Enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli* varies depending on the small subunit composition

Myat T. Lin¹, William D. Stone¹², Vishal Chaudhari¹, Maureen R. Hanson¹

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

Abstract

Rubisco catalyzes the first step in carbon fixation and has been a strategic target to improve photosynthetic efficiency. In plants, Rubisco is a complex made up of eight large subunits encoded by a chloroplast gene, *rbcL*, and eight small subunits expressed from a nuclear gene family and targeted to chloroplast stroma. Biogenesis of Rubisco in plants requires a chaperonin system composed of Cpn60α, Cpn60β and Cpn20, which helps fold the large subunit, and multiple chaperones including RbcX, Raf1, Raf2 and BSD2, which help the dimerization of the folded large subunits and subsequent assembly with the small subunits into L₈S₈ holoenzymes. A recent study successfully assembled functional Arabidopsis Rubisco in *Escherichia coli* by co-expressing the two subunits with Arabidopsis chaperonins and chaperones (Aigner et al., 2017). In this study, we modified the expression vectors used in that study and adapted them to express tobacco Rubisco by replacing the Arabidopsis genes with tobacco ones. Next, we surveyed the small subunits present in tobacco, co-expressed each with the large subunit and successfully produced active tobacco enzymes composed of different small subunits in *E. coli*. These enzymes produced in *E. coli* have carboxylation kinetics very similar to that of the native tobacco Rubisco. We also produced tobacco Rubisco with a recently discovered trichome small subunit in *E. coli* and found that it has a higher catalytic rate and a lower CO₂ affinity compared to the enzymes with other small subunits. Our improvements in the *E. coli* Rubisco expression system will allow us to probe features of both the chloroplast and nuclear-encoded subunits of Rubisco that affect its catalytic rate and CO₂ specificity.
Introduction

In photosynthetic organisms including plants, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco, EC 4.1.1.39) catalyzes the fixation of CO$_2$ from air to ribulose-1,5-bisphosphate (RuBP). Thus, Rubisco represents a major gateway for inorganic carbon to enter biosphere and has far-reaching impacts on a global scale. The carboxylation of RuBP by Rubisco is a well-known bottleneck in C$_3$ photosynthesis because of its slow catalytic rates with approximately 3 s$^{-1}$ turnover number ($k_{cat}$) (Whitney et al., 2011a). Rubisco also catalyzes a competing RuBP oxygenation, generating a toxic byproduct, 2-phosphoglycolate, which has to be recycled through the energy-intensive photorespiration pathway (Bauwe et al., 2012). 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% (Walker et al., 2016).

An important kinetic parameter of Rubisco is its CO$_2$/O$_2$ specificity factor ($S_{CO}$), 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 $k_{cat}$ and $K$ are the catalytic rate and Michaelis-Menten constant respectively (Jordan and Ogren, 1981). Due to mechanistic constraint, a Rubisco that binds its carboxylated intermediates tighter has a higher $S_{CO}$, a lower $K_C$ and a lower $k_{cat}^c$, leading to a strong direct relationship between $k_{cat}^c$ and $K_C$ as well as an inverse relationship between $k_{cat}^c$ and $S_{CO}$ (Tcherkez et al., 2006; Savir et al., 2010). Due to the lack of a CO$_2$-concentrating mechanism, C$_3$ Rubisco enzymes typically have higher $S_{CO}$ and lower $k_{cat}^c$ than C$_4$ enzymes. 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 (Evans, 1989).

The plant enzyme belongs to form I Rubisco and is composed of eight large subunits encoded by a single chloroplast gene and eight small subunits expressed from a family of nuclear genes. In contrast, dinoflagellates, many proteobacteria and archaea produce forms II and III Rubisco without small subunits (Tabita et al., 2008). The active site is formed at the dimeric interface between two large subunits and well conserved among all Rubisco forms. Cyanobacteria and many autotrophic prokaryotes also possess form I Rubisco, which has the same subunit composition as in the plant L$_8$S$_8$ holoenzyme. The large subunits 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 small subunits to form the L$_8$S$_8$ holoenzyme (Wilson and Hayer-Hartl, 2018). Form I Rubisco from many cyanobacteria can be assembled in E. coli (Gatenby et al., 1985; Tabita and Small, 1985) and tobacco chloroplasts (Lin et al., 2014; Long et al., 2018) without additional chaperones. On the other hand, plant Rubisco has strict assembly requirements that can only be met by closely related species (Whitney et al., 2015; Sharwood, 2017). Some red algae have been reported to have Rubisco with the highest $S_{CO}$ values, which could potentially improve plant photosynthesis, but their Rubisco failed to assemble into active enzyme in tobacco due to lack of compatible assembly factors (Whitney et al., 2001; Lin and Hanson, 2018).

Previous studies on maize 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) (Brutnell et al., 1999; Feiz et al., 2012; Feiz et al., 2014). Further structural studies have revealed the mechanism of Raf1 and BSD2 as well as RbcX in assisting the dimerization of large subunits and subsequent formation of the L$_8$ core before the formation of the final holoenzyme (Liu et al., 2010; Hauser et al., 2015; Aigner et al., 2017).
In a recent breakthrough for the assembly of plant Rubisco, Aigner and co-workers were able to produce functional Arabidopsis Rubisco in *E. coli* by co-expressing at least five additional proteins: two plant-specific chaperonin (Cpn60α and Cpn60β) and three chaperones (Raf1, Raf2 and BSD2) (Aigner et al., 2017). 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. Although Arabidopsis has been a great model plant in genetic studies, tobacco offers several advantages in Rubisco engineering since both nuclear and chloroplast genome transformations can be readily carried out in tobacco followed by subsequent field trials of the transgenic plants with modified Rubisco to access their photosynthetic performance when standard agricultural practices are employed. Thus, an *E. coli* expression system with improved assembly of tobacco Rubisco will be useful to identify enzymes with improved 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 (Aigner et al., 2017) to co-express tobacco chaperonins and chaperones. We also surveyed the expression of different small subunits in tobacco and co-expressed each of them in *E. coli* with the single tobacco chloroplast-encoded large subunit. We then quantified the tobacco Rubisco produced in *E. coli* with each small subunit and measured their carboxylation kinetics. We also found that the Rubisco produced in *E. coli* with a recently identified small subunit present in the secretory glands of tobacco trichomes (Laterre et al., 2017) has a considerably different kinetic profile.

**Results**

**Tobacco Rubisco was successfully expressed and assembled in *E. coli* using a modified two-vector system.**

The pET and pCDF expression vectors we used in this study are based on the system recently reported by (Aigner et al., 2017). We obtained pATC60αβ/C20 and pNtR1/AtR2/B2 that were used in that study to co-express three Arabidopsis chaperonins (Cpn60α1, Cpn60β1 and Cpn20), two Arabidopsis chaperones (Raf2 and BSD2) and one tobacco chaperone (Raf1). Instead of expressing the two tobacco Rubisco subunits from a third vector, we introduced the tobacco *rbcL* and *rbcS-S1* genes into pATC60αβ/C20 and pNtR1/AtR2/B2 vectors along with arabinose-inducible P*BAD* promoters to obtain pET-NtL-AtC60αβ20 and pCDF-NtR1/AtR2B-NtS1 respectively (Figure 1). We also introduced the tobacco *rbcX* gene into the latter since deleting At-rbcX was shown to reduce the yield of Arabidopsis Rubisco by up to 60% (Aigner et al., 2017).

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). We then followed the previously described growth protocol (Aigner et al., 2017) to initially induce the chaperonins and chaperones with IPTG for 3 hours at room temperature followed by the transfer of the *E. coli* cultures to a medium containing arabinose to induce the expression of tobacco Rubisco subunits for additional 16-18 hours at room temperature. We then analyzed the soluble extracts from three different *E. coli* strains on a native PAGE immunoblot with an antibody against Rubisco. We observed an L8S8 Rubisco complex band from the BL21(DE3) extract when the tobacco subunits were co-expressed with Arabidopsis chaperonins, Raf2, BSD2 and tobacco Raf1 (Figure 2). We also found that the Rosetta (DE3) strain was able to produce more L8S8 Rubisco than the two BL21(DE3) strains tested (Figure 3). Since none of the nine genes that were expressed in *E. coli*
are codon-optimized for the bacterium, the presence of tRNAs for rare codons in the Rosetta strain likely resulted in improved expression.

Figure 1. pET and pCDF E. coli expression vectors created in this study. (A) pET vectors express Nt-rbcL from P_{BAD} and cpn60α, cpn60β and either cpn20 or groES from P_{T7}. (B) pCDF vectors express Nt-rbcS from either P_{BAD} or P_{T7} and rbcX, raf1, raf2 and bsd2 from P_{T7}.

Figure 2. The native PAGE analysis of E. coli soluble extracts expressing genes from pET-NtL-AtC60αβ20 and pCDF-NtXR1Ar2B-NtS1 with the top panel showing Coomassie blue staining and the bottom panel showing the immunoblot with the antibody against Rubisco. The genes expressed are Nt-rbcL (P_{BAD}), Nt-rbcS-S1 (P_{BAD}), Nt-rbcX, Nt-raf1, At-raf2, At-bsd2, At-cpn60α1, At-cpn60β1 and At-cpn20.
The expression of tobacco Rubisco in *E. coli* was markedly improved when tobacco chaperonins were co-expressed.

Next, we replaced the Arabidopsis chaperonins (Cpn60α1, Cpn60β1 and Cpn20) with tobacco counterparts. Alternatively, we also overexpressed *E. coli* co-chaperonin GroES in place of tobacco Cpn20 since Aigner et al. (2017) showed that GroES was able to function with Arabidopsis Cpn60 to assemble Rubisco. We also tested the overexpression of tobacco RbcS with a T7 promoter (P<sub>T7</sub>) because the availability of RbcS was often limited in the assembly Arabidopsis Rubisco holoenzyme in *E. coli* (Aigner et al., 2017). All these changes resulted in dramatic improvements in tobacco Rubisco assembly in all three *E. coli* expression strains, especially when GroES was also overexpressed (Figure 3).

**Figure 3.** The native PAGE analysis of *E. coli* soluble extracts expressing genes from either pET-NtL-NtC60α20β or pET-NiL-NiC60αESβ and pCDF-NiXS1R1-AtR2B with the top panel showing Coomassie blue staining and the bottom panel showing the immunoblot with the antibody against Rubisco. The genes expressed are *Nt-rbcL* (P<sub>BAD</sub>), *Nt-rbcS-S1* (P<sub>T7</sub>), *Nt-rbcX*, *Nt-raf1*, *At-raf2*, *At-bsd2*, *Nt-cpn60α*, *Nt-cpn60β2* and either *Nt-cpn20* or *groES*. The lane marked with (*) was the Rosetta (DE3) extract expressing *At-cpn60α1*, *At-cpn60β2* and *At-cpn20*.

Subsequently, we replaced Arabidopsis *raf2* and *bsd2* genes in pCDF vectors with the tobacco versions and tested the expression of tobacco Rubisco in the BL21(DE3) strain. When the tobacco BSD2 was co-expressed instead of Arabidopsis BSD2, we observed increase in Rubisco expression although the improvement was not as dramatic as replacing Arabidopsis chaperonins with the tobacco ones (Figure 4). When Arabidopsis Raf2 was replaced with tobacco Raf2, there was no improvement in the assembly of Rubisco in *E. coli* (Figure 4). Although there appears to be no species specificity associated with Raf2 for Rubisco assembly, Raf2 is still important and necessary since the samples with lower tobacco Raf2 produced little Rubisco in *E. coli* (lanes marked with * in Figure 4).

**Figure 4.** The native PAGE analysis of BL21 (DE3) *E. coli* soluble extracts expressing genes from either pET-NiL-AtC60αβ20 or pET-NiL-NiC60αESβ and either pCDF-NiXS1R1-AtR2B, pCDF-NiXS1R1B-AtR2 or pCDF-NiXS1R1R2B with the top panel showing Coomassie blue staining and the bottom panel showing the immunoblot with the antibody against Rubisco. The genes expressed are *At-rbcL* (P<sub>BAD</sub>), *At-rbcS-S1* (P<sub>T7</sub>), *At-rbcX*, *At-raf1* and the other five genes as indicated on top of the lanes. The lanes marked with (*) were from *E. coli* strains expressing significantly less tobacco *raf2* due to difference in sequence near its start codon.
Functional Rubisco can be produced in *E. coli* expressing different tobacco small subunits.

In higher plants, the Rubisco small subunits are expressed from multiple nuclear genes, usually encoding slightly different amino acid sequences. In tobacco, 13 small subunit genes were previously identified (Gong et al., 2014). We analyzed their expression levels using publicly available RNA-Seq data generated from tobacco leaf samples (Sierro et al., 2014) and found that 11 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 (Figure 5A). Several of these genes encode identical mature small subunits, resulting in 7 unique small subunits (SSU-S1, SSU-S2, SSU-S5, SSU-T1, SSU-T2, SSU-T4a and SSU-T5) (Figure 5B).

Figure 5. Survey of the Rubisco small subunits in tobacco. (A) Comparison of transcript abundances estimated from an RNA-Seq experiment of tobacco leaf tissue (Sierro et al., 2014) with Kallisto software (Bray et al., 2016). The error bars are standard deviations of the abundances obtained from three SRA files. (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-SSU-T1 and Nt-SSU-T4a are indicated with red rectangles.
We co-expressed the tobacco rbcL and each of these 7 small subunits in BL21(DE3) together with two tobacco chaperonins (Cpn60α, Cpn60β), four tobacco chaperones (Raf1, Raf2, BSD2, RbcX) and E. coli GroES. We were able to detect an L8S8 Rubisco band similar to that from a tobacco leaf extract on a native PAGE immunoblot for 5 out of 7 small subunits. For E. coli co-expressing either Nt-rbcS-T1 or Nt-rbcS-T4a, a slower-migrating band was instead observed (Figure 6). Next, we used these E. coli extracts to measure their RuBP carboxylation activities using 14C bicarbonate in the absence of O2. We found that the E. coli soluble extract with each of the five small subunits that produced L8S8 Rubisco bands displayed RuBP carboxylation activities (Figure 6). Surprisingly, of the two E. coli extracts that did not produce a normal L8S8 Rubisco band on the native PAGE immunoblot, the one co-expressing SSU-T1 was also able to carboxylate RuBP (Figure 6). We then quantified the Rubisco active sites in each of these extracts using 14C-CABP binding assays to determine the yield of Rubisco from these E. coli cultures (Kubien et al., 2011). We found that the E. coli strain expressing SSU-T1 produced higher levels of Rubisco active sites than the strains expressing four other SSUs (Figure 7).

**Figure 6.** Comparison of tobacco Rubisco expressed in BL21 (DE3) with different small subunits. RuBP carboxylation rates shown on the top are averages of two measurements each using 14C bicarbonate. The native PAGE immunoblot in the bottom was obtained with the antibody against Rubisco.

**Figure 7.** The yields of tobacco Rubisco from BL21 (DE3) E. coli expressing different small subunits as determined by 14C-CABP binding assays. The yields were based on 45 mL cultures. The values plotted are averages of two measurements each.
We also measured the RuBP carboxylation rates at six different dissolved CO$_2$ levels for each E. coli extract and fitted the data to the standard Michaelis-Menten model to estimate their $K_C$ and $k_{cat}$ values. We found that the Rubisco produced in E. coli with each of the five small subunits (S1, S2, S5, T2 and T5) has an overall carboxylation kinetic profile that is very similar to the native Rubisco from the tobacco leaf although the enzymes produced in E. coli have slightly higher $K_C$ values (Figure 8). In contrast, the enzyme produced in E. coli with SSU-T1 displayed an inferior carboxylation activity due to its higher $K_C$ and lower $k_{cat}$ (Figure 8) in comparison to native Rubisco.

**Figure 8.** The kinetics of tobacco Rubisco expressed in E. coli with different small subunits compared to the native tobacco Rubisco in the absence of O$_2$. The Michaelis-Menten constants for CO$_2$ ($K_C$) and turnover numbers ($k_{cat}$) were obtained from nonlinear regression with the error bars showing 95% confidence intervals.

The Rubisco enzyme assembled in E. coli with a phylogenetically diverse small subunit found in the tobacco trichome has significantly different kinetic properties.

Recent studies discovered a phylogenetically diverse line of small subunits mainly expressed in nonphotosynthetic tissues of plants including tobacco trichomes (Morita et al., 2016; Laterre et al., 2017; Pottier et al., 2018). We co-expressed the trichome small subunit in BL21(DE3) with the same set of tobacco chaperonins and chaperones as described above. The corresponding E. coli extract had a normal L$_8$S$_8$ Rubisco band on the native PAGE immunoblot, and displayed the highest level of RuBP carboxylation activity as well as the highest level of Rubisco active sites (Figures 6 and 7). The tobacco trichome Rubisco produced in E. coli had carboxylation kinetics that are considerably different from the native enzyme found in mesophyll tissue and also different from those assembled in E. coli with other small subunits (Figure 8). Its higher turnover number and lower affinity for CO$_2$ are consistent with the recent findings concerning tobacco trichome Rubisco (Laterre et al., 2017).

**Discussion**

Aigner et al. (2017) recently overcame a major obstacle in functional studies of plant Rubisco by successfully expressing Arabidopsis Rubisco in E. coli. We modified two of the expression vectors used in their studies to enhance the expression of tobacco Rubisco in E. coli. Our two-vector system is compatible with E. coli expression systems with additional tRNAs for rare codons, which could be advantageous especially because none of the nine genes that are being co-expressed has codon usage optimized for E. coli. The ability to express plant Rubisco in E. coli means that both large and small subunits of this important enzyme can be rapidly modified for functional studies. Instead of expressing the two Rubisco subunits from a single vector, we expressed them from two different vectors in our two-vector system, which considerably reduces the number of vectors needed to express different combinations of large and small subunits.
We increased the assembly of tobacco Rubisco in *E. coli* by replacing Arabidopsis chaperonins and chaperones with tobacco versions. The most improvement was observed when the Arabidopsis Cpn60α, Cpn60β and Cpn20 with tobacco chaperonins Cpn60α, Cpn60β and the *E. coli* co-factor, GroES (Figure 4). In a previous report, co-expressing Arabidopsis *rbcL* and *raf1* in tobacco chloroplasts resulted in higher Rubisco levels than the levels achieved by expressing only Arabidopsis *rbcL* (Whitney et al., 2015). Similarly, tobacco BSD2 was able to provide additional improvement in tobacco Rubisco assembly in our current study. However, co-expressing tobacco Raf2 no longer improved the yield of tobacco Rubisco in *E. coli*. Raf2 was previously shown to be critical for Rubisco assembly in maize (Feiz et al., 2014) and Arabidopsis (Fristedt et al., 2018). Despite its importance, Arabidopsis Raf2 appeared to be fully functional in *E. coli* strains expressing tobacco Rubisco. Currently, Raf2 is the only Rubisco assembly chaperone, which mechanism has yet to be solved, and further structural studies of Raf2 in complex with Rubisco subunits will be critical to identify its role in Rubisco assembly.

Our current study shows that the RuBP carboxylation kinetics of tobacco Rubisco expressed in *E. coli* and the native enzyme from tobacco leaf are very similar but not identical (Figure 8). Clearly, the composition of small subunits in the native enzyme is heterogeneous, unlike the enzymes expressed in *E. coli*. Thus, it may be inappropriate to directly compare their kinetic properties. Interestingly, all five tobacco Rubisco, each with one of five different small subunits (S1, S2, S5, T2 and T5) expressed in *E. coli* have slightly higher *K_C* values than the native enzyme, indicating that there may be some fundamental differences in Rubisco assembled in *E. coli* other than the composition of small subunits (Figure 8). Rubisco in higher plants is known to undergo multiple conserved posttranslational modifications, and their exact roles are not well understood (Houtz et al., 2008). The conserved N-terminal acetylation of its large subunit and possibly other posttranslational modifications were found to be missing in the Arabidopsis Rubisco expressed in *E. coli* (Aigner et al., 2017). We hypothesize that some of these posttranslational modifications may be one of the strategies that plants have evolved to increase the CO₂ affinity of Rubisco as they adjusted to the declining global atmospheric CO₂ level over the last 20 million years prior to the industrial revolution (Pearson and Palmer, 2000). Clearly, further genetic, proteomic and biochemical studies will be necessary to address this issue. Despite the minor differences between tobacco Rubisco in *E. coli* and in tobacco leaves, we believe the two enzymes are sufficiently similar to make future functional studies informative.

Two tobacco small subunits, SSU-T1 and SSU-T4a, have unique residue substitutions, which may explain why they were unable to assemble proper Rubisco holoenzymes in *E. coli* (Figures 5 and 6). The substitution of conserved tryptophan 70 with arginine in SSU-T4a must have proven to be fatal because SSU-T4a and SSU-T2 are otherwise identical. Since the transcript level of SSU-T4a is negligible, it is likely that the nonfunctional SSU-T4a is physiologically irrelevant. On the other hand, SSU-T1 possesses the highest transcript level and has six unique residue substitutions. Surprisingly, the *E. coli* extract with SSU-T1 displayed some RuBP carboxylation activity with inferior kinetics despite lacking properly assembled holoenzyme (Figure 8). It is possible that some specific posttranslational modifications that are missing in either Rubisco subunit produced in *E. coli* are necessary for SSU-T1 to assemble into fully functional Rubisco.

We also found that the tobacco trichome small subunit was able assemble into functional Rubisco in *E. coli* with higher *k_cat* and *K_C* values. This trichome small subunit belongs to a phylogenetically distinct line of small subunits named T-type with only about 65% sequence identity to the canonical small subunits recently named M-type and usually expressed in non-photosynthetic tissues (Morita et al., 2016; Pottier et al., 2018). Previous studies with a rice T-type homolog (Morita et al., 2014) as well as the same tobacco trichome subunit (Laterre et al.,
indicated that these Rubisco enzymes exhibit higher $k_{cat}$ and $K_C$ values than those in mesophyll tissue. Although the active sites are located within dimers of large subunits in the holoenzyme, it has long been known that the small subunits can influence the carboxylation kinetics likely by modifying the conformation of the enzyme (Spreitzer, 2003). For example, studies in Chlamydomonas Rubisco indicated that changes along the interface between the large and small subunits resulted in an altered CO$_2$/O$_2$ specificity factor and other kinetic parameters (Spreitzer et al., 2005; Genkov and Spreitzer, 2009). 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 (Ishikawa et al., 2011).

Phylogenetic analysis of Rubisco has revealed that the substitution rates in the small subunits are much higher than those in the large subunit in higher plants (Andersson and Backlund, 2008). It is possible that diversification of the small subunits represents one strategy that plants have employed to adjust the kinetics of Rubisco as they adapt to different environmental conditions. Some information has been provided by analysis of Flaveria species, in which both C$_3$ and C$_4$ carbon fixation exists (Kapralov et al., 2011). Expression of variant and mutant small subunits 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 (Smith and Tabita, 2003; Mueller-Cajar and Whitney, 2008; Durao et al., 2015; Wilson et al., 2018), Rhodospirillum rubrum (Mueller-Cajar et al., 2007) and Methanococcoides burtonii (Wilson et al., 2016) and can be extended to plant Rubisco by co-expressing the required chaperonins and chaperones in E. coli. Recently, several studies measured detailed kinetics of Rubisco from a large number of plant species and identified potential kinetic switches in the large subunit (Galmés et al., 2014; Orr et al., 2016; Prins et al., 2016). Evolutionary analyses of the large subunits in C$_4$ plants also proposed residues that could be important in controlling the kinetics of Rubisco (Christin et al., 2008; Kapralov et al., 2012; Studer et al., 2014). However, due to enormous effort required to modify Rubisco in plants, only a few studies actually examined the effects of altering Rubisco in plants (Whitney et al., 1999; Whitney et al., 2011b). Our current study demonstrates that extension of the E. coli system to Rubisco enzymes from important crops should be feasible by incorporating appropriate species-specific assembly factors.
Materials and Methods

1. Construction of E. coli expression plasmids – We first cloned all the genes in Table 1 except for Nt-rbcL into BJFE holding vectors with a T7 promoter (P\(_{T7}\)), a ribosome binding site (RBS) and restriction sites (AscI, NotI) for rapid transfer into expression vectors. All the genes expressed in this study are summarized in Table 1, and the E. coli expression vectors in Table 2. The information on oligonucleotides used in generation of these vectors and full sequences of all vectors will be available upon request or at the NCBI website once the manuscript is published in a peer-reviewed journal.

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

*These accession numbers are from descriptions of blast results at [https://solgenomics.net/tools/blast/](https://solgenomics.net/tools/blast/) using Nitab v4.5 cDNA Edwards2017 database.
Table 2. Summary of plasmids used in the expression of tobacco Rubisco in *E. coli*. Only the final plasmids used in the *E. coli* expression strains are included here.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Genes Present*</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET-Nt-C60aβ20</td>
<td>Nt-rbcL (P_{BAD}), At-cpn60α1, At-cpn60β1, At-cpn20</td>
</tr>
<tr>
<td>pET-Nt-C60α20β</td>
<td>Nt-rbcL (P_{BAD}), Nt-cpn60α, Nt-cpn60β2, Nt-cpn20</td>
</tr>
<tr>
<td>pET-Nt-C60aESβ</td>
<td>Nt-rbcL (P_{BAD}), Nt-cpn60α, Nt-cpn60β2, groES</td>
</tr>
<tr>
<td>pCDF-NtXR1ArR2B-NtS1</td>
<td>Nt-rbcX, Nt-raf1, At-raf2, Nt-bsd2, Nt-rbcS-S1 (P_{BAD})</td>
</tr>
<tr>
<td>pCDF-NtXS1R1-ArR2B</td>
<td>Nt-rbcs-S1, Nt-rbcsX, Nt-raf1, At-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXS1R1B-ArR2</td>
<td>Nt-rbcs-S1, Nt-rbcsX, Nt-raf1, At-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXS2R1R2B</td>
<td>Nt-rbcs-S2, Nt-rbcsX, Nt-raf1, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXS5R1R2B</td>
<td>Nt-rbcs-S5, Nt-rbcsX, Nt-raf1, Nt-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXT1R1R2B</td>
<td>Nt-rbcs-T1, Nt-rbcsX, Nt-raf1, Nt-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXT2R1R2B</td>
<td>Nt-rbcs-T2, Nt-rbcsX, Nt-raf1, Nt-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXT4R1R2B</td>
<td>Nt-rbcs-T4a, Nt-rbcsX, Nt-raf1, Nt-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXT5R1R2B</td>
<td>Nt-rbcs-T5, Nt-rbcsX, Nt-raf1, Nt-raf2, Nt-bsd2</td>
</tr>
<tr>
<td>pCDF-NtXtricR1R2B</td>
<td>Nt-rbcs-trichome, Nt-rbcsX, Nt-raf1, Nt-raf2, Nt-bsd2</td>
</tr>
</tbody>
</table>

*Nt-rbcL* in the three pET vectors and *Nt-rbcs-S1* in pCDF-NtR1ArR2B-NtS1 and pCDF-NtXR1ArR2B-NtS1 vectors are under the arabinose-inducible promoter (P_{BAD}). All other genes are under the T7 promoter (P_{T7}).

(i) pET-Nt-C60αβ20 – *NtrbcL*-S2A which has alanine in place of serine at the second residue was amplified from tobacco DNA and ligated into NcoI-HindIII sites of pBAD-Dest49. Site-directed mutagenesis was next carried out to convert *NtrbcL*-S2A back into the wild-type sequence to obtain pBAD-Nt-rbcL vector. Next, *araC*-P_{BAD}-Nt-rbcL-term was amplified from the vector and ligated into NaeI sites of pArC60αβ/C20 (Aigner et al., 2017).

(ii) pET-Nt-C60α20β – The fragment between the two MluI sites in pET-Nt-C60αβ20 was removed to generate pET-Nt-C60αβ20-MluI vector. Nhel-RBS-Ntcpn60αa-P_{T7}-RBS-Ntcpn60β2-BsiWI was amplified from pRSF-Nt20αβ vector and ligated into XbaI and Acc65I sites of pET- Nt-C60αβ20-MluI to generate pET-Nt-C60αβ-2MluI. Site-directed mutagenesis was then carried out to remove the MluI site in Ntcpn60β2 to generate pET-Nt-C60αβ-1MluI. The fragment between the two MluI sites in pET-Nt-C60αβ20 was then inserted into pET-Nt-C60αβ-1MluI to obtain pET-Nt-Nt-C60αβ. MauBI-NotI fragment from BJFE-RBS-Ntcpn20 was then ligated into Ascl-NotI sites of pET-Nt-Nt-C60αβ to obtain pET-Nt-Nt-C60α20β.

(iii) pET-Nt-C60aESβ – MauBI-NotI fragment from BJFE-P_{T7}-RBS-groES was ligated into Ascl-NotI sites of pET-Nt-Nt-C60αβ to generate pET-Nt-Nt-C60αESβ.

(iv) pCDF-NtXR1ArR2B-NtS1 – pCDF-NtXR1ArR2B was obtained from Manajit Hayer-Hartl at Max Planck Institute of Biochemistry, Martinsried, Germany. Kpn2I-P_{BAD}-NtS1-B1002-Kpn2I was generated with overlapping PCR and ligated into the AgeI site of pCDF-NtXR1ArR2B to
obtain pCDF-NtR1Atr2B-NtS1. NcoI-NtrbcX-NotI was then ligated into pCDF-NtR1Atr2B-NtS1 to obtain pCDF-NtXR1Atr2B-NtS1.

(v) pCDF-NtXS1R1Atr2B-NtS1 was ligated into pCDF-NtR1Atr2B-NtS1. MauBI-NotI fragment from BJFE-P77-RBS-NtS1 was then ligated into AscI-NotI sites of pCDF-NtXR1Atr2B to obtain pCDF-NtXS1R1Atr2B.

(vi) pCDF-NtXS1R1B-Atr2 – IG2-NtB-IG3 was generated with overlapping PCR and ligated into SpeI-XhoI sites of pCDF-NtXS1R1Atr2B to obtain pCDF-NtXS1R1B-Atr2.

(vii) pCDF-NtXS1R1R2B – IG1-NtR2-IG2 was generated with overlapping PCR and ligated into PstI-SpeI sites of pCDF-NtXS1R1B-Atr2 to obtain pCDF-NtXS1R1R2B.

(viii) pCDF-NtX(SSU)R1R2B – NtR2 from pCDF-NtXS1R1R2B was ligated into PstI-XhoI sites of pCDF-NtXR1R2B to obtain pCDF-NtX(SSU)R1R2B with different tobacco small subunit genes after the Nt-rbcX gene.

2. Expression of tobacco Rubisco in E. coli – The E. coli cultures were grown at 37 °C 250 rpm overnight in LB medium and diluted 60-200 times into 6 mL or 45 mL LB. Ampicillin, spectinomycin and chloramphenicol were added at 150, 100 and 30 µg mL⁻¹ respectively to the growth medium as necessary. Once the OD₆₀₀ reached at about 0.3, the cultures were induced with 300 µM IPTG and grown at room temperature 250 rpm for 3 hours. They were then pelleted at 2500 rcf at room temperature for 5 minutes, resuspended in fresh LB medium containing 0.4% arabinose and grown at room temperature 250 rpm for 16-18 hours. They were then pelleted again and stored at -80 °C.

3. Native PAGE of E. coli extracts – The E. coli pellets were resuspended in 0.5 – 1.0 mL of lysis buffer consisting of 50 mM Tris-HCl pH 8, 10 mM MgCl₂, 1 mM EDTA, 2 mM DTT and protease inhibitor cocktail (Thermo Scientific) and broken by sonication on ice. For tobacco leaf sample, the leaf tissues were ground in a Wheaton homogenizer in 100 mM Bicine/NaOH pH 8.0, 10 mM MgCl₂, 1 mM EDTA, 1 mM ε-aminocaproic acid, 1 mM benzamidine, protease inhibitor cocktail, 1 mM phenylmethanesulfonyl fluoride, 1 mM KH₂PO₄, 2% w/v poly(ethylene glycol) 4000, 10 mM NaHCO₃ and 10 mM DTT. After the cell debris were removed by centrifugation at 16,000 rcf for 5 minutes, the protein concentrations in the supernatants were estimated by the Bradford assay. About 20 µg of total soluble proteins of each sample was loaded to 3-12% Bis-Tris 1.0 mm Invitrogen NativePAGE™ protein gels from ThermoFisher Scientific. Invitrogen NativePAGE™ running buffer from ThermoFisher Scientific was used as the running buffer with 0.002% Coomassie Brilliant Blue G-250 added to the cathode buffer. The native PAGE was run at 4 °C at 150-250 V for about 1.5 hours and then transferred to PVDF membranes with 0.45 um 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 secondary antibody in 2.5% milk in TBST buffer at room temperature and chemiluminescence was recorded with the ChemiDoc™ MP Imaging System from Bio-Rad.

4. Quantification of relative expression of rbcS genes in Tobacco – Publicly available SRA files used for tobacco genome sequencing (NCBI Bioproject -PRJNA208209) (Sierro et al., 2014) were utilized to quantitate 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 (Bray et al., 2016) with standard parameters.

5. Quantification of Rubisco active sites using $^{14}$C-CABP – A previously described protocol (Whitney and Sharwood, 2014) was followed to synthesize $^{14}$C-CABP from RuBP (Sigma-Aldrich part number 83895) and $^{14}$C-KCN (PerkinElmer). 40-100 µL of each sample was incubated with 7.2 nmol $^{14}$C-CABP with 5 mCi/mmol of specific activity for at least 20 minutes and applied to size-exclusion chromatography with Sephadex G50 Fine (Santa Cruz Biotechnology) as described previously (Kubien et al., 2011). The eluted fractions were mixed with Ultima Gold liquid scintillation cocktail from PerkinElmer and the activities of Rubisco-bound $^{14}$C-CABP were counted with a Beckman LS 6000IC scintillation counter.

6. Determination of RuBP carboxylation kinetics – About 15 mL of 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). 915 µL was then added to individual two-dram glass vials, which were then sealed with open top caps with bonded PTFE faced silicone liners (Wheaton part number W240842), and equilibrated again with CO$_2$-free N$_2$ gas for at least another 30 minutes at 25 °C. Six different concentrations of 50 µL of $^{14}$C bicarbonate were then added to the vials to achieve six final bicarbonate concentrations ranging from 10.55 mM to 177.82 mM. 15 µL of 26.7 mM RuBP (Santa Cruz Biotechnology, sc-214827A) was then added to each vial. 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 about 100 °C. Once almost all solutions in the vials were evaporated, 0.5 mL of ddH$_2$O was added to each vial to dissolve the leftover chemicals. Each vial was mixed with at least 3.5 mL of Ultima Gold liquid scintillation cocktail from PerkinElmer and the acid-stable $^{14}$C compounds in the vials were counted with a Beckman LS 6000IC scintillation counter. The catalytic rates at different [CO$_2$] concentrations were then fitted to the standard Michaelis-Menten equation with nonlinear regression in R software to obtain the $K_c$, $V_{max}$ and 95% confidence intervals. The value of $k_{cat}$ was then obtained by dividing $V_{max}$ with the Rubisco active sites in each sample.

Acknowledgments

We thank Manajit Hayer-Hartl from Max Planck Institute of Biochemistry in Martinsried, Germany, for providing us with their E. coli expression vectors. We also thank Douglas Orr, Elizabete Carmo-Silva and Martin Parry from Lancaster University in Lancaster, UK for advice on experiments to quantify Rubisco active sites and measure RuBP carboxylation kinetics. This study is funded by the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy to M.R.H. (Award Number: DE-SC0014339).

References


