An exploration of the folding and assembly requirements of diverse bacterial rubiscos

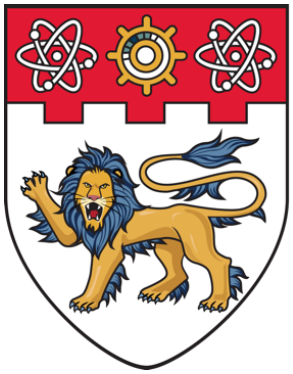
Ramya, Ramaswamy Chettiyan Seetharaman

2019

Ramya, R. C. S. (2019). An exploration of the folding and assembly requirements of diverse bacterial rubiscos. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/93623

https://doi.org/10.32657/10356/93623



NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

AN EXPLORATION OF THE FOLDING AND ASSEMBLY REQUIREMENTS OF DIVERSE BACTERIAL RUBISCOS

RAMASWAMY CHETTIYAN SEETHARAMAN RAMYA

SCHOOL OF BIOLOGICL SCIENCES

2019

AN EXPLORATION OF THE FOLDING AND ASSEMBLY REQUIREMENTS OF DIVERSE BACTERIAL RUBISCOS

RAMASWAMY CHETTIYAN SEETHARAMAN RAMYA

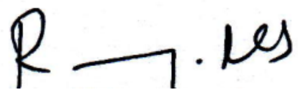
SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.



23/Jan/2019 Ramaswamy Chettiyan Seetharaman Ramya

. . . .

Date

SUPERVISOR DECLARATION STATEMENT

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

23/1/2019
Date

Associate Prof. Oliver Mueller-Cajar

Thesis Title:
An exploration of the folding and assembly requirements of diverse bacterial rubiscos

I have excluded the Authorship Attribution Statement declaration because the thesis does not contain any published material.

Ramaswamy Chettiyan Seetharaman Ramya

Date : 23/Jan/2019

Associate Prof. Oliver Mueller-Cajar

Date: 23/1/2019

ACKNOWLEDGEMENT

I express my true-hearted gratitude and respect to my supervisor, Assoc. Prof. Oliver Martin Mueller-Cajar. His genuine passion for science, inculcated a sense of responsibility and motivation in me to contribute my best my research. This thesis has been possible only because of his help and support.

I render deep gratitude to my Thesis Advisory Committee Members Assoc. Prof. Susana Geifman Shochat, Assoc. Prof. Peter Droge and Asst. Prof. Shashi Bhusan.

My sincere thanks to Dr. Tsai-Yi-Chin for helping me throughout my work, providing me all the required plasmids and giving moral support during hardships.

I give my special thanks to Dr. Tobias Wunder for mentoring me and providing insightful suggestions and being there always for scientific discussions. My deep felt thanks to Dr. Tiago Selao for enlightening me about the various aspects of Cyanobacteria and be readily available to address my clarifications and troubleshooting.

I thank all past and present OMC-lab members, Dev, Zhijun, Maybelle, Lynette, Zen, Jianann, Jed for sharing their knowledge and making lab environment a lively one.

My heartfelt thanks to Ashaa, Poorni, Ankita and Arvind for always being the wonderful persons of my life and supporting me through all times. I am ever indebted to the immense care and support provided by Nitin, LiuDi and Malini during my hard days and having them around in lab motivated to strive better. I would like to thank Yew kwang, Lavi, Apoorva, Gaurav, Mufeedha, my housemates and all my friends for their love and undying support. Special thanks to Bhargy for being an immense support throughout my PhD especially during thesis-writing times. This is the family I earned, and I Thank you for all the happy times together.

A big thanks to my Anna Raghulan, Anni Priya and pattukutti Prahan for the trust and belief in me. Without the support of whom this whole PhD journey wouldn't have been possible.

Appa and Amma, you guys are always my role model and inspiration to live life the best way. You might not be around, but you both have imparted enough wisdom to last for another lifetime. Miss you both.

Acknowledgements can’t do justice to this person in my life who has now taken the role of my husband, previously a friend, mentor, guide, partner-in-crime, Appa, Amma and switches these hats every day. Heartfelt thanks to Parjanya for his love, care, encouragements and always being around throughout this roller-coaster life.

I thank the financial support rendered by the Nanyang Technological University of Singapore in the form of Research Scholarship.

I express my sincere gratitude to all my teachers for their mentoring and helping me shape my career. Lastly, I thank all my friends and family for encouraging and supporting me always.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

3-PGA	3-phosphoglycerate
AAA+	ATPase associated with various cellular activities
ADP	Adenosine diphosphate
AfM	*Acidithiobacillus ferrooxidans*
ATP	Adenosine Triphosphate
BG 11	BlueGreen Medium
BSA	Bovine serum albumin
Bsd2	Bundle Sheath Defective 2
CA1P	2-Carboxy-arabinitol 1-phosphate
CBB	Calvin-Benson-Bassham
CCM	Carbon-concentrating mechanism
CDTA	cyclohexanediaminetetraacetic acid, low sodium
CmR	Chloramphenicol resistance
CO_2	Carbon dioxide
dNTP	Deoxynucleotide mix
DTT	Dithiothreitol
E. coli	Escherichia coli
ECL	Enhanced chemiluminescence
ECM	Activated Rubisco complex
EDTA	Ethylenediaminetetraacetic acid
G3P	Glyceraldehyde-3-phosphate
GT	Glucose tolerant
HRP	Horseradish peroxidase
L_8S_8	Form I Rubisco stoichiometry, 8 large and 8 small subunits
LB	Lysogeny broth
Mg^{2+}	magnesium ion
MOPS	3-morpholinopropane-1-sulfonic acid
NADH	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
OD_{730}	optical density at 730 nm
PCR	Polymerase chain reaction
PDBP	D-glycero-2, 3-pentodiulose-1, 5-bisphosphate
PS I	Photosystem I
PS II	Photosystem II
R.sphaeroides	*Rhodobacter sphaeroides*
Raf1	Rubisco accumulation factors-1
Raf2	Rubisco accumulation factors-2
rbcL	Rubisco large subunit gene
RbcL	Rubisco large subunit
rbcS	Rubisco small subunit gene
RbcS	Rubisco small subunit
rfc	Restriction free cloning

RLPs	Rubisco-like proteins
rpm	Revolutions per minute
RrM	*Rhodospirillum rubrum*
AfM	*Acidithiobacillus ferrooxidans*
RsCbbLS	*Rhodobacter sphaeroides* 2.4.1 Rubisco
RsCbbLSX	*Rhodobacter sphaeroides* 2.4.1 Rubisco system
RsCbbX	*Rhodobacter sphaeroides* 2.4.1 Rubisco activase
RT	Room temperature
Ru5P	Ribulose 5-phosphate
Rubisco	Ribulose 1, 5- bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1, 5-bisphosphate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC	Super Optimal Broth
***Syn*6803**	*Synechocystis* sp. PCC 6803
TAE	Tris base, acetic acid and EDTA
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline, 0.1% Tween 20
TIM-barrel	triose isomerase-like α-β barrel protein fold
TV	pGemT Vector
WT	Wild-type
XuBP	xyulose-1, 5-bisphosphate
Δ	Deletion
Δrbc	Rubisco-deletion

ABSTRACT

Ribulose 1, 5-bisphosphate carboxylase/oxygenase (Rubisco) is the most abundant enzyme representing the main gateway of inorganic carbon into the biosphere by catalysing CO_2-fixation during photosynthesis. In spite of its pivotal role, the enzyme has poor kinetics and substrate specificity. Structurally, Rubisco exists in different forms ranging from dimers of large subunits to hexa-decamers of large and small subunits. These structurally distinct Rubiscos utilize complex machinery for biogenesis including a multiverse of chaperones. It is our aim to devise a simple *in vitro* method to investigate the folding and assembly requirements for different forms of Rubisco. An understanding of *in vitro* Rubisco reconstitution will provide critical information on additional components required for successful Rubisco re-engineering in *in vivo* systems, knowledge of which will help in producing platforms for recombinant expression of heterologous Rubisco systems. To this end we have adopted the continuous spectrophotometric Rubisco activity assay to quantitatively evaluate folding and assembly kinetics in addition to specific chaperone requirements of different Rubisco systems. Chaperonin GroEL/GroES and Mg-ATP are necessary and sufficient for both dimeric and hexameric form II Rubisco, consisting only of large subunits, to fold and assemble from denatured subunits. It was observed that the dimeric *Rhodospirillum rubrum* form II Rubisco had a more rapid maturation rate compared to the hexameric *Acidithiobacillus ferrooxidans* form II Rubisco likely due to the requirement for higher order oligomerization of the hexamer. Interestingly, in contrast to the well-studied chaperone dependent form I cyanobacterial Rubisco, the assembly of the hexadecameric proteobacterial form I *Acidithiobacillus ferrooxidans* Rubisco occurs independently of ancillary proteins. Instead, it requires only the small subunits to assemble into the functional hexa-decamers following successful folding by GroEL/GroES. In conclusion, our method will permit systematic *in vitro* screening and determination of conditions that allow comparative assessment of folding and assembly of all forms of Rubiscos. Simultaneously, we are aiming at constructing Rubisco engineering platforms using a cyanobacterial system, a well-established model organism for photosynthetic research. We have generated cyanobacterial strains (termed cyano-red) using *Synechocystis* sp. PCC 6803 in which the form IC red-type Rubisco from *Rhodobacter sphaeroides* along with its activase is expressed, folded and assembled properly. Further biochemical and physiological characterization of this cyano-red strain will provide us

with novel insights on the biogenesis of different Rubiscos using this cyanobacterial platform.

1 INTRODUCTION

Today's world faces several challenges including rapid human population explosion, an exponential decrease in arable land and global climate crisis. Global efforts are directed towards tackling these problems. It is mandatory to ensure adequate food for this increasing population. Improving agricultural productivity is one of the important solution to provide food supply unintermittedly (Ort *et al.*, 2015). To achieve these methods such as traditional breeding, genetic engineering is being utilised. One important approach is to increase the photosynthetic efficiency in plants (Gornall *et al.*, 2010, Zhu *et al.*, 2010, Ziska Lewis *et al.*, 2012).

Photosynthesis is the basic process that drives life on earth. It provides food and fuel required for human sustenance. Photosynthesis is a complicated, but well-studied process involving several proteins. Proteins and protein engineering serve as an excellent approach for improving the efficiency of photosynthesis. Rubisco is a key protein (enzyme) involved in photosynthesis and is often the rate-limiting step of the process. Years of research have been invested to enhance this enzyme to reach the ulterior goal of improving photosynthesis. However, Rubisco is a complex holoenzyme requiring chaperones to achieve its active form, therefore involving protein folding. Protein folding helps proteins to fold to their correct three-dimensional structure to become biologically functional. Protein folding is sometimes followed by an assembly process to attain ultimate active form (Kim *et al.*, 2013). The key protein Rubisco needs folding chaperones and at times assembly chaperones as well (Bracher *et al.*, 2017). Despite the critical role of Rubisco, its folding and assembly requirements are still not understood completely on a broad spectrum. Hence in this thesis, we attempt to contribute towards enriching the biogenesis knowledge of Rubisco using *in vitro* methods. We are also reporting efforts towards the development of a cyanobacterial Rubisco chassis strain that we aim to use to evaluate different enzymes of interest.

1.1 Protein Folding

Proteins are organic molecules playing indispensable roles in all cellular functions, such as aiding in metabolic reactions and structural properties. Protein exists as catalysts, adaptors, transport molecules, or regulators to perform vital functions. For a protein to be functional, it needs to fold into a unique three-dimensional conformation. Primary structure of protein includes a linear sequence of amino acid residues that locally arrange to attain secondary structures in the form of α-helices and β-sheets. These secondary elements further get spatially arranged and stabilized by hydrogen, ionic, or Vanderwaal's interactions to form tertiary structures. By elaborate arrangements, folding and assembly, proteins achieve their biologically functional competent structure (Berg *et al.*, 2006, Nelson & Cox, 2008).

Christian Anfinsen conducted a series of experiments using ribonuclease to show that three-dimensional structure-related information is contained in the amino acid sequence of a polypeptide Native structure and folding of a protein are also contained in its primary amino acid residues. He revealed denatured proteins could refold spontaneously *in-vitro* (Anfinsen, 1973). However, Levinthal presented arguments stating the possibility of proteins undergoing multiple random searches to find their native conformation from a significantly large number of structures take an exceedingly long time (Levinthal, 1968).

From various studies and proposed models, it is showed that protein should reach a stable energy minimum structure at a realistic folding time. Hence, they follow a rugged funnel-like pathway advancing downhill towards an energy minimum native structure. This pathway may contain transient folding intermediates and misfolded states. Thus, protein folding does not necessarily be single defined path and proceeds through different mechanisms (Hartl *et al.*, 2011).

Protein folding is an important, ubiquitous, and complex bioprocess that ensures linear polypeptides transforms into functional conformations. It often happens spontaneously in a co-translational manner where one domain begins to fold while the other domain is still translating. Hydrophobic amino acids primarily drive the folding process based on non-covalent interactions. Many methods are used to study the folding of protein *in-vitro*. In a test-tube refolding condition, dilute unfolded proteins are driven to fold using many factors. Unlike these, the *in-vivo* environment is highly concentrated, causing

crowding and affects the protein folding and stability. In these cases, non-specific interactions occur and might favour aggregation. If folding is slower than aggregation, it can lead to adverse consequences in the cell. Protein misfolding to non-native structures leads to protein aggregation or formation of insoluble inclusion bodies causing several neurodegenerative diseases in human beings such as Alzheimer's, Parkinson's, Huntington's, Creutzfeld-Jakob or Gaucher's and lead to the collapse of cellular and molecular mechanisms (Hartl & Hayer-Hartl, 2002, Dobson, 2003, Hartl, 2017).

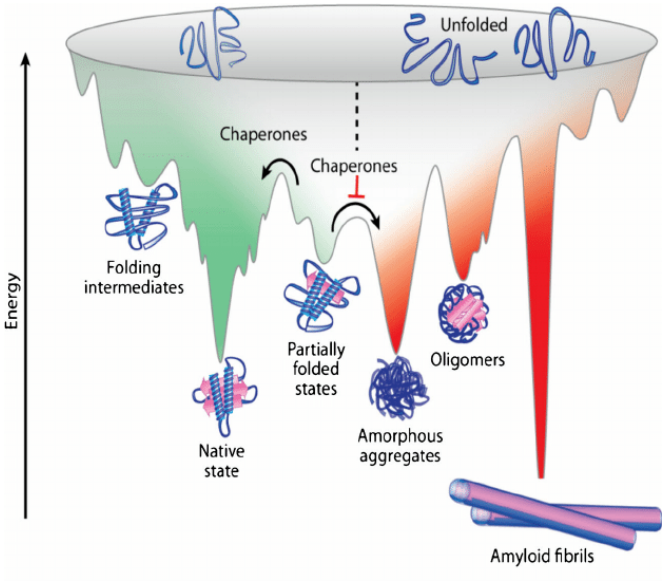


Figure 1.1 **Protein folding energy landscape:** Protein folding advances towards energy minimum native structure through an inverted funnel-shaped energy landscape. This path proceeds via kinetically trapped intermediates including partially and misfolded states. *Figure adapted from (Kim et al., 2013).*

1.1.1 Molecular Chaperones

Cells developed extensive machinery consisting of proteins termed molecular chaperones to aid in proper protein folding. Molecular chaperones help in the folding, assembly of cellular proteins and prevent its misfolding/aggregation, without being a part of the final protein structure. Molecular chaperones are found in all domains of life including Bacteria, Archaea, and Eukarya. Molecular chaperones act as catalysts for physiological protein folding using transient non-covalent interactions and regulate translocation along with protein complex formation. Chaperones take part right from the biosynthesis to proteostatic maintenance, folding, assembly, degradation of aggregates/misfolded proteins. They generally recognize the aggregation-prone hydrophobic regions of peptides/polypeptides and reversibly bind to these residues to prevent them from misfolding. They stabilize, regulate, and direct these non-native

polypeptides towards constructive folding/assembly pathways. In addition to this, chaperones can also rescue aggregated proteins or direct them towards degradation (Hartl, 1996, Bukau *et al.*, 2006). Complex networks of chaperone systems have been extensively studied for structural and mechanistic properties (Ben-Zvi & Goloubinoff, 2001).

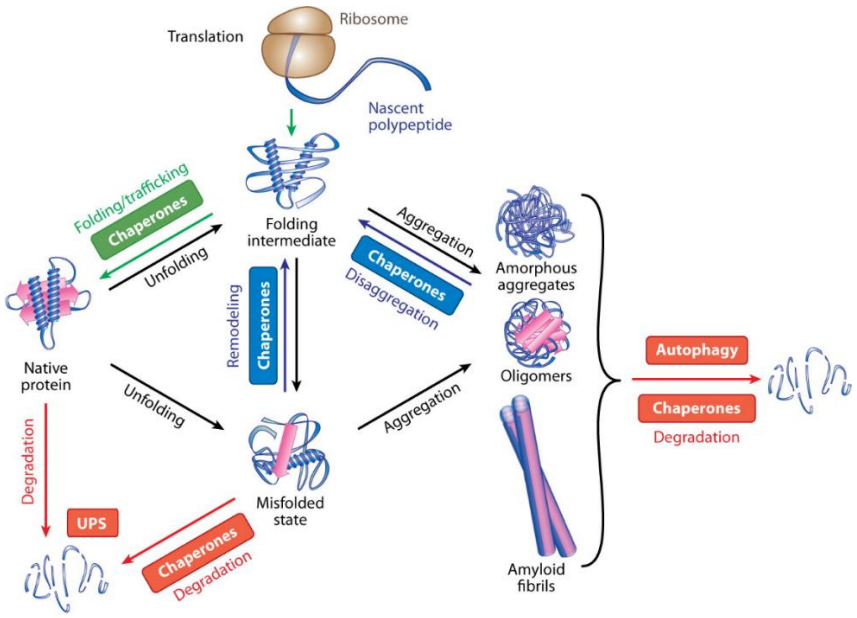


Figure 1.2 **Functions of Chaperones:** Molecular chaperones play a vital role in maintaining proteostasis balance by folding of translated polypeptides, remodelling of misfolded chains and degradation of aggregates. Figure adapted from (Kim *et al.*, 2013).

According to their function, molecular mass and sequence homology, chaperones are classified into different categories. Several chaperone families have been discovered and characterised according to their biochemistry, regulations and mechanisms of action. Some chaperones act in combination with co-chaperones. Most chaperones work in an ATP-dependant manner where ATP hydrolysis aids the protein folding (Hayer-Hartl *et al.*, 2016). Certain small heat shock proteins like Hsp27 can regulate substrate binding in an ATP-independent manner (Haslbeck *et al.*, 2005).

1.1.2 Heat shock proteins (Hsps)

Cell damage can lead to two kinds of reaction namely cell death that would eliminate the damaged cell to stop the progressing inflammation and a stress response that would help the cell to heal and survive. These two are jointly responsible for the cell survival after the stress caused by physical, chemical and environmental factors. During the

response times, the cells produce a group of chemicals called stress proteins or heat shock proteins (Hsps). These are found in all organisms (Lindquist, 1986). These are a family of chaperone proteins classified according to their size namely Hsp27, Hsp40 (DnaJ), Hsp60 (Cpn60, GroEL), Hsp70 (DnaK, BiP), Hsp90 and Hsp100 (Clp) families (Miernyk, 1999, Walter & Buchner, 2002). The primary function of Hsps is to guard the cells from stresses caused by variation in temperature, availability of nutrients, chemical response and irradiation. Hsps can also be involved in the immune responses of the body. If a cell is infected or tumorous, Hsps produces antigens which in turn activate the body's immune system and help them to clear the spoiled cells. Hsps are involved in the folding of proteins by inhibiting the misfolding and accelerating the refolding (Gething & Sambrook, 1992, Parsell & Lindquist, 1993). This regulating function is often based on how they interact with substrates using ATP domains (Hsp70, Hsp90) and sometimes without the ATP domain (Hsp27). Recent studies show that Hsps are also involved in mechanisms regulating apoptosis by a series of reactions of ATP binding, hydrolysis and nucleotide exchange .Hsps functions as regulators and survival promotors (Beere, 2004).

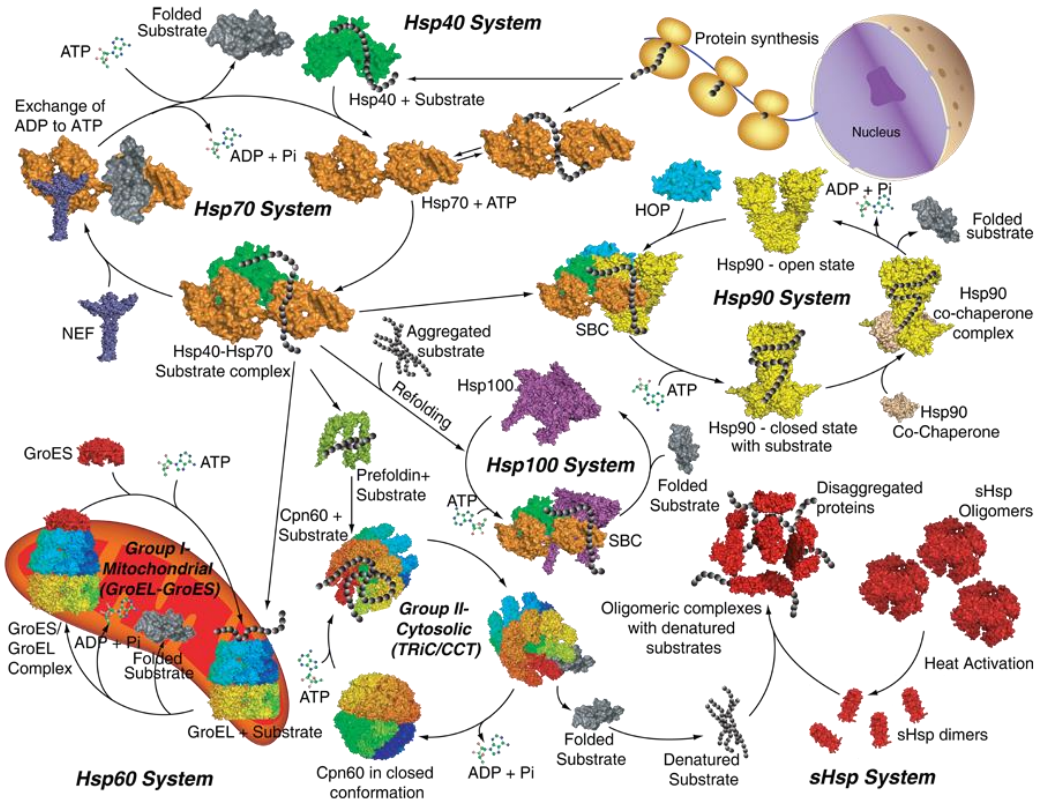


Figure 1.3 **Hsp System:** Hsps form a major network of chaperones involved in cell's function. The figure gives an overview of all the known Hsps with their respective functions and mechanisms. It also includes the downstream chaperonins associated with Hsps. Figure adapted from (R. *et al.*, 2012).

1.1.2.1 Hsp70

Hsp70, an ATP-dependant chaperone, belongs to a ubiquitous class of enzymes. They primarily interact with newly produced polypeptides and are extensively involved in protein transport, folding, refolding, and membrane translocation. Most of our understanding about the structure and function of Hsp70 is through studies conducted on the *E.coli* Hsp70 DnaK. However, with 50% structural homology between mammals and bacteria, Hsp70 is known to have comparable mechanistic functions across species (Sharma & Masison, 2009).

Hsp70 has two structural domains namely a 44 kDa N-terminal nucleotide-binding domain (NBD) and a 27 kDa C-terminal substrate-binding domain (SBD). The NBD, also called the ATP-binding domain, is divided into four sub-domains that form a nucleotide-binding cleft. One Mg^{2+} and two K^{+} ions connect the four sub-domains (Flaherty *et al.*, 1990). The SBD is divided into two sub-domains: the 18 kDa main sub-domain is made up of β-sheets that contain the substrate binding site and the second sub-domain comprises of an α-helical lid region (~10kDa) (Cegielska & Georgopoulos, 1989). A highly conserved hydrophobic linker domain that connects the NBD to SBD has shown to mediate interdomain communications.

ATP hydrolysis enhances substrate binding to Hsp70, which is further improved by closure of the binding cavity (Javid *et al.*, 2007, Morán Luengo *et al.*, 2019). Hsp70 can be called housekeeping proteins as they proof-read protein structure, they not only fold and assemble newly synthesized proteins but also refold misfolded proteins (Pratt & Toft, 2003). Hsp70 carries out this primary function of polypeptide folding as substrates bind and release at the expense of ATP hydrolysis. Substrate binding by Hsp70 is facilitated by NBD and SBD in two different conformations. The ADP-bound conformation enables substrate binding with high affinity and low exchange rates. In the ADP bound state, NBD and SBD are clasped together with the helical lid sub-domain of the SBD covering the protein-bound substrate-binding sub-domain. In contrast, the ATP bound conformation of Hsp70 facilitates fast exchange of substrates with low affinity. In this state, the two sub-domains of the SBD may detach and bind to different sites on NBD improving access to the substrate-binding site for swift, low affinity substrate binding (Genest *et al.*, 2019).

The stability of the low-affinity proteins in the ATP-bound state is enhanced and stabilized by co-chaperones such as Hsp40 and tetratricopeptide repeat (TPR). These co-chaperones bind to the C-terminal region of Hsp70 (Lu & Cyr, 1998). Minute errors during protein folding can lead to misfolded proteins, where proteins are not in their native state and incomplete folding of protein can lead to protein aggregation. These can give rise to mutant and malfunctioning proteins, which will eventually cause disease and pathogenic state. Hsp70 works in conjunction with Hsp40 to unfold misfolded protein to either degrade or refold them into right conformation, although much remains to be understood on how Hsp70 unfolds mis-folded proteins. Hsp70 also binds and stabilizes unfolded protein thus preventing aggregation until the protein folds into the right conformation. Furthermore, Hsp70 has shown to inhibit stress-induced apoptosis in cells and prevent neurodegenerative diseases associated with excessive apoptosis (Mosser *et al.*, 2000, Klucken *et al.*, 2004).

1.1.2.2 Hsp40

Hsp40 belong to the family of J-domain proteins (JDP) and play a vital role in assisting Hsp70. The *E.coli* Hsp40 DnaJ and the *S. cerevisiae* Hsp40 Ydj1 are well characterized for their structure and function. The J domain of Hsp40 contains a stretch of 70 residues that is conserved across human, *E.coli* and *S. cerevisiae* (Laufen et al., 1999). Thermodynamically unstable proteins rely on assistance from Hsp70 to stabilize and maintain their native form at the expense of continuous ATP hydrolysis. With low basal level ATPase activity, Hsp70 need constant support in performing their functions (Xu, 2018). Substrate binding can stimulate ATP hydrolysis by 2-10 folds. Nevertheless, interaction between the N-terminal J-domain of Hsp40 and the substrate-binding domain of Hsp70 can stimulate a 1000-fold ATP-hydrolysis, which enables Hsp70 to bind to a wider range of substrates than what it would have on its own. The C-terminal domain of Hsp40 binds to denatured proteins and hydrophobic residues of unfolded proteins and prevents protein aggregation (Szabo *et al.*, 1996).

ATP hydrolysis, protein binding and release cycle by chaperone DnaK is regulated by accessory proteins GrpE and DnaJ (Packschies *et al.*, 1997). GrpE is the nucleotide exchange factor for DnaK. It promotes the dissociation of ADP from nucleotide binding cleft of DnaK. GrpE and DnaJ act in concert to control the unfolded proteins shuttle in and out of the substrate binding domain of DnaK. They together regulate the nucleotide

bound state of DnaK. GrpE promotes the exchange of ADP for ATP and facilitates peptide release form DnaK substrate binding domain in an ATP-independent manner (Harrison, 2003).

1.1.2.3 Hsp90

Hsp90, a second notable member of the Hsp family, is one of the most abundant chaperones and plays a major role in signal transduction in many eukaryotic proteins. Furthermore, Hsp90 in conjunction with Hsp70 plays an essential part in protein folding, transport and maturation. Hsp90 is a dimer and each of its protomers has three structural domains: An N-terminal nucleotide binding domain (NTD), a middle domain (MD) which has been shown to enable interaction with Hsp70 of other species like *E.coli* and yeast (Genest *et al.*, 2015), and a C-terminal domain (CTD) which enables homo-dimerization. The NTD, also called the ATP-binding domain, has a conserved lid region and may bind to Hsp90 inhibitors and small peptides (Prodromou *et al.*, 1997). Like Hsp70, Hsp90 swings between ATP-bound and the ADP-bound state when the lid either closes the nucleotide-binding region or stays open, respectively. Although Hsp70 is involved in a wide range of protein folding and degradation, it must overcome several limitations inflicted by the binding/releasing mechanism. Thus, Hsp70 relays with Hsp90 to accomplish these objectives (Morán Luengo *et al.*, 2018). Hsp90 takes over downstream of Hsp70 in its chaperone activity of folding and maturation. Both Hsp70 and Hsp90 possess a TPR domain in their carboxy terminal. The co-chaperone Hop facilitates interaction by binding to the TPR domain of Hsp70 and Hsp90 thereby forming a bridge that would enable substrate to progress from Hsp70 to Hsp90 and finally towards its native form (Wegele *et al.*, 2006). Unlike Hsp70, Hsp90 can bind to folded proteins in the cytosol to assist their maturation and activation. Hsp90 is also involved in some cellular processes like the cell cycle, promoting cell survival etc. In recent years, the association of Hsp90 with disease progression had led it to be viewed as a potential therapeutic target in cancer and neurodegenerative diseases (Li & Buchner, 2013).

1.1.2.4 Hsp27

Hsp27, a small molecular weight heat shock protein is found predominantly in the cytosol. This is a heat inducible chaperone which lacks an ATP domain and functions

independent of ATP with the regulation done by mitogen-activated protein (MAP)-kinase-dependent phosphorylation and self-oligomerization. High expression of this protein results in a poor prediction of cancer. Phosphorylation and oligomerization of Hsp27 is necessary for the cellular existence. Some recent studies show that Hsp27 has direct implications on apoptosis (GARRIDO *et al.*, 1999, Castelli *et al.*, 2004).

1.1.3 Chaperonins

Several molecular chaperone families have been characterized thus far and of those chaperonins represent a well-characterized and conserved class of chaperones in prokaryotic and eukaryotic organisms. Chaperonins consist of a hollow double-ring complex where folding occurs in the central cavity preventing aggregation. Chaperonins assist protein folding in an ATP-dependent manner and are subdivided into Group I (GroEL, mitochondrial Hsp60) and Group II (TRiC/CCT, archaeal chaperonins). The Group I chaperonins generally work in concert with its co-chaperonin (Viale & Arakaki, 1994).

1.1.3.1 Group I Chaperonins

Group I chaperonins are found in the bacterial cytosol and eukaryotic organelles. These are an essential chaperonin found in bacteria. In *E.coli* it is involved in folding of 10-15% of cytosolic proteins.

1.1.3.1.1 GroEL-GroES chaperonin system

The GroEL monomer is approximately 60 kDa forming a large cylindrical homo-oligomeric complex consisting of two stacked heptameric rings to form an 800 kDa complex. The co-chaperonin GroES which is around 10 kDa forms a single heptameric ring which acts as a lid to the chaperonin GroEL. GroEL-GroES is an ATP-dependent folding system. GroEL provides a central cavity for isolation and binding of substrate to fold correctly. GroEL consists of three structurally different domains, namely apical, hinge-like intermediate, and equatorial domains. Substrate and GroES binding take place in the exposed hydrophobic surfaces of the apical domain which acts as entrance of the central cavity, whereas the hinge-like intermediate domain assists in the exchange of information between the apical and equatorial domains while inducing conformational changes upon ATP binding. The equatorial domain bears the ATP binding site and serves as inter-ring contacts within GroEL oligomers (Braig *et al.*, 1994, Sigler *et al.*, 1998, Hartl & Hayer-Hartl, 2002). The hollow cylindrical space formed by stacked GroEL heptameric subunits provides the central cavity for the folding of client

unfolded proteins by hydrophobic interactions. GroEL and GroES work cooperatively for the binding and release reactions of substrates. ATP and GroES binding to their respective sites are required for the functional folding cycle, where the GroES lid on top of GroEL cylinders makes the folding cavity to expand and facilitate folding. ATP binding and hydrolysis are responsible for transition between open and closed conformations which is receptive to unfolded protein binding and folded protein isolation respectively (Chandrasekhar *et al.*, 1986, Roseman *et al.*, 1996, Horovitz & Willison, 2005) The apical domain where GroES binds, forming the bullet-shaped GroEL-GroES complex, constitutes the cis cavity for non-native protein encapsulation and the opposing ring is termed the trans cavity. Interaction between GroEL and GroES is transient and dynamic, which is regulated by ATPase activity, and the substrate protein folds in ~6 seconds, which is the time required for hydrolysis of 7 ATP molecules at 25 ºC. Negative cooperativity between the two GroEL rings is responsible for the preferential binding of ATP and GroES to the cis rings. When ATP is absent, the apical domain is more hydrophobic, displaying a high affinity to protein substrates and in the presence of ATP, GroES binds to GroEL causing conformational changes to make the apical domain more hydrophilic to remove substrates(Langer *et al.*, 1992, Ranson *et al.*, 1998, Horovitz & Willison, 2005). GroEL-GroES can fold proteins up to ~60 kDa molecular weight. GroEL can repeat this coordinated folding cycles. Larger proteins (exceeding ~60 kDa) undergo multiple cycles of binding and release from GroEL.

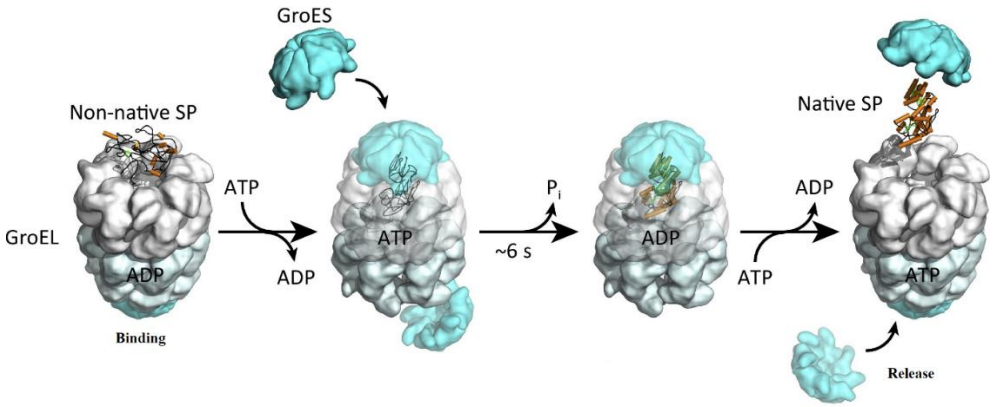


Figure 1.4 **GroEL/GroES protein folding mechanism:** Substrate binds to the apical domain of GroEL. Folding of the substrate takes place in the GroEL hollow cylindrical cage after the conformational changes initiated by ATP and GroES binding. Protein folds to its native structure and finally gets released. Figure adapted from (Hayer-Hartl *et al.*, 2016).

1.1.3.1.2 Hsp60

Hsp60 is also called HSPD1, 60 kDa Heat Shock Protein, Mitochondrial Matrix Protein P1 and Chaperonin 60. This is a nuclear genome encoded protein and functions as a signalling molecule for macrophage activation in the immune system and regulates the folding of the proteins formed and their assembly. Hsp60 plays a part in the body's autoimmunity and diseases like rheumatoid arthritis (Bukau & Horwich, 1998). Hsp60 chaperonin functions along with Hsp10 in maintaining the correct folding of the newly formed proteins under stressed conditions by averting the misfolding and stimulating the refolding and proper assembly of the folded proteins in the mitochondria. The main functional part of the chaperonin is a large heptameric ring that binds to an unfolded protein substrate which in turn binds to ATP and co-chaperonin Hsp10. When the proteins are folded successfully, hydrolysis of ATP happens and results in the dissociation of folded protein substrates and chaperonin rings along with ADP (Ohashi *et al.*, 2000).

1.1.3.2 Group II Chaperonins

Group II chaperonins structurally differ from Group I in exhibiting a built-in lid like helical protrusions in the apical domain that serves the purpose of co-chaperone. These are found in the cytoplasm of archaea and eukaryotes. Complete structural and mechanistic understanding of Group II Chaperonins is not available. Group II chaperonins are hetero-oligomeric complexes. They are made up of 8-9 subunits per ring, stacked directly opposite to one another. Substrate encapsulation is essential for folding by Group II chaperonins (Hartl *et al.*, 2011). The ATP utilization is very different in Group II chaperonins in comparison to Group I chaperonins. The first structure solved for the Group II chaperonins was the apical domain of thermosome α-subunit (Klumpp & Baumeister, 1998). This structure consisted of an apical domain, nucleotide-binding equatorial domain and an intermediate domain consisting of the catalytic aspartate for nucleotide hydrolysis.

1.1.3.2.1 TriC/CCT

TriC/CCT is a eukaryotic chaperonin consisting of eight subunits per ring of two identical rings ranging from 52-65 kDa. 5-10% of newly synthesized cytosolic proteins are folded by the TriC/CCT chaperonins. Different subunits of TriC/CCT are partitioned

according to their ATP binding affinities and net charge. This variation in functional and chemical properties of TriC/CCT helps in the folding of many substrate proteins. Different subunits have different binding specificities enabling the binding and folding of diverse proteins. These are multi-subunit, highly conserved chaperonin consisting of three distinct domains, namely apical, intermediate, and equatorial. Mutational studies reveal that generally four subunits out of eight shows higher affinity to ATP binding and are extremely important for folding (Gutsche *et al.*, 1999). The only exception is Saccharomyces cerevisiae, which requires all the eight subunits. Common substrates of TriC/CCT include actin and tubulin. Substrate recognition by TriC/CCT is by hydrophobic and electrostatic interactions (Zhang *et al.*, 2010). They are slower than GroEL in protein encapsulation and folding. TriC/CCT modulates protein aggregation and displays limited substrate specificity. It works in a concerted manner with upstream chaperones Hsp70/90 for substrate transfer and proteins prefoldin, Bag3 helps in the fine-tuning of its protein folding. TriC/CCT malfunction can cause several diseases, including sensory neuropathy. Cancer protein p53 is also a TriC/CCT substrate. TriC/CCT is involved in decreasing aggregation and cell cytotoxicity (Lopez *et al.*, 2015).

Overall, Group II chaperonins are an absolute necessity for maintaining proper cell physiology.

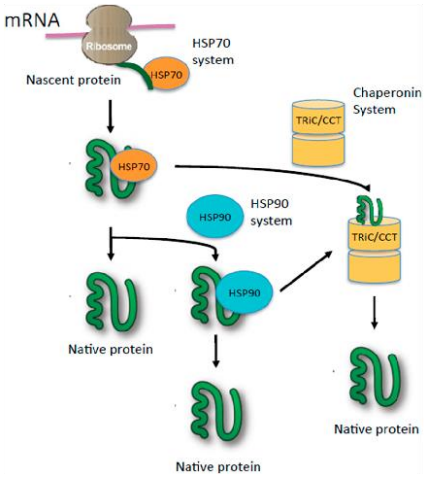


Figure 1.5 **Hsp70 and TRiC/CCT concerted model:** TriC/CCT can also work in concert with other chaperones. When a nascent protein interacts with Hsp70, it will either proceed folding by itself or sometimes delivers to downstream chaperones. Folding of these intermediates are then completed by TriC/CCT. Figure adapted from (Roh *et al.*, 2015).

1.1.4 Chaperone –mediated protein degradation

Cell proteostasis is maintained through a balance between protein translation, their function and proteolytic degradation, mediated by chaperones and proteolytic systems. Chaperone mediated autophagy plays a vital role in protein quality control, metabolic and transcriptional regulation, immune response, and can be triggered in response to cellular stress such as starvation (Kaushik & Cuervo, 2018). Selective endosomal micro-autophagy of mammalian proteins containing specific motifs, such as KFERQ-like peptide sequences, is mediated by the cytosolic chaperone Heat shock cognate (HSC 70) with its co-chaperone carboxyl terminus of HSC-70 interacting protein (CHIP), Hsp40 (also DNABJ1), and Hsp70-Hsp90 organizing protein (HOP), which transport them to the protein degradation machinery. HSC70 also contributes to macroautophagy in association with the BAG family of co-chaperones to selectively degrade Ubiquitin-positive aggregates (Kaushik & Cuervo, 2018).

The cellular protein quality control network comprises molecular chaperones which perform conformational repairs and ensure proper protein folding, together with a ubiquitin-proteasome system (UPS) and autophagy which get rid of the degraded and misfolded protein aggregates. A breakdown in this proteostasis network plays a crucial role in aging and in the progression of diseases associated with protein aggregation and misfolding. Upon stress stimulus, the chaperones are upregulated in order to transport the protein substrates for proteasomal degradation to the nucleus. A similar upregulation is induced by the cytosolic stress response, unfolded protein response in the endoplasmic reticulum, and mitochondria for lysosomal clearance. Some chaperones complexes, such as the Hsp70-Hsp110-Hsp40 ternary complex in mammalian cells, and Hsp100-Hsp70 complex in bacteria and fungi, act as disaggregases to break down the protein aggregates as a preparatory step for their proteolytic degradation by UPS (Balchin *et al.*, 2016).

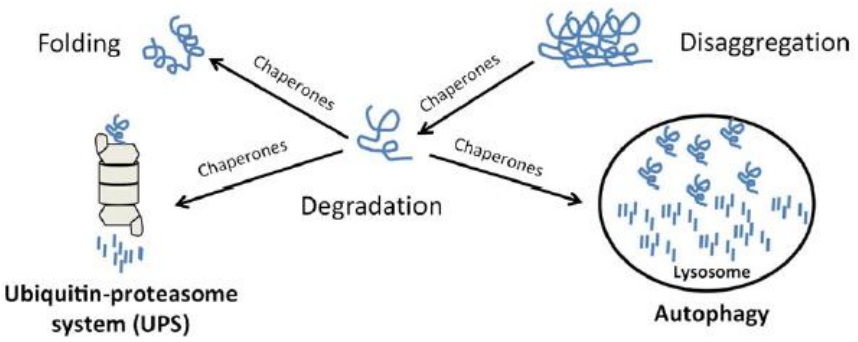


Figure 1.6 **Molecular chaperones mediate protein degradation:** Chaperones help in the degradation of misfolded polypeptides. They resort to Ubiquitin-proteasome system (UPS) or Autophagy. Figure adapted from(Fernández-Fernández *et al.*, 2017).

Chaperones act as either folding machines or degradation factors depending on the associated co-factors; binding with co-factors such as CHIP and BAG-1 would induce substrate degradation, whereas binding of Hop and Hip to the same docking sites would promote protein folding. CHIP can bind both Hsp70 and Hsp90 and modulate their chaperone activity. Once protein substrate is tagged for ubiquitination, CHIP binds to the E2 ubiquitin-conjugating enzymes, and the chaperone/CHIP complex acts as an E3 ubiquitin ligase to induce the degradation of the substrate (Esser *et al.*, 2004). From the prokaryotic Hsp70, i.e., archetypical DnaK, to the 13 human isoforms, Hsp70 chaperones are highly conserved and play a key role in protein degradation by the proteasome as well as autophagy. In mammalian cells, Hsp70 crucially contributes to protein folding, disaggregation, and degradation based on ATP-dependent conformational changes (Fernández-Fernández *et al.*, 2017).

1.2 Photosynthesis

The welfare of humankind revolves around a basic process, photosynthesis. It accounts for global food and fuel supply. This complex process is explained in a simple note as a fundamental biochemical process that utilizes light energy towards the synthesis of the organic compounds that constitute the planet's biomass. Photosynthetic organisms like plants, algae, some bacteria convert sunlight into carbohydrates using CO_2 and H_2O. This mechanism follows the stoichiometry given below:

$$6CO_2 + 6H_2O \rightarrow O_2 + C_6H_{12}O_6$$

The basic equation might seem deceptively simple, but its overall mechanism is complex involving numerous proteins and small molecules. The outcome of this process and the amount of energy stored is enormous. Free energy is stored annually by photosynthesis on Earth, which corresponds to the assimilation of more than 10^{10} tons of carbon into carbohydrate and other forms of organic matter (Field *et al.*, 1998). In a eukaryotic photosynthetic cell, it takes place in chloroplast whereas, in photosynthetic bacteria which lack chloroplasts, photosynthesis takes place in the cytoplasm. In cyanobacteria, photosynthetic electron transport machinery (with both photosystems I and II) is embedded in an internal membrane system known as the thylakoids. General photosynthetic bacteria have a variety of pigments called bacterio-chlorophylls and are often equipped with only one type of specialized photosystem. Under anaerobic conditions, many of these bacteria are able to use electron donors like H_2S, H_2SO_4, H_2SO_3 or H_2 instead of water (Berg *et al.*, 2012).

Photosynthesis takes place in two sequential stages of reactions called light-dependent reactions and light- independent reactions. The light-independent reactions are also known as dark reactions or Calvin-Benson-Bassham cycle (Raines, 2003, Berg *et al.*, 2012). The light-dependent reactions convert energy from sunlight into stored chemical energy. The chemical energy is stored in form of electron carrier molecule NADPH and energy currency molecule ATP. In light-independent reaction, this NADPH and ATP are used in the carbon-dioxide reduction to yield useful organic compounds (Berg *et al.*, 2012).

1.2.1 Light dependent reactions

This takes place in thylakoid membranes of the chloroplast (or the cytoplasm in cyanobacteria). There are four major protein complexes in this thylakoid membrane namely Photosystem II (PSII), the Cytochrome b6f complex, Photosystem I (PSI), and ATP synthase. These four complexes work in series towards ultimately forming the products ATP and NADPH. The process begins in PS II, where a chlorophyll molecule from the reaction centre absorbs a photon, an electron in this molecule gets excited and attains higher energy state. This higher energy state is very unstable, and this electron is transferred from one molecule to another. This leads to a series of redox reactions. The flow of electron is from the PS II to PS I via cytochrome b_6. The final acceptor of this

electron is $NADP^+$. The electron flow also drives acidification of the lumen which is utilized for photophosphorylation of ADP to ATP through the ATPase complex (Figure 1.1)(Nelson, 2004).

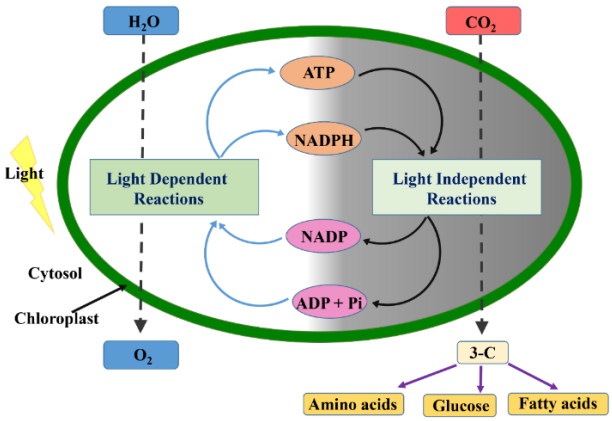


Figure 1.7 **Sequential steps of Photosynthesis:** Depicting sequential steps of photosynthesis. It shows that light-dependent reaction generates ATP and NADPH by utilising energy from sunlight. These molecules are used up by light-independent reactions to incorporate CO2 into RuBP to produce useful organic compounds.

1.2.2 Light independent reactions or The Calvin-Benson-Bassham Cycle

In this process energy from the light-dependent reaction is used to assemble carbohydrates from carbon-dioxide molecules. In chloroplasts the CBB cycle takes place in the stroma, a fluid-filled space outside the thylakoid membrane(Bassham, 2003). The process occurs in three major phases. It involves fixation of carbon, reduction reactions and regeneration of substrates. In the carbon fixation step, three molecules of atmospheric carbon-dioxide are incorporated into three molecules of substrate ribulose-1, 5- bisphosphate (RuBP) by the enzyme called Rubisco. This results in the formation of six molecules of the 3- carbon intermediate 3-Phosphoglycerate (3-PGA). Here starts the second phase of the cycle where 3-PGA is reduced to 6 molecules of Glyceraldehyde-3-phosphate (G-3-P) via the transition molecule 1, 3-bisphosphoglycerate. Five molecules of G-3-P in turn, complete the cycle by regenerating RuBP via 3 molecules of ribulose-5-phosphate (Ru5P). One molecule of G-3-P is diverted from the cycle towards organic compounds such as glucose, fructose, sucrose and many others. The whole process of the CBB cycle consumes six molecules of ATP and NADPH each. The key player in this cycle is enzyme Rubisco which accounts for CO_2 fixation and supports survival on earth (Figure 1.2) (Nelson, 2004).

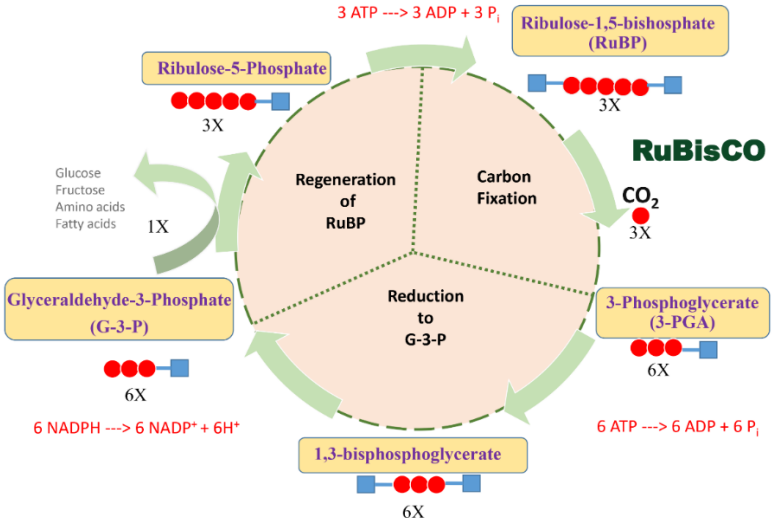


Figure 1.8 **Sequential steps of CBB cycle:** Depicting sequential steps of Calvin- Benson-Bassham cycle. This cycle is divided into three phases- carbon fixation, reduction of 3-PGA and regeneration of RuBP. Here red coloured balls depict the carbon atoms and blue squares the phosphate group.

1.3 Rubisco - The most abundant protein on earth

The most abundant enzyme in nature (Ellis, 1979) that catalyses the first step of carbon dioxide fixation in most photosynthetic and chemoautotrophic organisms. It acts as carboxylase of RuBP and produces useful products. This enzyme is inefficient and slow which is manifested in its low turnover number (Schneider *et al.*, 1992). Kinetic properties of Rubisco include a relatively low turnover rate (1-10 s^{-1}), poor affinity for CO_2 and competitive inhibition (and catalysis of a side-reaction) with O_2 (Tcherkez, 2006). These areas can be targeted to boost the competence of Rubisco. In general, from available literature it is accepted that rubisco is both sluggish and displays poor substrate specificity. Aspects that are still unclear regarding rubisco's behaviour are – (i) Origin of oxygen being a substrate. (ii) Lack of knowledge regarding proton transfers and exchanges. (iii) Why this low turnover and less specificity. (iv) Knowledge constrains regarding rubisco kinetics. Understanding all these aspects would help us to explain the enzyme's behaviour in a more balanced way. In a recent review, rubisco being inefficient has been elaborately discussed with newly available bioinformatic tools. According to this review, rubisco though being slow and confused, it is not very inefficient or promiscuous in comparison to other enzyme groups available. Practically enzymes can be as fast as 106 -107 s-1 (k_{cat}). Examples include carbonic anhydrase. In comparison to such exceeding values, rubisco only exhibits 1- 10 s-1 (k_{cat}) and hence

can be comparatively termed as slow enzyme. But from a recent analysis (BRENDA database), rubisco was found to be not particularly low and behaved good enough in comparison to 2000 other enzymes. Hence it can now be accepted that rubisco has moderate kinetics. It has evolved to respond to changing environment and newer tools available now has made these discoveries possible. But since there is still a lot of questions unanswered regarding rubisco, it has huge potential for improvement (Bathellier *et al.*, 2018, Flamholz *et al.*, 2019).

1.3.1 Structure and forms of Rubisco

The form I Rubisco holoenzyme has a molecular mass of 560 kDa. In land plants and green algae, the chloroplast genome contains the *rbcL* gene that encodes for 55-kDa large subunit and multiple nuclear-encoded RbcS genes encoding 15-kD small subunits.(Spreitzer & Salvucci, 2002). The large subunit is highly conserved throughout all phyla (Cohen *et al.*, 2006). The RbcL consists of two distinct domains, the N-terminal and C-terminal domain. The N-terminal α+β domain comprises about 150 amino acids and the C-terminal domain is a β8α8 triosephosphate isomerase (TIM) - barrel and is about 320 amino acids long. The large subunits RbcL are associated in a head to tail dimeric arrangement with an active site being formed by residues in N- domain of one sub-unit and C- domain of adjacent subunit (Whitney *et al.*, 2011). The RbcS subunits are more diverse in sequence, generally with only 30–40% homology. Their common core structure consists of four-stranded antiparallel β-sheets covered on one side by two helices. The most significant variation occurs in two distinct locations, the loop between β-strands A and B, and the C-terminal region which is elongated in enzymes of certain organisms. The diversity in small subunit structure appears to significantly contribute to functional differences between Rubisco enzymes (Hauser *et al.*, 2015).

Rubisco is classified into four forms namely form I, II,III and IV (Portis & Parry, 2007). Form I Rubisco is the most widespread and functionally active only when eight large RbcL and eight small RbcS subunits form a hexadecameric complex (L_8S_8). The form II Rubisco is found in proteobacteria and dinoflagellates and they consist of multiples of dimeric RbcL units $(L_2)n$. The first dimeric structure to be solved from form II was that of *Rhodospirillium rubrum* (Schneider *et al.*, 1986, Schneider *et al.*, 1990) Recently a hexameric form II Rubisco from *Rhodopseudomonas palustris* was reported (Satagopan *et al.*, 2014). Some phototrophic and chemoautotrophic organisms also

harbour operons of both form I and form II Rubiscos (Badger & Bek, 2008). Depending on CO_2 concentration of environment, they utilize either of two Rubiscos as form I functions in low CO_2 conditions and form II takes over at high CO_2 levels (Dangel & Tabita, 2015) The form III Rubisco is found only in archaea that lacks a CBB cycle but still catalyses the carboxylation reaction (Tabita *et al.*, 2008). They exist as L_2 or $(L_2)n$. The form IV Rubiscos are Rubisco like proteins (RLP) and so far, have been found to have a (L_2) arrangement. These are unable to catalyse the carbon fixation reaction as they lacked many essential active site residues (Figure 1.3) and one has been shown to function in the methionine salvage pathway (Ashida *et al.*, 2008). A RLP subfamily involved in the degradation pathway of four carbon sugar acids was recently discovered(Zhang *et al.*, 2016), but the majority of RLP proteins have no assigned function.

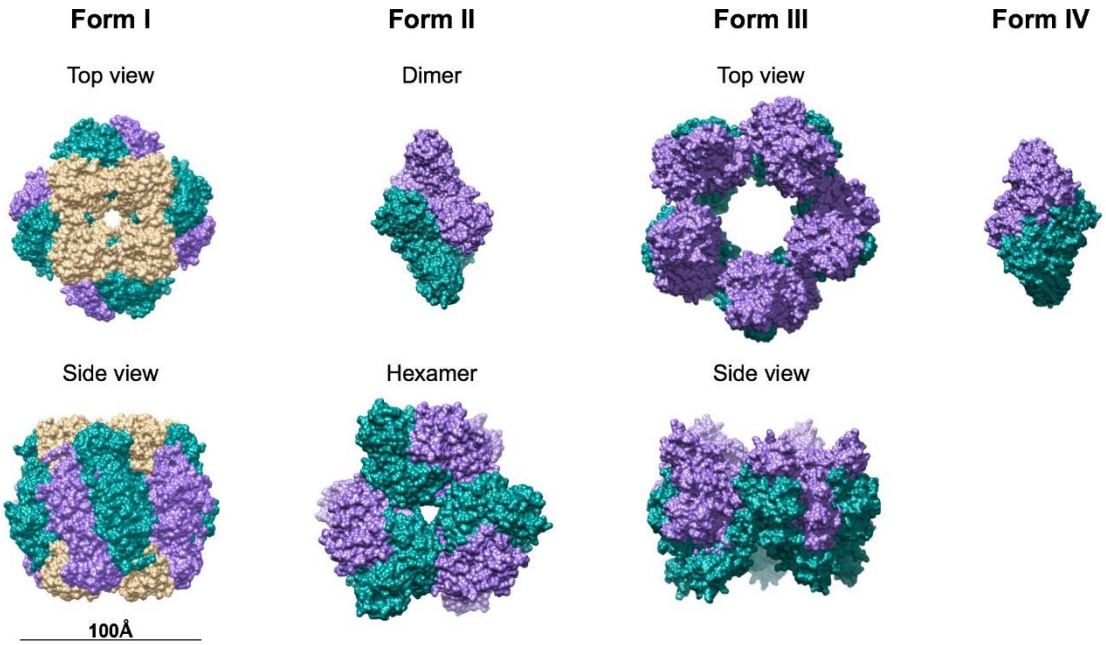


Figure 1.9 **Forms of Rubisco:** Depicting different forms of Rubisco with structures giving example for each distinct type. Each structure was retrieved from the PDB (Form I -1RCX, Form II dimer-5RUB, Form II hexamer- 4LF1, Form III- 1GEH, Form IV- 1YKW) and rendered using PyMOL. Rubisco exists in four distinct forms namely form I, II, III and IV.

Form I Rubisco is phylogenetically divided into green and red type clades (Tabita *et al.*, 2008). The green-type Rubisco is found in plants, algae and cyanobacteria and it can be further subdivided as form IA and form IB. The red-type Rubisco is found mainly in red algae, phototrophic bacteria and subdivided into form IC and form ID. Some red-type

red algal Rubiscos have higher CO_2/O_2 specificity compared to green type (Whitney *et al.*, 2001). They are considered kinetically superior to other Rubisco forms (Parry *et al.*, 2013). It is interesting to note that all sequences for small subunits belonging to red-type have longer C-terminal region. Red-type and green-type Rubisco can also be divided based on the presence or absence, respectively, of the small subunit βE-βF loop (Spreitzer, 2003) (Figure 1.4).

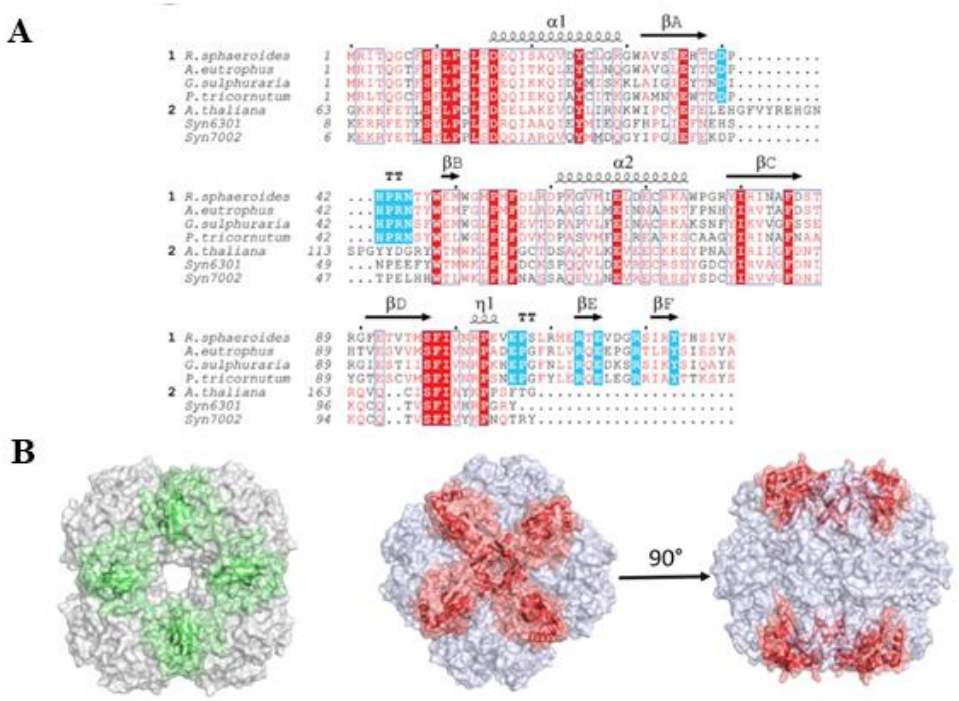


Figure 1.10 **Comparison of red-type and green-type Rubisco:** A. Showing the sequence alignment of representative red-type (1) and green-type (2) Rubisco small subunit sequences.(Joshi et al., 2015). B. Showing the RbcS subunit arrangements in the crystal structures of green- and red-type Rubiscos from Synechococcus sp. PCC 6301 and Galdieria sulphuraria (PDB codes 1RBL and 4F0K respectively).

In plants and green algae Rubisco large subunit is encoded in chloroplast whereas Rubisco small subunit is encoded by nuclear genome (Hartman & Harpel, 1994). Prokaryotic organisms generally consist of form IA, IB, and IC rubisco whereas eukaryotic organisms are predominantly form ID or IB rubisco. The folding of prokaryotic rubisco's large subunit is performed by chaperonins GroEL/GroES while chaperones Cpn60/20/10 drives the folding of eukaryotic rubisco (Bhat *et al.*, 2017). Prokaryotic rubisco functions in a wider range of metabolic contexts compared to the photosynthetic eukaryotic plant enzymes (Liu *et al.*, 2017).

Rubisco displays a rich molecular diversity in its occurrence. It exists in all domains of life, from versatile bacteria and Archaea to higher plants and eukaryotes. These diverse Rubiscos are placed in different clades. The four well-studied forms of Rubisco are detailed out in previous sections. By exploring evolutionary variations in Rubiscos, we discover new properties and occurrences for this ancient enzyme. Long it was established that form II Rubiscos were dimers until the firm assignment of a hexameric oligomeric state for form II Rubisco in *Rhodopseudomonas palustris* (RpM) (Satagopan *et al.*, 2014). The first snapshot of form II Rubisco catalysis was gleaned from the crystal structure RpM Rubisco complexed with the transition-state analogue 2-carboxyarabinitol 1,5-bisphosphate (CABP). This crystal structure revealed a unique arrangement of active sites where the C-terminus is pointing away from the active site, whereas in CABP bound Form I structures this element is folded over the active site. Another newly characterized CABP-bound hexameric form II Rubisco from GWS1B, an uncultivated bacterium of the *Gallionellaceae* family in β-proteobacteria exhibited the same C-terminal conformation (Varaljay *et al.*, 2016).

Study of Rubisco from organisms inhabiting extreme environments enhance the existing diversity of Rubisco by differing structurally and sometimes functionally. They showcase distinct properties that were not spotted earlier in model organisms. One such example is the Rubisco of thermophilic methanogenic archaea *Methanococcoides burtonii* which assembles into oligomers when stimulated by substrate RuBP and it shows a distinct Rubisco Assembly Domain (RAD) consisting of 29 amino acids in proximity to the C-terminal of large subunit (Gunn & Valegard, 2017). Although previously classified as a form II enzyme, this new structural information about MbR Rubisco places it in a sub-group of form III Rubiscos called form IIIB. This includes other Rubiscos from the Methanosarcinales order which also possess the RAD and better reflect the role of MbR Rubisco (Gunn & Valegard, 2017). It is likely that as more Rubisco structures become available, the traditional Rubisco findings will require constant revision to accommodate new information uncovered (Figure 1.5).

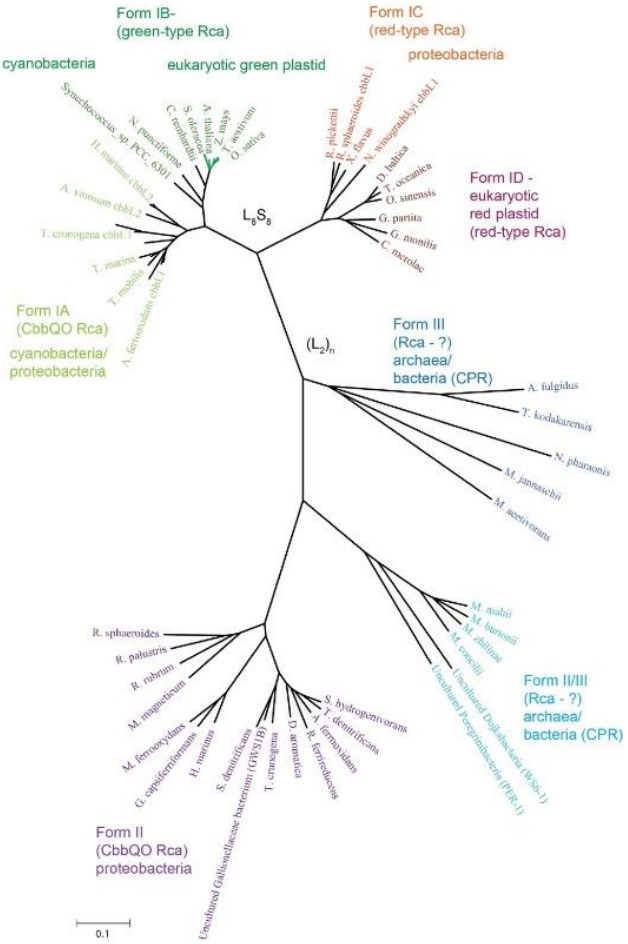


Figure 1.11 **Phylogenetic tree of carboxylating Rubiscos:** An updated evolutionary tree of Rubisco forms showing recently added Form II/III to the traditional classification. Form IV Rubisco is not shown here as they do not catalyse RuBP dependent CO2 fixation. 57 sequences of Rubisco collected from public databases were aligned using CLUSTALW alignment and evolutionary tree was produced using neighbourhood joining in Mega 7 software suite.

1.3.2 Photorespiration and CCMs (Carbon Concentrating Mechanisms)

Rubisco displays poor substrate specificity as it can also perform the oxygenation of RuBP with O_2 to form one molecule of 3-PGA and one molecule of 2-PG (Phosphoglycolate) (Andersson, 2008). The 2-PG must be recycled to 3-PGA via a wasteful process known as photorespiration using ATP and NADPH, wherein it also releases a previously fixed CO_2 molecule (Bauwe *et al.*, 2010). The net efficiency of photosynthesis is reduced by up to 50% by this photorespiratory pathway, which also severely affects the plant's water-use efficiency. Photorespiration also reduces crop productivity by at least 30% (Zhu *et al.*, 2010). To overcome the slow carboxylation velocity, plants produce Rubisco in huge amounts, likely making it the most abundant protein on earth (Ellis, 1979). This poor substrate discrimination is owed to the properties of the two gases. They are of similar size, unbalanced abundances in

occurrence and Rubisco originally evolved in a CO_2 rich atmosphere. In contrast the current atmosphere is O_2 rich. To succumb to this shortcoming, many photosynthetic organisms devised unique mechanisms to either concentrate CO_2 concentration around Rubisco or increase Rubisco's CO_2/O_2 specificity (Whitney *et al.*, 2011). Some plants evolved biochemical pathways to concentrate CO_2 like in C4 and crassulacean acid metabolism (CAM) plants. In C4 plants leaf's adaptation is termed Kranz anatomy and Rubisco is localized in bundle sheath cells that lack oxygen evolving photosynthetic complexes, and fixed CO_2 in the form of C4 acids such as malate is shuttled to these bundle sheaths (Leegood, 2002). CAM plants function similarly, but utilize a temporal separation of CO_2 uptake and fixation (Dodd *et al.*, 2002).

Another concept is the biophysical carbon-concentrating mechanisms (CCMs) found in some proteobacteria, all cyanobacteria and the majority of unicellular algae. Specialized compartments called carboxysomes in cyanobacteria (Rae *et al.*, 2013) and pyrenoids (Meyer & Griffiths, 2013) in algae concentrate Rubisco, which enables to possibility of establishing a high concentration of CO_2 around Rubisco to compensate for its poor kinetic properties (Espie & Kimber, 2011). Inorganic carbon is transported into the cells by bicarbonate and CO_2 pumps and subsequently into these compartments. These compartments contain carbonic anhydrase that are localized with Rubisco to interconvert bicarbonate and CO_2 (Badger *et al.*, 1998)

1.3.3 Rubisco reaction and regulation mechanism

Rubisco is not catalytically competent *in vivo* once it is assembled. It needs to be modulated by an activation process to attain catalytic competency. The Apo-form (termed E) must be carbamylated at an essential lysine residue at the catalytic site (Lys 201 as per numbering in spinach Rubisco) by a non-substrate CO_2. It is then stabilized by binding to co-factor magnesium ion (Mg^{2+}). This catalytically active ternary complex is termed as ECM (Enzyme Rubisco + CO_2 needed for carbamylation + Mg^{2+} co-factor) (Cleland *et al.*, 1998, Parry *et al.*, 2008) (Figure 1.6).

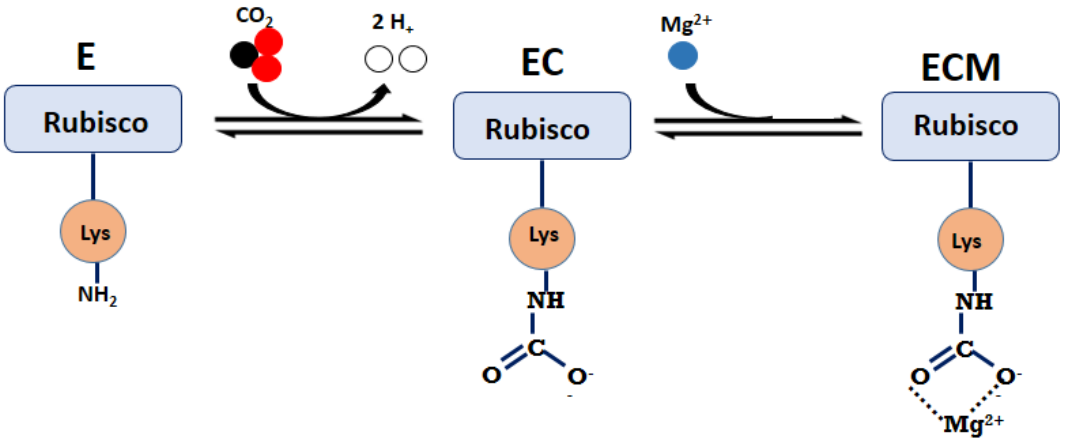


Figure 1.12 **Rubisco activation:** Non-activated Rubisco (E) is carbamylated by non-substrate CO2 (EC) to form activated Rubisco (ECM).

Rubisco can form inactive complexes with several sugar phosphates including its own substrate RuBP. Rubisco depends on molecular chaperones called Rubisco activases to modulate its activity and control its activation state. The mechanism of action of these activases varies depending on different organisms. The Rubisco enzyme gets inhibited and forms EI, where it needs synergistic action of Rubisco activase to cope with a variety of inhibitory mechanisms. Some of these inhibitory mechanisms are regulated by nature such as the nocturnal inhibition by 2- carboxy-arabinitol 1- phosphate (CAIP) or inhibition of decarbamylated Rubisco by its substrate RuBP, which in turn produces intrinsic side reaction by-products like xylulose-1, 5- bisphosphate (XuBP) and D-glycero-2, 3-pentodiulose-1, 5-bisphosphate (PDBP) (Mueller-Cajar *et al.*, 2014, Hazra *et al.*, 2015). Inhibitors bind either before or after carbamylation and block the active site of the enzyme, preventing carbamylation and/or substrate binding. The removal of tightly bound inhibitors from the catalytic site of the carbamylated and decarbamylated forms of Rubisco requires Rubisco activase and the hydrolysis of ATP. The motor protein Rubisco activase restores catalytic competence by facilitating conformational changes in Rubisco that promote the release of inhibitory compounds from the active site. In this way, Rubisco activase ensures the Rubisco active site is not blocked by inhibitors and it is free for carbamylation or to participate directly in catalysis (Parry *et al.*, 2008) (Figure 1.7 Removal of the inhibitor by Rubisco activase is ATP dependent, consistent with its classification as a molecular machine belonging to the AAA^+ (ATPases associated with various cellular activities) superfamily of P-loop ATPases. The canonical higher plant Rca consists of an N-domain bearing the Walker A and B

motifs followed by a more divergent C-domain with a conserved structural core composed of two helical hairpins with a left-handed twist (Henderson *et al.*, 2011). It is a thermo-labile enzyme that is active in its hexameric form. The N-terminal α/β sub-domain and a C-terminal α-helical sub-domain with a nucleotide-binding site located between the domains that display considerable conformational flexibility (Hasse *et al.*, 2015).

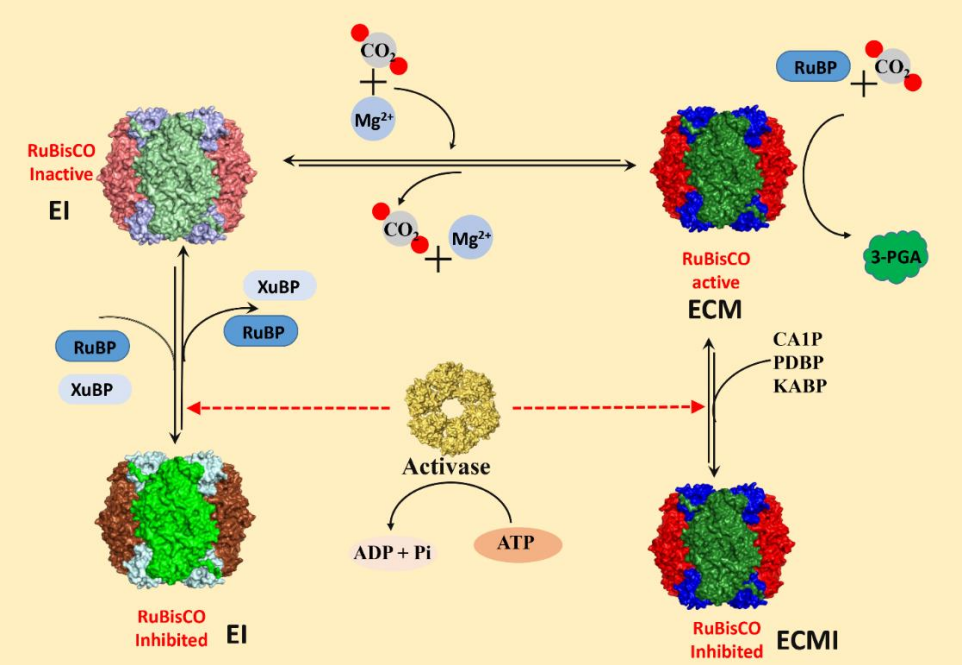


Figure 1.13 **Activation of Rubisco by its activase:** A scheme depicting the regulation of Rubisco activity by Rubisco activase. The inactive, active and inhibited forms of Rubisco is shown along with the inhibitors. It shows that activase is an ATP dependent motor protein.

1.3.4 Remodelling by Rubisco activases

There are currently three known types of activases. They are the green-type activase Rca, which activates green-type Rubisco (Stotz *et al.*, 2011), red-type CbbX for red-type Rubisco (Mueller-Cajar *et al.*, 2011) and a third type of activase for chemoautotrophic bacteria (Tsai *et al.*, 2015) (Figure 1.8). Rca from tobacco forms a hexameric arrangement in its functional state. The N-terminal domain operating in conjunction with helical insertion of C-terminal functions for Rubisco recognition. Loop segments exposed towards central pore of hexamer are required for ATP-dependent remodelling of Rubisco, resulting in the release of inhibitory sugar (Stotz *et al.*, 2011). In case of red-type it is allosterically regulated by the binding of RuBP to a conserved pocket in the C-terminal α-helical subdomain. The surface-accessible C-terminal

peptide of RbcL subunit is transiently pulled into central pore of CbbX, mediated by CbbX ATPase, resulting in the release of inhibitory RuBP (Mueller-Cajar *et al.*, 2011, Bhat *et al.*, 2017). The third class of Rubisco CbbQO isoforms system of activases is found in chemo-lithoautotrophic bacteria. The CbbQ and CbbO form hetero-oligomers to function as specific activases for two structurally distinct Rubisco types. CbbQ and CbbO work conjunctionally as a motor and substrate adaptor respectively and bind Rubisco through a von Willebrand factor A domain on CbbO. In *Acidithiobacillus ferrooxidans* the CbbQ1O1 activates form I Rubisco wherein the CbbQ2O2 is specific for a form II enzyme encoded in the same genome (Tsai *et al.*, 2015).

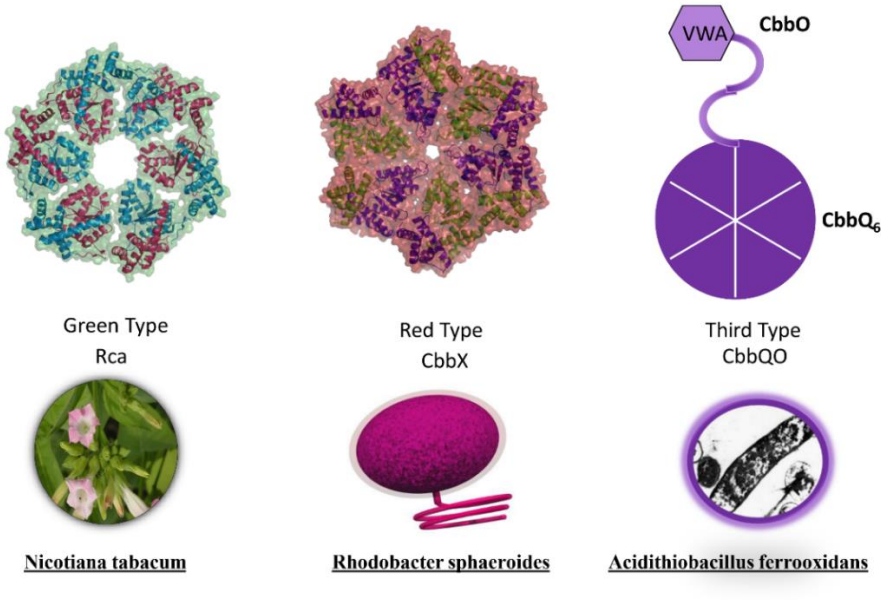


Figure 1.14 **Types of Rubisco activases:** Depicting the hexameric forms of different Rubisco activases. The green-type activase is from tobacco plant, the red-type from proteobacterium of Rhodobacter genus and CbbQO from proteobacterium of Acidithiobacillus genus.

1.4 Biogenesis of Rubisco

For decades biotechnological ambitions to enhance Rubisco function have been hampered by the complicated biogenesis requirements of the enzyme, particularly that from eukaryotic sources (Mueller-Cajar & Whitney, 2008). The biogenesis of Rubisco obligatorily involves folding chaperonins and often requires additional assembly chaperones to form the functional Rubisco holoenzyme. Folding of RbcL subunits into assembly competent folded subunit is achieved by the assistance of folding chaperonins GroEL and GroES (or the homologues found in the relevant compartments). Sometimes,

assembly to functional holoenzyme requires additional assistance from assembly chaperones (Bracher *et al.*, 2017).

A total of three assembly chaperones RbcX, Rubisco accumulation factor Raf1 and Raf2 have been identified for green-type form I Rubisco (Liu *et al.*, 2010, Feiz *et al.*, 2012, Feiz *et al.*, 2014, Wheatley *et al.*, 2014, Hauser *et al.*, 2015, Fristedt *et al.*, 2018). RbcX and Raf1 occur in organisms encoding form IB Rubisco such as plants, cyanobacteria and green algae. Raf2 is encoded in organisms with green-type form IA Rubisco and certain eukaryotic organisms with form IB Rubisco.

Cyanobacterial form IB Rubisco is encoded by an operon consisting of *rbcL* and *rbcS* genes, which are often separated by the *rbcX* gene. The protein product of the *rbcX* gene is for proper Rubisco biosynthesis and assembly (Onizuka *et al.*, 2004, Emlyn-Jones *et al.*, 2006). It has been demonstrated that cyanobacterial RbcL protein is folded by GroEL/GroES complex and retains high affinity for this chaperonin until its C-terminal fragment is captured by the RbcX dimer (Saschenbrecker *et al.*, 2007). After the release of folded RbcL from GroEL, RbcX stabilizes RbcL dimers and facilitates $RbcL_8$ core assembly. Incorporation of RbcS into holoenzyme results in decreased affinity of RbcX to RbcL, which leads to release of C-terminal fragment of large subunit and formation of fully functional Rubisco enzyme. Specific assembly chaperones may be required more generally in the formation of complex oligomeric structures when folding is closely coupled to assembly (Liu *et al.*, 2010). RbcX proteins are functionally alike and composed of two dimers. Two highly conserved regions were identified, one at the central groove of each dimer, which contains several hydrophobic residues. The other conserved region is a polar surface region. The central groove region is essential and facilitates the formation of soluble RbcL subunits, whereas the polar surface region ensures that RbcL subunits are properly arranged. It is found that conserved peptide (EIKFEFD), derived from C terminus of RbcL, and specifically interacts with the central groove of RbcX dynamically (Saschenbrecker *et al.*, 2007) (Figure 1.9).

Identification of Raf1 in *Zea mays* was reported by Feiz et al.,(Feiz *et al.*, 2012). Raf1 from *Synechococcus elongatus* acts downstream of chaperonin-assisted RbcL folding by stabilizing RbcL antiparallel dimers for assembly into $RbcL_8$ complexes with four Raf1 dimers bound. Raf1 displacement by RbcS results in holoenzyme formation. Raf1

fulfils a role similar to that of assembly chaperone RbcX, thus suggesting that functionally redundant factors ensure efficient Rubisco biogenesis (Hauser *et al.*, 2015).

Raf2 is a member of PCD family and it is required for Rubisco accumulation in maize. It could play a role in chaperoning RbcS for imported into chloroplast and also assembly with RbcL (Feiz *et al.*, 2014) In a recent structural analysis of bacterial Raf2 homolog encoded in the α-carboxysomes operon, co-expression with Rubisco and GroEL/GroES in *E. coli* increased the concentration of soluble, assembled Rubisco (Wheatley *et al.*, 2014). Recently, *A. thaliana* Rubisco in *E. coli* by was successfully expressed by co-expressing chaperonins Cpn60/Cpn20 and assembly factors, Raf1, Raf2, RbcX, and Bsd2 (Aigner & Wilson, 2017). The exact mechanism of Raf2 is yet to be established.

Along with the above-mentioned chaperones, chloroplast-specific bundle sheath defective2 protein (Bsd2) (Brutnell *et al.*, 1999) and RbcX-II (Bracher *et al.*, 2015) also play roles in biogenesis of different types of Rubisco. Bsd2 is present in both bundle sheath and mesophyll chloroplasts and it is necessary for Rubisco accumulation in maize. It takes part in the initial biogenesis of Rubisco. It is proposed that Raf2, Raf1 and Bsd2 form transient complexes with Rubisco small subunit, which in turn assembles with large subunit as it is released from chaperones (Feiz *et al.*, 2014). RbcX exists in two isoforms in eukaryotes, RbcX-I and RbcX-II. RbcX-I is closely related to cyanobacterial RbcX and green algae Chlamydomonas reinhardtii contains only RbcX-II. RbcX-II supports the assembly of Rubisco by forming arc-shaped dimers for binding-terminal sequence of RbcL. It works relatively with reduced activity than RbcX-I and appears to have adapted to Rubisco clients due to co-evolution (Bracher *et al.*, 2015). It has been shown assembly chaperone RbcX is not essential form IA cyanobacteria. It is found to be crucial only in form IB cyanobacteria (Emlyn-Jones *et al.*, 2006). It is demonstrated that *Synechococcus elongatus* 7942 Rubisco can be properly assembled by tobacco chloroplast chaperones without the intervention of either cyanobacterial RbcX or CcmM35 (carboxysomal protein) in presence of elevated CO_2 (Occhialini *et al.*, 2015). The mechanism of Rubisco biosynthesis in plants is more complex and not yet fully understood. Recently a breakthrough of successfully expressing plant Rubisco *A. thaliana* in *E. coli* by co-expressing chaperonins Cpn60/Cpn20 and assembly factors RbcX, Raf1, Raf2 and Bsd2 has been demonstrated (Aigner & Wilson, 2017).

The genomes of organisms encoding red-type form IC Rubisco do not contain sequence homologues of assembly factors RbcX, Bsd2, Raf1 or Raf2, pointing to an alternative mechanism of Rubisco assembly. On co-expression in *E. coli*, the RbcL and RbcS subunits from the red-type proteobacterium *Rhodobacter sphaeroides* assembled to holoenzyme with very high efficiency. Red-type form IC RbcL subunits in proteobacterium *Rhodobacter sphaeroides* assembles with help of small subunit and chaperone-independent manner. The C-terminal β-hairpin extension of the red-type RbcS is critical for this efficient assembly and these β-hairpins of four RbcS subunits form an eight-stranded β-barrel that protrudes into the central solvent channel of RbcL core complex. The two β-barrels stabilize complexes through multiple interactions with RbcL subunits. Chimeric green-type RbcS carrying C-terminal β-hairpin renders assembly of cyanobacterial Rubisco independent of RbcX (Joshi *et al.*, 2015). The assembly requirements of red-type eukaryotic form ID Rubiscos have not been discovered.

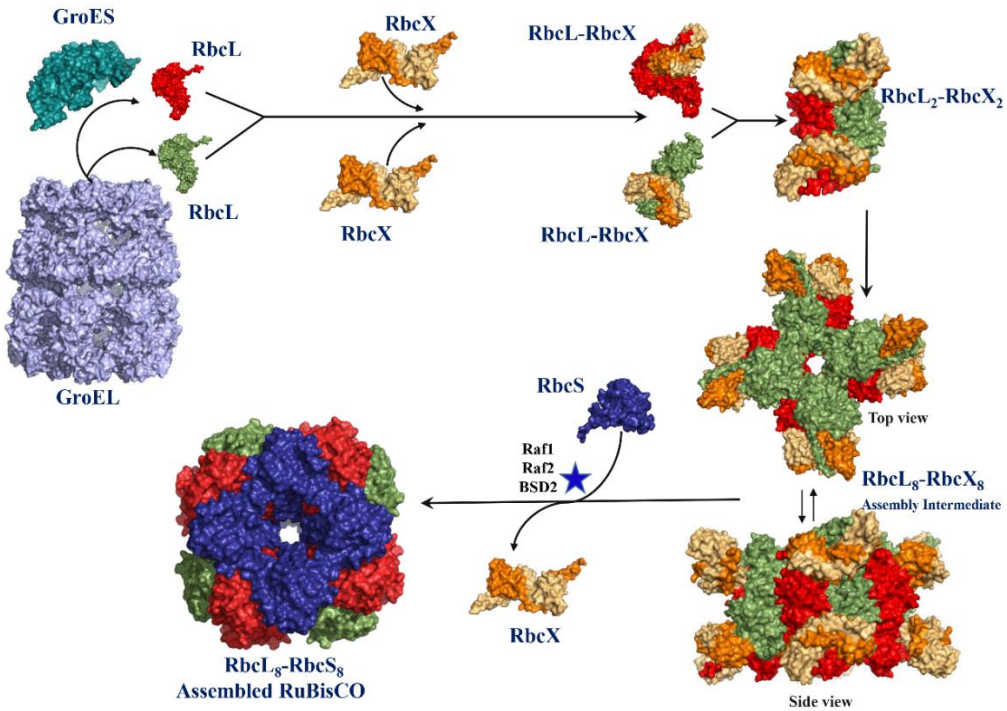


Figure 1.15 **Folding and assembly of Rubisco:** Depicting sequential organisation of chaperones dependent folding and assembly of green-type Rubisco. Structures of GroEL & GroES (1AON), RbcL& RbcX intermediate (3RG6), RbcX (2PEO) and RbcL8S8 (8RUB) was retrieved from PDB and rendered using PyMOL.

1.5 Cyanobacterial platform for *in vivo* studies

1.5.1 Cyanobacteria- The creator of the oxygen rich atmosphere

Cyanobacteria serve as a model organism for photosynthetic research because of its ease in genetic manipulation. It is particularly very useful when aiming at improving photosynthetic processes like CO_2 fixation offering a rich resource for diverse strategies for bioengineering of the organism. They are also quantitatively extremely important organisms on earth accounting for ~25% of organic carbon fixation (Whitton & Potts, 2012). The emergence of oxygenic photosynthesis in ancient Cyanobacteria represents one of the most impressive microbial innovations in Earth's history, and oxygenic photosynthesis is the source of O_2 in the atmosphere today. Cyanobacteria presumably provided the first large-scale biotic source of oxygen on early Earth (Hamilton *et al.*, 2015).

As explained earlier, cyanobacteria have a special adaptation called carboxysomes for surviving in conditions of low CO_2. Carboxysomes are a group of specialized selectively permeable polyhedral proteinaceous bacterial micro-compartments that encapsulate Rubisco. It performs a set of adaptations collectively termed the CO_2-concentrating mechanism (CCM). As already outlined in section 1.3.2 the purpose of the CCM is to support effective CO_2 fixation by enhancing the substrate supply towards Rubisco. The carboxysome is modelled as an icosahedron. The pentameric shell proteins make up vertices and pseudo hexameric shell proteins assemble to form facets. Both types of shell proteins possess central pores of distinct sizes and charge, which presumably allow permeation of bicarbonate, RuBP and 3PGA. Negatively charged bicarbonate can easily pass through the positively charged pores, it is converted to CO_2 gas by internal carbonic anhydrase and then captured by Rubisco. This results in locally increased concentration of CO_2 around encapsulated Rubisco. There are two types of carboxysome (α and β) based on cyanobacterial type with form IA in α-carboxysomes and form IB in β-carboxysomes. It is shown that α-carboxysomes are found in other proteobacteria as well whereas the β-carboxysomes are restricted only to cyanobacteria. The shell proteins of α and β-carboxysomes are Cso1A/C and CcmK2 whose sequence similarities range from 44-57% (Kinney *et al.*, 2011, Zarzycki *et al.*, 2013).

Rubisco engineering of cyanobacteria was first reported comprising a cyanobacterial mutant where Rubisco was substituted with that from the photosynthetic anaerobe *Rhodospirillum rubrum*. This mutant showed carboxylase gene is an essential gene for

both photoautotrophic and photoheterotrophic growth. A mutant lacking endogenous carboxylase gene can be constructed only by supplementing it with functional replacement gene. The growth of cyanorubrum mutant was 2-3 times slower than wild-type *Synechocystis* sp. PCC 6803. This attempt failed to produce a superior cyanobacterial strain but was intended to serve as a system to screen libraries for carboxylase mutants for substrate specificity (Pierce *et al.*, 1989).

A subsequent attempt was made to construct a cyanobacterium that can serve as a target organism for studying a selection of variant plant Rubisco. The obstacle to studying higher plant enzymes is its inability to produce an active assembled Rubisco. This can be overcome by functionally expressing the cyanobacterial enzymes encoded by rbcL and rbcS genes clustered in a single transcriptional unit in *Escherichia coli*. Added to this calculated specificity factor for cyanobacterium enzyme is half or less the value of higher plant enzymes. Hence cyanobacterial species can serve as an intermediate to study plant Rubisco (Amichay *et al.*, 1993). As retarded growth and oxygen sensitivity was reported earlier, a new attempt was made to develop means for screening ability of foreign Rubisco to fold and assemble functionally in cyanobacteria using *Synechococcus* PCC 7942. They assumed that shortcomings of cyanorubrum strain are due to inferior catalytic and assembly properties of *R.rubrum* in *Synechocystis sp*. PCC 6803. Hence, they used the *Synechococcus* PCC 7942 and introduced foreign Rubisco by encoding plasmid in conjunction with inactivation of cyanobacterial *rbcLS* operon on the chromosome by replacement with an antibiotic resistance gene. The generation of this new system for expressing foreign Rubisco in Synechococcus PCC7942 provided information about folding and assembly of Rubisco and gave new insights into the plasticity of the carbon concentrating and fixing mechanisms of cyanobacteria (Emlyn-Jones *et al.*, 2001).

A major difference between form I and form II Rubisco is the presence of the small subunit RbcS in form I. The evolutionary significance of small subunit has remained a mystery. The families of *rbcS* genes that encode small subunits were among the first plant nuclear genes cloned, sequenced, and studied for expression. Generally, the form I has higher specificity to CO_2 than form II. This gives a skeptical thought of whether more divergent small subunits make a substantial contribution to differences in kinetic properties. Among form I Rubiscos, red-type has superior kinetic properties than green-type. Small subunits of all red-type Rubiscos have a long C- terminus that forms two additional β strands (E and F). The resulting βE–βF loop resides in the central-solvent

channel of the holoenzyme. Regardless of the reasons for their origin, small-subunits, today, comprises a phylogenetically diverse set of polypeptides. Analysis of hybrid enzymes has shown small subunits can influence Rubisco catalytic efficiency and specificity. Hence small subunits should be genetically engineered to contemplate the design of a better Rubisco (Spreitzer, 2003).

1.6 *In vitro* reconstitution of Rubisco forms

It is established that RbcL of Rubisco undergoes chaperone-assisted folding and assembly. The folding chaperone GroEL along with its co-factor GroES entraps and folds the RbcL subunit in a cage-like structure where GroEL acts as the folding chamber and GroES is its enclosing lid. The first and major discovery in the avenue of research displaying *in vitro* systems for Rubiscos was done using the dimeric form II Rubisco, which showed that the form II Rubisco from *R. rubrum* merely required the chaperonin system GroEL and GroES to fold and assemble the functional dimeric enzyme (Goloubinoff *et al.*, 1989). Almost two decades later, another significant discovery of the assembly chaperone RbcX was made for cyanobacterial Rubiscos where RbcX acts downstream of chaperonins assisted RbcL folding by interacting and stabilizing RbcL subunits for effective assembly to the octameric $RbcL_8$ core and *in vitro* demonstrations of the same was put forth (Liu *et al.*, 2010). Following this one more assembly chaperone Raf1 aiding in folding and assembly of cyanobacterial Rubisco was discovered which functions similar to RbcX using different interaction sites (Hauser *et al.*, 2015). It was later shown that red-type Rubiscos can assemble into functional forms with only the help of small-subunits (Joshi *et al.*, 2015) (Figure 1.10).

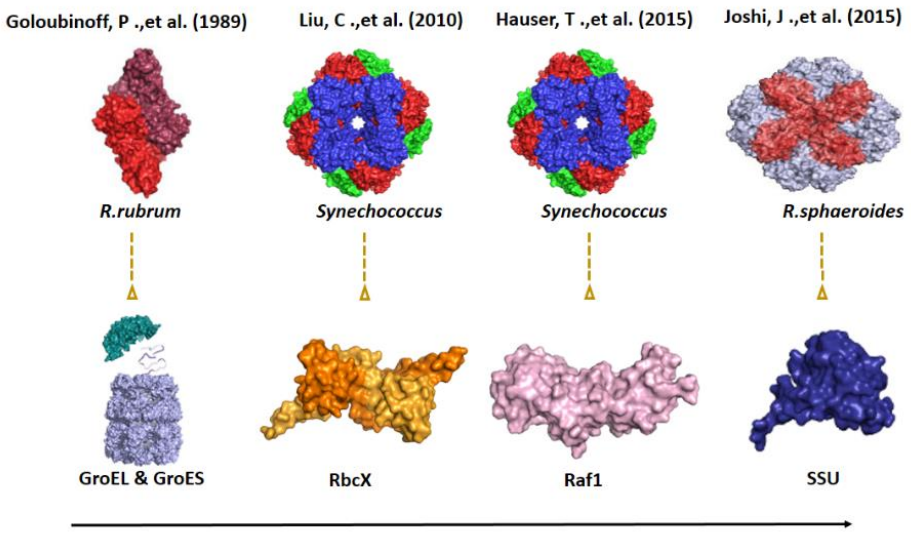


Figure 1.16 ***In vitro* folding and assembly of Rubisco:** Depicting the key milestones regarding the journey of *in vitro* reconstitution studies regarding Rubisco.

1.7 Aim and scope of the study

1.7.1 Understanding Rubisco folding and assembly by *in vitro* methods

Form II Rubisco requires merely the chaperonin system whereas form I enzymes generally have more complex assembly requirements including the need for dedicated assembly chaperones. It is seen that majority of *in vitro* experiments for form II Rubiscos are performed using dimeric Rubiscos while recent structural addition, hexameric form II Rubisco are yet unexplored regarding their assembly requirements. Similarly, of the diverse Rubisco forms the most widely studied are form IB Rubiscos found in eukaryotic green plastids and cyanobacteria whereas proteobacterial form IA Rubiscos are less investigated. Hence, to bridge these research gaps and cumulatively understand Rubisco systems, we aimed to obtain insights on the folding and assembly requirements of hexameric form II Rubiscos and hexadecameric form IA Rubisco. It was also observed that well established *in vitro* reconstitution of Rubisco has so far only used the radioactive Rubisco assay which is laborious and reduces experimental throughput. We currently lack the understanding of how to implement the necessary folding machinery of any eukaryotic Rubisco (irrespective of green or red-type) in *Escherichia coli* (Mueller-Cajar & Whitney, 2008). Hence, we tried to perform the GroEL/ES assisted refolding using a technically simpler procedure and by quantifying the enzymatic activity spectrophotometrically. This was aimed at rapid assessment of the effect of various assembly chaperones from different sources (such as RbcX, Raf1 and Raf2

(Hauser *et al.,* 2015) on diverse Rubisco large subunits using *in vitro* assays with purified components. The rationale behind this attempt is to refold and assemble the different forms of Rubisco in an easier manner and to monitor the stepwise refolding. It will help us to devise a novel method that can be extended and used for all forms of Rubisco. This method once established will be of great use in understanding the efficient refolding and assembly of different Rubisco systems. Once the method is developed, we are aiming to perform the *in vitro* assays for the form IB cyanobacterial rubisco to evaluate and understand if this method can be extended to these assembly chaperone dependent systems also. Our long-term goal is to systematically establish conditions that allow certain Rubisco large subunits to achieve a folded and assembled state *in vitro*. We will then know if and what additional components are required for successful Rubisco production in the cyanobacterial or other systems such as plant chloroplasts.

1.7.2 Engineering a novel *in vivo Synechocystis* sp. PCC 6803 cyanobacterial platform for Rubisco assembly

Most of the photosynthetic organisms depend on Rubisco for their carbon fixation. Though it is a vital enzyme, Rubisco has poor catalytic efficiency and its activities necessitate the energetically wasteful photorespiratory pathway. Efforts to enhance carbon fixation in plants have traditionally focused on Rubisco or on approaches that can help to remedy Rubisco's undesirable traits. This has made Rubisco a popular target for improving photosynthetic efficiency (Zarzycki *et al.,* 2013). Attempts to engineer cyanobacterial Rubisco in plants did not yield expected outcomes because of problems in Rubisco folding and assembly (Whitney *et al.,* 2011). Efforts made for modification of cyanobacteria did not provide insights into the concept of a fitter strain than the wild type (Pierce *et al.,* 1989). The aim of this project was to engineer the *Synechocystis* sp. PCC 6803 strain as an intermediary platform for expressing and studying Rubisco systems in chloroplast given the biochemical similarities between plant chloroplasts and *Synechocystis* sp. PCC 6803 (Emlyn-Jones *et al.,* 2001). We are interested in establishing Cyanobacteria as a platform for expression of different kinds of Rubisco systems (Rubisco and activases) to verify and investigate their compatibility, activity, efficiency in the context of a photosynthetic cell. The long-term aim is the generation and examination of tools for synthetic biology to study photosynthetic carbon assimilation of global food supply and industrially important Rubisco-dependent

organisms like plants and cyanobacteria. This would also provide a scope for improving the overall carbon concentration mechanism and carbon fixation.

2 MATERIALS AND METHODS

2.1 Materials

All the chemicals used in the current study were of at least analytical grade. Chemicals other than the ones listed here were purchased from Sigma-Aldrich.

2.1.1 General chemicals

Table 1: General chemicals

Chemical	Brand
30 % Acryl /Bis solution 37.5:1	BioRad
Agarose (molecular grade)	Vivantis
Acetic acid (Glacial)	Merck
Ammonium per sulphate (molecular grade)	Promega
Ampicillin	Gold Biotechnology
Bacto agar	BD Difco
Bleach 5X concentrate	GB bleach
Chloramphenicol	Gold Biotechnology
Coomassie brilliant blue G-250	Applied Chemisty
Creatine kinase	Roche
DTT, Cleland's reagent	Thermo Scientific
dNTP Set 100 mM solutions	Fermentas
EDTA, disodium salt di-hydrate	BioRad
Ethidium bromide	BioRad
Ethanol 99.5-99.8%, absolute	Merck
Glycerol	Affymetrix
D-Glucose	Sigma – Aldrich
HisPur^{TM} Ni-NTA resin	Thermo Scientific
IPTG	Gold Biotechnology
Kanamycin sulphate	Gibco
Methanol, GR, min 99.8%	Merck
Magnesium chloride hexahydrate	MP Biomedicals
Proteinase K	Promega
Phosphoenolpyruvate (PEP)	Roche
SDS, Sodium dodecyl sulphate	Hoefer

Skim milk powder	Sigma-Aldrich
Sodium phosphate, monobasic	Affymetrix
Sodium chloride	Merck
Sodium hydroxide, analysis grade pellets, 99%	Merck
Sodium thiosulphate	Merck
SYBR Safe ® DNA gel stain	Invitrogen
Tryptone	BD Difco
TEMED	BioRad
Tris- HCl Base	BioRad
X-gal	Fermentas
Yeast extract	Affymetrix

2.1.2 Buffers and Media

Buffers

Table 2: Buffers

Buffer	Components
Coomassie staining solution	0.1% serva Coomassie blue G-250, 40% ethanol, 8% acetic acid
Coomassie de- staining solution	10% ethanol, 7% acetic acid
Electrophoresis migration buffer	120 mM Tris-HCl, 134 mM glycine, 0.1 % SDS
Lysis buffer A	50 mM Tris-HCl pH 8.0, 50 mM NaCl
Lysis buffer B	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM imidazole
Elution buffer	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 200 mM imidazole
Native loading buffer	62.5 mM Tris-HCl, 40% glycerol, 0.1% Bromophenol blue
PonceauS staining solution	0.1% (w/v) PonceauS, 5% acetic acid
TAE	40 mM Tris-acetate, pH 8.5, 2 mM EDTA
TBS	50 mM Tris-HCl pH 8.0, 137 mM NaCl, 2.7 mM KCl
TBS-T	TBS, 0.1% v/v TWEEN 2.0

Western transfer buffer	50 mM Tris-HCl pH 8.0, 20% methanol, 192 mM glycine

Media

Media were prepared with deionised, double distilled water and subsequently autoclaved.

LB medium:

Table 3: LB medium components.

Components	Amount (in g/L)
Trptone	10
Yeast extract	5
NaCl	10
Agar (for solid)	15

pH 7.1 was adjusted using sodium hydroxide.

BG 11 medium:

Table 4: BG 11 medium components.

Stock	Components	Amount
100x BG-11	$NaNO_3$	149.60 g/L
	$MgSO_4.7H_2O$	7.49 g/L
	$CaCl_2.2H_2O$	3.60 g/L
	Citric acid	0.60 g/L
	0.25 M NaEDTA, pH 8.0	1.12 mL/L
	Trace minerals	100.0 mL/L
	Deionized water	Up to 1 L
Trace minerals	H_3BO_3	2.860 g/L
	$MnCl_2.4H_20$	1.810 g/L
	$ZnSO_4.7H_20$	0.222 g/L
	$Na_2MoO_4.2H_2O$	0.390 g/L
	$CuSO_4.5H_2O$	0.079 g/L
	$Co(NO_3)_2.6H_20$	0.049 g/L

	Deionized water	Up to 1L
1000 x	Ammonium iron(III) citrate, brown	0.60g/100 mL
1000 x	Na_2CO_3	2.00 g/100 mL
1000 x	K_2HPO_4	3.05 g/100 mL

To obtain 1 L of BG 11, 987 mL of deionized water was combined with 10 mL of 100x BG 11 and 1 mL each of 1000 x ammonium iron (III) citrate brown, 1000 x sodium carbonate and 1000 x Dipotassium phosphate. For making solid media 15.0 g/L agar and 3 g/L sodium thiosulphate were added. After autoclaving, the medium was cooled to 55°C before adding glucose or antibiotics if required.

2.1.3 Molecular biology materials

All the primers were ordered from IDT technologies. Other materials used are listed below:

- Pfu polymerase, Taq polymerase, Phusion polymerase
- *Nde*I, *Bgl*II, *Nco*I, *Sac*II, *Xba*I, *Hind*III, *Dpn*I
- T4 DNA Ligase
- Calf intestine phosphatase (CIP)
- Pyruvate kinase/Lactate dehydrogenase from Rabbit muscle
- Glyeraldehyde -3-P dehydrogenase
- Creatine P-kinase from rabbit muscle
- 5x DNA loading dye
- GeneRuler 1 kb DNA ladder
- MiniPrep Plasmid Kit
- Wizard SV Gel and PCR Clean-Up System
- pGemT vector system

The following table gives the list of primers used in the study:

Table 5: List of primers used.

Primer	Purpose	Sequence
AfacRAF	To clone the assembly chaperone from *Acidithiobacillus ferrooxidans*	AfacRAF_For 5'-ATCCGCGGTGGAATGAACGCTGTAAAGGATCTG- 3'
		AfacRAF_Rev 5'- ATAAGCTTTTAGGTTCTTGCCGTAGCGGC-3'
Syn6301_RbcX	To clone the assembly chaperone RbcX from *Synechococcus elongatus PCC. 6301*	Syn6301_RbcX For 5'-ATCCGCGGTGGAATGCAATTTATGGGTACAGCC- 3'
		Syn6301_RbcX For 5'-ACGCGGCCGCTCAATCCGCATGGGAGGCATT- 3'
Syn6301_Raf1	To clone the assembly chaperone Raf1 from *Synechococcus elongatus PCC. 6301*	Syn6301_Raf1 For 5'-ATCCGCGGTGGAATGAATGCTCTTGCGATGCGTGAG- 3'
		Syn6301_Raf1 Rev 5'- ATAAGCTTTCAAACTGCCAAGGTGTTGCG-3'
Syn6301_S	To clone the Rubisco small subunit from *Synechococcus elongatus PCC. 6301*	Syn6301_S For 5'-ACCATATGATGAGCATGAAAACTCTGCCC-3'
		Syn6301_S Rev 5'- ATGGATCCTTAGTAgCGGCCGGGACGATG-3'
RsCbbLS	To clone the large and small subunit from	RsCbbLS_For 5’-TCGAGGCGGCTGCCGCCGT-3’

	Rhodobacter sphaeroides	RsCbbLS_Rev 5’CGAGATCTTCAGCGGACGATGCTGTGC-3’
RsCbbS_sdm	Site directed mutagenesis to remove endogenous BglII site	RsCbbS_sdm_For 5’-CGACGCGCAGATATCGGCGCA-3’
		RsCbbS_sdm_Rev 5’-TGCGCCGATATCTGCGCGTCG-3’
RsCbbX	*To clone the activase from Rhodobacter sphaeroides*	RsCbbX_For 5’GCATCGTCCGCTGAAGATCTTCCTTATGACCGACGCGGCAA-3’
		RsCbbX_Rev 5’GGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGTCATCTGGCCAGAGCCTCC-3’
psbAII	*To clone the whole of promoter region*	GCATCGTCCGCTGAAGATCTTCCTTGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACATGACCGACGCGGCAACG
Syn6803RbcL	To clone the large subunit of *Synechocystis pcc 6803*	Syn6803RbcL_For 5’-ATGGTACAAGCCAAAGCAGGGTTT AAGG-3’
		Syn6803RbcL_Rev 5’-TTA GAG GGT ATC CAT GGC CTC GAA-3’
W1	To check the segregation of the cyanobacteria	5’-TGTCATCTATAAGCTTCGTG-3’
W2		5’-ATCCGCCGGCAGACGTTCTTCC-3’
CmR	To amplify the chloramphenicol cassette	CmR_For 5’CAAATACGGTCGTCCTCTGCTTGGTTGATGCCATGGAGAGTAAAATCCTCAG-3’
		CmR_Rev 5’GACACTTGGCCAAAACCCGGAAGTTTACGCCCCGCCCTG-3’

2.1.4 Protein chemistry materials

This part explains protein expression analysis using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), Native PAGE, Western blot and immuno-detection.

- 5X SDS loading dye (225 mM Tris-HCl pH 6.8, 50% glycerol, 5% w/v SDS, 5% β-mercaptoethanol, 0.01% w/v Bromophenol blue)
- PageRuler Plus Prestained Protein Ladder 10-250 kDa
- Nitrocellulose membrane
- Ponceau S staining solution (0.1% w/v Ponceau S, 5% acetic acid)
- Anti-RbcL, Anti-Flag, Anti-CbbX, Anti-mouse, horseradish peroxidase conjugated Anti-rabbit
- Enhanced chemiluminescence (ECL) detection solution

2.1.5 *In vitro* reconstitution assay materials

- 20 mM MOPS/KOH (pH 8), 100 mM Potassium chloride, 5 mM magnesium chloride
- 6M Guanidinium hydrochloride
- Assay buffer (20 mM MOPS/KOH,100 mM KCl, 5 mM $MgCl_2$)
- 1 mg/mL Bovine Serum Albumin (BSA), 10 mM DTT, 1 mM EDTA
- 0.6 µM GroEL, 1.2 µM GroES (chaperones)

2.1.6 Rubisco assay materials

- Coupling enzyme (Creatine P-kinase, Glyceraldehyde-3-P dehydrogenase, 3-phosphoglycerate kinase and Triose-P isomerase/Glycerol-3-dehydrogenase)
- 40 mM $NaHCO_3$
- 1 mM NADH
- 1 mM ATP
- 1 mM RuBP
- 10 mM Creatin-P
- 1 mM DTT

2.1.7 General instruments

Table 6: General instruments.

Instrument	Brand
Akta Purifier Frac900, Akta micro	GE health care
Bench top microcentrifuge	Kintaro
Cimarec stirrer	Thermo Scientific
CPS controller	Shimadzu
Excella e25 Incubator	New Brunswick
Innovo 44R	New Brunswick
Forma 900 (-80^{o} C)	Thermo Scientific
MaxQ 6000 incubator	Thermo Scientific
Minicentrifuge	Green liv
Mini-protean tetra system	Bio-Rad
MonoQ HR 10/100	GE health care
Nanodrop 2000	Thermo scientific
Professional TR10 Thermocycler	BioMetra
Power pac Basic	BioRad
Power pac HC	BioRad
PL-600 pH/mV/Temp meter	Ezodo
Rocker 400	Rocker
TB2 thermo block	BioMetra
Thermomixer comfort	Eppendorf
Trans-Blot	Invitrogen
Semi – dry transfer cell	Invitrogen
SourceQ 15 and 30	GE health care
Sonics Sonicator	Vibra cell
Superdex 200 16/600, 10/300, 3.2/30	GE health care
UV-1800 spectrophotometer	Shimadzu
Vortex genie 2	Scientific industries
Z300K refrigerated bench top centrifuge	Hermle

2.2 Molecular biology strategies

2.2.1 Amplification of DNA by Polymerase Chain Reaction (PCR)

Amplification of genes from plasmids as well as site-directed mutagenesis was performed using high fidelity polymerase, Pfu polymerase (Thermo Scientific) with proofreading activities. Plasmid/Genomic DNA was used as a template for PCR. To amplify coding regions of interest, primers incorporated with restriction sites were designed. The PCR was set up as 20 µL reaction mix on ice with appropriate concentration of constituents pipetted into PCR tube, reagents were mixed by short centrifugation. The constituents of the PCR mix and PCR program for amplification is given below:

Table 7: PCR reaction mix

Component	Volume (µL)
MilliQ water	12
Template (plasmid or genomic)	2
Primers (10 µM)	2 × 0.5
Buffers (according to the polymerase)	2
dNTP's (2 mM)	2
Polymerase enzyme	1

The PCR steps were repeated for 30 cycles. PCR product was mixed with 5µL of 5x DNA loading dye and applied to 1% TAE- agarose gel supplemented with 5 µL of SYBR Safe DNA stain. GeneRuler 1kb was loaded as the marker. Agarose gel electrophoresis was performed in TAE buffer at 100 V for 40 minutes. DNA fragments were then extracted from the gel using Wizard SV Gel and PCR clean-up kit according to manufacturer's instructions.

Table 8: PCR cycle conditions.

Steps	Temperature (^{o}C)	Time
Initial denaturation	95	5 min
Denaturation	95	30 sec
Annealing	(depends on Tm)	30 sec

Extension	72	2 min
Final extension	72	5 min
Pause	10	-

2.2.2 pGem T Vector Cloning

PCR product was polyadenylated at both 5' and 3' ends to allow cloning of PCR product into pGem T. Polyadenylation was done by adding 0.2 µL of Taq polymerase and 2 µL of 2 mM dATP to amplified DNA and incubated at 72^{o} C for 30 minutes. Polyadenylated PCR product was fractionated using agarose gel electrophoresis as explained earlier (Section 2.2.1) and subsequent gel extraction was done. Ligation of polyadenylated PCR product into pGem T was carried out as per manufacturer's instructions. The ligation mixture was transformed into DH5α and plated on ampicillin LB plate supplemented with X-gal. It enables blue-white screening of pGem T and only white colonies containing the insert was further inoculated to LB medium and grown overnight in shaking incubator at 37^{o} C, 220 rpm. The plasmids were then purified, and concentrations were measured using Nanodrop 2000 spectrophotometer at 260 nm. The plasmids were then sent for DNA sequencing (First base sequencing, Singapore) to verify the correct protein- encoding sequences.

2.2.3 Site-directed Mutagenesis

Site-directed mutagenesis was used to introduce point mutations in the existing plasmids. QuikChange protocol (Stratagene) was used according to manufacturer's instructions. Primers used were designed to have 12 complementary nucleotides to the plasmid at the 3' and 5' end with the centre of the primer containing the desired nucleotide changes. Forward and reverse primers were reverse complementary to each other. DpnI treatment was used to remove the original plasmids. 1 µL of DpnI was added to the mutagenesis PCR product followed by incubation at 37^{o} C for one hour. Dpn1 treated PCR product was subsequently transformed into DH5α, colonies seen were inoculated and purified plasmid was checked using sequencing for desired mutations.

2.2.4 Restriction digestion and ligation

0.5- 2 µg of DNA was digested using 20-50 U of restriction (single or double digestions) in 20 µL reactions with buffers and incubated at 37^{o} C for 1-3 hours. For single

digestion, 5 U of calf intestinal phosphatase (CIP) was added and incubated at 37 °C for 1 hour or overnight. Restriction digestion of plasmid produce linearized plasmids (vector) and DNA fragments flanked with respective overhangs. These digested products were separated by 1% agarose gel electrophoresis and purified using Wizard SV Gel and PCR Clean-Up system. Concentrations of the products were measured using a NanoDrop. Subsequently, ligation reaction between vector and purified DNA fragments was set up at room temperature for 3 hours. The ligation mixture was calculated based on 1:3 molar ratio of vector to insert. The ligation reaction mixture consists of the following:

Table 9: Ligation reaction mix.

Ligase buffer (10 X)	2
Vector	6
Insert	2
T4 DNA ligase	0.5
Water	9.5

2.2.5 Chemical transformation

50 µL chemically competent *Escherichia coli* DH5α cells were thawed on ice and 100 ng of plasmid was added to these cells. This mixture was incubated on ice for 30 minutes. After incubation, the cells were heat-shocked at 42 º C for 90 seconds and placed back on the ice for 2 minutes. Recovery media (LB or SOC) was added and the cells were incubated in a shaking incubator at 37 º C, 1200 rpm for one hour. Cells were plated onto selection agar plates supplemented with respective antibiotics and incubated overnight at 37º C.

2.2.6 Restriction free cloning

Restriction-free cloning (aka overlap extension PCR cloning or ligation independent cloning) is a PCR-based method for creation of custom DNA plasmids. Essentially, it allows for the insertion of any sequence into any position within any plasmid, independent of restriction enzyme recognition sites or homologous recombination sites within these sequences. To accomplish this, a pair of hybrid primers were designed

containing a complementary sequence to both the desired insert and the target plasmid. These primers were used to amplify the insert from an appropriate source using high-fidelity PCR conditions. The resulting product was purified and used as 'mega-primer' in the secondary PCR reaction, with the target plasmid acting as a template. The plasmid is amplified in both directions, and the mega-primers act as long single-stranded overhangs that allow the complementary strands of the plasmid to anneal, forming a nicked hybrid molecule. DpnI is used to degrade any parental plasmid (based on its methylation), and the final product is used to transform competent bacterial strains normally.

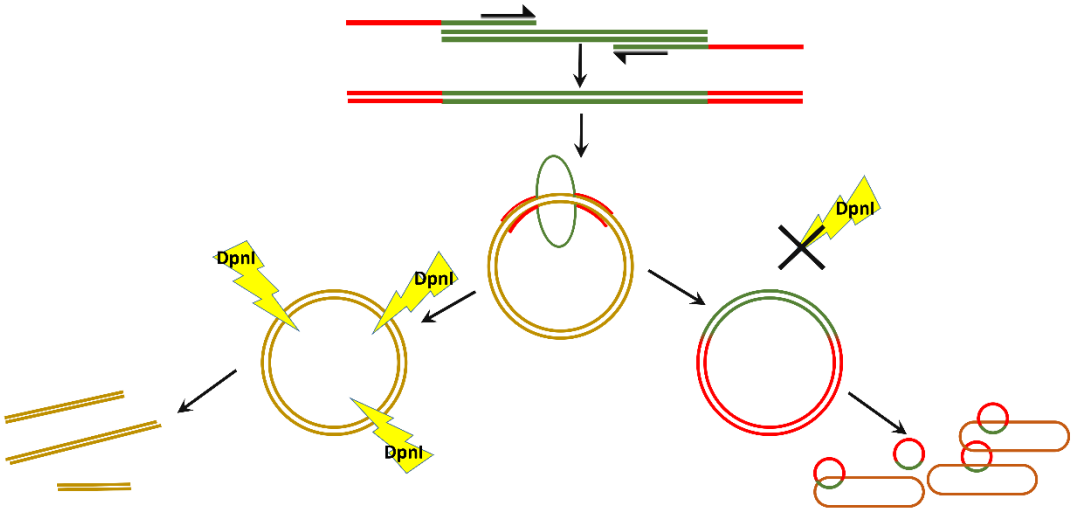


Figure 2.1 **Restriction free cloning:** Picture depicting the sequential steps of restriction free cloning. Green colour represents the desired insert and red colour the overlap from the plasmid. The light brown lines show the circular desired plasmid.

The primary PCR was performed first followed by the secondary reactions. The conditions for primary PCR reaction are given below in Table 10:

Table 10: Primary PCR mix for restriction free cloning.

Component	Volume (µL)
MilliQ water	31.5
Template (plasmid DNA containing the desired insert)	2
Primers (10 µM)	2 × 2.5
10 X Pfu Buffer	5
dNTP's (2 mM)	5
Pfu Polymerase	1.5

Table 11: Primary PCR conditions for restriction free cloning.

Steps	**Temperature (^{o} C)**	**Time**	**Number of cycles**
Initial denaturation	95	5 min	1
Cycle denaturation	95	30 sec	25
Cycle annealing	Depends on primer Tm	30 sec	
Cycle extension	72	1min/kb	
Final extension	72	5 min	1
Pause	10	-	1

After the primary PCR, the product of the reaction was called the mega-primer and used for the subsequent secondary PCR reactions. The conditions of secondary PCR are state below:

Table 12: Secondary PCR mix for restriction free cloning.

Component	Volume (µL)
MilliQ water	12.9
Vector	1 (35 ng)
Megaprimer	20 x 35 ng x bp of insert / bp of vector
5 X Phusion polymerase buffer	4
dNTP's (2 mM)	0.4
Pfu Polymerase	0.2
DMSO	0.6

Table 13: Secondary PCR conditions for restriction free cloning.

Steps	Temperature (º C)	Time	Number of cycles
Initial denaturation	95	5 min	1
Cycle denaturation	95	30 sec	20
Cycle annealing	Depends on primer Tm	30 sec	
Cycle extension	72	25 sec	
Final extension	72	5 min	1
Pause	10	-	1

2.2.7 Cloning strategies for assembly chaperones used in the study

Cloning and construction of the plasmids used for the *in vitro* folding and assembly study (other than those detailed below and in the respective purification results) was performed and kindly provided by our Post-doc Yi-Chin-Tsai.

2.2.7.1 Cloning details of AfacRAF into the pHue vector

The coding sequence of *AfacRAF* for Rubisco accumulation factor2 of *Acidithiobacillus ferrooxidans* (Gene ID: AFE_1681, Uniprot: B7JB20) was amplified from its genomic DNA with the flanking 5'SacII and 3'HindIII restriction site using the primers

AfacRAF_For and AfacRAF_Rev (Table 5). The PCR product was subsequently polyA tailed, cloned into vector pGem-T (Section 2.2.2) and subsequently digested with SacII and HindIII. Linearized insert and vector fragments were separated by agarose gel electrophoresis and purified using Wizard Plus SV Gel and PCR Clean-up (using manufacturer's instructions). Target insert and pHue vector were ligated at room temperature and transformed into competent *E.coli* DH5α. The expression plasmid was designated as pHue_AfacRAF and correct insert sequence was verified by DNA sequencing (First base, Singapore).

2.2.7.2 Cloning details of Syn6301_RbcX into the pHue vector

The coding sequence of *Syn6301_RbcX* for assembly chaperone of *Synechococcus elongatus PCC. 6301* (Uniprot: Q31N04) was amplified from its genomic DNA with the flanking 5'SacII and 3'NotI restriction site using the primers Syn6301RbcX_For and Syn6301RbcX _Rev (Table 5). The PCR product was polyA tailed, cloned into vector pGem-T (Section 2.2.2) and subsequently digested with SacII and NotI. Linearized insert and vector fragments were separated by agarose gel electrophoresis and purified using Wizard Plus SV Gel and PCR Clean-up (using manufacturer's instructions). Target insert and pHue vector were ligated at room temperature and transformed into competent *E.coli* DH5α. The expression plasmid was designated as pHue_Syn6301_RbcX and correct insert sequence was verified by DNA sequencing (First base, Singapore).

2.2.7.3 Cloning details of Syn6301_Raf1 into the pHue vector

The coding sequence of *Syn6301_Raf1* for assembly chaperone of *Synechococcus elongatus PCC. 6301* (Uniprot: Q5N472) was amplified from its genomic DNA with the flanking 5'SacII and 3'HindIII restriction site using the primers Syn6301Raf1_For and Syn6301Raf1 _Rev (Table 5). The PCR product was polyA tailed, cloned into vector pGem-T (Section 2.2.2) and subsequently digested with SacII and NotI. Linearized insert and vector fragments were separated by agarose gel electrophoresis and purified using Wizard Plus SV Gel and PCR Clean-up (using manufacturer's instructions). Further ligation and transformation into competent *E.coli* DH5α was carried out as explained above for Syn6301_RbcX . The expression plasmid was

designated as pHue_Syn6301_Raf1 and correct insert sequence was verified by DNA sequencing (First base, Singapore).

2.3 Protein Biochemistry Methods

2.3.1 Protein quantification using Bradford assay/ Nanodrop

Protein mixtures were quantified by Bradford assay (BioRad) after calibration against known amounts of bovine serum albumin. 2-10 μg of BSA/ purified proteins were added to 500 μL autoclaved H_2O and 500 μL Bradford reagent, the absorbance of the sample was measured at 595 nm on a spectrophotometer and concentration was calculated.

Also, pure proteins were quantified by their absorbance at 280 nm by Nanodrop and the concentration was calculated using the extinction coefficients calculated with the program ProtParam. Final concentration of the protein was calculated using extinction coefficient and molecular weight based on the following formula:

Concentration of the protein (mg/mL) = (A_{280}/ε) X Molecular weight

A_{280}-Absorbance measured by Nanodrop at 280nm

ε – Extinction coefficient $M^{-1}cm^{-1}$

2.3.2 SDS- PAGE

In order to analytically separate proteins based on apparent molecular weight, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used. Gels were prepared as per standard protocol. Samples were mixed with 5x SDS-sample loading dye, boiled for 5 minutes at 95°C, and centrifuged for 1 minute at room temperature at 16000 g before loading to 12.5% or 16% SDS polyacrylamide gel. 5 μL of the pre-stained marker was loaded alongside the samples. Gel electrophoresis was performed at a starting voltage of 120 V followed by 160 V for the rest of the run.

2.3.3 Native- PAGE

6% Native polyacrylamide gel was prepared as per standard protocol and gel electrophoresis was performed similarly to previously described SDS-PAGE protocol. Supernatant of the samples from cell lysate or the protein was mixed with 5X Native loading buffer and loaded onto 6% Native polyacrylamide gel. Purified Rubisco proteins

were loaded along with samples as control and these were separated at 180 volts for 2 hours. The entire procedure was performed at 4 °C.

2.3.4 Coomassie staining of SDS-PAGE and Native-PAGE

Coomassie blue dye integrates with cationic, nonpolar, and hydrophobic side chains of a polypeptide. Coomassie blue staining was used for the visualization of protein bands in polyacrylamide gels. Protein bands were fixed and stained in the gel matrix by incubating the gels for 1 h with Coomassie blue staining solution. Background staining was removed by incubation with Coomassie de-staining solution for 1-3 h.

2.3.5 Western blotting and immuno-detection

Western blot followed by specific immuno-detection was employed for specific identification of protein bands after SDS and native PAGE. Proteins were transferred to a nitrocellulose membrane in a semi-dry western blot apparatus (BioRad) in transfer buffer at a current of 15 volts per gel for 30 minutes. Transferred proteins were stained with Ponceau S to ensure efficient transfer and de-stained using milliQ water before proceeding with the immuno-detection. Nitrocellulose membranes were blocked for 1 hour at room temperature in TBS with 2% skim milk powder. The primary antibody anti-RbcL, anti-CbbX or anti- FLAG was diluted to 1:5000 in TBS-T and incubated with the membrane for 1 hour at room temperature. The membrane was washed thrice with TBS-T for 10 minutes. The secondary antibody peroxidase-conjugated goat anti-mouse or anti-rabbit was diluted to 1:10000 in TBS-T and incubated with the membrane for 1hour at room temperature. The membrane was washed again (as above) and bound HRP-coupled antibodies were detected by enhanced chemiluminescence (ECL) staining with a 1:1 mix of ECL detection reagent 1 and ECL detection reagent 2. Chemiluminescence was detected on LAS-3000 Imaging System from Fuji.

2.3.6 Small-scale protein expression test

Plasmids carrying the desired constructs were chemically transformed into competent BL21/BL21_pBAD (for chaperonins GroEL/GroES and are under arabinose-inducible pBAD promoter) cells. A single colony was inoculated in 2 mL of LB media containing respective antibiotics and was grown overnight at 37 °C in an Innova44 shaker incubator shaking at 220 rpm. To two fresh tubes each of 5 mL of LB medium containing the respective antibiotics, 50 µL of the overnight culture was transferred and further grown

at 37 ° C for 3- 4 hours until it reaches an optical density of ~0.6 measured at 600 nm (OD_{600}). Then the tubes were shifted to a reduced temperature of 23 ° C and grown further for 0.5 – 1 hour. One tube was induced with 0.5 mM IPTG (0.2% (w/v) L-arabinose was added 30 minutes prior to the addition of IPTG in case of BL21_pBAD) and uninduced sample was used as the control. The culture was grown overnight at 23 °C, 220 rpm. OD600 of the overnight culture was measured and cells were harvested by centrifugation by spinning at 4000 g for10 minutes. Supernatant was collected and re-suspended in 1 mL Lysis buffer (50 mM Tris, 50 mM NaCl, pH 8).1 mM PMSF was added and the cells were sonicated at 30% amplitude for 50 s using 10 s pulses alternated with 50 s of pause. Samples were collected for analysis using SDS-PAGE. Total lysate (L) of 80 µL was added to 20 µL of 5x SDS loading buffer. 100 µL of Lysate was centrifuged at 4 ° C 16000 rpm for 10 minutes and 80 µL of the supernatant (S) + was mixed with 20 µL of 5x SDS loading buffer. Excess supernatant was removed, and the pellet (P) was re-suspended in 125 µL of 1x SDS loading buffer. Samples L (total protein), S (soluble fraction), and P (insoluble fraction) were heated at 95 ° C for 5 minutes for protein denaturation and centrifuged for 2 min before being loaded to 12.5% or 16% SDS-PAGE gels. 5 µL of pre-stained marker (PageRuler™ Plus Prestained Protein Ladder) or 5 µL unstained marker (PageRuler™ Unstained Protein Ladder) was loaded along with the samples to give estimates of protein sizes. SDS-PAGE gel was run at 100 V until the dye passes stacking part and then the voltage was increased to 160V. The protein bands were visualized using coomassie staining, both staining and de-staining were performed at room temperature using the rocking platform.

2.3.7 Protein Expression and purification

His_6Ub-fusion constructs using pHue expression vector (Catanzariti *et al.*, 2004) was transformed into BL21 cells and these cells carrying the desired constructs were inoculated in LB media containing respective antibiotics and grown until the culture reaches OD_{600} = 0.6 (~4 hours) at 37 ºC in an Innova44 shaker incubator shaking at 220 rpm. Subsequently, the temperature was reduced to 23 ºC and grown for another 1 hour. Later the cultured was induced with 0.5 mM IPTG. Following induction, the culture was grown for ~16 hours/overnight at 220 rpm at 23 °C. Overnight culture was harvested by centrifugation using a JA-10 centrifuge (Beckman Coulter) at 4000 g for 20 minutes at 4 °C. Cell pellets were either stored at -20 °C or directly lysed for protein purification. After centrifugation, the supernatant was discarded and the pellet containing the

expressed protein was re-suspended in 40 mL of Lysis buffer (50 mM Tris-HCl, 50 mM NaCl, 10 mM imidazole). 0.3 mg/mL of lysozyme was added to the pellet and incubated on ice for 30 minutes to aid lysis. 1 mM PMSF for protease inhibition was added just before sonication. Sonication of the sample was performed at 30% amplitude with 15 seconds bursts with 30 seconds interval between bursts for total burst of 5 minutes. After sonication, the sample was centrifuged to get rid of cell debris using the rotor JA-25.5 at 22000 rpm for 30 minutes at 4 °C. Affinity chromatography column packed with Ni-NTA resins was equilibrated using lysis buffer and kept at 4 °C prior to binding the supernatant obtained from above centrifugation. The column was washed with wash buffer prior to eluting the His_6Ub-fusion protein using elution buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, and 200 mM imidazole). The Ubiquitin tag was cleaved off from the pooled protein fractions using deubiquitylating enzyme Usp2. After cleaving the tag by overnight incubation with Usp2, the protein solution was applied to Anion exchange columns (Source15Q or MonoQ 10/100) for further purification using linear salt gradients of NaCl. Subsequently, the protein was applied to size exclusion chromatography column Superdex 200 16/600 column if required. The final fractions of pure proteins were pooled and concentrated in an Amicon Ultra-15 centrifugal filter units. Concentrated proteins are used for biochemical studies immediately or the protein was stored in aliquots at -80°C after flash-freezing in liquid N2 (or with 5% glycerol) for future use.

2.3.7.1 Protein purification from inclusion bodies

His_6Ub-fusion constructs using pHue expression vector was transformed into BL21 cells and these cells were inoculated in LB media containing respective antibiotics and grown until the culture reaches OD_{600} = 0.6 (~4 hours) at 37 ºC in an Innova44 shaker incubator shaking at 220 rpm. The temperature was then reduced to 23 ºC after 4 hours and grown for another 1 hour. Cultured was induced with 0.5 mM IPTG. Following induction, the culture was grown for ~16 hours at 220 rpm at 23 °C. Overnight culture was harvested by centrifugation using a JA-10 centrifuge with JA 10 rotor (Beckman Coulter) at 4000 g for 20 minutes at 4 °C. Cell pellets were either stored at -20 °C or directly lysed for protein purification. After centrifugation, the supernatant was discarded, and the pellet was re-suspended in 40 mL of Lysis buffer (50 mM Tris-HCl, 50 mM NaCl, 10 mM imidazole). 0.3 mg/mL of lysozyme was added to the pellet and

incubated on ice for 30 minutes to aid lysis. 1 mM EDTA and 1 mM PMSF was added just before sonication. Sonication of the sample was performed at 30% amplitude with 15 seconds bursts with 30 seconds interval between bursts for a total burst of 5 minutes. After sonication, the sample was centrifuged to get rid of cell debris using the rotor JA-25.5 at 22000 rpm for 30 minutes at 4 °C. The supernatant was discarded, and the pellet was used for purification. Pellet was transferred to a mortar and homogenised using a pestle by applying manual pressure in inclusion body buffer A. Re-suspended until the pellet completely dissolves. 30 mL of wash buffer A was added to the re-suspended pellet and centrifuged in a J-25 centrifuge with JA25.50 rotor (Beckman Coulter) at 20,000 x *g* for 30 minutes. The re-suspension and centrifugation step were repeated with inclusion body wash buffer B. The final supernatant was discarded, and the purified inclusion bodies were dissolved in 20 ml of denaturation buffer. Concentration of the protein was quantified by measuring their absorbance at 280 nm by Nanodrop, then using the formula, Concentration of the protein (mg/mL) = (A_{280}/Extinction coefficient) X Molecular weight. Protein sample was diluted to 1 mg/ml in 10ml of denaturation buffer and dialyzed against 5 L refolding buffer for 16 hrs/overnight at 4 °C for refolding. Refolded proteins were checked for precipitation and centrifuged at 4000 x *g* for 10 minutes, pooled and concentrated. The concentrated samples were flash frozen in liquid N_2 and stored at -80°C.

2.3.8 Analytical size-exclusion chromatography

Analytical size exclusion chromatography of proteins was performed using Akta microsystem (GE Healthcare) on a Superdex200 PC3.2/30 increase column equilibrated with buffer A. The column was calibrated with globular proteins of known molecular weight (ovalbumin 43 kDa, conalbumin 75 kDa, aldolase 158 kDa, ferritin 440 kDa, thyroglobulin 669 kDa).50 µL of the protein sample was injected to the column and analysis was performed at room temperature. Flow rate of 0.075 ml/min in respective buffer was used and UV signal for the protein was detected at 280 nm. 50µL fractions showing positive curves for the protein were collected and analysed with SDS-PAGE or Native-PAGE gels.

2.4 Enzymatic assays

2.4.1 ATPase assay for GroEL/GroES

ATPase activity of chaperonins GroEL/GroES was measured spectrophotometrically using a coupled assay monitoring the oxidation of NADH in buffer containing 100 mM Bis-Tris pH 7.5, 200 mM KCl, 20 mM MgCl2, 2 mM phosphoenolpyruvate, 0.5 mM NADH, 2 mM ATP, 20 U/ml pyruvate kinase, 30 U/ml lactate dehydrogenase (Kreuzer & Jongeneel, 1983). To this mix required volumes of GroEL was added and assayed immediately. The oxidation of NADH, which is directly coupled to the generation of ADP, was followed by the reduction in the absorbance at 340 nm on a UV-1800 spectrophotometer (Shimadzu). The final volume of the reaction was 100 µl. Two different concentrations of GroEL was used for the measurement, 0.5 µM and 1 µM.

2.4.2 Rubisco assay

Rubisco assay was performed as described by Kubien 2011 (Kubien et al., 2011) with minor modifications. This assay measures the oxidation of NADH spectrophotometrically in a 100 µL reaction volume. The assay was performed in a buffer (100 mM Tricine pH 8.0, 5 mM MgCl2) with 3 µL coupling enzymes mixture (Creatine P-kinase, Glyceraldehyde-3-P dehydrogenase, 3-phosphoglycerate kinase and Triose-P isomerase/Glycerol-3-P dehydrogenase), 20-60 mM $NaHCO_3$, 1 mM NADH, 1 mM ATP, 10 mM Creatin-P, 1 mM DTT, 1 mM RuBP, with differing concentrations of Rubisco. The oxidation of NADH is directly coupled to the generation of 3-phosphoglycerate, was followed by the reduction in the absorbance at 340 nm on a spectrophotometer. ECM complex is made by incubating 100 µM Rubisco active sites with 20-60 mM $NaHCO_3$, 10 mM $MgCl_2$

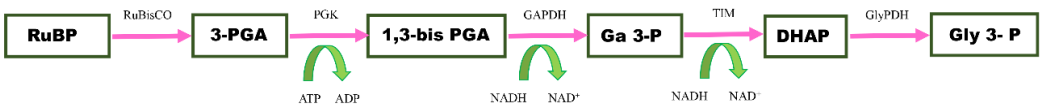


Figure 2.2 **Representation of reactions that occurred during a Rubisco activity assay:** For each carboxylation of RuBP, two molecules of 3-Phosphoglycerate (3-PGA) were produced by Rubisco which were phosphorylated by 3-PGA kinase (PGK). Glyceraldehyde-3-P dehydrogenase (GAPDH) reduces the bis-PGA to an aldehyde (Ga-3-P). Triose-phosphate isomerase (TIM) produces dihydroxyacetone (DHAP) from Ga-3-P, glycerol-p dehydrogenase (GlyPDH) reduces DHAP to Glycerol 3-Phosphate (Gly 3-P). Four NADH being oxidized per RuBP, leading to a decrease in absorbance at 340nm.

2.4.3 *In vitro* reconstitution assay for pure proteins

Individual subunits of Rubisco, Rubisco large subunit and the Rubisco small subunit were mixed together in 1:5 ratio (generally 0.5 µM large subunit and 2.5 µM small subunit in the final Rubisco assay mix) and incubated with 4 mM $NaHCO_3$, 20 mM $MgCl_2$ for 30 minutes. The incubated mix was assayed for rubisco activity as explained in section 2.4.2.

2.4.4 *In vitro* Rubisco folding and assembly assay

2.4.4.1 Electrophoretic *In vitro* Rubisco folding assay

This assay was performed for both form I and form II rubiscos. In the case of form I rubisco which consisted of both large and small subunits, the large subunit was denatured, and small subunit was added later in a stepwise protocol. The presence or absence of rubisco small subunit is mentioned in the respective experiments. In general, rubisco (20 µM) was denatured in 6 M guanidine HCl and diluted 100-fold into ice-cold refolding buffer (20 mM MOPS/KOH- pH 8, 100 mM KCl, 5 mM $MgCl_2$, 10 mM DTT) containing 0.6 µM GroEL (oligomer) (additional components such as $NaHCO_3$, $MgCl_2$, RuBP are included wherever necessary), followed by 10 minutes incubation at 25 °C. Rubisco un-bound to GroEL was removed by centrifugation at 16,000 g for 10 minutes at 4 °C. The GroEL bound rubisco was first quantified to find the proportion of the denatured rubisco using densitometer. The supernatant containing GroEL-bound Rubisco was transferred to new tubes and incubated with 1.2 µM GroES (oligomer) for 5 minutes, ancillary proteins were added if desired. Samples were collected before the start of folding. Refolding was initiated by the addition of 4 mM ATP and end-point samples were collected approximately after 30-60 minutes. The samples were analysed by 6% native- PAGE followed by immunoblotting with anti-Rubisco antibody. Densitometric analysis was performed to quantify the amount of unfolded Rubisco bound to GroEL, in this case samples were collected without ATP addition and analysed using SDS-PAGE.

2.4.4.2 Spectrophotometric Rubisco assembly assay

Rubisco activity was measured using a variation of the coupled enzymatic Rubisco assay described in Section 2.4.2, performed at 25 °C where oxidation of NADH was monitored. Denaturing and binding of Rubisco was done as explained above for the

electrophoretic Rubisco folding assay. Refolding buffer incubated with GroES/ ancillary proteins was added to the Rubisco reaction mix (totally 100 µL) containing 100 mM MOPS/KOH pH 8, 5 mM $MgCl_2$, coupling enzymes, 1 mM NADH, 10 mM phosphocreatine, 40 mM $NaHCO_3$, 1 mM DTT/ATP/RuBP, 0.1 µM Rubisco and assayed continuously for biochemical activity. The assay was performed in triplicates.

2.5 Cyanobacterial growth methods

2.5.1 Construction of *Synechocystis* sp. PCC 6803 (Syn6803) cyanobacterial strains

2.5.1.1 pPD-Flag- *Rs*CbbLS

The *RsCbbLS* fragment was obtained by double digesting the pET30bRscbbLS plasmid (Mueller-Cajar *et al.*, 2011) with NdeI and BglII and was subsequently cloned to the plasmid pPD-Flag (Hollingshead *et al.*, 2012). pPD-FLAG contains the Synechocystis psbAII promoter, a sequence encoding the FLAG tag and flanking sequences for homologous recombination that allows insertion of tagged constructs into the Synechocystis genome in place of the psbAII gene This plasmid was designated as pPD-Flag-*Rs*CbbLS. The psbAII promoter sequence (Table 5) was synthesized and introduced upstream to *Rs*CbbL in pPD-Flag-*Rs*CbbLS. The whole plasmid with promoter region was introduced into *Syn*6803 by cyanobacterial transformation as explained in section 2.5.4. The pPD-Flag-*Rs*CbbLS was expected to integrate into the genome of *Syn*6803 by homologous recombination. This transformation efficiency largely depends on the amount of the DNA introduced and incubation time for which the plasmid was allowed to integrate into the *Syn*6803.

2.5.1.2 pPD-Flag- *Rs*CbbLSX

The next step was the introduction of activase encoding *cbbX* gene into the plasmid system using restriction-free cloning (rfc) to the downstream of the 3' of the Rubisco small subunit gene *cbbS*. The pPD-Flag-*Rs*CbbLS was used as the vector backbone to introduce *cbbX* gene. The insert *cbbX* gene was purified by double digesting the pHueRscbbX plasmid (Mueller-Cajar *et al.*, 2011) and inserted downstream to the large and small subunit in pPD-Flag-*Rs*CbbLS by rfc. After the secondary PCR reaction of rfc, the resulting product was treated with *Dpn*I and heat transformed into DH5α cells. The colonies obtained were re-streaked and the presence of the insert was confirmed

using colony-PCR followed by DNA sequencing. The resulting plasmid designated as pPD-Flag-*Rs*CbbLSX was then introduced into *Syn*6803.

2.5.1.3 pPD-Flag- *Rs*CbbLS- psbAII- *Rs*CbbX

The next construct included an additional promoter region for the *cbbX* gene, as the expression of CbbX using pPD-Flag-*Rs*CbbLSX plasmid was not obseved in the western blot analysis using anti-CbbX primary antibody. This new construct was generated by restriction-free cloning. The psbAII promoter was introduced in front of CbbX in the plasmid pPD-Flag-*Rs*CbbLSX. This plasmid was designated as pPD-Flag-*Rs*CbbLS- psbAII- *Rs*CbbX and introduced into *Syn*6803.

2.5.2 Construction of Syn6803 cyanobacterial endogenous Rubisco knock-out strains

The overall strategy to construct this strain was to disrupt the *Syn*6803 large subunit by a chloramphenicol resistance cassette in the strains constructed above.

2.5.2.1 Amplification of *Syn*6803 large subunit

The large subunit RbcL of *Syn*6803 (Gene ID: 952593) was amplified using PCR. Genomic DNA of *Syn*6803 was used as the template for PCR and the primers Syn6803RbcL_For and Syn6803RbcL_rev was used. The annealing temperature was 55° C for 30 seconds and the extension temperature of 72° C for 2 minutes was used. After the PCR run, the product was analysed using agarose gel electrophoresis.

2.5.2.2 pGEM -T cloning of *Syn*6803 large subunit

The resulting plasmid from *Syn*6803 RbcL amplification was introduced into the pGem T vector as explained in Section 2.2.2. Blue and white colonies screening resulted in colonies carrying the desired insert. The plasmid of the colonies was extracted and used later for the construction of knock-out strain.

2.5.2.3 Amplification of the chloramphenicol-resistance cassette (CmR)

The chloramphenicol cassette was amplified separately using PCR reactions to introduce it into the *Syn*6803 RbcL. Plasmid pSK9 (source Dr. Tiago Selao) (Kuchmina *et al.*, 2012) was used as the DNA template and the primers CmRfor and CmRrev were used. The PCR programme was set up as explained earlier with an annealing

temperature of 55° C for 20 seconds and extension temperature of 72° C for 25 seconds (as per the manufacturer instructions of Phusion polymerase, time was used). After the PCR run, the product was analysed using agarose gel electrophoresis.

2.5.2.4 Insertion of CmR into *Syn*6803 RbcL

The amplified CmR cassette was introduced into the *Syn*6803 RbcL carried by pGem T using restriction-free cloning (Figure 2.3). The primers designed for the rfc facilitated the insertion of CmR in between the amplified *Syn*6803 RbcL gene (Table 5). This is done to increase the efficiency of the endogenous Rubisco knock-out. After the construction of pGem T-CmR- *Syn*6803 RbcL, it was directly transformed into pPD-Flag- *Rs*CbbLS, pPD-Flag- *Rs*CbbLSX and pPD-Flag- *Rs*CbbLS- psbAII- *Rs*CbbX. In this case, BG 11 medium for the starting culture was substituted with 5 mM glucose.

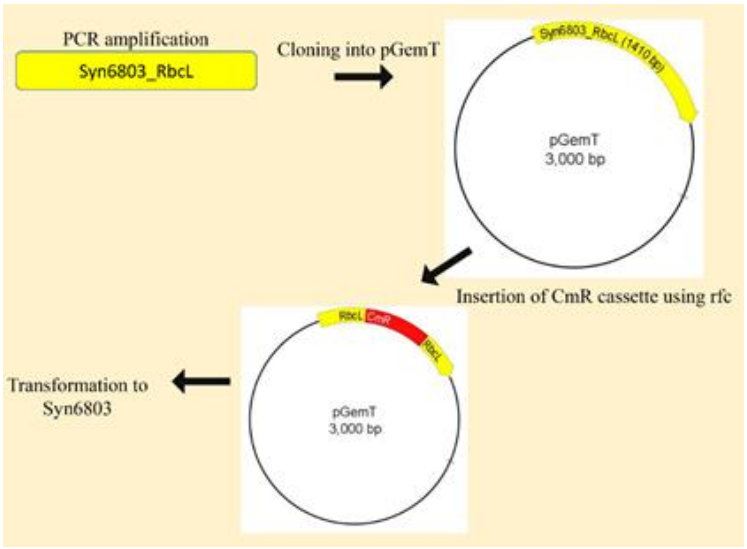


Figure 2.3 **Depiction of knock-out strategy:** Picture depicting the stepwise procedure involved in making the knockout strain of cyanobacteria.

2.5.3 Growing liquid cultures

Cyanobacterial strain used for the study *Synechocystis* sp. PCC 6803 (*Syn*6803) was wild type (WT) - glucose tolerant (GT) strain, kindly provided by Dr. Tiago Selao. A modified Erlenmeyer flask was used to prepare a liquid culture of *Syn*6803. The flask was attached to a glass capillary tube, through a cork. A flexible tubing attached to a 0.2 µM millipore filter was in turn connected to the capillary tube (Figure 2.4). The whole unit was autoclaved before use. For initial culture, a loop full of culture (WT-GT) was inoculated to BG 11 medium, substituted with glucose. While inoculation it was noted

that several clumps of cells were clearly seen in the medium. After inoculation, the set-up was placed in a static growth chamber with continuous bubbling of air, illumination of 15–25 µE/m^2s and incubated at 30° C (Eaton-Rye, 2011).

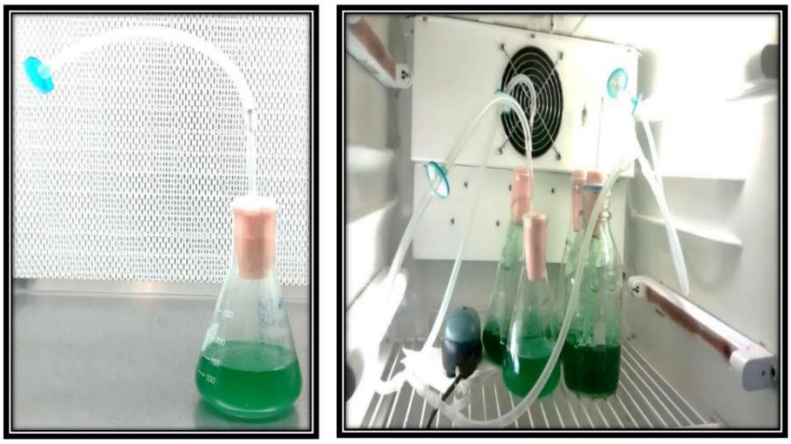


Figure 2.4 **Growth of Cyanobacterial strains:** The image shows the set-up for the growth of the WT-GT cyanobacterial species in BG 11 with 5 mM glucose and no antibiotics. Once the set-up is made, it is maintained in the incubator shown for 2-3 days.

2.5.4 Cyanobacterial transformation

The liquid culture takes 2-3 days to grow. The cyanobacterial culture was harvested before/or at OD_{730nm} of 0.5 as the transformation efficiency declines in late log phase. Before starting the transformation procedure, all work surfaces and equipment were sterilized. Liquid cultures from the flasks were transferred to sterile Falcon tubes. The cells were centrifuged at 3000 g for 5 minutes at room temperature. Fresh medium was supplemented after discarding the supernatant. Then 0.5 mL of culture was aliquoted to sterile test tubes with culture tube caps. About 5-10 µg of plasmid DNA is added to the culture. One tube without the desired plasmid served as control. The tubes were then placed in the growth chamber at 30° C for 6 hours with a light intensity of 15–25 µE/m^2s with occasional shaking of the tubes. After 6 hours the cells were plated on a BG 11 agar plate without antibiotics. It was grown for 24 hours in the growth chamber. Then the cells grown on the plate were completely scraped off and plated on a new BG 11 agar plate containing half the concentration of desired antibiotics (as 12.5 µg/mL of kanamycin) (Eaton-Rye, 2011). The transformants after successful transformation are a result of random homologous DNA recombination events into the host genome. Chromosomal segregation is ensured by selection using antibiotic resistance.

2.5.5 Selection using antibiotics to ensure chromosomal segregation

When colonies appear on the plates containing half concentration of antibiotics, they are plated on fresh BG 11 agar plates containing increased concentrations of antibiotics. This is done to select the segregated transformants. As *Syn*6803 cells contain multiple copies of their genome, it is essential to confirm that all copies contain the desired insert. The *Rs*CbbLS insertion occurs at the locus of psbAII gene, knocking out the redundant, non-essential psbAII gene. In colony-PCR, the signal will be specific for the wild-type state of the targeted genome integration site. In the case of pPD-Flag, the integration of the heterologous gene occurs at the psbAII locus via homologous recombination. Absence of this wild-type-specific PCR signal suggests that all genome copies carry only the *Rs*CbbLS insertion. To obtain a genetically homogeneous recombinant strain, segregation with selection pressure is necessary, by continuously streaking the colonies onto plates containing increasing level of antibiotics. To ensure complete segregation it is necessary to re-streak the transformants approximately every 2-3 weeks, gradually increasing the antibiotic concentration in the subsequent re-streaks. Transformants generally take 5-7 months to completely segregate. If they fail to segregate, it shows that an essential gene has been disrupted (Eaton-Rye, 2011, lange *et al.*, 2011, Reinsvold *et al.*, 2011)

2.5.6 Colony PCR to check segregation

The most efficient way to check the segregation of the transformants is to perform a colony PCR (Eaton-Rye, 2011, Reinsvold *et al.*, 2011). The *Syn*6803 colonies were added to the mix of 40 mM Tris-HCl pH 8.8 with 0.1% triton X using an inoculation loop. It was then heated up to 95° C for 5 minutes followed by centrifugation at 13,000 g for 5 minutes. The following PCR mix was prepared using the *Syn*6803 as the template:

Table 14: Colony PCR mix.

Components	Volume (µL)
Water	4.5
Colony	2
W1for (10 µM)	2.5
W1rev (10 µM)	2.5
5x Taq Buffer	5
dNTP's (2 mM)	2.5
Taq polymerase	1

The PCR programme was set up as explained earlier with an annealing temperature of 55° C for 30 seconds and an extension temperature of 72° C for 1 minute. After the PCR run, the product was analysed using agarose gel electrophoresis.

2.5.7 Sample preparation of *Syn*6803 for protein expression studies

After a considerable number of re-streaks (up to 200 µg/mL of antibiotics), colonies from the plate were inoculated into BG11 medium and grown at 30° C for 2-4 days. The culture was centrifuged at 4000 rpm for 10 minutes and the pellet was re-suspended in 20 mM phosphate buffer. The cells were again centrifuged at maximum speed for 10 minutes. Supernatant was discarded, and the sample was mixed with protease inhibitors (20 µL/mL). The cells were then mixed with glass beads (150-212 µm diameter). This was done to shear the cyanobacterial cell wall and break the cells. Cells were mixed carefully by alternate vortexing and keeping in ice. This was repeated four times. After thorough mixing of the glass beads and cells, 20 mM phosphate buffer pH 7.5 was added to prepare SDS-PAGE or native-PAGE samples. Then the mix was incubated at room temperature for one hour. This sample was analysed by SDS-PAGE or Native-PAGE.

3 FOLLOWING THE *IN VITRO* RECONSTITUTION OF FORM II RUBISCOS IN REAL - TIME

The essential enzyme Rubisco serving as the gateway for inorganic carbons to enter biosphere is notorious for exhibiting poor kinetics and a dual substrate specificity. This made Rubisco a long-standing candidate towards achieving agricultural improvements and biotechnological applications. Rubisco exhibits a rich molecular diversity in occurrence. It is prevalent in all domains of life and made up of large subunits displaying complex multimeric assemblies. This assembly requires a pool of additional factors and follows varied mechanisms (Bracher *et al.*, 2017). A comprehensive toolbox is required to understand these diverse forms of Rubisco and their chaperonic requirements. Apart from assembly, Rubisco requires chaperonins for folding its subunits into correct three-dimensional structures. Many years of research has enlightened us with respect to various mechanisms for folding and assembly. Despite the extensive studies done there are still unexplored assembly pathways for newly discovered Rubisco, the knowledge of which is mandatory for Rubisco synthetic biology.

The first and major discovery in this avenue was done using a dimeric form II Rubisco, where through *in vitro* reconstitution experiments it was established that chaperonins GroEL and GroES were necessary for folding and assembling dimeric form II Rubisco (Goloubinoff *et al.*, 1989). Almost two decades later, discovery of assembly chaperone RbcX for cyanobacterial Rubiscos was proposed and its mechanistic involvement in biogenesis of Rubisco was detailed out (Saschenbrecker *et al.*, 2007, Liu *et al.*, 2010). Similarly plant and cyanobacterial chaperone Raf1 exhibiting a different scaffold was described (Feiz *et al.*, 2014, Hauser *et al.*, 2015). Kinetically superior red-type Rubiscos can assemble into functional forms with only the help of small-subunits (Joshi *et al.*, 2015). Although a lot of research on complex biogenesis of Rubisco enzyme has been done for form IB Rubiscos, other Rubisco forms are mostly unstudied for their assembly requirements. This includes the recent structural addition, hexameric form II Rubisco (Satagopan *et al.*, 2014, Tsai *et al.*, 2015) and proteobacterial Rubisco systems.

Available Rubisco folding, and assembly *in vitro* assays are generally a two-step assay that does not allow to follow the stepwise folding and assembly. We aimed to understand and appreciate the diversity of Rubisco by developing simple, rapid *in vitro* methods to monitor the stepwise assembly of Rubisco. In this study we describe a simple and robust method consisting of electrophoretic and spectrophotometric analyses to understand Rubisco biogenesis *in vitro*. With this method, we monitored the entire folding and assembly process without resorting to radioactive assay.

In this chapter we outline the development and detail the application of this method to study folding and assembly of the dimeric form II *Rhodospirillum rubrum* Rubisco and later to investigate the assembly requirements of hexameric form II *Acidithiobacillus ferrooxidans* Rubisco.

3.1 Purification of Form II Rubiscos

3.1.1 Purification of dimeric Form II Rubisco from the proteobacteria *Rhodospirillum rubrum* (RrM)

The dimeric form II Rubisco from *Rhodospirillum rubrum* was heterologously expressed in *E .coli* using pHueRrcbbM (Tsai *et al.*, 2015) as an N-terminal fusion with a hexa-histidine tagged ubiquitin (His6Ub) to improve solubility. This plasmid was kindly provided by our lab post-doc Dr. Yi-Chin-Tsai. Protein Expression of RrM was confirmed by small-scale protein expression tests, which was followed by large scale purification for soluble proteins. A representative illustration of the large-scale purification is shown in Figure 3.1. The His_6 ubiquitin tagged RrM was purified using an immobilized metal affinity chromatography (IMAC) column followed by precise cleavage of hexa-histidine Ubiquitin from the N-terminus using ubiquitin specific protease (USP2). The overnight His_6 ubiquitin tag cleaved protein was subjected to Anion exchange chromatography (AEC) column for further purification. Purify of the protein was analysed using SDS-PAGE. After two steps of purification, the concentration of RrM pure protein was measured by Nanodrop absorbance at 280 nm and calculated using extinction coefficients.

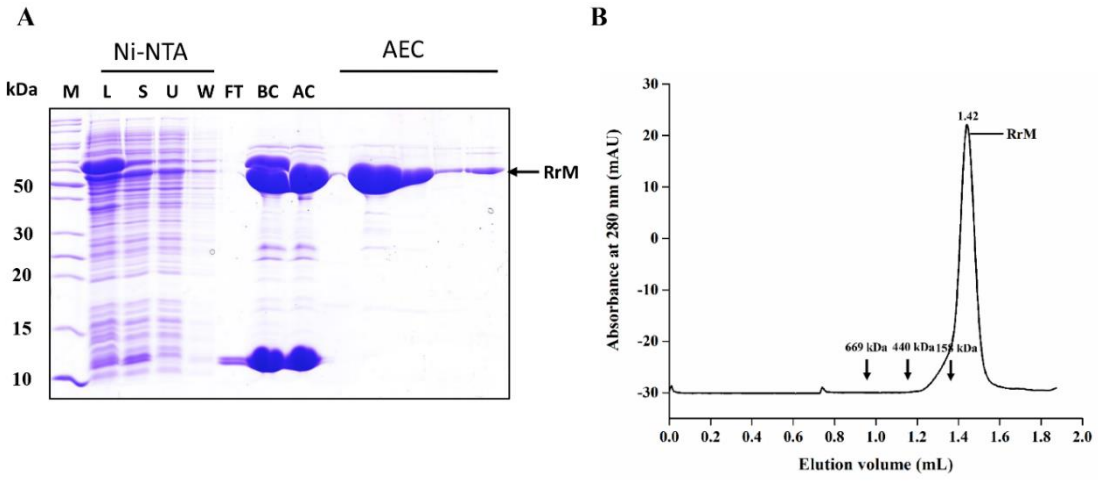


Figure 3.1 **Purification of dimeric form II *Rhodospirillum rubrum* (RrM) Rubisco**: (A) 12.5% SDS-PAGE showing the successive purification steps of RrM Rubisco using immobilized metal affinity chromatography IMAC (labelled as Ni-NTA) and Anion exchange chromatography (AEC) column Resource15Q. M-Molecular weight marker, L-Lysate, S- Soluble, U-Unbound, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC- After cleavage using ubiquitin specific protease (USP2), FT- Flow through from AEC column. (B) Analytical size exclusion chromatogram of RrM using SuperDex 200 16/600 column showing a single peak confirms homogeneity in the purified protein.

3.1.2 Purification of hexameric Form II Rubisco from the proteobacteria *Acidithiobacillus ferrooxidans* (AfM)

Genomic DNA of *Acidithiobacillus ferrooxidans* was used for the cloning of hexameric form II Rubisco and this plasmid was cloned by our post-doc Dr. Yi-Chin-Tsai (pHueAfcbbM) (Tsai *et al.*, 2015). Three-step purification was carried out for *Acidithiobacillus ferrooxidans* (AfM) Rubisco as shown in Figure 3.2. Following the Immobilized metal affinity chromatography and cleaving of His_6 ubiquitin tag, the protein was purified using anion exchange column. The fractions containing the protein from AEC column were concentrated and applied to size- exclusion chromatography column SuperDex 200 16/600 for further purification. The final concentration of pure protein measured using Nanodrop was ~18-20 mg/mL.

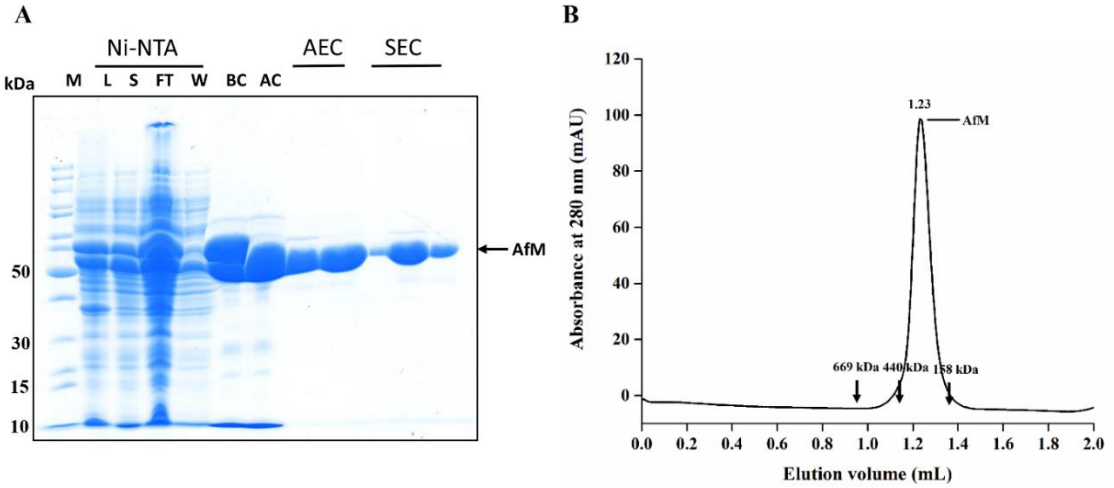


Figure 3.2 **Purification of hexameric form II *Acidithiobacillus ferrooxidans* (AfM) Rubisco:** (A) 12.5% SDS-PAGE showing the successive purification steps of AfM Rubisco using immobilized metal affinity chromatography IMAC (labelled as Ni-NTA), Anion exchange chromatography (AEC) column Resource15Q and Size exclusion chromatography (SEC) column SuperDex 200 16/600. M-Molecular weight marker, L-Lysate, S- Soluble, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC- After cleavage using ubiquitin specific protease (USP2), FT- Flow through from AEC column. (B) Analytical size exclusion chromatogram of AfM using SuperDex 200 16/600 column showing a single peak confirms homogeneity of the purified protein

3.2 Recombinant expression and purification of *E.coli* chaperonins

3.2.1 Purification of *E.coli* chaperonin GroEL

E.coli chaperonin GroEL cloned in pET30b vector was provided by our post doc Dr. Yi-Chin-Tsai and purified using three step chromatography techniques. The supernatant obtained from overnight culture as described in methods section was applied to equilibrated Source 30Q Anion exchange column. Fractions obtained from the Source

30Q column purification was analysed by 12.5% SDS-PAGE and the fractions containing GroEL were pooled together and dialysed overnight in Buffer A (20 mM Tris-HCl, 50 mM NaCl, pH 8.0). The dialysed sample was concentrated and applied to Anion Exchange chromatography Mono 10/100 GL column. The fractions from MonoQ 10/100 was again analysed by SDS-PAGE prior to next step of purification. The concentrated protein sample was finally applied to the size exclusion SuperDex 200 16/600 and pure protein was obtained (Figure 3.3). Concentration of pure protein was ~30 mg/mL. The ATPase activity of purified *E.coli* chaperonin GroEL was measured using continuous enzyme-coupled spectrophotometric assays performed at 25 °C as described in methods section 2.4.1.

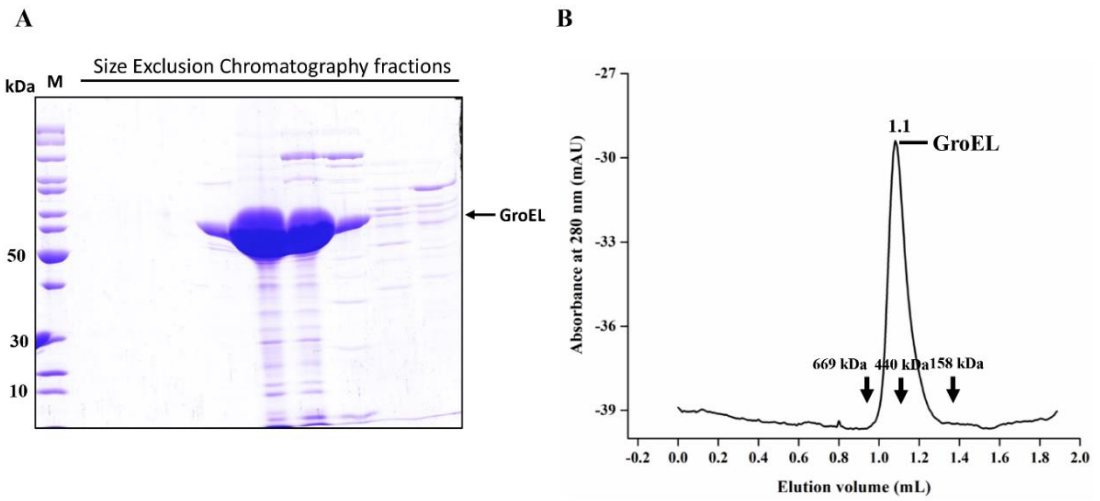


Figure 3.3 **Purification of E.coli chaperonin GroEL:** (A) 12.5% SDS-PAGE showing the final step purification step of *E.coli* chaperonin GroEL using Size exclusion chromatography (SEC) column SuperDex 200 16/600. (B) Analytical size exclusion chromatogram of GroEL using SuperDex 200 16/600 column showing a single peak confirms homogeneity in the purified protein.

3.2.2 Purification of *E.coli* co-chaperonin GroES

E.coli co-chaperonin GroES was purified using Immobilized metal affinity chromatography and Anion Exchange chromatography Source 15Q column (Figure 3.4). SDS-PAGE analysis of the final fractions confirmed the purity of the protein and it was concentrated to an average of 40 mg/mL.

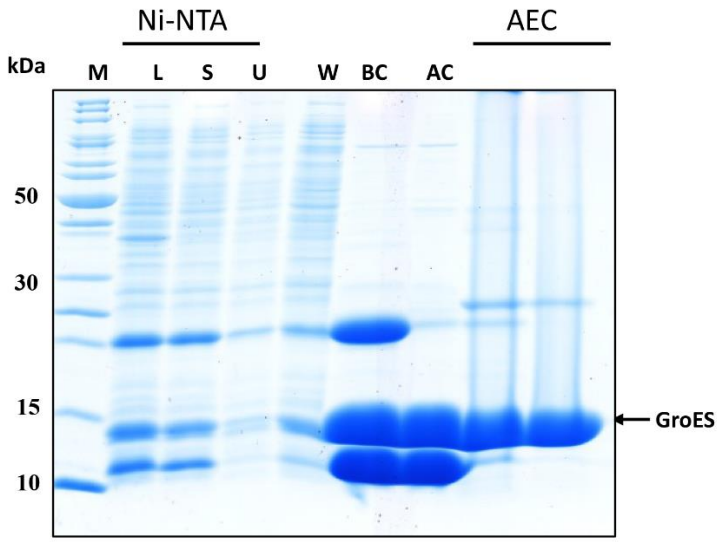


Figure 3.4 **Purification of *E.coli* co-chaperonin GroES**: 12.5% SDS-PAGE showing the two-step purification of *E.coli* chaperonin GroES using IMAC and AEC. M-Molecular weight marker, L-Lysate, S- Soluble, U-unbound, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC-After cleavage using ubiquitin specific protease (USP2).

3.3 Developing the *In vitro* assay to study Rubisco folding and assembly in real time

In our study design, Rubiscos from different organisms were heterologously expressed in *E.coli*, followed by purification using chromatographic techniques to obtain soluble proteins. Purified proteins were first tested *in vitro* for activity using the coupled spectrophotometric Rubisco assay (Kubien *et al.*, 2011). The functional Rubiscos were then denatured by chemical method (Guanidine hydrochloride) and later supplied stepwise with additional factors and chaperones to evaluate the minimum requirement for folding and assembly. The folded and assembled proteins were analysed electrophoretically and assayed for biochemical activity (Figure 3.5).

We applied this developed study design to the classic dimeric form II *Rhodospirillum rubrum* and after validation, explored the biogenesis of the newly described hexameric form II *Acidithiobacillus ferrooxidans* Rubisco.

Analysis of purified proteins using Native-PAGE and SDS-PAGE documented the homogeneity and purity of proteins respectively. The *Rhodospirillum rubrum* Rubisco (RrM) and *Acidithiobacillus ferrooxidans* Rubisco (AfM) migrated consistent with their previously confirmed dimeric and hexameric states in the non-denaturing PAGE in

compliance to literature (Tsai *et al.*, 2015). The analytical size-exclusion column indicated RrM and AfM proteins exist in single oligomeric states (Figure 3.6). *E. coli* chaperonins GroEL and GroES used in further experiments were tested for biochemical activity.

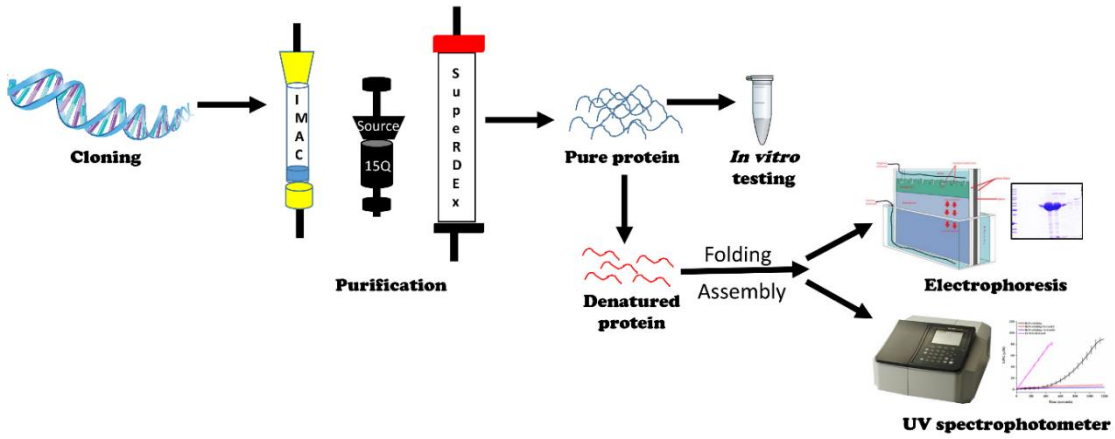


Figure 3.5 **Schematic representation of *in vitro* assay design**: Different cloning techniques as explained in methods section were employed to clone Rubiscos from its genomic DNA. Following successful cloning, the constructs were heterologously expressed in *E.coli* host and tested for protein expression. After confirming the expression, proteins were purified using various chromatographic techniques. Purified proteins were tested *in vitro* for activity and functional proteins were further used for folding and assembly studies. Denaturation of pure proteins was performed as explained earlier (Section 2.4.3.1) and necessary components are supplied stepwise for effective folding & assembly of Rubiscos. Electrophoretic analysis employing native-PAGE and the coupled spectrophotometric Rubisco activity assay experiments were set up to understand the mechanisms *in vitro*. These analyses permit a dissection of Rubisco biogenesis requirement.

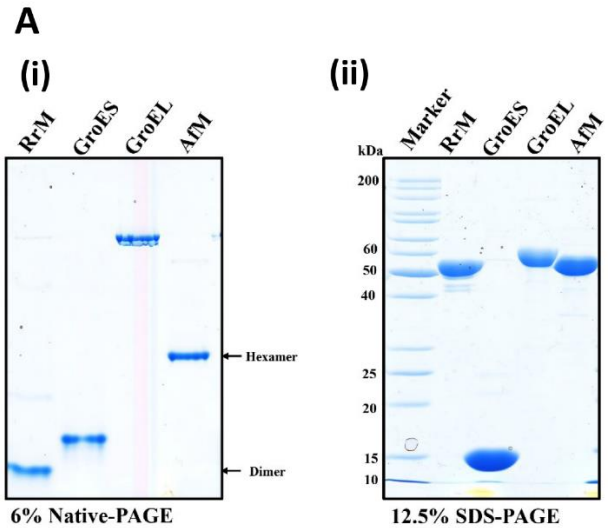


Figure 3.6 **Electrophoretic analysis of purified proteins**: 6% Native-PAGE and 12.5% SDS-PAGE (Coomassie staining) analysis documenting the homogeneity (RrM- dimer, AfM- hexamer, GroEL -

14mer and GroES- heptamer) and purity of proteins (corresponding to respective molecular weights) further used in the *in vitro* protocol, 4 µg of total protein per lane.

Rubiscos were denatured and bound to the chaperonin GroEL for folding. The GroEL bound Rubisco was quantified using densitometry prior to initiation of folding. ATP was added to the refolding buffer to initiate the GroEL-ES reaction cycle (Hartl *et al.*, 2011) and samples were collected to analyse the folding and assembly process (Figure 3.7).

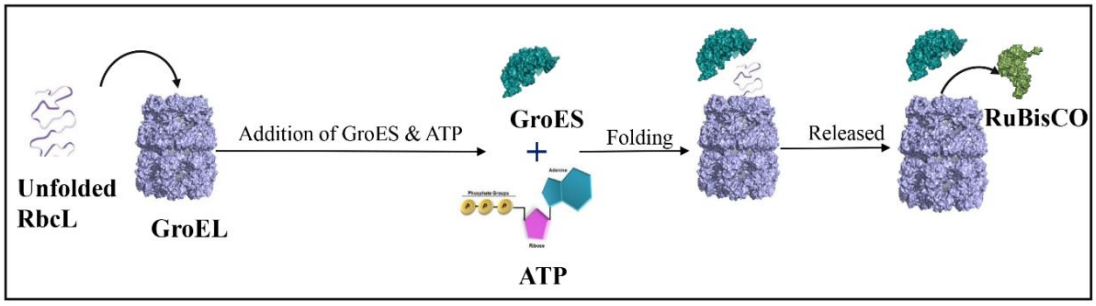
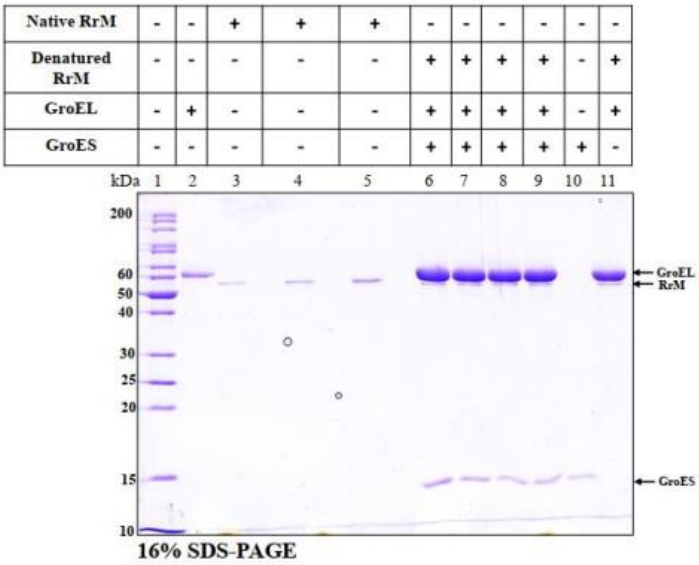
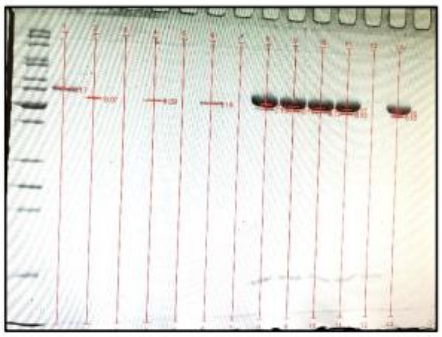


Figure 3.7 **Schematic representation of denaturing and refolding protocol**: Scheme depicting the stepwise procedure involved in the refolding assay. The 0.2 µM denatured Rubisco is bound to the chaperonin GroEL (0.6 µM). The co-chaperonin GroES (1.2 µM) is added to the mix to assist refolding. The refolding of the denatured Rubisco is initiated by the addition of 4 mM ATP. Finally, the refolded Rubisco is released from the GroEL-ES cage and assayed for functionality and oligomeric details.

3.4 SDS-PAGE analysis followed by densitometer measurements to quantify GroEL bound dimeric RrM

Prior to performing refolding experiments, it was critical to determine the concentration of denatured Rubisco bound to chaperonin GroEL. This would then allow a comparison of maturation kinetics. Denaturing and refolding of dimeric RrM was performed as explained earlier. The denatured RrM (0.2 µM) was bound by adding it to rapidly vortexing refolding buffer containing GroEL (0.6 µM); and GroEL bound RrM samples were centrifuged to remove unbound aggregates. 1.2 µM GroES was added to the samples , which were then subjected to SDS-PAGE and subsequent western blotting. In the coomassie stained SDS-PAGE two distinct bands showed the GroEL and bound RrM as separate proteins. These bands were quantified for band intensity using densitometer which showed ~57% of denatured RrM was trapped by the GroEL chaperonin (Figure 3.8). This indicates ~0.1 µM of denatured RrM is available to undergo GroEL-GroES chaperonins assisted folding *in vitro*.





Trace OD X nm	Quantity in ng
0.07596	200 (known)
0.03993	100 (known)
0.03787	50 (known)
0.03846	67.50 (unknown)
0.03839	67.22 (unknown)
0.03813	66.24 (unknown)

Figure 3.8 **GroEL bound denatured RrM quantification**: (A) 16% SDS-PAGE showing the native RrM in lanes 3, 4, 5 with 50 ng, 100 ng and 200 ng loaded respectively. Lanes 6-9 are GroEL bound RrM samples showing two distinct bands corresponding to their molecular weights. Samples labelled 1-4 are the same facilitating duplication of the reactions. The coomassie stained SDS-PAGE image was then used to quantify the amount of RrM present in the individual samples. Densitometer was employed to quantify the band intensity and native RrM was used as control for the experiment. (B) Bands selected for intensity quantification by densitometer. (C) Intensities quantified for the control (known) native RrM and the unknown RrM bound to GroEL. In this experiment ~ 57% of denatured RrM (~ 0.1 µM) was bound to GroEL and was available for folding.

3.5 Coupled spectrophotometric Rubisco assay shows the assembly of RrM into functional dimers

Following the quantification of denatured RrM bound to GroEL, next step was to study the folding and assembly using the spectrophotometric Rubisco assay. RrM Rubisco denatured in 6 M guanidine hydrochloride solution was diluted into 0.6 µM GroEL containing refolding buffer, which was later provided with 1.2 µM GroES and Rubisco activity assay was utilized to follow the folding and assembly of RrM. Rubisco specific coupled spectrophotometric assay was performed. Denaturing and binding of Rubisco was done as explained above. Refolding buffer incubated with GroES/ ancillary proteins

was added to the Rubisco reaction mix (totally 100 µL) containing 100 mM MOPS/KOH pH 8, 5 mM $MgCl_2$, coupling enzymes, 1 mM NADH, 10 mM phosphocreatine, 40 mM $NaHCO_3$, 1 mM DTT/ATP/RuBP, 0.1 µM Rubisco and assayed continuously for biochemical activity. In the assay it was seen that until ten minutes there was a lag in activity indicating RrM is in process of folding and assembly. At ~10 minutes Rubisco activity was seen, showing the RrM is successfully assembled to functional dimers. This assay shows the assembly to functional dimers require both chaperonins GroEL and GroES. Native RrM served as a positive control and reactions in the absence of either GroEL or GroES were used as negative control (Figure 3.9). Within 10 minutes of refolding, high yield (~89%) of native Rubisco activity was attained.

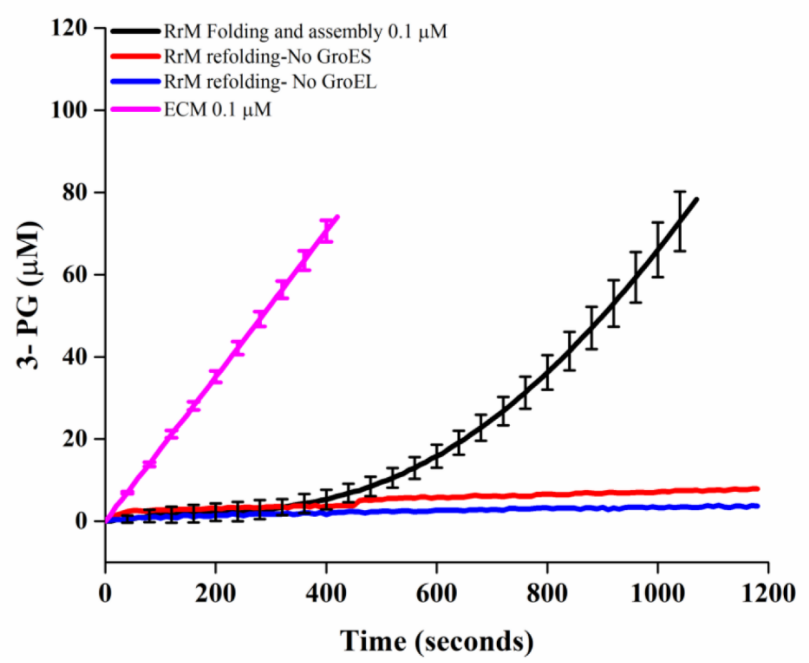


Figure 3.9 **Following Rubisco assembly in real time using the spectrophotometric Rubisco assay:** Stepwise folding and assembly monitored by Rubisco assay. 0.1 µM of denatured Rubisco is folding and assembly to functional dimers. Negative controls in the absence of either chaperonin shows the folding and assembly requires both chaperonins GroEL- GroES. It takes about ten minutes to assemble into functional dimers. Refolding yield in comparison to the native enzyme is ~90%. Error bars indicate the mean and standard deviation of at least three independent experiments. ECM stands for activated Rubisco complex.

3.6 Native-PAGE analysis shows that *in vitro* folding and assembly of dimeric RrM requires chaperonins GroEL and GroES

A similar set-up of denaturing and binding to GroEL as explained earlier was performed in the absence of Rubisco assay components and analyzed using Native-PAGE where the protein will retain its higher order structure due to non-denaturing conditions and immunoblotting of the Native-PAGE gel was performed with anti-RbcL antibodies. After the addition of GroES, 4 mM ATP was added to start the GroEL-GroES reaction cycle. The end-point samples analysed using Native-PAGE (coomassie staining and Immunoblotting against Rubisco antibody) showed that before addition of ATP, the unfolded RrM Rubisco is completely bound to GroEL and migrates as high molecular weight complex. With advancement in time, the GroEL bound Rubisco undergoes folding and is subsequently delivered from GroEL-GroES cage. At end-point sample (after 30 minutes), Rubisco is no longer bound to GroEL and released completely to dimers (Figure 3.10). Rubisco activity assay was utilized to evaluate the functionality of these formed dimers.

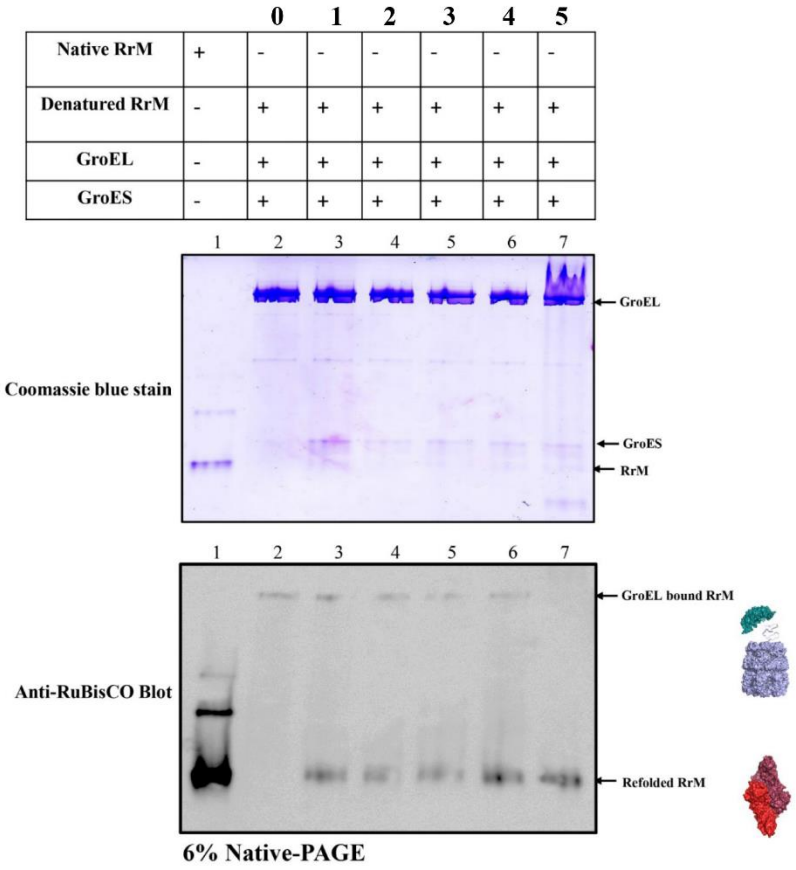


Figure 3.10 **Native-PAGE analysis of folding and assembly of dimeric RrM:** Denaturing of Rubisco and binding it to folding chaperonin GroEL (0.4 µM) was performed as explained previously. The samples were analysed using Native-PAGE (Coomassie staining and immunoblotting against Rubisco antibody).

Lane 1 shows the native RrM Rubisco used as a positive control. Upon addition of 4 mM ATP, the process of folding and assembly was initiated (Lane 3). Samples were collected before the addition ATP (Lane 2) and after 30 minutes (Lanes 3-7). Duplicates from the same reaction were collected at different time points. It was observed that before addition of ATP, there is no folding and the Rubisco is bound to GroEL (Lane 2). After ATP addition, folding was initiated, and assembled Rubisco was released from GroEL-ES cage. At 30 minutes (Lane 7), Rubisco was no longer bound to GroEL and was folded and assembled to dimers.

3.7 Spectrophotometric Rubisco assay shows hexameric form II AfM Rubisco matures more slowly than dimeric form II RrM Rubisco

After successfully establishing the *in vitro* folding and assembly assay and employing it to study the dimeric form II Rubisco, we attempted to investigate the recently discovered hexameric form II Rubisco from the proteobacteria *Acidithiobacillus ferrooxidans* (AfM) for assembly requirements. Following our results from dimeric RrM indicating that presence of GroEL, GroES and ATP was sufficient for folding and assembly, we set-up similar experimentations for AfM. Coupled spectrophotometer Rubisco activity assay showed hexameric AfM also has an initial lag and starts becoming functional but matures slower than dimeric RrM achieving ~40% of reconstitution (Figure 3.11). The requirement to assemble as hexamers was responsible for the slower maturation kinetics observed compared to its dimeric counterpart.

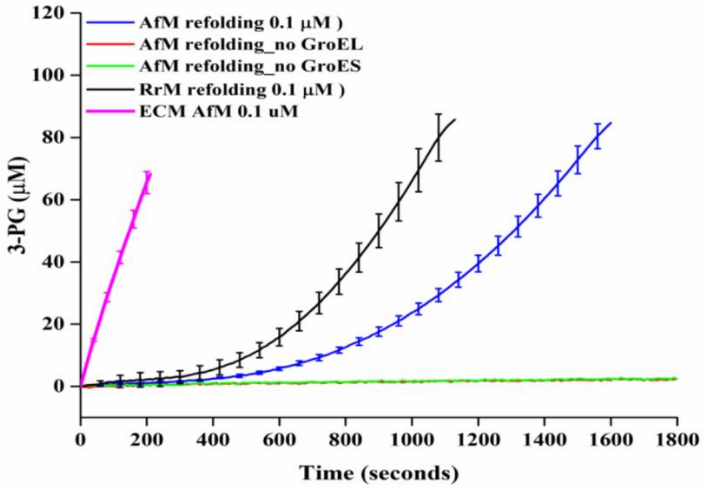


Figure 3.11 **Hexameric form II Rubisco matures slower than dimeric form II RrM:** 0.1 µM of denatured Rubisco is folding and assembly to functional hexamers. Spectrophotometric Rubisco assay shows the folding and assembly of hexameric form II Rubisco takes about 14 minutes and refolding yield is 40%. In comparison, the hexameric form II Rubisco has slower maturation rate attributed to its requirement of higher oligomerisation for functionality. Negative controls in the absence of either chaperonin shows the folding and assembly requires both chaperonins GroEL- GroES. Error bars indicate the mean and standard deviation of at least three independent experiments.

3.8 Assembly of hexameric form II AfM Rubisco proceeds via dimeric intermediates

In the native-PAGE analysis of end-point samples, we observed dimers of AfM formed subsequent to ATP addition and these dimers assembled into hexamers. Figure 3.13 lane 6 shows the hexamers are present in maximum concentration for these set of samples containing GroEL, GroES and ATP in the refolding buffer. For the later samples, hexamers are not seen, indicating that the formed AfM Rubisco is tending to rebind to the GroEL-ES cage. To prevent the formed hexamers to rebind to the GroEL-ES, we supplemented the refolding buffer with substrate RuBP as substrate-mediated assembly of Rubisco dimers has been reported previously for the Form II/III enzyme of *Methanococcoides burtonii* (Alonso *et al.*, 2009) as well as the red-type Form IC enzyme from *Rhodobacter sphaeroides*. In the presence of substrate, the formed hexamers were stable and higher in concentration. There was also a steady increase in the hexamer accumulation (Figure 3.12 lanes 11-15). But in contrast to the dimeric RrM, the GroEL-bound denatured Rubisco did not completely undergo folding, the dimeric intermediates did not entirely assemble to hexamers.

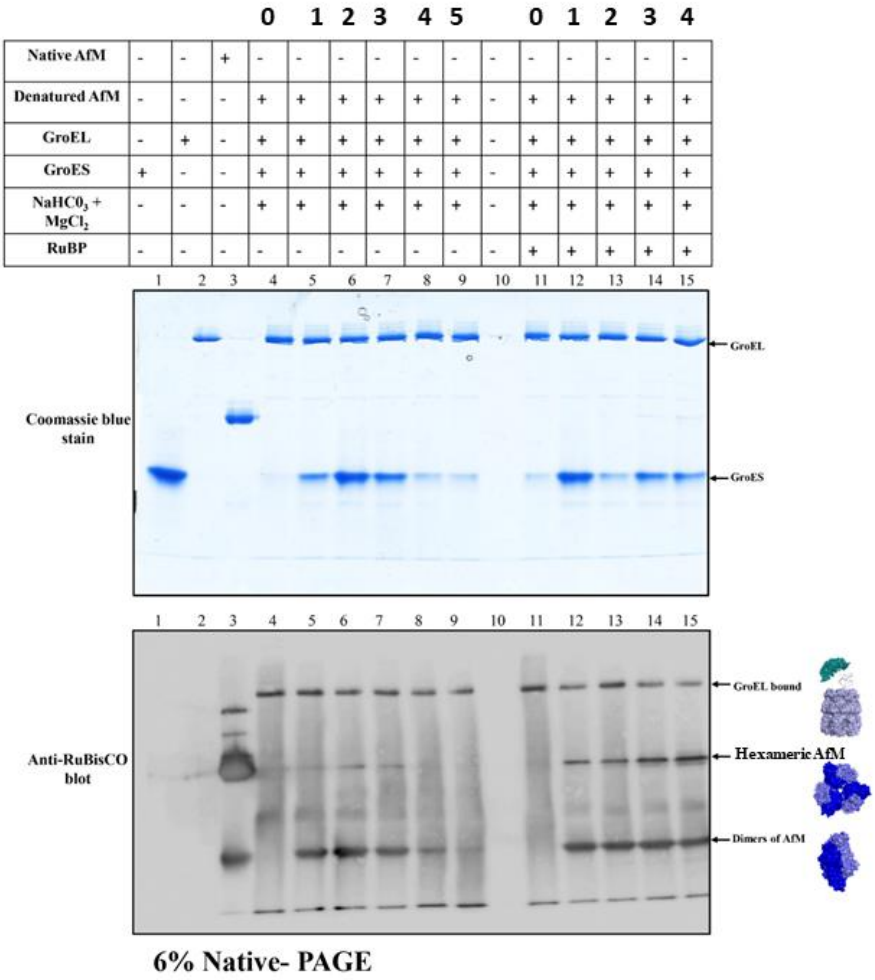


Figure 3.12 **Native-PAGE analysis of hexameric form II AfM shows assembly proceeds via dimeric intermediates:** Native-PAGE (Coomassie staining and immunoblotting against Rubisco antibody) analysis of *in vitro* folding and assembly in hexameric form II Rubisco shows hexameric Rubisco is formed only in the presence of substrate 1 mM RuBP (lanes 12-15) and in absence of which the folded Rubisco tends to rebind to folding chaperonin GroEL (Lanes 5-9). Lane 1 and 2 are native GroES and GroEL chaperonins respectively. Non-denatured AfM Rubisco in Lane 3 serves as a positive control. Lane 4 and Lane 11 are samples before ATP addition. Lanes 5-9 are samples in the absence of substrate RuBP. Multistep assembly proceeds via dimeric intermediates (Lanes 12-15), protein bands corresponding to dimers, folded hexamers and GroEL-bound Rubiscos are indicated.

3.9 Discussion and Conclusions

Owing to Rubisco's rich diversity in occurrence, a complete biogenesis toolbox for all Rubisco forms is not available. Majority of chaperone dependent *in vitro* Rubisco experiments are conducted with classic model client proteins and assembly of recently discovered and minor players of CO_2 fixation is poorly studied. In this work we aimed to bridge this gap and explore the biogenesis of hexameric form II Rubisco. Firstly, we developed a complete *in vitro* protocol using the previously established denaturing and refolding methods and combining it with the coupled spectrophotometric Rubisco assay to study folding and assembly in real-time. With this method, we monitored the entire folding and assembly process without resorting to radioactive assays.

By applying this study design to dimeric form II *Rhodospirillum rubrum* (RrM) Rubisco, we observed the assembly *in vitro* requires only chaperonins GroEL and GroES. A successful recovery of ~90% of native Rubisco activity was achieved. This demonstrated the validity of the method and it can be used to study other complex Rubisco forms.

We studied the interesting chemoautotroph *Acidithiobacillus ferrooxidans* (Af) which possess multiple Rubisco encoding operons including a form II (AfM) and two Form I Rubiscos ($AfLS_c$ & $AfLS_q$). Hexameric form II AfM Rubisco exhibited disparate assembly requirements compared to dimeric RrM. We identified that the presence of substrate RuBP is mandatory for AfM to assemble into stable hexamers *in vitro*, in the absence of which rebinding to chaperonin GroEL occurs. The assembly transition to hexamers proceeds via sequential oligomerization which includes prominent dimeric intermediates. These dimeric intermediates were less in concentration and our attempts to isolate them was unsuccessful. Hence it remains to be unclear if these intermediates are functional, but our hypothesis is that these dimers are not functional. AfM assembly exhibited a refolding yield of ~40% suggesting that AfM monomers are not completely folded by GroEL-GroES and they remain bound to the chaperonin cage. In addition to this, the maturation rate of hexameric AfM is slower compared to RrM. This could be attributed to the higher order oligomeric state required.

4 DETAILED STUDY OF *IN VITRO* FOLDING AND ASSEMBLY IN FORM IA RUBISCO FROM *ACIDITHIOBACILLUS FERROOXIDANS* (AfLS)

In the previous chapter we detailed a simple, rapid *in vitro* protocol to study folding and assembly of Rubisco in real-time and successfully demonstrated the method for two form II Rubiscos from *Rhodospirillum rubrum* (RrM) and *Acidithiobacillus ferrooxidans* (AfM). It was confirmed that chaperonins GroEL-GroES was sufficient for dimeric RrM whereas hexameric AfM form II required substrate RuBP to stabilize the assembly. Following this we are attempting to understand the structurally more complex hexa-decameric form I Rubisco using similar experimental setup.

In this study, we aimed to understand the *in vitro* folding and assembly requirements of Form IA *Acidithiobacillus ferrooxidans* (AfLS). As previously shown this organism contains multiple CO_2 fixation operons encoding Form II, which was already discussed in our previous chapter and two types of form IA Rubiscos (Badger & Bek, 2008). The form IAs include a carboxysomal associated Rubisco which will be referred to as AfLcSc/ form IAc in this thesis and form IAq associated with QO type activases (will be referred as AfLqSq hereafter). It was previously proposed that alpha-carboxysomal Rubisco accumulation factor acRAF acts as an assembly chaperone for form IA Rubiscos (Wheatley *et al.*, 2014). A crystal structure has been solved but thorough mechanistic and *in vitro* experiments are not available. It is critical to understand the minimum requirements for folding and assembly of form I Rubiscos to inform biotechnologically relevant synthetic biology aspects of Rubisco. In all cases when prokaryotic Rubiscos have been produced in higher plant chloroplasts only low yields of properly folded and assembled enzyme have been achieved (Whitney *et al.*, 1999, Cai *et al.*, 2014, Wilson *et al.*, 2016, Long *et al.*, 2018).

This chapter will focus on form IA Rubiscos as *in vitro* folding and assembly of examples of other form I Rubiscos such as form IB, form IC have already been described (Liu *et al.*, 2010, Hauser *et al.*, 2015, Joshi *et al.*, 2015, Aigner & Wilson, 2017). We will detail out experiments performed, difficulties faced, and remedial attempts carried out in building the complete toolbox of proteins necessary for the study.

4.1 Purification of Form IA *Acidithiobacillus ferrooxidans* (AfLS) Rubisco as individual large and small subunits

4.1.1 Purification of large subunit (AfLq) associated with CO2 fixing CbbQO operon of form IAq AfLS

Large subunit AfLq of AfLS was cloned as individual subunit from genomic DNA of *Acidithiobacillus ferrooxidans* into pHue vector and provided by our Post-doc Dr. Yi-Chin-Tsai (Tsai *et al.*, 2015). Protein expression was validated using small scale protein-expression tests as explained in methods section. Two-step purification was carried out for this heterologously expressed in in *E.coli* BL21 as a N-terminal fusion with a hexa-histidine tagged ubiquitin (His6Ub). Following IMAC, the cleaved protein was subjected to Source 15Q AEC column. SDS-PAGE analysis of the fractions showed the protein was pure (Figure 4.1) and the respective fractions were pooled and concentrated for future use. Purified AfLq existed as an octamer when tested using analytical size-exclusion chromatography and native-PAGE. The protein concentration was estimated using a Nanodrop, measuring the absorbance at 280 nm.

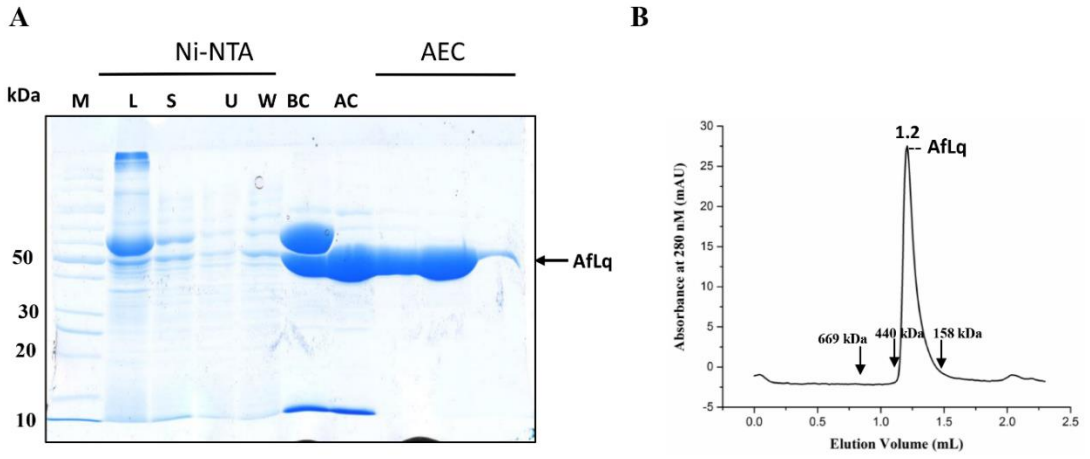


Figure 4.1 **Purification of form IAq Rubisco large subunit AfLq from *Acidithiobacillus ferrooxidans* (AfLS)**: (A) SDS-PAGE showing the successive purification steps of AfLq Rubisco large subunit using immobilized metal affinity chromatography IMAC (labelled as Ni-NTA) and Anion exchange chromatography (AEC) column Source15Q. M-Molecular weight marker, L-Lysate, S- Soluble, U-Unbound, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC- After cleavage using ubiquitin specific protease (USP2).(B) Analytical size exclusion chromatogram of AfLq using SuperDex 200 16/600 column showing a single peak confirms homogeneity in the purified protein.

4.1.2 Purification of small subunit (AfSq) associated form IAq AfLS

The small subunit AfSq cloned into pHue vector (provided by Yi-Chin-Tsai) with a hexa-histidine (His_6) ubiquitin tag. Small scale protein expression test confirmed the expressed AfSq. After passing through IMAC column and cleavage of ubiquitin tag, the protein was applied to Anion exchange column, but it did not bind to the column and collected as flow-through. The flow-through was again passed through Ni-NTA to get get AfSq. Because of the similar molecular weight of ubiquitin (8.6 kDa) and AfSq (13.5 kDa) getting rid of ubiquitin completely was not possible. (Figure 4.2).

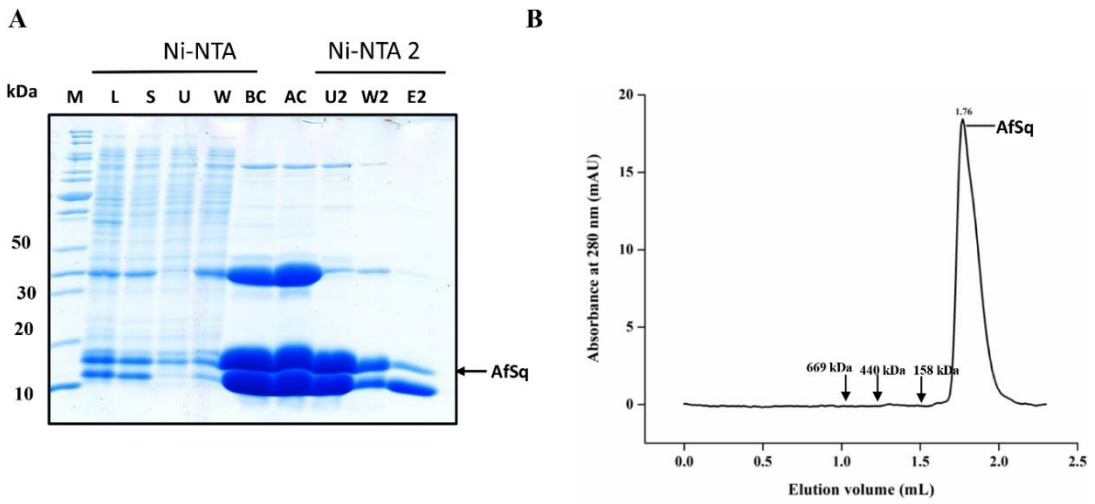


Figure 4.2 **Purification of form IAq Rubisco small subunit AfSq from *Acidithiobacillus ferrooxidans* (AfLS)**: (A) SDS-PAGE showing the successive purification steps of AfSq Rubisco small subunit using immobilized metal affinity chromatography IMAC (labelled as Ni-NTA & Ni-NTA 2). M-Molecular weight marker, L-Lysate, S- Soluble, U-Unbound, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC- After cleavage using ubiquitin specific protease (USP2), U2- Unbound from second Ni-NTA, W2- Wash from second Ni-NTA, E2- Elute from second Ni-NTA.(B) Analytical size exclusion chromatogram of AfSq using SuperDex 200 16/600 column showing a single peak.

4.1.3 Cloning and purification of Alpha- carboxysomal Rubisco assembly chaperone (AfacRAF) from proteobacteria *Acidithiobacillus ferrooxidans*

An assembly chaperone for form IAc Rubisco was recently identified in *Thiomonas intermedia*. Co-expression of a specific pterin-4α-carbinolamine dehydratase (PCD) like protein, occurring downstream to the α-carboxysomal Rubisco gene cluster in *Halothiobacillus neapolitanus* and its Rubisco with GroELS in *E.coli* lead to increased soluble and assembled Rubisco in comparison to absence of this PCD-like protein. This PCD- like protein was termed acRAF (α- carboxysomal Rubisco accumulation factor). These acRAF lack the N-terminal extensions required for PCD activity (Wheatley *et al.*,

2014). Previously Raf 2 was shown to perform assembly functions in maize (Feiz *et al.*, 2012, Feiz *et al.*, 2014). Sequences of this predicted acRAFs of form IA Rubisco carrying organisms were retrieved from NCBI database employing BLAST function and aligned to study its position in the genome and regions of conservation. It was observed that the protein was always found in association with the carboxysomal genes with a maximum conservation in its N-terminal domain. It is believed that the N-terminal domain plays an important role in the protein's functionality.

We analysed the Rubisco operons of *Acidithiobacillus ferrooxidans* for similar protein using bioinformatics tools (Figure 4.3). It was seen that a similar gene with ~40% was found in the carboxysomal gene cluster and we called it AfacRAF. We generated an I-tasser model (Yang *et al.*, 2015) for AfacRAF to compare it structurally with acRAF. According to the bioinformatics tools, AfacRAF was sequentially and structurally like acRAF. Hence, we decided to clone and purify AfacRAF to be used in the *in vitro* assays.

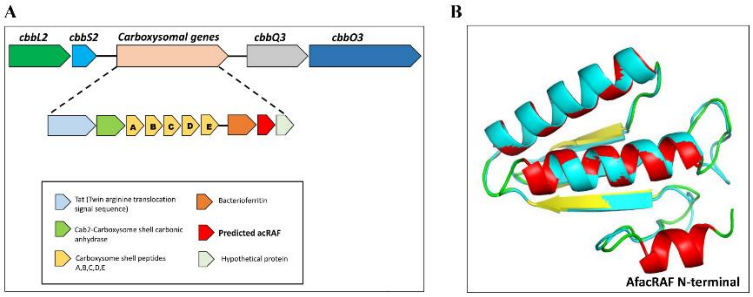


Figure 4.3 **Bioinformatic analysis of AfacRAF**: (A) Assembly chaperone like sequence was found in the carboxysomal Rubisco cluster of Acidithiobacillus ferrooxidans. (B) Structural alignment of assembly chaperone acRAF with I-tasser model of AfacRAF using PyMol software.

pHue vector was used to clone as explained in section 2.2.7.1, followed by protein-expression studies and first step of purification using Ni-NTA column was done. The overnight cleaved protein was applied to the AEC column and similar to AfSq, it did not bind to the column. The flow-through was again passed through equilibrated Ni-NTA column and AfacRAF was eluted. When the eluted AfacRAF was tested using analytical size exclusion chromatography, it showed two peaks. Hence third step of purification using SuperDex 200 16/600 column was done to obtain AfacRAF in purest form (Figure 4.4). Fractions from SuperDex 200 16/600 were froze with glycerol for further use.

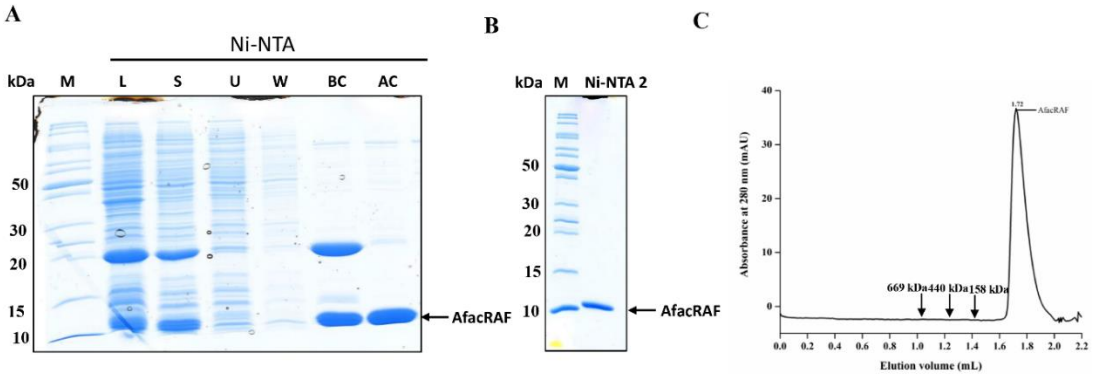


Figure 4.4 **Purification of Alpha- carboxysomal Rubisco assembly chaperone (AfacRAF) from proteobacteria *Acidithiobacillus ferrooxidans* (AfLS):** (A) SDS-PAGE showing the purification steps of AfacRAF using immobilized metal affinity chromatography IMAC (labelled as Ni-NTA). M-Molecular weight marker, L-Lysate, S- Soluble, U-Unbound, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC- After cleavage using ubiquitin specific protease (USP2), U2-Unbound from second Ni-NTA, W2- Wash from second Ni-NTA. (B) Coomassie stained SDS-PAGE showing the second Ni-NTA eluted sample of AfacRAF. (C) Analytical size exclusion chromatogram of AfacRAF using SuperDex 200 16/600 column showing a single peak.

4.1.4 Purification of carboxysomal related large subunit AfLc from form IAc AfLS

After small scale protein expression studies of carboxysomes associated Rubisco large subunit AfLc of *Acidithiobacillus ferrooxidans* cloned in pET30b vector (provided by Yi-Chin-Tsai), two-step purification was carried out. Soluble extracts generated from large scale *E.coli* culture was applied to AEC Source 30Q column and fractions from this column were analysed using SDS-PAGE. Fractions carrying the protein were pooled and dialysed in Buffer A overnight. Size-exclusion chromatography was performed to obtain pure AfLc protein (Figure 4.5).

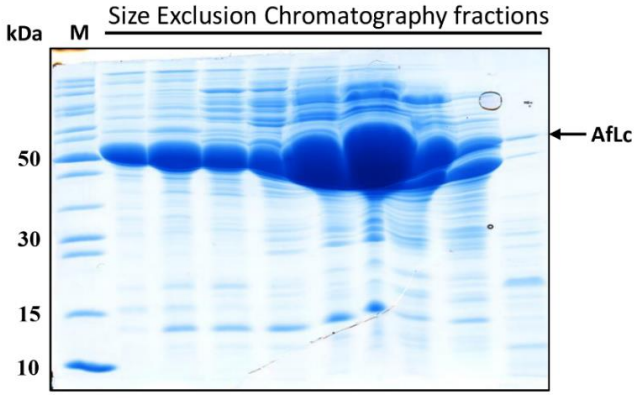


Figure 4.5 **Purification of carboxysomal related large subunit AfLc from form IAc AfLS**: (A) SDS-PAGE showing the final step purification step of large subunit AfLc using Size exclusion chromatography (SEC) column SuperDex 200 16/600.

4.1.5 Unsuccessful attempts to purify small subunit AfSc from form IAc AfLS

To complete the pool of purified individual Rubisco subunits protein for form IA AfLS Rubisco required the carboxysomally associated small subunit AfSc. Small scale protein expression test was performed for small subunits AfSc in vector pET24b. This test showed that AfSc is not expressed (Figure 4.6 A), AfSc was then cloned into pHue vector and protein expression test showed the protein is expressed (Figure 4.6 B). We carried out large scale purification with soluble *E.coli* extract. Following IMAC, the overnight cleaved samples were applied through AEC Source 15Q column. AEC column AfSc fractions were concentrated and applied to SuperDex 200 16/600 column. As shown in Figure 4.7 A, when the peak fractions from size-exclusion column was analysed using SDS-PAGE, there was very negligible amount of protein. Hence, we decided to purify AfSc from inclusion bodies. Purification of AfSc from inclusion bodies was performed as explained in methods section 2.3.7.1. It was observed that during overnight refolding of AfSc, all of the protein precipitated. We tried reducing the time of refolding, but it was unsuccessful to prevent complete precipitation (Figure 4.7 B). Substituting the refolding buffer with ions such as magnesium chloride or factors such as L-Arginine also failed. Gradient denaturing using guanidinium hydrochloride prior to refolding was attempted, but precipitation was unavoidable. Co-expressing AfSc with other carboxysomal cluster protein csoS2 was also not favourable (Chaijarasphong *et al.*, 2016).

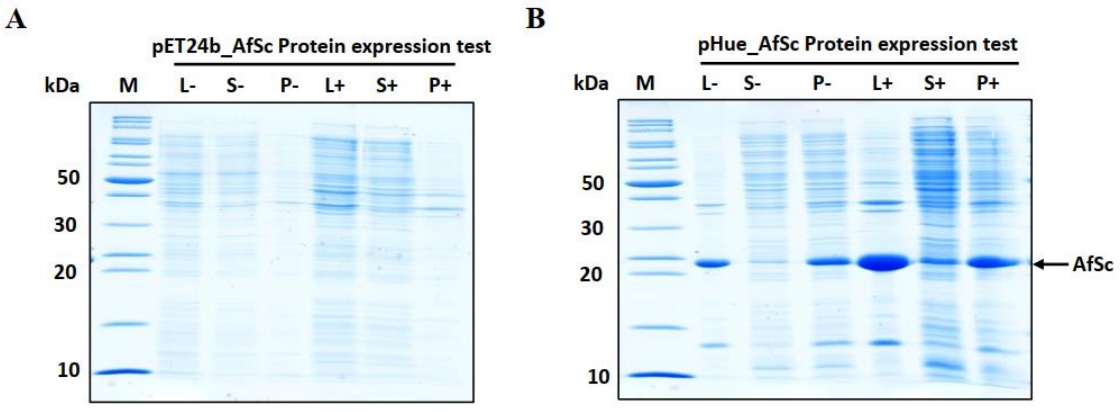


Figure 4.6 **Protein expression test of carboxysomal related small subunit AfSc from form IAc AfLS**: (A) Protein expression test using pET24b vector was unsuccessful. (B) SDS-PAGE showing the protein expression test of small subunit AfSc, it is seen that the protein is found in both soluble and insoluble fraction. In both (A) and (B) L-, S-, P- are Lysate, Soluble and Pellet in the absence of IPTG and L+, S+, P+ are Lysate, Soluble and Pellet in the presence of 0.5 mM IPTG.

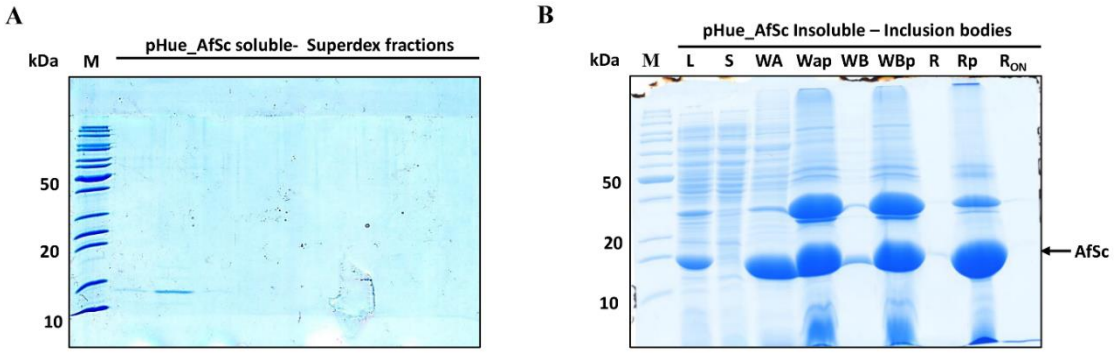


Figure 4.7 **Purification of carboxysomal related small subunit AfSc from form IAc AfLS**: (A) AfSc was attempted to be purified from soluble fraction. SDS-PAGE analysis of fractions from size exclusion chromatography SuperDex 200 16/600 column. (B) Inclusion body purification of AfSc shows that the protein precipitates. L-Lysate, S-Soluble, Wash- first wash, Wap- pellet from first wash, WB- second wash, WBp- pellet from second wash, R-Refolded AfSc, Rp- pellet from refolded AfSc, R_{ON}- refolded overnight.

As AfSc was not able to refold in the refolding buffer and precipitated, we thought that AfSc might need some accessory proteins to aid in refolding. Feiz et al hypothesized that Raf2 may play such a function (Feiz *et al.*, 2014). AfacRAF might be helping AfSc to refold *in vivo.* Hence, we tried to add AfacRAF to the refolding buffer. Binding the denatured 0.1 µM AfSc to refolding buffer containing 0.6 µM AfacRAF was performed in the same manner as described for binding the denatured RbcL to the GroEL chaperonin in section 2.4.3.1.When these samples were run in native-PAGE, AfSc was found in the pellet and AfacRAF alone was not helpful (Figure 4.8 A). Figure 4.8 B shows the addition of 0.6 µM large subunit AfLc to the refolding buffer was also unsuccessful. Hence purification of AfSc from both soluble fractions and inclusion bodies was not possible.

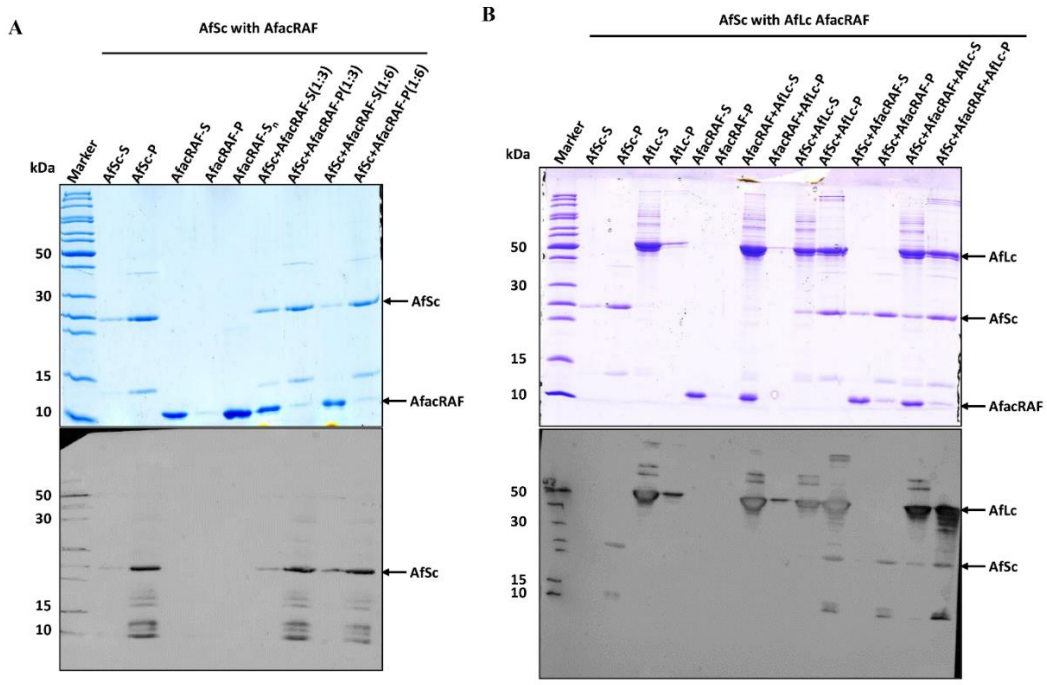


Figure 4.8 **Attempts to rescue precipitation of carboxysome related small subunit AfSc from form IAc AfLS using AfacRAF and large subunit AfLc:** (A) AfacRAF (0.3 µM, 0.6 µM) was supplemented in the refolding buffer to aid 0.1 µM AfSc, it was seen that even in the presence of AfacRAF in two different ratios (1:3, 1:6) AfSc pelleted (lanes 7-8). In labelling, 1:3, 1:6 represents ratios AfSc:AfacRAF. (B) Supplementing with AfacRAF (0.3 µM, 0.6 µM) and large subunit AfLc (0.6 µM) also did not help to prevent precipitation (lanes 14-15). Anti-Rubisco antibody was used for the western blotting for both Figure 4.8 A and B.

Folding and assembly of form I Rubiscos is more elaborate compared to form II as these have additional small subunits to form the hexa-decameric complex. The purity and size of the proteins was confirmed using SDS-PAGE and native-PAGE (Figure 4.9) prior to using them in the *in vitro* assays.

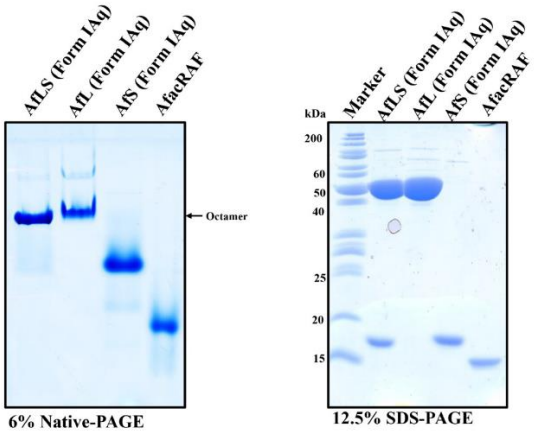


Figure 4.9 **Native-PAGE and SDS-PAGE analysis of purified AfLS Rubisco subunits:** Native-PAGE analysis shows the purified proteins are monodispersed and SDS-PAGE confirms the correct size and purity, 4 µg of total protein per lane.

4.2 *In vitro* reconstitution of purified individual Rubisco subunits shows the proteins are functional

We first performed *in vitro* reconstitution Rubisco assays with individual folded subunits and compared the Rubisco activity with native enzymes. The reconstitution assay was performed as described in the methods section 2.4.3. When AfLq large subunit of form IAq Rubisco (L_8 core) was added to an excess of AfSq small subunit, they showed carboxylation activity (0.7 s^{-1}) confirming their functionality. Individual subunits by themselves were not active which served as negative control for the experiments (Figure 4.10 A). In the same way, we tested the carboxysomal associated AfLc large subunit. Since we failed to purify its small subunit AfSc, the reconstitution assay was performed with AfSq small subunit. The pairwise identity between AfSq and AfSc is 56%. It was observed that AfLc large subunit mixed with AfSq small subunit was functional (carboxylation rate of 1 s^{-1}) and had comparable kinetics to native AfLcSc Rubisco under the experimental conditions (Figure 4.10 B). Following the functional *in vitro* reconstitution experiments, the subunits were mixed together and analysed using native-PAGE to test if they formed higher oligomeric complexes. But we could not distinguish higher bands on the gel (Figure 4.10 A and B insets), wherein a western blot could have resolved the issue. Due to unavailability of specific antibody, western blot was not performed.

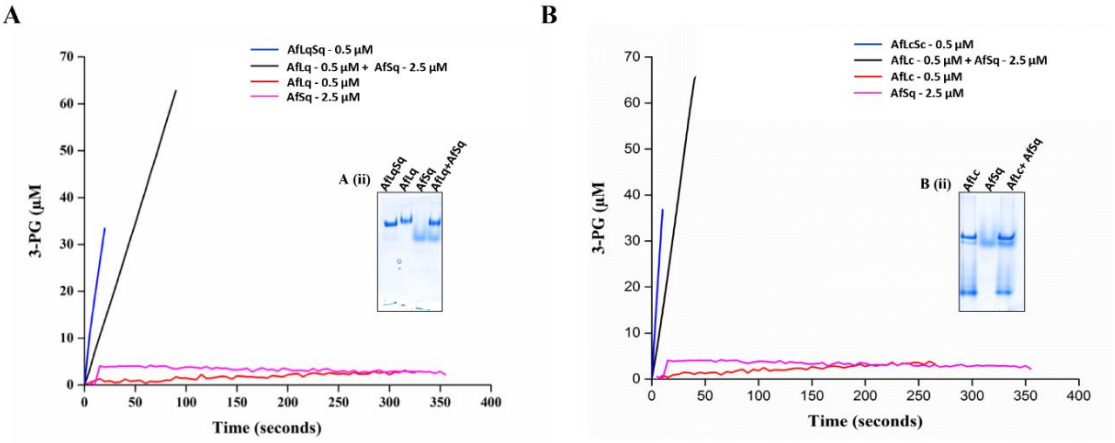


Figure 4.10 ***In vitro* reconstitution of individual subunits of AfLS Rubisco proteins**: (A) Spectrophotometric Rubisco assay shows purified AfLq large subunit mixed with AfSq small subunit was functional. A (ii) shows native-PAGE analysis of the same where higher molecular weight bands were not observed. (B) Carboxysomal large subunit AfLc can be successfully reconstituted to functional enzyme with AfSq small subunit. B (ii) Similar to AfLqSq Rubisco, carboxysomal AfLc also does not show evidence in native-PAGE. In both (A) and (B) the individual subunits by themselves are used as negative control and native AfLS Rubisco serves as positive control.

4.3 *In vitro* folding and assembly of Form IAq AfLS Rubisco

4.3.1 Assembly of form IAq Rubisco large subunit octameric core ($AfLq_8$) can occur in the absence of ancillary proteins

Assembly of structurally more complex form I Rubiscos of plants and cyanobacteria differs from form II in requiring additional chaperones such as RbcX, Raf1, Raf2 and Bsd2 to assemble into hexa-decamers (Brutnell *et al.*, 1999, Onizuka *et al.*, 2004, Saschenbrecker *et al.*, 2007, Liu *et al.*, 2010, Feiz *et al.*, 2014, Wheatley *et al.*, 2014, Hauser *et al.*, 2015, Aigner & Wilson, 2017). Red-type form IC Rubisco from *Rhodobacter sphaeroides* can assemble with the aid of its small subunit alone (Joshi *et al.*, 2015). The mechanism of certain assembly chaperones is well studied whereas Raf2 (denoted as AfacRAF in this thesis) belonging to Form IA Rubisco remains elusive. In this part of our study, we hoped to elucidate the function of AfacRAF in Form IA *Acidithiobacillus ferrooxidans* (AfLS) *in vitro* folding and assembly.

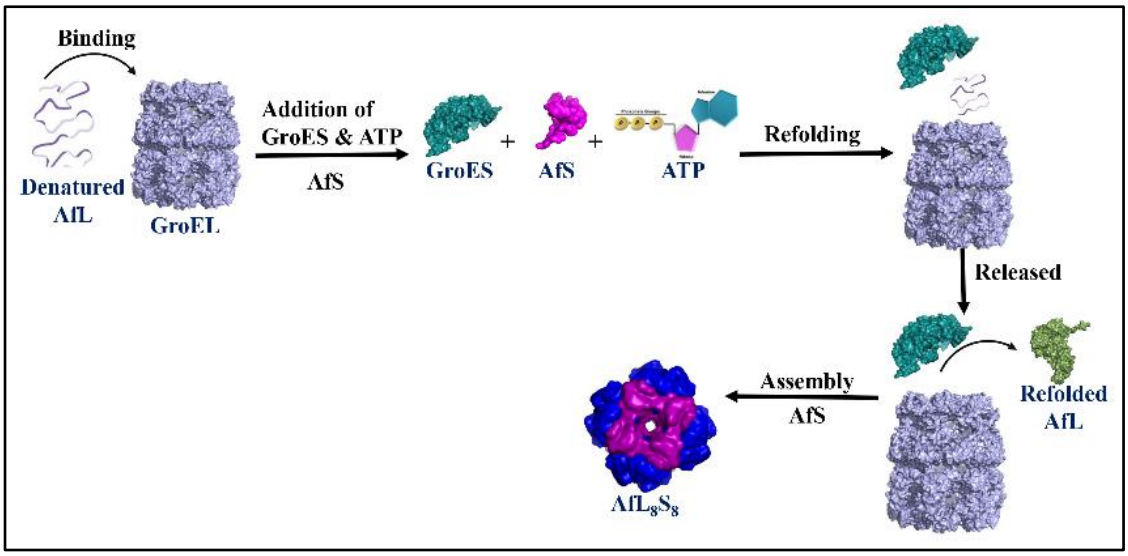


Figure 4.11 **Schematic representation of denaturing and refolding protocol for form I Rubisco**: Scheme depicting the stepwise procedure involved in the refolding assay. The 0.2 µM denatured Rubisco is bound to the chaperonin GroEL (0.6 µM). The co-chaperonin GroES (1.2 µM) and small subunit (1 µM) were added to the mix to assist refolding. The refolding of the denatured Rubisco is initiated by the addition of 4 mM ATP. Finally, the refolded Rubisco is released from the GroEL-ES cage and assayed for functionality and oligomeric state.

We followed the previously established denaturing and binding steps for *in vitro* refolding assay in form II Rubiscos for this form IA Rubisco as well. Electrophoretic analysis using native -PAGE was performed for the samples collected before the addition of ATP and end-point samples after undergoing GroEL-GroES cycling. Multiple end-points samples were collected to have duplicates of reactions. In the non-denaturing PAGE as expected we noticed there was no octameric AfLq before ATP addition (Figure 4.12 lane 5) and accumulation of species consistent to octamers were seen only after ATP was added (Figure 4.12 lanes 6-9). Western blotting of the native-PAGE gel with anti-RbcL antibody confirmed that the octamers formed were AfLq. The interesting outcome of this experiment was that these octamers were formed in absence of any additional proteins. The denatured AfLq folded with the aid of *E.coli* chaperonins GroEL-GroES and self-assembled into octameric cores. We also tested the consequences of AfSq small subunit addition to the *in vitro* folding and assembly setup (Figure 4.12 lanes 11-15). In contrast to the form IB Rubiscos which require RbcX or Raf1 to assemble *in vitro* (Liu *et al.*, 2010, Hauser *et al.*, 2015), form IAq Rubisco was able to assemble in absence of ancillary factors (Figure 4.12).

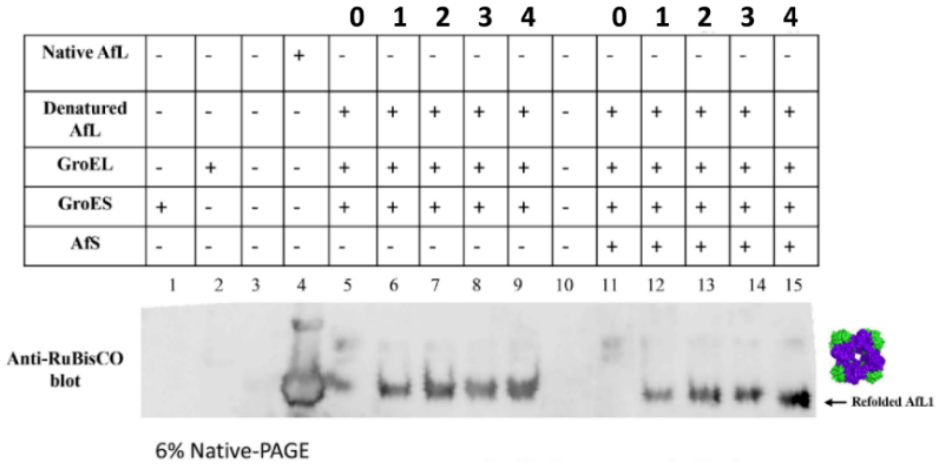


Figure 4.12 ***In vitro*** **Form IAq folding and assembly does not require chaperones:** AfL_q large subunits can assemble to octameric core (L_8) independent of ancillary proteins (labelled as refolded AfL1). Lane 1 and Lane 2 are the chaperonins GroES and GroEL respectively. Lane 4 is native AfLq serving as positive control for the experiment. Lane 5 is the sample before ATP addition to initiate folding. Lanes 6-9 does not contain the small subunit AfSq whereas the lanes 11-15 consists of the small subunit. The addition of small subunit does not majorly affect the assembly process. 0.1 µM denatured AfLq, 0.6 µM GroEL, 1.2 µM GroES, 1µM AfSq and 4 mM ATP were used in the experiment shown here.

4.3.2 Addition of AfacRAF to the folding and assembly of form IAq Rubisco does not affect the assembly process *in vitro*

Wheatley et. al. has proposed that AfacRAF is the assembly chaperone for form IA Rubiscos. But our observations from *in vitro* experiments suggested self-assembly of AfLq octameric core. Hence, we investigated the assembly process by supplementing with 1 µM AfacRAF to comprehend its functional significance to *in vitro* assembly. AfacRAF was added to the refolding buffer along with co-chaperonin GroES. Similar set of samples as the above experiment was collected and analysed using native-PAGE. The addition of AfacRAF to the assembly process did not enhance the octameric $AfLq_8$ formation (Figure 4.13 lanes 11-15). These results confirmed Form IA Rubisco does not require ancillary proteins to assist its assembly and this self-assembly is not inhibited or enhanced by the addition of ancillary proteins (Figure 4.13).

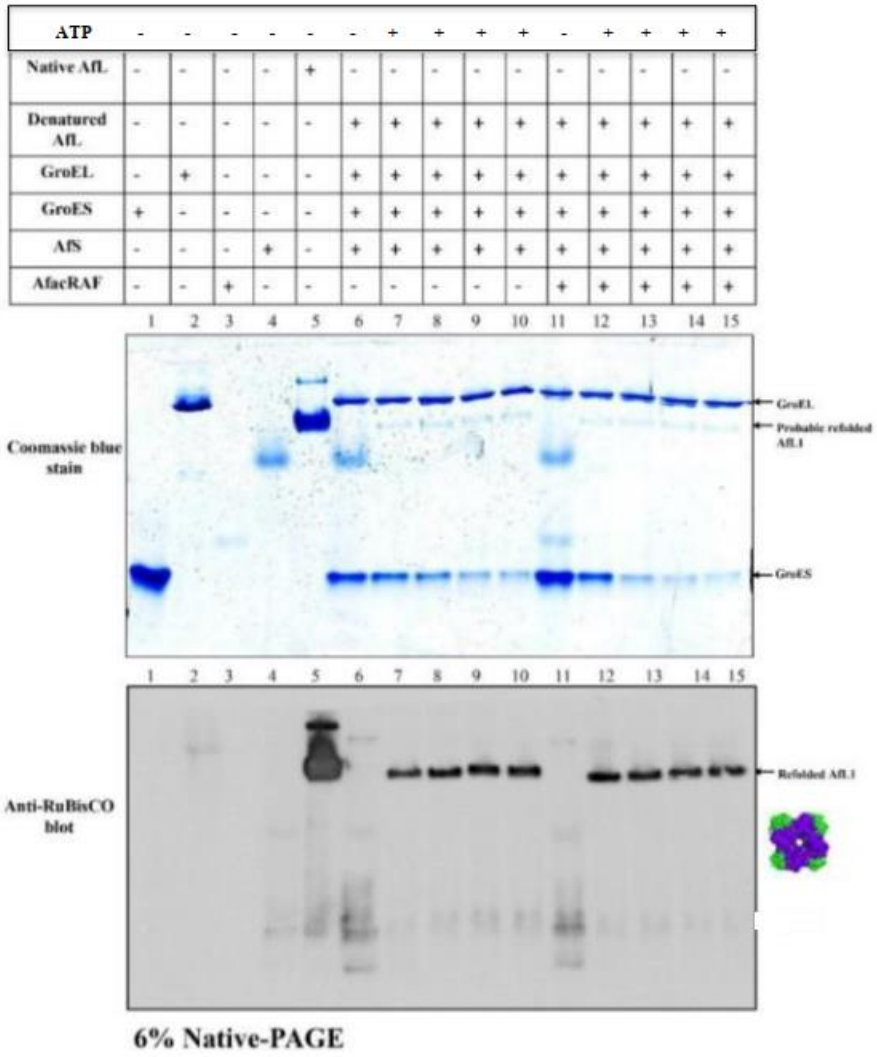


Figure 4.13 ***In vitro*** **Form IAq assembly is not affected by addition of assembly chaperone AfacRAF:** Available literature shows the *in vivo* requirement of assembly chaperones for form IA Rubisco, but in *in vitro* conditions the addition of recombinantly purified assembly chaperone AfacRAF to the system does not change/enhance the assembly. Lane 1 and Lane 2 are the chaperonins GroES and GroEL respectively. Lane 6 and 11 are the sample before ATP addition. Lanes 7-9 and Lanes 12-15 is after sample addition showing the folded AfLq. 0.1 µM denatured AfLq, 0.6 µM GroEL, 1.2 µM GroES, 1µM AfSq, 1 µM AfacRAF and 4 mM ATP were used in the experiment shown here.

4.3.3 Spectrophotometric Rubisco assay shows, despite spontaneous assembly into octameric cores, Form IAq Rubisco needs the small subunit AfSq to become catalytically functional

From the above experiments we concluded that form IAq Rubisco can accumulate higher molecular weight complexes consistent with the octameric L_8 core and western blotting confirmed that these were $AfLq_8$. The logical next step was to test if these octamers are functional. Owing to this we used the coupled spectrophotometric Rubisco assay to analyse the activity. The assay was performed as explained earlier and we found

that denatured AfLq was able to be folded by GroEL/ES and produce functional Rubisco in our spectrophotometric assay if the small subunit was provided (Figure 4.14). In summary form IAq large subunits can self-assemble to octamers core and mere addition of small subunits is enough to make the self-assembled octamers enzymatically functional.

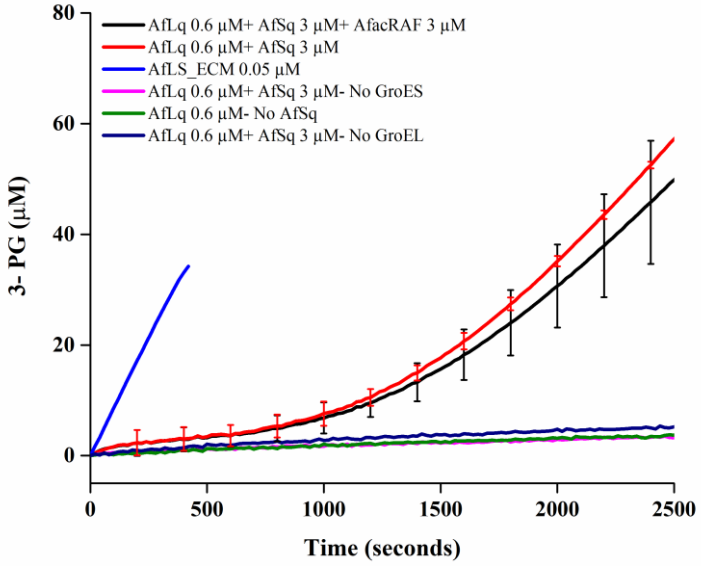


Figure 4.14 **AfLqSq folded by GroEL/ES produced functional Rubisco in spectrophotometric assay if the small subunit was provided**: Spectrophotometry based Rubisco assay shows that small subunit is necessary for functionality though not needed for large subunit octamer (L_8) assembly. It requires at least 20 minutes for functional hexa-decamers to appear.

4.4 Existence of AfLc large subunit dimers - Preliminary analysis of purified AfLc using analytical size exclusion column shows AfLc might exist as a mixture of dimeric and octameric conformations

When we analysed purified large subunit AfLc of carboxysome associated Form IAc Rubisco using analytical size exclusion column chromatography, we noticed that the large subunit AfL_c eluted as two distinct peaks (Figure 4.15 A). These two peaks were collected as separate fractions and analysed using non-denaturing PAGE (Figure 4.15 B). Of the two peaks, the smaller peak protein runs lower in the native-PAGE which was of comparable size to dimeric Rubisco large subunit. When this fraction was run along with a form II dimeric RrM Rubisco, both run at comparable sizes. This was the first time that form I large subunits were isolated as smaller oligomeric units and we further investigated these dimers.

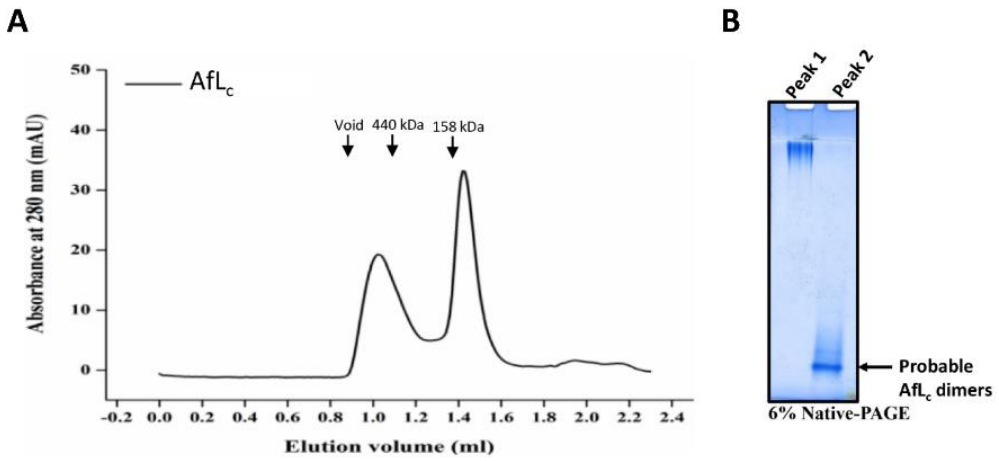


Figure 4.15 **Form IAc Rubisco large subunits AfLc can be isolated as dimers:** (A) Analytical gel-filtration of purified Form IAc Rubisco large subunit (AfL_c) shows the protein elutes as two distinct peaks. (B)Two peaks from the analytical column was applied to Native-PAGE and peak 2 runs consistent to a dimer size.

Following the analysis with native-PAGE, we compared the analytical size-exclusion chromatograms of AfLq (Figure 4.16 A) and AfLc (Figure 4.16 B) large subunits. It is noted that AfLq and AfLc has sequence pairwise % identity of ~85%. Despite the similarity, the overlay of the chromatograms did not show coinciding peaks. Instead we noticed AfLq gives a single peak at 1.18 mL corresponding to a molecular weight of approximately 440 kDa (octamer whereas AfLc shows a peak at 1.42 mL, corresponding to a molecular weight of ~ 120 kDa (Figure 4.16 B and C) which is much smaller in size. This analysis assured us the eluted second peak can be a dimer.

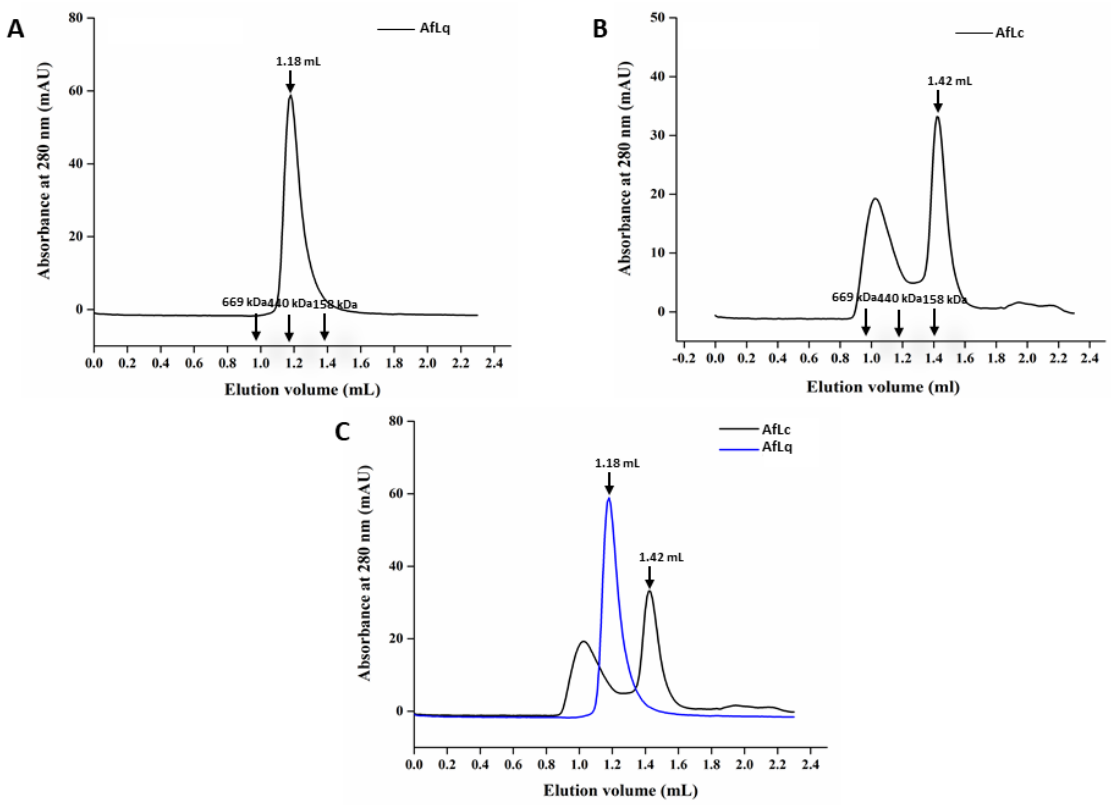


Figure 4.16 **Comparison of AfLc dimers and AfLq analytical size-exclusion chromatograms:** (A) Analytical size-exclusion chromatogram of AfLq showing a single monodispersed octamer population with a peak at 1.18 mL (~440 kDa molecular weight). (B) AfLc elutes as two peaks and the smaller peak at 1.42 mL (~120 kDa molecular weight). (C) Overlay of AfLq and AfLc chromatograms does not show coinciding peaks. 50 µL of samples AfLq and AfLc were injected into Superdex 200 increase 3.2/30 GL analytical gel filtration column using an ÄKTAmicro chromatography system.

4.5 Carbamylation drives oligomerisation of AfLc dimers to octamers

4.5.1 AfLc dimers mixed with AfSq small subunits and assembly chaperone AfacRAF does not yield octamers

To investigate if the AfLc dimers can be assembled into octamers, we mixed the dimers with AfSq small subunits and analysed using native-PAGE for higher bands on the gel. Since the assembly chaperone AfacRAF locus occurs in the carboxysomal Rubisco gene cluster, we tested if AfacRAF can drive the octamer assembly. The AfLc dimers were collected from analytical size exclusion column and incubated with AfS1, AfacRAF or both together to promote interaction and assembly. After brief incubation, the samples were subjected to native-PAGE (Figure 4.17). To our surprise neither of the protein was able to effectively assemble the dimers to octamers or other intermediates. This shows there was some other unknown mechanism that helps the octamer core assembly.

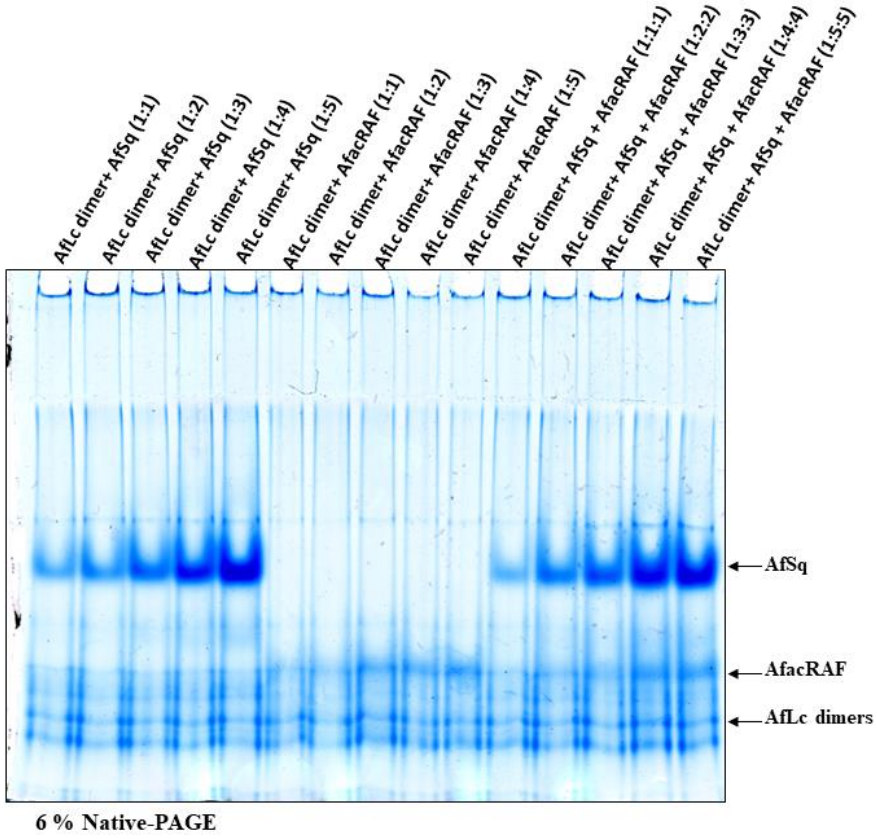


Figure 4.17 **Chaperone AfacRAF and AfSq small subunit did not promote assembly of AfLc into octamers:** Isolated AfLc dimers (5 µM) were incubated with small subunit AfSq (5 µM, 10 µM, 15 µM, 20 µM, 25 µM respectively to different ratios indicated) and AfacRAF (5 µM, 10 µM, 15 µM, 20 µM, 25 µM respectively to different ratios indicated) to drive assembly. The proteins were incubated in the ratios indicated in the labelling. Lanes 1-5 have 5 µM AfLc dimers mixed with AfSq in 1:1, 1:2, 1:3, 1:4 and 1:5 ratios. Lanes 6- 10 have 5 µM AfLc dimers mixed with AfacRAF in 1:1, 1:2, 1:3, 1:4 and 1:5 ratios. Lanes 11-15 contain have 5 µM AfLc dimers mixed with AfSq and AfacRAF in 1:1:1, 1:2:2, 1:3:3, 1:4:4 and 1:5:5 ratios. In all the ratios, there was no higher bands observed.

4.5.2 AfLc dimers form higher molecular weight oligomers (probable octamers) in the presence of sodium bicarbonate and magnesium chloride

According to previous studies, substrate can induce assembly of Rubisco large subunits L_2 to L_{10} in Form II/III *Methanococcoides burtonii* (Alonso *et al.*, 2009). Hence with the aim to promote assembly, we prepared the AfLc samples as carbamylated form (ECM), incubated with substrate RuBP (ER) and stably inhibited carbamylated form with CABP (ECMC). The incubated samples were analysed by native-PAGE. Figure 4.18 lane 3 and 7 shows that incubation with substrate RuBP failed to form the octamers whereas dimers were spontaneously assembling to octamers in the presence of activation factors $NaHCO_3$ and Mg^{2+}. Presence of additional ligand such as CABP did not enhance the assembly yield (Figure 4.18 lane 4 and 8). These results cumulatively

suggest that dimers of form IAc Rubisco can assemble into octameric core *in vitro* by carbamylation of the active site Lys-201 followed by binding of the activating metal Mg^{2+} (Figure 4.18 lane 2). Ligands, AfacRAF and AfSq small subunits appear not to be significant in this case. Additional verification of this effect should involve analysis of Lys-201 amino acid substituted variants which should not be able to assemble in these experiments.

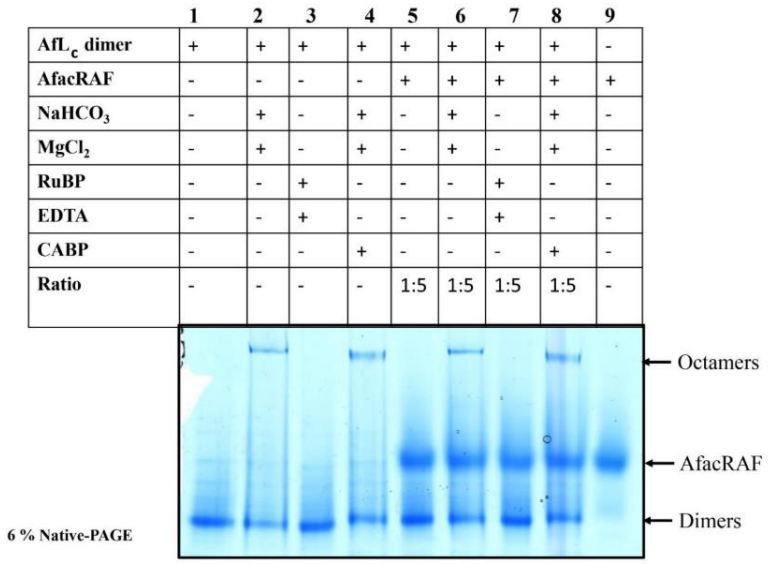


Figure 4.18 **Form IAc Rubisco dimers can be assembled in vitro to octamers:** 5 µM AfLc dimer peaks from the analytical column was separately collected and mixed with 25 µ M AfacRAF, 40 mM $NaHCO_3$, 100 mM $MgCl_2$, 20 mM RuBP, 50 mM EDTA, 4 mM CABP as indicated in the lanes and analysed using Native-PAGE (coomassie staining). The dimers formed higher molecular weight oligomers in the presence of $NaHCO_3$ and $MgCl_2$ (lane 2), whereas it was not found when only RuBP and EDTA were present (lane 3). The higher molecular weight oligomers, probable octamers were not formed when only assembly chaperone was present in the absence of $NaHCO_3$ and $MgCl_2$ (lane 5) but can be seen when the Rubisco activation factors are added (lane 6) Ratio 1:5 in the labelling represents the ratio of AfLc dimers: AfacRAF.

4.5.3 Addition of assembly chaperones from other Rubiscos does not affect the octamer assembly

We examined the effect of adding Rubisco assembly chaperones from different organisms to AfLc dimers to octamers setup to observe for any alterations to the assembly. We tested form IB cyanobacterial assembly chaperones RbcX and Raf1. There was no effect seen when these incubated samples were run in native-PAGE (Figure 4.19). The octamers formed spontaneously in the presence of $NaHCO_3$ and $MgCl_2$ were unaffected by other chaperones thus exhibiting stability.

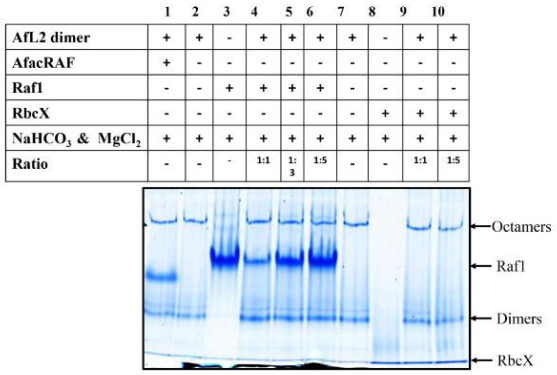


Figure 4.19 **Assembled octamers are stable and not affected by other assembly chaperones:** AfLc dimers assembled to octamers were supplemented with non-native form IB assembly chaperone, the octamer remained stable. The octamers formed were not affected by cyanobacterial assembly chaperones RbcX and Raf1. Ratios 1:1, 1:3, 1:5 represents the ratios of AfLc dimers: Assembly chaperones Raf1/RbcX.

4.5.4 Native-PAGE western blot and Analytical size exclusion chromatography confirms the AfLc dimers to octamers assembly

Preliminary findings from above experiments prompted us to confirm if these higher molecular weight bands were octamers. We resorted to analytical size exclusion chromatography and native-PAGE followed by western blot with Anti-RbcL antibody. Firstly, we isolated the AfLc dimers (Figure 4.20 A) and incubated with sodium bicarbonate and magnesium chloride. The carbamylated AfLc dimers were applied to analytical size exclusion column (Figure 4.20 B). Two peaks were seen at 1.16 mL corresponding to a molecular weight of approximately 440 kDa and 1.42 mL corresponding to a molecular weight of approximately 120 kDa as the earlier higher molecular weight peak shifted to 1.16 mL (molecular weight ~440 kDa). From above experiments it was already known that 1.42 mL (molecular weight ~120 kDa) peak corresponds to AfLc dimers, whereas the peak 1.16 mL (molecular weight ~440 kDa) could belong to octamer. This peak was of comparable size to AfLq large subunits (1.18 mL) showing a peak comparable to molecular weight ~440 kDa showed in Figure 4.20 C.

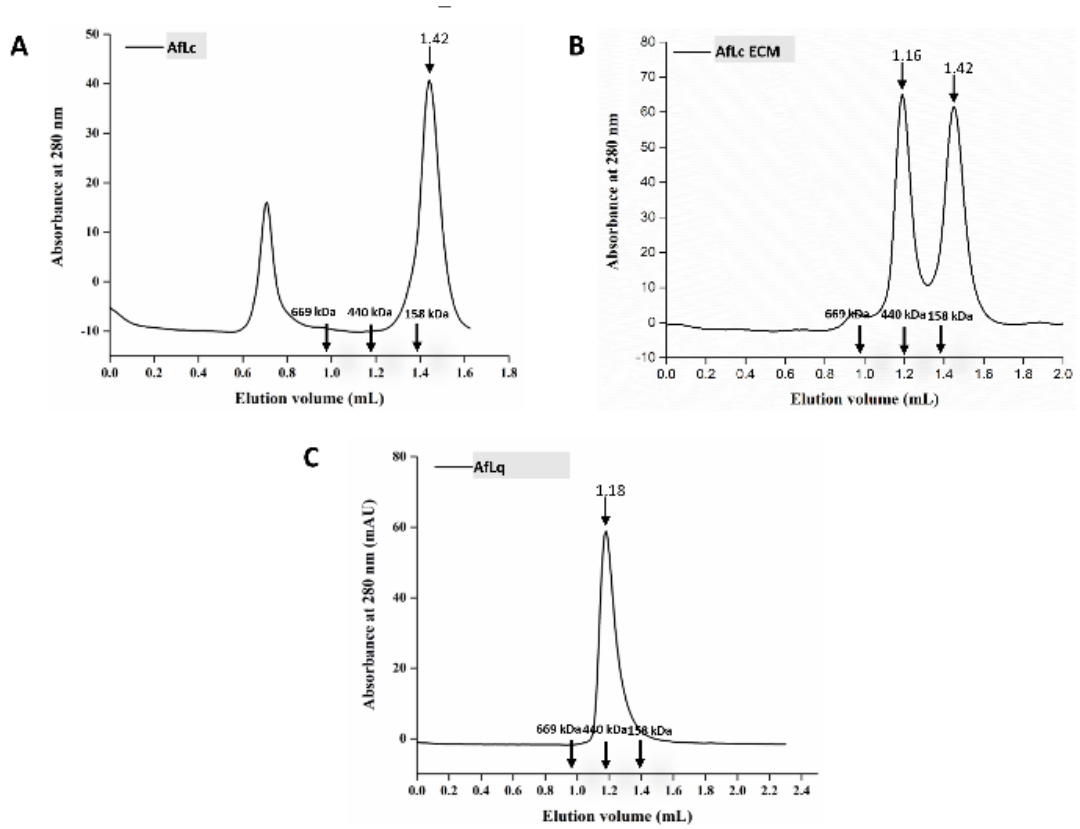


Figure 4.20 **Analytical size exclusion chromatography of carbamylated AfLc dimers**: Carbamylated AfLc dimers elutes out at 1.16 mL (~440 kDa molecular weight) which shows they are octamers. 50 µL of the protein sample was injected to the analytical column.

Native-PAGE followed by western blotting with anti-RbcL antibodies confirmed that the higher bands seen were octamers. On the blot it is seen that the dimers first assemble to tetramers and then finally become octamers (Figure 4.21). Lane 4 (Figure 4.21) consists of the purified AfLcSc hexadecameric rubisco. It was loaded on the lane as a positive control. But it seems the hexadecamer tends to dissociate into dimers. This substantiates the hypothesis made in the thesis that AfLc dimers require activation factors Mg^{2+} and CO_2 for remaining stable. Since Lane 4 did not contain the activation factors Mg^{2+} and CO_2, the hexadecameric AfLcSc dissociated to lower oligomers (dimers).

The combined results presented here demonstrated that AfLc large subunits can be isolated out as dimers and these dimers can assemble to stable octamers when carbamylated.

	1	2	3	4	5	6	7	8	9	10
AfL2 dimer	+	-	-	-	+	+	+	+	+	+
AfS1	-	+	-	-	-	-	-	-	-	+
AfacRAF	-	-	+	-	+	+	+	+	+	-
$NaHCO_3$ & $MgCl_2$	-	-	-	-	+	+	+	+	+	+
AfL2S2	-	-	-	+	-	-	-	-	-	-

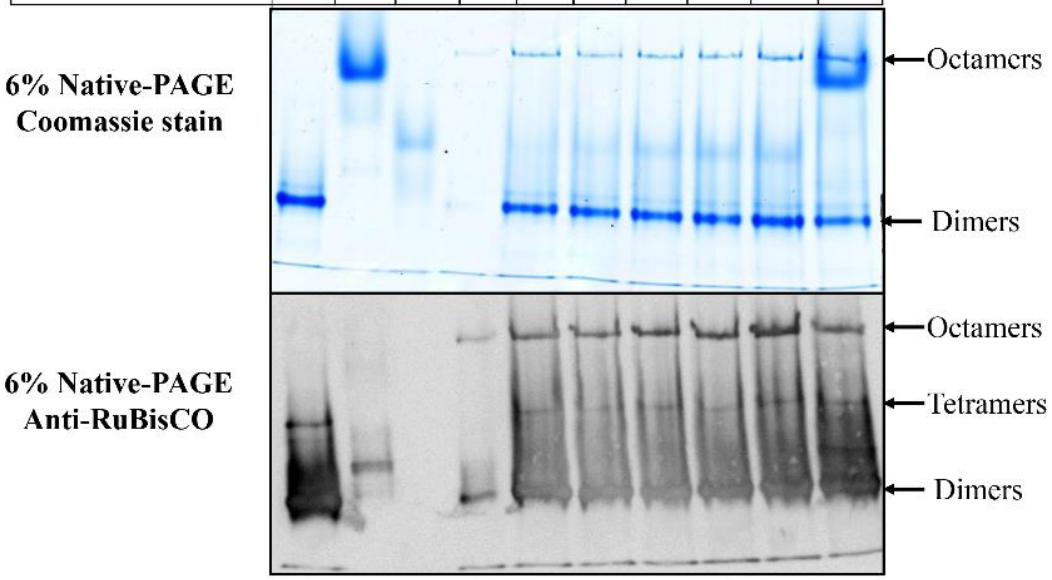


Figure 4.21 **Native-PAGE followed by western blotting confirms the AfLc dimers assembly to octamers:** Carbamylated AfLc dimers assembles to octamers through tetramer intermediates. Lanes 5-10 shows the tetramers and octamers. Anti-RbcL primary antibody was used to perform the western blotting. 5 µM AfL2 dimers, 25 µM AfSq, 25 µM AfacRAF, 40 mM NaHCO3 and 100 mM MgCl2 was added to the respective reactions.

4.6 Discussion and Conclusions

In vitro refolding studies of Form IAq Rubisco showed the purification of Rubisco large subunit AfLq exhibits a stable single octameric state. Assembly of form IAq Rubisco large subunit octamers took place in the absence of ancillary proteins. The addition of assembly chaperone AfacRAF did not influence the assembly process. The assembled octameric core was functional upon addition of small subunit. Earlier studies demonstrated the critical *in vivo* requirement of AfacRAF protein, but its function was not clearly defined. From our experiments it is now clear that AfacRAF is not required *in vitro* and their functionality might be upstream to the chaperonin in biogenesis (that is on the pathway from the ribosome to the chaperonin).

Carboxysomal Form IAc Rubisco from *Acidithiobacillus ferrooxidans* purified as separate large subunits presented us two oligomeric populations. The lower molecular weight oligomers were dimers of large subunit AfL_c and they assembled to octameric core L_8 in the presence of activation factors CO_2 and Mg^{2+}. This interesting chaperone-independent octameric assembly driven by activation factors tempted us to further explore the competency of these dimers. Despite our efforts, we were unable to purify the associated small subunits AfS_c owing to precipitation problems and this restricted us from understanding the further assembly to hexa-decamers. The biochemical characterization of the isolated dimers should also be further investigated.

5 ATTEMPTS TO RE-EXPLORE THE FOLDING AND ASSEMBLY OF FORM IB CYANOBACTERIAL RUBISCO BY *IN VITRO* METHODS

Primary CO2 fixation enzyme Rubisco responsible for biosphere's exclusive inorganic to organic carbon conversion exists in multiple forms such as form I, II, III, II/III, IV requiring extensive auxiliary proteins for correct folding and assembly to functional enzyme (Hauser *et al.*, 2015, Bracher *et al.*, 2017) Kinetic inadequacies, inefficiency in substrate recognition and significant energy investments for ancillary chaperones make Rubisco a valid target for crop improvements. Developing a better screening system for the regularly discovered molecular chaperone requirements for Rubiscos is inevitable. The recent breakthrough of functional expression of plant Rubisco in *E.coli* with the aid of multiple chaperones has been achieved through positive in vitro reconstitution of plant Rubisco is not thus far accomplished, and this could be due to inadequate understanding of Rubisco chaperones mechanistically (Aigner & Wilson, 2017). To reach towards this goal of gaining insights about chaperones, cyanobacterial Rubiscos can be instrumental as this pathway is best explained for them.

Form II Rubiscos requires only the chaperonins GroEL-GroES to fold and assemble into a functional form, and we have shown in the previous chapter that Form IA Rubiscos also have relaxed assembly requirements. In stark contrast, form IB Rubiscos require assembly chaperones functionally downstream to folding to achieve correctly assembled oligomeric states. Multiple Rubisco assembly chaperones were identified and studied of which RbcX and Raf1 are of interest to us for this chapter as they co-exist in the green lineage of form IB *Synechococcus* cyanobacterial Rubiscos.

Both assembly chaperones RbcX and Raf1 aid the large subunit to assemble but mechanistically they bind and interact differently to large subunit dimer surface. RbcX is a homo-dimer that stabilizes the octameric Rubisco large subunits before binding to Rubisco small subunits and facilitate proper assembly (Saschenbrecker *et al.*, 2007). Raf1 is functionally a dimer that binds to the large subunit and embraces one side of it (Hauser *et al.*, 2015)

In this chapter, we plan to re-explore the form IB cyanobacterial Rubisco through *in vitro* experiments and applying our refolding protocol. Firstly, we describe the purification of cyanobacterial Rubisco and other related proteins. Following successful purification, *in vitro* functionality was tested. Attempts to refold cyanobacterial Rubisco was not achieved in the available time-frame.

5.1 Purification of cyanobacterial *Synechococcus elongatus PCC. 6301* (Syn6301 Rubisco proteins

5.1.1 Purification of *Synechococcus elongatus PCC. 6301* (Syn6301) Rubisco Syn6301F345I_LS with mutation F354I in large subunit

The Rubisco from cyanobacteria Syn6301 carrying the mutation F345I in the Rubisco large subunit was cloned as pTrcSynLF345IS (Wunder *et al.*, 2018),), as this mutation was previously shown to enhance the solubility of cyanobacterial Rubisco expressed in *E.coli* (Mueller-Cajar & Whitney, 2008). *E.coli* BL21 cells with over-expressed chaperonins GroEL-GroES was used for this as it was already shown that GroEL-GroES were obligatory for the proper folding of cyanobacterial proteins in heterologous *E.coli* expression systems. Successful production of the recombinant protein was achieved by the addition of arabinose and IPTG to the liquid cultures for induction. Purification protocol employing three steps was followed. Supernatant samples from overnight induced culture were applied to Source30Q anion exchange column, and eluted fractions were analysed by SDS-PAGE. Samples containing the protein was dialyzed overnight in buffer A, prior to applying to MonoQ column. Fractions from MonoQ column was concentrated and finally applied to SuperDex 200 size exclusion column. Representative SDS-PAGE gel after SuperDex 200 column is shown in Figure 5.1, where bands corresponding to Syn6301_L large subunit and Syn6301_S are seen.

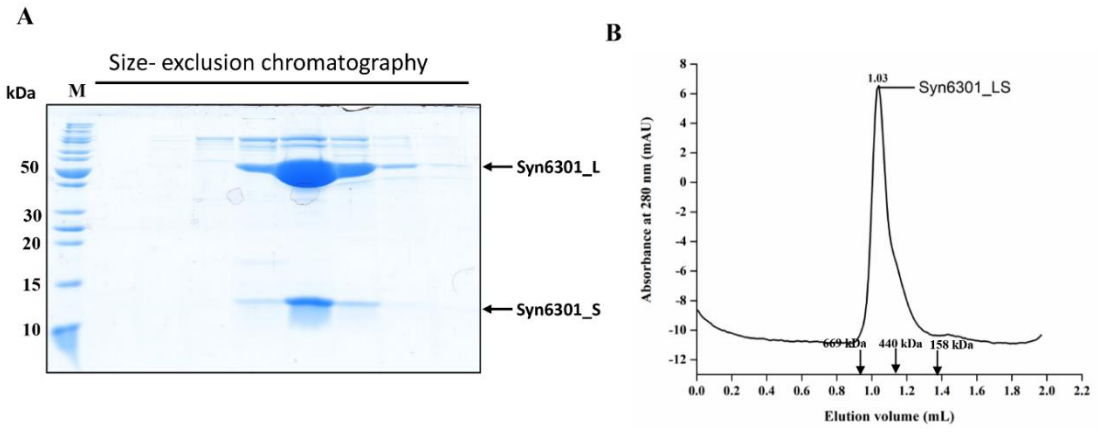


Figure 5.1 **Purification of form IB Rubisco Syn6301_LS from Synechococcus elongatus PCC. 6301 (Syn6301):** (A) SDS-PAGE showing final purification step of Syn6301_LS Rubisco using Size-exclusion chromatography. M-Molecular weight marker. (B) Analytical size exclusion chromatogram of Syn6301_LS using SuperDex 200 16/600 column showing a single peak.

5.1.2 Purification of *Synechococcus elongatus PCC. 6301* (Syn6301) Rubisco as individual large Syn6301F345I_L and small Syn6301_S subunits

Purification of *Synechococcus elongatus PCC. 6301* (Syn6301) individual large Syn6301F345I_L (pTrcSynLF345I- (Wunder *et al.*, 2018)) subunit was similar to Syn6301_LS following same purification steps (Figure 5.2). The final concentration of Syn6301_L was low (2 mg/mL) as quantified by measuring absorbance at 280 nm in comparison to other large subunits purified for this study.

The insert *Syn6301_S* was purified by double digesting the pUC57Syn6301_S plasmid (provided by Zhijun) with NdeI and BamHI using primers in Table 5. It was subsequently restricted and ligated to the plasmid pET11a. This plasmid was designated as pET11a_Syn6301_S, and correct insert sequence was verified by DNA sequencing (First base, Singapore). This small subunit Syn6301_S cloned in pET11a vector was tested for protein expression, and it was seen the protein was mostly present in insoluble pellet fractions. Hence, we attempted to purify Syn6301_S from inclusion bodies, but it was unsuccessful owing to precipitation issues. We then tested the Syn6301_S protein expression using vector pHue and observed fractional Syn6301_S in the soluble fraction. The supernatant from overnight *E.coli* culture was subjected to Affinity chromatography, Anion exchange chromatography, and Size -exclusion columns to obtain the protein (Figure 5.3). Two distinct bands were seen in the size exclusion fractions, ubiquitin is likely to co-elute and even when these fractions were pooled and passed through MonoQ columns, the ubiquitin band was not separated out. Owing to incomplete cleavage of ubiquitin tag and as well the similar molecular weight of ubiquitin and Syn6301_S (13.3 kDa) getting rid of ubiquitin completely was not possible.

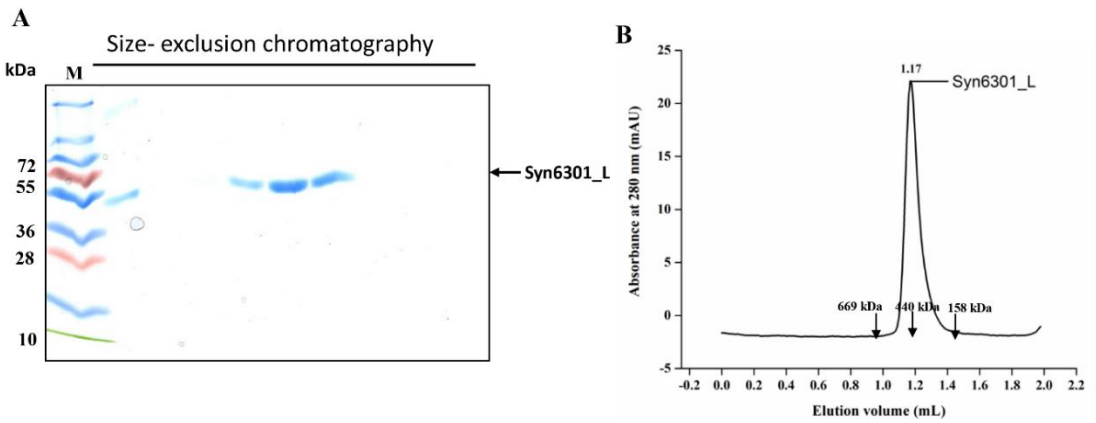


Figure 5.2 **Purification of form IB Rubisco Syn6301_L large subunit from *Synechococcus elongatus PCC. 6301* (Syn6301)**: (A) SDS-PAGE showing final purification step of Syn6301_LS Rubisco using Size-exclusion chromatography. M-Molecular weight marker. (B) Analytical size exclusion chromatogram of Syn6301_LS using SuperDex 200 16/600 column showing a single peak.

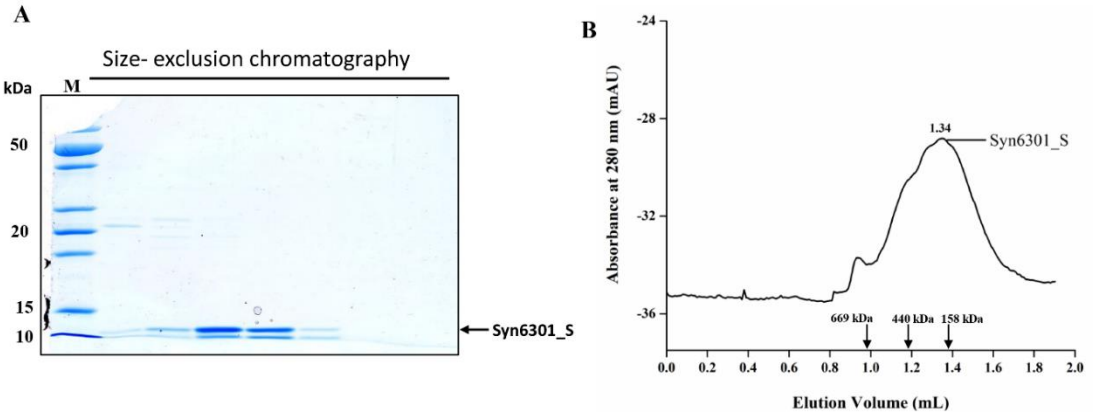


Figure 5.3 **Purification of form IB Rubisco Syn6301_S small subunit from Synechococcus elongatus PCC. 6301 (Syn6301):** (A) SDS-PAGE showing final purification step of Syn6301_LS Rubisco using Size-exclusion chromatography. M-Molecular weight marker. (B) Analytical size exclusion chromatogram of Syn6301_LS using SuperDex 200 16/600 column showing predominantly a single peak.

5.1.3 Assembly chaperones Syn6301_RbcX and Syn6301_Raf1 of *Synechococcus elongatus PCC. 6301* (Syn6301) were expressed and purified

Both the chaperones were cloned in pHue vector, as explained in section 2.2.7.2, 2.2.7.3, and heterologously expressed in *E.coli*. Following test protein expression studies, the supernatant of the overnight culture was firstly applied to Immobilized metal ion chromatography, and fractions were cleaved of His_6 tagged ubiquitin. The cleaved samples were applied to the Source 15Q column, and fractions collected were analyzed by SDS-PAGE. Syn6301_Raf1 assembly chaperone was pure after the second step

(Figure 5.4 B) whereas Syn6301_RbcX required an additional step of Size-exclusion chromatography (Figure 5.4 A).

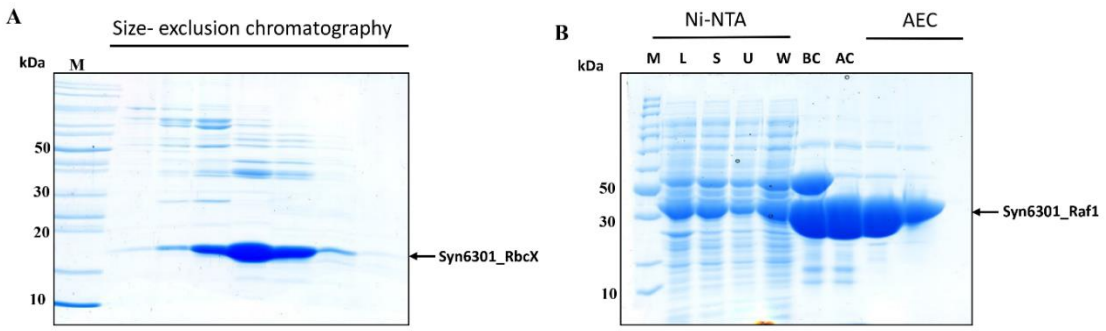


Figure 5.4 **Purification of assembly chaperones Syn6301_RbcX and Syn6301_Raf1 of form IB *Synechococcus elongatus PCC. 6301* (Syn6301)**: (A) SDS-PAGE showing final purification step of assembly chaperone Syn6301_RbcX using Size-exclusion chromatography. M-Molecular weight marker. (B) SDS-PAGE showing the successive purification steps of assembly chaperone Syn6301_Raf1 using immobilized metal affinity chromatography IMAC (labelled as Ni-NTA) and Anion exchange chromatography (AEC) column Source15Q. M-Molecular weight marker, L-Lysate, S- Soluble, U-Unbound, W-Wash, BC-Before Cleavage using ubiquitin specific protease (USP2), AC- After cleavage using ubiquitin specific protease (USP2).

5.2 *In vitro* mixing experiments of *Synechococcus elongatus PCC. 6301* (Syn6301) Rubisco proteins

5.2.1 Spectrophotometric Rubisco reconstitution assay with purified Syn6301 proteins shows formation of functional holoenzyme

The purified individual subunits were mixed and checked for *in vitro* reconstitution. *In vitro* reconstitution assay was performed as explained earlier in the methods section. From the assay, it was seen the carboxylation rate of native Syn6301_LS was low (0.121 s^{-1}), which could be due to small subunits falling off from the holo-enzyme. It was observed that Syn6301_L and syn6301_S mixed together exhibited carboxylation validating the successful formation of the active enzyme. Negative controls of individual subunits by themselves did not show activity (Figure 5.5).

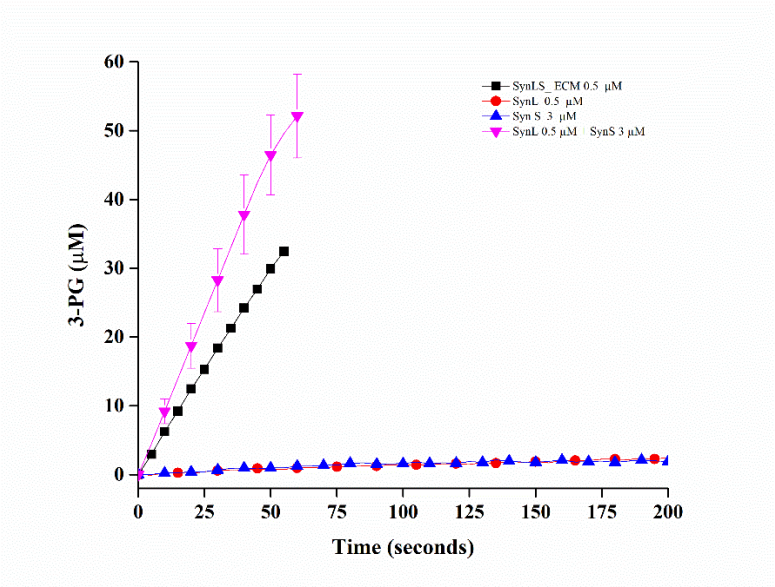


Figure 5.5 ***In vitro*** **reconstitution of individual subunits of Syn6301 Rubisco proteins**: (A) Spectrophotometric Rubisco assay shows purified Syn6301_L large subunit mixed with Syn6301_S small subunit was functional. Individual subunits by themselves are used as negative control and native Syn6301_LS Rubisco serves as positive control.

5.2.2 Native-PAGE analysis of Syn6301 *Synechococcus elongatus PCC. 6301* (Syn6301) proteins

Following the indications from the above experiment, native-PAGE analysis of the samples was performed. As seen in Figure 5.6 native-PAGE analysis of individual proteins and mixed purified proteins show Rubisco Syn6301_LS runs as a hexa-decamer with Syn6301_L large subunit octamer running slightly lower to Syn6301_LS. Small subunit Syn6301_S (octamer) and assembly chaperones Syn6301_RbcX (dimer) and Syn6301 Raf1 (dimer) also run in their respective oligomeric states. When Syn6301_L large and Syn6301_S small subunits were mixed, incubated with sodium bicarbonate and magnesium chloride prior to application to native-PAGE, it was seen that they did not run as a higher band corresponding to Syn6301_LS. This shows the large and small subunits by itself cannot be reconstituted and might require accessory proteins. But higher bands were not seen even after supplementing with assembly chaperones Syn6301_RbcX or Syn6301_Raf1 (Figure 5.6 right panels).

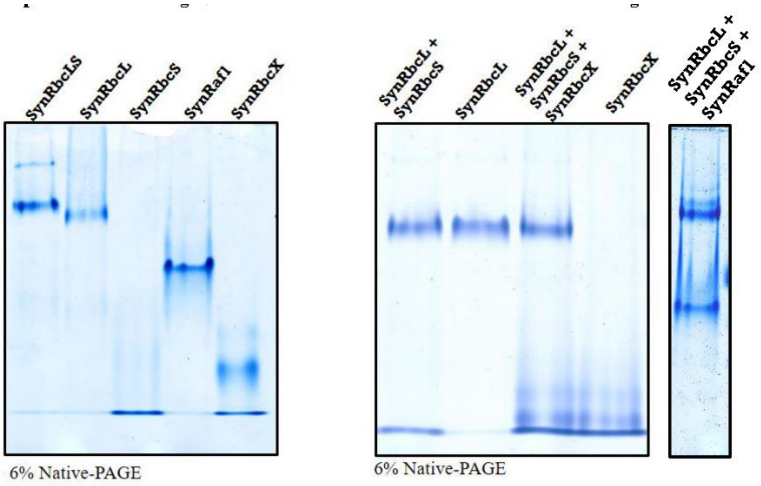


Figure 5.6 **Native-PAGE analysis of *in vitro* reconstituted individual subunits of Syn6301 Rubisco proteins**: Native-PAGE analysis of Syn6301 individual proteins shows that Syn6301_LS, Syn6301_L, Syn6301_S, Syn6301_RbcX and Syn6301_Raf1 runs in their corresponding oligomeric states, whereas in vitro reconstitution experiments does not show higher bands.

5.2.3 Compatibility of large subunits AfLq and Syn6301_L with small subunits from other organisms investigated

The ultimate aim of this project was to express Rubisco by parts *in vivo* in cyanobacterial platform where-in producing kinetically enhanced organisms. *In vitro* reconstitution protocol was utilized to unravel best candidate large subunit, small subunit, and chaperones to be expressed *in vivo*. With the Rubisco proteins purified, we first studied the compatibility of form IA *Acidithiobacillus ferrooxidans* AfLq with different small subunits. It was interesting to explore the compatibility of assembly chaperone not requiring Form IAq AfLq with other Rubisco small subunits that require a chaperone. As seen in Figure 5.7 - A(ii) 0.5 µM AfLq can form functional Rubisco with 2.5 µM form IB *Synechococcus* Small subunit (similarity between the sequences is 37%) but was unable to form functional Rubisco with the eukaryotic Chlamydomonas small subunit CrS (0.1 µM AfLq with 0.8 µM CrS).

Similarly, large subunit Syn6301_L of *Synechococcus elongatus* PCC. 6301 to different small subunits compatibility experiments were done. It was found that 0.5 µM Syn6301_RbcL forms functional Rubisco with 0.8 µM eukaryotic Chlamydomonas small subunit CrS (Figure 5.7 A (i)).

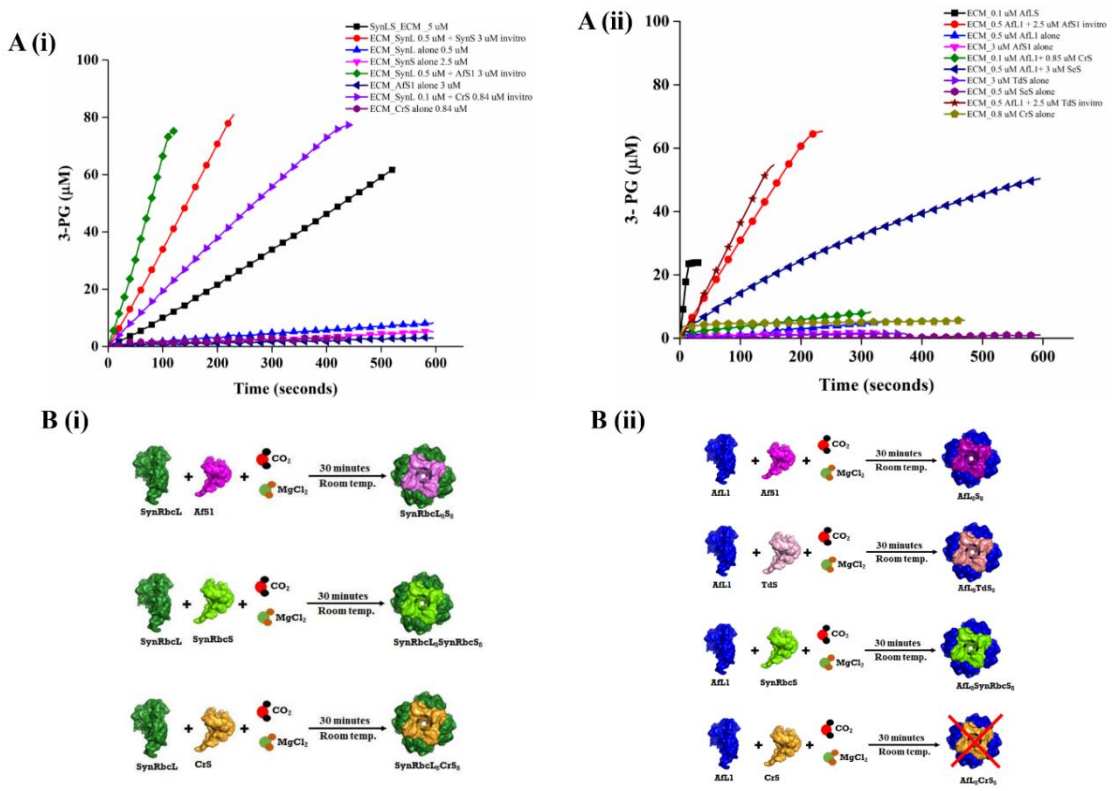


Figure 5.7 **Spectrophotometric Rubisco assay of *Acidithiobacillus ferrooxidans* Rubisco large subunit AfLq and *Synechococcus elongatus PCC. 6301* large subunit Syn6301_L with different small subunits:** A(i) & B(i) *In vitro* reconstitution shows 0.5 µM Syn6301_L can form functional Rubisco with 0.8 µM eukaryotic Chlamydomonas small subunit. A(ii) & B(ii) 0.5 µM AfLq can form functional Rubisco with 2.5 µM form IB *Synechococcus* Small subunit (similarity between the sequences is 37%), but unable to form functional Rubisco with the 0.8 µM eukaryotic Chlamydomonas small subunit.

Finally, we performed the refolding experiments for form IB *Synechococcus elongatus* PCC. 6301. Despite being able to demonstrate the *in vitro* reconstitution of form IB Rubisco using octameric cores and small subunits, attempts to denature and refold failed, possibly due to concentration issues with the Syn6301_L large subunit. Multiple experimental approaches were carried out to optimize these experiments but were unable to yield functional enzymes. More elaborate stepwise folding and assembly steps with different chaperones and additional factors should be adopted.

5.3 Discussion and Conclusions

A vast amount of research time and resources were invested in studying Rubisco for its irreplaceable function and economic importance. Of these, the attempts to study folding and assembly requirements of different Rubisco forms have been limited. Substantial attention was given to cyanobacterial form IB Rubiscos as they display the vast extent of sequence and structural similarity to plant Rubiscos. Understanding cyanobacterial systems can pave the way to comprehend plant Rubiscos better. Unlike the more straightforward form II Rubisco requiring only chaperonins GroEL-GroES, form IB Rubiscos have been established to require a vast array of assembly chaperones. Significant efforts have been put to reconstitute form IB Rubiscos *in vitro*.

In this chapter, we demonstrated that form IB Rubisco proteins of *Synechococcus elongatus* PCC. 6301 carrying the mutation F345I can be cloned and successfully purified as holoenzyme and as separate large subunit Syn6301_L & small subunit Syn6301_S complying to previous literature. Assembly chaperones from form IB cyanobacterial *Synechococcus elongatus* PCC. 6301 Syn6301_RbcX and Syn6301_Raf1 were also purified by expressing in *E.coli*. Form IB *Synechococcus elongatus* PCC. 6301 Rubisco large subunit (containing the mutation F345I) and small subunit can reconstitute *in vitro* to form functional holoenzyme. This observation confirms that the purified large and small units were functional entities.

Greater interest of this project was to produce functionally fitter Rubiscos in cyanobacterial *in vivo* platform. As a starting point to this, we tried to find out different combinations of large and small subunits of Rubisco, which can show better carboxylation activity compared to native ones. From these results, it was observed that form IAq *Acidithiobacillus ferrooxidans* Rubisco large subunit AfLq was compatible with form IB *Synechococcus elongatus* PCC. 6301 Rubisco small subunit. Form IB *Synechococcus elongatus* PCC. 6301 Rubisco large subunit (containing the mutation F345I) was compatible with Form IA *Acidithiobacillus ferrooxidans* Rubisco small subunit and eukaryotic green-algae *Chlamydomonas reinhardtii*.

Additionally, we have tried to adapt the previously explained folding and assembly protocol to form IB Rubisco but was unsuccessful. Recognition of critical factors,

conditions, chaperones to achieve successful refolding was not further investigated in this thesis.

6 *IN VIVO* ENGINEERING OF *SYNECHOCYSTIS* SP. PCC 6803 (Syn6803) TO ESTABLISH A CYANOBACTERIAL RUBISCO ENGINEERING PLATFORM

Rubisco kinetics is diverse among species (Tcherkez *et al.*, 2006, Savir *et al.*, 2010) and Rubisco forms. It is interesting to note that proteobacteria often possess multiple Rubisco forms. For instance, *Rhodobacter sphaeroides* 2.1.4 belonging to the α-proteobacterial phylum possesses both Form IC and Form II Rubiscos, providing it with adaptability to differing atmospheres. The expression of both Rubiscos is related to the metabolic and redox states of the cell. It also depends on the CO_2 and O_2 levels in the surroundings. Form II Rubisco will be favoured at high CO_2 and low O_2 while Form IC comes into work when the environment becomes more aerobic with lesser CO_2. In contrast, currently, it appears that cyanobacteria always possess only a single green-type Form I Rubisco making them less diverse than the former (Badger & Bek, 2008, Zarzycki *et al.*, 2013). Cyanobacteria depend on their efficient carbon-concentrating mechanism for coping with low CO_2 levels. This leads to the expenditure of energy, which may not be required if a different Rubisco enzyme was present. This observation could provide useful outputs in a synthetic biology context.

6.1 Production of red-type Rubisco in *Synechocystis* sp. PCC 6803

As a first step towards engineering a heterologous Rubisco system into *Synechocystis* sp. PCC 6803 (hereafter *Syn*6803), the red-type Rubisco from *Rhodobacter sphaeroides* (hereafter *Rs*) was used. This attempt can give us insights on the folding and assembly of the red-type Rubisco in an organism containing green-type Rubisco. The sequence identity between these two sequences was ~50% (Figure 6.1.)

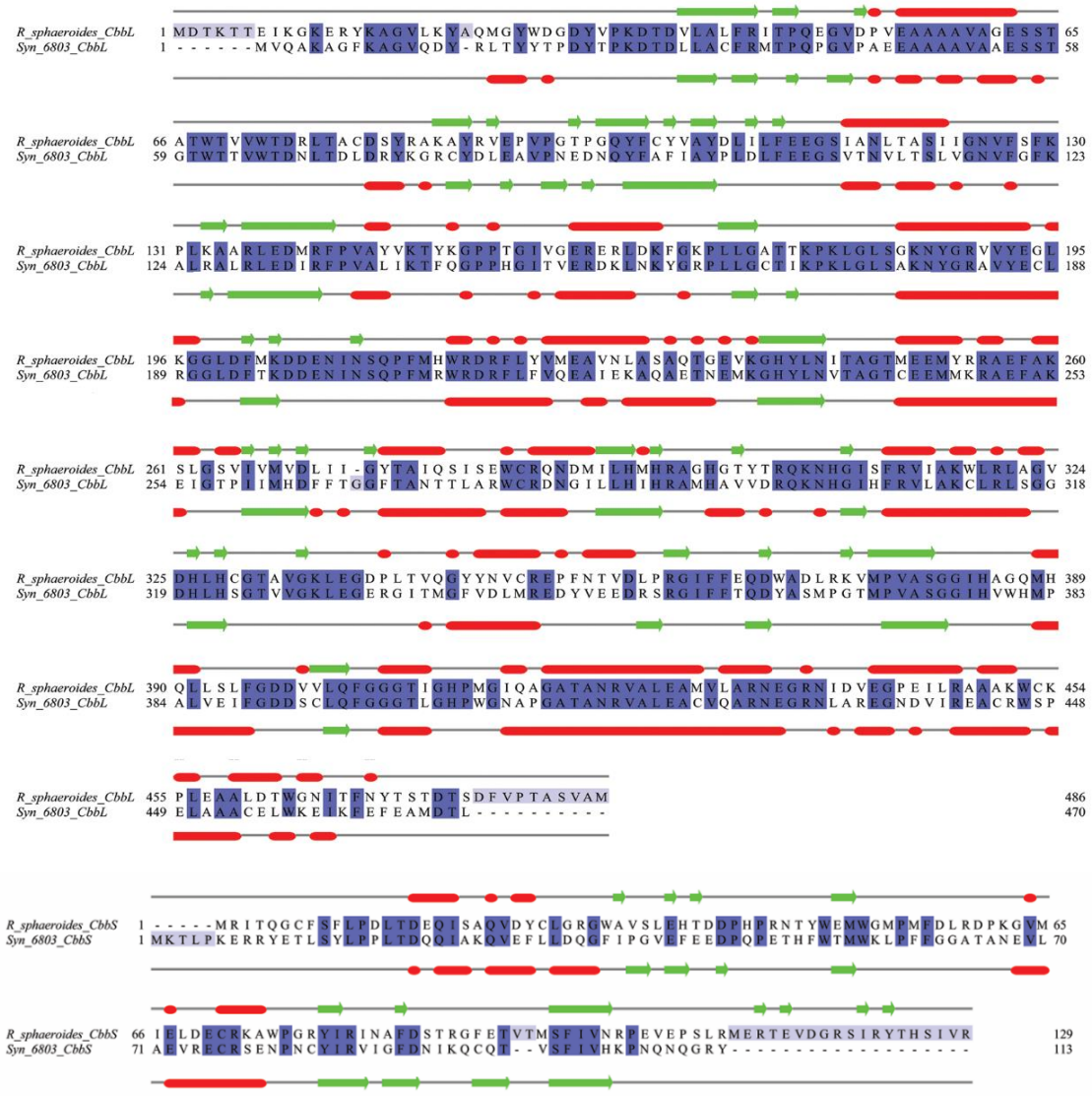


Figure 6.1 **Sequence alignment of CbbL and CbbS from the red-type (Rs) and green-type (Syn6803) Rubisco:** Coloured based on percentage identity. Alignment made using ClustalO (Sievers *et al.*, 2011) and Figure was generated using Jalview (Waterhouse *et al.*, 2009). Secondary structure (Red tubes and green arrows represent helices and sheets respectively) depiction based on PDB Id 4F0K (top) and 1RBL (bottom).

6.1.1 Introducing a red-type Rubisco system into S*ynechocystis* sp. PCC 6803

Firstly, only the large subunit CbbL and the small subunit CbbS were cloned into the vector pPD-Flag (Hollingshead *et al.*, 2012). pPD-FLAG contains the *Synechocystis* psbA II promoter, a sequence encoding the FLAG tag and flanking sequences for homologous recombination that allows insertion of tagged constructs into the *Synechocystis* genome in place of the psbA II gene.). The presence of the correct *Rs*CbbLS was confirmed by analytical restriction digestion using specific restriction enzymes and subsequent sequencing of the plasmid (Figure 6.2).

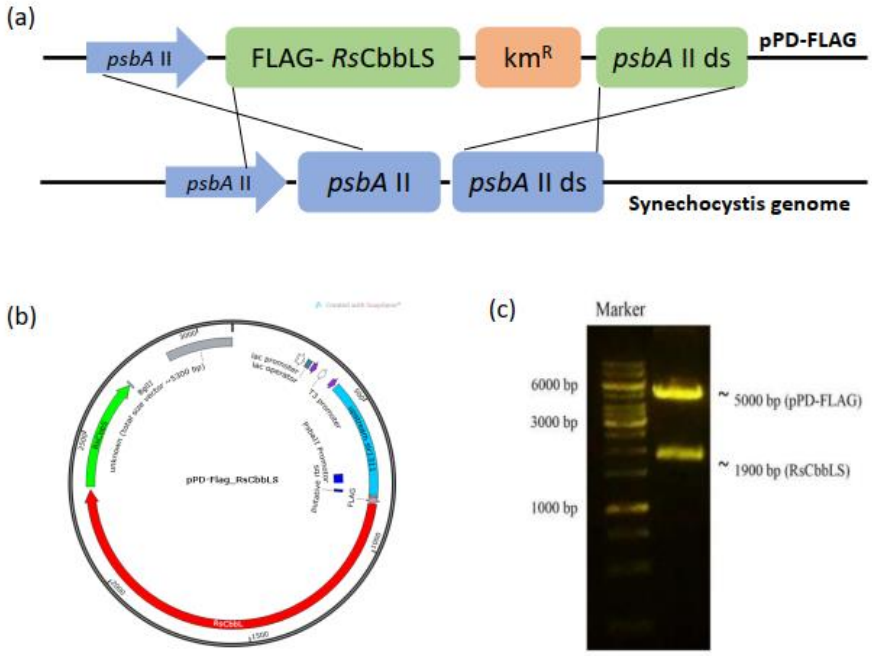


Figure 6.2 **Restriction digestion analysis of pPD-Flag carrying *Rs*CbbLS**: (a) Cartoon depiction of the homologous recombination event, Upon transformation into Synechocystis the two regions of homology align with their counterpart regions in the Synechocystis genome. A homologous recombination event occurs, inserting the FLAG- tagged RsCbbLS into the cyanobacterial genome (b) Expression plasmid map made using Snapgene software (c)Agarose gel picture showing the restriction digestion of the construct pPD-Flag_ *Rs*CbbLS

The wild-type cyanobacterial strain that is glucose tolerant (GT) was grown in liquid BG 11 medium containing 5 mM glucose with illumination provided and maintained at 30° C. The culture is grown until the optical density (OD) of 2.5 before a transformation is performed. It takes 2-3 days to grow, once the inoculation is done (Eaton-Rye, 2011). The successfully made pPD-Flag_*Rs*CbbLS construct was then introduced into the wild-type *Syn*6803. After the transformation of the pPD-Flag-*Rs*CbbLS into *Syn*6803 cells,

they were grown on the BG 11 agar plates. The transformants after successful transformation are a result of homologous DNA recombination events into the host genome. Chromosomal segregation is ensured by selection using antibiotic resistance. First the transformed culture was grown in agar plates without antibiotics and after 24 hours it was re-streaked on agar plates containing 12.5 µg/mL of kanamycin (Figure 6.3 a) The colonies from this plate were subsequently grown in increasing concentrations for kanamycin (50 µg/ mL, 100 µg/ mL) (Figure 6.3 b and c).

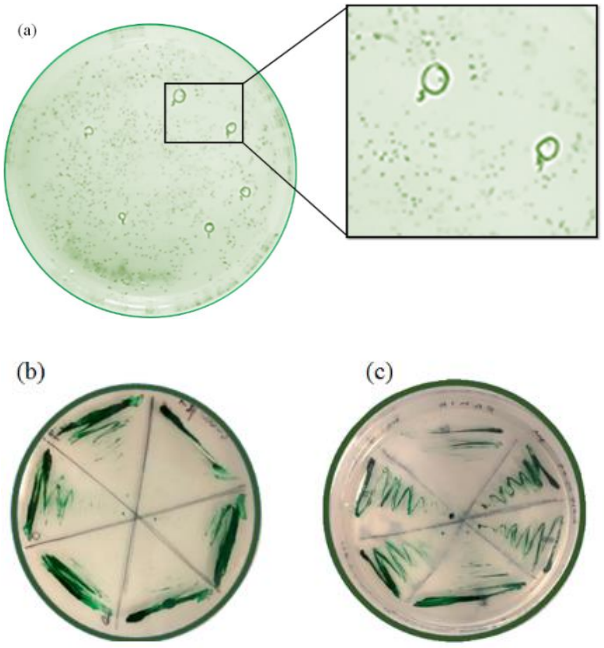


Figure 6.3 (a) **Cyanobacterial strain colonies:** Representative image shows the Syn6803 colonies on a BG 11 agar plate with 12.5 µg/mL of kanamycin. The circles indicate the colonies that were picked for re-streaking in subsequently increased concentrations of kanamycin. (b) and (c) Representative image shows the colonies in the re-streaking process in subsequently increased concentrations of 50 µg/mL and 100 µg/mL of kanamycin, respectively.

After multiple re-streaking with increasing kanamycin concentrations (up to 200 µg/mL), the samples were prepared for testing the protein expression. Individual samples from each of the colonies were preliminarily analysed by immuno-blotting using anti-flag antibody (data not shown). It was seen that the *Rs*CbbL$_8$S$_8$ was successfully expressed in the host *Syn*6803. This shows that the red-type Rubisco can be expressed and accumulated in a green type system. The next step was to introduce the activase *Rs*CbbX into the cyanobacterial system carrying the *Rs*CbbL$_8$S$_8$. The construct was done using the technique of restriction free cloning (rfc). After the primary PCR, the product called the megaprimer was used for subsequent secondary PCR

reactions as primer (Figure 6.4. a). After the secondary PCR reaction, Dpn1 treatment was done and transformed into the DH5α cells. The presence of the desired insert was confirmed by a colony PCR (Figure 6.4. b) and the insert was confirmed by subsequent sequencing results.

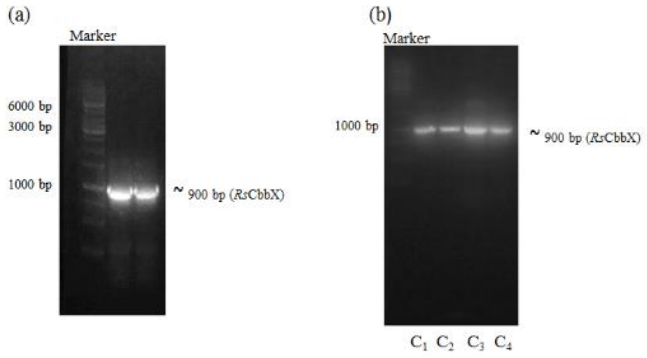


Figure 6.4 **Agarose gel images showing *Rs*CbbX:** (a) Agarose gel image showing the product of primary rfc reactions (megaprimer). The desired product *Rs*CbbX was seen in the correct size (b) Agarose gel image of DH5α colony PCR showing the presence of the desired insert *Rs*CbbX after transformation.

After the introduction of the activase, the plasmid was transformed into the cyanobacterial *Syn*6803 system. Colonies were seen in the BG 11 agar plates and were re-streaked in solid medium with increasing concentrations of kanamycin (200 µg/mL) and samples were analysed. It was found that the CbbX expression could not be demonstrated by immunoblotting. We introduced an additional promoter region for the activase and this next construct was made using restriction free cloning for the introduction of extra psbAII promoter region (Eriksson *et al.*, 2000, Iwaki *et al.*, 2006, Hollingshead *et al.*, 2012). The promoter region which is ~125 bp long was synthesized as a megaprimer and inserted in the plasmid carrying *Rs*CbbLSX by secondary PCR reactions. The presence of the promoter was confirmed by colony PCR and subsequent sequencing results. The new plasmid of the desired construct was transformed into the WT-GT strain of *Syn*6803. After successful transformation colonies were seen in the BG 11 agar plates. It was noticed that only five colonies were seen in contrary to the previous constructs which had ~100 colonies. The five colonies were picked and similar re-streaking on increasing concentrations of kanamycin was performed. The colonies grew well and was undergoing segregation process.

After the cyanobacterial cells were grown with considerable increased concentrations of antibiotic (maximum of 200 µg/mL), the cell extracts from all the produced strains

were preliminarily analysed for the protein expression using 12.5 % SDS-PAGE. Purified *R. sphaeroides* Rubisco (3 µg) was loaded for comparison as the protein standard and wild-type *Synechocystis* sp. PCC 6803 was used as the negative control. The concentration of protein in the *Synechocystis* strains were quantified using Bradford assay after TCA precipitation for each sample, the same amount of the cyanobacterial total protein (2 µg) was loaded per lane in the SDS-PAGE gel (Figure 6.5.a). The samples were analysed using immunoblotting applying anti-flag antibody. Specific signals seen in the newly constructed strains confirmed the expression of *Rs*CbbL$_8$S$_8$ in the host *Syn*6803 (Figure 6.5 b). The samples were then analysed using 6% native-PAGE to check the assembly competence of *R. sphaeroides* Rubisco and it indicated the Rubisco was expressed. The native-PAGE gel was further analysed using immunoblotting using anti-Flag and anti-RbcL antibody to confirm that the assembled protein in the cyanobacterial cells is the desired red-type Rubisco (Figure 6.6)

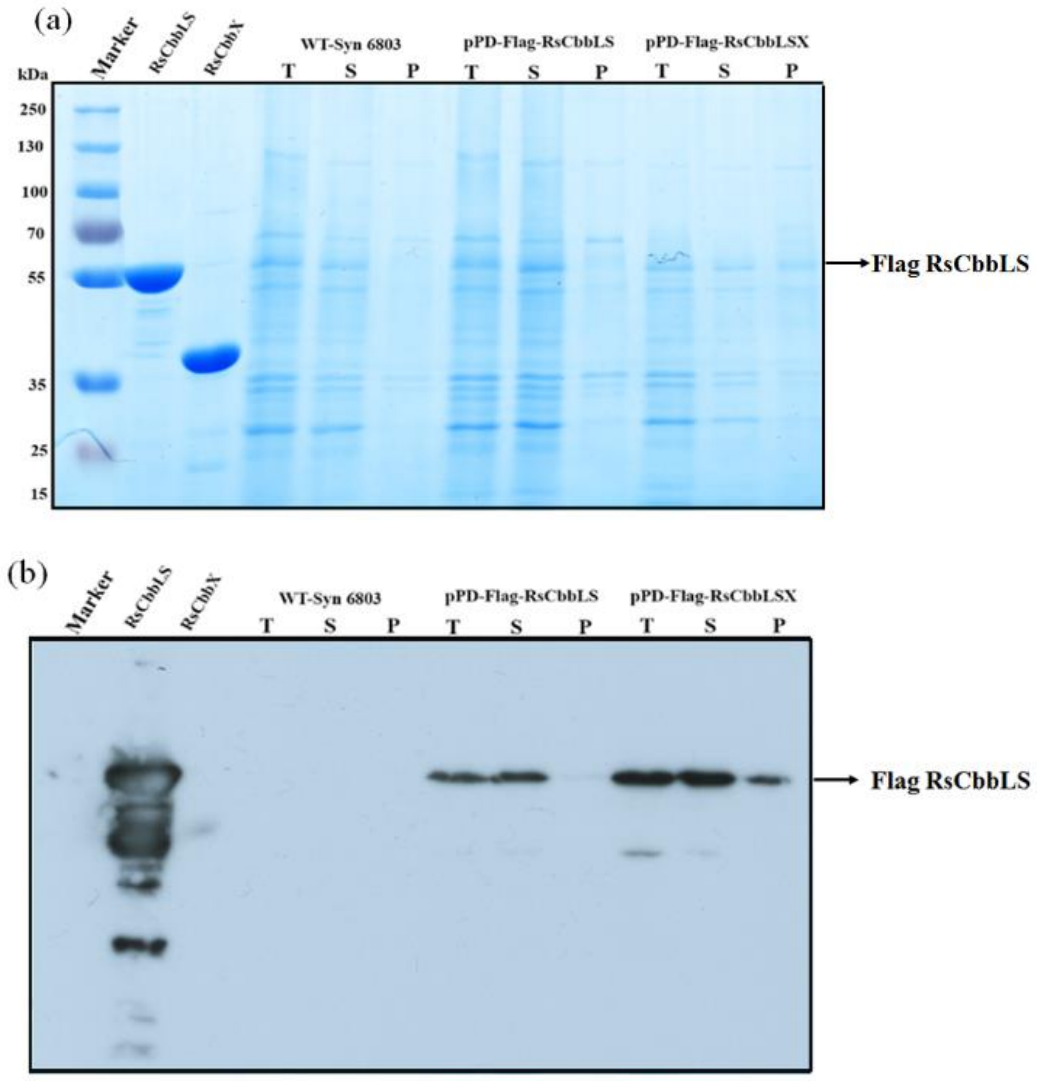


Figure 6.5 (a) **SDS-PAGE gel of cyanobacterial cell extracts:** 12.5% SDS-PAGE gel of cell extracts of wild-type and transformed cyanobacterial strains analysed. T: Total cell lysate, S: Supernatant with the soluble protein and P: Pellet.

(b) **Immunoblot of Cyanobacterial strains:** Anti – flag Western blot of 12.5% SDS-PAGE gel showing the expression of desired *Rs*CbbLS in the cyanobacterial *Syn*6803 system. 2 µg of each fraction the total protein from the sample was loaded.

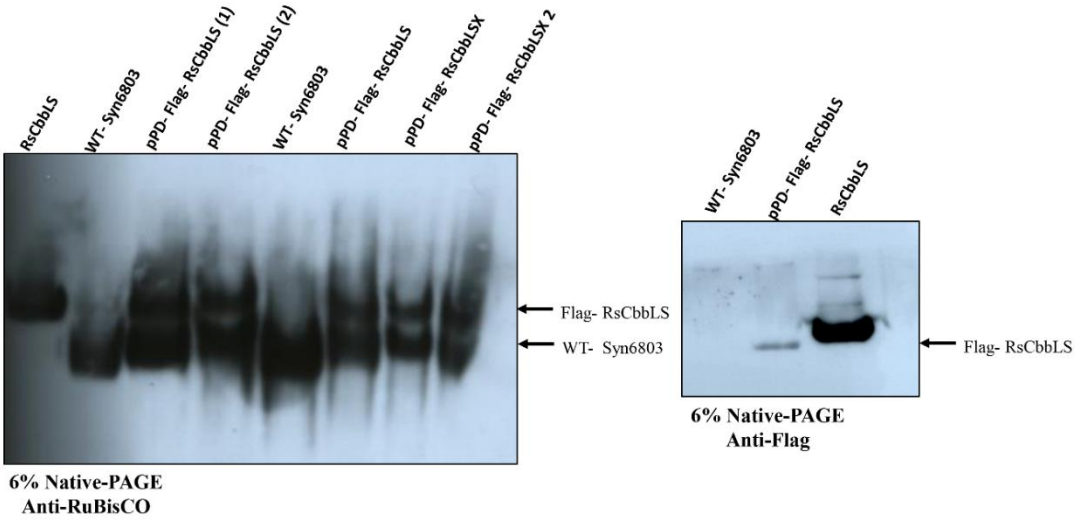


Figure 6.6 **Native-PAGE gel of cyanobacterial cell extracts:** Protein expression analysed using 6% native-PAGE gel with 3 µg of purified protein for positive control and 2 µg of protein in test samples loaded to each lane showing that the RsCbbL8S8 is expressed in the cyanobacterial strains designed. . Left panel is anti-RbcL (1:5000) western blot of 6% native-PAGE gel with 3 µg of purified protein for positive control and 2 µg of protein in test samples loaded to each lane, showing the desired RsCbbLS is assembled in the cyanobacterial strains designed. It shows two signals for the Rubisco clearly stating that the RsCbbL8S8 is assembled in the cyanobacterial constructs along with the endogenous Rubisco. Anti-rabbit (1: 10,000) was used as the secondary antibody. Right panel is anti-Flag (1:5000 dilution) western blot of 6% native-PAGE gel with 3 µg of purified protein for positive control and 2 µg of protein in test samples loaded to each lane, showing that the RsCbbL8S8 is assembled in the cyanobacterial strains designed. Anti-mouse (1: 10,000) was used as the secondary antibody. The negative control wild-type Synechocystis does not give a signal corresponding to the FLAG epitope.

A critical concept in cyanobacterial culture is checking for complete segregation, procedure explained in methods section. The colonies from all the above strains were checked for segregation. It was found that only the cyanobacterial strain with the *Rs*CbbLS is fully segregated showing the WT- specific PCR signal (Figure 6.7). The other constructs were still undergoing segregation.

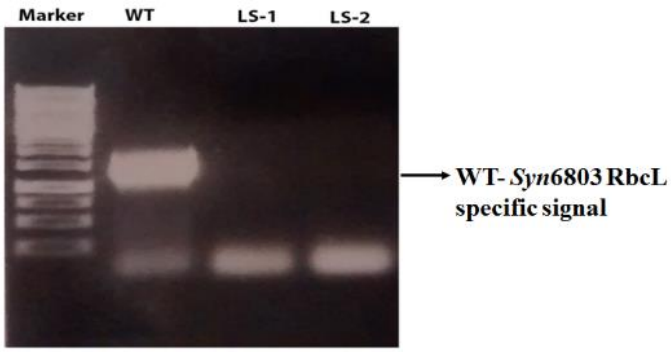


Figure 6.7 **Agarose gel image demonstrating completely segregated Cyanobacterial strain:** Gel image showing the complete segregation of the colonies from the *Syn*6803 carrying the *Rs*CbbLS.

6.1.2 Knockout of the endogenous cyanobacterial Rubisco

The strategy for knocking out the endogenous Rubisco from *Syn*6803 was performed as explained in methods section 2.5.2. The genomic DNA was extracted from the wild-type *Syn*6803 and used as the template for amplification of the large subunit RbcL from *Synechocystis* sp. PCC 6803 (Figure 6.8 a). The amplified *Syn*6803RbcL was introduced into the sub-cloning vector pGem T as explained earlier. The blue-white screeing of the colonies was performed and the subsequent plasmid was isolated, the insert was confirmed by sequencing. The chloramphenicol resistance cassette (CmR) was amplified from the pSK9 plasmid (Figure 6.8 b) (Kuchmina *et al.*, 2012) that encodes for CmR. Restriction free cloning was used to introduce the CmR in between the amplified RbcL carried by pGem T. The plasmid carrying the CmR interrupting the RbcL gene of *Syn*6803 was transformed into the strains generated above to knockout the endogenous Rubisco. This strategy aims on revealing if the knock-out of the endogenous *Syn*6803 Rubisco can be complemented by the heterologous recombination of the red-type *Rs*CbbLS_*Rs*CbbX system taking over the role in carbon fixation. A pre-condition is that the implemented Rubisco-activase system is functional in the context of the cyanobacterial cell. Knock-out mutants will be screened by taking advantage of CmR introduced.

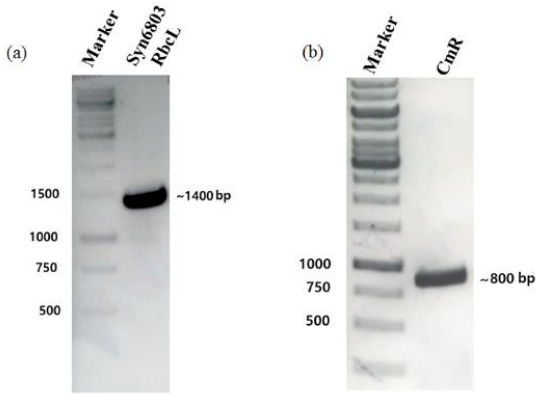


Figure 6.8 **Agarose gel image showing the PCR amplification of Syn6803 large subunit RbcL (a) and CmR (b) from the pSK9 plasmid**.

After transformation, the cells were streaked to BG 11 agar plates containing both the antibiotics kanamycin and chloramphenicol. It was seen that the colonies took longer time to appear than the normal strains.

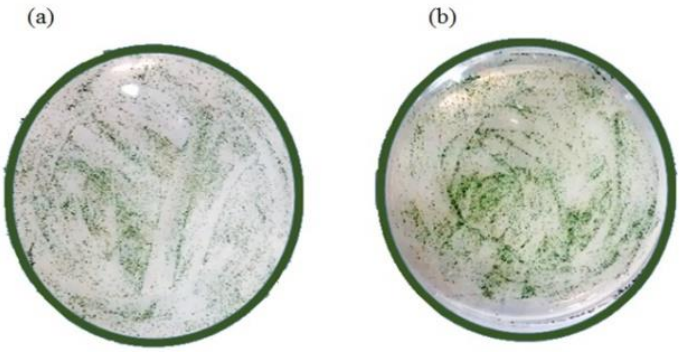


Figure 6.9 **Knock-out Cyanobacterial strain colonies:** The image shows the colonies in the re-streaking process in BG 11 agar plates containing 25 µg/mL kanamycin and 15 µg/mL chloramphenicol. (a) and (b) represents the knockout of *Rs*CbbLSX and *Rs*CbbLS-psbAII-*Rs*CbbX, respectively.

The colonies were checked for segregation using colony-PCR and it was found that they have not started to segregate yet. Hence, they were subjected to further increasing concentrations of antibiotics to promote the process. After confirming the segregation, the samples can be tested for protein expression. The growth of knock-out cyano-red strains was obstructed due to stress caused by increased concentrations of antibiotic (100 µg/mL of kanamycin and 60 µg/mL of chloramphenicol). It showed that functional supplementation of the red-type Rubisco did not restore Rubisco function.

6.2 Discussion and Conclusions

Cyanobacteria are photosynthetic organisms playing a pivotal role in global carbon fixation, making them the target organism for improving the highly essential but inefficient carbon fixation enzyme Rubisco. The higher aim of this project is to generate cyanobacterial strains that are fitter than the wild-type under specific conditions such as elevated levels of CO_2. While this goal is long term and not immediately achievable, the work performed here represents the progress towards these ends. The cyanobacterial system is being used as the platform for Rubisco engineering. This work describes the construction of cyanobacterial strains supplemented with a functional replacement for Rubisco. Firstly, the strains were made by introducing the functional replacement along with the endogenous Rubisco. Then the knock-out of endogenous Rubisco was attempted by the insertional inactivation of the RbcLS from *Synechocystis* 6803 (*Syn*6803) after supplementing it with the equivalent genes from *Rhodobacter sphaeroides* 2.4.1 (*Rs*) encoding the functional holoenzyme. It should be noted that an earlier attempt employed the form II *Rhodospirillum rubrum* Rubisco *(Pierce et al., 1989, Amichay et al., 1993)* and the attempt in this study involves form I red-type *Rhodobacter sphaeroides* 2.4.1 Rubisco.

We have thus far successfully demonstrated that the red-type Rubisco could be produced, folded, and assembled in the cyanobacterial host (cyano-red). This is of particular interest since red-type Rubiscos from eukaryotic phytoplankton have been shown to have much higher CO_2/O_2 specificity values than those from organisms harboring green-type Rubiscos including the higher plants (Whitney *et al.*, 2001, Andrews & Whitney, 2003). Demonstration that a bacterial homolog of these "superior" enzymes can be produced in cyanobacteria bodes well for future work involving the more challenging eukaryotic enzymes.

Although red-type Rubiscos often possess desirable kinetic properties, they are always found associated with genes encoding the red-type Rubisco activase CbbX. It hence appears evident that to take full advantage of red-type enzymes *in-vivo*, the associated activase will need to be co-expressed. Although we were able to obtain transformants encoding both red-type Rubisco and CbbX, we have so far been unable to demonstrate CbbX production by immunoblotting. Nevertheless, we decided to include an additional psbAII promoter for CbbX to achieve higher expression levels (Hollingshead *et al.*,

2012). It was noticed that very few colonies appeared in the BG 11 agar plate even after incubation for more than 4 days. This showed that the introduction of additional promoter could have possibly put the cyanobacteria upon stress. After the appearance of colonies, it took time for them to grow more significant enough to be carried on for re-streaking process. When the sample was preliminarily analyzed for protein expression using 12.5% SDS-PAGE, it showed that all the new cyano-red strains could produce the desired *Rs* CbbLS proteins (Figure 6.5.). Immunoblot analysis of Native-PAGE gels with anti-Flag and Anti-RbcL showed that the desired holoenzyme complex of *Rs*CbbLS is expressed and assembled in the cyano-red strains (Figure 6.6).

When the samples from pPD-Flag-*Rs*CbbLSX and pPD-Flag-*Rs*CbbLS-psbAII2-*Rs*CbbX were analysed using colony PCR, it was found that they are not yet fully segregated through all chromosome copies. Hence the re-streaking with increased antibiotics selection will be continued for these strains to get segregated entirely. The next was the knock-out of the endogenous Rubisco from the new strains carrying the fully assembled red-type Rs Rubisco which will now serve as the functional supplement. This part of the project is still in progress and may eventually depend on CO_2 supplementation of the cells if CbbX is not produced in significant amounts.

7 SUMMARY AND FUTURE WORK

Rubisco (Ribulose 1, 5-bisphosphate carboxylase/oxygenase) represents the main gateway of inorganic carbon entry into the biosphere by catalysing the carbon (CO_2)-fixation step in photosynthesis. Despite its pivotal role, it is an enzyme with poor kinetic parameters and substrate specificity. However, a great diversity of kinetic parameters is encountered across the natural world, engendering hope of encountering enzymes more suited for environmental conditions. Rubisco biogenesis has not been completely understood yet. Therefore, through *in vitro* methods we have developed a method to understand folding and assembly of Rubisco.

In this thesis, the *in vitro* reconstitution of Rubisco was focused on studying the chaperonin- assisted folding and assembly requirements of diverse Rubiscos using electrophoretic analysis and spectrophotometer-based assays. This helps us to rapidly evaluate the folding and assembly process unlike the laborious radioactivity-based techniques used so far. A protocol for following the *in vitro* reconstitution of form II Rubiscos in real-time was established. It was demonstrated that form II dimeric Rubisco can fold and assemble with only the chaperonins GroEL and GroES whereas the assembly of hexameric form II Rubisco is promoted by the substrate RuBP *in vitro*. Form II hexameric *A. ferrooxidans* (AfM) maturation is slower and assembly proceeds via dimeric intermediates. Form IAq *Acidithiobacillus ferrooxidans* Rubisco large subunit can assemble independently to octamers in the absence of ancillary proteins and requires only small subunit to become functional. Successfully isolated form IAc Rubisco large subunit dimers can assemble to octamers with the help of Rubisco activation factors CO_2 and Mg^{2+}.

In summary, our work explored the variance in folding and assembly requirements of major Rubisco forms comprising the recent advancements. Comprehensive understanding of diverse Rubisco forms is critical to engineer strategies to overcome bottlenecks of Rubisco which can be harnessed for further agricultural improvements and biotechnological innovations. The next step in this project will be the biochemical characterization of form IAc large subunit dimers and adapting the protocol to other forms of Rubiscos. This extension might involve including additional chaperones and co-factors. This attempt can result in interesting findings and novel procedure.

Simultaneously, we have shown that red-type Rubisco can be successfully expressed, folded and assembled in a cyanobacterial host. Further direction of this project is to characterize if the supplemented Rubisco is functional or not. This will be assessed by characterizing the cyano-red's ability to grow photoautotrophically. If it is found that

the strain can grow photoautotrophically, then this cyanobacterial system can serve as an expression system to verify efficient compatibilities between any Rubisco and activase in the photosynthetic active cell. It can reveal whether other Rubisco systems can also rescue cyanobacteria with knocked-out endogenous Rubisco more efficiently than the form II Rubisco used earlier in the literature because the red-type we substituted has superior catalytic properties than *Rhodospirillum rubrum.* If found that this Rubisco also assembles competently in the cyanobacteria strain and functionally be active, then it could be concluded that other efficient Rubisco systems can also be supplemented to the cyanobacterial strain to obtain a kinetically better-performing organism than the wild-type.

8 REFERENCES

Aigner H & Wilson RH (2017) Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. **358**: 1272-1278.

Alonso H, Blayney MJ, Beck JL & Whitney SM (2009) Substrate-induced assembly of Methanococcoides burtonii D-ribulose-1, 5-bisphosphate carboxylase/oxygenase dimers into decamers. *Journal of Biological Chemistry* **284**: 33876-33882.

Amichay D, Levitz R & Gurevitz M (1993) Construction of a Synechocystic PCC6803 mutant suitable for the study of variant hexadecameric ribulose bisphosphate carboxylase/oxygenase enzymes. *Plant molecular biology* **23**: 465-476.

Andersson I (2008) Catalysis and regulation in Rubisco. *Journal of experimental botany* **59**: 1555-1568.

Andrews TJ & Whitney SM (2003) Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants. *Archives of biochemistry and biophysics* **414**: 159-169.

Anfinsen CB (1973) Principles that Govern the Folding of Protein Chains. *Science* **181**: 223-230.

Ashida H, Saito Y, Nakano T, de Marsac NT, Sekowska A, Danchin A & Yokota A (2008) RuBisCO-like proteins as the enolase enzyme in the methionine salvage pathway: functional and evolutionary relationships between RuBisCO-like proteins and photosynthetic RuBisCO. *Journal of experimental botany* **59**: 1543-1554.

Badger MR & Bek EJ (2008) Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *Journal of experimental botany* **59**: 1525-1541.

Badger MR, Andrews TJ, Whitney SM, Ludwig M, Yellowlees DC, Leggat W & Price GD (1998) The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. *Canadian Journal of Botany* **76**: 1052-1071.

Balchin D, Hayer-Hartl M & Hartl FU (2016) In vivo aspects of protein folding and quality control. *Science* **353**: aac4354.

Bassham JA (2003) Mapping the carbon reduction cycle: a personal retrospective. *Photosynthesis research* **76**: 35-52.

Bathellier C, Tcherkez G, Lorimer GH & Farquhar GD (2018) Rubisco is not really so bad. *Plant, cell & environment* **41**: 705-716.

Bauwe H, Hagemann M & Fernie AR (2010) Photorespiration: players, partners and origin. *Trends in Plant Science* **15**: 330-336.

Beere HM (2004) `The stress of dying': the role of heat shock proteins in the regulation of apoptosis. *Journal of Cell Science* **117**: 2641-2651.

Ben-Zvi AP & Goloubinoff P (2001) Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *Journal of structural biology* **135**: 84-93.

Berg JM, Tymoczko JL & Stryer L (2006) Biochemistry. 5th. *New York: WH Freeman* **38**: 76.

Berg JM, Tymoczko JL & Stryer L (2012) *Biochemistry*. W.H. Freeman, Basingstoke.

Bhat JY, Thieulin-Pardo G, Hartl FU & Hayer-Hartl M (2017) Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Frontiers in Molecular Biosciences* **4**.

Bhat JY, Miličić G, Thieulin-Pardo G, *et al.* (2017) Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Molecular Cell* **67**: 744-756.e746.

Bracher A, Whitney SM, Hartl FU & Hayer-Hartl M (2017) Biogenesis and Metabolic Maintenance of Rubisco. *Annu Rev Plant Biol*.

Bracher A, Hauser T, Liu C, Hartl FU & Hayer-Hartl M (2015) Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from Chlamydomonas reinhardtii. *PloS one* **10**: e0135448.

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL & Sigler PB (1994) The crystal structure of the bacterial chaperonln GroEL at 2.8 Å. *Nature* **371**: 578.

Brutnell TP, Sawers RJ, Mant A & Langdale JA (1999) BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the rbcL gene of maize. *The Plant cell* **11**: 849-864.

Bukau B & Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**: 351-366.

Bukau B, Weissman J & Horwich A (2006) Molecular Chaperones and Protein Quality Control. *Cell* **125**: 443-451.

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

Castelli C, Rivoltini L, Rini F, Belli F, Testori A, Maio M, Mazzaferro V, Coppa J, Srivastava PK & Parmiani G (2004) Heat shock proteins: biological functions and clinical application as personalized vaccines for human cancer. *Cancer Immunology, Immunotherapy* **53**: 227-233.

Catanzariti A-M, Soboleva TA, Jans DA, Board PG & Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science* **13**: 1331-1339.

Cegielska A & Georgopoulos C (1989) Functional domains of the Escherichia coli dnaK heat shock protein as revealed by mutational analysis. *Journal of Biological Chemistry* **264**: 21122-21130.

Chaijarasphong T, Nichols RJ, Kortright KE, Nixon CF, Teng PK, Oltrogge LM & Savage DF (2016) Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *Journal of Molecular Biology* **428**: 153-164.

Chandrasekhar GN, Tilly K, Woolford C, Hendrix R & Georgopoulos C (1986) Purification and properties of the groES morphogenetic protein of Escherichia coli. *J Biol Chem* **261**: 12414-12419.

Cleland WW, Andrews TJ, Gutteridge S, Hartman FC & Lorimer GH (1998) Mechanism of Rubisco: The Carbamate as General Base. *Chemical reviews* **98**: 549-562.

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

Dangel AW & Tabita FR (2015) CbbR, the Master Regulator for Microbial Carbon Dioxide Fixation. *Journal of Bacteriology* **197**: 3488.

Dobson CM (2003) Protein folding and misfolding. *Nature* **426**: 884-890.

Dodd AN, Borland AM, Haslam RP, Griffiths H & Maxwell K (2002) Crassulacean acid metabolism: plastic, fantastic. *Journal of experimental botany* **53**: 569-580.

Eaton-Rye JJ (2011) Construction of gene interruptions and gene deletions in the cyanobacterium Synechocystis sp. strain PCC 6803. *Methods Mol Biol* **684**: 295-312.

Ellis RJ (1979) The most abundant protein in the world. *Trends in Biochemical Sciences* **4**: 241-244.

Emlyn-Jones D, Woodger FJ, Price GD & Whitney SM (2006) RbcX can function as a rubisco chaperonin, but is non-essential in Synechococcus PCC7942. *Plant & cell physiology* **47**: 1630-1640.

Emlyn-Jones DE-J, Whitney S & Price D (2001) Substitution of foreign Rubiscos in the cyanobacterium, Synechococcus PCC7942. *Science Access* **3**: -.

Eriksson J, Salih GF, Ghebramedhin H & Jansson C (2000) Deletion mutagenesis of the 5′ psbA2 region in Synechocystis 6803: identification of a putative cis element involved in photoregulation. *Molecular Cell Biology Research Communications* **3**: 292-298.

Espie GS & Kimber MS (2011) Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynthesis research* **109**: 7-20.

Esser C, Alberti S & Hohfeld J (2004) Cooperation of molecular chaperones with the ubiquitin/proteasome system. *Biochimica et biophysica acta* **1695**: 171-188.

Feiz L, Williams-Carrier R, Wostrikoff K, Belcher S, Barkan A & Stern DB (2012) Ribulose-1,5-Bis-Phosphate Carboxylase/Oxygenase Accumulation Factor1 Is Required for Holoenzyme Assembly in Maize. *The Plant cell* **24**: 3435.

Feiz L, Williams-Carrier R, Belcher S, Montano M, Barkan A & Stern DB (2014) A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *The Plant Journal* **80**: 862-869.

Fernández-Fernández MR, Gragera M, Ochoa-Ibarrola L, Quintana-Gallardo L & Valpuesta JM (2017) Hsp70 – a master regulator in protein degradation. *FEBS Letters* **591**: 2648-2660.

Field CB, Behrenfeld MJ, Randerson JT & Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237-240.

Flaherty KM, DeLuca-Flaherty C & McKay DB (1990) Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**: 623-628.

Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, Oltrogge LM, Alves R, Savage D & Milo R (2019) Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* **58**: 3365-3376.

Fristedt R, Hu C, Wheatley N, *et al.* (2018) RAF2 is a Rubisco Assembly Factor in Arabidopsis thaliana. *The Plant journal : for cell and molecular biology*.

GARRIDO C, BRUEY J-M, FROMENTIN A, HAMMANN A, ARRIGO AP & SOLARY E (1999) HSP27 inhibits cytochrome c-dependent activation of procaspase-9. *The FASEB Journal* **13**: 2061-2070.

Genest O, Wickner S & Doyle SM (2019) Hsp90 and Hsp70 chaperones: Collaborators in protein remodeling. *Journal of Biological Chemistry* **294**: 2109-2120.

Genest O, Hoskins JR, Kravats AN, Doyle SM & Wickner S (2015) Hsp70 and Hsp90 of E. coli Directly Interact for Collaboration in Protein Remodeling. *Journal of Molecular Biology* **427**: 3877-3889.

Gething M-J & Sambrook J (1992) Protein folding in the cell. *Nature* **355**: 33-45.

Goloubinoff P, Christeller JT, Gatenby AA & Lorimer GH (1989) Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342**: 884.

Gornall J, Betts R, Burke E, Clark R, Camp J, Willett K & Wiltshire A (2010) Implications of climate change for agricultural productivity in the early twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 2973-2989.

Gunn LH & Valegard K (2017) A unique structural domain in Methanococcoides burtonii ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) acts as a small subunit mimic. **292**: 6838-6850.

Gutsche I, Essen L-O & Baumeister W (1999) Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *Journal of Molecular Biology* **293**: 295-312.

Hamilton TL, Bryant DA & Macalady JL (2015) The role of biology in planetary evolution: cyanobacterial primary production in low oxygen Proterozoic oceans. *Environmental microbiology*.

Harrison C (2003) GrpE, a nucleotide exchange factor for DnaK. *Cell Stress Chaperones* **8**: 218-224.

Hartl FU (1996) Molecular chaperones in cellular protein folding. *Nature* **381**: 571-579.

Hartl FU (2017) Protein Misfolding Diseases. *Annual Review of Biochemistry* **86**: 21-26.

Hartl FU & Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**: 1852-1858.

Hartl FU, Bracher A & Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**: 324-332.

Hartman FC & Harpel MR (1994) Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu Rev Biochem* **63**: 197-234.

Haslbeck M, Franzmann T, Weinfurtner D & Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nature Structural & Molecular Biology* **12**: 842-846.

Hasse D, Larsson AM & Andersson I (2015) Structure of Arabidopsis thaliana Rubisco activase. *Acta crystallographica Section D, Biological crystallography* **71**: 800-808.

Hauser T, Popilka L, Hartl FU & Hayer-Hartl M (2015) Role of auxiliary proteins in Rubisco biogenesis and function. *Nature Plants* **1**: 15065.

Hauser T, Bhat JY, Milicic G, Wendler P, Hartl FU & Bracher A (2015) Structure and mechanism of the Rubisco-assembly chaperone Raf1. **22**: 720-728.

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

Hazra S, Henderson JN, Liles K, Hilton MT & Wachter RM (2015) Regulation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase: PRODUCT INHIBITION, COOPERATIVITY, AND MAGNESIUM ACTIVATION. *The Journal of biological chemistry* **290**: 24222-24236.

Henderson JN, Kuriata AM, Fromme R, Salvucci ME & Wachter RM (2011) Atomic resolution x-ray structure of the substrate recognition domain of higher plant ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase. *The Journal of biological chemistry* **286**: 35683-35688.

Hollingshead S, Kopečná J, Jackson PJ, Canniffe DP, Davison PA, Dickman MJ, Sobotka R & Hunter CN (2012) Conserved Chloroplast Open-reading Frame ycf54 Is Required for Activity of the Magnesium Protoporphyrin Monomethylester Oxidative Cyclase in Synechocystis PCC 6803. *The Journal of biological chemistry* **287**: 27823-27833.

Horovitz A & Willison KR (2005) Allosteric regulation of chaperonins. *Current Opinion in Structural Biology* **15**: 646-651.

Iwaki T, Haranoh K, Inoue N, *et al.* (2006) Expression of foreign type I ribulose-1,5-bisphosphate carboxylase/ oxygenase (EC 4.1.1.39) stimulates photosynthesis in cyanobacterium Synechococcus PCC7942 cells. *Photosynthesis research* **88**: 287-297.

Javid B, MacAry PA & Lehner PJ (2007) Structure and Function: Heat Shock Proteins and Adaptive Immunity. *The Journal of Immunology* **179**: 2035-2040.

Joshi J, Mueller-Cajar O, Tsai YC, Hartl FU & Hayer-Hartl M (2015) Role of small subunit in mediating assembly of red-type form I Rubisco. *J Biol Chem* **290**: 1066-1074.

Kaushik S & Cuervo AM (2018) The coming of age of chaperone-mediated autophagy. *Nature Reviews Molecular Cell Biology* **19**: 365-381.

Kim YE, Hipp MS, Bracher A, Hayer-Hartl M & Hartl FU (2013) Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry* **82**: 323-355.

Kinney JN, Axen SD & Kerfeld CA (2011) Comparative analysis of carboxysome shell proteins. *Photosynthesis research* **109**: 21-32.

Klucken J, Shin Y, Masliah E, Hyman BT & McLean PJ (2004) Hsp70 Reduces α-Synuclein Aggregation and Toxicity. *Journal of Biological Chemistry* **279**: 25497-25502.

Klumpp M & Baumeister W (1998) The thermosome: archetype of group II chaperonins. *FEBS Letters* **430**: 73-77.

Kreuzer KN & Jongeneel CV (1983) [9] Escherichia coli phage T4 topoisomerase. *Methods in Enzymology,* Vol. 100 p.^pp. 144-160. Academic Press.

Kubien DS, Brown CM & Kane HJ (2011) Quantifying the amount and activity of Rubisco in leaves. *Methods in molecular biology (Clifton, NJ)* **684**: 349-362.

Kuchmina E, Wallner T, Kryazhov S, Zinchenko V & Wilde A (2012) An expression system for regulated protein production in Synechocystis sp. PCC 6803 and its application for construction of a conditional knockout of the ferrochelatase enzyme. *Journal of biotechnology* **162**: 75-80.

lange c, soppa j & griese m (2011) Ploidy in cyanobacteria. *FEMS Microbiology Letters* **323**: 124-131.

Langer T, Pfeifer G, Martin J, Baumeister W & Hartl FU (1992) Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *The EMBO Journal* **11**: 4757-4765.

Laufen T, Mayer MP, Beisel C, Klostermeier D, Mogk A, Reinstein J & Bukau B (1999) Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proceedings of the National Academy of Sciences* **96**: 5452-5457.

Leegood RC (2002) C4 photosynthesis: principles of CO2 concentration and prospects for its introduction into C3 plants. *Journal of experimental botany* **53**: 581-590.

Levinthal C (1968) Are there pathways for protein folding? *J Chim Phys* **65**: 44-45.

Li J & Buchner J (2013) Structure, function and regulation of the hsp90 machinery. *Biomedical journal* **36**: 106-117.

Lindquist S (1986) THE HEAT-SHOCK RESPONSE. *Annual Review of Biochemistry* **55**: 1151-1191.

Liu C, Young AL, Starling-Windhof A, *et al.* (2010) Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**: 197.

Liu D, Ramya RCS & Mueller-Cajar O (2017) Surveying the expanding prokaryotic Rubisco multiverse. *FEMS Microbiology Letters* **364**.

Long BM, Hee WY, Sharwood RE, *et al.* (2018) Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nature Communications* **9**: 3570.

Lopez T, Dalton K & Frydman J (2015) The Mechanism and Function of Group II Chaperonins. *Journal of Molecular Biology* **427**: 2919-2930.

Lu Z & Cyr DM (1998) Protein Folding Activity of Hsp70 Is Modified Differentially by the Hsp40 Co-chaperones Sis1 and Ydj1. *Journal of Biological Chemistry* **273**: 27824-27830.

Meyer M & Griffiths H (2013) Origins and diversity of eukaryotic CO2-concentrating mechanisms: lessons for the future. *Journal of experimental botany* **64**: 769-786.

Miernyk JA (1999) Protein folding in the plant cell. *Plant Physiol* **121**: 695-703.

Morán Luengo T, Mayer MP & Rüdiger SGD (2019) The Hsp70–Hsp90 Chaperone Cascade in Protein Folding. *Trends in Cell Biology* **29**: 164-177.

Morán Luengo T, Kityk R, Mayer MP & Rüdiger SGD (2018) Hsp90 Breaks the Deadlock of the Hsp70 Chaperone System. *Molecular Cell* **70**: 545-552.e549.

Mosser DD, Caron AW, Bourget L, Meriin AB, Sherman MY, Morimoto RI & Massie B (2000) The Chaperone Function of hsp70 Is Required for Protection against Stress-Induced Apoptosis. *Molecular and Cellular Biology* **20**: 7146-7159.

Mueller-Cajar O & Whitney SM (2008) Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynthesis Research* **98**: 667-675.

Mueller-Cajar O & Whitney SM (2008) Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynthesis research* **98**: 667-675.

Mueller-Cajar O & Whitney Spencer M (2008) Evolving improved Synechococcus Rubisco functional expression in Escherichia coli. *Biochemical Journal* **414**: 205-214.

Mueller-Cajar O, Stotz M & Bracher A (2014) Maintaining photosynthetic CO2 fixation via protein remodelling: the Rubisco activases. *Photosynthesis research* **119**: 191-201.

Mueller-Cajar O, Stotz M, Wendler P, Hartl FU, Bracher A & Hayer-Hartl M (2011) Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* **479**: 194-199.

Nelson DL (2004) Lehninger principles of biochemistry. *Oxidative phosphorylation and photophosphorylation*.

Nelson DL & Cox MM (2008) LEHNINGER PRINCIPLES OF BIOCHEMISTRY.

Occhialini A, Lin MT, Andralojc PJ, Hanson MR & Parry MA (2015) Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO. *The Plant journal : for cell and molecular biology*.

Ohashi K, Burkart V, Flohé S & Kolb H (2000) Cutting Edge: Heat Shock Protein 60 Is a Putative Endogenous Ligand of the Toll-Like Receptor-4 Complex. *The Journal of Immunology* **164**: 558-561.

Onizuka T, Endo S, Akiyama H, Kanai S, Hirano M, Yokota A, Tanaka S & Miyasaka H (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 & cell physiology* **45**: 1390-1395.

Ort DR, Merchant SS, Alric J, *et al.* (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences* **112**: 8529-8536.

Packschies L, Theyssen H, Buchberger A, Bukau B, Goody RS & Reinstein J (1997) GrpE accelerates nucleotide exchange of the molecular chaperone DnaK with an associative displacement mechanism. *Biochemistry* **36**: 3417-3422.

Parry MA, Keys AJ, Madgwick PJ, Carmo-Silva AE & Andralojc PJ (2008) Rubisco regulation: a role for inhibitors. *Journal of experimental botany* **59**: 1569-1580.

Parry MA, Andralojc PJ, Scales JC, Salvucci ME, Carmo-Silva AE, Alonso H & Whitney SM (2013) Rubisco activity and regulation as targets for crop improvement. *Journal of experimental botany* **64**: 717-730.

Parsell DA & Lindquist S (1993) THE FUNCTION OF HEAT-SHOCK PROTEINS IN STRESS TOLERANCE: DEGRADATION AND REACTIVATION OF DAMAGED PROTEINS. *Annual Review of Genetics* **27**: 437-496.

Pierce J, Carlson TJ & Williams JG (1989) A cyanobacterial mutant requiring the expression of ribulose bisphosphate carboxylase from a photosynthetic anaerobe. *Proceedings of the National Academy of Sciences* **86**: 5753-5757.

Portis AR, Jr. & Parry MA (2007) Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynthesis research* **94**: 121-143.
Pratt WB & Toft DO (2003) Regulation of Signaling Protein Function and Trafficking by the hsp90/hsp70-Based Chaperone Machinery. *Experimental Biology and Medicine* **228**: 111-133.

Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW & Pearl LH (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**: 65-75.

R. RK, N. S. N, S. P. A, Sinha D, Veedin Rajan VB, Esthaki VK & D'Silva P (2012) HSPIR: a manually annotated heat shock protein information resource. *Bioinformatics* **28**: 2853-2855.

Rae BD, Long BM, Badger MR & Price GD (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiology and Molecular Biology Reviews* **77**: 357-379.

Raines CA (2003) The Calvin cycle revisited. *Photosynthesis Research* **75**: 1-10.

Ranson NA, White HE & Saibil HR (1998) Chaperonins. *Biochemical Journal* **333**: 233.

Reinsvold RE, Jinkerson RE, Radakovits R, Posewitz MC & Basu C (2011) The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. *Journal of Plant Physiology* **168**: 848-852.

Roh SH, Kasembeli M, Bakthavatsalam D, Chiu W & Tweardy DJ (2015) Contribution of the Type II Chaperonin, TRiC/CCT, to Oncogenesis. *Int J Mol Sci* **16**: 26706-26720.
Roseman AM, Chen S, White H, Braig K & Saibil HR (1996) The Chaperonin ATPase Cycle: Mechanism of Allosteric Switching and Movements of Substrate-Binding Domains in GroEL. *Cell* **87**: 241-251.

Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU & Hayer-Hartl M (2007) Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**: 1189-1200.

Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU & Hayer-Hartl M (2007) Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco. *Cell* **129**: 1189-1200.

Satagopan S, Chan S, Perry LJ & Tabita FR (2014) Structure-function studies with the unique hexameric form II ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco) from Rhodopseudomonas palustris. *Journal of Biological Chemistry* **289**: 21433-21450.
Savir Y, Noor E, Milo R & Tlusty T (2010) Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proceedings of the National Academy of Sciences* **107**: 3475.

Schneider G, Lindqvist Y & Lundqvist T (1990) Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from Rhodospirillum rubrum at 1.7 Å resolution. *Journal of Molecular Biology* **211**: 989-1008.

Schneider G, Lindqvist Y & Branden CI (1992) RUBISCO: structure and mechanism. *Annual review of biophysics and biomolecular structure* **21**: 119-143.

Schneider G, Lindqvist Y, Brändén C-I & Lorimer G (1986) Three-dimensional structure of ribulose-1,5-bisphosphate carboxylase/oxygenase from Rhodospirillum rubrum at 2.9 Å resolution. *The EMBO Journal* **5**: 3409-3415.

Sharma D & Masison DC (2009) Hsp70 Structure, Function, Regulation and Influence on Yeast Prions. *Protein and Peptide Letters* **16**: 571-581.

Sievers F, Wilm A, Dineen D*, et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Vol. 7 p.^pp. 539.

Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA & Horwich AL (1998) Structure and function in GroEL-mediated protein folding. *Annu Rev Biochem* **67**: 581-608.

Spreitzer RJ (2003) Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of biochemistry and biophysics* **414**: 141-149.

Spreitzer RJ & Salvucci ME (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annual review of plant biology* **53**: 449-475.

Stotz M, Mueller-Cajar O, Ciniawsky S, Wendler P, Hartl FU, Bracher A & Hayer-Hartl M (2011) Structure of green-type Rubisco activase from tobacco. *Nature structural & molecular biology* **18**: 1366-1370.

Szabo A, Korszun R, Hartl FU & Flanagan J (1996) A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO J* **15**: 408-417.

Tabita FR, Hanson TE, Satagopan S, Witte BH & Kreel NE (2008) Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 2629-2640.

Tcherkez GG, Farquhar GD & Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences* **103**: 7246-7251.

Tsai Y-CC, Lapina MC, Bhushan S & Mueller-Cajar O (2015) Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat Commun* **6**.

Varaljay VA, Satagopan S, North JA, Witte B, Dourado MN, Anantharaman K, Arbing MA, McCann SH, Oremland RS & Banfield JF (2016) Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environmental microbiology*.

Viale AM & Arakaki AK (1994) The chaperone connection to the origins of the eukaryotic organelles. *FEBS letters* **341**: 146-151.

Walter S & Buchner J (2002) Molecular chaperones--cellular machines for protein folding. *Angewandte Chemie (International ed in English)* **41**: 1098-1113.

Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

Wegele H, Wandinger SK, Schmid AB, Reinstein J & Buchner J (2006) Substrate Transfer from the Chaperone Hsp70 to Hsp90. *Journal of Molecular Biology* **356**: 802-811.

Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D & Yeates TO (2014) Structure and identification of a pterin dehydratase-like protein as a RuBisCO assembly factor in the alpha-carboxysome. *Journal of Biological Chemistry*.

Whitney SM, Houtz RL & Alonso H (2011) Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco. *Plant Physiol* **155**: 27-35.

Whitney SM, von Caemmerer S, Hudson GS & Andrews TJ (1999) Directed Mutation of the Rubisco Large Subunit of Tobacco Influences Photorespiration and Growth. *Plant Physiology* **121**: 579.

Whitney SM, Baldet P, Hudson GS & Andrews TJ (2001) Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *The Plant Journal* **26**: 535-547.

Whitton BA & Potts M (2012) Introduction to the cyanobacteria. *Ecology of Cyanobacteria II*,p.^pp. 1-13. Springer.

Wilson RH, Alonso H & Whitney SM (2016) Evolving Methanococcoides burtonii archaeal Rubisco for improved photosynthesis and plant growth. *Scientific Reports* **6**: 22284.

Wunder T, Cheng SLH, Lai SK, Li HY & Mueller-Cajar O (2018) The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat Commun* **9**: 5076.

Xu H (2018) Cochaperones enable Hsp70 to use ATP energy to stabilize native proteins out of the folding equilibrium. *Scientific Reports* **8**: 13213.

Yang J, Yan R, Roy A, Xu D, Poisson J & Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12**: 7-8.

Zarzycki J, Axen SD, Kinney JN & Kerfeld CA (2013) Cyanobacterial-based approaches to improving photosynthesis in plants. *Journal of experimental botany* **64**: 787-798.

Zhang J, Baker ML, Schröder GF, *et al.* (2010) Mechanism of folding chamber closure in a group II chaperonin. *Nature* **463**: 379.

Zhang X, Carter MS, Vetting MW, *et al.* (2016) Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars. *Proceedings of the National Academy of Sciences* **113**: E4161.

Zhu X-G, Long SP & Ort DR (2010) Improving Photosynthetic Efficiency for Greater Yield. *Annual Review of Plant Biology* **61**: 235-261.

Ziska Lewis H, Bunce James A, Shimono H, *et al.* (2012) Food security and climate change: on the potential to adapt global crop production by active selection to rising atmospheric carbon dioxide. *Proceedings of the Royal Society B: Biological Sciences* **279**: 4097-4105.