DISSECTION OF *SYNECHOCOCCUS* RUBISCO LARGE SUBUNIT SECTIONS INVOLVED IN HETEROLOGOUS HOLOENZYME FORMATION IN *ESCHERICHIA COLI*

ONG WEI CHI

MASTER OF SCIENCE

FACULTY OF SCIENCE
UNIVERSITI TUNKU ABDUL RAHMAN
JULY 2020

DISSECTION OF *SYNECHOCOCCUS* RUBISCO LARGE SUBUNIT SECTIONS INVOLVED IN HETEROLOGOUS HOLOENZYME FORMATION IN *ESCHERICHIA COLI*

By

ONG WEI CHI

A dissertation submitted to the Department of Chemical Science,
Faculty of Science,
Universiti Tunku Abdul Rahman,
in partial fulfillment of the requirements for the degree of
Master of Science
July 2020

ABSTRACT

**DISSECTION OF *Synechococcus* RUBISCO LARGE SUBUNIT SECTIONS
INVOLVED IN HETEROLOGOUS HOLOENZYME FORMATION IN
*Escherichia coli***

Ong Wei Chi

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes fixation of

atmospheric CO_2 into organic carbon ribulose-1,5-bisphosphate (RuBP) in the

Calvin cycle of photosynthesis. Yet, its slow catalysis and inability to differentiate

CO_2 from O_2 for fixation into RuBP make it a target for genetic engineering to

improve its catalytic properties as a prospect for improving photosynthetic

efficiency to raise crop yield. Unfortunately, formation of functional Rubiscos in

heterologous host has been a challenge due to chaperone incompatibility. In

Escherichia coli, prokaryotic Rubiscos form holoenzyme whereas eukaryotic

Rubiscos form insoluble aggregates. As studies have shown, GroEL-GroES

chaperonin mediates folding of Rubisco large subunit (RbcL) in *E. coli*, it is

hypothesized that GroEL does not recognize eukaryotic RbcL as substrate protein.

A previous study reported a few regions of cyanobacterial RbcL from

Synechococcus PCC6301 to be important for successful holoenzyme formation in

E. coli. This study aims to further narrow-down the potential GroEL recognition

(GR) regions by breaking down these regions into six smaller regions (each with

25 amino acid residues), replacing them with their counterpart sequence of green

algal *Chlamydomonas reinhardtii* RbcL, and checking their influences on holoenzyme formation in *E. coli*. If swapped regions of *Synechococcus* RbcL are important for are important for GR, no formation of chimeric Rubisco should be observed. Therefore, six chimeric Rubiscos were constructed. Besides, examined RbcL regions were screened for potential GR sites based on the hydropathicity (GRAVY) value of GroES mobile loop sequence. Moreover, as non-assembly could also be due to global protein instability imparted by the structural destabilization effect of mutations, any loss of interaction arises from mutations were also predicted. Assembly analysis, based on native polyacrylamide gel electrophoresis (PAGE) showed residues 248-272, 273-297, 348-372 and 423-447 could be essential for GR whereas residues 373-397 and 398-422 are not. Moreover, site-directed Rubisco mutants with single mutations were created to examine their individual impacts on Rubisco assembly. Substitution of residues 348-372 of *Synechococcus* RbcL by the *Chlamydomonas* counterpart introduced eight mutations and resulted in non-assembly. Interestingly, these eight mutations did not result in non-assembly individually but some of them reduced the amount of assembled enzyme.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to my supervisor, Dr. Lim Boon Hoe for giving me the opportunity to engage in this project. His patience, guidance and insightful advices have been an encouragement for me to overcome the challenges and difficulties encountered throughout the course of project.

I would also like to pay my gratitude to my co-supervisor, Dr. Wong Hann Ling for offering constructive suggestions and assessing the work progress.

I would also like to thank the laboratory officers especially Mr. Ooh Keng Fei, Mr Leong Thung Lim and Mr Seou Chi Kien for their assistance and technical support.
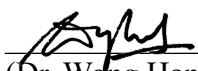
This dissertation/thesis entitled "**DISSECTION OF *SYNECHOCOCCUS* RUBISCO LARGE SUBUNIT SECTIONS INVOLVED IN HETEROLOGOUS HOLOENZYME FORMATION IN *ESCHERICHIA COLI*"** was prepared by ONG WEI CHI and submitted as partial fulfillment of the requirements for the degree of Master of Science at Universiti Tunku Abdul Rahman.

Approved by:

Lim Boon Hoe
_____

(Dr. Lim Boon Hoe)   Date: … 21 Oct 2020
Assistant Professor/Supervisor
Department of Chemical Science
Faculty of Science
Universiti Tunku Abdul Rahman

_____

(Dr. Wong Hann Ling)   Date:… 21.10.2020
Associate Professor/Co-supervisor
Department of Biological Science
Faculty of Science
Universiti Tunku Abdul Rahman

FACULTY OF SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: ___21 Oct 2020_____

SUBMISSION OF DISSERTATION

It is hereby certified that **Ong Wei Chi** (ID No: **17ADM05230**) has completed this dissertation entitled "**DISSECTION OF *Synechococcus* RUBISCO LARGE SUBUNIT SECTIONS INVOLVED IN HETEROLOGOUS HOLOENZYME FORMATION IN *Escherichia coli***" under the supervision of Dr. Lim Boon Hoe (Supervisor) from the Department of Chemical Science, Faculty of Science, and Dr. Wong Hann Ling (Co-Supervisor) from the Department of Biological Science, Faculty of Science.

I understand that University will upload softcopy of my dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

_____
(Ong Wei Chi)

DECLARATION

I hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

Name   Ong Wei Chi

Date   21/10/2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2PG	2-phosphoglycolate
3PGA	3-phosphoglycerate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
AMP-PNP	β-Υ-imidoadenosine 5'-phosphate
Bicine	2-(Bis(2-hydroxyethyl)amino)acetic acid
CBB	Coomassie brilliant blue
CO_2	Carbon dioxide
DTT	Dithiothreitol
G3P	Glyceraldehydes-3-phosphate
GR	GroEL Recognition
GRAVY	Grand average of hydrophobicity
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Ω	Relative specificity of Rubisco for CO_2 versus O_2
k_{cat}	Carboxylation turn-over rate
LB	Luria-bertani
NADPH	Nicotinamide adenine dinucleotide phosphate
O_2	Oxygen
PCR	Polymerase chain reaction
PAGE	Polyacrylamide gel electrophoresis
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase

RuBP	Ribulose-1,5-bisphosphate
RbcL	Rubisco large subunit
RbcS	Rubisco small subunit
RPM	Revolution per minute
SDS	Sodium dodecyl sulphate
TBS	Tris-buffered saline
TBST	Tris-buffered saline-Tween20
Vmax	Maximal carboxylation rate
VDW	Van der Waals

CHAPTER 1

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), is a CO_2 fixing enzyme found in photoautotrophs and chemoautotrophs. It catalyzes the first step of Calvin cycle, fixing of inorganic CO_2 into organic carbon ribulose-1,5-bisphosphate (RuBP) (Ellis, 1979; Hauser et al., 2015). Carboxylation of RuBP produces two molecules of three-carbon 3-phosphoglycerate (3PGA), which is then metabolized in the Calvin cycle for sugar biosynthesis. However, Rubiscos are slow enzymes with carboxylation turnover rates in a range of 1 s^{-1} to 13 s^{-1} (Greene et al., 2007). Moreover, its inability to differentiate CO_2 from O_2 makes it a bifunctional enzyme capable of fixing O_2 other than CO_2 onto RuBP. Oxygenation of RuBP is undesired for photosynthesis as it results in the production of a two-carbon compound, phosphoglycolate (2PG), which lead to photorespiration for its recycling into 3PGA. Photorespiration is energetically costly and lead to loss of fixed carbon, hence, greatly decreases net photosynthesis (Wingler et al., 2000; Wilson and Hayer-Hartl, 2018). Slow catalytic rate and oxygenase activity of Rubisco make it one of the rate-limiting factors of photosynthesis, thus, a target for genetic engineering for better kinetic properties as a prospect for raising photosynthetic efficiency to increase crop productivity (Long et al., 2006; Parry et al., 2013).

Rubiscos exist in four different forms in nature. Form I Rubisco is comprised of eight large (RbcL, 50-55 kDa) and small subunits (RbcS, 12-15 kDa) arranged as a L_8S_8 hexadecamer (~560 kDa) is the most abundant form (Spreitzer and Salvucci, 2002; Andersson and Backlund, 2008). They are found in proteobacteria, cyanobacteria, algae and plants (Tabita et al., 2008). Form II Rubiscos are found in purple non-sulfur bacteria and eukaryotic dinoflagellates while Form III in archaea (Andersson, 2008; Tabita et al., 2008). Both forms exist as different number of L_2 dimer pair, $(L_2)_n$. Form IV Rubiscos are Rubisco-like proteins (RLP) that share similar tertiary structure but lack of key conserved catalytic residues for Rubisco catalytic reactions (Tabita et al., 2008).

Rubiscos have conserved structures and key catalytic residues but varying catalytic efficiencies (Andersson and Backlund, 2008). Two important kinetic parameters to describe and compare catalytic performance are 1) relative specificity for CO_2 and O_2 in term of CO_2/O_2 specificity factor (Ω) and 2) maximal carboxylation rate (Vmax). Nevertheless, trade-off between Ω and Vmax is observed as Rubisco with high Ω has low Vmax, and vice versa (Tcherkez et al., 2006). On contrary, some of the Rubiscos from non-green algae do not follow such inverse relationship. They have high CO_2/O_2 specificity coupled with a slightly lower carboxylation velocity, making them more catalytically efficient than plant Rubisco (Whitney et al., 2001). It is postulated that if C3 Rubiscos are replaced with Rubisco from red alga *Griffithsia monilis*, 30% of increment in C3 crops yield may be produced (Zhu et al., 2004).

For improving the kinetic performance of Rubisco, extensive studies of Rubisco have been conducted in order to engineer a catalytically improved Rubisco and identify the key determinants governing its catalytic efficiency (Spreitzer et al., 2005; Sharwood et al., 2008; Genkov and Spreitzer, 2009; Ishikawa et al., 2011; Whitney et al., 2011a). Nevertheless, unsuccessful Rubiscos holoenzyme formation in phylogenetically distant host limits the study of Rubiscos and selection of host. Such inability of heterologous expression stems from the chaperone incompatibility and need of additional auxiliary factors for their complex biogenesis (Mueller-Cajar and Whitney, 2008). Rubisco biogenesis is a multi-step process that necessitates different kinds of chaperones for proper folding and assembly into L_2, $(L_2)_4$ and final L_8S_8 hexadecamer (Kolesinski et al., 2014; Hauser et al., 2015; Wilson and Hayer-Hartl, 2018).

Directed evolution of Rubisco in *E. coli* has shown to be a high-throughput tool for evolving Rubisco variants with improved catalytic performance and providing structure-function insights (Parikh et al., 2006; Mueller-Cajar and Whitney, 2008; Wilson et al., 2018). *E. coli* has faster growth rate and higher transformation efficiency ($\sim 10^{10}$ transformants/µg plasmid) that is at least three orders of magnitude higher than photosynthetic microbes (Smith and Tabita, 2003), therefore, allowing large scale sampling of Rubisco mutants in a relatively shorter time (Parikh et al., 2006). However, genetic manipulation of eukaryotic Rubiscos in this genetically malleable host is hampered by chaperone incompatibility.

Unlike prokaryotic Rubiscos which are able to form holoenzyme in *E. coli*, eukaryotic Rubiscos form insoluble aggregates (Bradley et al., 1986; Gatenby et al., 1987; Koay et al., 2016). Only recently were eukaryotic Rubisco from tobacco and *Arabidopsis* able to form soluble enzyme in *E. coli* by co-expressing with their cognate chaperones from chloroplast (Aigner et al., 2017; Lin et al., 2019). Meanwhile, formation of active non-green algae Rubiscos in *E. coli* or tobacco chloroplast are still impossible, hinting that molecular chaperone system in these two organisms do not meet its biogenesis requirement (Whitney et al., 2001; Lin and Hanson, 2018). In *E. coli*, folding of Rubiscos large subunit monomers is mediated by GroEL-GroES chaperonin system (Goloubinoff et al., 1989b). Studies have reported that GroEL-GroES chaperonin recognizes prokaryotic Rubiscos like Form II Rubisco from *Rhodospirillum rubrum* and Form I cyanobacterial Rubisco (Goloubinoff et al., 1989b; Mueller-Cajar and Whitney, 2008; Koay et al., 2016).

GroEL chaperonin is a tetradecamer (~800 kDa) that consists of two identical heptameric rings stacked back-to-back (Xu et al., 1997). Each seven-membered ring has a central cavity in which the substrate protein fold. GroEL works with co-chaperonin GroES, a single-ring heptamer, in an ATP-dependent manner. As substrate protein binds to multiple GroEL apical domains through hydrophobic interactions (Farr et al., 2000), it is suggested that a GroEL substrate protein carries multiple GroEL binding sites. In addition, given that GroEL only interacts with its substrate proteins with non-native conformations, it is believed that GroEL binding regions expose in their non-native state but buried in native

4

states and become inaccessible to GroEL (Stan et al., 2006). Taking GroES as the best substrate of GroEL, the sequence of GroES mobile loop was used as a reference to look for hydrophobic patches in substrate proteins with similar chemical identity and approximate grand average of hydropathicity (GRAVY) value as potential GroEL binding sites (Chaudhuri and Gupta, 2005; Stan et al., 2006).

Previously, an attempt was made to pinpoint the regions of *Synechococcus* PCC6301 RbcL important for successful holoenzyme formation in *E. coli* by creating chimeric Rubisco with regions of *Synechococcus* RbcL swapped to their counterparts in eukaryotic RbcL from green alga *Chlamydomonas reinhardtii* (Koay et al., 2016). As *Chlamydomonas* Rubisco failed to form active enzyme in *E. coli*, it was presumed that loss of regions *Synechococcus* RbcL important for successful holoenzyme formation, when substituted by its *Chlamydomonas* counterpart sequence, would result in non-assembly. Using this approach, a few regions of *Synechococcus* RbcL that might be important for successful holoenzyme formation were suggested (Koay et al., 2016). As an extension of the previous work, the present study aims to further narrow down these reported regions. As GroEL chaperonin mediates folding of bacterial RbcL in *E. coli*, these reported sections might carry potential GroEL recognition (GR) sites for interacting with the GroEL chaperonin to enter GroEL mediated folding pathway. Moreover, primary structures of examined RbcL regions are screened for hydrophobic patch with approximate GRAVY value of GroES mobile-loop sequence as potential GR sites.

Other than loss of GroEL recognition for proper folding, non-assembly of Rubisco variants could be due to structural destabilizing effect of mutations introduced by *Chlamydomonas* counterpart, which lead to protein instability. Therefore, any potential loss of interaction arises from mutations are predicted by *in silico* analysis.

The specific objectives are:

1. To narrow down the possible regions harbouring the GroEL binding sites by creating chimeric bacterial/eukaryotic Rubiscos with 25 amino acid residues of *Synechococcus* PCC6301 RbcL swapped to their counterparts from *Chlamydomonas* RbcL.

2. To examine the importance of single amino acid residues for the GroEL recognition (GR) /assembly by creating mutants with single mutations by site-directed mutagenesis.

3. To identify possible GroEL recognition (GR) sites by *in silico* analysis of wild-type *Synechococcus* PCC6301 and *Chlamydomonas* RbcL primary structures.

4. To identify potential loss of interaction resulted from mutations by *in silico* analysis.

CHAPTER 2

LITERATURE REVIEW

2.1   Rubisco: Enzyme for CO_2 fixation in photosynthesis

Photoautotrophs utilize light energy and inorganic CO_2 through photosynthesis for biomass accumulation to support their life and sustain all life on earth directly or indirectly (Boyer, 2006; Evans, 2013). In plants, photosynthesis involves two separate reactions, namely light reaction and dark reaction, taking place in different location in chloroplast to complete whole process of carbohydrate synthesis (Fig 2.1) (Boyer, 2006; Long et al., 2015). The light reactions, which are sometimes being referred as light-dependent reactions, take place at thylakoid in chloroplast where light energy is captured and converted to NADPH and ATP by electron transport chain and ATP synthase in thylakoid membrane. Energy harvested from light reaction then power the Calvin–Benson–Bassham cycle of dark reaction, also known as light independent reaction, in stroma to incorporate inorganic CO_2 into the biosphere for starch and sucrose biosynthesis (Boyer, 2006; Long et al., 2015). Calvin cycle is comprised of three steps: carbon fixation, sugar reduction and RuBP regeneration.

Rubiscos catalyze the first step of Calvin cycle, fixation of CO_2 to five-carbon ribulose-1,5-bisphosphate (RuBP) and produces two molecules of three-

carbon 3-phosphoglycerate (3PGA), which are then reduced to triose sugar,

glyceraldehyde 3-phosphate (G3P) for both regeneration of RuBP and carbohydrate

synthesis (Hauser et al., 2015). Given its role in CO_2 fixation, Rubisco is considered

as the gateway of inorganic CO_2 into biosphere and its capability to sequester

inorganic CO_2 dictates the efficiency of photosynthetic CO_2 assimilation (Parry et

al., 2007; Wilson and Hayer-Hartl, 2018).
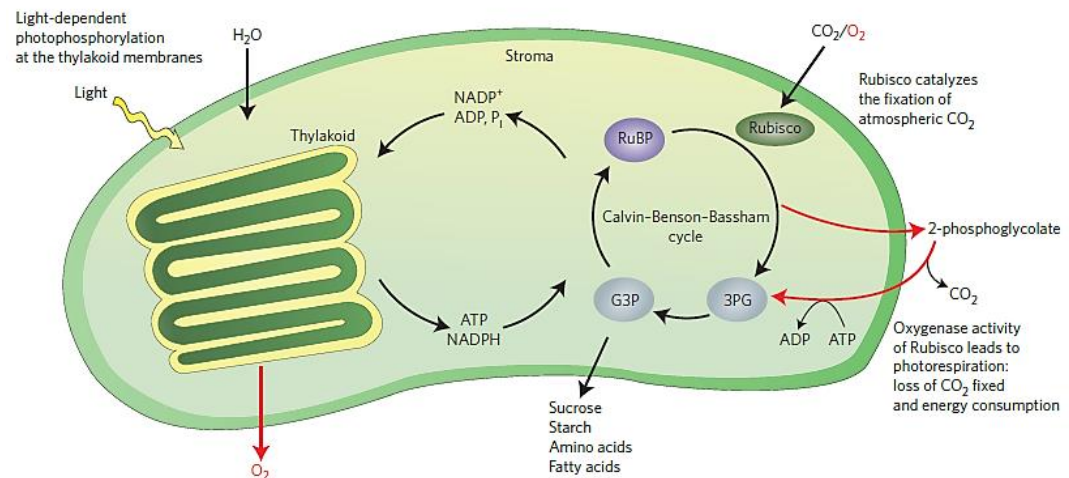
Figure 2.1: Light and dark reaction of photosynthesis in chloroplast. Light reaction happens in thylakoid where light energy is captured and converted to ATP and NADPH. In stroma, energy harvested from the light reaction powers the Calvin cycle. Atmospheric CO_2 is fixed into RuBP by Rubisco to produce two molecules of 3PG, which are reduced into G3P for RuBP regeneration and carbohydrate synthesis. Other than CO_2, Rubisco also fixes O_2 into RuBP and produces one 3PG and one 2PG, which must be recycled into 3PG to re-enter the Calvin cycle through photorespiration (Figure adopted from Hauser et al. 2015).

2.1.1 Slow and bifunctional catalysis of Rubiscos limit the photosynthetic CO_2 assimilation and resource-use efficiency

Rubiscos' catalytic efficiency determines photosynthetic CO_2 assimilation rate in C3 crops. However, slow catalytic rate and bifunctional catalysis of Rubisco render CO_2 assimilation one of the rate-limiting factors of photosynthesis (Long et al., 2006; Parry et al., 2013; Ort et al., 2015). Rubisco from plants has slow catalytic rate of ~3 per active site per second (Whitney et al., 2011b). Besides, it is a bifunctional enzyme capable of fixing CO_2 or O_2 into RuBP (Fig. 2.1). Inability of Rubisco to fully distinguish between CO_2 and O_2 greatly decreases the photosynthetic efficiency by diverting the flux into photorespiration. Oxygenation of RuBP results in one molecule of 3PGA and one molecule of two-carbon compound, phosphoglycolate (2PG), which must be recycled into 3PGA to re-enter Calvin cycle (Wingler et al., 2000; Betti et al., 2016). Plants metabolize 2PG through a photorespiratory pathway consists of a series of enzymatic reactions taking place in chloroplast, peroxisomes, mitochondria and cytosol to regenerate 3PGA (Bauwe et al., 2010). However, recycling of 2PG through photorespiration is energetically costly, requiring significant amount of energy generated by the light reaction, 3.25 mol ATP and 2 mol NADPH per molecule of 2PG (Wingler et al., 2000). Furthermore, one molecule of 3PGA is recovered from two molecules of 2PG, and one out of four carbon atoms from the two 2PG is released as photorespiratory CO_2, thereby, significantly reducing net photosynthesis (Wingler et al., 2000; Bauwe et al., 2010; Timm et al., 2016).

Photorespiratory carbon flow is exceptionally high in C3 plants. At 25 °C and current atmospheric CO_2 concentration, plants lose ~30% of organic carbon formed in C3 photosynthesis via photorespiration (Zhu et al., 2010). This carbon loss further increases to more than 50% at high temperature and scarcity of water (Long et al., 2006). This is because as temperature rise, CO_2/O_2 specificity factor of Rubisco decrease and the solubility of CO_2 reduced more greatly than O_2 (Peterhansel et al., 2008). Moreover, plant reduces gaseous exchange by stomatal closure when temperature is high and water is scarce to decrease loss of water through transpiration (Parry et al., 2007; Peterhansel et al., 2008). Under these situations, $[CO_2]$ around Rubiscos deplete over time, relative $[O_2]$ becomes high and favor oxygenation (Zhu et al., 2008).

Rubisco's catalytic performance also determines the maximum efficiency of photosynthesis in its use of light, water, and N resources (Morell et al., 1992; Carmo-Silva et al., 2015). Competing oxygenase activity of Rubisco reduces light-use efficiency, which is particularly important under light limiting condition, by diverting part of limited light energy into photorespiration, hence decrease CO_2 assimilation. Moreover, the slow speed and presence of oxygenase activity of Rubiscos necessitate large amounts of their production to achieve adequate photosynthesis. As a result, Rubisco make up ~ 30-50% of total soluble leaf protein, account for 10-30% of total nitrogen in leaves (Ellis, 1979; Carmo-Silva et al., 2015).

11

2.1.2 Engineering of Rubisco for better photosynthetic CO_2 assimilation efficiency to raise crop yield

In order to meet global food supply and biofuel demand of growing world population, more than 85% increment of agricultural productivity in 2013 is required by 2050 (Long et al., 2015). Unfortunately, current crop productivity is stagnating and conventional breeding techniques are unlikely to bring further improvement in crop yield. In the past, crop productivities were raised substantially by agronomic management, smart plant architecture for better light capturing efficiency, and selection of cultivar with higher harvest index that allows partition of higher proportion of total biomass generated by photosynthesis into the harvestable product (Ort et al., 2015; Simkin et al., 2019). However, these approaches and use of nitrogenous fertilizers seem to have reached their theoretical maxima (Long et al., 2006). Now, the remaining hope of crop yield improvement is to increase photosynthetic efficiency for better conversion of intercepted light energy into biomass (Long et al., 2006; Zhu et al., 2010). For C3 plants, theoretical maximum solar energy conversion efficiency achievable is 4.6%, but less than 50% of this efficiency is attained in field (Zhu et al., 2008). Feasibility of improving photosynthesis as a means to raise crop yield is supported by CO_2 enrichment studies, which show crops grown under elevated $[CO_2]$ have increased leaf photosynthetic CO_2 uptake and improved yield (Long et al., 2006). Several targets have been identified to increase photosynthetic efficiency in the aspects of 1) reduce or eliminate photorespiration, 2) increase capacity of photon use at light saturation and 3) maintain or increase maximum efficiency of photosynthesis at

light-limitation. These may involve conversion of C3 crops to C4 or increase Rubisco CO_2/O_2 specificity to reduce photorespiration, increase maximum carboxylation rate of Rubisco and rate of RuBP regeneration for higher capacity to utilize NADPH and ATP produced in high light, and decrease oxygenase activity of Rubisco for more productive use of light energy under light limiting condition (Long et al., 2006; 2015; Ort et al., 2015). Among all, improving kinetic properties of Rubiscos is regarded as one of the prime targets. Rubisco with higher catalytic rate, specificity and affinity for CO_2 would increase photosynthetic efficiency and crop yield (Parry et al., 2003; 2011). A more kinetically efficient Rubisco would also give better CO_2 assimilation with better use of light energy, less-water loss, and less nitrogen investment in Rubisco (Andrews and Whitney, 2003).

Different forms of Rubiscos exist in nature. They have conserved architecture and catalytic reactions but vary in kinetic properties. Interestingly, plant Rubiscos are not the most catalytically efficient forms. These two findings have spurred study of Rubiscos and genetic engineering to improve its catalytic properties. It is postulated that if the present C3 crop Rubisco could be replaced by the Rubisco from the red alga *G. monilis*, a 27% increase in daily canopy carbon gain could be attained (Zhu et al., 2004).

2.2   Four Forms of Rubisco

Four forms, namely Form I, II, III and IV of Rubiscos are found in diverse autotrophs (Tabita et al., 2008). Among all, Form I Rubiscos found in

proteobacteria, cyanobacteria, plants and algae are the most abundant form (Tabita et al., 2008). It is a hexadecamer (L_8S_8, ~560 kDa) comprised of eight large subunits (RbcL, 50-55 kDa) and eight small subunits (RbcS, 12-15 kDa) (Knight et al., 1990; Spreitzer and Salvucci, 2002; Andersson and Backlund, 2008). Four RbcL dimers are arranged into $(L_2)_4$ core, then capped by two groups of four RbcS at top and bottom, respectively, to form a L_8S_8 complex (Fig. 2.2C) (Knight et al., 1990; Whitney et al., 2011b). They are further classified into green type from cyanobacteria, green algae and plant, and red type from non-green algae, except for dinoflagellate, and some photosynthetic bacteria (Andrews and Whitney, 2003). In plant and green algae, large subunits are chloroplast encoded while small subunits are nuclear encoded from a multigene family (Tabita et al., 2008; Spreitzer and Salvucci, 2002). Meanwhile, in prokaryotes and non-green algae, both *Rbc*L and *Rbc*S genes are arranged together in an operon (Tabita et al., 2008; Spreitzer and Salvucci, 2002).

Form II Rubiscos are found in purple non-sulfur bacteria, chemoautotrophic bacteria and eukaryotic dinoflagellates while Form III in archaea (Tabita et al., 2008; Andersson, 2008). Both forms consist of only large subunits in different numbers of dimeric pairs, $(L_2)_n$, from L_2 to $(L_2)_5$ complexes (Whitney et al., 2011b). Form IV Rubiscos are Rubisco-like proteins (RLP) that shares similar tertiary structure but lack of key conserved catalytic residues for Rubisco catalytic reactions (Tabita et al., 2008).

14

2.2.1   Conserved large subunit architecture and active site framework

Rubiscos from different organisms show variation in primary sequence but have conserved key catalytic residues and three-dimensional structure. The functional unit of all forms of Rubiscos is a large subunit dimer (L_2) composed of two RbcL arranged head to tail, forming two active sites at the interface (Fig. 2.2B). RbcL in all forms have a conserved structure of one amino-terminal domain (~150 residues) and a larger carboxy-terminal domain (~320 residues) (Fig. 2.2A) (Whitney et al., 2011b). The amino-terminal domain consists of four to five-stranded mixed β-sheets with helices on one side of the sheet whereas the carboxy-terminal consists of eight consecutive βα-units arranged into an α/β barrel (Andersson and Backlund, 2008). At the L-L interface, active site are brought together by conserved catalytic residues mainly from α/β barrel of C-terminal of one RbcL, and a few from N-terminal of adjacent RbcL (Whitney et al., 2011b).

15
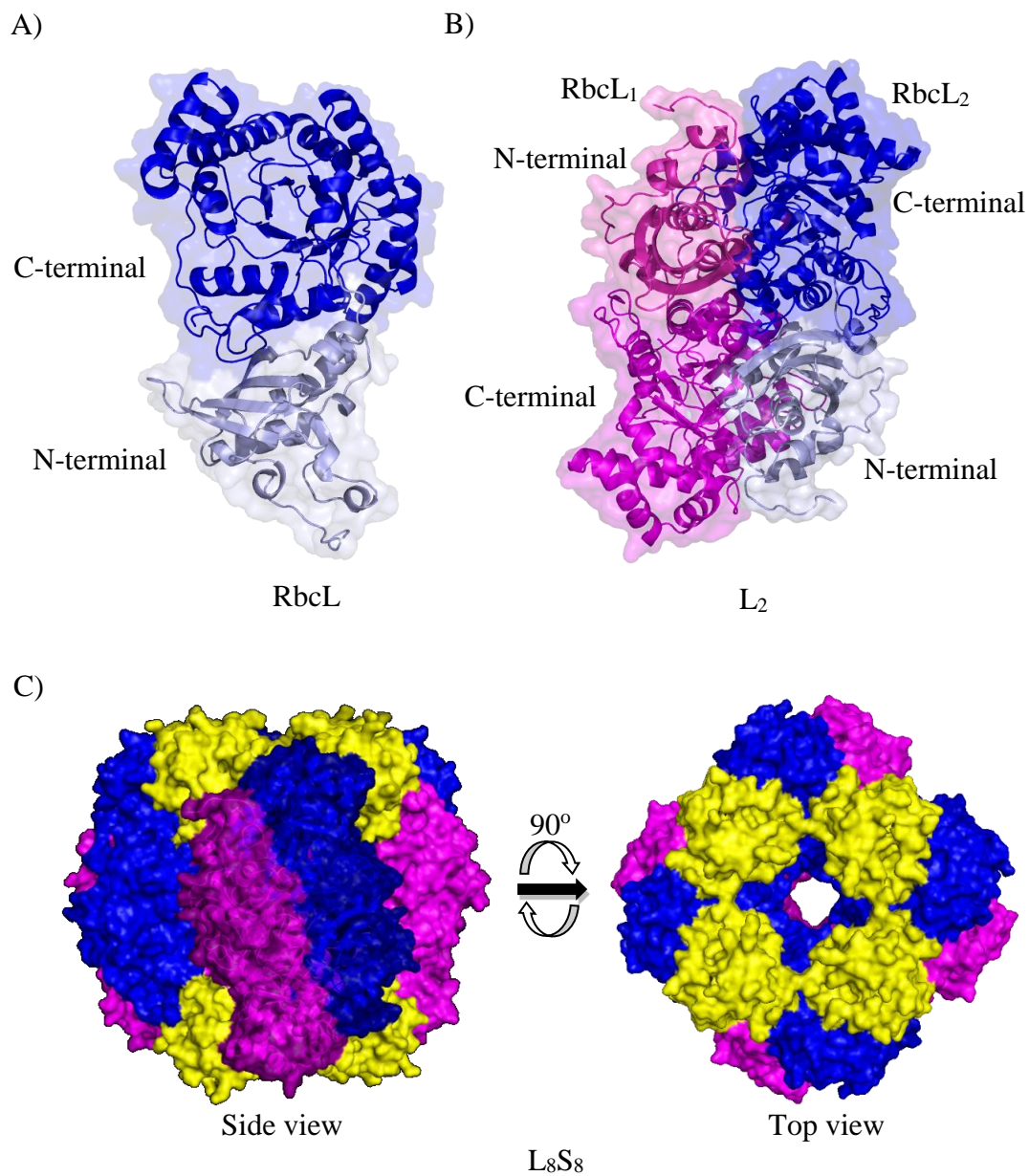
Figure 2.2: Structure of Form I Rubisco. A) Rubisco large subunit (RbcL), B) large subunit dimer (L_2) and C) L_8S_8 of Form I Rubisco. RbcL consists of one large C-terminal and a smaller N-terminal. Two RbcL arranged head to tail and form L_2 dimer. RbcL pairings within L_2 dimers are colored distinctly in magenta and blue whereas tetrameric RbcS capping both ends of the L_8 complex are colored in yellow. Images are generated using PyMOL v0.99.

2.2.2 Catalytic mechanism

Catalysis of carboxylation and oxygenation of RuBP require prior activation of Rubiscos that involve carbamylation of amino group of a conserved lysine residue (Lys 201, based on spinach numbering) in the active site by non-substrate CO_2 molecules and stabilization of carbamate by divalent ion, Mg^{2+} (Parry et al., 2003; Andersson, 2008). Resultant ternary complex (Rubisco.CO_2.Mg^{2+}) is the catalytically active state of Rubisco as the carbamate group is directly involved in both carboxylation and oxygenation of RuBP (Parry et al., 2003; 2008). Without carbamylation, binding of substrates to the active site lock Rubisco in a closed, unproductive form (Parry et al., 2008). Besides, there are naturally occurring sugar phosphates resemble the transition state analogue acting as inhibitors to Rubiscos. They bind tightly to active site before/after carbamylation, results in inactivation of Rubisco and prevent carbamylation/substrate binding (Parry et al., 2008). Reactivation of Rubisco requires assistance form Rubisco activase for removing inhibitor.

Carbamylated Rubiscos catalyze carboxylation in five discrete steps: 1) enolization of RuBP, 2) carboxylation/oxygenation of the 2,3-enediolate intermediate of RuBP, 3) hydration of resulting ketone, 4) carbon-carbon scission, and 5) stereospecific protonation of resulting carboxylate of one of the 3-PGA produced. The first step of carboxylation catalyzed by Rubisco is the formation of 2,3-enediol intermediate of the RuBP. This is then followed by carboxylation where C2,C3-enediol reacts with CO_2 at the C2 position, leading to the formation of a six-

carbon intermediate, which is later hydrolytically cleaved to form two molecules of 3-PGA (Andersson, 2008; Sharwood, 2017). Oxygenation occurs when O_2 instead of CO_2 is added to 2,3-enediol intermediate of RuBP, following hydration and cleavage produce one molecules of 3-PGA and one molecules of 2PG (Andersson, 2008).

2.2.3 Catalytic properties of Rubiscos

Two parameters important for Rubisco catalytic performance are 1) maximum carboxylation rate (Vc) and 2) CO_2/O_2 specificity factor (Ω), which describes the efficiency to distinguish between CO_2 and O_2. CO_2/O_2 specificity factor (Ω) is the ratio of carboxylation efficiency (Vc/Kc) to oxygenation efficiency (Vo/Ko) at equal concentration of CO_2 and O_2, $\Omega = VcKo/VoKc$, where Vc and Vo are the maximal velocity (V_{max}) for carboxylation and oxygenation, and Kc and Ko are the Michaelis constant for CO_2 and O_2 (Chen and Spreitzer, 1992; Spreitzer and Salvucci, 2002; Andersson, 2008). A higher CO_2/O_2 specificity factor indicates greater efficiency of CO_2 to compete with O_2 (Read and Tabita, 1994). With respect to photosynthesis, the CO_2/O_2 specificity factor reflects the balance between the carboxylase and oxygenase activities of Rubisco, therefore, balance between photosynthesis and photorespiration (Whitney et al., 2001).

Rubiscos from different sources show varying kinetic properties. In general, Form I Rubisco has higher specificity for CO_2 than Form II Rubisco (Spreitzer, 2003). Among Form I Rubisco, eukaryotic red algae have the highest Ω of 180-240,

18

plants and green algae have an intermediate Ω of 60-100, whereas cyanobacteria and photosynthesizing α-proteobacteria have lowest Ω of about 5-40 (Andersson, 2008). However, a trade-off between Ω and V_c is observed where Rubisco with high Ω is generally accompanied by a low V_c and vice versa (Tcherkez et al., 2006). It is argued that, despite being sluggish, Rubiscos have achieved tremendous work for being able to fix CO_2 under readily available O_2 molecules in the solution at 25°C, and they may almost have been perfectly adapted to their varying subcellular CO_2/O_2 ratio and thermal condition by natural evolution (Tcherkez et al., 2006).

Therefore, if such a trade-off is unavoidable, increase of Ω or V_c at the expense of another does not necessarily confer better catalytic performance. High Ω, by itself, does not necessarily give rise to higher photosynthesis if V_c is too low, and vice versa (Whitney et al., 2001; Spreitzer and Salvucci, 2002; Zhu et al., 2004). For example, transgenic tobacco with faster cyanobacterial Rubisco grow at a slower rate than wild type tobacco and require higher CO_2 levels for growth, as its low specificity for CO_2 and high K_c lead to high oxygenase activity and photorespiration (Lin et al., 2014).

Some of Form I Rubiscos from the non-green algae, particularly rhodophytes, do not follow this trade-off. They have specificity two- or more fold higher than those of higher-plant Rubiscos and sometimes coupled with high or slightly lower V_c, making them more kinetically efficient (Read and Tabita, 1994; Whitney et al., 2001). In fact, it is proposed that although *G. monolis* Rubisco has

a slightly lower Vc and Vc/Kc than that of C3 crop, its large specificity (Ω =167; C3 plant has average Ω of 92.5) would allow it to outperform the tobacco enzyme in achieving higher photosynthetic rate per unit Rubisco by improving light-use efficiency (Zhu et al., 2004).

2.3   Past attempts on Rubiscos

Although large subunits contain the active sites for catalytic activity, studies have shown that small subunits can affect the catalytic properties of the enzyme (Spreitzer and Salvucci, 2002; Spreitzer, 2003; Andersson, 2008). RbcS are more divergent than RbcL. One particular region of the small subunit that exhibits high variability in sequence and length is the βA-βB loop. RbcS from green algae and land-plant usually have longer βA–βB loops (21 or more residues) whereas those from some prokaryotes and all nongreen algae have shorter loop (10 residues). Rubisco enzymes with enlarged βA–βB loops, as in land plants and green algae, generally have higher CO_2/O_2 specificity values than enzymes with normal βA–βB loops. Spreitzer et al. (2001) performed mutagenesis of small subunit βA-βB loop, by changing residues in the loop region to alanine, and showed that βA-βB loop which is distant from the active site influences catalytic efficiency and CO_2/O_2 specificity. Furthermore, *Chlamydomonas* Rubisco with five amino acid residues mutated (V221C, V235I, C256F, K258R, and I265V) along with replacement of its βA-βB loop with shorter βA-βB loop of spinach small subunit resulted in 12-17% of increment in CO_2/O_2 specificity and spinach-like carboxylation and oxygenation kinetic constants (Spreitzer et al., 2005).

20

Besides, there have been studies which successfully constructed hybrid Rubiscos. One hybrid comprising sunflower (*Helianthus annuus*) large subunit and tobacco small subunit, retains catalytic properties that are comparable to the native sunflower and tobacco Rubiscos (Sharwood et al., 2008). Moreover, functional hybrids with 3-11% increases in CO_2/O_2 specificity and an almost normal V_{max} value were obtained by transforming *C. reinhardtii* mutant lacking RbcS gene with small subunits from spinach, *Arabidopsis*, and sunflower respectively (Genkov et al., 2010). Another instance is a hybrid of RbcL from rice (*Oryza sativa*) with RbcS from C4 plant sorghum (*Sorghum bicolor*), displaying higher k_{cat} but lower affinity and slightly lower specificity for CO_2 (Ishikawa et al., 2011).

Although past extensive studies have advanced our understanding of catalytic reaction mechanism and structure-function relationships of Rubisco, non-active site residues that can influence the catalytic properties of Rubisco have not been fully determined and no "better" Rubisco has been rationally designed (Mueller-Cajar and Whitney, 2008; Wilson et al., 2018). Therefore, more comprehensive understanding of this enzyme is required in order to engineering catalytically more efficient Rubisco. On the other hand, artificial laboratory evolution of Rubisco has been shown to be a highly useful approach towards improving the kinetics of enzymes and providing previously unknown structure-function relationships (Mueller-Cajar and Whitney, 2008; Wilson et al., 2018). Using hosts with high transformation efficiencies, high throughput screening of library of Rubisco mutants created by random mutagenesis under selection pressure

for improved fitness had been established (Mueller-Cajar and Whitney, 2008; Wilson et al., 2016; 2018).

2.3.1 *E. coli* **as a bio-selection system for directed evolution of Rubisco**

Smith and Tabita (2003) demonstrated the first directed evolution of Rubisco in a bacterial bioselection system using Rubisco-deletion (Δrbc) mutants of *Rhodobacter capsulatus* (Paoli et al., 1998). A library of *Synechococcus* PCC6301 Rubisco mutants was created by random mutagenesis, transformed into *R. capsulatus* and selected under two different CO_2 levels, 1.5% and 5%. Around $5x10^3$ of transformants were screened. One mutant carries single amino acid substitution in large subunit, F342V (based on *Synechococcus* numbering) showed about 2-fold greater affinity for RuBP (K_{RuBP}) relative to wild-type Rubisco. In addition, a G176D mutation exhibits higher K_{RuBP} but lower K_C and slower carboxylase activity than wild-type (Smith and Tabita, 2003; 2004). Interestingly, glycine 176 is one of the conserved residues in spinach Rubisco and locates near the interface between the large subunit dimer. This approach provided insight that residue distant from the active site and near the interface can affect kinetic properties.

An *E. coli*-based selection system were later demonstrated for directing the evolution of *Synechococcus* Rubisco PCC6301 (Parikh et al., 2006). Taking the advantage that 8 of the 11 Calvin cycle enzyme are naturally expressed in *E. coli*, partial Calvin cycle was constructed in *E. coli* by introducing the *Synechococcus*

PCC7492 phosphoribulokinase (PRK) (Parikh et al., 2006). PRK converts central intermediate in the pentose-phosphate pathway, ribulose-5-phosphate, irreversibly into RuBP which *E. coli* does not metabolize, thereby leading to a metabolic dead-end. Accumulation of RuBP is toxic to cell and finally results in growth arrest. The RuBP-toxicity could be alleviated by Rubisco, to convert RuBP into 3-PG, reverting the metabolic flux to glycolysis.

Using this Rubisco-dependent *E. coli*, randomly mutagenized Rubisco mutants were selected by 3 rounds of generation under different CO_2 level and non-permissive amount of PRK. Three hypermorphs (M259T, A8S/M259T and M259T/F342S), previously identified by Smith and Tabita, (2003), exhibit 5-fold improvement in expression and ~12-28% improvements in carboxylation efficiency (k_c^{cat}/K_c) (Parikh et al., 2006; Greene et al., 2007). The improved expression in *E. coli* was assumed to be due to reduced propensity to misfold and/or by enhanced interaction with the GroES–GroEL chaperonins (Greene et al., 2007).

Furthermore, an *E. coli* strain with glyceraldehyde-3-phosphate dehydrogenase (*gap*A) deletion was used to develop another Rubisco-dependent *E. coli* (RDE) selection system (Mueller-Cajar et al., 2007). Knockout of *gap*A gene leads to blockage of glycolysis and hence inhibits *E. coli* Δ*gap*A from using glucose as carbon source. Through co-expression of phosphoribulokinase and Rubisco, glycolytic blockage is bypassed by creating a metabolic bridge between ribulose-5-phosphate, produced from sugar substrates by pentose phosphate pathway above

23

blockage, and the glycolytic intermediate, PGA, below the blockage. Using Rubisco-dependent MM1 (ΔgapA)-PRK *E. coli* cells, a few *R. rubrum* Rubisco (Mueller-Cajar et al., 2007) and *Synechococcus* PCC6301 mutants (Mueller-Cajar and Whitney, 2008) were selected under PRK inducer concentration that inhibit growth of cell with wild-type Rubiscos. However, biochemical characterization show that improved fitness was due to improved folding and/or assembly instead of kinetic properties.

Higher transformation efficiency of *E. coli* (up to 10^{10} transformants/µg plasmid), at least three orders of magnitude greater than *R. capsulatus* (Smith and Tabita, 2003), enable high-throughput of selection. About 3×10^5 of colonies were screened each round (Parikh et al., 2006).

Other than selecting Rubisco variants with better kinetic properties, this approach provides insight into structure-function relationship of Rubisco (Smith and Tabita, 2004; Mueller-Cajar et al., 2007). Evolved *R. rubrum* with mutation at residue His-44 & Asp-117 led to reductions in Ω and k_c^{cat}/K_c. Sequence comparison pointed out that these two amino acid residues are conserved in Form II Rubisco but not in Form I Rubisco and structural analysis then reveal a hydrogen bond between them (Mueller-Cajar et al., 2007). Another *Synechococcus* PCC7002 Rubisco mutant M6-5 carries E49V and D82G in RbcS improved the specific carboxylation activity by 85% (Cai et al., 2014). Structural and kinetic analysis of the evolved variants revealed a previously unexplored conserved hydrogen bond

that is functionally linked to sustaining CO_2/O_2 specificity (Cai et al., 2014). Moreover, a few evolved *T. elongatus* BP1 Rubisco variants with RbcS mutations that cluster around the RbcL-RbcS interface displayed improved k_{cat}, k_{cat}/K_c and specificity (Wilson et al., 2018). This finding again showed that residues away from the catalytic site play role in enzyme kinetics (Wilson et al., 2018).

Wilson et al. (2016) demonstrated the first proof of concept that directed evolution of Rubisco in *E. coli* as a useful approach to evolve mutants with improvement in all the catalytic parameters and stimulate photosynthesis in leaf chloroplasts. By directing the evolution of non-photosynthetic *Methanococcoides burtonii* archael Rubisco (MbR) using MM1-prk RDE system, two mutants, MbR-E138V and MbR-K332E, with improved CO_2-fixation speed, CO_2-affinity and specificity for CO_2 were selected. When transplanted into a tobacco transplastomic line, these two mutants supported faster CO_2 assimilation rate and growth than wild-type MbR in tobacco (Wilson et al., 2016).

2.4   Chaperone incompatibility hinders heterologous expression of Rubisco

One challenge encountered by heterologous expression of foreign Rubisco is limited understanding of folding and assembly requirement for their complex biogenesis (Liu et al., 2010). Rubisco biogenesis is a multistep process that involves different chaperones for its folding and assembly (Kolesinski et al., 2014; Hauser et al., 2015; Wilson and Hayer-Hartl, 2018). Attempts of expressing Rubiscos in phylogenetically distant host result in formation of insoluble aggregates owing to

chaperone incompatibility (Gutteridge and Gatenby, 1995; Sharwood, 2017). For instances, plastid transformation of tobacco with non-green algal Rubiscos from *Galdieria sulphuraria*, diatom *Phaeodactylum tricornutum* and *G. monolis*, separately, demonstrated that both RbcL and RbcS subunits were successfully expressed in chloroplast but failed to form functional enzymes (Whitney et al., 2001; Lin and Hanson, 2018). Similarly, eukaryotic Rubiscos from maize, wheat (*Triticum aestivum*), tobacco (*Nicotiana tabacum*) and green alga *Chlamydomonas*, form insoluble aggregates in *E. coli* (Gatenby et al., 1987; Cloney et al., 1993; Whitney and Sharwood, 2007; Koay et al., 2016).

On the other hand, Form II Rubisco from *R. rubrum* and Form I cyanobacterial Rubisco successfully fold and assemble into functional Rubisco oligomers in tobacco and *E. coli* (Gatenby et al., 1985; Tabita and Small, 1985; Bradley et al., 1986; Goloubinoff et al., 1989b; Whitney and Andrews, 2001; Whitney and Sharwood, 2007; Lin et al., 2014). Besides, RbcL from sunflower and various *Flaveria* species Rubisco successfully folded and assembled with tobacco RbcS in tobacco chloroplast (Sharwood et al., 2008; Whitney et al., 2011a). Successful formation of holoenzymes indicates that their biogenesis requirement are satisfied in the host.

Recently, formation of soluble *Arabidopsis thaliana* and tobacco Rubiscos in *E. coli* were made possible with co-expression of their cognate chloroplast chaperones. Five chaperones, Cpn60/Cpn10/Cpn20, bundle sheath defective-2

(BSD2), Rubisco accumulation factors 1 (Raf1) and 2 (Raf2) and RbcX are required for their productive folding and assembly in *E. coli* (Aigner et al., 2017; Lin et al., 2019). These requirements of species-specific factor for biogenesis of plant Rubiscos in *E. coli* imply that their biogenesis is more complex than bacterial and cyanobacterial Rubiscos (Feiz et al., 2012). However, until now, biogenesis of more catalytically efficient red algae Rubisco is not understood and it has been shown that tobacco chloroplast lacks compatible factors for their biogenesis (Lin and Hanson, 2018).

2.5 Molecular chaperones mediate *de novo* protein folding and assembly

In cells, small proteins (fewer than 150 amino acids) exhibit two-state folding behaviour where they transit directly between non-native and native state (Fenton and Horwich, 2003). They are capable of efficient folding into native form *in vitro* when rapidly diluted from denaturant. On the other hand, large proteins (>150 amino acids) do not follow such folding behaviour. They, especially ones with multiple domains or with complex topology, tend to populate an ensemble of kinetically trapped meta-stable misfolded or partially folded intermediates (Fenton and Horwich, 2003; Lin and Rye, 2006). These folding intermediates have exposed hydrophobic residues and unstructured backbone that make them aggregation-prone, as they interact with each other through intermolecular hydrophobic interaction and inter-chain hydrogen bonding to form aggregates (Chaudhuri et al., 2009; Hartl and Hayer-hartl, 2009). To avoid misfolding and aggregation, they require assistance from a group of proteins called molecular chaperones for their

de novo folding to acquire native conformations (Hartl and Hayer-Hartl, 2002).

Molecular chaperones mediate correct folding by interacting with the exposed

hydrophobic region to stabilize and keep polypeptides from aggregation (Hayer-

Hartl et al., 2016). Subsequent of folding, another kind of chaperone named

assembly chaperone assists assembly of monomers into oligomer (Hartl and Hayer-

hartl, 2009; Bracher et al., 2011; 2015; Hauser et al., 2015). Different kinds of

chaperones present in cytosol, chloroplast and mitochondria to mediate folding of

nascent polypeptides and downstream assembly (Hayer-Hartl et al., 2016).

2.5.1   GroEL-GroES chaperonin in *E. coli*

Chaperonins are a family of molecular chaperone that have cavity to

accommodate substrate protein and allow folding in isolation (Hartl and Hayer-

Hartl, 2002). Bacterial GroEL chaperonin and its co-chaperonin GroES in *E. coli*

is the most-studied. GroEL protects non-native substrate proteins from irreversible

aggregation and improves the final yield of productive folding (Aoki et al., 2000).

GroEL is a tetradecamer (~800 kDa) consists of 14 identical subunits arranged as

two structurally identical heptameric rings, stacked back-to-back (Xu et al., 1997).

Each ring has a central cavity that is 45 Å in diameter and ~40 Å in height with

hydrophobic lining and house substrate protein up to 57 kDa (Chaudhuri et al., 2009;

Horwich, 2013).

GroEL subunit (57 kDa) consists of three domains: apical domain,

intermediate domain and equatorial domain (Fig. 2.3). Apical domain locates at the

entrance of cavity and carries substrate protein and co-chaperonin GroES binding sites, intermediate domain connects both apical and equatorial domain, which carry ATP binding site (Hayer-Hartl et al., 2016). GroEL works with co-chaperonin GroES, a single-ring heptamer of identical 10 kDa subunits, in ATP-dependent manner (Xu et al., 1997). Each GroES subunit folds into single domain and has a mobile loop (residues 17–33) that is unstructured in the isolated GroES but adopts a β-turn structure when forms contact with the GroEL apical domain (Xu et al., 1997). GroEL mediates folding of one substrate at a time in its cavity. Substrate protein binds to GroEL in a multivalent manner, which it binds to multiple apical domains in GroEL for efficient binding (Farr et al., 2000; Elad et al., 2007; Natesh et al., 2018).

GroEL rings that stacked together work alternatively. Reaction cycle starts with binding of ATP and polypeptide to one open GroEL ring, subsequent binding of GroES to GroEL seal the cavity and form a folding active asymmetric GroEL-GroES-(ATP)$_7$ complex (Fig. 2.4) (Horwich, 2013; Wilson and Hayer-Hartl, 2018). Binding of ATP and GroES trigger a rigid body movement of GroEL ring (referred as the *cis*-ring) that enlarges the volume of cavity 2-fold (from 85,000 Å^3 to 175,000 Å^3; diameter and height of cavity increased to ~80 Å and ~85 Å respectively) and changes the hydrophobic cavity lining into hydrophilic (Xu et al., 1997; Li et al., 2009). In the encapsulated cavity of *cis*-ring, folding of polypeptide occurs in a brief time (~2-7 s) while ATP hydrolyze. Binding of ATP and polypeptide to the opposite open GroEL ring (known as *trans*-ring) then induces the dissociation of

the *cis*-ring to discharge GroES, ADP and polypeptide, whether it is folded or not (Wilson and Hayer-Hartl, 2018). Protein that is not folded rebinds GroEL for further round of folding until it reaches native state. A new *cis*-ring forms when GroES binds to former *trans*-ring and allows folding of new polypeptide.

GroEL is able to differentiate its substrate protein in non-native and native state, which it no longer recognizes (Stan et al., 2006; Liu et al., 2010). *In vitro* study showed GroEL form stable binary complexes with substrate proteins in their non-native form but not native form (Viitanen et al., 1992). GroEL recognizes its substrate protein by the exposed hydrophobic surfaces in non-native state, whereas in native state, these hydrophobic regions are buried in the interior, therefore become inaccessible to GroEL (Stan et al., 2006; Horwich, 2013).
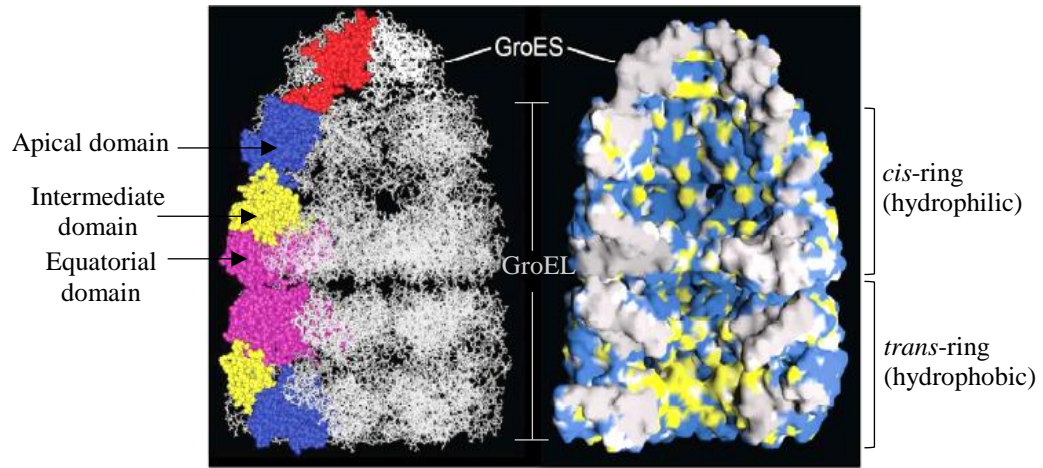
Figure 2.3: Structure of asymmetric GroELGroES-ADP₇ complex. Left: apical, intermediate and equatorial domains of two GroEL subunits in the *cis-* and *trans-* ring of the tetradecameric GroEL are colored in magenta, yellow and blue, respectively, and one GroES subunit is colored in red. Right: *Trans*-ring has a hydrophobic cavity lining whereas *cis*-ring has a hydrophilic lining. Solvent-excluded surfaces of subunit interface in gray, hydrophobic side-chains in yellow, and charged and polar side-chain in blue. Figure adopted from Wilson and Hayer-Hartl (2018) and amended here.
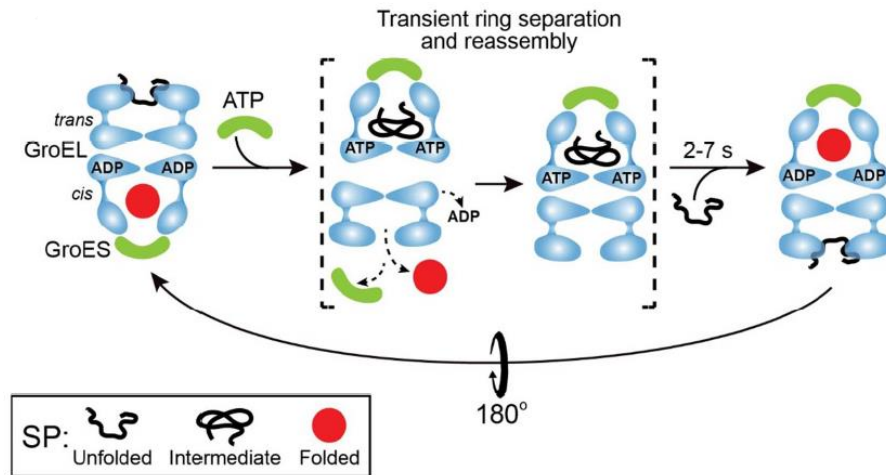
Figure 2.4: Reaction cycle of GroEL-GroES chaperonin. Substrate protein (SP) interacts with apical domains of *trans*-ring. Subsequent binding of ATP to SP-bound ring leads to changes of ring's conformation and triggers ADP and GroES to dissociate from former *cis*-ring. After that, GroES binds to the SP-bound ring, forming a new *cis*-ring. In the *cis*-ring, SP folds while ATP hydrolyze (2-7 s). Figure adopted from Wilson and Hayer-Hartl (2018).

2.5.2   Binding between GroEL and substrate protein

Mutagenesis of GroEL identified two helices, helices H (residues 230-244) and I (residues 256-271), and an underlying extended loop segment (residues 199–209) of cavity facing apical domain as polypeptide binding sites (Fenton et al., 1994). These three regions made up of mainly hydrophobic residues that form hydrophobic contact with the substrate protein. Studies have reported binding of short peptides to the groove between H and I (Chatellier et al., 1999; Li et al., 2009). Moreover, cryo-electron microscopy (cryo-EM) studies of GroEL-substrate complexes showed substrates form contact with these three regions (Elad et al., 2007; Natesh et al., 2018).

Although the polypeptide binding sites on GroEL cavity lining, and the nature of interaction between GroEL and its substrate are known, the exact way GroEL recognizes its substrate protein and GR factor carried by substrate protein are not elucidated (Chaudhuri and Gupta, 2005; Stan et al., 2005). GroEL is a promiscuous enzyme that interacts with 300 *E. coli* proteins diverse in the aspects of structures and functions (Houry et al., 1999; Li et al., 2009). The binding of substrate proteins to GroEL does not appear to be generally dependent on a particular amino acid sequence or specific structural motif (Lin and Rye, 2006).

Building on the idea that GroES is the best substrate of GroEL, hydrophobic segments of GroES mobile loop (GGIVLTG) were used as standard to look for mobile loop-like hydrophobic patches in the primary structure of *in vivo* and *in vitro*

substrates as potential GR sites (Chaudhuri and Gupta, 2005; Stan et al., 2005; 2006). It is suggested that one characteristic of GroEL substrate protein is presence of multiple hydrophobic patches for multivalent binding to GroEL apical domains to form stable GroEL-SP complex. In addition, these potential GroEL binding hydrophobic patches should be exposed in non-native state but buried in native state (Chaudhuri and Gupta, 2005; Stan et al., 2005; 2006). Chaudhuri and Gupta (2005) looked for mobile loop-like hydrophobic patches with approximate or higher grand average of hydropathicity (GRAVY) value than that of GroES mobile loop (GRAVY>1.5). On the other hand, Stan et al. (2005; 2006) reported that known GroEL substrate proteins have two to four mobile-loop like sequences as GroEL binding motifs, each separated from another by a ''spacer'' of at least 10 residues. However, the predicted binding motifs were not verified experimentally in these studies.

2.5.3   GroEL-GroES mediates folding of RbcL in *E. coli*

In *E. coli*, folding intermediate of Rubisco is highly aggregation prone and this off-pathway is kinetically favoured without assistance of GroEL chaperonin (Gutteridge and Gatenby, 1995). Dependence of Rubisco on GroEL-GroES chaperonin is evidenced by increased yield of holoenzyme with overexpression of GroEL and abolished production of holoenzyme by mutations in GroEL or GroES (Goloubinoff et al., 1989b). Furthermore, *in vitro* reconstitution of Form II dimeric Rubisco from *R. rubrum* requires all GroEL, GroES and Mg-ATP (Goloubinoff et al., 1989a; Lin and Rye, 2004). Besides, red type Form I Rubisco from

proteobacterium *Rhodobacter sphaeroides* can be folded by GroEL-GroES in *E. coli* as well (Joshi et al., 2015).

Form I cyanobacterial Rubiscos are able to form soluble holoenzyme in *E. coli*, however, in a small amount (2%) as most of the translated polypeptides form insoluble aggregate (Whitney and Sharwood, 2007; Mueller-Cajar and Whitney, 2008). Yield of soluble Rubisco (L_8S_8) can be increased 9-fold by overexpression of GroEL-GroES chaperonin complex, which indicates that level of soluble Rubisco in *E. coli* is limited by post-translational folding by GroEL-GroES (Greene et al., 2007). RbcL binds exclusively to GroEL in a 1:1 ratio to the *cis*-ring and following binding of co-chaperonin GroES allows RbcL monomer to fold in the encapsulated cavity of asymmetric GroEL-GroES-$(ATP)_7$ (van Duijn et al., 2006; 2007). Folded cyanobacterial RbcL subunits have partially disordered C-terminal tail (residues 413-475) that retains high affinity for binding of GroEL. The folded cyanobacterial RbcL then assemble into L_2, $(L_2)_4$ core and finally L_8S_8. In contrast, folded RbcL of *R. rubrum* able to reach a state no longer recognized by GroEL (Liu et al., 2010).

Assembly of L_2 and L_8 of cyanobacterial Rubisco is facilitated by another chaperone, RbcX. Co-expression of RbcX further increases the yield of soluble cyanobacterial L_8S_8 hexadecamer (Onizuka et al., 2004; Saschenbrecker et al., 2007). However, Rubisco of *Synechococcus* PCC7942 does not require RbcX for assembly (Emlyn-Jones et al., 2006). $RbcX_2$ is a homodimer of 15 kDa and

functions as molecular staple to stabilize the L_2 dimer and facilitate L_8 core formation. Hydrophobic cleft of $RbcX_2$ recognize sequence motif EIKFEF(E/D) on the flexible C-terminal of RbcL, at the same time, interact with N-terminal of another RbcL and form a RbcL-$RbcX_2$ complex. (Liu et al., 2010; Bracher et al., 2011; 2015). Upon formation of $RbcL_8$-$(RbcX_2)_8$, spontaneously folded RbcS displaces $RbcX_2$ and form L_8S_8.

2.5.4 Non-substrate protein become GroEL recognizable

Although GroEL chaperonin folds many proteins, it does not fold all. For instances, *E. coli* dihydrofolate reductase (DHFR) is not substrate protein of GroEL chaperonin. On the other hand, murine DHFR shows a strong interaction with GroEL. Interestingly, two *E. coli* DHFR mutants EcDHFR-i936 and EcDHFR-i7136 with residues 36 and 37 (L-N) and 136−139 (V-F-S-E) replaced by surface loop of murine DHFR respectively, become GroEL recognizable (Clark et al., 1996). It is not clear whether GroEL interacts directly with the amino acids in these loops or mutations in *E.coli* DHFR have resulted in structural changes that allow GroEL to interact with other regions of the protein (Clark et al., 1996).

2.6 Rationale behind this study

In the case of Rubisco, prokaryotic Rubiscos is hypothesized to be carrying GR region that is absent in eukaryotic Rubiscos for entry into GroEL mediated folding pathway. A few studies have reported single or a few amino acid mutations affect biogenesis of Rubiscos in tobacco and *E. coli* without affecting the steady-

state mRNA levels of RbcL and RbcS (Greene et al., 2007; Parry et al., 2013). As an attempt to pinpoint regions of cyanobacterial *Synechococcus* PCC6301 large subunit important for successful holoenzyme formation in *E. coli*, Koay et al. (2016) created chimeric Rubisco with 50 residues of *Synechococcus* RbcL swapped to their counterpart in eukaryotic RbcL from green algae *C. reinhardtii* sequentially. There are only 85 residues different between *Synechococcus* and *Chlamydomonas* RbcL (Appendix A). Yet, *Chlamydomonas* Rubisco failed to form holoenzyme in *E. coli*, indicating that their folding and/or assembly requirements are not satisfied in *E. coli*. One of the possible causes is *Chlamydomonas* RbcL is not recognized by GroEL chaperonin in *E. coli*, therefore, it misfolds and forms insoluble aggregates. When three stretches of 50 amino acids of *Synechococcus* PCC6301 Rubisco large subunit (i.e. residues 248-297, 348-397 and 398-447, parts of C-terminal domain) were replaced with counterparts in *Chlamydomonas* RbcL (residues 251-300, 351-400 and 401-450), polypeptides failed to assemble into functional structure, L_8S_8. This suggested that the replaced regions of *Synechococcus* RbcL might be important for interaction with *E. coli* GroEL chaperonin for proper folding and assembly (Koay et al., 2016). It is noteworthy that non-assembly could be caused by protein instability, which is imparted by the structural destabilizing effect of mutations introduced by *Chlamydomonas* counterpart. To narrow down the range, this study aims to examine the importance of 25 residues of each stretch (i.e. residues 248-272, 273-297, 348-372, 373-397, 398-422, and 423-447) for functional expression in *E. coli* by replacing them with corresponding residues in *Chlamydomonas* RbcL. If GR regions or residues are identified, it might be possible

to allow more eukaryotic Rubisco to form functional enzymes in *E. coli* by incorporating the recognition sequence, provided that no deleterious effect on the structural stability and functionality are exerted by these recognition residues.

CHAPTER 3

MATERIALS AND METHODS

3.1   Swapping of 25 amino acid residues of *Synechococcus elongatus* PCC6301 Rubisco to *Chlamydomonas reinhardtii* Rubisco large subunit counterpart

In previous study, when three stretches of 50 amino acids of *Synechococcus* PCC6301 Rubisco large subunit (RbcL), residues 248-297, 348-397 and 398-447 (parts of C-terminal domain), were replaced with counterpart in *Chlamydomonas* RbcL, residues 251-300, 351-400 and 401-450, polypeptides failed to assemble into functional structure, L_8S_8 (Koey et al., 2016). This finding suggested that the replaced regions of *Synechococcus* RbcL are important for interaction with *E. coli* GroEL chaperonin for proper folding and assembly. To narrow the range, this study examines the importance of 25 residues of each stretches, residues 248-272, 273-297, 348-372, 373-397, 398-422, 423-447 for functional expression in *E. coli* by replacing them with corresponding residues in *Chlamydomonas* RbcL. Therefore, six plasmids harbouring chimeric *Rbc*L-*Rbc*S operons coding for recombinant RbcL and wild-type *Synechococcus* RbcS were constructed. Figure 3.1 shows the schematic diagram of (A) wild-type *Chlamydomonas* RbcL coding sequence, (B) *Synechococcus Rbc*L-*Rbc*S operon and chimeric *Rbc*L-*Rbc*S operons constructed in this study, (C) pTrcSynL(Chl251-275)S, (D) pTrcSynL(Chl276-300)S, (E)

pTrcSynL(Ch351-375)S, (F) pTrcSynL(Chl376-400)S, (G) pTrcSynL(Chl401-425)S and (H) pTrcSynL(Chl426-450)S. Expression and assembly of chimeric Rubiscos in *E. coli* were then analysed to identify which regions are important.
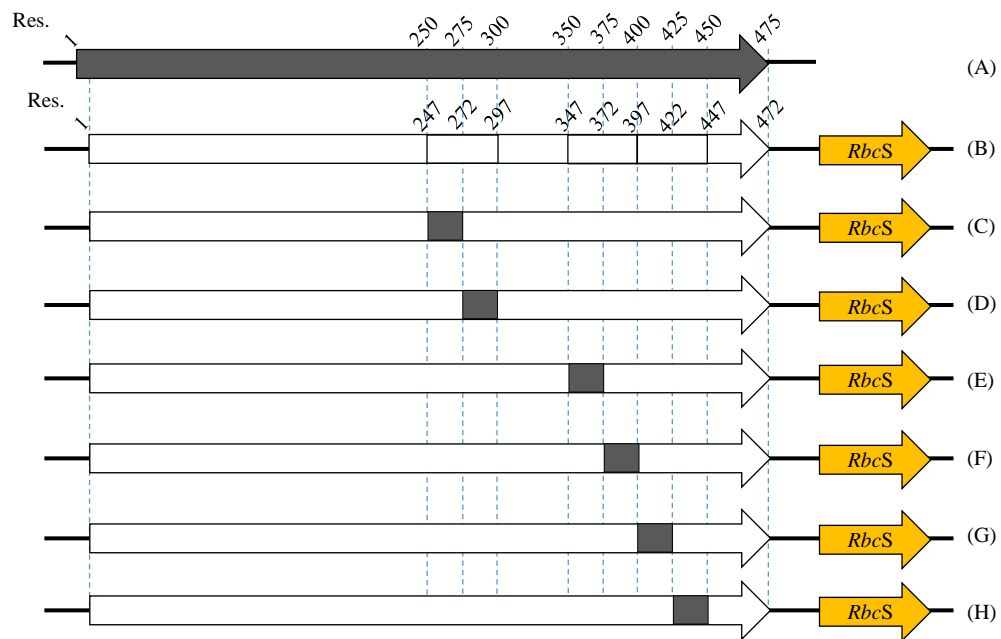
Figure 3.1: Schematic diagram of *Rbc*L-*Rbc*S operon. A) *Chlamydomonas* RbcL coding gene. (B) *Synechococcus* PCC6301 wild-type *Rbc*L-*Rbc*S operon in plasmid pTrcSynLS. Chimeric *Rbc*L-*Rbc*S operons with coding sequences for the 25 residues, (C) 248-272, (D) 273-297, (E) 348-372, (F) 373-397, (G) 398-422, (H) 423-447 replaced by counterpart in (A), respectively. Codon numbers (Res.) instead of nucleotide positions are indicated in the *Rbc*L genes.

3.1.1 Workflow for molecular cloning of chimeric *Rbc*L-*Rbc*S operon

All six plasmids harbouring recombinant *Rbc*L-*Rbc*S operon were constructed by similar workflow. For each chimeric operon, two fragments of genes of interests were amplified from respective templates. In order to join two fragments and vector backbone in correct orientation, different restriction site were added to primer pairs so that joining sites have complementary sticky ends after restriction digestion. Type IIS restriction enzyme, *BsmB*I, which cut at few bases after its recognition site was useful in this study for joining the two gene fragments without causing any unintentional changes to polypeptide sequences. Amplified gene fragments digested with restriction enzymes were then ligated into vector backbone, pTrcHisB of pTrcSynLS (Mueller-Cajar and Whitney, 2008). *E. coli* XL-1 Blue cells were then transformed with ligation mixture and plated on ampicillin agar plate for selection. Colonies grown on selective plate were screened to identify the positive transformants by colony PCR. After that, plasmids were extracted and sent for DNA sequencing.

3.1.2 Polymerase chain reaction (PCR) amplification of target genes fragments

For each constructs, two separate PCR reactions were carried out to amplify gene fragments from respective DNA templates. Each PCR reaction mixture, in final volume of 50 µL, contained 10 pg of DNA template, 1X ViBuffer A (100 mM Tris-HCl (pH 9.1), 500 mM KCl, 0.1% Triton™X-100), 1.5 mM MgCl$_2$, 0.1 µM of

primer pair, 0.2 mM dNTPs, and 2.5 u Vivantis *Pfu* DNA Polymerase. Amplification was conducted using condition presented in Table 3.1.

Table 3.1: PCR condition for gene fragments amplification.

Initial denaturation	95ºC, 3 min.	
Denaturation	95ºC, 30 sec.	
Annealing	55ºC, 30 sec.	25 cycles
Extension	72ºC, 2 min.	
Final extension	72ºC, 5 min.	
Hold	10ºC, 3 min.	

Table 3.2 shows the templates, primer pairs (sequences in Appendix B) for fragments amplification of each constructs and the expected amplicon size. Take construct pTrcSynL(Chl251-275)S as an example, primer pair SynN-BsmBI and Chl275-rev-BsmBI were used to amplify a gene fragment of 845 bp from template pTrcSynL(Chl251-300)S (Koay et al., 2016). Meanwhile, another primer pair, Syn275-fwd-BsmBI and SynSS-C-PstI amplified another fragment of 1075 bp from pTrcSynLS (Mueller-Cajar and Whitney, 2008).

PCR products were analysed by gel electrophoresis using 1% agarose gel and 1 kb DNA ladder for size determination. Amplicons with correct size were then purified using GeneJet PCR Purification Kit from Thermo Scientific.

Table 3.2: Templates and primer pairs used for amplification of gene fragments.

Construct	Template[a]	Primer pairs[b]	Fragment size (bp)
pTrcSynL(Chl251-275)S	pTrcSynL(Chl251-300)S	SynN-BsmBI & Chl275-rev-BsmBI	845 bp
	pTrcSynLS	Syn275-fwd-BsmBI & SynSS-C-PstI	1075 bp
pTrcSynL(Chl276-300)S	pTrcSynLS	SynN-BsmBI & Syn275-rev-BsmBI	853 bp
	pTrcSynL(Chl251-300)S	Chl275-fwd-BsmBI & SynSS-C-PstI	1071 bp
pTrcSynL(Chl351-375)S	pTrcSynL(Chl351-400)S	SynN-BsmBI & Chl375-rev-BsmBI	1159 bp
	pTrcSynLS	Syn375-fwd-BsmBI & SynSS-C-PstI	761bp
pTrcSynL(Chl376-400)S	pTrcSynLS	SynN-BsmBI & Syn375-rev-BsmBI	1156 bp
	pTrcSynL(Chl351-400)S	Chl375-fwd-BsmBI & SynSS-C-PstI	763bp
pTrcSynL(Chl401-425)S	pTrcSynL(Chl401-450)S	SynN-BsmBI & Chl425-rev-BsmBI	1294 bp
	pTrcSynLS	Syn425-fwd-BsmBI & SynSS-C-PstI	622 bp
pTrcSynL(Chl426-450)S	pTrcSynLS	SynN-BsmBI & Syn425-rev-BsmBI	1306 bp
	pTrcSynL(Chl401-450)S	Chl425-fwd-BsmBI & SynSS-C-PstI	618 bp

[a]pTrcSynLS is from Mueller-Cajar and Whitney (2008) while the other templates are from Koay et al. (2016).

[b]primer sequences are in Appendix B

3.1.3 Restriction digestion of amplicons and preparation of vector backbone

Three kinds of restriction enzymes, *Nco*I (Thermo Scientific), *BsmB*I (Thermo Scientific) and *Pst*I (Thermo Scientific) were used for digesting the PCR products and preparation of vector backbone. Reaction mixture contained 100 ng to 1 µg of DNA, 1X Tango buffer (33 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/mL BSA) and 10 U of enzyme in final volume of 30 µL. In the case of *BsmB*I, 1.0 mM of dithiothreitol (DTT) was added into the reaction mixture. Reaction mixture was incubated 37 °C for 1 hour and 30 minutes for restriction digestion.

Purified PCR products were either digested by *BsmB*I or double digested by *BsmB*I and *Pst*I, according to the restriction sites in the primer pair. After incubation, PCR products were purified using GeneJet PCR Purification Kit.

On the other hand, plasmid pTrcSynLS were digested with both *Nco*I and *Pst*I for backbone preparation. Digested pTrcSynLS was proceed to gel electrophoresis, followed by excision of vector backbone (4292 bp) and purification by Thermo Scientific GeneJet Gel Extraction Kit.

3.1.4 Ligation

Digested fragments were cloned into vector in a temperature-cycle fashion. The amount of vector and fragments used in each reaction followed a molar ratio of 1 (vector):3 (each of the two fragments). For each construct, 30 µL of ligation mixtures contained 100 ng vector, both digested fragments (amount follow molar ratio), 1X T4 DNA ligase buffer (50 mM Tric-HCl, 5 mM DTT, 10 mM MgCl$_2$, 1 mM ATP, pH 7.6) and 6 U NxGen™ T4 DNA Ligase. Ligation was conducted using a thermal-cycler under condition presented in Table 3.3.

Table 3.3: Ligation condition using thermal-cycler

Ligation Condition	Cycle
10 ℃, 1 min.	1
10 ℃, 30 sec. 30 ℃, 30 sec. 22 ℃, 5 min.	20
65 ℃, 3 min. 16 ℃, 1 min.	1

3.1.5 Preparation of electrocompetent cells and transformation via electroporation

Starter culture of XL-1 Blue *E. coli* was grown overnight in 10 mL of LB broth supplemented with 50 µg/mL tetracycline at 37 °C, 200 rpm. After overnight incubation, 3 mL of starter culture was introduced into a fresh 150 mL of LB broth containing 50 µg/mL tetracycline. The culture was then grown to OD_{600} of 0.5 at 37 °C, 200 rpm. Bacterial culture was then distributed evenly into three 50 mL conical tubes, and centrifuged at 5000 *g*, 4 °C for 10 minutes. Then, supernatant was discarded and cell pellet was re-suspended with 50 mL of pre-chilled 10% glycerol. Centrifugation step was repeated twice with 25 mL and 12.5 mL of 10% glycerol used for re-suspension. Cell suspensions were then pooled together, centrifuged, and the pellet was re-suspended in small volume of 10% glycerol. Electrocompetent cells were then distributed into aliquots of 50 µL, kept at -80 ℃ and ready to use for transformation. Electrocompetent cells were transformed with 1 µL of ligation mixture by electroporation at 1800 V using 0.1 cm of electrocuvette. Transformed cells was then cultured in 1 mL of LB broth and incubated at 37 °C, 200 rpm for 1 hour before plating on LB plate supplemented with 100 µg/mL ampicillin and subsequent incubation at 37 °C for 16 hours.

3.1.6 Colony screening by colony PCR

Colonies grown on the selective plate were selected randomly for colony PCR to screen for positive clones. Primer pairs, Trc3 and Trc5 (sequence in Appendix B), anneal to vector backbone and amplify the insert ligated into the

vector. Plasmid pTrcSynLS (Mueller-Cajar and Whitney, 2008) was used as positive control. Trc3 and Trc5 primers will amply a 1955 bp long fragment, which carries wild-type *Rbc*L-*Rbc*S operon, from it. Positive clones would show amplicon that is same size as positive control. Each PCR reaction mixture contained, 0.1 µM of each Trc3 and Trc5 primer, 0.2 mM dNTPs, single colony as DNA template, 1X complete KCl buffer (pH 8.8, 0.01% Tween 20, 1.5 mM $MgCl_2$) and 0.5 u DFS-Taq DNA Polymerase (Bioron) in final volume of 10 µl . Running condition is shown in Table 3.4.

Table 3.4: Colony PCR reaction condition

Initial denaturation	95ºC, 10 min.
Denaturation	95ºC, 30 sec.
Annealing	55ºC, 30 sec. ⎤ 25 cycles
Extension	72ºC, 2 min. ⎦
Hold	10ºC, 3 min.

3.1.7 Plasmid extraction and DNA sequencing

Plasmid of positive clones was extracted using Plasmid DNA Extraction Kit (Favorgen) and sent to MyTACG Sdn. Bhd. for DNA sequencing.

3.2   Expression of wild-type and chimeric Rubiscos in *E. coli* and protein extraction

All six correctly constructed plasmids were transformed into XL-1 Blue *E. coli* cells for expression and extraction. Besides, *E. coli* was transformed with plasmid pTrcSynLS (Mueller-Cajar and Whitney, 2008) to serve as positive control for following analysis. On the other hand, *E. coli* cells without plasmid served as negative control. Single colony was grown in 10 mL LB broth containing 100 µg/mL ampicillin at 37 °C, 200 rpm for 16 hours. Then, 3 mL of overnight culture was transferred into fresh 150mL LB broth with 100 µg/mL ampicillin and propagated to OD_{600} of 0.5 at 37 °C, 200 rpm. In the case of negative control, 50 µg/mL of tetracycline was used in place of ampicillin. Then, cultures were kept on ice for 1 hour prior to induction of Rubisco expression with 0.5 mM IPTG at 37 °C, 200 rpm for 16 hours. Next, induced cells were harvested by centrifugation at 5000 *g* for 5 minutes and pelleted cells were re-suspended in chilled extraction buffer (50 mM Bicine/NaOH (pH 8.0), 10 mM $NaHCO_3$, 2 mM DTT and 10 mM $MgCl_2$) to 10% (w/v) cells suspensions. Cells were then sonicated for 5 minutes, with 10 seconds pulses, at 40% amplitude to obtain crude lysates.

3.3   Expression and assembly analysis of Rubiscos by Polyacrylamide Gel Electrophoresis (PAGE)

Total cellular protein in the crude lysate was separated by SDS-PAGE and native-PAGE to check if chimeric Rubisco expressed and assembled in *E. coli*.

Vertical gel electrophoresis system (MV-10DSYS from Major Science) was used for handcasting polyacrylamide gels (1 mm thick) and running the electrophoresis.

3.3.1 Rubisco expression analysis by denaturing SDS-PAGE

Denaturing SDS-PAGE were performed to check expression of RbcL and RbcS in *E. coli*. Gel comprised 2 mL of 4% stacking gel (150 mM Tris-HCl, pH 8.45, 4% polyacrylamide/N,N'-methylene-bis-acrylamide (37.5:1), 0.1% SDS, 0.05% APS, and 0.1% TEMED) and 5 mL of 12% resolving gel (12% polyacrylamide/N,N'-methylene-bis-acrylamide (19:1), 13.6% sucrose, 0.1% SDS, 750 mM Tris-HCl, pH 8.45, 0.05% APS, and 0.1% TEMED). Running buffer used consists of 25 mM Tris, 192 mM glycine and 1% SDS. Mixture of crude lysate and loading buffer (100 mM DTT, 0.05% bromophenol blue, 5% SDS, and 30% sucrose) at 3:2 ratio was prepared and boiled for 5 minutes. Next, 4 µL of boiled sample mixtures and equal volume of pre-stained protein ladder were loaded into the gel for stacking and separation at 100 V, 15 mA until dye front reach the end of resolving gel. Resolving gel was then stained with Coomassie Brilliant Blue solution (0.25% Coomassie Brilliant Blue R-250, 50% methanol and 7% acetic acid) for 30 minutes with agitation. Stained gel was destained in distilled water overnight and then in destaining solution (40% methanol and 7% acetic acid) with agitation until protein bands can be seen clearly. Gel images were captured using Bio-Rad ChemiDoc™ Imaging System.

3.3.2 Rubisco assembly analysis by native-PAGE

Assembly of chimeric Rubisco in XL-1 Blue cells was checked by native gel electrophoresis with 4% stacking gel (130 mM Tris-HCl, pH 6.8, 4% polyacrylamide/N,N'-methylene-bis-acrylamide (37.5:1), 0.05% APS, and 0.1% TEMED) and 7.5% resolving gel (7.5% polyacrylamide/N,N'-methylene-bis-acrylamide (37.5:1), 375 mM Tris-HCl, pH 8.8, 0.05% APS, and 0.05% TEMED). Running buffer used consists of 25 mM Tris and 192 mM glycine. Cell lysates were mixed with sample loading buffer (50% glycerol, 150 mM Tris-HCl, and 0.25% bromophenol blue) at a 4:1 ratio and 25 µL of sample mixtures were loaded to the gel. Proteins were separated at 100 V, 15 mA for 6 hours. Resolving gel was then stained, destained and imaged as described in 3.3.1.

3.4 Western blot

To further validate expression and assembly of chimeric Rubiscos results obtained from protein gel electrophoresis, duplicate gels of SDS-PAGE and Native-PAGE were prepared for electro-blotting and chemiluminescent detection. Blotting was carried out in form of wet transfer using Electro Blot System (MEBM10 from Major Science), and Towbin buffer (25 mM Tris-HCl, 192 mM glycine and 20% methanol) as transfer buffer. Directly after electrophoresis, SDS-PAGE and Native-PAGE resolving gels were equilibrated in Towbin buffer and SDS-PAGE running buffer respectively, for 1 hour with agitation. On the other hand, nitrocellulose membranes (0.45 µm), filter papers and foam pads were soaked in transfer buffer for 10 minutes. Prior to transfer, transfer "sandwich" was assembled by placing

membrane and gel between few sheets of filter papers, followed by foam pads. Transfer sandwiches were then placed in the transfer tank, which had filled with pre-chilled transfer buffer and immersed in an ice bath. Next, separated proteins were transferred from gels to membranes at 30 V for 2 hours. After transfer, membranes were incubated in blocking solution (TBST buffer (150 mM NaCl, 20 mM Tris-HCl, pH 7.6 and 0.1% v/v Tween 20) with 5% v/v skim milk) for 1 hour with agitation. Blocked membranes were then proceeded to overnight probing with rabbit anti-*Synechococcus* PCC6301 Rubisco polyclonal antibodies (reconstituted to 74 mg/mL, diluted in 1:100,000 in blocking solution) at 4 °C. Next, membranes were washed with TBST buffer for 15 minutes three times. Washed membranes were then incubated with of Pierce ® Goat Anti-Rabbit IgG, horseradish peroxidase conjugated (reconstituted to 0.8 mg/mL, diluted to 1:6250 in TBST buffer) for 1 hour at room temperature. Following incubation, membranes were washed again for 15 minutes three times. At last, membranes were incubated with Pierce ® ECL Western Blotting Substrate (Thermo Scientific) for 2 minutes and signals were captured by Bio-Rad ChemiDoc^{TM} Imaging System.

3.5   Single amino acid mutation of *Synechococcus* RbcL

When comparing residues 348-397 of *Synechococcus* to *Chlamydomonas* RbcL, 10 different amino acids were found. These differences have led to non-assembly, thus, impacts of individual residues on folding and assembly were examined by creating mutant with single amino acid mutations.

3.5.1 Site-directed mutagenesis for single amino acid mutation

Mutations were integrated into primer pairs (Appendix C) and used to amplify entire plasmid, which carries *Synechococcus* PCC6301 wild-type *Rbc*L-*Rbc*S operon. Table 3.5 lists the mutants and their primer pairs. Each reaction mixture contains 1 ng of DNA template pTrcSynLS (Mueller-Cajar and Whitney, 2008), 1X ViBuffer S (160 mM $(NH_4)_2SO_4$, 500 mM Tris-HCl (pH 9.2), 17.5 mM $MgCl_2$ and 0.1% Triton^{TM}X-100), 0.4 mM dNTPs, 0.56 µM of forward and reverse primers and 2.5 u Vivantis *Pfu* DNA Polymerase in final volume of 25 µL. Amplification was performed according to Table 3.6.

Following amplification, reaction mixture was digested with restriction enzyme *Dpn*I at 37 °C overnight. Addition of *Dpn*I was to cleave the template plasmid pTrcSynLS at its methylated recognition site. *Dpn*I will not digest non-methylated PCR product. Digestion of plasmid pTrcSynLS reduces its transformation efficiency into *E. coli* in the following transformation. This then reduce the chances of selection of colonies transformed with pTrcSynLS instead of desired mutants, as they are all ampicillin resistant.

After *Dpn*I digestion, *E. coli* cells were then transformed with 1 µL of digested product and incubated at 37 °C for an hour. Next, cells were plated on LB agar plate containing 100 µg/mL ampicillin for selection, followed by 16 hours incubation at 37 °C. Colonies were selected randomly for plasmids extraction and extracted plasmids were sent for DNA sequencing.

Table 3.5: Single amino acid substitution in wild-type *Synechococcus* RbcL.

Mutation in RbcL	Primer pairs
Glutamic acid at 348 to aspartic acid (E348D)	Fwd-E348D, Rev-E348D
Histidine 350 to tyrosine (H350Y)	Fwd-H350Y, Rev-H350Y
Isoleucine 351 to valine (I351V)	Fwd-I351V, Rev-I351V
Alanine 352 to lysine (A353K)	Fwd-A353K, Rev-A353K
Valine 359 to isoleucine (V359I)	Fwd-V359I, Rev-V359I
Phenylalanine 360 to tyrosine (F360Y)	Fwd-F360Y, Rev-F360Y
Alanine 366 to cysteine (A366C)	Fwd-A366C, Rev-A366C
Leucine 372 to methionine (L372M)	Fwd-L372M, Rev-L372M
Serine 395 to alanine (S395A)	Fwd-S395A, Rev-S395A
Valine 396 to cysteine (V396C)	Fwd-V396C, Rev-V396C

Table 3.6: PCR reaction condition of site-directed mutagenesis.

Initial denaturation	95ºC, 30 sec.
Denaturation	95ºC, 1 min.
Annealing	55ºC, 1 min.
Extension	72ºC, 14 min.
Hold	10ºC, 3 min.

Denaturation, Annealing, Extension — 18 cycles

3.5.2 Expression and assembly analysis of single amino acid Rubisco mutant in *E. coli*

Plasmids harbouring the single amino acid Rubisco mutants were transformed into *E. coli*, respectively, and expression of mutants were induced by IPTG, followed by protein extraction (same method as 3.2). Expression and assembly of each mutants were then analysed by SDS-PAGE, native-PAGE and Western blot in the same way as chimeric Rubiscos (as described in 3.3.1, 3.3.2 and 3.4).

3.6 Bioinformatics analysis and predictions

Synechococcus and *Chlamydomonas* RbcL were screened for localized hydrophobic patches (GRAVY>1.5) of seven residues (Chaudhuri and Gupta, 2005) as potential GroEL binding sites. GRAVY values were calculated based on Kyte and Doolittle (1982) hydrophobicity scale. Besides, any loss of inter- or intra-subunit interaction in RbcL and with RbcS (distance within 4 Å) resulted by mutations were predicted using PyMOL^{TM} v0.99.

RESULTS

4.1 Molecular cloning of chimeric *Rbc*L-*Rbc*S operon

4.1.1 PCR amplification of target genes fragments

Amplicons were analysed by gel electrophoresis, using 1 kb DNA ladder as a reference for size estimation. Gel electrophoresis image (Fig. 4.1) shows that single band in expected size was amplified in each PCR reaction.
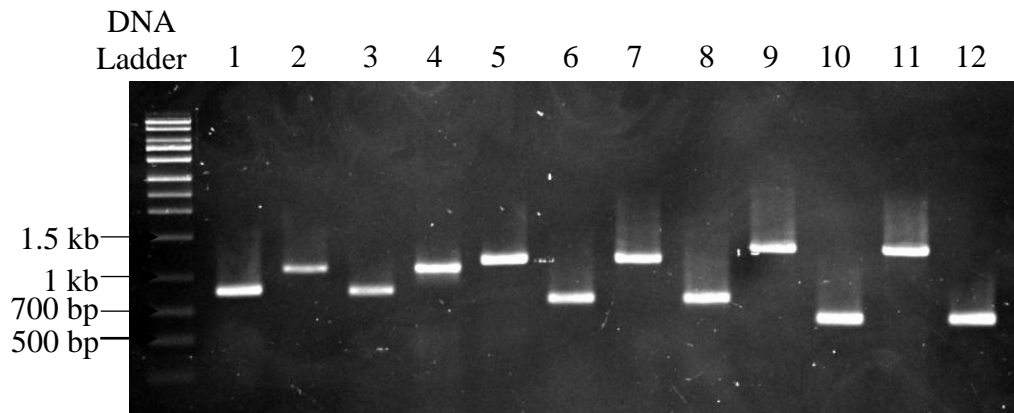
Figure 4.1: Agarose gel electrophoresis of PCR products required for constructing six chimeric operons. Lanes 1 and 2 are fragments for pTrcSynL(Chl251-275)S: 845 bp and 1075 bp. Lanes 3 and 4 are fragments for pTrcSynL(Chl276-300)S: 853 bp and 1071 bp. Lanes 5 and 6 are fragments for pTrcSynL(Chl351-375)S: 1159 bp and 761 bp. Lanes 7 and 8 are fragments for pTrcSynL(Chl376-400)S: 1156 bp and 763 bp. Lanes 9 and 10 are fragments for pTrcSynL(Chl401-425)S: 1294 bp and 622 bp. Lanes 11 and 12 are fragments for pTrcSynL(Chl426-450)S: 1306 bp and 618 bp. (Refer to Table 3.2)

4.1.2 Colony screening by colony PCR

Using construct pTrcSynL(Chl251-275)S as an example, Figure 4.2 shows the results of colony PCR using primers Trc3 and Trc5 (sequence in Appendix B) which flanks the whole *Rbc*L-*Rbc*S operon. Plasmid pTrcSynLS (Mueller-Cajar and Whitney, 2008) carrying the wild-type *Rbc*L-*Rbc*S operon (WT) served as positive control. Positive clones (C1-C2, C9-C12), showed single band of 1955 bp, same as positive control.
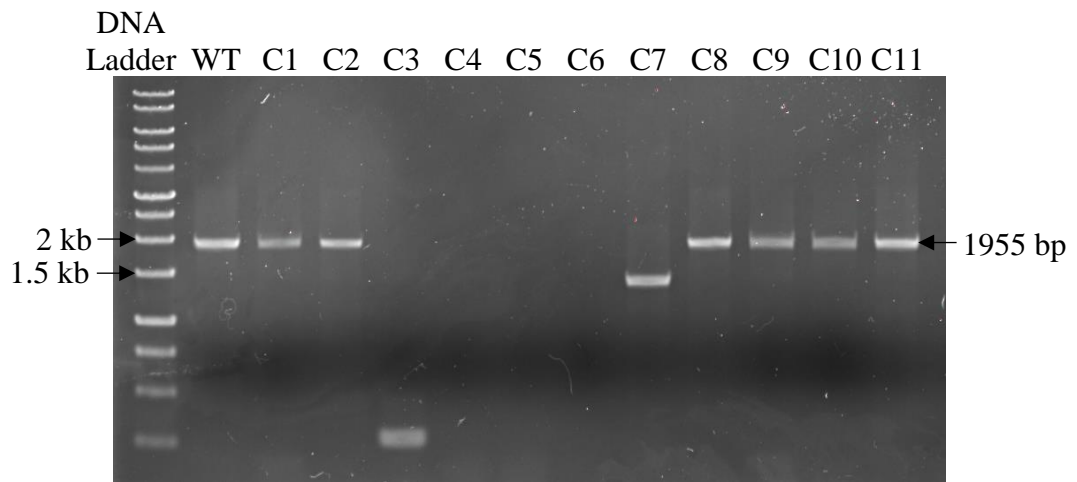

Figure 4.2: Screening of pTrcSynL(Chl251-275)S positive clones by colony PCR. Agarose gel electrophoresis of PCR products of positive control (WT) and 11 colonies (C1-C11).

4.2 Analysis of Expression and Assembly of chimeric Rubiscos (25-amino acid sectional swaps) in *E. coli*

4.2.1 Rubisco expression analysis by denaturing SDS-PAGE and Western Blot

Crude lysates from *E. coli* harbouring wild-type (from pTrcSynLS, as positive control), mutant Rubiscos, and no plasmid (as negative control) were denatured and analyzed by SDS-PAGE and Western blot. SDS-PAGE (Fig. 4.3A) showed that two bands, one slightly higher than 55 kDa while another one located between 10 kDa and 15 kDa, were observed for cells transformed with wild-type and mutant Rubiscos but absent in negative control. Western blot (Fig. 4.3B) analysis further validated that these two bands were large (LS) and small (SS) subunits and no signal was detected for negative control. Therefore, large and small subunits were successfully expressed in *E. coli*.

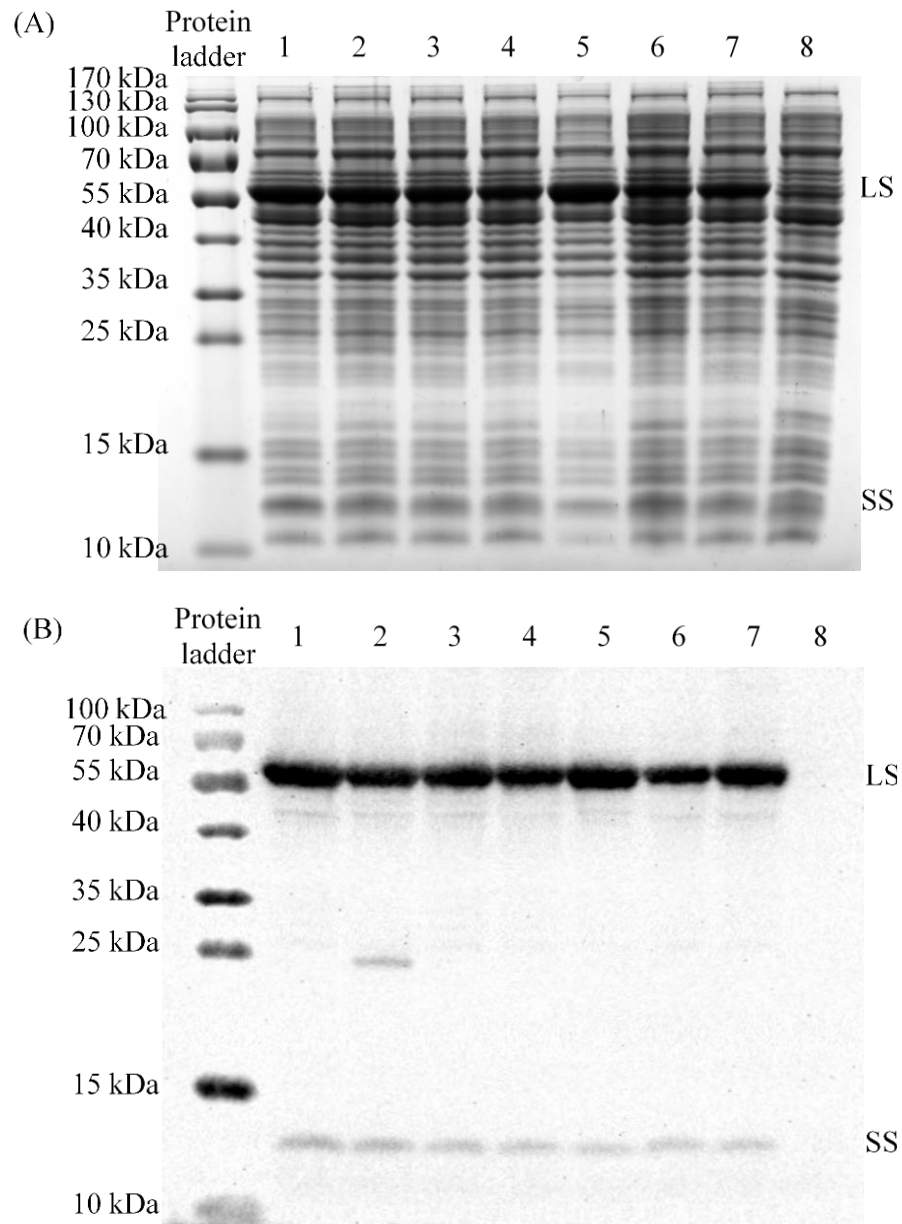**Figure 4.3: Rubiscos expression analysis in XL-1 Blue *E. coli*.** (A) SDS-PAGE and (B) Western blot analysis of Rubisco expression in *E. coli* transformed with (1) pTrcSynLS; (2) pTrcSynL(Chl251-275)S; (3) pTrcSynL(Chl276-300)S; (4) pTrcSynL(Chl351-375)S; (5) pTrcSynL(Chl376-400)S; (6) pTrcSynL(Chl401-425)S; (7) pTrcSynL(Chl426-450)S ; (8) no plasmid. LS and SS denote RbcL and RbcS respectively.

4.2.2   Rubisco assembly analysis by native-PAGE and Western Blot

Crude lysate of *E. coli* harbouring wild-type and Rubisco mutants, and negative control were subjected to native-PAGE and Western blot for assembly checking. Results from both analysis were consistent (Fig. 4.4). Assembly was observed for (1) positive control, (5) pTrcSynL(Chl376-400)S and (6) pTrcSynL(Chl401-425)S only, though there was reduction in holoenzyme for (5) pTrcSynL(Chl376-400)S.
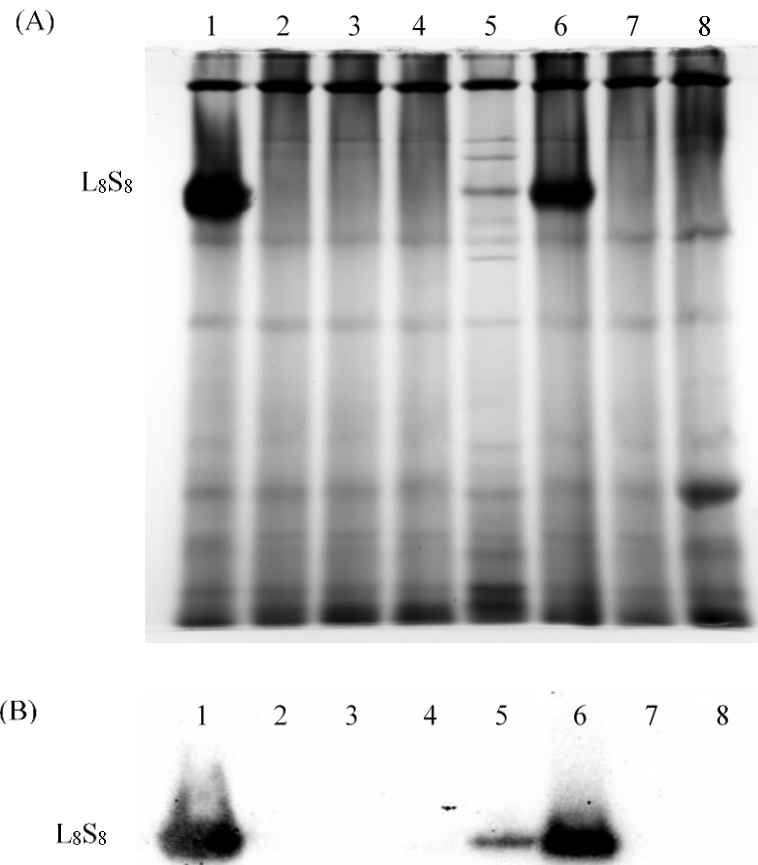
Figure 4.4: Rubiscos assembly analysis in XL-1 Blue *E. coli*. (A) Native-PAGE and (B) Western blot analysis of Rubisco assembly in *E. coli* transformed with (1) pTrcSynLS; (2) pTrcSynL(Chl251-275)S; (3) pTrcSynL(Chl276-300)S; (4) pTrcSynL(Chl351-375)S; (5) pTrcSynL(Chl376-400)S; (6) pTrcSynL(Chl401-425)S; (7) pTrcSynL(Chl426-450)S ; (8) no plasmid.

58

4.3 Expression and Assembly analysis of site-directed Rubisco mutants

4.3.1 Expression of site-directed mutant by SDS-PAGE and Western blot

SDS-PAGE (Fig. 4.5A) and Western Blot (Fig. 4.5B) analysis showed that both subunits were expressed in *E. coli* for all site-directed mutants.
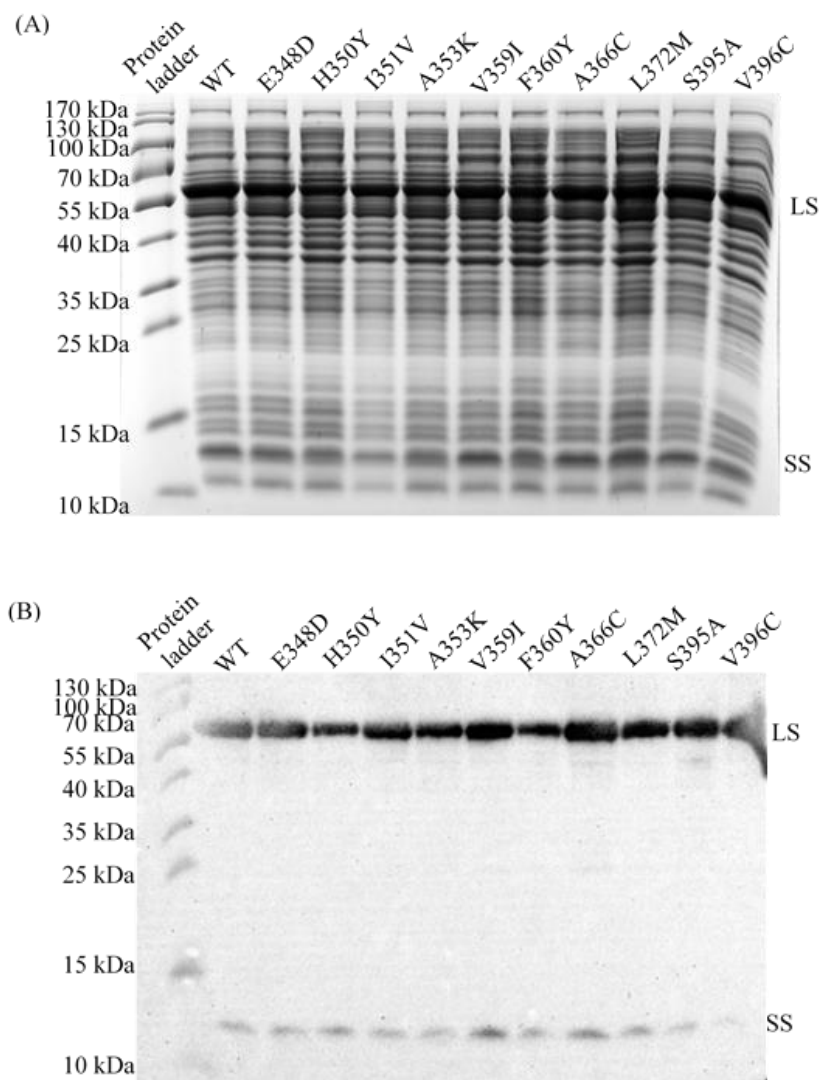
Figure 4.5: Analysis of expression of site-directed Rubisco mutants in XL-1 Blue *E. coli*. (A) SDS-PAGE and (B) Western blot analysis. LS and SS denote RbcL and RbcS respectively.

4.3.2 Assembly analysis by Native-PAGE and Western Blot

Both native-PAGE (Fig. 4.6A) and Western blot (Fig. 4.6B) showed all mutants assembled in *E. coli* though there was variation in the amount of holoenzyme among the mutants.

(A)

(B)
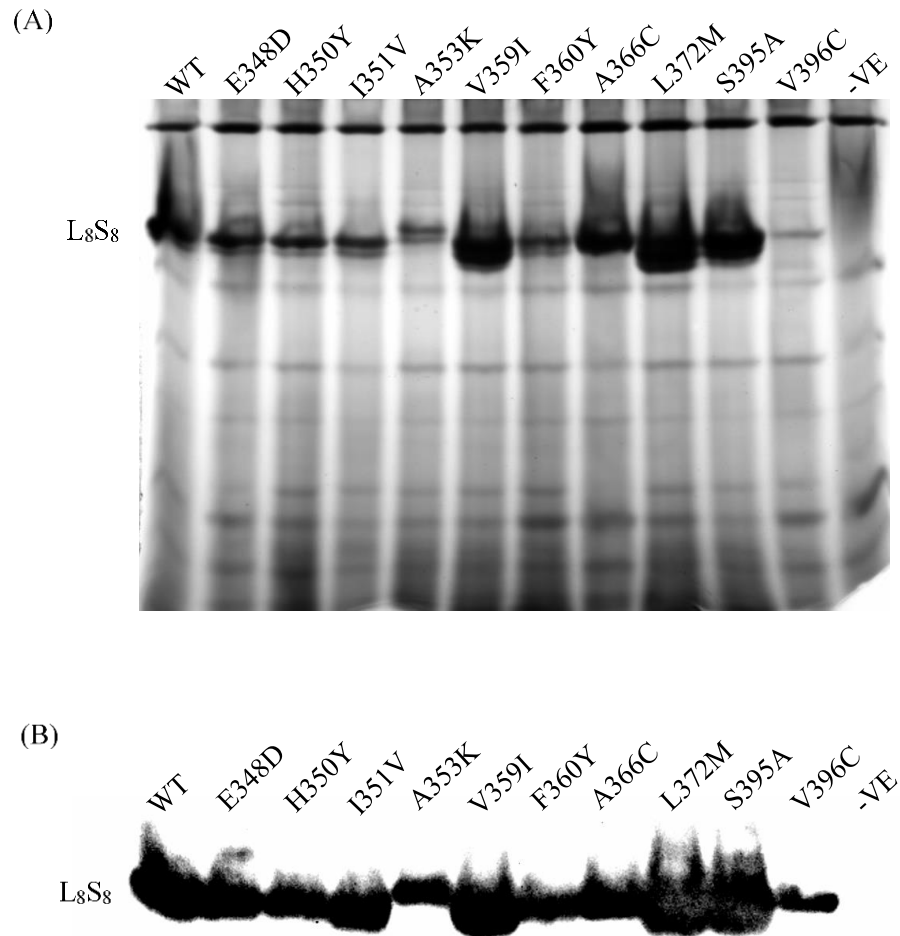
Figure 4.6: Assembly analysis of site-directed Rubisco mutants in XL-1 Blue *E. coli*. (A) Native-PAGE and (B) Western blot analysis. WT and –VE denote *E. coli* transformed with wild-type *Synechococcus* Rubisco (as positive control) and no plasmid (as negative control) respectively.

4.4   Screening of hydrophobic patches on *Synechococcus* PCC6301 and *Chlamydomonas* RbcL (GRAVY>1.5)

Table 4.1: Screening of localized hydrophobic patches (GRAVY>1.5) of seven residues along *Synechococcus* and *Chlamydomonas reinhardtii* RbcL.

Synechococcus PCC6301 RbcL			*Chlamydomonas* RbcL		
Residue Number	Patch sequence	GRAVY Value	[a]Residue Number	Patch sequence	GRAVY Value
257 - 263	LGMPIIM	2.085714	259 - 265	ELGVPII	1.642857
261 - 267	IIMHDFL	1.542857	260 - 266	LGVPIIM	2.414286
370 - 376	GVLPVAS	1.6	387 - 393	MPALVEI	1.585714
371 - 377	VLPVASG	1.6	388 - 394	PALVEIF	1.714286
384 - 390	MPALVEI	1.585714	389 - 395	ALVEIFG	1.885714
385 - 391	PALVEIF	1.714286			
386 - 392	ALVEIFG	1.885714			
419 - 425	VALEACV	2.114286			

[a]*Chlamydomonas* numbering is 3 amino acids more than *Synechococcus*

4.5 Loss of inter- and intra-subunit interaction in RbcL and with RbcS caused by mutations

Table 4.2: Loss of interaction (≤ 4 Å) in the respective chimeric mutants.

Mutation	[a]Loss of interaction (≤ 4 Å)				
pTrcSynL(Chl251-275)S					
E252V	W281				
F253C	W211	F215	R212	E256	L257
M259V	A219				
F266Y	-				
A269G	H264	D265	T268	G270	T272 (Another RbcL)
pTrcSynL(Chl276-300)S					
T276S	M247	L277	E245 (Another RbcL)		
K279I	Q146	T275			
W280Y	A251	E252	C281	N284	V286
V286L	I262				
pTrcSynL(Ch351-375)S					
E348D	K96 (RbcS)				
H350Y	I351	E352			
I351V	D344	L345	H350	Q363	W365
A353K	-				
V359I	-				
F360Y	-				
L372M	I152	E155	M166	H322	D394	S395
pTrcSynL(Chl376-400)S					
S395A	P165	F391	G392	V396	L397
V396C	S395				
pTrcSynL(Chl401-425)S					
T415A	G192	A411	E451	L6 (RbcS)	
pTrcSynL(Chl426-450)S					
V425T	-				
Y435A	R436				
I441V	A423	A427	R432	L434	E437	L442
L442I	V381	W382	P385	G438	
E444S	R432				
G446C	-				

[a]Loss of interaction is defined as residue contact distance more than 4 Å in *Chlamydomonas* compared to in *Synechococcus*.

CHAPTER 5

DISCUSSION

Heterologous expression of Rubiscos in *E. coli* is hampered by chaperone incompatibility where only Form II Rubisco from *R. rubrum* and Form I cyanobacterial Rubisco have been capable of folding and assembly into holoenzyme using GroEL-GroES chaperonin (Tabita and Small, 1985; Goloubinoff et al., 1989b; Parikh et al., 2006; Mueller-Cajar and Whitney, 2008). In contrast, eukaryotic Rubiscos from algae and plants form insoluble aggregates owing to chaperone incompatibility (Bradley et al., 1986; Gatenby et al., 1987; Whitney and Sharwood, 2007; Koay et al., 2016). Studies have shown evidences of GroEL-GroES chaperonin mediates folding of Rubisco large subunits (RbcL) in *E. coli*. For instances, *in vitro* reconstitution of *R. rubrum* and cyanobacterial Rubiscos demonstrated that folding of RbcL is dependent on GroEL-GroES chaperonin and these prokaryotic Rubiscos are recognized by GroEL chaperonin as stringent substrate proteins (Goloubinoff et al., 1989a; Lin and Rye, 2004; Liu et al., 2010). Furthermore, over-expression of GroEL-GroES substantially increases the level of soluble Rubiscos *in vivo* (Goloubinoff et al., 1989b; Mueller-Cajar and Whitney, 2008). Therefore, it is hypothesized that large subunits of Rubiscos from *R. rubrum* and cyanobacteria carry GR sequence while eukaryotic isoforms do not, hence, they misfold and aggregate.

In order to identify the potential GroEL binding sites on *Synechococcus* PCC6301 RbcL, six chimeric Rubiscos with stretches of 25 amino acid residues of RbcL replaced by counterpart in *Chlamydomonas* RbcL, which is presumed to be GroEL unrecognizable. If regions carrying GroEL binding site are lost, GroEL no longer recognizes RbcL of chimeric Rubiscos for their proper folding into tertiary structure, thus resulting in non-assembly. SDS-PAGE analysis showed both large and small subunits of chimeric Rubisco were expressed in *E. coli* (Fig. 4.3), thus, ensuring any observed non-assembly was not due to non-expression. Moreover, their expression levels were relatively similar (Fig. 4.3) Native-PAGE and western blot analysis revealed that swapping of residues 373-397 and 398-422 still allowed folding and assembly into chimeric Rubisco, while for the other chimeras, no holoenzyme were detectable (Fig. 4.4). Hence, GR sequences are most likely not located in these two regions (residues 373-397 and 398-422) or at least were not disrupted to significantly abrogate interaction with GroEL. Notably, when intensity of L_8S_8 in native-PAGE and Western blot are compared, Rubisco mutant from pTrcSynL(Chl376-400)S showed reduced amount of enzyme relative to wild type *Synechococcus* PCC6301 Rubisco (Fig. 4.4). On contrary, Rubisco from pTrcSynL(Chl401-425)S had comparable amount of assembled enzyme to wild-type (Fig. 4.4).

For the remaining four mutants, which are from pTrcSynL(Chl251-275)S, pTrcSynL(Chl276-300)S, pTrcSynL(Ch351-375)S, and pTrcSynL(Chl426-450)S, which showed no formation of holoenzyme (Fig. 4.4), these help to narrow down

the potential GroEL binding sites on *Synechococcus* RbcL to residues 248-297, 348-372 and 423-447. It may be possible that at least three GroEL binding sites are located in these three discrete regions, given the nature of multivalent interaction between GroEL chaperonin and its substrate protein (Farr et al., 2000; Chen et al., 2013). As a stringent substrate of GroEL, Rubiscos bind to a minimum of three consecutive apical domains of a GroEL ring for efficient binding (Farr et al., 2000). Binding of Rubisco to GroEL mutant with less than three consecutive binding-competent apical domain resulted in significant reduced amount of binary complex of Rubisco and GroEL mutant, as compared to the wild-type GroEL (Farr et al., 2000). Consistent with this finding, cryo-electron microscopy (cryo-EM) of GroEL-RbcL complex showed C-terminal domain of RbcL folding intermediate in contact with three consecutive apical domains while N-terminal in contact with one apical domain (Natesh et al., 2018). Therefore, it could be possible that the loss of any favourable interactions of these three regions (residues 248-297, 348-372 and 423-447), which are parts of the C-terminal domain, disrupted consecutive binding of RbcL to apical domains. Consequently, binding between GroEL and RbcL, if not abolished, may be greatly reduced to an extent that RbcL captured by GroEL folded and assembled but at undetectable amount of L_8S_8.

Another possible cause for non-assembly is mutations introduced by the *Chlamydomonas* counterpart, instead of resulting in loss of GR site, disrupt intra-subunit interaction or even exert disturbance to tertiary structure. Therefore, despite captured by GroEL chaperonin, chimeric RbcL monomer folded into disordered

structure that was unable to assemble into final hexadecamer. Moreover, as there are three states of assembly, L_2, $(L_2)_4$ and L_8S_8, mutations that affect the inter-subunit interactions of any of these assembled complex could eventually lead to non-assembly. For mutants shown non-assembly, *Chlamydomonas* counterparts introduced four to eight different residues. Table 4.2 shows potential loss of residual interaction (within 4 Å) upon mutation. These mutations can contribute to beneficial, neutral or deleterious effect (Guo et al., 2004). Only one of all mutations involved in non-assembly is located at interface between RbcL and RbcS: E348D created by swapping of residues 348-397 to 351-400 of *Chlamydomonas*, results in replacement of carboxyethyl group of Glu-348 by a shorter carboxymethyl group of Asp, which might disrupt salt-bridge with K96 in RbcS (Koay et al., 2016).

As for potential interaction with GroEL, GroEL recognizes substrate protein by exposed hydrophobic regions in their non-native state, and several studies have attempted to identify potential GroEL binding sites on GroEL substrates by sequence approach (Chaudhuri and Gupta, 2005; Stan et al., 2005). Considering that mobile loop of co-chaperonin GroES and substrate proteins bind to same binding site on GroEL apical domain, hydrophobic segments of GroES mobile loop (GGIVLTG) (GRAVY:1.5143) was used as standard to look for mobile loop-like hydrophobic patches on substrate protein as potential GroEL binding site. Besides, it is believed that GroEL substrate proteins possess multiple such hydrophobic patches to interact with multiple GroEL apical domains. Therefore, residues 248-

297, 348-397 and 398-447 of *Synechococcus* RbcL were screened for potential GR

sites with GRAVY value of more than 1.5 (Table 4.1; Fig. 5.1).

Among residues 248-272 of *Synechococcus* RbcL, hydrophobic patch

consisting of residues 257-267 was identified as potential GroEL binding site

(Table 4.1; Fig. 5.1A). Swapping of residues 248-272 to residues 251-275 of

Chlamydomonas RbcL generated Rubisco mutant (from pTrcSynL(Chl251-275)S)

with five mutations relative to wild-type *Synechococcus*. Two of the mutations,

M259V and F266Y, affect the patch pattern (Table 4.1; Fig. 5.1A). In fact, M259T

has shown to elevate the amount of *Synechococcus* PCC6301 Rubisco holoenzyme

formed in *E. coli* (Greene et al., 2007). Both M259T and M259V influence the

GRAVY value but do not lead to loss of hydrophobic patch (GRAVY>1.5) whereas

F266Y reduces the GRAVY value (from 1.54 to 1.12), though the hydrophobic

patch is still there, but shortened to residue 256-263 (Fig. 5.1A). Meanwhile,

mutation A269G has been reported to improve the fitness of Rubisco dependent *E.

coli* (Mueller-Cajar and Whitney, 2008). Therefore, non-assembly is more likely

caused by remaining two mutations, E252V and/or F253C, which show potential

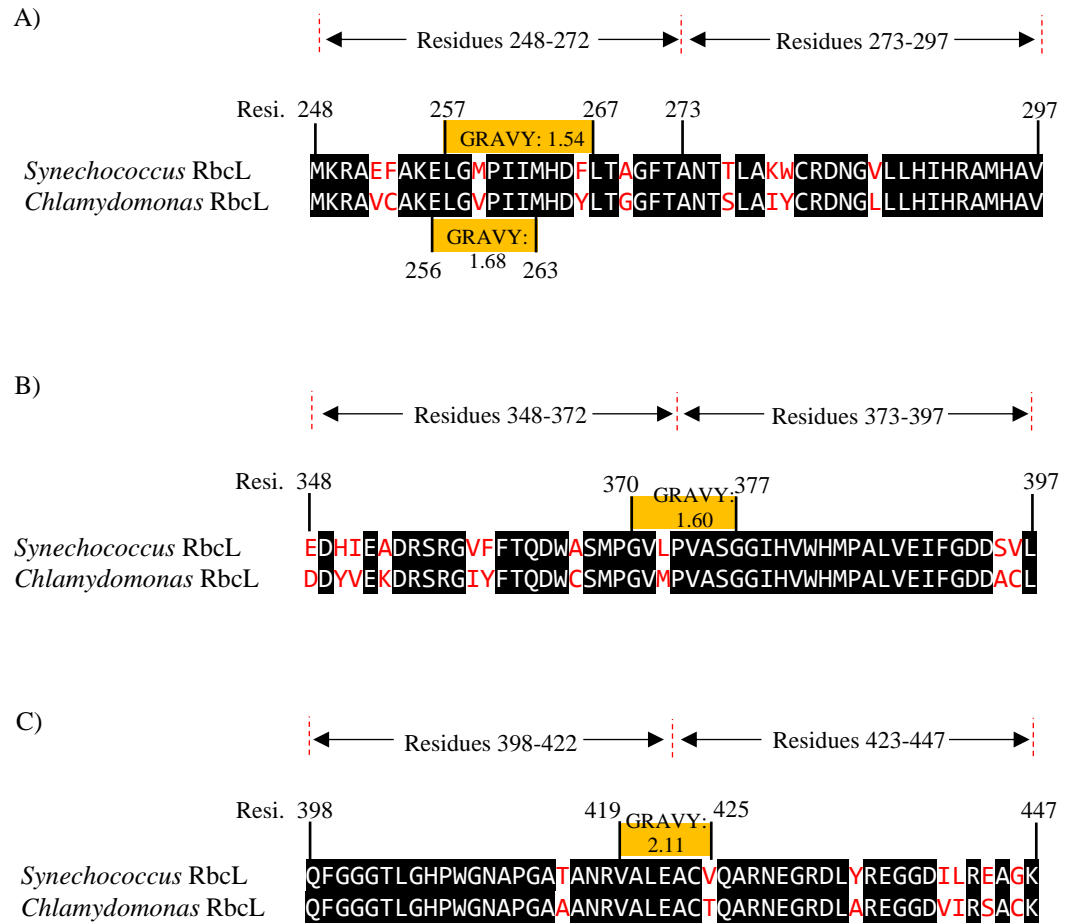loss of interaction with multiple residues within the RbcL monomer (Fig. 5.2).

Figure 5.1: Hydrophbic patches (GRAVY>1.5) of *Synechococcus* and *Chlamydomonas* RbcL. Hydrophbic patches found on RbcL as potential GroEL binding sites. Numbering based on *Synechococcus* RbcL.
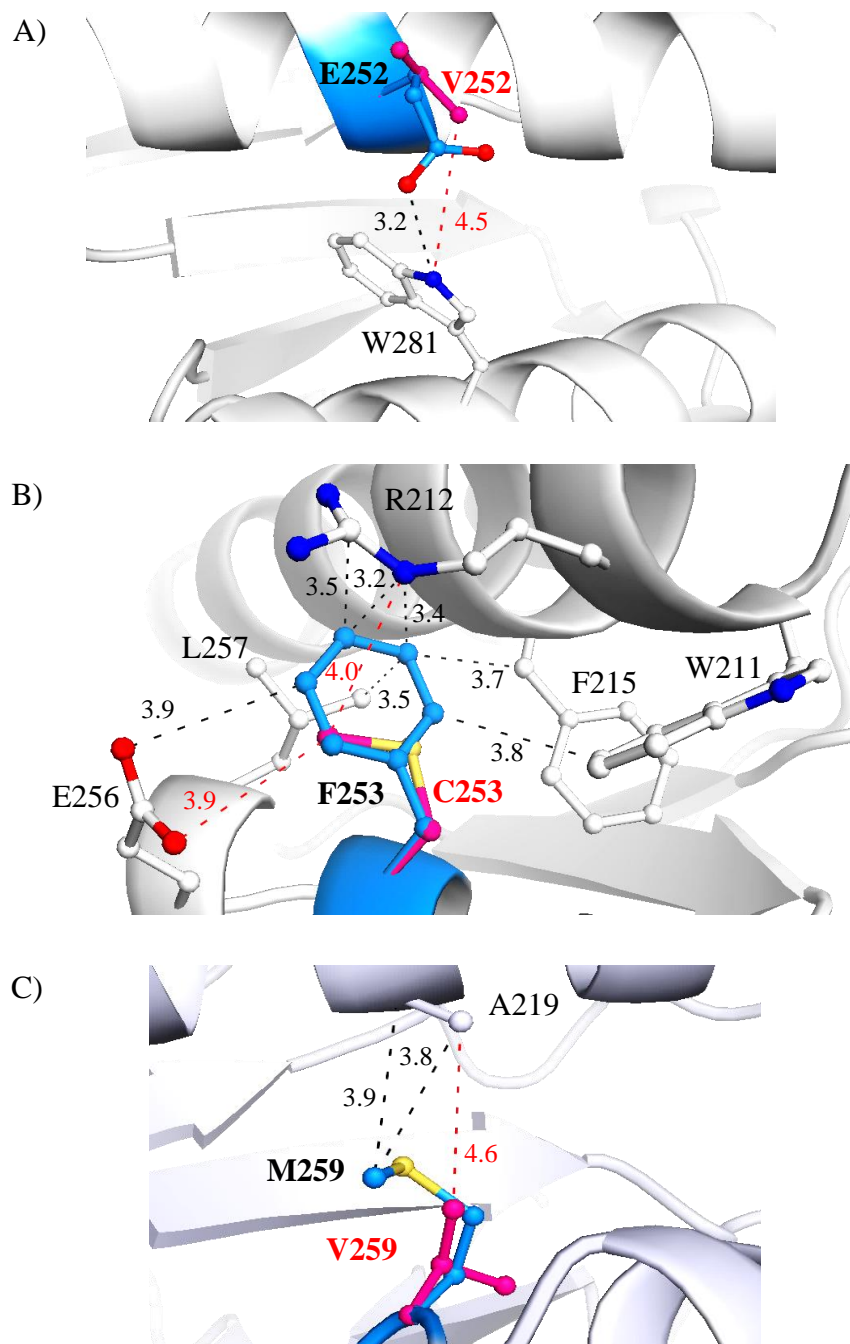
Figure 5.2: Potential loss of interactions introduced by swapping of residues 248-272 of *Synechococcus* RbcL. A) E252V, B) F253C and C) M259V lead to loss of interaction to another residue (within 4 Å). *Synechococcus* Rubisco (PDB ID 1RBL) was superimposed on *Chlamydomonas* Rubisco (1GK8). Carbon atoms of mutated *Synechococcus* residues are coloured *marine*. Carbon atoms of their corresponding *Chlamydomonas* residues are coloured *hot pink*. Nitrogen atoms are *blue*, oxygen atoms *red*, and sulphur atoms *gold*. Also shown are *Synechococcus* residues (carbon atoms *light grey*), which have potentially weakened/lost interaction in the mutants (the smallest distance measurements are shown between these residues and the mutated residues).

On the other hand, no GroES mobile-loop like patch was identified in residues 273-297 (Table 4.1; Fig. 5.1A). Swapping of this region resulted in Rubisco mutant (from pTrcSynL(Chl276-300)S) with four mutations, T276S, K279I, W280Y and V286L. Potentially, non-assembly was due to destabilizing effect of these mutations instead of loss of GroEL binding site (Fig. 5.3).
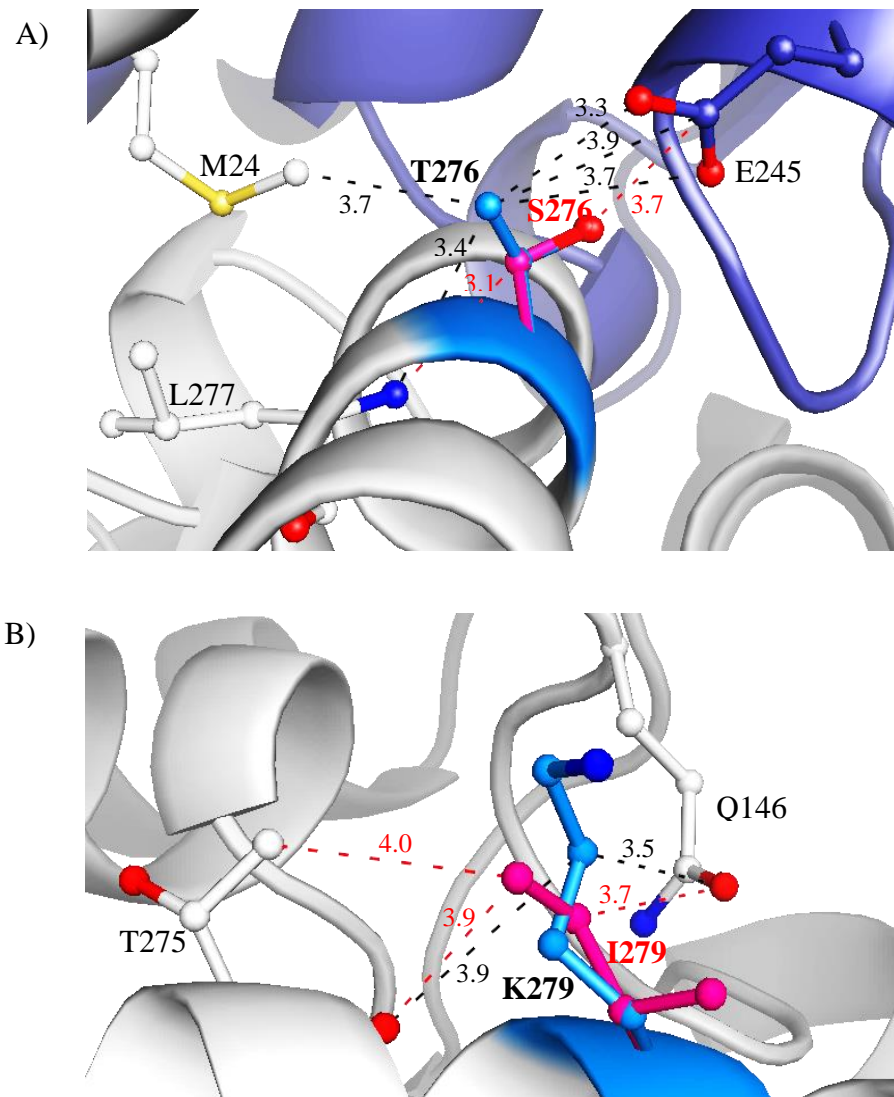
Figure 5.3: Potential loss of interactions introduced by swapping of residues 273-297 of *Synechococcus* RbcL. A) T276S, B) K279I, C) W280Y and D) V286L lead to loss of interaction to another residue (within 4 Å). *Synechococcus* Rubisco (PDB ID 1RBL) was superimposed on *Chlamydomonas* Rubisco (1GK8). Carbon atoms of mutated *Synechococcus* residues are coloured *marine*. Carbon atoms of their corresponding *Chlamydomonas* residues are coloured *hot pink*. Nitrogen atoms are *blue*, oxygen atoms *red*, and sulphur atoms *gold*. Also shown are *Synechococcus* residues (carbon atoms *light grey*), which have potentially weakened/lost interaction in the mutants (the smallest distance measurements are shown between these residues and the mutated residues).

On the other hand, swapping of residues 348-372 give rise to eight mutations. Of which, L372M could result in the loss of a hydrophobic patch of residues 370-377 (Table 4.1; Fig. 5.1B). As for residues 398-447, one hydrophobic patch of residues 419-425 was identified (Table 4.1; Fig. 5.1C). Swapping of residues 423-447 (mutant pTrcSynL(Chl425-450)S) resulted in six mutations. V425T significantly decrease the hydrophobicity (GRAVY value from 2.11 to 1.41) and lead to loss of the hydrophobic patch (Fig 5.1C). Four other mutations potentially lead to non-assembly through loss of interactions, which impart structural destabilization (Fig. 5.4). One of these mutations, E444S, could cause loss of intra-subunit salt-bridge with R432 in the RbcL monomer (Fig. 5.4D).
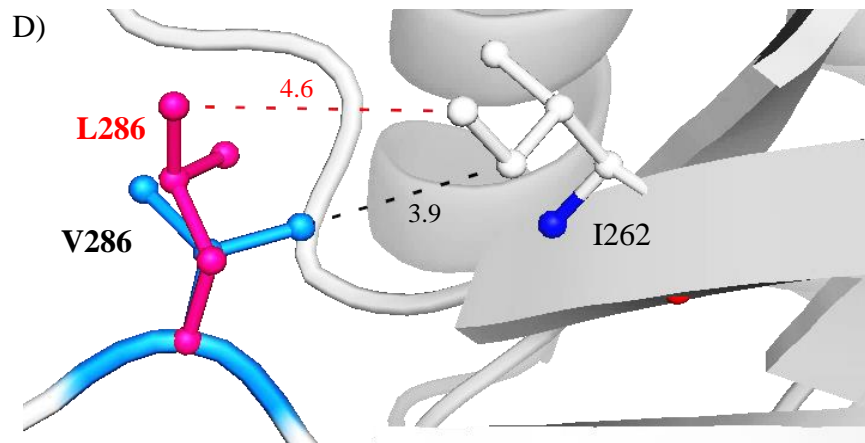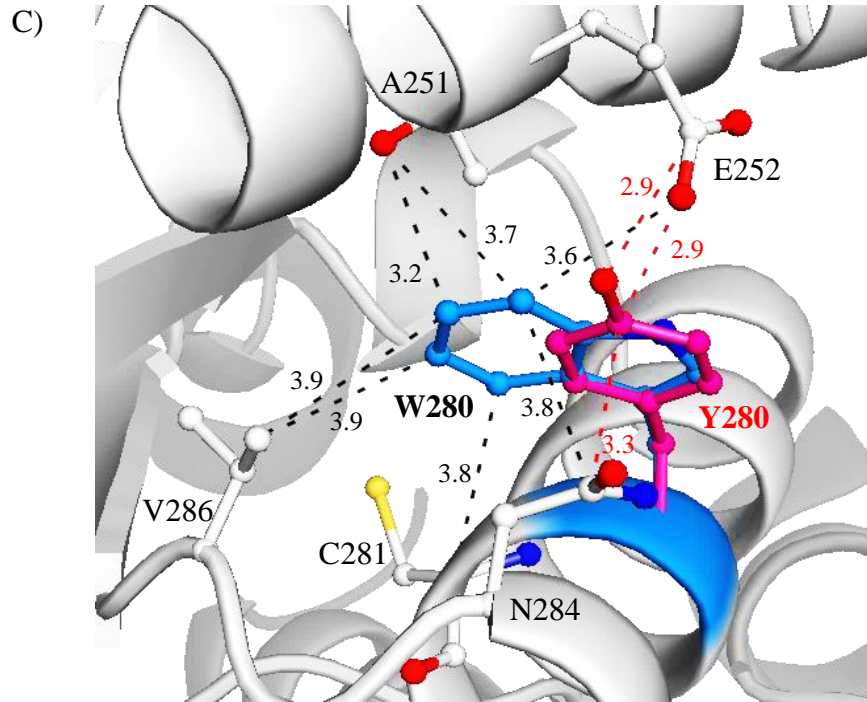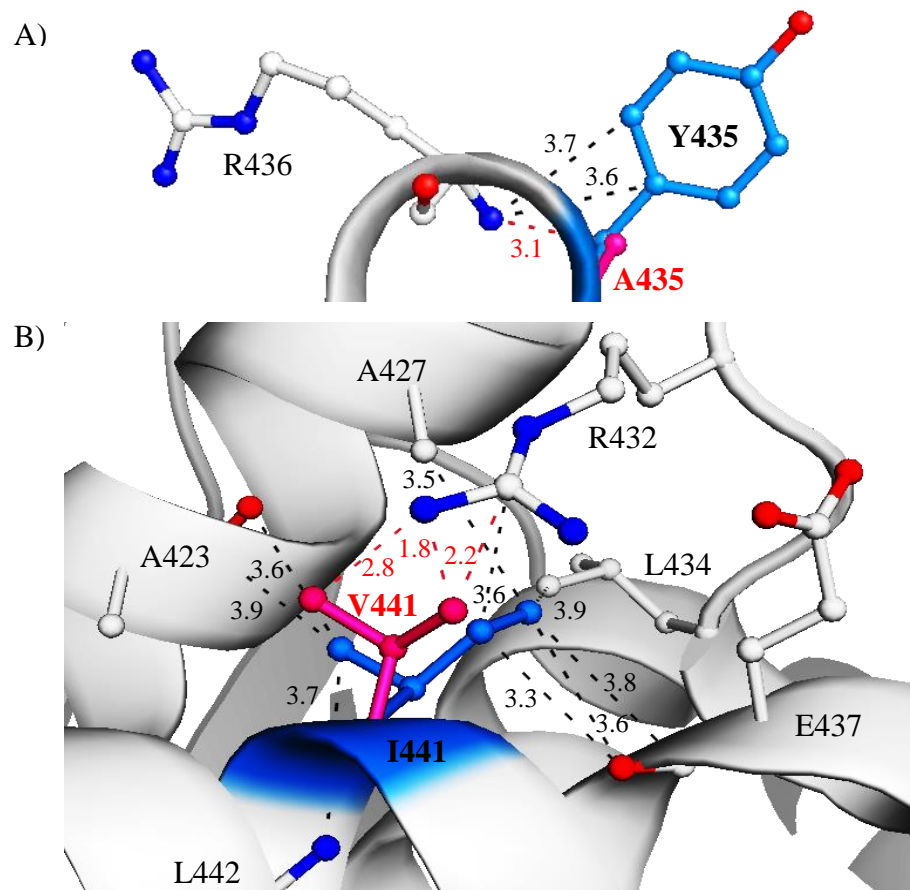
C)

D)
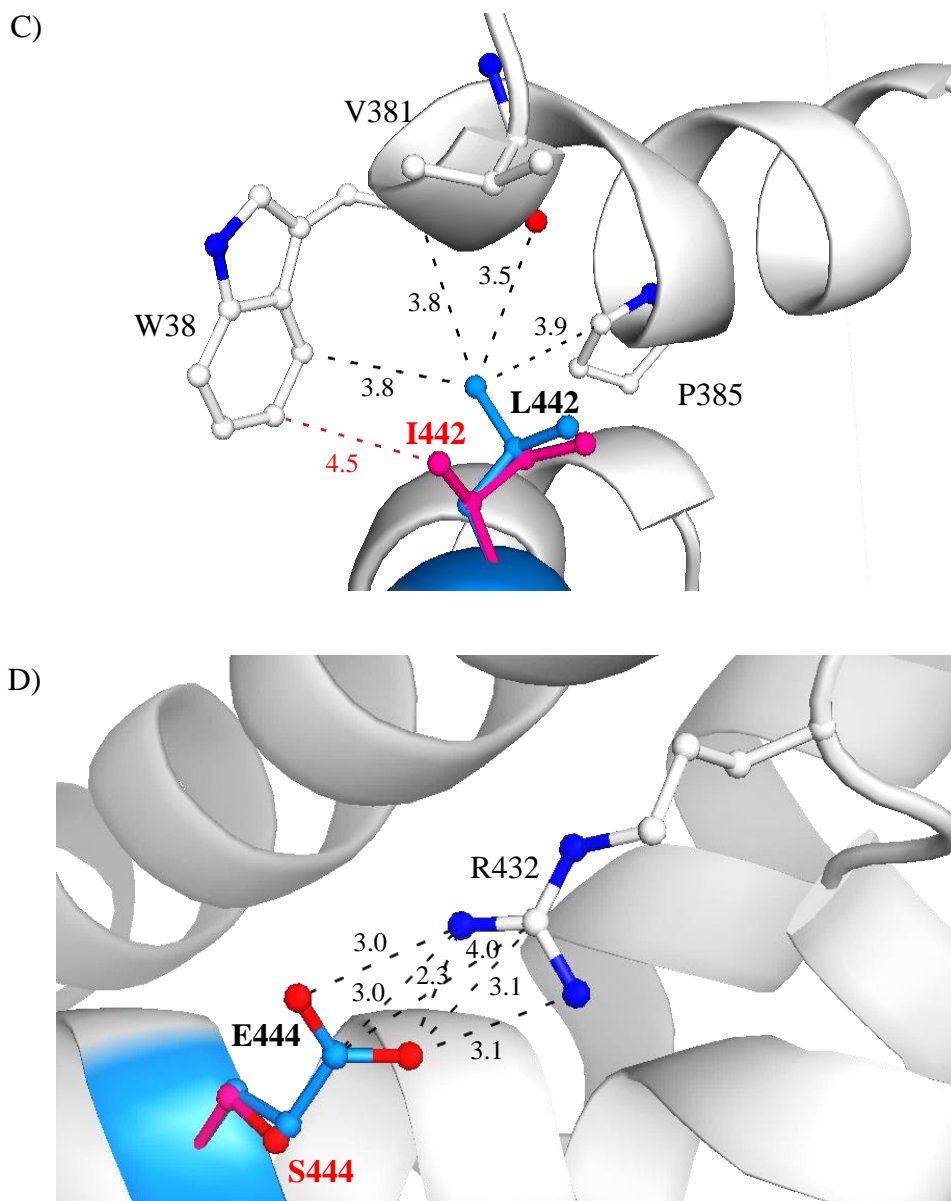
Figure 5.4: Potential loss of interactions introduced by swapping of residues 423-447 of *Synechococcus* RbcL. A) Y435A, B) I441V, C) L442I and D) E444S lead to loss of interaction to another residue (within 4 Å). *Synechococcus* Rubisco (PDB ID 1RBL) was superimposed on *Chlamydomonas* Rubisco (1GK8). Carbon atoms of mutated *Synechococcus* residues are coloured *marine*. Carbon atoms of their corresponding *Chlamydomonas* residues are coloured *hot pink*. Nitrogen atoms are *blue*, oxygen atoms *red*, and sulphur atoms *gold*. Also shown are *Synechococcus* residues (carbon atoms *light grey*), which have potentially weakened/lost interaction in the mutants (the smallest distance measurements are shown between these residues and the mutated residues).

In order to examine the independent effect of mutations on the functional expression, ten Rubisco mutants with single residue mutation were created by site-directed mutagenesis. Again, SDS-PAGE analysis show both large and small subunits of Rubisco mutants were expressed in *E. coli* (Fig. 4.5). In the sectional swaps, when two of these ten mutations, S395A and V396C, were made together, there was reduced holoenzyme formation (Fig. 4.4; mutant pTrcSynL(Chl376-400)S). Native-Page and Western blot analysis of site-directed mutants showed that S395A had little effect on the amount of Rubisco holoenzyme but V396C showed reduced amount, as compared to wild-type Rubisco (Fig. 4.6). Therefore, V396C accounts for reduced holoenzyme formation of mutant with swapped residues 373-397. Referring to crystal structure of wild-type Rubisco (PDB ID 1RBL), V396 is in the core of folded monomer. The exact reason V396C caused reduced amount of assembled enzyme is unknown. It is possible that this mutation influenced holoenzyme thermal stability negatively, as have been reported before for other single mutants in RbcL (Du et al., 2000; Genkov et al., 2006; Genkov and Spreitzer, 2009). The level of mRNA expression might be a factor, but SDS PAGE analysis indicates that V396C has similar amount of subunit to wild type (Fig. 4.5). Indeed, previous studies also found single mutations in RbcL that could affect its protein expression in *E. coli* without affecting its steady-state level mRNA (Parikh et al., 2006; Greene et al., 2007; Mueller-Cajar and Whitney, 2008).

Another eight single mutations were constructed to find out mutations responsible for non-assembly when residues 348-372 were swapped. Interestingly,

all single mutations permitted assembly into L_8S_8 complexes (Fig. 4.6). As mentioned previously, L372M would result in loss of the predicted GroEL binding site. However, it showed increased amount of assembled L_8S_8 to wild-type Rubisco whereas four other mutations, E348D, H350Y, I351V and A353K, showed reduced amount. Reduced amount of L_8S_8 observed caused by E348D could be due to disruption of salt-bridge with K96 in small subunit as mentioned previously (Fig. 5.5). Hence, non-assembly of Rubisco mutant from pTrcSynL(Chl351-375)S could be attributed to collective impact of these four mutations.
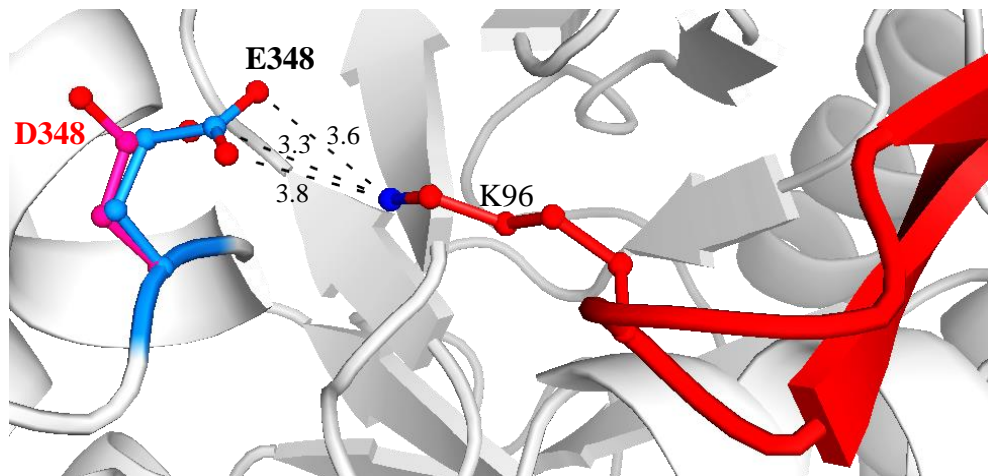
Figure 5.5: Potential loss of interaction (within 4 Å) resulted from E348D. *Synechococcus* Rubisco (PDB ID 1RBL) was superimposed on *Chlamydomonas* Rubisco (1GK8). Carbon atoms of mutated *Synechococcus* residues are coloured *marine*. Carbon atoms of their corresponding *Chlamydomonas* residues are coloured *hot pink*. Nitrogen atoms are *blue*, oxygen atoms *red*, and sulphur atoms *gold*. Also shown are *Synechococcus* residues (carbon atoms *light grey*), which have potentially weakened/lost interaction in the mutants (the smallest distance measurements are shown between these residues and the mutated residues).

Future work

Although GR sites on *Synechococcus* PCC6301 RbcL are not directly identified, potential regions carrying them are narrowed down to residues 248-297, 348-372 and 423-447. As non-assembly caused by mutations in these swapped regions could be due to loss of GR site or destabilizing effect on structure, future work could aim to elucidate residue(s) truly contributing to GroEL binding via biochemical assays. One method is *in vitro* proteinase K protection experiment (Liu et al., 2010). This method could give insights if non-assembly of the four sectional mutants is due to denied entry into the GroEL-GroES chaperonin. In this method, denatured RbcL of chimeric Rubisco will be first incubated with GroEL and a non-hydrolysable ATP analogue, β-ϒ-imidoadenosine 5'-phosphate (AMP-PNP, which prevents dissociation of GroES and substrate from cis-ring), followed by addition of GroES and proteinase K into the mixture for further incubation. If the RbcL mutant can bind to GroEL-GroES and be captured, it will be proteinase K-protected and the RbcL subunit will be observed in subsequent SDS-PAGE analysis of the mixture, and non-assembly can then be attributed to the destabilizing effect of the mutations instead of loss of GroEL interaction. In contrast, if the RbcL mutant cannot bind with GroEL, it will be digested by proteinase K, indicating that the non-assembly is caused by loss of GroEL interaction. After which, short peptides of the mutated regions can be generated, and incubated with GroEL *in vitro* for NMR and chemical shift analysis to determine the GroEL binding sites (e.g. Yagi-Utsumi et al., 2013).

CHAPTER 6

CONCLUSION

Six regions of *Synechococcus* RbcL were examined for their importance for holoenzyme formation in *E. coli* by swapping to their counterparts in *Chlamydomonas* RbcL. Non-assembly resulted from swapping indicates that GroEL binding sites might reside in the mutated regions of *Synechococcus* RbcL. Native-PAGE and Western blot analysis revealed that residues 373-397 and 398-422 are not essential for GR as it did not affect assembly of Rubisco in *E. coli*. Meanwhile, swapping of residues 248-272, 273-297, 348-372 and 423-447 of *Synechococcus* RbcL led to non-assembly. Hence, they potentially carry GroEL binding sites. Using a sequence-based approach, GroES mobile loop-like hydrophobic patches were found in residues 248-272, 348-372 and 423-447 respectively. Mutations introduced by swapping would either disrupt the hydrophobic pattern or lead to loss of patches. However, non-assembly could be due to structural destabilizing effect of mutations other than loss of GroEL binding site. Eight mutations introduced by swapping of residues 348-372 resulted in non-assembly. Therefore, eight Rubisco mutants with single mutations were constructed to check individual impact of mutations. Surprisingly, all eight mutants managed to assemble in *E. coli*. Mutation L372M was predicted to result in loss of the predicted GroEL binding site and contribute to non-assembly when residues 348-

372 were swapped. Yet, L372M single mutation showed no impact on the assembly. Meanwhile, four mutations, E348D, H350Y, I351V and A353K, showed reduced amount of assembled enzyme. In relation to non-assembly resulted from swapping of residues 348-372, these mutations might be responsible when introduced together, presumably by destabilizing the enzyme structure or binding of RbcL to GroEL. However, true reasons for non-assembly require further examination. Nonetheless, potential GroEL binding sites on *Synechococcus* RbcL are narrowed down to residues 248-297, 348-372 and 423-447.

REFERENCES

Aigner, H. et al., 2017. Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science*, 358(6368), pp. 1272–1278.

Andersson, I., 2008. Catalysis and regulation in Rubisco. *Journal of Experimental Botany*, 59(7), pp. 1555–1568.

Andersson, I. and Backlund, A., 2008. Structure and function of Rubisco. *Plant Physiology and Biochemistry*, 46(3), pp. 275–291.

Andrews, T.J. and Whitney, S.M., 2003. Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants. *Archives of Biochemistry and Biophysics*, 414(2), pp. 159–169.

Aoki, K. et al., 2000. GroEL binds artificial proteins with random sequences. *Journal of Biological Chemistry*, 275(18), pp. 13755–13758.

Bauwe, H., Hagemann, M. and Fernie, A.R., 2010. Photorespiration: players, partners and origin. *Trends in Plant Science*, 15(6), pp. 330–336.

Betti, M. et al., 2016. Manipulating photorespiration to increase plant productivity: Recent advances and perspectives for crop improvement. *Journal of Experimental Botany*, 67(10), pp. 2977–2988.

Boyer, R., 2006. *Concepts in biochemistry*, 3rd ed. New Jersey: Wiley.

Bracher, A. et al., 2015. Structural analysis of the rubisco-assembly chaperone RbcX-II from *Chlamydomonas reinhardtii*. *PLoS ONE*, 10(8), pp. 1–17.

Bracher, A., Starling-Windhof, A., Hartl, F.U. and Hayer-Hartl, M., 2011. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nature Structural and Molecular Biology*, 18(8), pp. 875–880.

Bradley, D., Van Der Vies, S.M. and Gatenby, A.A., 1986. Expression of cyanobacterial and higher-plant ribulose 1,5-bisphosphate carboxylase genes in *Escherichia coli. Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 313(1162), pp. 447–458.

Cai, Z., Liu, G., Zhang, J. and Li, Y., 2014. Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. *Protein and Cell*, 5(7), pp. 552–562.

Carmo-Silva, E., Scales, J.C., Madgwick, P.J. and Parry, M.A.J., 2015. Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant, Cell and Environment*, 38(9), pp. 1817–1832.

Chatellier, J., Buckle, A.M. and Fersht, A.R., 1999. GroEL recognises sequential and non-sequential linear structural motifs compatible with extended β-strands and α-helices. *Journal of Molecular Biology*, 292(1), pp. 163–172.

Chaudhuri, T.K. and Gupta, P., 2005. Factors governing the substrate recognition by GroEL chaperone: A sequence correlation approach. *Cell Stress and Chaperones*, 10(1), pp. 24–36.

Chaudhuri, T.K., Verma, V.K. and Maheshwari, A., 2009. GroEL assisted folding of large polypeptide substrates in *Escherichia coli*: Present scenario and assignments for the future. *Progress in Biophysics and Molecular Biology*, 99(1), pp. 42–50.

Chen, D.H. et al., 2013. Visualizing GroEL/ES in the act of encapsulating a folding protein. *Cell*, 153(6), pp. 1354–1365.

Chen, Z. and Spreitzer, R.J., 1992. How various factors influence the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Photosynthesis Research*, 31(2), pp. 157–164.

Clark, A.C., Hugo, E. and Frieden, C., 1996. Determination of regions in the dihydrofolate reductase structure that interact with the molecular chaperonin GroEL. *Biochemistry*, 35(18), pp. 5893–5901.

Cloney, L.P., Bekkaoui, D.R. and Hemmingsen, S.M., 1993. Co-expression of plastid chaperonin genes and a synthetic plant Rubisco operon in *Escherichia coli*. *Plant Molecular Biology*, 23(6), pp. 1285–1290.

Du, Y.C., Hong, S. and Spreitzer, R.J., 2000. RbcS suppressor mutations improve the thermal stability and CO_2/O_2 specificity of rbcL-mutant ribulose-1,5-bisphosphate carboxylase/oxygenase. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26), pp. 14206–14211.

Elad, N. et al., 2007. Topologies of a Substrate Protein Bound to the Chaperonin GroEL. *Molecular Cell*, 26(3), pp. 415–426.

Ellis, R.J., 1979. The most abundant protein in the world. *Trends in Biochemical Sciences*, 4(11), pp. 241–244.

Emlyn-Jones, D., Woodger, F.J., Price, G.D. and Whitney, S.M., 2006. RbcX can function as a Rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant and Cell Physiology*, 47(12), pp. 1630–1640.

Evans, J.R., 2013. Improving photosynthesis. *Plant Physiology*, 162(4), pp. 1780–1793.

Farr, G.W. et al., 2000. Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell*, 100(5), pp. 561–573.

Feiz, L. et al., 2012. Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation Factor1 is required for holoenzyme assembly in maize. *The Plant Cell*, 24(8), pp. 3435–3446.

Fenton, W.A. and Horwich, A.L., 2003. Chaperonin-mediated protein folding: Fate of substrate polypeptide. *Quarterly Reviews of Biophysics*, 36(2), pp. 229–256.

Fenton, W.A., Kashi, Y., Furtak, K. and Norwich, A.L., 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, 371(6498), pp. 614–619.

Gatenby, A.A., Van der Vies, S.M. and Bradley, D., 1985. Assembly in *E. coli* of a functional multi-subunit ribulose bisphosphate carboxylase from a blue-green alga. *Nature*, 314(6012), pp. 617–620.

Gatenby, A.A., van der Vies, S.M. and Rothstein, S.J., 1987. Co-expression of both the maize large and wheat small subunit genes of ribulose-bisphosphate carboxylase in *Escherichia coli*. *European Journal of Biochemistry*, 168(1), pp. 227–231.

Genkov, T., Du, Y.C. and 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(2), pp. 167–174.

Genkov, T., Meyer, M., Griffiths, H. and 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(26), pp. 19833–19841.

Genkov, T. and Spreitzer, R.J., 2009. Highly conserved small subunit residues influence Rubisco large subunit catalysis. *Journal of Biological Chemistry*, 284(44), pp. 30105–30112.

Goloubinoff, P., Christeller, J.T., Gatenby, A.A. and Lorimer, G.H., 1989a. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature*, 342(6252), pp. 884–889.

Goloubinoff, P., Gatenby, A.A. and Lorimer, G.H., 1989b. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature*, 337(6202), pp. 44–47.

Greene, D.N., Whitney, S.M. and Matsumura, I., 2007. Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. *Biochemical Journal*, 404(3), pp. 517–524.

Gutteridge, S. and Gatenby, A.A., 1995. Rubisco synthesis, assembly, mechanism, and regulation. *Plant Cell*, 7(7), pp. 809–819.

Hartl, F.U. and Hayer-hartl, M., 2009. Converging concepts of protein folding *in vitro* and *in vivo*. *Nature structural & molecular biolog*y, 16(6), pp. 574–582.

Hartl, F.U. and Hayer-Hartl, M., 2002. Protein folding. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science*, 295(5561), pp. 1852–1858.

Hauser, T., Popilka, L., Hartl, F.U. and Hayer-hartl, M., 2015. Role of auxiliary proteins in Rubisco biogenesis and function. *Nature plants*, 1(6), pp. 1–11.

Hayer-Hartl, M., Bracher, A. and Hartl, F.U., 2016. The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences*, 41(1), pp. 62–76.

Horwich, A.L., 2013. Chaperonin-mediated protein folding. *Journal of Biological Chemistry*, 288(33), pp. 23622–23632.

Houry, W.A. et al., 1999. Identification of *in vivo* substrates of the chaperonin GroEL. *Nature*, 402(6758), pp. 147–154.

Ishikawa, C. et al., 2011. Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic rice. *Plant Physiology*, 156(3), pp. 1603–1611.

Joshi, J. et al., 2015. Role of small subunit in mediating assembly of red-type Form I Rubisco. *Journal of Biological Chemistry*, 290(2), pp. 1066–1074.

Knight, S., Andersson, I. and Brändén, C.I., 1990. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2·4 Å resolution. Subunit interactions and active site. *Journal of Molecular Biology*, 215(1), pp. 113–160.

Koay, T.W., Wong, H.L. and Lim, B.H., 2016. Engineering of chimeric eukaryotic/bacterial Rubisco large subunits in *Escherichia coli*. *Genes and Genetic Systems*, 91(3), pp. 139–150.

Kobayashi, N. et al., 1999. NMR analysis of the binding of a rhodanese peptide to a minichaperone in solution. *Journal of Molecular Biology*, 292(1), pp. 181–190.

Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M. and Szczepaniak, A., 2014. Rubisco Accumulation Factor 1 from *ThermoSynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and *in vitro*. *FEBS Journal*, 281(17), pp. 3920–3932.

Li, Y., Gao, X. and Chen, L., 2009. GroEL recognizes an amphipathic helix and binds to the hydrophobic side. *Journal of Biological Chemistry*, 284(7), pp. 4324–4331.

Lin, M.T. et al., 2014. A faster Rubisco with potential to increase photosynthesis in crops. *Nature*, 513(7519), pp. 547–550.

Lin, M.T. and Hanson, M.R., 2018. Red algal Rubisco fails to accumulate in transplastomic tobacco expressing *Griffithsia monilis* RbcL and RbcS genes. *Plant Direct*, 2(2), p. e00045.

Lin, M.T., Stone, W.D., Chaudhari, V. and Hanson, M.R., 2019. Enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli* varies depending on the small subunit composition. *BioRxiv*, p. 562223.

Lin, Z. and Rye, H.S., 2004. Expansion and compression of a protein folding intermediate by GroEL. *Molecular Cell*, 16(1), pp. 23–34.

Lin, Z. and Rye, H.S., 2006. GroEL-mediated protein folding: Making the impossible, possible. *Critical Reviews in Biochemistry and Molecular Biology*, 41(4), pp. 211–239.

Liu, C. et al., 2010. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature*, 463(7278), pp. 197–202.

Long, S.P., Marshall-Colon, A. and Zhu, X.G., 2015. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell*, 161(1), pp. 56–66.

Long, S.P., Zhu, X.G., Naidu, S.L. and Ort, D.R., 2006. Can improvement in photosynthesis increase crop yields? *Plant, Cell and Environment*, 29(3), pp. 315–330.

Morell, M.K., Paul, K., Kane, H.J. and John Andrews, T., 1992. Rubisco: Maladapted or misunderstood? *Australian Journal of Botany*, 40(5), pp. 431–441.

Mueller-Cajar, O., Morell, M. and Whitney, S.M., 2007. Directed evolution of Rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochemistry*, 46(49), pp. 14067–14074.

Mueller-Cajar, O. and Whitney, S.M., 2008. Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. *Biochemical Journal*, 414(2), pp. 205–214.

Natesh, R. et al., 2018. A two-domain folding intermediate of RuBisCO in complex with the GroEL chaperonin. *International Journal of Biological Macromolecules*, 118, pp. 671–675.

Onizuka, T. et al., 2004. The rbcX gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant and Cell Physiology*, 45(10), pp. 1390–1395.

Ort, D.R. et al., 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences of the United States of America*, 112(28), pp. 8529–8536.

Paoli, G.C., Vichivanives, P. and Tabita, F.R., 1998. Physiological control and regulation of the *Rhodobacter capsulatus cbb* operons. *Journal of Bacteriology*, 180(16), pp. 4258–4269.

Parikh, M.R., Greene, D.N., Woods, K.K. and Matsumura, I., 2006. Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E.coli*. *Protein Engineering, Design and Selection*, 19(3), pp. 113–119.

Parry, M.A.J. et al., 2003. Manipulation of Rubisco: The amount, activity, function and regulation. *Journal of Experimental Botany*, 54(386), pp. 1321–1333.

Parry, M.A.J. et al., 2011. Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *Journal of Experimental Botany*, 62(2), pp. 453–467.

Parry, M.A.J. et al., 2013. Rubisco activity and regulation as targets for crop improvement. *Journal of Experimental Botany*, 64(3), pp. 717–730.

Parry, M.A.J. et al., 2008. Rubisco regulation: A role for inhibitors. *Journal of Experimental Botany*, 59(7), pp. 1569–1580.

Parry, M.A.J., Madgwick, P.J., Carvalho, J.F.C. and Andralojc, P.J., 2007. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *Journal of Agricultural Science*, 145(1), pp. 31–43.

Peterhansel, C., Niessen, M. and Kebeish, R.M., 2008. Metabolic engineering towards the enhancement of photosynthesis. *Photochemistry and Photobiology*, 84(6), pp. 1317–1323.

Read, B.A. and Tabita, F.R., 1994. High substrate specificity factor ribulose bisphosphate carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant cyanobacterial rubisco containing "algal" residue modifications. *Archives of Biochemistry and Biophysics*, 312(1), pp. 210–218.

Saschenbrecker, S. et al., 2007. Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco. *Cell*, 129(6), pp. 1189–1200.

Sharwood, R.E., 2017. Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phytologist*, 213(2), pp. 494–510.

Sharwood, R.E., Von Caemmerer, S., Maliga, P. and Whitney, S.M., 2008. The catalytic properties of hybrid rubisco comprising tobacco small and sunflower large subunits mirror the kinetically equivalent source rubiscos and can support tobacco growth. *Plant Physiology*, 146(1), pp. 83–96.

Simkin, A.J., López-Calcagno, P.E. and Raines, C.A., 2019. Feeding the world: Improving photosynthetic efficiency for sustainable crop production. *Journal of Experimental Botany*, 70(4), pp. 1119–1140.

Smith, S.A. and 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(3), pp. 557–569.

Smith, S.A. and Tabita, F.R., 2004. Glycine 176 affects catalytic properties and stability of the *Synechococcus* sp. strain PCC6301 ribulose-1,5-bisphosphate carboxylase/oxygenase. *Journal of Biological Chemistry*, 279(24), pp. 25632–25637.

Spreitzer, R.J., 2003. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics*, 414(2), pp. 141–149.

Spreitzer, R.J., Esquivel, M.G., Du, Y.C. and McLaughlin, P.D., 2001. Alanine-scanning mutagenesis of the small-subunit βA - βB loop of chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase: Substitution at Arg-71 affects thermal stability and CO_2/O_2 specificity. *Biochemistry*, 40(19), pp. 5615–5621.

Spreitzer, R.J., Peddi, S.R. and Satagopan, S., 2005. Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proceedings of the National Academy of Sciences of the United States of America*, 102(47), pp. 17225–17230.

Spreitzer, R.J. and Salvucci, M.E., 2002. RUBISCO: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme. *Annual Review of Plant Biology*, 53(1), pp. 449–475.

Stan, G., Brooks, B.R., Lorimer, G.H. and Thirumalai, D., 2005. Identifying natural substrates for chaperonins using a sequence-based approach. *Protein Science*, 14(1), pp. 193–201.

Stan, G., Brooks, B.R., Lorimer, G.H. and 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 Sciences of the United States of America*, 103(12), pp. 4433–4438.

Tabita, F.R. et al., 2008. 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(7), pp. 1515–1524.

Tabita, F.R. and Small, C.L., 1985. Expression and assembly of active cyanobacterial ribulose-1,5-bisphosphate carboxylase/oxygenase in *Escherichia coli* containing stoichiometric amounts of large and small subunits. *Proceedings of the National Academy of Sciences of the United States of America*, 82(18), pp. 6100–6103.

Tcherkez, G.G.B., Farquhar, G.D. and Andrews, T.J., 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences of the United States of America*, 103(19), pp. 7246–7251.

Timm, S., Florian, A., Fernie, A.R. and Bauwe, H., 2016. The regulatory interplay between photorespiration and photosynthesis. *Journal of Experimental Botany*, 67(10), pp. 2923–2929.

van Duijn, E. et al., 2006. Tandem mass spectrometry of intact GroEL - substrate complexes reveals substrate-specific conformational changes in the *trans* ring. *Journal of the American Chemical Society*, 128(14), pp. 4694–4702.

van Duijn, E., Heck, A.J.R. and van der Vies, S.M., 2007. Inter-ring communication allows the GroEL chaperonin complex to distinguish between different substrates. *Protein Science*, 16(5), pp. 956–965.

Viitanen, P. V, Gatenby, A.A. and Lorimer, G.H., 1992. Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Science*, 1(3), pp. 363–369.

Whitney, S.M. and Andrews, T.J., 2001. Plastome-encoded bacterial ribulose-1, 5-bisphosphate carboxylase/oxygenase (RubisCO) supports photosynthesis and growth in tobacco. *Proceedings of the National Academy of Sciences*, 98(25), pp.14738-14743.

Whitney, S.M., Baldet, P., Hudson, G.S. and John Andrews, T., 2001. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant Journal*, 26(5), pp. 535–547.

Whitney, S.M., et al., 2011a. Isoleucine 309 acts as a C_4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in *Flaveria*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(35), pp. 14688–14693.

Whitney, S.M., Houtz, R.L. and Alonso, H., 2011b. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiology*, 155(1), pp. 27–35.

Whitney, S.M. and Sharwood, R.E., 2007. Linked rubisco subunits can assemble into functional oligomers without impeding catalytic performance. *Journal of Biological Chemistry*, 282(6), pp. 3809–3818.

Wilson, R.H., Alonso, H. and Whitney, S.M., 2016. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Scientific Reports*, 6(1), pp. 1–11.

Wilson, R.H. and Hayer-Hartl, M., 2018. Complex Chaperone Dependence of Rubisco Biogenesis. *Biochemistry*, 57(23), pp. 3210–3216.

Wilson, R.H., Martin-Avila, E., Conlan, C. and Whitney, S.M., 2018. An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO_2-fixation kinetics. *Journal of Biological Chemistry*, 293(1), pp. 18–27.

Wingler, A., Lea, P.J., Quick, W.P. and Leegood, R.C., 2000. Photorespiration: Metabolic pathways and their role in stress protection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1402), pp. 1517–1529.

Xu, D., Lin, S.L. and Nussinov, R., 1997. Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *Journal of Molecular Biology*, 265(1), pp. 68–84.

Yagi-Utsumi, M. et al., 2013. NMR characterization of the interaction of GroEL with amyloid β as a model ligand. *FEBS Letters*, 587(11), pp. 1605–1609.

Zhu, X.-G., Long, S.P. and Ort, D.R., 2010. Improving Photosynthetic Efficiency for Greater Yield. *Annual Review of Plant Biology*, 61(1), pp. 235–261.

Zhu, X.G., Long, S.P. and Ort, D.R., 2008. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology*, 19(2), pp. 153–159.

Zhu, X.G., Portis, A.R. and Long, S.P., 2004. Would transformation of C_3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant, Cell and Environment*, 27(2), pp. 155–165.

Appendix A

Differences in primary sequence between *Synechococcus* PCC6301 RbcL and *Chlamydomonas* RbcL

```
                        10        20        30        40        50
                        |         |         |         |         |
Synechococcus RbcL  M~P~~KTQSAAGYKAGVKDYKLTYYTPDYTPKDTDLLAAFRFSPQPGVPA
Chlamydomonas RbcL  MVPQTETKAGAGFKAGVKDYRLTYYTPDYVVRDTDILAAFRMTPQPGVPP

                        60        70        80        90        100
                        |         |         |         |         |
Synechococcus RbcL  DEAGAAIAAESSTGTWTTVWTDLLTDMDRYKGKCYHIEPVQGEENSYFAF
Chlamydomonas RbcL  EECGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYDIEPVPGEDNQYIAY

                        110       120       130       140       150
                        |         |         |         |         |
Synechococcus RbcL  IAYPLDLFEEGSVTNIITSIVGNVFGFKAIRSLRLEDIRFPVALVKTFQG
Chlamydomonas RbcL  VAYPIDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYVKTFVG

                        160       170       180       190       200
                        |         |         |         |         |
Synechococcus RbcL  PPHGIQVERDLLNKYGRPMLGCTIKPKLGLSAKNYGRAVYECLRGGLDFT
Chlamydomonas RbcL  PPHGIQVERDKLNKYGRGILGCTIKPKLGLSAKNYGRAVYECLRGGLDFT

                        210       220       230       240       250
                        |         |         |         |         |
Synechococcus RbcL  KDDENINSQPFQRWRDRFLFVADAIHKSQAETGEIKGHYLNVTAPTCEEM
Chlamydomonas RbcL  KDDENVNSQPFMRWRDRFLFVAEAIYKAQAETGEVKGHYLNATAGTCEEM

                        260       270       280       290       300
                        |         |         |         |         |
Synechococcus RbcL  MKRAEFAKELGMPIIMHDFLTAGFTANTILAKWCRDNGVLLHIHRAMHAV
Chlamydomonas RbcL  MKRAVCAKELGVPIIMHDYLTGGFTANTSLAIYCRDNGLLLHIHRAMHAV

                        310       320       330       340       350
                        |         |         |         |         |
Synechococcus RbcL  IDRQRNHGIHFRVLAKCLRLSGGDHLHSGTVVGKLEGDKASTLGFVDLMR
Chlamydomonas RbcL  IDRQRNHGIHFRVLAKALRVSGGDHLHSGTVVGKLEGEREVTLGFVDLMR

                        360       370       380       390       400
                        |         |         |         |         |
Synechococcus RbcL  EDHIEADRSRGVFFTQDWASMPGVLPVASGGIHVWHMPALVEIFGDDSVL
Chlamydomonas RbcL  DDYVEKDRSRGIYFTQDWCSMPGVMPVASGGIHVWHMPALVEIFGDDACL

                        410       420       430       440       450
                        |         |         |         |         |
Synechococcus RbcL  QFGGGTLGHPWGNAPGATANRVALEACVQARNEGRDLYREGGDILREAGK
Chlamydomonas RbcL  QFGGGTLGHPWGNAPGAAANRVALEACTQARNEGRDLAREGGDVIRSACK

                        460       470
                        |         |
Synechococcus RbcL  WSPELAAALDLWKEIKFEFETMDKL
Chlamydomonas RbcL  WSPELAAACEVWKEIKFEFDTIDKL
```

*Chlamydomonas RbcL has 3 additional residues at N-terminus and 85 different residues from *Synechococcus* PCC6301 RbcL. Numbering based on *Chlamydomonas* RbcL.

Appendix B

Primer designed for construction of chimeric Rubiscos (Swapping of 25 residues)

Primer	Primer sequence
SynN-BsmBI	5'-AATAAGGAGG<u>CGTCTC</u>CCATG*C*CCAAGACGCAATCTG-3'
SynSS-C-PstI	5'-TGGTACCAG<u>CTGCAG</u>ATCTCGACTTAGTA*G*CGGCCGGGACG-3'
Syn275-fwd-BsmBI	5'- ACTTCTTGA<u>CGTCTC</u>GTTTCACAGCCAACACCACCT-3'
Syn275-rev-BsmBI	5'-TTGCCAAGGT<u>CGTCTC</u>GGCGGTGAAACCAG-3'
Syn375-fwd-BsmBI	5'-GGCGTGCT<u>CGTCTC</u>TGCTTCCGGTGGTATC-3'
Syn375-rev-BsmBI	5'-GGATAC<u>CGTCTC</u>AAGCAACTGGCAGCACGC-3'
Syn425-fwd-BsmBI	5'-CGAAC<u>CGTCTC</u>GCTCTTGAAGCTTGCG-3'
Syn425-rev-BsmBI	5'-CGTTCCGAG<u>CGTCTC</u>CACAAGCTTCCAAGG-3'
Chl275-fwd-BsmBI	5'-CTTAACAGGT<u>CGTCTC</u>ACCGCTAACACTTCATTAGC-3'
Chl275-rev-BsmBI	5'-GAAGTGT<u>CGTCTC</u>TGAAACCACCTGTTAAGTAGTCGTGC-3
Chl375-fwd-BsmBI	5'-CAGGTGTTA<u>CGTCTC</u>TTGCTTCAGGTGGTATTCACG-3'
Chl375-rev-BsmBI	5'-CGTGAATA<u>CGTCTC</u>GAAGCAACCGGCATAACACCTG-3'
Chl425-fwd-BsmBI	5'-CCGTGTAGCTC<u>CGTCTC</u>CTTGTACTCAAGC-3'
Chl425-rev-BsmBI	5'-GAGTACA<u>CGTCTC</u>AAGAGCTACACGGTTAG-3'
Trc5	5'-GAGGTATATATTAATGTATCG-3'
Trc3	5'-ATCTTCTCTCATCCGCCA-3'

*Restriction sites are underlined.

Appendix C

Primer pairs designed for construction of ten site-directed Rubiscos mutants

Mutants	Primer pairs
E348D GAA>GAT	Fwd-E348D 5'-GACTTGATGCGCGATGACCACATCGAAGC-3' Rev-E348D 5'-GCTTCGATGTGGTCATCGCGCATCAAGTC-3'
H350Y CAC>TAT	Fwd-H350Y 5'-GATGCGCGAAGACTATATCGAAGCTGACC-3' Rev-H350Y 5'-GGTCAGCTTCGATATAGTCTTCGCGCATC-3'
I351V ATC>GTT	Fwd-I351V 5'-GATGCGCGAAGACCACGTTGAAGCTGACCGCAGC-3' Rev-I351V 5'-GCTGCGGTCAGCTTCAACGTGGTCTTCGCGCATC-3'
A353K GCT>AAA	Fwd-A353K 5'-CGAAGACCACATCGAAAAAGACCGCAGCCGTGGG-3' Rev-A353K 5'-CCCACGGCTGCGGTCTTTTTCGATGTGGTCTTCG-3'
V359I GTC>ATT	Fwd-V359I 5'-GACCGCAGCCGTGGGATTTTCTTCACCCAAGATTGG-3' Rev-V359I 5'-CCAATCTTGGGTGAAGAAAATCCCACGGCTGCGGTC-3'
F360Y TTC>TAT	Fwd-F360Y 5'-GCAGCCGTGGGGGTCTATTTCACCCAAGATTGG-3' Rev-F360Y 5'-CCAATCTTGGGTGAAATAGACCCCACGGCTGC-3'
A366C GCG>TGT	Fwd-A366C 5'-CACCCAAGATTGGTGTTCGATGCCGGGCG-3' Rev-A366C 5'-CGCCCGGCATCGAACACCAATCTTGGGTG-3'
L372M CTG>ATG	Fwd-L372M 5'-GTCGATGCCGGGCGTGATGCCGGTTGCTTCCG-3' Rev-L372M 5'-CGGAAGCAACCGGCATCACGCCCGGCATCGAC-3'

S395A TCC>GCA	Fwd-S395A 5'-GGAAATCTTCGGTGATGACGCAGTTCTCCAGTTCGGTG-3' Rev-S395A 5'-CACCGAACTGGAGAACTGCGTCATCACCGAAGATTTCC-3'
V396C GTT>TGT	Fwd-V396C 5'-CTTCGGTGATGACTCCTGTCTCCAGTTCGGTGG-3' Rev-V396C 5'-CCACCGAACTGGAGACAGGAGTCATCACCGAAG-3'

*Nucleotides highlighted in red introduce mutations