

Dissection of *Synechococcus* Rubisco Large Subunit Sections Involved in Holoenzyme Formation in *Escherichia coli* by Combinatorial Section Swapping and Sequence Analyses

(Pembahagian *Synechococcus* Rubisco Seksyen Subunit Besar Terlibat dalam Pembentukan holoenzim dalam *Escherichia coli* oleh Seksyen Kombinatori Tertukar dan Jujukan Analisis)

YEE HUNG YEAP, TENG WEI KOAY, HANN LING WONG & BOON HOE LIM*

ABSTRAK

Engineering the CO₂-fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) to improve photosynthesis has long been sought. Rubisco large subunits (RbcL) are highly-conserved but because of certain undefined sequence differences, plant Rubisco research cannot fully utilise the robust heterologous Escherichia coli expression system and its GroEL folding machinery. Previously, a series of chimeric cyanobacteria Synechococcus elongatus Rubisco, incorporated with sequences from the green alga Chlamydomonas reinhardtii, were expressed in E. coli; differences in RbcL sections essential for holoenzyme formation were pinpointed. In this study, the remaining sections, presumably not crucial for holoenzyme formation and also the small subunit (RbcS), are substituted to further ascertain the possible destabilising effects of multiple section mutations. To that end, combinations of Synechococcus RbcL Sections 1 (residues 1-47), 2 (residues 48-97), 5 (residues 198-247) and 10 (residues 448-472), and RbcS, were swapped with collinear Chlamydomonas sections and expressed in E. coli. Interestingly, only the chimera with Sections 1 and 2 together produces holoenzyme and an interaction network of complementing amino acid changes is delineated by crystal structure analysis. Furthermore, sequence-based analysis also highlighted possible GroEL binding site differences between the two RbcLs.

Keywords: *Chaperone; Chlamydomonas reinhardtii; protein assembly; ribulose bisphosphate carboxylase/oxygenase (Rubisco); Synechococcus elongatus PCC6301*

ABSTRAK

Kajian untuk mengubah suai ribulosa-1,5-bisfosfat karboksilase/oksigenase (Rubisco) bagi memperbaiki proses fotosintesis adalah usaha yang telah lama dijalankan. Subunit- besar Rubisco amat konservatif tetapi disebabkan perbezaan jujukan asid amino yang tertentu, Rubisco tumbuh-tumbuhan tidak dapat dikaji dengan menggunakan sistem pengekspresan Escherichia coli yang serba-boleh serta mekanisme penglipatan GroEL-nya. Sebelum ini, satu siri Rubisco kimerik yang menggabungkan jujukan daripada cyanobacteria Synechococcus elongatus dengan alga hijau Chlamydomonas reinhardtii telah diekspreskan ke dalam E. coli; dalam uji kaji tersebut, perbezaan yang merangkumi seksyen RbcL yang mustahak dalam pembentukan holoenzim telah ditentukan. Dalam uji kaji ini, seksyen lain yang mungkin tidak penting untuk pembentukan holoenzim, bersama-sama subunit kecil (RbcS) telah digantikan untuk menentukan kemungkinan kesan ketidakstabilan akibat mutasi seksyen berbilang. Untuk itu, kombinasi Synechococcus RbcL Seksyen 1 (residu 1-47), 2 (residu 48-97), 5 (residu 198-247) dengan 10 (residu 448-472) dan RbcS, telah digantikan dengan seksyen Chlamydomonas yang kolinear dan diekspreskan dalam E. coli. Kesimpulannya, hanya kimera yang ditukarkan kedua-dua Seksyen 1 dan 2 dapat membentuk holoenzim dan rangkaian interaksi yang meliputi perubahan asid amino yang saling melengkapi berdasarkan analisis struktur kristal telah dikemukakan. Selain itu, analisis berasaskan jujukan asid amino juga menunjukkan bahawa perbezaan tapak ikatan GroEL yang mungkin bagi RbcL.

Kata kunci: *Chaperone; Chlamydomonas reinhardtii; himpunan protein; ribulosa-1,5-bisfosfat karboksilase/oksigenase (Rubisco); Synechococcus elongatus PCC6301*

INTRODUCTION

Ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco, E.C 4.1.1.39) is responsible for the sustainability of our biosphere, as it is the first enzyme in the Calvin cycle that carries out carbon fixation with the purpose of assimilating atmospheric carbon dioxide into organic molecules for cellular metabolism and biomass

accumulation (Andersson & Backlund 2008; Campbell & Ogren 1992). Rubisco makes up about half of the total protein composition in C3 leaf cells and approximately one third of the total soluble protein in C4 leaves, thus hinting at its importance (Ellis 1979; Feller et al. 2008). Different forms of the multi-subunit enzyme have been elucidated, including the hexadecameric form I, the

dimeric form II, the multi-dimeric form III and the multi-dimeric Rubisco-like form IV (Tabita et al. 2008, 2007a). All these are putatively derived from an ancestral protein in a methanogenic archaeon (Tabita et al. 2007b). Hexadecameric Form I Rubisco has long been the subject of intensive study, due to it being the major form found in higher plants and many microorganisms, with the end-goal often being the successful engineering of a 'better' Rubisco (Andersson & Backlund 2008; Genkov et al. 2006; Gutteridge et al. 1993; Spreitzer et al. 1995). Form I Rubisco consists of eight large subunits (RbcL), which are arranged as a tetramer of dimers. These are capped by eight small subunits (RbcS) at the top and bottom of the RbcL octamer (Figure 1(A)). Each holoenzyme possesses eight active sites, which are formed by the C-terminal domain of one RbcL and the N-terminal domain of an adjacent RbcL in each dimer (Curmi et al. 1992; Knight et al. 1990; Taylor et al. 2001).

Generally, Rubisco catalyses the carboxylation of ribulose-1,5-bisphosphate (RuBP) with CO_2 , producing two molecules of 3-phosphoglycerate, which contributes toward photosynthetic growth, but this 'schizophrenic' bifunctional enzyme sometimes catalyses oxygenation of RuBP with O_2 (up to a third of the time under atmospheric conditions), producing 2-phosphoglycolate, which has to be metabolized in an unfavourable ATP-consuming CO_2 -releasing photorespiratory process (Laing et al. 1974; Ogren 1984; Peterhansel et al. 2008). This difficulty in discriminating between competing substrates CO_2 and O_2 , which is attributed to their electrostatic similarities (Kannappan & Gready 2008), restricts photosynthetic growth considerably in most plants, leading to losses in biomass yields (Bainbridge et al. 1995; Parry et al. 2007). Indeed, the specious diversity of Rubisco kinetic properties and evolution of CO_2 -concentrating mechanisms found in particular microalgae, cyanobacteria and higher plants point at shortcomings within the enzyme that are preventing maximum carbon fixation (Chen & Spreitzer 1992; Meyer et al. 2012; Tcherkez et al. 2006). Oft times a high CO_2/O_2 specificity is compromised by a low carboxylation rate and vice versa (Chen & Spreitzer 1992; Tcherkez et al. 2006; Whitney et al. 2011).

The search for a 'better' Rubisco began when kinetic properties of Rubiscos from various photosynthetic organisms were elucidated, with Rubiscos from thermophilic red algae having the highest CO_2/O_2 specificity, alongside an acceptable albeit lower carboxylation rate compared to crop species (Jordan & Ogren 1981; Tcherkez et al. 2006; Whitney et al. 2011). It was calculated that the successful substitution of Rubisco from C_3 crop plant with that of the red alga *Griffithsia monilis* (CO_2/O_2 specificity factor of 167) can well increase photosynthetic yield by more than 25% (Long et al. 2006; Zhu et al. 2004). Unfortunately, Rubisco subunits from phylogenetically distant organisms often fail to assemble, which has limited the diversity of Rubisco genes available for genetic engineering.

When the form I Rubiscos of non-green alga *Galdieria sulphuraria* and diatom *Phaeodactylum tricornutum* were introduced into tobacco, the holoenzymes could not form, with the possible cause being an incompatible chaperone-mediated folding pathway (Whitney et al. 2001). It is only recently that plant Rubiscos were functionally expressed in *E. coli* by co-expressing their chaperone counterparts (Aigner et al. 2017). Mutagenesis and directed evolution studies involving localized changes of the holoenzyme on the other hand yielded mixed results, with the most promising being those involving Rubisco-dependent *E. coli* selection systems whereby there are trade-offs between functional expression and kinetic efficiencies (Mueller-Cajar & Whitney 2008; Parikh et al. 2006; Smith & Tabita 2003; Wilson et al. 2017).

Sequence and structure-based approaches to investigate non-formation of the holoenzyme would involve analysing effects of each residue substitution on intra-molecular interactions that affects protein conformation, while also assessing solvent accessibility changes in the RbcL sequence for each progression of the subunit assembly up to the order of the complete holoenzyme (Stan et al. 2006). This is critical as placing hydrophobic residues on solvent-exposed surfaces might lead to protein aggregation while having a polar residue within the hydrophobic core might be destabilizing to the protein structure (Cordes & Sauer 1999; Pakula & Sauer 1990). However, there are cases whereby polar residues within the interior of folded proteins contribute greater stability via hydrogen bonding and van der Waals interactions (Pace 2001). In fact, buried polar groups in mutant human lysozymes exact a low-energy cost that is compensated by forming hydrogen bonds to overcome the loss of hydrophobic effects on protein stability (Takano et al. 2001). Furthermore, the number of residue substitutions is also a factor to consider. With *Synechococcus* and *Chlamydomonas* RbcLs having a total of 85 residue differences, there should be a limit to the number of substitutions a wild-type Rubisco can tolerate before both function and stability are lost (Bloom et al. 2005; Burger et al. 2006).

In a previous study, an effort was made to determine the RbcL domains of cyanobacteria Rubisco that are essential for successful holoenzyme formation in *Escherichia coli* (Koay et al. 2016). Sections spanning the whole coding sequence of cyanobacterial *rbcL* from *Synechococcus* PCC6301 were swapped sequentially with homologous sections of the eukaryotic *rbcL* gene from *Chlamydomonas reinhardtii* to generate ten chimeric mutants. In the end, it was established that four *Synechococcus* RbcL sections with residues 1-47, 48-97, 198-247 and 448-472 could be separately swapped without nullifying holoenzyme formation in *E. coli* (Koay et al. 2016).

In this follow-up study, we attempt to determine the extent to which *Synechococcus* RbcL can be phylogenetically substituted to *Chlamydomonas* RbcL

and still form holoenzyme in *E. coli*. To do so, various combinations of *Synechococcus* RbcL sections involving the aforementioned residues 1-97, 198-247 and 448-472 were swapped with their collinear sections in *Chlamydomonas* RbcL. As it turns out, *Synechococcus* RbcL can only tolerate the combined swapping of residues 1-47 and 48-97, whereby holoenzyme formation for this mutant is detectable in *E. coli* lysate via non-denaturing PAGE and immunoblotting, but in a greatly reduced amount. Furthermore, to determine whether changing RbcS to that of *Chlamydomonas* can structurally complement the amino acid changes in RbcL, a chimeric *Synechococcus* Rubisco with all four RbcL sections substituted and lacking holoenzyme formation had its RbcS substituted, but that did not restore holoenzyme formation.

MATERIALS AND METHODS

CHIMERIC *SYNECHOCOCCUS* RBCL CONSTRUCTION AND *CHLAMYDOMONAS* RBCS SUBSTITUTION

Chimeric Rubisco plasmids (Figure 1(B)) were mainly created by using primers and existing plasmids from Koay et al. (2016) to amplify selected fragments, which were then digested with restriction enzymes as indicated in the primer names and then ligated into the pTrcHisB vector backbone of NcoI/PstI-digested pTrcSynLS (Mueller-Cajar & Whitney 2008). pL¹⁺¹⁰S was constructed by ligating fragments from pTrcSynL(Chl1-50)S (Koay et al. 2016) and pTrcSynL(Chl451-475)S (Koay et al. 2016), which were amplified with primer pairs ChlN-NcoI/Syn150-rev-BsmBI (Koay et al. 2016) and Syn150-fwd-BsmBI/SynSS-C-PstI (Koay et al. 2016), respectively. pL¹⁺²⁺⁵⁺¹⁰S had fragments from pTrcChlLS-SynSS (a plasmid with *Synechococcus* *rbcL* replaced by *Chlamydomonas*, unpublished), pTrcSynL(Chl200-250)S (Koay et al. 2016) and pTrcSynL(Chl451-475)S (Koay et al. 2016), which were amplified with primer pairs ChlN-NcoI/Chl100-rev-BsmBI (Koay et al. 2016), Syn100-fwd-BsmBI/Syn400-rev-BsmBI (Koay et al. 2016) and Syn400-fwd-BsmBI/SynSS-C-PstI (Koay et al. 2016), respectively. pL¹⁺²S had fragments from pL¹⁺²⁺⁵⁺¹⁰S and pTrcSynLS (Mueller-Cajar & Whitney 2008), which were amplified with primer pairs ChlN-NcoI/Syn150-rev-BsmBI (Koay et al. 2016) and Syn150-fwd-BsmBI/SynSS-C-PstI (Koay et al. 2016), respectively. pL¹⁺²⁺¹⁰S had fragments from pL¹⁺²⁺⁵⁺¹⁰S and pTrcSynL(Chl451-475)S (Koay et al. 2016), which were amplified with primer pairs ChlN-NcoI/Syn150-rev-BsmBI (Koay et al. 2016) and Syn150-fwd-BsmBI/SynSS-C-PstI (Koay et al. 2016), respectively. pL¹⁺²⁺⁵⁺¹⁰S^C had fragments containing chimeric *rbcL* from pL¹⁺²⁺⁵⁺¹⁰S and *Chlamydomonas* *rbcS* from pSS1-ITP (Genkov et al. 2010), which were amplified with primer pairs ChlN-NcoI (Koay et al. 2016)/SynSSN-link-rev-BsmBI (5'-CTTTGGGCAGAGTTCGTCTCCT

CATGATTCAGAAATC-3') and ChlSS-N-fwd-BsmBI (5'-CTCAGGCCGTCTCGATGATGGTCTGGACC-3')/ChlSS-C-rev-PstI(5'-CTCAGATCGCTGCAGATCTCGACTTACACGGAGCGCTTG-3'). Constructed plasmids were electroporated into *E. coli* XL-1 Blue, which were then selected at 37°C overnight on LB plates with 200 µg/mL ampicillin. Plasmids were verified by size screening on an agarose gel, restriction enzyme mapping and sequencing of the bicistronic *rbcL-rbcS* operon.

PROTEIN EXPRESSION AND ASSEMBLY ANALYSES OF CHIMERIC RUBISCOS

Essentially, Rubisco expression and protein analyses were carried out similar to Koay et al. (2016). Briefly, XL-1 Blue *E. coli* were electroporated with the verified *rbcL-rbcS* constructs and selected on 200 µg/mL ampicillin LB plates at 37°C to obtain single colonies, which were picked and grown overnight at 37°C in 10 mL LB cultures (100 µg/mL ampicillin). These starter cultures were then transferred to fresh 100 µg/mL ampicillin LB broths and grown to OD₆₀₀ of 0.5 at 37°C. Rubisco expression was induced with 0.5 mM IPTG for 16 h, after which, *E. coli* cells were harvested, resuspended to 10% (w/v) in ice-cold extraction buffer (50 mM Bicine-NaOH, pH8.0, 10 mM MgCl₂, 10 mM NaHCO₃, 2 mM DTT) and sonicated. Following that, total cellular proteins were resolved on 7.5% native-PAGE gel, or denatured by mixing with SDS-containing loading buffer and boiled for 30 min before resolving on 12% SDS-PAGE gels. For Western blotting, resolved proteins were probed using rabbit anti-*Synechococcus* PCC6301 Rubisco IgG (Parikh et al. 2006) and for the detection of *Chlamydomonas* RbcS in SDS-PAGE, additional rabbit anti-*Chlamydomonas* RbcS IgG (Esquivel et al. 2013) was used.

SCANNING OF *CHLAMYDOMONAS* AND *SYNECHOCOCCUS* RBCL SEQUENCES FOR GROEL-BINDING MOTIFS AND HYDROPHOBIC PATCHES

GroEL substrate-protein binding motifs are postulated to have amino acid functional groups characteristic of the GroES mobile loop, which binds in the hydrophobic groove formed by two helices of the GroEL apical domain (Stan et al. 2004). The most elementary motif is P_HHH, whereby 'P' refers to (p)olar amino acids Asn, Gln, Gly, His, Pro, Ser and Thr, '_' refers to any of the twenty amino acids, and 'H' refers to (h)ydrophobic amino acids Ala, Cys, Ile, Leu, Met, Phe, Trp, Tyr and Val (Stan et al. 2006). A further criterion applied in the search for binding motifs is that these residues should become inaccessible upon transition to the native assembled Rubisco oligomer (Stan et al. 2006).

A second approach to look for GroEL recognition sites was by screening *Synechococcus* and *Chlamydomonas* RbcL for local hydrophobic patches with average hydropathy indices (Kyte & Doolittle 1982) greater or similar to that of the GroEL-binding seven-residue GroES mobile loop region, which has amino acid identity GGIVLTG

and hydropathy index of 1.514 (Chaudhuri & Gupta 2005; Kumar et al. 2012).

RESULTS

STRUCTURAL POSITION OF THE SWAPPED LARGE SUBUNIT SECTIONS AND CONSTRUCTS OF CHIMERIC RUBISCOS

Residues changed in Rubisco large subunit Sections 1, 2 and 10 are mainly situated on the holoenzyme surface (Figure 1(A)). In addition, because Sections 1 and 10 are at the N- and C-terminal ends of the large subunit, respectively, these sections were predicted to be flexible and more tolerable of amino acid changes, therefore, a chimeric *Synechococcus* Rubisco was created with these two terminal end sections swapped to *Chlamydomonas* while retaining the *Synechococcus* small subunit ($L^{1+10}S$ in Figure 1(B)). Next, because Sections 1 and 2 are adjacent in the protein primary structure and comprise over 60% of the 150-amino acid ferredoxin-like fold N-terminal domain, these sections were predicted to have complementing

residue interactions that might warrant changing the interacting functional groups simultaneously, therefore, a chimeric *Synechococcus* Rubisco was made with only these two sections substituted with that of *Chlamydomonas* ($L^{1+2}S$ in Figure 1(B)). A chimera was also constructed replacing mainly the surface sections (Sections 1, 2 and 10) ($L^{1+2+10}S$ in Figure 1(B)) since perturbations on the holoenzyme surface is predicted to be less destabilizing to its structural core. Another chimera was also created with all four sections (Sections 1, 2, 5 and 10) substituted ($L^{1+2+5+10}S$ in Figure 1(B)) followed by a further chimera, which had the *Chlamydomonas* Rubisco small subunit ($L^{1+2+5+10}S^C$ in Figure 1(B)) to ascertain whether accompanying small subunit changes can complement large subunit alterations caused by our sectional swaps.

PROTEIN EXPRESSION

For all chimeric Rubisco constructs, the mutant genes were expressed in *E. coli*, with the RbcL and RbcS proteins detected in each transformant cell lysate when resolved by

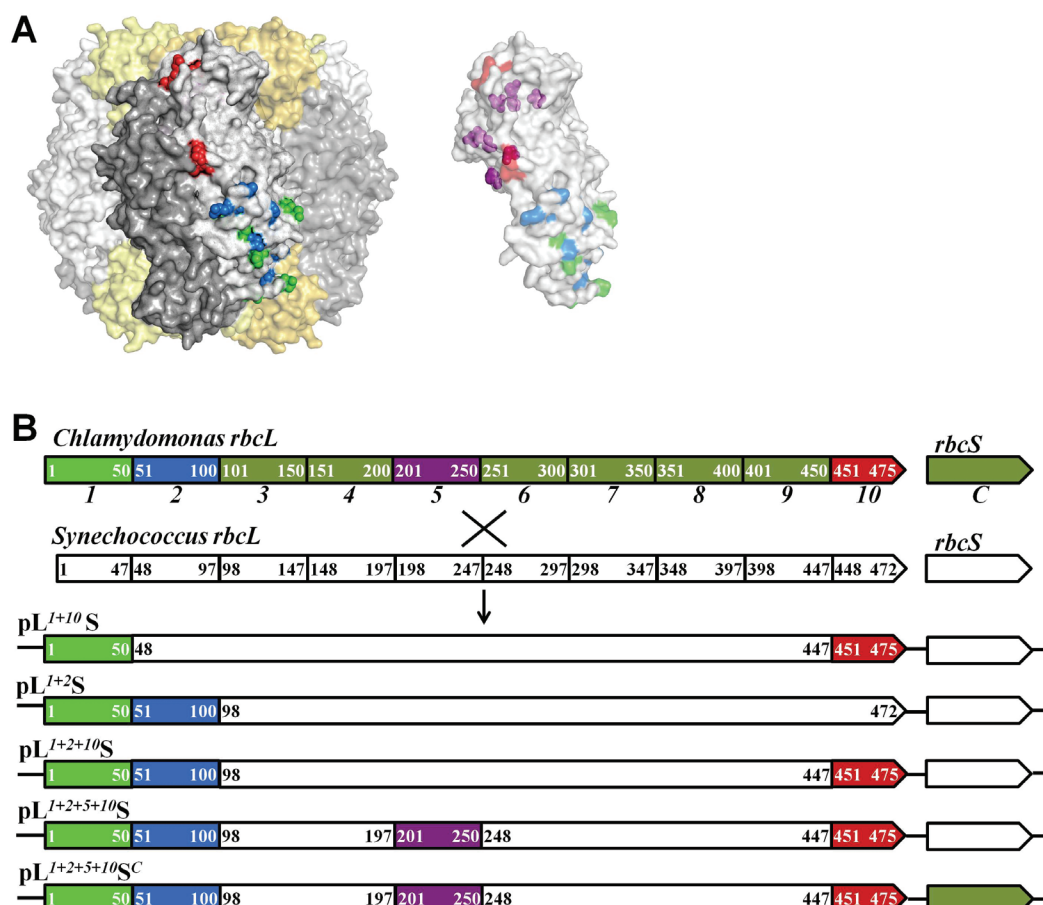


FIGURE 1. A) Location of amino acids mutated in the holoenzyme of *Synechococcus* Rubisco (PDB ID 1RBL). Left, complete hexadecameric holoenzyme with large subunits in silver and white and small subunits in gold and yellow. Mutated residues are indicated (green, Section 1; blue, Section 2; purple, Section 5; red, Section 10) for only one subunit. Right, a single subunit is shown as transparent surface with Section 5 residues as solid spheres (purple) and B) constructs of the chimeric Rubiscos with section numbers denoted below the *Chlamydomonas* genes. Amino acid numbers are demarcated in each section, in lieu of base pairs

SDS-polyacrylamide gel electrophoresis, blotted and probed with anti-Rubisco antibodies (Figure 2). RbcL levels only differed slightly among the various Rubisco transformants based on SDS-PAGE, albeit these differences become seemingly striking upon immunoblotting, most likely because of alterations to the epitope identity and antibody affinity in some of the mutants as a result of changing their protein sequences (Figure 2). However, for reasons that have yet to be ascertained, RbcS levels were lower for cells with the mutant constructs compared to the one with the wild-type construct (Figure 2), but because the *rbcL* and *rbcS* genes were expressed as a bicistronic transcript, it is reasonable to eliminate disparate transcript stabilities and levels as a cause for the reduced RbcS production.

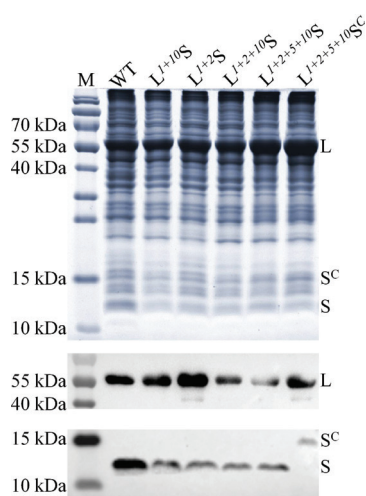


FIGURE 2. SDS-PAGE (top) and Western blot (middle and bottom) analyses of total cellular protein from *E. coli* expressing *Synechococcus* Rubisco (WT) or the various chimeric mutants. Protein marker (M) with band sizes (in kDa) labelled was run concurrently. *Synechococcus* large (L) and small (S) subunits and *Chlamydomonas* small subunit (S^C) are indicated

PROTEIN ASSEMBLY

Remarkably, only the chimeric Rubisco with RbcL sections 1 and 2 swapped (L¹⁺²S) formed complete hexadecameric holoenzyme at a detectable level on native-polyacrylamide gel and immunoblotting, though the amount of assembled holoenzyme is markedly less than that of wild type (Figure 3). The absence of other chimeric holoenzymes and the reduced amount of L¹⁺²S suggested that the various mutations are destabilizing and combining these mutations abrogates holoenzyme formation. Perhaps other complementing amino acid substitutions in RbcL sections that were not probed in the current study are required.

BIOINFORMATICS ANALYSES

GroEL binding motifs were found throughout RbcL, specifically in Sections 1, 2, 3, 4, 6, 9 and 10 (Supplementary File 1). When considering only binding motifs that are buried in the native hexadecameric Rubisco, *Chlamydomonas* RbcL has twelve putative GroEL binding motifs whereas *Synechococcus* RbcL only has nine (Table 1). The three additional motifs in *Chlamydomonas* RbcL are in Sections 6 and 9 (Table 1). Because binding motifs need to be ten residues apart on the primary structure in order to bind to adjacent GroEL subunits (Stan et al. 2004), the three additional motifs possibly extend the Sections 6 and 9 motifs to encompass extra contact sites with *Chlamydomonas* chaperones but distort the proper contacts with bacterial chaperones. Conversely, *Synechococcus* RbcL has twenty eight hydrophobic patches with hydrophathy indices greater than that of the GroES mobile loop, which is the best GroEL substrate, whereas *Chlamydomonas* RbcL only has sixteen (Supplementary File 2). These seven-residue hydrophobic patches could also hint at binding sites for GroEL. The fourteen additional patches in *Synechococcus* span Sections 2 (into 3), 3, 4,

TABLE 1. Substrate protein binding motifs in RbcLs that are buried in the native state

Section	Residue range	<i>Synechococcus</i> Sequence	<i>Chlamydomonas</i> Sequence
1	31-35	TDLLA	TDILA
	51-55	GAAIA	GAAVA
2	92-96	NSYFA	NQYIA
	93-97	SYFAF	QYIAY
3	138-142	PVALV	PPAYV
4	183-187	GRAVY	GRAVY
	275-279	TTLAK	TSLAI
6	276-280	TLAKW	SLAIY
	284-288	NGVLL	NGLLL
9	412-416	PGATA	PGAAA
	417-421	NRVAL	NRVAL
10	450-454	PELAA	PELAA

Residue numbering is based on *Synechococcus* RbcL. Binding motifs are bolded

6, 7, 8 and 9, while the two additional *Chlamydomonas* patches are in Sections 4 and 6 (Table 2). Notwithstanding the disparity in the number of GroEL binding sites found, both sequence-based methods in this study mainly preclude sites from Sections 1, 2, 5, and 10.

DISCUSSION

Here we show that when *Synechococcus* PCC6301 RbcL sections previously defined as non-determinants for holoenzyme formation (Koay et al. 2016) were simultaneously swapped with that of *Chlamydomonas*, the various mutant combinations mainly resulted in additive effects that nullified holoenzyme production. Only the chimeric Rubisco having RbcL Sections 1 and 2 substituted together, totalling 25 residue substitutions, could form in the *E. coli* transformants, albeit at reduced levels compared to wild-type Rubisco (Figure 3). This implies that even though Sections 1 (residues 1-47), 2 (residues 48-97), 5 (residues 198-247) and 10 (residues 448-472) of *Synechococcus* RbcL may be separately non-critical for holoenzyme assembly, they do contain residues that provide a certain stability to the overall protein conformation.

The structural basis for why residue changes in the aforementioned sections influence holoenzyme stability can be deduced by comparing the Rubisco structures of *Chlamydomonas* (PDB ID 1GK8) and *Synechococcus* (PDB ID 1RBL). The large subunit N-terminus has a different conformation between *Synechococcus* and

Chlamydomonas (Figure 4). Because substitutions in this region comprise of S9A and A10G (*Chlamydomonas* RbcL numbering) occur in Section 1 chimeras, which includes all chimeric proteins in this study, the chimeric large subunits are predicted to mimic the *Chlamydomonas* N-terminus structural arrangement, which could cause losses of van der

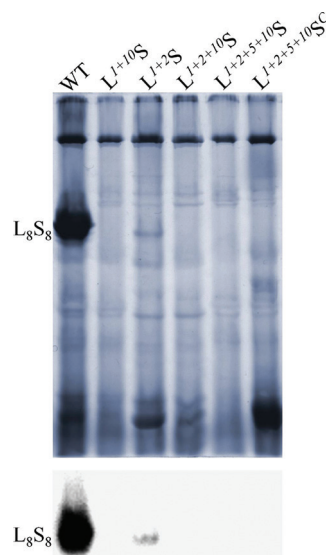


FIGURE 3. Native-PAGE (top) and Western blot (bottom) analyses of total cellular protein from *E. coli* expressing *Synechococcus* Rubisco (WT) or the various chimeric mutants. Assembled hexadecameric holoenzyme (L_8S_8) is indicated

TABLE 2. Sequences with hydrophaticities higher than the GroES mobile loop (1.514) and are present in either only *Synechococcus* or *Chlamydomonas* RbcL

Section	Residue range	<i>Synechococcus</i>		<i>Chlamydomonas</i>	
		Sequence	Hydropathy index	Sequence	Hydropathy index
2/3	93-99	SYFAFIA	1.6571	QYIAYVA	0.8857
	94-100	YFAFIAY	1.5857	YIAYVAY	1.2000
	95-101	FAFIAYP	1.5429	IAYVAYP	1.1571
	96-102	AFIAYPL	1.6857	AYVAYPI	1.1571
3	112-118	NILTSIV	1.7143	NMFTSIV	1.2000
	135-141	IRFPVAL	1.5714	LRIPPAY	0.1571
	136-142	RFPVALV	1.5286	RIPPAYV	0.2143
	137-143	FPVALVK	1.6143	IPPAYVK	0.3000
	139-145	VALVKTF	1.7429	PAYVKTF	0.1857
4	165-171	PMLGCTI	1.4286	GLLGCTI	1.8714
6	256-262	ELGMPII	1.3143	ELGVPII	1.6429
	261-267	IIMHDFL	1.5429	IIMHDYL	0.9571
7	337-343	ASTLGFV	1.5286	EVTLGFV	1.4857
8	370-376	GVLVAS	1.6000	GVMPVAS	1.3286
	371-377	VLPVASG	1.6000	VMPVASG	1.3286
9	419-425	VALEACV	2.1143	VALEACT	1.4143

Residue numbering is based on *Synechococcus* RbcL. Sequences with hydrophathy indices greater than 1.514 are bolded

Waals contacts between the large subunit residues and the *Synechococcus* small subunit residues His-39 in β -strand A and Phe-63 in the loop between β -strand B and α -helix B (Figure 4). A previous alanine-scanning mutagenesis study on *Chlamydomonas* small subunit demonstrated compelling evidence that β -strand A has a role in holoenzyme stability, whereby an E43A mutant enzyme was unstable at an elevated temperature of 35°C (Genkov & Spreitzer 2009). An F81A (Phe-63 in *Synechococcus*) mutant *Chlamydomonas* enzyme also had a decrease in thermal stability (Genkov & Spreitzer 2009). Therefore, it is unsurprising that our mutant chimeras with alterations that potentially disrupt complementing residue interactions between the large subunit N-terminus and small subunit have impaired protein stability. To test whether these disruptions can be restored by concomitant changes to the small subunit, we also constructed a chimeric enzyme with the small subunit swapped, but curiously, assembled holoenzyme still could not form (Figure 3).

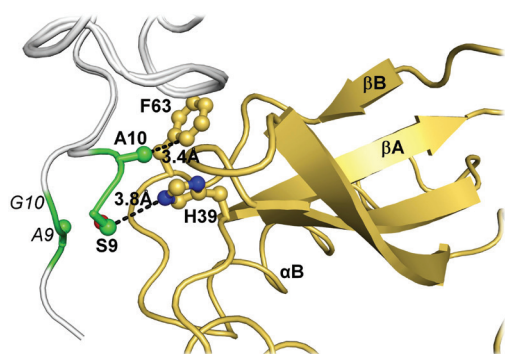


FIGURE 4. Overlaid *Synechococcus* (PDB ID 1RBL) and *Chlamydomonas* (PDB ID 1GK8) large subunits (white ribbons), with differing crucial N-terminus Section 1 residues (green) which interacts with the *Synechococcus* small subunit (gold). Critical residues are represented with ball and sticks (blue, nitrogen; red, oxygen) and bolded (*Synechococcus*) or italicised (*Chlamydomonas*). Large subunit residues are numbered based on the *Chlamydomonas* sequence. Measured distances (in Å) for large/small subunit contacts which are lost upon mutation are indicated

In addition to Section 1, changes in other sections are also predicted to perturb amino acid side-chain interactions. Large subunit Leu-73 (*Chlamydomonas* RbcL numbering), which is mutated to Gly-73 in the Section 2 chimeras, is expected to have weakened, if not lost van der Waals contacts between the significantly diminished glycine functional group and the large subunit N-terminus (Figure 5). Another Section 2 mutation, K83R (*Chlamydomonas* RbcL numbering) on β -strand C potentially lengthens the distance between residue 83 and the pyrrole ring of Pro-104, which is positioned between β -strand D and α -helix C (Figure 6). Interestingly, Pro-104 is hydroxylated in the *Chlamydomonas* enzyme, albeit replacement of this residue with alanine did not cause any obvious structural

disruptions despite its close proximity to the interface between interdimeric large subunits (Rasineni et al. 2017). Within the Section 2 chimeras, β -strand C also has a H86D (*Chlamydomonas* RbcL numbering) mutation, which might cause destabilizing repulsive forces with the similarly negatively-charged Glu-88 on the same secondary structural element (Figure 6). Indeed, our results accord with a previous study whereby several mutations on β -strand C, including D86R and P89A, halved the amount of holoenzymes produced relative to wild type (Ott et al. 2000).

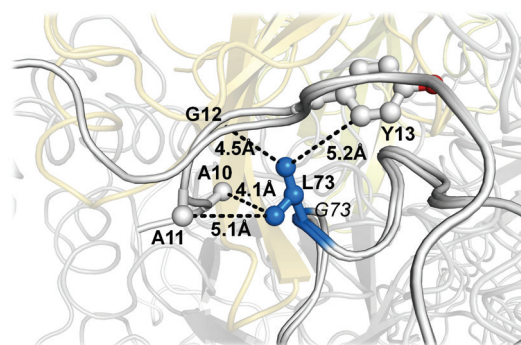


FIGURE 5. Comparison of residue 73 (blue) interactions in aligned large subunits (white ribbons) of *Synechococcus* (PDB ID 1RBL) and *Chlamydomonas* (PDB ID 1GK8). Distances (in Å) for contacts with N-terminus residues (represented with white ball and sticks; oxygen atom is red) that might be lost in the Section 2 mutants are denoted. Large subunit residues are numbered based on the *Chlamydomonas* sequence. *Synechococcus* residues are in bold whereas *Chlamydomonas* are in italics

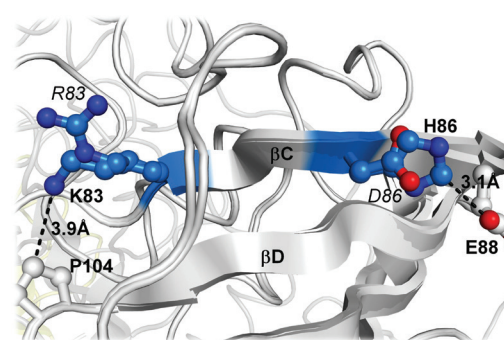


FIGURE 6. Differences in β C residues (light blue; nitrogen atoms are dark blue; oxygen atoms are red) between aligned large subunits (white) of *Synechococcus* (PDB ID 1RBL) in bold and *Chlamydomonas* (PDB ID 1GK8) in italics. Contacts with neighbouring residue (represented with white ball and sticks; oxygen atoms, red) that might be lost in the Section 2 mutants are measured (in Å). Large subunit residues are numbered based on the *Chlamydomonas* sequence

For Section 10 chimeras, as a consequence of amino acid substitution L459C, abolished van der Waals contact between the branched isobutyl side chains of Leu-459 on large subunit α -helix H and Leu-445 on α -helix G is

proposed (*Chlamydomonas* numbering for all residues) (Figure 7). In *Chlamydomonas* Rubisco, Cys-459 is disulfide bonded to Cys-449 whereby site-directed mutagenesis showed a noticeable structural instability to the holoenzyme upon disruption of this covalent bond (Marin-Navarro & Moreno 2006). Because the Section 10 chimeras have Gly-449 instead, the remaining unpaired Cys-459 still renders the enzyme susceptible to redox modulation which intensifies the proteolytic degradation of the enzyme (Marin-Navarro & Moreno 2006).

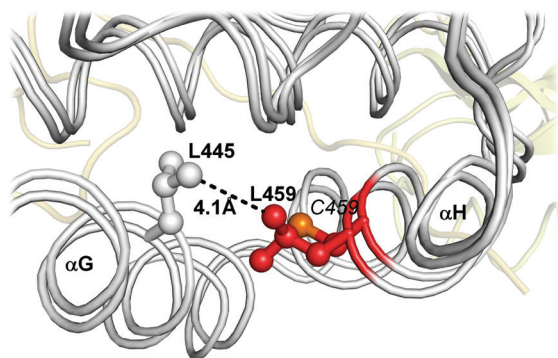


FIGURE 7. Comparison of residue 459 (red) in aligned large subunits (white ribbons) of *Synechococcus* (PDB ID 1RBL) (in bold) and *Chlamydomonas* (PDB ID 1GK8) (in italics). Contact with Leu-445 (white ball and sticks), which is denoted in Å, might be lost in the Section 10 mutants. Large subunit residues are numbered based on the *Chlamydomonas* sequence

One could attempt to deduce whether a particular mutant protein would successfully form holoenzyme or not based on mutational strain threshold by simply counting the number of amino acid changes since protein instability generally increases with mutations (Bloom et al. 2005). However, such an approach would miss the mark in this case because while Chimera $L^{1+2}S$, which has as many as 25 residue substitutions encompassing Sections 1 and 2, is capable of holoenzyme formation, Chimera $L^{1+10}S$, which has only 18 substitutions, could not produce assembled holoenzyme (Figure 3). Therefore, we infer that there must

be some thermodynamically favourable complementary bond or interaction that is engendered between Sections 1 and 2 when both sections assumed the *Chlamydomonas* identity. Indeed, upon closer inspection of the crystal structures, there is a network of interaction between *Chlamydomonas* large subunit residues Arg-21 (Lys-21 in *Synechococcus*) to Glu-51 (Asp-51 in *Synechococcus*) and Pro-50 (Ala-50 in *Synechococcus*), which is absent in the *Synechococcus* structure (Figure 8). These interactive forces may be responsible for stabilizing the folded large subunit.

It is noteworthy that expression of mutant Rubisco large subunits is almost indistinguishable from wild-type in the *E. coli* transformants despite extensive modifications to the protein sequence (Figure 2). This suggests that transcription and translation of the mutant large subunits are unaffected even though the large subunit has an autoregulatory role in translational arrest by binding to its own mRNA transcript (Cohen et al. 2006; Wostrikoff & Stern 2007).

In this study, sequence-based approaches were also undertaken to rule out Sections 1, 2, 5 and 10 from harbouring crucial differences in binding sites for the GroEL chaperone which could be key determinants for enabling bacterial Rubisco assembly in *E. coli* (Goloubinoff et al. 1989; Lin & Rye 2006; Saschenbrecker et al. 2007) while preventing assembly of eukaryotic Rubisco (Cloney et al. 1993; Koay et al. 2016). Scanning the primary structures of *Synechococcus* and *Chlamydomonas* RbcL identified putative GroEL substrate-protein binding motifs (Stan et al. 2006, 2004) in Sections 1, 2, 3, 4, 6 and 9 of the large subunit, with distinctions between the two species limited to Sections 1, 6 and 9, which correspond to *Synechococcus* residues 1-47, 248-297 and 398-447 (Supplementary File 1). Binding motifs should also be buried when the substrate protein is already in its properly-folded assembled native conformation to avoid unnecessary binding and refolding of it by the chaperone complex (Stan et al. 2006). With this additional criterion, we further narrowed down the possible important binding motifs to within Sections 6 and 9 (Table 1).

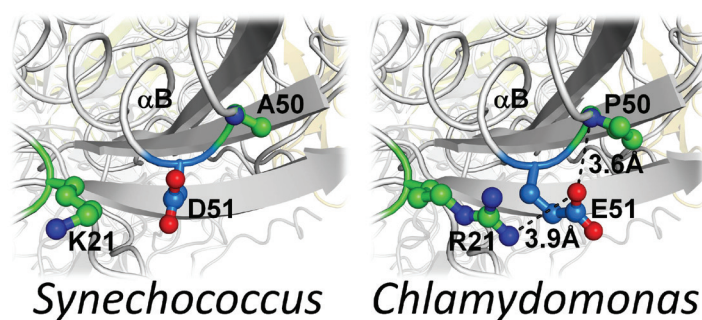


FIGURE 8. Comparison between *Synechococcus* (PDB ID 1RBL) and *Chlamydomonas* (PDB ID 1GK8) of critical Sections 1 (green; nitrogen atoms in blue) and 2 (blue; oxygen atoms in red) residues. These residues form a network of interactions in *Chlamydomonas*, with contact distances shown in Å

Another bioinformatics approach is to evaluate the RbcL sequence for seven-residue hydrophobic segments that closely match the hydrophobicity of the seven-residue GroES mobile loop with sequence GGIVLTG, which presents hydrophobic interactions with the GroEL apical domain (Chaudhuri & Gupta 2005). The GroES loop has a hydropathy index of 1.514 (Chaudhuri & Gupta 2005), thus hydrophobic patches with hydropathicity greater than 1.514 were evaluated (Table 2). From this analysis, differences in potential GroEL binding sites between *Synechococcus* and *Chlamydomonas* are found in Sections 2, 3, 4, 6, 7, 8 and 9 (Table 2). For Section 2 sites, they extend into Section 3, therefore, the demarcation of Sections 2 and 3 are less well-defined for these sites (Table 2). Hence, distinctions in definable possible GroEL-binding sites between the two species are confined to Sections 3, 4, 6, 7, 8 and 9, which correspond to *Synechococcus* residues 98-147, 148-197, 248-297, 298-347, 348-397 and 398-447.

From this study, protein assembly analysis of combinatorial mutations indicates that mutations in Sections 1, 2, 5 and 10 are destabilizing to the holoenzyme but there are possible complementary interactions between Sections 1 and 2, which could partially offset the destabilizing forces. A correlation between the present bioinformatics analyses and chimeric Rubisco holoenzyme formation in Koay et al. (2016) is observed, whereby key determinants for successful *Synechococcus*, but not *Chlamydomonas* Rubisco formation in *E. coli* can be discounted from Sections 1, 2, 5 and 10 of RbcL. The two bioinformatics methods (Chaudhuri & Gupta 2005; Stan et al. 2006) also concur and seemingly suggest that Sections 6 and 9 could be future targets for investigating disparities between *Chlamydomonas* and *Synechococcus* Rubisco formation in *E. coli*.

ACKNOWLEDGEMENTS

We are grateful to Dr. Oliver Martin Mueller-Cajar from Nanyang Technology University, Singapore for providing the *Synechococcus* PCC6301 Rubisco construct pTrcSynLS. We also thank Professor F. Robert Tabita, Ohio State University, for his generosity in providing us with anti-*Synechococcus* PCC6301 Rubisco antibodies. Special thanks is due to Professor Robert J. Spreitzer of University of Nebraska-Lincoln in providing us with anti-*Chlamydomonas* Rubisco antibodies and the *Chlamydomonas* small subunit construct pSS1-ITP. This work was supported by Universiti Tunku Abdul Rahman Research Fund (6200/L87 and 6200/L14).

REFERENCES

Aigner, H., Wilson, R.H., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U. & Hayer-Hartl, M. 2017. Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* 358: 1272-1278.

Andersson, I. & Backlund, A. 2008. Structure and function of Rubisco. *Plant Physiology and Biochemistry* 46: 275-291.

Bainbridge, G., Madgwick, P.J., Parmar, S., Mitchell, R., Paul, M., Pitts, J., Keys, A.J. & Parry, M.A.J. 1995. Engineering Rubisco to change its catalytic properties. *Journal of Experimental Botany* 46: 1269-1276.

Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C. & Arnold, F.H. 2005. Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Science of the United States of America* 102: 606-611.

Burger, R., Willensdorfer, M. & Nowak, M.A. 2006. Why are phenotypic mutation rates much higher than genotypic mutation rates? *Genetics* 172: 197-206.

Campbell, W.J. & Ogren, W.L. 1992. Light activation of Rubisco by Rubisco activase and thylakoid membranes. *Plant Cell Physiology* 33: 751-756.

Chaudhuri, T.K. & Gupta, P. 2005. Factors governing the substrate recognition by GroEL chaperone: A sequence correlation approach. *Cell Stress Chaperones* 10: 24-36.

Chen, Z. & Spreitzer, R.J. 1992. How various factors influence the CO₂/O₂ specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Photosynthesis Research* 31: 157-164.

Cloney, L.P., Bekkaoui, D.R. & Hemmingsen, S.M. 1993. Co-expression of plastid chaperonin genes and a synthetic plant Rubisco operon in *Escherichia coli*. *Plant Molecular Biology* 23: 1285-1290.

Cohen, I., Sapir, Y. & Shapira, M. 2006. A conserved mechanism controls translation of Rubisco large subunit in different photosynthetic organisms. *Plant Physiology* 141: 1089-1097.

Cordes, M.H. & Sauer, R.T. 1999. Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Science* 8: 318-325.

Curmi, P.M.G., Cascio, D., Sweet, R.M., Eisenberg, D. & Schreuder, H. 1992. Crystal structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase/oxygenase from tobacco refined at 20-Å resolution. *Journal of Biological Chemistry* 267: 16980-16989.

Ellis, R.J. 1979. The most abundant protein in the world. *Trends in Biochemical Science* 4: 241-244.

Esquivel, M.G., Genkov, T., Nogueira, A.S., Salvucci, M.E. & Spreitzer, R.J. 2013. Substitutions at the opening of the Rubisco central solvent channel affect holoenzyme stability and CO₂/O₂ specificity but not activation by Rubisco activase. *Photosynthesis Research* 118: 209-218.

Feller, U., Anders, I. & Mae, T. 2008. Rubiscolytics: Fate of Rubisco after its enzymatic function in a cell is terminated. *Journal of Experimental Botany* 59: 1615-1624.

Genkov, T., Du, Y.C. & Spreitzer, R.J. 2006. Small subunit cysteine-65 substitutions can suppress or induce alterations in the large-subunit catalytic efficiency and holoenzyme thermal stability of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics* 451: 167-174.

Genkov, T. & Spreitzer, R.J. 2009. Highly conserved small subunit residues influence Rubisco large subunit catalysis. *Journal of Biological Chemistry* 284: 30105-30112.

Genkov, T., Meyer, M., Griffiths, H. & Spreitzer, R.J. 2010. Functional hybrid Rubisco enzymes with plant small subunits and algal large subunits: Engineered rbcS cDNA for expression in *Chlamydomonas*. *Journal of Biological Chemistry* 285: 19833-19841.

Goloubinoff, P., Gatenby, A.A. & Lorimer, G.H. 1989. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* 337: 44-47.

- Gutteridge, S., Rhoades, D.F. & Herrmann, C. 1993. Site-specific mutations in a loop region of the C-terminal domain of the large subunit of ribulose biphosphate carboxylase/oxygenase that influence substrate partitioning. *Journal of Biological Chemistry* 268: 7818-7824.
- Jordan, D.B. & Ogren, W.L. 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* 291: 513-515.
- Kannappan, B. & Greedy, J.E. 2008. Redefinition of Rubisco carboxylase reaction reveals origin of water for hydration and new roles for active-site residues. *Journal of the American Chemical Society* 130: 15063-15080.
- Knight, S., Andersson, I. & Brändén, C.I. 1990. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. *Journal of Molecular Biology* 215: 113-160.
- Koay, T.W., Wong, H.L. & Lim, B.H. 2016. Engineering of chimeric eukaryotic/bacterial Rubisco large subunits in *Escherichia coli*. *Genes & Genetic Systems* 91: 139-150.
- Kumar, V., Punetha, A., Sundar, D. & Chaudhuri, T.K. 2012. In silico engineering of aggregation-prone recombinant proteins for substrate recognition by the chaperonin GroEL. *BMC Genomics* 13: S22.
- Kyte, J. & Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105-132.
- Laing, W.A., Ogren, W.L. & Hageman, R.H. 1974. Regulation of soybean net photosynthetic CO₂ fixation by the interaction of CO₂, O₂, and ribulose 1,5-diphosphate carboxylase. *Plant Physiology* 54: 678-685.
- Lin, Z. & Rye, H.S. 2006. GroEL-mediated protein folding: Making the impossible, possible. *Critical Reviews in Biochemistry and Molecular Biology* 41: 211-239.
- Long, S.P., Zhu, X-G., Naidu, S.L. & Ort, D.R. 2006. Can improvement in photosynthesis increase crop yields? *Plant, Cell and Environment* 29: 315-330.
- Marin-Navarro, J. & Moreno, J. 2006. Cysteines 449 and 459 modulate the reduction-oxidation conformational changes of ribulose 1,5-bisphosphate carboxylase/oxygenase and the translocation of the enzyme to membranes during stress. *Plant, Cell and Environment* 29: 898-908.
- Meyer, M.T., Genkov, T., Skepper, J.N., Jouhet, J., Mitchell, M.C., Spreitzer, R.J. & Griffiths, H. 2012. Rubisco small-subunit α -helices control pyrenoid formation in *Chlamydomonas*. *Proceedings of the National Academy of Science of the United States of America* 109: 19474-19479.
- Mueller-Cajar, O. & Whitney, S.M. 2008. Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. *Biochemical Journal* 414: 205-214.
- Ogren, W.L. 1984. Photorespiration: Pathways, regulation, and modification. *Annual Review of Plant Physiology and Plant Molecular Biology* 35: 415-442.
- Ott, C.M., Smith, B.D., Portis, A.R. & Spreitzer, R.J. 2000. Activase region on chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase. *Journal of Biological Chemistry* 275: 26241-26244.
- Pace, C.N. 2001. Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* 40: 310-313.
- Pakula, A.A. & Sauer, R.T. 1990. Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface. *Nature* 344: 363-364.
- Parikh, M.R., Greene, D.A., Woods, K.K. & Matsumura, I. 2006. Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E. coli*. *Protein Engineering, Design and Selection* 19: 113-119.
- Parry, M.A.J., Madgwick, P.J., Carvalho, J.F.C. & Andralojc, P.J. 2007. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *Journal of Agricultural Science* 145: 31-43.
- Peterhansel, C., Niessen, M. & Kebeish, R.M. 2008. Metabolic engineering towards the enhancement of photosynthesis. *Photochemistry and Photobiology* 84: 1317-1323.
- Rasineni, G.K., Loh, P.C. & Lim, B.H. 2017. Characterization of *Chlamydomonas* ribulose-1,5-bisphosphate carboxylase/oxygenase variants mutated at residues that are post-translationally modified. *Biochimica et Biophysica Acta* 1861(2): 79-85.
- Saschenbrecker, S., Bracher, A., Rao, K.V., Rao, B.V., Hartl, F.U. & Hayer-Hartl, M. 2007. Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* 129: 1189-1200.
- Smith, S.A. & Tabita, F.R. 2003. Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. *Journal of Molecular Biology* 331: 557-569.
- Spreitzer, R.J., Thow, G. & Zhu, G. 1995. Pseudoreversion substitution at large-subunit residue 54 influences the CO₂/O₂ specificity of chloroplast ribulose biphosphate carboxylase/oxygenase. *Plant Physiology* 109: 681-685.
- Stan, G., Brooks, B.R., Lorimer, G.H. & Thirumalai, D. 2006. Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. *Proceedings of the National Academy of Science of the United States of America* 103: 4433-4438.
- Stan, G., Brooks, B.R., Lorimer, G.H. & Thirumalai, D. 2004. Identifying natural substrates for chaperonins using a sequence-based approach. *Protein Science* 14: 193-201.
- Tabita, F.R., Hanson, T.E., Satagopan, S., Witte, B.H. & Kreeel, N.E. 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philosophical Transactions of the Royal Society B* 363: 2629-2640.
- Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J. & Chan, S. 2007a. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiology and Molecular Biology Reviews* 71(4): 576-599.
- Tabita, F.R., Satagopan, S., Hanson, T.E., Kreeel, N.E. & Scott, S.S. 2007b. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of Experimental Botany* 59: 1515-1524.
- Takano, K., Yamagata, Y. & Yutani, K. 2001. Contribution of polar groups in the interior of a protein to the conformational stability. *Biochemistry* 40: 4853-4858.
- Taylor, T.C., Backlund, A., Bjorhall, K., Spreitzer, R.J. & Andersson, I. 2001. First crystal structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*. *Journal of Biological Chemistry* 276: 48159-48164.
- Tcherkez, G.G.B., Farquhar, G.D. & Andrews, T.J. 2006. Despite slow catalysis and confused substrate specificity, all ribulose biphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Science of the United States of America* 103: 7246-7251.

- Whitney, S.M., Baldet, P., Hudson, G.S. & Andrews, T.J. 2001. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant Journal* 26: 535-547.
- Whitney, S.M., Houtz, R.L. & Alonso, H. 2011. Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiology* 155: 27-35.
- Wilson, R.H., Martin-Avila, E., Conlan, C. & Whitney, S.M. 2017. An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO₂-fixation kinetics. *Journal of Biological Chemistry* 293: 18-27.
- Wostrikoff, K. & Stern, D. 2007. Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. *Proceedings of the National Academy of Science of the United States of America* 104: 6466-6471.
- Zhu, X.G., Portis, A.R. & Long, S.P. 2004. Would transformation of C3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant, Cell and Environment* 27: 155-165.
- Yee Hung Yeap, Teng Wei Koay & Boon Hoe Lim*
Department of Chemical Science
Universiti Tunku Abdul Rahman
31900 Kampar, Perak Darul Ridzuan
Malaysia
- Hann Ling Wong
Department of Biological Science
Universiti Tunku Abdul Rahman
31900 Kampar, Perak Darul Ridzuan
Malaysia

*Corresponding author; email: bhlm@utar.edu.my

Received: 14 Mac 2018

Accepted: 4 June 2018

Supplementary file 1: GroEL substrate-protein binding motif (SPBM) in *Synechococcus elongatus* PCC6301 RbcL and *Chlamydomonas reinhardtii* RbcL and their solvent accessibilities

<i>Synechococcus</i> RbcL			Solvent accessibility ^b
RbcL_section	Residue range ^a	Sequence	
1	31-35	TDLLA	
	51-55	GAAlA	
2	92-96	NSYFA	
	93-97	SYFAF	
3	138-142	PVALV	
4	183-187	GRAVY	
6	284-288	NGVLL	
9	417-421	NRVAL	
10	450-454	PELAA	
<i>Chlamydomonas</i> RbcL			Solvent accessibility ^b
RbcL_section	Residue range ^a	Sequence	
1	27-31	PDYVV	
	34-38	TDILA	
2	54-58	GAAVA	
	95-99	NQYIA	
	96-100	QYIAY	
3	141-145	PPAYV	
4	186-190	GRAVY	
	278-282	TSLAI	
6	279-283	SLAIY	
	287-291	NGLLL	
9	415-419	PGAAA	
	420-424	NRVAL	
10	453-457	PELAA	

^aTo align *Synechococcus* RbcL to *Chlamydomonas* RbcL, add 3 to the *Synechococcus* residue numbers.

^bSolvent accessibilities for each residue in the SPBM were predicted using Swiss-Pdb Viewer 4.1.0 at the parameter of $0\% \leq S < 5\%$, $5\% \leq S < 10\%$, $10\% \leq S < 15\%$, $15\% \leq S < 20\%$, $20\% \leq S < 25\%$, $25\% \leq S < 30\%$, $30\% \leq S < 35\%$, $35\% \leq S < 40\%$, $40\% \leq S < 45\%$, and $45\% \leq S < 50\%$, which identified residues from the interior to the surface, based on a single folded RbcL subunit. These are represented by the following color scale:



SUPPLEMENTARY FILE 2: Average hydropathy index of seven-residue segments
on *Chlamydomonas* and *Synechococcus* RbcL

^a Residue range	<i>Chlamydomonas</i> RbcL		<i>Synechococcus</i> RbcL	
	Sequence	Hydropathy index	Sequence	Hydropathy index
1-7	MVPQTET	-0.5571	MPKTQSA	-0.9714
2-8	VPQTETK	-1.3857	^b _____	^b _____
3-9	PQTETKA	-1.7286	PKTQSAA	-0.9857
4-10	QTETKAG	-1.5571	KTQSAAG	-0.8143
5-11	TETKAGA	-0.8000	^b _____	^b _____
6-12	ETKAGAG	-0.7571	^b _____	^b _____
7-13	TKAGAGF	0.1429	TQSAAGY	-0.4429
8-14	KAGAGFK	-0.3143	QSAAGYK	-0.9000
9-15	AGAGFKA	0.5000	SAAGYKA	-0.1429
10-16	GAGFKAG	0.1857	AAGYKAG	-0.0857
11-17	AGFKAGV	0.8429	AGYKAGV	0.2571
12-18	GFKAGVK	0.0286	GYKAGVK	-0.5571
13-19	FKAGVKD	-0.4143	YKAGVKD	-1.0000
14-20	KAGVKDY	-1.0000	KAGVKDY	-1.0000
15-21	AGVKDYR	-1.0857	AGVKDYK	-1.0000
16-22	GVKDYRL	-0.8000	GVKDYKL	-0.7143
17-23	VKDYRLT	-0.8429	VKDYKLT	-0.7571
18-24	KDYRLTY	-1.6286	KDYKLTY	-1.5429
19-25	DYRLTYT	-1.2571	DYKLTYY	-1.1714
20-26	YRLTYYT	-0.8571	YKLTYYT	-0.7714
21-27	RLTYTTP	-0.9000	KLTYYTP	-0.8143
22-28	LTYYTPD	-0.7571	LTYYTPD	-0.7571
23-29	TYTTPDY	-1.4857	TYTTPDY	-1.4857
24-30	YYTPDYV	-0.7857	YYTPDYT	-1.4857
25-31	YTPDYVV	0.0000	YTPDYTP	-1.5286
26-32	TPDYVVR	-0.4571	TPDYTPK	-1.9000
27-33	PDYVVRD	-0.8571	PDYTPKD	-2.3000
28-34	DYVVRDT	-0.7286	DYTPKDT	-2.1714
29-35	YVVRDTD	-0.7286	YTPKDTD	-2.1714
30-36	VVRDTD	0.1000	TPKDTD	-1.4429
31-37	VRDTDIL	0.0429	PKDTDLL	-0.8000
32-38	RDTDILA	-0.3000	KDTDLLA	-0.3143
33-39	DTDILAA	0.6000	DTDLLAA	0.5000
34-40	TDILAAF	1.5000	TDLLAAF	1.4000
35-41	DILAAFR	0.9571	DLAAFR	0.8571
36-42	ILAAFRM	1.7286	LLAAFRF	1.7571
37-43	LAAFRMT	0.9857	LAAFRFS	1.1000
38-44	AAFRMTP	0.2143	AAFRFSP	0.3286
39-45	AFRMTPQ	-0.5429	AFRFSPQ	-0.4286
40-46	FRMTPQP	-1.0286	FRFSPQP	-0.9143
41-47	RMTPQPG	-1.4857	RFSPQPG	-1.3714
42-48	MTPQPGV	-0.2429	FSPQPGV	-0.1286
43-49	TPQPGVP	-0.7429	SPQPGVP	-0.7571
44-50	PQPGVPP	-0.8714	PQPGVPA	-0.3857
45-51	QPGVPPE	-1.1429	QPGVPAD	-0.6571
46-52	PGVPPEE	-1.1429	PGVPADE	-0.6571
47-53	GVPPEEC	-0.5571	GVPADEA	-0.1714
48-54	VPPEECG	-0.5571	VPADEAG	-0.1714
49-55	PPEECGA	-0.9000	PADEAGA	-0.5143

Continued SUPPLEMENTARY FILE 2

50-56	PEECGAA	-0.4143	ADEAGAA	-0.0286
51-57	EECGAAV	0.4143	DEAGAAI	0.3571
52-58	ECGAAVA	1.1714	EAGAAIA	1.1143
53-59	CGAAVAA	1.9286	AGAAIAA	1.8714
54-60	GAAVAAE	1.0714	GAAIAAE	1.1143
55-61	AAVA AES	1.0143	AAIAAES	1.0571
56-62	AVAAESS	0.6429	AIAAESS	0.6857
57-63	VAAESST	0.2857	IAAESST	0.3286
58-64	AAESSTG	-0.3714	AAESSTG	-0.3714
59-65	AESSTGT	-0.7286	AESSTGT	-0.7286
60-66	ESSTGTW	-1.1143	ESSTGTW	-1.1143
61-67	SSTGTWT	-0.7143	SSTGTWT	-0.7143
62-68	STGTWTT	-0.7000	STGTWTT	-0.7000
63-69	TGTWTTV	0.0143	TGTWTTV	0.0143
64-70	GTWTTVW	-0.0143	GTWTTVW	-0.0143
65-71	TWTTVWT	-0.0571	TWTTVWT	-0.0571
66-72	WTTVWTD	-0.4571	WTTVWTD	-0.4571
67-73	TTVWTDG	-0.3857	TTVWTDL	0.2143
68-74	TVWTDGL	0.2571	TVWTDLL	0.8571
69-75	VWTDGLT	0.2571	VWTDLLT	0.8571
70-76	WTDGLTS	-0.4571	WTDLLTD	-0.2429
71-77	TDGLTSL	0.2143	TDLLTDM	0.1571
72-78	DGLTSLD	-0.1857	DLLTDMD	-0.2429
73-79	GLTSLDR	-0.3286	LLTDMDR	-0.3857
74-80	LTS LDRY	-0.4571	LTDMDRY	-1.1143
75-81	TSLDRYK	-1.5571	TDMDRYK	-2.2143
76-82	SLDRYKG	-1.5143	DMDRYKG	-2.1714
77-83	LDRYKGR	-2.0429	MDRYK GK	-2.2286
78-84	DRYKGR C	-2.2286	DRYK GK C	-2.1429
79-85	RYKGR CY	-1.9143	RYK GK CY	-1.8286
80-86	YKGR CYD	-1.7714	YK GK CYH	-1.6429
81-87	KGR CYDI	-0.9429	KG GK CYHI	-0.8143
82-88	GRCYDIE	-0.8857	GK GK CYHIE	-0.7571
83-89	RCYDIEP	-1.0571	KCYHIEP	-0.9286
84-90	CYDIEPV	0.1857	CYHIEPV	0.2286
85-91	YDIEPVP	-0.4000	YHIEPVQ	-0.6286
86-92	DIEPVP G	-0.2714	HIEPVQ G	-0.5000
87-93	IEPVPGE	-0.2714	IEPVQGE	-0.5429
88-94	EPVPGED	-1.4143	EPVQGEE	-1.6857
89-95	PVPGEDN	-1.4143	PVQGEEN	-1.6857
90-96	VPGEDNQ	-1.6857	VQGEENS	-1.5714
91-97	PGEDNQY	-2.4714	QGEENSY	-2.3571
92-98	GEDNQYI	-1.6000	GEENSYF	-1.4571
93-99	EDNQYIA	-1.2857	EENSYFA	-1.1429
94-100	DNQYIAY	-0.9714	ENSYFAF	-0.2429
95-101	NQYIAYV	0.1286	NSYFAFI	0.9000
96-102	QYIAYVA	0.8857	SYFAFIA	1.6571
97-103	YIAYVAY	1.2000	YFAFIAY	1.5857
98-104	IAYVAYP	1.1571	FAFIAYP	1.5429
99-105	AYVAYPI	1.1571	AFIAYPL	1.6857
100-106	YVAYPID	0.4000	FIAYPLD	0.9286
101-107	VAYPIDL	1.1286	IAYPLDL	1.0714
102-108	AYPIDLF	0.9286	AYPLDLF	0.8286

Continued SUPPLEMENTARY FILE 2

103-109	YPIDLFE	0.1714	YPLDLFE	0.0714
104-110	PIDLFE	-0.1429	PLDLFEE	-0.2429
105-111	IDLFEEG	0.0286	LDLFEEG	-0.0714
106-112	DLFEEGS	-0.7286	DLFEEGS	-0.7286
107-113	LFEEGSV	0.3714	LFEEGSV	0.3714
108-114	FEEGSVT	-0.2714	FEEGSVT	-0.2714
109-115	EEGSVTN	-1.1714	EEGSVTN	-1.1714
110-116	EGSVTNM	-0.4000	EGSVTNI	-0.0286
111-117	GSVTNMF	0.5000	GSVTNIL	1.0143
112-118	SVTNMFT	0.4571	SVTNILT	0.9714
113-119	VTNMFTS	0.4571	VTNILTS	0.9714
114-120	TNMFTSI	0.5000	TNILTSI	1.0143
115-121	NMFTSIV	1.2000	NILTSIV	1.7143
116-122	MFTSIVG	1.6429	ILTSIVG	2.1571
117-123	FTSIVGN	0.8714	LTSIVGN	1.0143
118-124	TSIVGNV	1.0714	TSIVGNV	1.0714
119-125	SIVGNVF	1.5714	SIVGNVF	1.5714
120-126	IVGNVFG	1.6286	IVGNVFG	1.6286
121-127	VGNVFGF	1.3857	VGNVFGF	1.3857
122-128	GNVFGFK	0.2286	GNVFGFK	0.2286
123-129	NVFGFKA	0.5429	NVFGFKA	0.5429
124-130	VFGFKAL	1.5857	VFGFKAI	1.6857
125-131	FGFKALR	0.3429	FGFKAIR	0.4429
126-132	GFKALRA	0.2000	GFKAIRS	-0.0714
127-133	FKALRAL	0.8000	FKAIRSL	0.5286
128-134	KALRALR	-0.2429	KAIRSLR	-0.5143
129-135	ALRALRL	0.8571	AIRSLRL	0.5857
130-136	LRALRLE	0.1000	IRSLRLE	-0.1714
131-137	RALRLED	-0.9429	RSLRLED	-1.3143
132-138	ALRLEDL	0.2429	SLRLEDI	-0.0286
133-139	LRLEDLR	-0.6571	LRLEDIR	-0.5571
134-140	RLEDLRI	-0.5571	RLEDIRF	-0.7000
135-141	LEDLRIP	-0.1429	LEDIRFP	-0.2857
136-142	EDLRIPP	-0.9143	EDIRFPV	-0.2286
137-143	DLRIPPA	-0.1571	DIRFPVA	0.5286
138-144	LRIPPAY	0.1571	IRFPVAL	1.5714
139-145	RIPPAYV	0.2143	RFPVALV	1.5286
140-146	IPPAYVK	0.3000	FPVALVK	1.6143
141-147	PPAYVKT	-0.4429	PVALVKT	1.1143
142-148	PAYVKTF	0.1857	VALVKTF	1.7429
143-149	AYVKTFV	1.0143	ALVKTFQ	0.6429
144-150	YVKTFVG	0.7000	LVKTFQG	0.3286
145-151	VKTFVGP	0.6571	VKTFQGP	-0.4429
146-152	KTFVGPP	-0.1714	KTFQGP	-1.2714
147-153	TFVGPPH	-0.0714	TFQGP	-1.1714
148-154	FVGPPHG	-0.0286	FQGP	-1.1286
149-155	VGPPHGI	0.2143	QGPPHGI	-0.8857
150-156	GPPHGIQ	-0.8857	GPPHGIQ	-0.8857
151-157	PPHGIQV	-0.2286	PPHGIQV	-0.2286
152-158	PHGIQVE	-0.5000	PHGIQVE	-0.5000
153-159	HGIQVER	-0.9143	HGIQVER	-0.9143
154-160	GIQVERD	-0.9571	GIQVERD	-0.9571
155-161	IQVERDK	-1.4571	IQVERDL	-0.3571

Continued SUPPLEMENTARY FILE 2

156-162	QVERDKL	-1.5571	QVERDLL	-0.4571
157-163	VERDKLN	-1.5571	VERDLLN	-0.4571
158-164	ERDKLNK	-2.7143	ERDLLNK	-1.6143
159-165	RDKLNKY	-2.4000	RDLLNKY	-1.3000
160-166	DKLNKYG	-1.8143	DLLNKYG	-0.7143
161-167	KLNKYGR	-1.9571	LLNKYGR	-0.8571
162-168	LNKYGRG	-1.4571	LNKYGRP	-1.6286
163-169	NKYGRGL	-1.4571	NKYGRPM	-1.9000
164-170	KYGRGLL	-0.4143	KYGRPML	-0.8571
165-171	YGRGLLG	0.0857	YGRPMLG	-0.3571
166-172	GRGLLGC	0.6286	GRPMLGC	0.1857
167-173	RGLLGCT	0.5857	RPMLGCT	0.1429
168-174	GLLGCTI	1.8714	PMLGCTI	1.4286
169-175	LLGCTIK	1.3714	MLGCTIK	1.1000
170-176	LGCTIKP	0.6000	LGCTIKP	0.6000
171-177	GCTIKPK	-0.5000	GCTIKPK	-0.5000
172-178	CTIKPKL	0.1000	CTIKPKL	0.1000
173-179	TIKPKLG	-0.3143	TIKPKLG	-0.3143
174-180	IKPKLGL	0.3286	IKPKLGL	0.3286
175-181	KPKLGLS	-0.4286	KPKLGLS	-0.4286
176-182	PKLGLSA	0.3857	PKLGLSA	0.3857
177-183	KLGLSAK	0.0571	KLGLSAK	0.0571
178-184	LGLSAKN	0.1143	LGLSAKN	0.1143
179-185	GLSAKNY	-0.6143	GLSAKNY	-0.6143
180-186	LSAKNYG	-0.6143	LSAKNYG	-0.6143
181-187	SAKNYGR	-1.8000	SAKNYGR	-1.8000
182-188	AKNYGRA	-1.4286	AKNYGRA	-1.4286
183-189	KNYGRAV	-1.0857	KNYGRAV	-1.0857
184-190	NYGRAVY	-0.7143	NYGRAVY	-0.7143
185-191	YGRAVYE	-0.7143	YGRAVYE	-0.7143
186-192	GRAVYEC	-0.1714	GRAVYEC	-0.1714
187-193	RAVYECL	0.4286	RAVYECL	0.4286
188-194	AVYECLR	0.4286	AVYECLR	0.4286
189-195	VYECLRG	0.1143	VYECLRG	0.1143
190-196	YECLRGG	-0.5429	YECLRGG	-0.5429
191-197	ECLRGGL	0.1857	ECLRGGL	0.1857
192-198	CLRGGLD	0.1857	CLRGGLD	0.1857
193-199	LRGGLDF	0.2286	LRGGLDF	0.2286
194-200	RGGLDFT	-0.4143	RGGLDFT	-0.4143
195-201	GGLDFTK	-0.3286	GGLDFTK	-0.3286
196-202	GLDFTKD	-0.7714	GLDFTKD	-0.7714
197-203	LDFTKDD	-1.2143	LDFTKDD	-1.2143
198-204	DFTKDDE	-2.2571	DFTKDDE	-2.2571
199-205	FTKDDEN	-2.2571	FTKDDEN	-2.2571
200-206	TKDDENV	-2.0571	TKDDENI	-2.0143
201-207	KDDENVN	-2.4571	KDDENIN	-2.4143
202-208	DDENVNS	-2.0143	DDENINS	-1.9714
203-209	DENVNSQ	-2.0143	DENINSQ	-1.9714
204-210	ENVNSQP	-1.7429	ENINSQP	-1.7000
205-211	NVNSQPF	-0.8429	NINSQPF	-0.8000
206-212	VNSQPFM	-0.0714	INSQPFQ	-0.8000
207-213	NSQPFMR	-1.3143	NSQPFQR	-2.0857
208-214	SQPFMRW	-0.9429	SQPFQRW	-1.7143

Continued SUPPLEMENTARY FILE 2

209-215	QPFMRWR	-1.4714	QPFQRWR	-2.2429
210-216	PFMRWRD	-1.4714	PFQRWRD	-2.2429
211-217	FMRWRDR	-1.8857	FQRWRDR	-2.6571
212-218	MRWRDRF	-1.8857	QRWRDRF	-2.6571
213-219	RWRDRFL	-1.6143	RWRDRFL	-1.6143
214-220	WRDRFLF	-0.5714	WRDRFLF	-0.5714
215-221	RDRFLFV	0.1571	RDRFLFV	0.1571
216-222	DRFLFVA	1.0571	DRFLFVA	1.0571
217-223	RFLFVAE	1.0571	RFLFVAD	1.0571
218-224	FLFVAEA	1.9571	FLFVADA	1.9571
219-225	LFVAEAI	2.2000	LFVADAI	2.2000
220-226	FVAEAIY	1.4714	FVADAIH	1.2000
221-227	VAEAIYK	0.5143	VADAIHK	0.2429
222-228	AEAIYKA	0.1714	ADAIHKS	-0.4714
223-229	EAIYKAQ	-0.5857	DAIHKSQ	-1.2286
224-230	AIYKAQA	0.1714	AIHKSQA	-0.4714
225-231	IYKAQAE	-0.5857	IHKSQAE	-1.2286
226-232	YKAQAET	-1.3286	HKSQAET	-1.9714
227-233	KAQAETG	-1.2000	KSQAETG	-1.5714
228-234	AQAETGE	-1.1429	SQAETGE	-1.5143
229-235	QAETGEV	-0.8000	QAETGEI	-0.7571
230-236	AETGEVK	-0.8571	AETGEIK	-0.8143
231-237	ETGEVKG	-1.1714	ETGEIKG	-1.1286
232-238	TGEVKGH	-1.1286	TGEIKGH	-1.0857
233-239	GEVKGHY	-1.2143	GEIKGHY	-1.1714
234-240	EVKGHYL	-0.6143	EIKGHYL	-0.5714
235-241	VKGHYLN	-0.6143	IKGHYLN	-0.5714
236-242	KGHYLN	-0.9571	KGHYLVN	-0.6143
237-243	GHYLNAT	-0.5000	GHYLVNT	-0.1571
238-244	HYLNATA	-0.1857	HYLVNTA	0.1571
239-245	YLNATAG	0.2143	YLVNTAP	0.3857
240-246	LNATAGT	0.3000	LVNTAPT	0.4714
241-247	NATAGTC	0.1143	NVTAPTC	0.2857
242-248	ATAGTCE	0.1143	VTAPTCE	0.2857
243-249	TAGTCEE	-0.6429	TAPTCEE	-0.8143
244-250	AGTCEEM	-0.2714	APTCEEM	-0.4429
245-251	GTCEEMM	-0.2571	PTCEEMM	-0.4286
246-252	TCEEMMK	-0.7571	TCEEMMK	-0.7571
247-253	CEEMMKR	-1.3000	CEEMMKR	-1.3000
248-254	EEMMKRA	-1.4000	EEMMKRA	-1.4000
249-255	EMMKRAV	-0.3000	EMMKRAE	-1.4000
250-256	MMKRAVC	0.5571	MMKRAEF	-0.5000
251-257	MKRAVCA	0.5429	MKRAEFA	-0.5143
252-258	KRAVCAK	-0.2857	KRAEFAK	-1.3429
253-259	RAVCAKE	-0.2286	RAEFAKE	-1.2857
254-260	AVCAKEL	0.9571	AEFAKEL	-0.1000
255-261	VCAKELG	0.6429	EFAKELG	-0.4143
256-262	CAKELGV	0.6429	FAKELGM	0.3571
257-263	AKELGVP	0.0571	AKELGMP	-0.2714
258-264	KELGVPI	0.4429	KELGMPI	0.1143
259-265	ELGVPII	1.6429	ELGMPII	1.3143
260-266	LGVPIIM	2.4143	LGMPIIM	2.0857
261-267	GVPIIMH	1.4143	GMPIIMH	1.0857

Continued SUPPLEMENTARY FILE 2

262-268	VPIIMHD	0.9714	MPIIMHD	0.6429
263-269	PIIMHDY	0.1857	PIIMHDF	0.7714
264-270	IIMHDYL	0.9571	IIMHDFL	1.5429
265-271	IMHDYLT	0.2143	IMHDFLT	0.8000
266-272	MHDYLTG	-0.4857	MHDFLTA	0.4143
267-273	HDYLTGG	-0.8143	HDFLTAG	0.0857
268-274	DYLTGGF	0.0429	DFLTAGF	0.9429
269-275	YLTGGFT	0.4429	FLTAGFT	1.3429
270-276	LTGGFTA	0.8857	LTAGFTA	1.2000
271-277	TGGFTAN	-0.1571	TAGFTAN	0.1571
272-278	GGFTANT	-0.1571	AGFTANT	0.1571
273-279	GFTANTS	-0.2143	GFTANTT	-0.2000
274-280	FTANTSL	0.3857	FTANTTL	0.4000
275-281	TANTSLA	0.2429	TANTTLA	0.2571
276-282	ANTSLAI	0.9857	ANTTLAK	-0.2000
277-283	NTSLAIY	0.5429	NTTLAKW	-0.5857
278-284	TSLAIYC	1.4000	TTLAKWC	0.2714
279-285	SLAIYCR	0.8571	TLAKWCR	-0.2714
280-286	LAIYCRD	0.4714	LAKWCRD	-0.6714
281-287	AIYCRDN	-0.5714	AKWCRDN	-1.7143
282-288	IYCRDNG	-0.8857	KWCRDNG	-2.0286
283-289	YCRDNGL	-0.9857	WCRDNGV	-0.8714
284-290	CRDNGLL	-0.2571	CRDNGVL	-0.2000
285-291	RDNGLLL	-0.0714	RDNGVLL	-0.0143
286-292	DNGLLLH	0.1143	DNGVLLH	0.1714
287-293	NGLLLHI	1.2571	NGVLLHI	1.3143
288-294	GLLLHIH	1.3000	GVLLHIH	1.3571
289-295	LLLHIHR	0.7143	VLLHIHR	0.7714
290-296	LLHIHRA	0.4286	LLHIHRA	0.4286
291-297	LHIHRAM	0.1571	LHIHRAM	0.1571
292-298	HIHRAMH	-0.8429	HIHRAMH	-0.8429
293-299	IHRAMHA	-0.1286	IHRAMHA	-0.1286
294-300	HRAMHAV	-0.1714	HRAMHAV	-0.1714
295-301	RAMHAVI	0.9286	RAMHAVI	0.9286
296-302	AMHAVID	1.0714	AMHAVID	1.0714
297-303	MHAVIDR	0.1714	MHAVIDR	0.1714
298-304	HAVIDRQ	-0.6000	HAVIDRQ	-0.6000
299-305	AVIDRQR	-0.7857	AVIDRQR	-0.7857
300-306	VIDRQRN	-1.5429	VIDRQRN	-1.5429
301-307	IDRQRNH	-2.6000	IDRQRNH	-2.6000
302-308	DRQRNHG	-3.3000	DRQRNHG	-3.3000
303-309	RQRNHGI	-2.1571	RQRNHGI	-2.1571
304-310	QRNHGIH	-1.9714	QRNHGIH	-1.9714
305-311	RNHGIHF	-1.0714	RNHGIHF	-1.0714
306-312	NHGIHFR	-1.0714	NHGIHFR	-1.0714
307-313	HGIHFRV	0.0286	HGIHFRV	0.0286
308-314	GIHFRVL	1.0286	GIHFRVL	1.0286
309-315	IHFRVLA	1.3429	IHFRVLA	1.3429
310-316	HFRVLAK	0.1429	HFRVLAK	0.1429
311-317	FRVLAKA	0.8571	FRVLAKC	0.9571
312-318	RVLAKAL	1.0000	RVLAKCL	1.1000
313-319	VLAKALR	1.0000	VLAKCLR	1.1000
314-320	LAKALRM	0.6714	LAKCLRL	1.0429

Continued SUPPLEMENTARY FILE 2

315-321	AKALRMS	0.0143	AKCLRLS	0.3857
316-322	KALRMSG	-0.3000	KCLRLSG	0.0714
317-323	ALRMSGG	0.2000	CLRLSGG	0.5714
318-324	LRMSGGD	-0.5571	LRLSGGD	-0.2857
319-325	RMSGGDH	-1.5571	RLSGGDH	-1.2857
320-326	MSGGDHL	-0.3714	LSGGDHL	-0.1000
321-327	SGGDHLH	-1.1000	SGGDHLH	-1.1000
322-328	GGDHLHS	-1.1000	GGDHLHS	-1.1000
323-329	GDHLHSG	-1.1000	GDHLHSG	-1.1000
324-330	DHLHSGT	-1.1429	DHLHSGT	-1.1429
325-331	HLHSGTV	-0.0429	HLHSGTV	-0.0429
326-332	LHSGTVV	1.0143	LHSGTVV	1.0143
327-333	HSGTVVG	0.4143	HSGTVVG	0.4143
328-334	SGTVVGK	0.3143	SGTVVGK	0.3143
329-335	GTVVGKL	0.9714	GTVVGKL	0.9714
330-336	TVVGKLE	0.5286	TVVGKLE	0.5286
331-337	VVGKLEG	0.5714	VVGKLEG	0.5714
332-338	VGKLEGE	-0.5286	VGKLEGD	-0.5286
333-339	GKLEGER	-1.7714	GKLEGDK	-1.6857
334-340	KLEGERE	-2.2143	KLEGDKA	-1.3714
335-341	LEGEREV	-1.0571	LEGDKAS	-0.9286
336-342	EGEREVT	-1.7000	EGDKAST	-1.5714
337-343	GEREVTL	-0.6571	GDKASTL	-0.5286
338-344	EREVTLG	-0.6571	DKASTLG	-0.5286
339-345	REVTLGF	0.2429	KASTLGF	0.3714
340-346	EVTLGFV	1.4857	ASTLGFV	1.5286
341-347	VTLGFVD	1.4857	STLGFVD	0.7714
342-348	TLGFVDL	1.4286	TLGFVDL	1.4286
343-349	LGFVDLM	1.8000	LGFVDLM	1.8000
344-350	GFVDLMR	0.6143	GFVDLMR	0.6143
345-351	FVDLMRD	0.1714	FVDLMRE	0.1714
346-352	VDLMRDD	-0.7286	VDLMRED	-0.7286
347-353	DLMRDDY	-1.5143	DLMREDH	-1.7857
348-354	LMRDDYV	-0.4143	LMREDHI	-0.6429
349-355	MRDDYVE	-1.4571	MREDHIE	-1.6857
350-356	RDDYVEK	-2.2857	REDHIEA	-1.7000
351-357	DDYVEKD	-2.1429	EDHIEAD	-1.5571
352-358	DYVEKDR	-2.2857	DHIEADR	-1.7000
353-359	YVEKDRS	-1.9000	HIEADRS	-1.3143
354-360	VEKDRSR	-2.3571	IEADRSR	-1.5000
355-361	EKDRSRG	-3.0143	EADRSRG	-2.2000
356-362	KDRSRGI	-1.8714	ADRSRGV	-1.1000
357-363	DRSRGIY	-1.5000	DRSRGVF	-0.9571
358-364	RSRGIYF	-0.6000	RSRGVFF	-0.0571
359-365	SRGIYFT	-0.0571	SRGVFFT	0.4857
360-366	RGIYFTQ	-0.4429	RGVFFTQ	0.1000
361-367	GIYFTQD	-0.3000	GVFFTQD	0.2429
362-368	IYFTQDW	-0.3714	VFFTQDW	0.1714
363-369	YFTQDWC	-0.6571	FFTQDWA	-0.1714
364-370	FTQDWCS	-0.5857	FTQDWAS	-0.6857
365-371	TQDWCSM	-0.7143	TQDWASM	-0.8143
366-372	QDWCSMP	-0.8429	QDWASMP	-0.9429
367-373	DWCSMPG	-0.4000	DWASMPG	-0.5000

Continued SUPPLEMENTARY FILE 2

368-374	WCSMPGV	0.7000	WASMPGV	0.6000
369-375	CSMPGVM	1.1000	ASMPGVL	1.2714
370-376	SMPGVMP	0.5143	SMPGVLP	0.7857
371-377	MPGVMPV	1.2286	MPGVLPV	1.5000
372-378	PGVMPVA	1.2143	PGVLPVA	1.4857
373-379	GVMPVAS	1.3286	GVLPVAS	1.6000
374-380	VMPVASG	1.3286	VLPVASG	1.6000
375-381	MPVASGG	0.6714	LPVASGG	0.9429
376-382	PVASGGI	1.0429	PVASGGI	1.0429
377-383	VASGGIH	0.8143	VASGGIH	0.8143
378-384	ASGGIHV	0.8143	ASGGIHV	0.8143
379-385	SGGIHVW	0.4286	SGGIHVW	0.4286
380-386	GGIHVWH	0.0857	GGIHVWH	0.0857
381-387	GIHVWHM	0.4143	GIHVWHM	0.4143
382-388	IHVWHMP	0.2429	IHVWHMP	0.2429
383-389	HVWHMPA	-0.1429	HVWHMPA	-0.1429
384-390	VWHMPAL	0.8571	VWHMPAL	0.8571
385-391	WHMPALV	0.8571	WHMPALV	0.8571
386-392	HMPALVE	0.4857	HMPALVE	0.4857
387-393	MPALVEI	1.5857	MPALVEI	1.5857
388-394	PALVEIF	1.7143	PALVEIF	1.7143
389-395	ALVEIFG	1.8857	ALVEIFG	1.8857
390-396	LVEIFGD	1.1286	LVEIFGD	1.1286
391-397	VEIFGDD	0.0857	VEIFGDD	0.0857
392-398	EIFGDDA	-0.2571	EIFGDDS	-0.6286
393-399	IFGDDAC	0.6000	IFGDDSV	0.4714
394-400	FGDDACL	0.5000	FGDDSVL	0.3714
395-401	GDDACLQ	-0.4000	GDDSVLQ	-0.5286
396-402	DDACLQF	0.0571	DDSVLQF	-0.0714
397-403	DAQLQFG	0.5000	DSVLQFG	0.3714
398-404	ACLQFGG	0.9429	SVLQFGG	0.8143
399-405	CLQFGGG	0.6286	VLQFGGG	0.8714
400-406	LQFGGGT	0.1714	LQFGGGT	0.1714
401-407	QFGGGTL	0.1714	QFGGGTL	0.1714
402-408	FGGGTLG	0.6143	FGGGTLG	0.6143
403-409	GGGTLGH	-0.2429	GGGTLGH	-0.2429
404-410	GGTLGHP	-0.4143	GGTLGHP	-0.4143
405-411	GTLGHPW	-0.4857	GTLGHPW	-0.4857
406-412	TLGHPWG	-0.4857	TLGHPWG	-0.4857
407-413	LGHPWGN	-0.8857	LGHPWGN	-0.8857
408-414	GHPWGNA	-1.1714	GHPWGNA	-1.1714
409-415	HPWGNAP	-1.3429	HPWGNAP	-1.3429
410-416	PWGNAPG	-0.9429	PWGNAPG	-0.9429
411-417	WGNAPGA	-0.4571	WGNAPGA	-0.4571
412-418	GNAPGAA	-0.0714	GNAPGAT	-0.4286
413-419	NAPGAAA	0.2429	NAPGATA	-0.1143
414-420	APGAAAN	0.2429	APGATAN	-0.1143
415-421	PGAAANR	-0.6571	PGATANR	-1.0143
416-422	GAAANRV	0.1714	GATANRV	-0.1857
417-423	AAANRVA	0.4857	ATANRVA	0.1286
418-424	AANRVAL	0.7714	TANRVAL	0.4143
419-425	ANRVALE	0.0143	ANRVALE	0.0143
420-426	NRVALEA	0.0143	NRVALEA	0.0143

Continued SUPPLEMENTARY FILE 2

421-427	RVALEAC	0.8714	RVALEAC	0.8714
422-428	VALEACT	1.4143	VALEACV	2.1143
423-429	ALEACTQ	0.3143	ALEACVQ	1.0143
424-430	LEACTQA	0.3143	LEACVQA	1.0143
425-431	EACTQAR	-0.8714	EACVQAR	-0.1714
426-432	ACTQARN	-0.8714	ACVQARN	-0.1714
427-433	CTQARNE	-1.6286	CVQARNE	-0.9286
428-434	TQARNEG	-2.0429	VQARNEG	-1.3429
429-435	QARNEGR	-2.5857	QARNEGR	-2.5857
430-436	ARNEGRD	-2.5857	ARNEGRD	-2.5857
431-437	RNEGRDL	-2.3000	RNEGRDL	-2.3000
432-438	NEGRDLA	-1.4000	NEGRDLY	-1.8429
433-439	EGRDLAR	-1.5429	EGRDLYR	-1.9857
434-440	GRDLARE	-1.5429	GRDLYRE	-1.9857
435-441	RDLAREG	-1.5429	RDLYREG	-1.9857
436-442	DLAREGG	-0.9571	DLYREGG	-1.4000
437-443	LAREGGD	-0.9571	LYREGGD	-1.4000
438-444	AREGGDV	-0.9000	YREGGDI	-1.3000
439-445	REGGDVI	-0.5143	REGGDIL	-0.5714
440-446	EGGDVIR	-0.5143	EGGDILR	-0.5714
441-447	GGDVIRS	-0.1286	GGDILRE	-0.5714
442-448	GDVIRSA	0.1857	GDILREA	-0.2571
443-449	DVIRSAC	0.6000	DILREAG	-0.2571
444-450	VIRSACK	0.5429	ILREAGK	-0.3143
445-451	IRSACKW	-0.1857	LREAGKW	-1.0857
446-452	RSACKWS	-0.9429	REAGKWS	-1.7429
447-453	SACKWSP	-0.5286	EAGKWSP	-1.3286
448-454	ACKWSPE	-0.9143	AGKWSPE	-1.3286
449-455	CKWSPEL	-0.6286	GKWSPEL	-1.0429
450-456	KWSPELA	-0.7286	KWSPELA	-0.7286
451-457	WSPELAA	0.0857	WSPELAA	0.0857
452-458	SPELAAA	0.4714	SPELAAA	0.4714
453-459	PELAAAC	0.9429	PELAAAL	1.1286
454-460	ELAAACE	0.6714	ELAAALD	0.8571
455-461	LAAACEV	1.7714	LAAALDL	1.9000
456-462	AAACEVW	1.1000	AAALDLW	1.2286
457-463	AACEVWK	0.2857	AALDLWK	0.4143
458-464	ACEVWKE	-0.4714	ALDLWKE	-0.3429
459-465	CEVWKEI	-0.0857	LDLWKEI	0.0429
460-466	EVWKEIK	-1.0000	DLWKEIK	-1.0571
461-467	VWKEIKF	-0.1000	LWKEIKF	-0.1571
462-468	WKEIKFE	-1.2000	WKEIKFE	-1.2000
463-469	KEIKFEF	-0.6714	KEIKFEF	-0.6714
464-470	EIKFEFD	-0.6143	EIKFEFE	-0.6143
465-471	IKFEFDT	-0.2143	IKFEFET	-0.2143
466-472	KFEFDTI	-0.2143	KFEFETM	-0.5857
467-473	FEFDTID	-0.1571	FEFETMD	-0.5286
468-474	EFDTIDK	-1.1143	EFETMDK	-1.4857
469-475	FDTIDKL	-0.0714	FETMDKL	-0.4429

^aResidue numbering is based on Chlamydomonas RbcL

^bSynechococcus RbcL is shorter by 3 residues at the N-terminus compared to Chlamydomonas RbcL