functional multi-subunit ribulose bisphosphate carboxylase from a blue-green alga. *Nature* **314**, 617-620, doi:10.1038/314617a0 (1985).

13 Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. & Hanson, M. R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**, 547-550, doi:10.1038/nature13776 (2014).
14 Long, B. M. *et al.* Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 3570, doi:10.1038/s41467-018-06044-0 (2018).
15 Tabita, F. R. & Small, C. L. Expression and assembly of active cyanobacterial ribulose-1,5-bisphosphate carboxylase/oxygenase in *Escherichia coli* containing stoichiometric amounts of large and small subunits. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 6100-6103, doi:10.1073/pnas.82.18.6100 (1985).
16 Whitney, S. M., Birch, R., Kelso, C., Beck, J. L. & Kapralov, M. V. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3564-3569, doi:10.1073/pnas.1420536112 (2015).
17 Sharwood, R. E., Ghannoum, O., Kapralov, M. V., Gunn, L. H. & Whitney, S. M. Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C_3 photosynthesis. Nat. Plants **2**, 16186, doi: 10.1038/nplants.2016.186 (2016).
18 Lin, M. T. & Hanson, M. R. Red algal Rubisco fails to accumulate in transplastomic tobacco expressing *Griffithsia monilis RbcL* and *RbcS* genes. *Plant Direct* **2**, e00045, doi:10.1002/pld3.45 (2018).
19 Whitney, S. M., Baldet, P., Hudson, G. S. & Andrews, T. J. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* **26**, 535-547 (2001).
20 Brutnell, T. P., Sawers, R. J., Mant, A. & Langdale, J. A. BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the *rbcL* gene of maize. *Plant Cell* **11**, 849-864 (1999).
21 Feiz, L. *et al.* A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J.* **80**, 862-869, doi:10.1111/tpj.12686 (2014).
22 Feiz, L. *et al.* Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435-3446, doi:10.1105/tpc.112.102012 (2012).
23 Fristedt, R. *et al.* RAF2 is a RuBisCO assembly factor in *Arabidopsis thaliana*. *Plant J.* **94**, 146-156, doi:10.1111/tpj.13849 (2018).
24 Aigner, H. *et al.* Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**, 1272-1278, doi:10.1126/science.aap9221 (2017).
25 Hauser, T. *et al.* Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22**, 720-728, doi:10.1038/nsmb.3062 (2015).
26 Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197-202, doi:10.1038/nature08651 (2010).
27 Kolesinski, P. *et al.* Insights into eukaryotic Rubisco assembly - Crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Bba-Gen Subjects* **1830**, 2899-2906, doi:10.1016/j.bbagen.2012.12.025 (2013).
28 Schmidt, J. A., McGrath, J. M., Hanson, M. R., Long, S. P. & Ahner, B. A. Field-grown tobacco plants maintain robust growth while accumulating large quantities of a bacterial cellulase in chloroplasts. *Nat. Plants* **5**, 715-721, doi:10.1038/s41477-019-0467-z (2019).
29 South, P. F., Cavanagh, A. P., Liu, H. W. & Ort, D. R. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* **363**, doi:10.1126/science.aat9077 (2019).
30 Laterre, R., Pottier, M., Remacle, C. & Boutry, M. Photosynthetic trichomes contain a specific Rubisco with a modified pH-dependent activity. *Plant Physiol.* **173**, 2110-2120, doi:10.1104/pp.17.00062 (2017).

31 Wilson, R. H., Thieulin-Pardo, G., Hartl, F. U. & Hayer-Hartl, M. Improved recombinant expression and purification of functional plant Rubisco. *FEBS Lett.* **593**, 611-621, doi:10.1002/1873-3468.13352 (2019).
32 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207-234, doi:10.1016/j.pep.2005.01.016 (2005).
33 Gong, L., Olson, M. & Wendel, J. F. Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol. Biol. Evol.* **31**, 2624-2636, doi:10.1093/molbev/msu207 (2014).
34 Sierro, N. *et al.* The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* **5**, 3833, doi:10.1038/ncomms4833 (2014).
35 Morita, K., Hatanaka, T., Misoo, S. & Fukayama, H. Identification and expression analysis of non-photosynthetic Rubisco small subunit, *OsRbcS1*-like genes in plants. *Plant Gene* **8**, 26-31, doi:10.1016/j.plgene.2016.09.004 (2016).
36 Pottier, M., Gilis, D. & Boutry, M. The Hidden Face of Rubisco. *Trends Plant Sci*. **23**, 382-392, doi:10.1016/j.tplants.2018.02.006 (2018).
37 Conlan, B. *et al*. BSD2 is a Rubisco-specific assembly chaperone, forms intermediary hetero-oligomeric complexes, and is nonlimiting to growth in tobacco. *Plant Cell Environ.* **42**, 1287-1301 doi:10.1111/pce.13473 (2019).
38 Houtz, R. L., Magnani, R., Nayak, N. R. & Dirk, L. M. Co- and post-translational modifications in Rubisco: unanswered questions. *J. Exp. Bot.* **59**, 1635-1645, doi:10.1093/jxb/erm360 (2008).
39 Morita, K., Hatanaka, T., Misoo, S. & Fukayama, H. Unusual small subunit that is not expressed in photosynthetic cells alters the catalytic properties of rubisco in rice. *Plant Physiol.* **164**, 69-79, doi:10.1104/pp.113.228015 (2014).
40 Spreitzer, R. J. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* **414**, 141-149, doi:10.1016/S0003-9861(03)00171-1 (2003).
41 Genkov, T. & Spreitzer, R. J. Highly conserved small subunit residues influence rubisco large subunit catalysis. *J. Biol. Chem.* **284**, 30105-30112, doi:10.1074/jbc.M109.044081 (2009).
42 Spreitzer, R. J., Peddi, S. R. & Satagopan, S. Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17225-17230, doi:10.1073/pnas.0508042102 (2005).
43 Ishikawa, C., Hatanaka, T., Misoo, S., Miyake, C. & Fukayama, H. Functional incorporation of sorghum small subunit increases the catalytic turnover rate of Rubisco in transgenic rice. *Plant Physiol.* **156**, 1603-1611, doi:10.1104/pp.111.177030 (2011).
44 Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275-291, doi:10.1016/j.plaphy.2008.01.001 (2008).
45 Kapralov, M. V., Kubien, D. S., Andersson, I. & Filatov, D. A. Changes in Rubisco kinetics during the evolution of C_4 photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Mol. Biol. Evol.* **28**, 1491-1503, doi:10.1093/molbev/msq335 (2011).
46 Yamada, K., Davydov, II, Besnard, G. & Salamin, N. Duplication history and molecular evolution of the *rbcS* multigene family in angiosperms. *J. Exp. Bot.* **70**, 6127-6139, doi:10.1093/jxb/erz363 (2019).
47 Smith, S. A. & Tabita, F. R. Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. *J. Mol. Biol.* **331**, 557-569, doi:10.1016/s0022-2836(03)00786-1 (2003).
48 Mueller-Cajar, O., Morell, M. & Whitney, S. M. Directed evolution of rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochemistry* **46**, 14067-14074, doi:10.1021/bi700820a (2007).
49 Wilson, R. H., Alonso, H. & Whitney, S. M. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci. Rep.* **6**, 22284, doi:10.1038/srep22284 (2016).

50 Zhou, Y. & Whitney, S. Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction. *Int. J. Mol. Sci.* **20**, doi:10.3390/ijms20205019 (2019).
51 Hanson, M. R., Gray, B. N. & Ahner, B. A. Chloroplast transformation for engineering of photosynthesis. *J. Exp. Bot.* **64**, 731-742, doi:10.1093/jxb/ers325 (2013).
52 Orr, D. J. & Carmo-Silva, E. Extraction of RuBisCO to determine catalytic constants. *Methods Mol. Biol.* **1770**, 229-238, doi: 10.1007/978-1-4939-7786-4_13 (2018).
53 Whitney, S. M. & Sharwood, R. E. Plastid transformation for Rubisco engineering and protocols for assessing expression. *Methods Mol. Biol.* **1132**, 245-262, doi:10.1007/978-1-62703-995-6_15 (2014).
54 Kubien, D. S., Brown, C. M. & Kane, H. J. Quantifying the amount and activity of Rubisco in leaves. *Methods Mol. Biol.* **684**, 349-362, doi:10.1007/978-1-60761-925-3_27 (2011).
55 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525-527, doi:10.1038/nbt.3519 (2016).

Correspondence and requests for materials should be addressed to M.R.H.

Acknowledgments

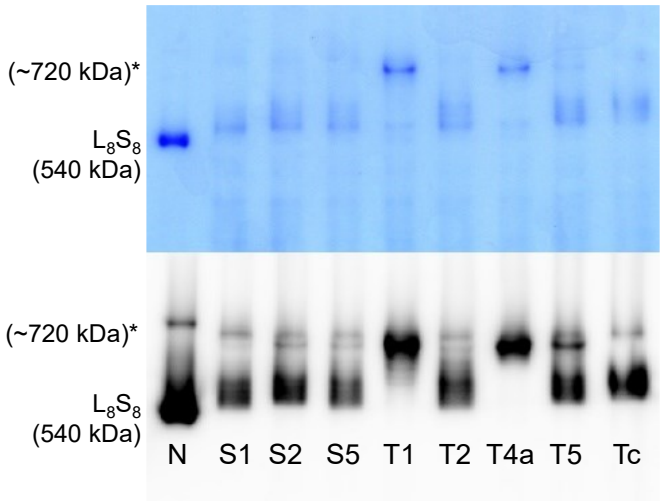
We thank Manajit Hayer-Hartl from Max Planck Institute of Biochemistry in Martinsried, Germany, for providing us with her lab's *E. coli* expression vectors. We also thank Douglas Orr, Elizabete Carmo-Silva and Martin Parry from Lancaster University in Lancaster, UK for providing us with purified RuBP and advice on experiments to quantify Rubisco active sites and measure RuBP carboxylation kinetics. M.T.L. and W.D.S. are supported by a grant to M.R.H. from the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (Award Number: DE-SC0020142) and M.R.H. and V.C. by the National Science Foundation (Award number MCB-1642386).

Author Contributions

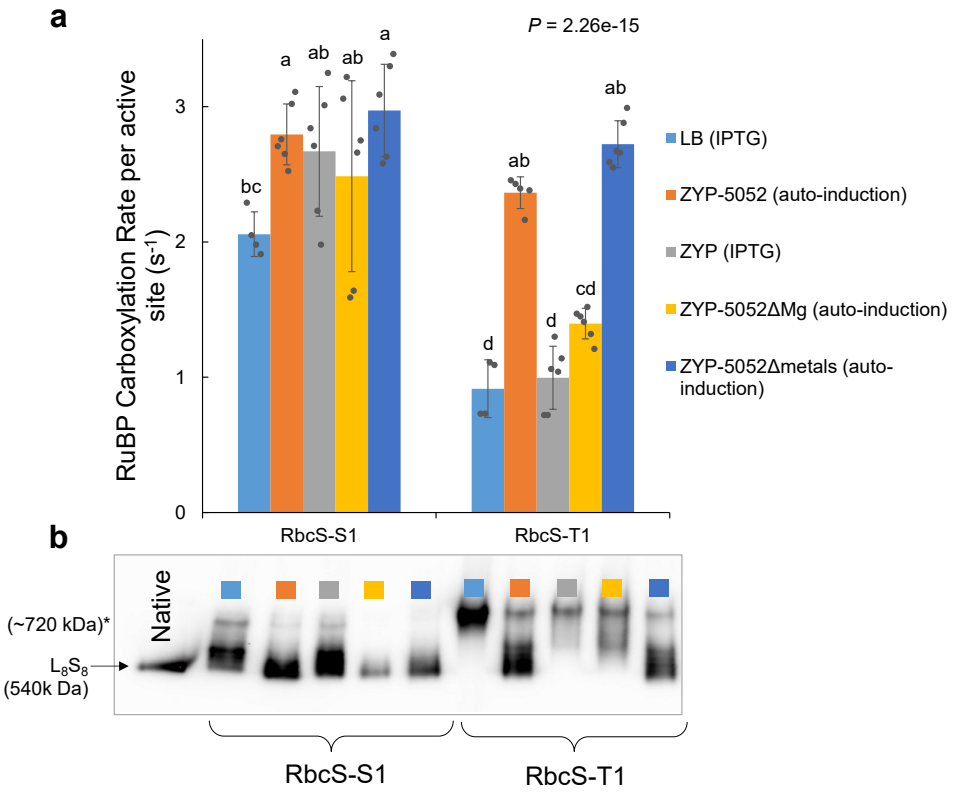
M.T.L. and M.R.H. conceived research. All authors designed experiments. M.T.L., W.D.S. and V.C. performed the experiments. All authors analysed the data and contributed to writing the manuscript.

Competing Interests Statement

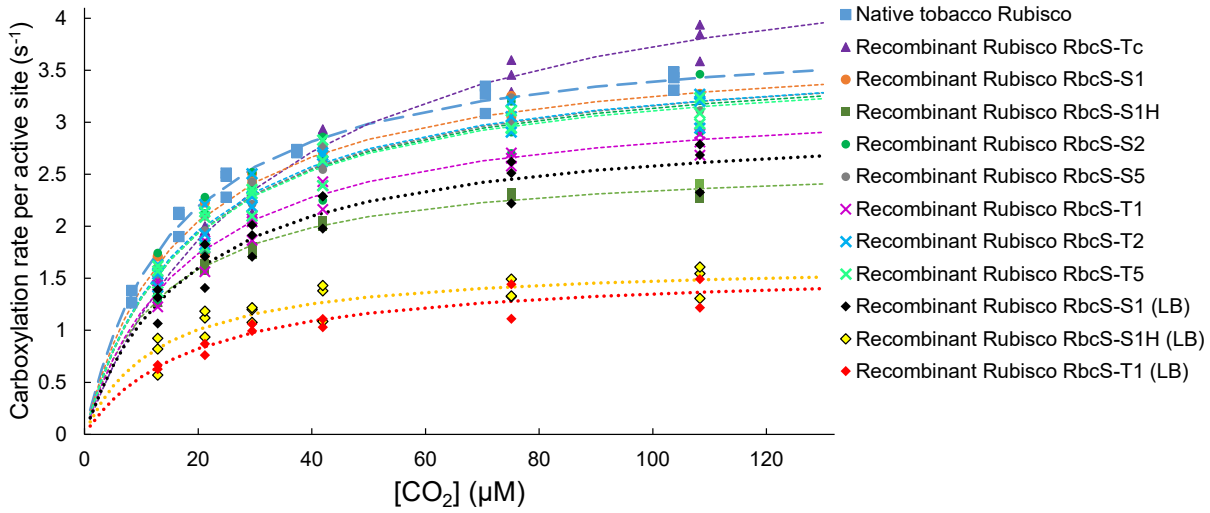
The authors declare no competing financial interests.



Extended Data Fig. 1 | Native PAGE immunoblots of tobacco Rubisco expressed in Rosetta (DE3) *E. coli* with different small subunits. Nt-Cpn60α, Nt-Cpn60β, GroES, Nt-RbcX, Nt-Raf1, At-Raf2 and Nt-Bsd2 were co-expressed. 9 µg of total soluble extract was loaded for each sample. 3 µg of total soluble extract from a young tobacco leaf (N) was also included as a control. The protein expressions were auto-induced in ZYP-5052 medium at 23 °C for 18-20 h. The top panel shows Coomassie blue staining, and the bottom panel shows the immunoblot using the antibody against Rubisco. The bands for L_8S_8 Rubisco (540 kDa) and chaperonin-RbcL complex (~720 kDa, indicated with asterisks) are marked next to each panel. The native PAGE was performed once for the same set of samples and multiple times for samples with RbcS-S1 and RbcS-T1 with similar results.



Extended Data Fig. 2 | Comparison of tobacco Rubisco expressed in BL21 Star™ (DE3) *E. coli* using different culture media. a, RuBP carboxylation rates at 42 μM [CO_2] were measured from incorporation of ^{14}C into RuBP in the absence of O_2 at 25 °C. Rubisco active sites were quantified with bound ^{14}C-CABP separated in size-exclusion chromatography. The error bars represent the mean values and standard deviations of measurements from three to six *E. coli* growth experiments for each condition. Data were analyzed with one-way ANOVA, and the P-value was obtained from F-statistics. Tukey's honest significance test was then carried, and samples with *P* value > 0.05 are indicated with the same letter. **b**, The native PAGE immunoblot of the same samples in **a** using the antibody against wheat Rubisco. The three additional media tested are the buffered medium without additional carbon source (ZYP), the auto-induction medium without Mg (ZYP-5052ΔMg) and the auto-induction medium without trace metals (ZYP-5052Δmetals). The native PAGE was performed once for the same set of samples.



Extended Data Fig. 3 | The enzyme kinetics of tobacco Rubisco expressed in *E. coli* with different small subunits compared to the native tobacco Rubisco. RuBP carboxylation rates were measured from incorporation of ^{14}C into RuBP in the absence of O_2 at 25 °C. At-Cpn60α, At-Cpn60β, At-Cpn20, Nt-RBCX, Nt-RAF1, At-RAF2 and Nt-BSD2 were co-expressed in BL21 Star™ (DE3) either in LB or ZYP-5052 auto-induction medium at 23 °C for 18-20 h. [CO_2] in the reaction mixtures ranged from 12.9 to 108.3 μM for the *E. coli* extracts and 8.3 to 103.7 μM for tobacco leaf extracts. Rubisco active sites were quantified with bound ^{14}C-CABP separated in size-exclusion chromatography. The data were fitted to the Michaelis-Menten equation with nonlinear regression. The fitted models are shown as dotted lines for the *E. coli* samples and dashed line for the tobacco leaf samples.

Extended Data Fig 4. Summary of the genes expressed in this study.

Genes	Source Organisms	Protein Products	Accession Numbers
Nt-rbcL	*Nicotiana tabacum*	Rubisco large subunit	NP_054507
Nt-rbcS-S1	*Nicotiana tabacum*	Rubisco small subunit	XP_016481946
Nt-rbcS-S2	*Nicotiana tabacum*	Rubisco small subunit	Nitab4.5_0000682g0090.1*
Nt-rbcS-S5	*Nicotiana tabacum*	Rubisco small subunit	Nitab4.5_0000659g0110.1*
Nt-rbcS-T1	*Nicotiana tabacum*	Rubisco small subunit	XP_016446652
Nt-rbcS-T2	*Nicotiana tabacum*	Rubisco small subunit	Nitab4.5_0000912g0250.1*
Nt-rbcS-T4a	*Nicotiana tabacum*	Rubisco small subunit	Nitab4.5_0000912g0270.1*
Nt-rbcS-T5	*Nicotiana tabacum*	Rubisco small subunit	Nitab4.5_0020038g0010.1*
Nt-rbcS-Tc	*Nicotiana tabacum*	Rubisco small subunit	XP_016450905
Nt-rbcX	*Nicotiana tabacum*	RbcX	XP_016503320
Nt-raf1	*Nicotiana tabacum*	Rubisco accumulation factor 1	XP_016475347
Nt-raf2	*Nicotiana tabacum*	Rubisco accumulation factor 2	XP_016516208
Nt-bsd2	*Nicotiana tabacum*	Bundle sheath defective 2	XP_016474402
Nt-cpn60α	*Nicotiana tabacum*	chaperonin 60alpha	XP_016509320
Nt-cpn60β2	*Nicotiana tabacum*	chaperonin 60beta	XP_016490465 (2 aa difference)
Nt-cpn20	*Nicotiana tabacum*	cochaperonin 20	XP_016489864
At-raf2	*Arabidopsis thaliana*	Rubisco accumulation factor 2	At5g51110
At-bsd2	*Arabidopsis thaliana*	Bundle sheath defective 2	At3g47650
At-cpn60α1	*Arabidopsis thaliana*	chaperonin 60alpha	At2g28000
At-cpn60β1	*Arabidopsis thaliana*	chaperonin 60beta	At1g55490
At-cpn20	*Arabidopsis thaliana*	cochaperonin 20	At5g20720
groES	*Escherichia coli*	cochaperonin GroES	NP_418566

*These accession numbers are from descriptions of blast results at https://solgenomics.net/tools/blast/ using Nitab v4.5 cDNA Edwards2017 database.

Extended Data Fig 5. Summary of plasmids used in the expression of tobacco Rubisco in *E. coli*.

Plasmids	Genes Present*
pET-*Nt*L/*At*C60αβ/C20	*Nt-rbcL* (P_{BAD}), *At-cpn60α1*, *At-cpn60β1*, *At-cpn20*
pET-*Nt*L/*Nt*C60α/C20/C60β	*Nt-rbcL* (P_{BAD}), *Nt-cpn60α*, *Nt-cpn20*, *Nt-cpn60β2*
pET-*Nt*L/*Nt*C60α/ES/C60β	*Nt-rbcL* (P_{BAD}), *Nt-cpn60α*, *groES*, *Nt-cpn60β2*
pET-*At*C60αβ/C20/*Nt*L	*At-cpn60α1*, *At-cpn60β1*, *At-cpn20*, *Nt-rbcL*
pET-*Nt*C60α/C20/*Nt*L/C60β	*Nt-cpn60α*, *Nt-cpn20*, *Nt-rbcL*, *Nt-cpn60β2*
pET-*Nt*C60α/ES/*Nt*L/C60β	*Nt-cpn60α*, *groES*, *Nt-rbcL*, *Nt-cpn60β2*
pCDF-*Nt*X/R1/*At*R2/B2/*Nt*S1	*Nt-rbcX*, *Nt-raf1*, *At-raf2*, *At-bsd2*, *Nt-rbcS-S1* (P_{BAD})
pCDF-*Nt*X/*Nt*S1/R1/*At*R2/B2	*Nt-rbcX*, *Nt-rbcS-S1*, *Nt-raf1*, *At-raf2*, *At-bsd2*
pCDF-*Nt*X/*Nt*S1/R1/*At*R2/*Nt*B2	*Nt-rbcX*, *Nt-rbcS-S1*, *Nt-raf1*, *At-raf2*, *Nt-bsd2*
pCDF-*Nt*X/*Nt*S1/R1/R2/B2	*Nt-rbcX*, *Nt-rbcS-S1*, *Nt-raf1*, *Nt-raf2*, *Nt-bsd2*
pCDF-*Nt*X/*Nt*rbcS/R1/*At*R2/B2	*Nt-rbcX*, *Nt-rbcS*, *Nt-raf1*, *At-raf2*, *Nt-bsd2*
pCDF-*Nt*X/*Nt*rbcS/R1/R2/B2	*Nt-rbcX*, *Nt-rbcS*, *Nt-raf1*, *Nt-raf2*, *Nt-bsd2*

**Nt-rbcL* in the first three pET vectors and *Nt-rbcS-S1* in pCDF-*Nt*X/R1/*At*R2/B2/*Nt*S1 vectors are under the arabinose-inducible promoter (P_{BAD}). All other genes are under the T7 promoter (P_{T7}).