

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**EVOLUTIONARY ESCAPE FROM THE RUBISCO ACTIVASE
REQUIREMENT**

GUO ZHIJUN

SCHOOL OF BIOLOGICAL SCIENCES

2019

EVOLUTIONARY ESCAPE FROM THE RUBISCO ACTIVASE REQUIREMENT

GUO ZHIJUN

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.

.....23/01/2019.....
Date


.....
Guo ZhiJun

SUPERVISOR DECLARATION STATEMENT

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.

.....23/01/2019.....
Date

..........
Associate Professor Oliver Mueller-Cajar

AUTHORSHIP ATTRIBUTION STATEMENT

AUTHORSHIP ATTRIBUTION STATEMENT

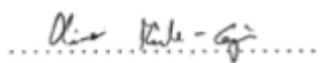
Thesis Title: EVOLUTIONARY ESCAPE FROM THE RUBISCO ACTIVASE REQUIREMENT

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



Guo ZhiJun

Date : 23/01/2019



Associate Professor Oliver Mueller-Cajar

Date: 23/01/2019

ACKNOWLEDGEMENTS

First of all, I would to express my sincere gratitude for my supervisor, Associate Professor Oliver Mueller-Cajar, for his guidance and patience. His passion for science and Rubisco has been inspiring. I would also like to thank the members of the OMC lab for their support and help throughout these four years. These awesome great people include, Candace Tsai, Tobias Wunder, Nitin Loganathan, Devendra Shivhare, Liu Di, Ramya, Jed Ng, How Jianan, Oh Zhen guo and Lynette Liew. I would express my gratitude towards my students, Ong Weilin, Kingston Tan and Jules Ng, who have provided technical support on the various aspect of this project.

I would like to thank Associate Professor Whitney Spencer for his assistance in performing the radioactive assay. I am very thankful and honor for the opportunity to learn from one of the best Rubisco scientists. I would also like Sally Buck, Amy Bing and Zhuo Yu for their help during my stay in Spencer's lab.

I am also very grateful to Asst. Prof Xavier Roca and his wonderful lab members for their generosity in sharing their lab equipment and bottles for my screening experiments. I would express my thanks for the beautiful ladies of this lab; Liu Jun, Liu Hongfei, Tan SzeJing, Malini Bhadra, Jessie Ho and Liu Shufang. Their support and friendship have been one of the most valuable things I gain in this PhD journal.

Lastly, I would like to express my thanks to my family, friends and beloved Boo Zhaozhi for their support and love through the years.

#Rubiscoislove.

TABLE OF CONTENTS

STATEMENT OF ORIGINALITY	2
SUPERVISOR DECLARATION STATEMENT	3
AUTHORSHIP ATTRIBUTION STATEMENT.....	4
ACKNOWLEDGEMENTS.....	5
LIST OF FIGURES	14
LIST OF TABLES.....	18
ABBREVIATIONS	20
ABSTRACT	22
1 INTRODUCTION	24
1.1 Calvin-Benson-Bassham Cycle	24
1.2 Photorespiration and carbon concentrating mechanisms.....	25
1.3 Structure of Rubisco	27
1.3.1 The different forms of Rubisco.....	27
1.3.2 Structure of Rubisco large and small subunit	30
1.4 Reaction mechanism of Rubisco.....	33
1.4.1 Activation of Rubisco	33
1.4.2 Conformational changes during catalysis	33
1.4.3 The carboxylation reaction	34
1.4.4 Rubisco's side reactions.....	34
1.4.5 Sugar phosphate inhibitors.....	36
1.5 Diverse kinetic properties of Rubisco.....	37
1.6 Maintenance of Rubisco's activity.....	41
1.6.1 Rubisco activase.....	41
1.6.1.1 The Red-type Rubisco activase remodels red-type Form I Rubisco.	41
1.6.1.2 Green-type Rubisco activase acts on green-type form IB Rubisco ..	42
1.6.1.3 CbbQO Rubisco activase is associated with green-type form IA	
Rubisco	44
1.7 Biogenesis and assembly of form I Rubisco.....	46
1.8 Directed evolution of Rubisco	47

1.8.1	Directed evolution of Rubisco using Rubisco deficient <i>R. capsulatus</i> and <i>R. eutropha</i> strains	48
1.8.2	Directed evolution of Rubisco using Rubisco-dependent <i>E. coli</i>	50
1.8.2.1	Directed evolution of Form II Rubisco using RDE	51
1.8.2.2	Directed evolution of Form IBC cyanobacteria Rubisco using RDE	51
1.8.2.3	Directed evolution of Form III Rubisco using RDE	52
1.9	Aim and scope of the study	53
2	MATERIAL AND METHODS	55
2.1	Bacteriological Methods	55
2.1.1	Preparation of Chemical Competent <i>E. coli</i> cells	55
2.1.1.1	Chemical Transformation	56
2.1.2	Preparation of Electrocompetent MM1 cells	56
2.1.2.1	Electroporation.....	57
2.2	Molecular Biology	57
2.2.1	Polymerase Chain Reaction (PCR).....	57
2.2.2	pGemT Vector Cloning.....	57
2.2.3	Site-directed Mutagenesis	60
2.2.4	Restriction digestion and ligation	60
2.2.5	DNA Precipitation	61
2.3	Direct Evolution Methodology	61
2.3.1	pTrcHisB Plasmids	61
2.3.2	pTrcDuet Plasmid	62
2.3.3	Reconstruction of PCC7942PRK expression plasmid.....	64
2.3.4	Construction of pAC ^{BAD} PRK ⁶⁸⁰³ and pAC ^{BAD} PRK ⁷⁰⁰² vectors.....	65
2.3.5	Establishment of RDE and RADE system.....	66
2.3.6	Mutant Library construction	67
2.3.7	Gene design of <i>A. ferrooxidans</i> Rubisco and activase.....	68
2.3.8	Selection in RDE and RADE	69
2.4	Chimeric Rubisco Methodology	70
2.4.1	Gene Design.....	70

2.4.1.1	Rubisco Large subunit Gene Design.....	70
2.4.1.2	Rubisco Small Subunit Gene Design.....	70
2.4.2	Cloning of Synthesized Genes into pTrcHisB	71
2.4.3	Chimeric Rubisco.....	71
2.4.3.1	Restriction-free cloning of chimeric Rubisco	71
2.5	Protein Biochemistry	72
2.5.1	Protein expression test and SDS-PAGE	72
2.5.2	Western blot and Immunodetection	74
2.5.3	Native-PAGE	74
2.5.4	Rubisco purification.....	75
2.5.5	Activase purification.....	76
2.5.6	Protein concentration determination by absorbance at 280 nm	77
2.5.7	Protein concentration determination by Bradford assay.....	77
2.6	Enzymatic assay.....	77
2.6.1	RuBP and CABP synthesis	78
2.6.2	Spectrophotometric Rubisco assay	78
2.6.2.1	Kinetics of RuBP-inhibited Rubisco complexes.....	79
2.6.3	Radioisotopic Rubisco activity assay.....	80
2.6.3.1	Rubisco activation by CO ₂ -Mg ²⁺	80
2.6.3.2	Rubisco activity measurement using ¹⁴ CO ₂ incorporation	80
2.6.3.3	Quantification of Rubisco content using ¹⁴ CABP binding	82
2.6.3.4	Determination of <i>K_m(RuBP)</i>	82
2.6.3.5	Rubisco CO ₂ /O ₂ specificity (<i>S_{C/O}</i>).....	83
2.6.3.6	Statistical Analysis.....	84
3	Directing the evolution of Rubisco in the presence and absence of its activase in MM1 <i>E. coli</i>	85
3.1	Three different PRK enzymes were used in establishing RDE and RADE selection system	89
3.1.1	Reconstruction of the PRK ⁷⁹⁴² containing plasmid	90

3.1.2	Cloning of PRK ⁶⁸⁰³ and PKR ⁷⁰⁰² into the pAC ^{BAD} vector	91
3.2	Protein expression of Rubisco and activase in pTrcHisB and pTrcDuet vector 91	
3.2.1	Soluble protein expression of <i>Rs</i> -Rubisco and activase in pTrcHisB and pTrcDuet 92	
3.2.2	Soluble expression of <i>Af</i> -Rubisco and activase in pTrcHisB and pTrcDuet 94	
3.3	Development of Rubisco and activase dependent <i>Escherichia coli</i> system .97	
3.3.1	pAC ^{BAD} PRK ^{DM} had limited success in the development of RDE and RADE for the different Rubisco and activase.....	97
3.3.1.1	Presence of functional and compatible pair of Rubisco and activase are required for survival in RADE selection.....	99
3.3.1.2	RADE and RDE screening condition of <i>Af</i> -Rubisco and activase were not achieved using MM1-pAC ^{BAD} PRK ^{DM}	102
3.3.1.3	MM1-pAC ^{BAD} PRK ^{DM} proved ineffective in the screening of mutant Rubisco libraries	104
3.4	A high throughput selection screen for Rubisco in the presence and absence of activase using MM1-pAC ^{BAD} PRK ⁶⁸⁰³	104
3.4.1	RDE and RADE screening using MM1-pAC ^{BAD} PRK ⁶⁸⁰³	105
3.4.2	Repeatedly mutated residues were obtained from RDE and RADE screening 110	
3.5	Discussion	112
4	Directed evolution of activase-requiring Rubisco reveals facile escape from sugar phosphate inhibition.....	114
4.1	Directed evolution of Rubisco using MM1-pAC ^{BAD} PRK ⁶⁸⁰³ resulted in Rubisco variants with repeatedly mutated residues	114
4.1.1	Directed evolution of <i>Rs</i> -Rubisco using the RDE system	117
4.1.2	Directed evolution of <i>Af</i> -Rubisco using RDE system.....	119
4.1.3	Directed evolution of <i>Rs</i> -Rubisco in the presence of its activase using the RADE system.....	121

4.2	Preliminary biochemical characterization of Rubisco variants using soluble <i>E. coli</i> extract and spectrophotometric assay	123
4.2.1	Characterization of <i>R_s</i> -Rubisco variants using soluble <i>E. coli</i> extract	125
4.2.2	Characterizing <i>A_f</i> -Rubisco variants using soluble <i>E. coli</i> extract.....	129
4.3	Purification of selected <i>R_s</i> -Rubisco and <i>A_f</i> -Rubisco variants.....	132
4.4	Characterization of selected purified <i>R_s</i> -Rubisco and <i>A_f</i> -Rubisco variants	134
4.4.1	Characterization of purified <i>R_s</i> -Rubisco variants	136
4.4.2	Characterization of mutant <i>A. ferrooxidans</i> Rubisco using purified Rubisco	140
4.5	Summary of characterizing Rubisco variants using spectrophotometric assay	144
5	Different molecular adaption of Red- and Green- type rubisco.....	146
5.1	Soluble expression levels of selected variants are variable	146
5.2	<i>R_s</i> -Rubisco and <i>A_f</i> -Rubisco variants showed diversity in their kinetic properties.....	152
5.2.1	Rubisco variants had diverse carboxylation turnover rate.....	155
5.2.2	Majority of Rubisco variants had reduced affinity towards CO ₂	157
5.2.3	Most selected Rubisco variants exhibited a reduced carboxylation efficiency	158
5.2.4	The CO ₂ /O ₂ specificity of <i>R_s</i> -Rubisco and <i>A_f</i> -Rubisco variants were similar to wild type	158
5.2.5	Determination of <i>K_m</i> (<i>RuBP</i>) for variant Rubisco.....	159
5.3	Evolved Rubisco variants target different regions of Rubisco	162
5.3.1	Conformational changes during transition between open-to-closed state in Rubisco	165
5.3.1.1	Ordering of two loops from the N-terminal domain during active site closure	167
5.3.1.2	Loop 6 functions as a lid that closes over the active site.....	169
5.3.1.3	The C-terminus folds over loop 6 during active site closing.....	170

5.3.2	Mutations found in the 60s loop caused increased inhibitor release rates and displayed trade-offs in kinetic properties	173
5.3.2.1	Pro50 substitution may influence the position of Thr65 in the active site	175
5.3.2.2	Ser61 substitutions negatively influence catalysis via Glu60.....	176
5.3.2.3	Tyrosine 24 mutation increased the flexibility of the 60's loop and N-terminus	177
5.3.2.4	Thr71, Thr75 and Arg81 interact with the 3 ₁₀ helix residues of the 60s loop	179
5.3.2.4.1	The effect of Thr71 substitutions is likely mediated through the 60s loop and loop 8.....	180
5.3.2.4.2	T75A variant had the lowest affinity towards CO ₂	181
5.3.2.4.3	Arg81 substitutions may influence positions of N-terminus loop and 60s loop	182
5.3.3	Mutations targeting the C-terminus and C-terminal α -helical subdomain resulted in increased rate of inhibitor release	184
5.3.3.1	Substitutions at K463 and I465 of the C-terminal strand may influence inhibitor release.....	185
5.3.3.1.1	Substitution at Lys463 alone may not be sufficient to allow rapid inhibitor release.....	186
5.3.3.1.2	A single substitution at Ile465 can increase the rate of inhibitor release	187
5.3.3.2	K41 and P305 are a pair of interacting residues located near the latch site of the C-terminus.....	190
5.3.3.2.1	Substitutions of Lys41 may affect active site closure via C-terminal tail	191
5.3.3.2.2	Pro305 substitution negatively influence C-terminal tail and loop 6 placement	193
5.3.3.3	Thr392 and Glu438 are potentially a pair of interacting residues ..	199
5.3.3.3.1	Thr392 substitution could affect inhibitor binding via loop 7 ..	199

5.3.3.3.2	Glu438 substitution may affect the position of loop 7 and loop 8	201
5.3.4	Mutations affecting the dimer-dimer interface	203
5.3.4.1	Substitution at residue 105 may influence inhibitor release indirectly through the β C- α C loop residues	205
5.3.4.2	The effect for F220L substitution might be mediated through long distance to the active site	206
5.3.5	Small subunit mutation clustering around β -strand D affects the rate of inhibitor release.....	210
5.3.5.1	Substitutions of Phe91 and Thr95 in <i>R</i> s-Rubisco may influence rate of inhibitor release via the 60s loop.....	212
5.3.5.2	V112 substitution may influence β C- β D loop's position.....	213
5.4	Discussion	215
5.4.1	Loops involved in substrate binding and active site closure were influenced in Rubisco variants.....	216
5.4.2	Why did nature evolve Rubiscos to form a tightly inhibited complex?	218
5.5	Conclusions.....	220
6	Directed evolution of Chimeric Plant/cyanobacterial Rubisco in <i>E. coli</i> ...	222
6.1	Chimeric <i>Syn</i> / <i>Os</i> -Rubisco	223
6.1.1	Synthesizing <i>Syn</i> -Rubisco and <i>Os</i> -Rubisco gene	223
6.1.2	<i>Syn</i> -Rubisco and <i>Os</i> -Rubisco presented with different solubilities in <i>E. coli</i>	225
6.1.3	Chimeric Rubisco had different stability in their holoenzyme complex	227
6.1.3.1	Varying solubility of the chimeric Rubisco with single section replaced with <i>Os</i> -Rubisco	228
6.1.4	Chimeric Rubisco consisting of more than one fragment of <i>Os</i> -Rubisco had different solubility in <i>E. coli</i>	233
6.1.5	Assembly of soluble chimeric Rubisco were different.....	234

6.2	Biochemical characterization of purified chimeric Rubisco.....	237
6.2.1	Soluble chimeric Rubiscos show diversity in their Rubisco activity..	237
6.2.1.1	SynL(OsLA)S exhibits characteristic similar to plant Rubisco.....	238
6.3	Directed evolution of chimeric Rubisco	241
6.3.1	Small numbers of variants were obtained from directed evolution of chimeric Rubisco	242
6.3.2	SO-CA1 and SO-CA2 may have evolved away from plant Rubisco..	243
6.4	Discussion	247
6.4.1	Higher plant Rubisco's insolubility in <i>E. coli</i> may be caused by a small number of residues	247
6.4.2	Rubisco can be easily evolved to have a higher rate of inhibitor release 248	
6.4.3	Future prospects of directing the evolution of Rubisco	249
7	Summary and Conclusion	251
8	Appendix.....	256
8.1	Appendix I: List form I Rubisco from the phylogenetic tree.....	256
8.2	Appendix II: Combination of selective media for optimizing RADE screening	257
8.3	Appendix III: Reoccurring mutations between <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco variants	263
8.4	Appendix IV: List of publication.....	265
9	References.....	266

LIST OF FIGURES

Figure 1-1: The Calvin-Benson-Bassham Cycle (simplified).	25
Figure 1-2: Structures of the four different forms of Rubisco.	29
Figure 1-3: Phylogenetic tree of form I Rubisco large subunit sequences.	30
Figure 1-4: Structure of Rubisco large and small subunits.	32
Figure 1-5: Rubisco's reaction mechanism and inhibitors.	35
Figure 1-6: Regulation of Rubisco's activation state.	37
Figure 1-7: Structures of the different Rubisco activase classes.	45
Figure 1-8: Screening of Rubisco using the Rubisco-dependent <i>Escherichia coli</i> system	53
Figure 2-1: Representative vector map for pTrcHisB and pTrcDuet.	64
Figure 2-2: Gene design of <i>A. ferrooxidans</i> Rubisco and activase.	69
Figure 2-3: Spectrophotometric Rubisco assay.	79
Figure 3-1: Overview of MM1 <i>E. coli</i> screening system and compatibility between different Rubisco and activase.	88
Figure 3-2: Sequence alignment of the PRK from three different cyanobacteria, <i>Synechococcus</i> PCC 7942/6301, <i>Synechocystis</i> PCC 6803 and <i>Synechococcus</i> PCC 7002.	90
Figure 3-3: Overview of small-scale protein expression test for Rubisco and/or activase in pTrcHisB and pTrcDuet.	92
Figure 3-4: Soluble expression of <i>Rs</i> -Rubisco and its activase in pTrcHisB and pTrcDuet.	93
Figure 3-5: Protein expression of form I <i>Af</i> -Rubisco (AfLS) and its activase (AfQ1O1) in pTrcHisB and pTrcDuet vectors.	95
Figure 3-6: Protein expression of form II <i>Af</i> -Rubisco (AfM) and its activase (AfQ2O2) in pTrcHisB and pTrcDuet vector.	96
Figure 3-7: Representation of Rubisco-dependent <i>E. coli</i> selection using MM1- pAC ^{BAD} PRK ^{DM}	98

Figure 3-8: Representation of Rubisco and activase dependent <i>E. coli</i> selection using MM1-pAC ^{BAD} PRK ^{DM}	100
Figure 3-9: Screening of <i>Af</i> -Rubisco and activase in RADE selective media using MM1-pAC ^{BAD} PRK ^{DM} cells.	103
Figure 3-10: Representative plate images for RDE screening using MM1-pAC ^{BAD} PRK ⁶⁸⁰³	108
Figure 3-11: Screening of Rubisco and/or activase under RADE system using MM1-pAC ^{BAD} PRK ⁶⁸⁰³	109
Figure 4-1: Mutations in <i>Rs</i> -Rubisco obtained from RDE and RADE screening.	115
Figure 4-2: Mutations in <i>Af</i> -Rubisco obtained from RDE screening.	116
Figure 4-3: Position of repeatedly mutated residues of <i>Rs</i> - and <i>Af</i> -Rubisco variants.	117
Figure 4-4: Representation of Rubisco activation assay using soluble <i>E. coli</i> extract.	123
Figure 4-5: Representative SDS-PAGE analysis of soluble <i>E. coli</i> extract from <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco variants.	124
Figure 4-6: Rubisco activity of fully activated Rubisco complexes (ECM) of <i>Rs</i> -Rubisco variants assayed using soluble <i>E. coli</i> extract.	127
Figure 4-7: Time course of spontaneous activation of RuBP-inhibited Rubisco complexes (ER) of <i>Rs</i> -Rubisco variants assayed with soluble <i>E. coli</i> extract....	128
Figure 4-8: Rubisco activity of fully activated ECM complex of <i>Af</i> -Rubisco variants assayed using soluble <i>E. coli</i> extract.....	130
Figure 4-9: Time course of spontaneous activation of RuBP-inhibited Rubisco complexes (ER) <i>Af</i> -Rubisco variants assayed using soluble <i>E. coli</i> extract.	131
Figure 4-10: Representative purification of <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco.....	133
Figure 4-11: Kinetics of slow-binding inhibitors of Rubisco.	135
Figure 4-12: Rubisco activity of ECM and ER complexes assayed with purified <i>Rs</i> -Rubisco variants.....	139
Figure 4-13: Rate of RuBP-inhibitor (k_{obs}) release for <i>Rs</i> -Rubisco variant.	140

Figure 4-14: Rubisco activity ECM complex and time course of spontaneous activation of ER complex of <i>Af</i> -Rubisco variants assayed using purified protein.	143
Figure 4-15: Rate of inhibitor release for <i>Af</i> -Rubisco variants.....	144
Figure 5-1: The expression level of <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco as a percentage of soluble <i>E. coli</i> protein using <i>E. coli</i> XL1-Blue cells.....	149
Figure 5-2: Representative gel of <i>Rs</i> -Rubisco variants' expression in <i>E. coli</i> XL1-Blue	150
Figure 5-3: Representative gel of <i>Af</i> -Rubisco variants' expression in <i>E. coli</i> XL1-Blue.	151
Figure 5-4: The carboxylation turnover rate (k_{cat}^c) and Michaelis constant for CO ₂ (K_C) of <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco variants.	156
Figure 5-5: The CO ₂ /O ₂ specificity ($S_{C/O}$) of <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco variants. ...	159
Figure 5-6: The Michaelis constant for RuBP, $K_m(RuBP)$, of <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco variants.....	161
Figure 5-7: Sequence alignments of selected green-type and red-type Rubisco large subunits.	163
Figure 5-8: Substrate-binding sites and conformational changes during the transition between open and closed state in form I Rubisco.....	172
Figure 5-9: Sequence conservation of residues from two selected regions of the N-terminal domain of Rubisco large subunit.	174
Figure 5-10: Position and interaction of selected residues residing or interacting with the 60s loop.	178
Figure 5-11: Interaction of selected residues with 3 ₁₀ helix residues of the 60s loop.	183
Figure 5-12: Sequence conservation of the C-terminal residues.	186
Figure 5-13: Substitution of C-terminal residues increased the rate of inhibitor release.	189
Figure 5-14: Sequence conservation of residues at position 41 and 305 of Rubisco large subunit.....	191
Figure 5-15: Position and interactions of Lys41 and Pro305 in <i>Hn</i> -Rubisco.....	196

Figure 5-16: Position and interactions of Arg41 and Lys305 in Spinach Rubisco.....	198
Figure 5-17: Structural analysis of residues at position 392 and 438 of the C-terminal domain.....	202
Figure 5-18: The conservation and effect of substituting residues from the dimer-dimer interface.....	204
Figure 5-19: Interaction and position of selected residues at the large subunit dimer-dimer interface.	209
Figure 5-20: Sequence alignment of selected Rubisco small subunits.....	211
Figure 5-21: Position and interaction of small subunit residues.....	214
Figure 6-1: Sequence alignment and synthetic constructs encoding <i>Synechococcus</i> PCC6301 and <i>Oryza sativa</i> Rubisco.....	224
Figure 6-2: Protein expression and western blot of pTrcSynL ^{F345I} S and pTrcOsL ^{F345I} S in <i>E. coli</i>	226
Figure 6-3: Representation of the different chimeric Rubiscos generated.	228
Figure 6-4: Western blot analysis of the expression of chimeric Rubisco with single region substitution with Os-Rubisco with and without the supplement of GroEL/ES in <i>E. coli</i>	231
Figure 6-5: Native-PAGE and western blot of induced soluble extracts from chimeric Rubisco in the presence and absence of GroEL/ES.....	232
Figure 6-6: Protein expression of chimeric Rubisco, SynL(OsL ^{AE})OsS in presence and absence of GroEL/ES.....	234
Figure 6-7: Summary of soluble chimeric Rubisco.	236
Figure 6-8: Rubisco activation assay of chimeric Rubisco.....	240
Figure 6-9: Representative gel of the purified chimeric SynL(OsL ^A)S and its evolved Rubisco variant.	243
Figure 6-10: Rubisco activation assay of the evolved SO-CA1 and SO-CA2 Rubisco.	245
Figure 6-11: Structural analysis of the mutated residues in SO-CA2.....	247

LIST OF TABLES

Table 1-1: Kinetic properties of the Rubisco from various species.....	40
Table 2-1: Composition of media and buffer used for the preparation of competent cells.	55
Table 2-2 Primers used for Site-directed mutagenesis.	59
Table 2-3: List of pTrcHisB and pTrcDuet plasmids used in this study.	63
Table 2-4: Components for optimization of RADE selective media.	67
Table 2-5: Buffer for protein biochemistry.....	75
Table 3-1: Relative fitness of MM1-pAC ^{BAD} PRK ^{DM} expressing Rubisco and/or activase in the RADE system.....	102
Table 3-2: Relative fitness of Rubisco and activase in MM1 cells expressing either PRK ⁷⁰⁰² or PRK ⁶⁸⁰³	105
Table 3-3: Relative fitness of Rubisco and activase in MM1-pAC ^{BAD} PRK ⁶⁸⁰³ cells screened on selective media containing (RDE) and lacking (RADE) 0.5 mM IPTG.	107
Table 3-4: Mutant libraries of Rubisco in pTrcHisB and pTrcDuet and the screening conditions for each library.	110
Table 3-5: Summary of mutations from various mutant library screened.	111
Table 4-1: <i>Rs</i> -Rubisco variants with repeatedly mutated residues obtained from RDE screening.	118
Table 4-2: <i>Rs</i> -Rubisco variants with unique mutation(s) obtained from RDE screening.	119
Table 4-3: <i>Af</i> -Rubisco variants with repeatedly mutated residues obtained from RDE screening.	120
Table 4-4: <i>Af</i> -Rubisco variants containing unique mutation(s) from RDE screening.	121
Table 4-5: <i>Rs</i> -Rubisco variants with repeatedly mutated residues obtained from RADE screening.	122

Table 4-6: <i>Rs</i> -Rubisco variants with unique mutations obtained from RADE screening.	122
Table 4-7: Summary of kinetic parameters purified <i>Rs</i> -Rubisco variants assayed using spectrophotometric assay	138
Table 4-8: Summary of kinetic properties of <i>Af</i> -Rubisco variants assayed using purified Rubisco.	142
Table 5-1: Summary of the kinetic properties and expression level of <i>Rs</i> -Rubisco variants.	153
Table 5-2: Summary of the kinetic properties and expression level of <i>Af</i> -Rubisco variant.	154
Table 5-3: Interacting region of selected Rubisco variants with the phenotype of increased k_{obs}	164
Table 5-4: Role and location of active site residues from N-terminal and C-terminal domain.....	167
Table 6-1: The chimeric Rubisco libraries generated and the screening conditions. .	241
Table 6-2: Chimeric Rubisco variants obtained from screening.	242
Table 6-3: Summary of the kinetic properties of SynL(OsL ^A)S and its evolved variants.	246

ABBREVIATIONS

2PG	2-Phosphoglycolate
3PGA	3-phosphoglycerate
AAA+	ATPase associated with various cellular activities
<i>Af</i> -Rubisco	<i>Acidithiobacillus ferrooxidans</i> Rubisco
Amp	Ampicilin
ATP	Adenosine Triphosphate
AtRca	Agave tequiliana Rubisco activase
Bsd2	Bundle Sheath Defective 2
CA1P	2-Carboxy-arabinitol 1-phosphate
CABP	2- carboxyarabinitol-1,5-bisphosphate
CAM	Crassulacean acid metabolism
CBB	Calvin-Benson-Bassham
CbbX	Red-type Rubisco activase
CCM	Carbon-concentrating mechanism
Chl	chloramphenicol
CO ₂	Carbon dioxide
CTBP	2-carboxytetritol-1,4-bisphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
ECL	Enhanced chemiluminescence
ECM	Activated Rubisco complex
ECMI	Inhibited Rubisco complex (carbamylated)
EI	Inhibited Rubisco complex (non-carbamylated)
ER	RuBP-inhibited Rubisco complex
EX	XuBP-inhibited Rubisco complex
G3P	Glyceraldehyde-3-phosphate
<i>gapA</i>	Glyceraldehyde 3-phosphate dehydrogenase gene
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Hsp70	Heatshock Protein 70
IPTG	Isopropyl- beta-D-thiogalactopyranoside
Kan	kanamycin
K_C	Michaelis constant for CO ₂
k_{cat}	maximum carboxylation rate
k_{obs}	Rate of inhibitor release
$K_m(RuBP)$	Michaelis constant for RuBP
L ₂	Rubisco large subunit dimer
L ₈	Rubisco large subunit octomer core
L ₈ S ₈	Form I Rubisco stoichiometry, 8 large and 8 small subunits
Mg ²⁺	Magnesium ion
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NtRca	Nicotiana tabacum Rubisco activase

<i>O. sativa</i>	<i>Oryza sativa</i>
OD ₆₀₀	Optical density at 600 nm
<i>Os-Rubisco</i>	<i>Oryza sativa</i> Rubisco
OsRbcLS	<i>Oryza sativa</i> Rubisco
PCR	Polymerase chain reaction
PDBP	D-glycero-2,3-pentodiulose-1,5-bisphosphate
PMSF	Phenylmethylsulfonylfluoride
PRK	Phosphoribulokinase
PRK ⁶⁸⁰³	<i>Synechocystis</i> PCC 6803 phosphoribulokinase
PRK ⁷⁰⁰²	<i>Synechococcus</i> PCC 7002 phosphoribulokinase
<i>prkA</i>	Phosphoribulokinase gene
pTrcRsLSRsX	pTrcDuet containing wild type <i>Rhodobacter sphaeroides</i> 2.4.1 Rubisco and activase
<i>R. rubrum</i>	<i>Rhodospirillum rubrum</i>
<i>R. sphaeroides</i>	<i>Rhodobacter sphaeroides</i>
RADE	Rubisco and activase dependent <i>E. coli</i>
Raf1	Rubisco accumulation factors-1
Raf2	Rubisco accumulation factors-2
<i>rbcL</i>	Rubisco large subunit gene
RbcL	Rubisco large subunit
<i>rbcS</i>	Rubisco small subunit gene
RbcS	Rubisco small subunit
Rca	Rubisco activase
RDE	Rubisco-dependent <i>E. coli</i>
RLPs	Rubisco-like proteins
<i>Rs-Rubisco</i>	<i>Rhodobacter sphaeroides</i> 2.4.1 Rubisco
RsCbbLS, RsLS	<i>Rhodobacter sphaeroides</i> 2.4.1 Rubisco
RsCbbX	<i>Rhodobacter sphaeroides</i> 2.4.1 Rubisco activase
RsRca	<i>Rhodobacter sphaeroides</i> 2.4.1 Rubisco activase
RT	Room temperature
Ru5P	Ribulose 5-phosphate
Rubisco	Ribulose 1,5- bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1, 5-bisphosphate
S _{C/O}	CO ₂ /O ₂ specificity
Syn-Rubisco	<i>Synechococcus</i> PCC 6301 Rubisco
SynLS	<i>Synechococcus</i> PCC 6301 Rubisco
TIM-barrel	triose isomerase-like α - β barrel protein fold
XuBP	xyulose-1, 5-bisphosphate
Δ	Deletion

ABSTRACT

Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the rate-limiting carbon fixation step in photosynthesis. Despite its essential role, Rubisco is inefficient and possesses sluggish kinetics. In addition, Rubisco from many species (including plants) are prone to inhibition by sugar phosphates and require assistance from Rubisco activase to facilitate inhibitor release. Thus, Rubisco has been a target for improvement using directed evolution. Here, we performed directed evolution of Rubisco from *Rhodobacter sphaeroides* and *Acidithiobacillus ferrooxidans* using a **R**ubisco-**d**ependent *E**sch*erichia *coli* system (RDE). Similar to plant Rubisco, *R. sphaeroides* and *A. ferrooxidans* Rubiscos form inhibited complexes with RuBP and require an activase. In addition, *Rs*-Rubisco was also evolved in presence of its activase using **R**ubisco and **a**ctivase **d**ependent *E. coli* (RADE) system where the presence of functional and compatible Rubisco and activase pair are required for survival. Rubisco variants with reoccurring mutated residues were obtained from screening and 22 of these were biochemically characterized using radioactive and spectrophotometric Rubisco assays. Many variants isolated during RDE screening no longer form stable inhibited complexes with RuBP and exhibited impaired kinetics. However, the CO₂/O₂ specificity of these variants was largely unchanged compared to the wild type. In contrast, a Rubisco variant obtained from RADE screening still formed stable inhibited complexes while having a modest reduction in kinetics. Structural analysis of the variants with increased rates of inhibitor release revealed that common mutations were clustered around elements involved in substrate binding (60s loop) and active site closure (C-terminus). The observed mutations likely altered the local conformation such that the active site becomes less rigid and this facilitated rapid inhibitor release. Simultaneously, plant/cyanobacteria (*Oryza sativa*/ *Synechococcus*) chimeric Rubiscos were successfully produced in *E. coli*. Unlike plant Rubisco, cyanobacterial Rubisco does not form inhibited complexes with sugar phosphates and the organisms do not generally encode an activase. One chimeric Rubisco exhibited a catalytic fallover effect and was also able to form inhibited complexes with RuBP. In addition, the inhibited chimeric Rubisco could be recognized and remodeled by plant activase. The directed evolution of this plant-like chimeric Rubisco in the RDE screening system yielded variants that no longer form stable inhibited complexes. Therefore, relief from the activase dependency can easily be discovered using directed evolution methodologies. Using a

combination of rational and evolutionary methods to modify Rubisco, we obtained insights into the underlying principles governing the shortcomings of the enzyme and its protein-protein interactions with accessory proteins.

1 INTRODUCTION

As the world population is estimated to increase by another two billion by 2050, there is a need to ensure global food security (Ziska et al., 2012). Crop productivity has to increase to meet these global food demands as well as for other purposes such as biofuel production and animal feed (Edgerton, 2009). The uncertainties in global climate change and the decrease in arable land pose additional challenges for improving crop productivity (Whitney et al., 2011). Strategies to improve crop yield are targeted at supercharging photosynthesis, as the cumulative rate of photosynthesis is the primary determinant of crop yield (Ainsworth et al., 2005; Whitney et al., 2011). Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco; EC4.1.1.39) is a major target in attempts to improve crop yield, as its low catalytic turnover rate ($3\text{--}5\text{ s}^{-1}$ in plants) is often the limiting step in photosynthesis (Parry et al., 2008). Engineering Rubisco with improved catalytic efficiency will contribute towards improving crop productivity by improving water and nitrogen usage.

1.1 Calvin-Benson-Bassham Cycle

The Calvin-Benson-Bassham (CBB) Cycle is an evolutionarily conserved and major pathway for inorganic carbon entry biosphere entry into the biosphere (Tabita, 1988). It is known as the dark reactions in photosynthesis and Rubisco is the enzyme responsible for catalyzing CO_2 fixation (Raines, 2003). Rubisco converts approximately 10^{11} tons of CO_2 into organic biomass per annum (Field et al., 1998). Therefore, almost all life on earth is either directly or indirectly dependent on the CBB cycle and Rubisco for their survival.

The CBB cycle is utilized by both photosynthetic and chemoautotrophic organism for CO_2 fixation. Photosynthetic organisms are dependent on the light reactions of photosynthesis to produce nicotinamide adenine dinucleotide phosphate (NADPH) and adenosine triphosphate (ATP). Whereas, chemoautotrophic organisms generate ATP and nicotinamide adenine dinucleotide (NADH) from inorganic oxidation compounds. Overall, these molecules are utilized in the CBB cycle for carbon fixation and production of carbohydrates. In brief, the CBB cycle proceeds in three phases, a carbon fixation phase, a reduction phase, and a regeneration phase (Figure 1-1). In the carbon fixation phase, Rubisco incorporates atmospheric CO_2 in the pentose sugar, Ribulose-1,

5-bisphosphate (RuBP), to produce two molecules of a 3-carbon intermediate 3-phosphoglycerate (3PGA). The 3PGA is then further reduced through a series of reactions in the second phase to yield glyceraldehyde-3-phosphate (G3P). G3P would be used for the synthesis of other carbohydrates, amino acid and fatty acids. In the regeneration phase, a series of carbon rearrangement are performed to generate ribulose 5-phosphate (Ru5P) from G3P. Phosphorylation of Ru5P by phosphoribulokinase (PRK) will regenerate RuBP and it can then be utilized for another cycle of carbon fixation.

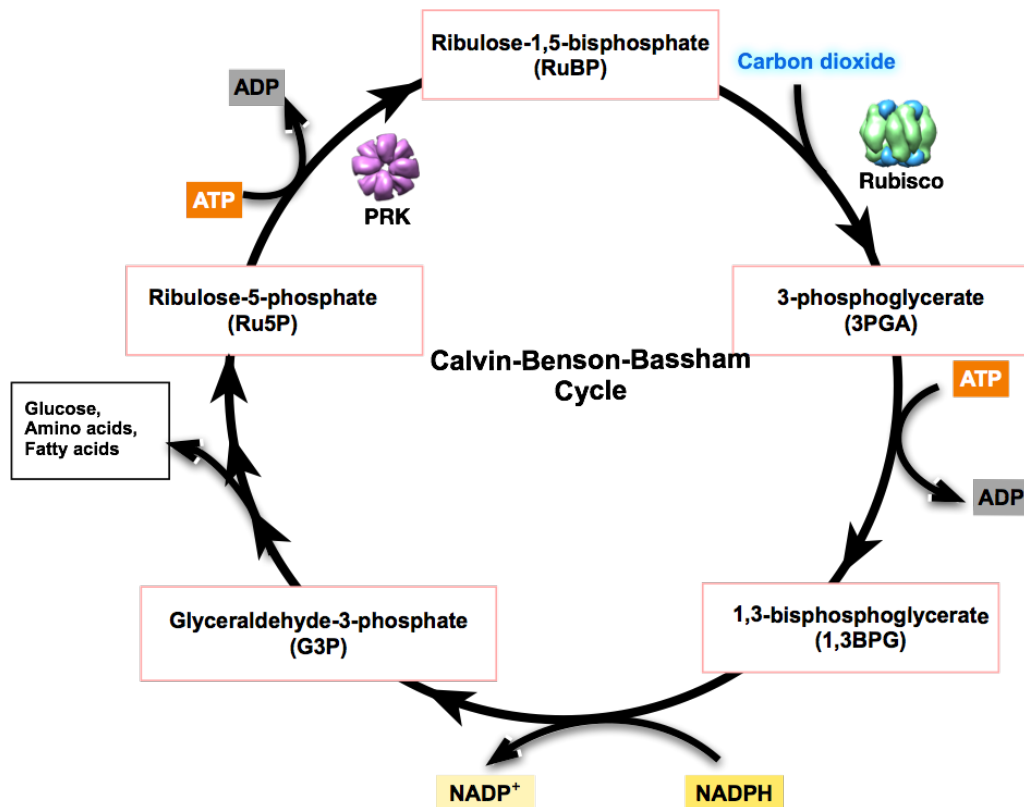


Figure 1-1: The Calvin-Benson-Bassham Cycle (simplified).

1.2 Photorespiration and carbon concentrating mechanisms

Rubisco has poor substrate specificity and can catalyze both carboxylation and oxygenation reactions of RuBP with CO_2 and O_2 respectively. During the oxygenation reaction, one molecule of 3PGA and one molecule of 2-phosphoglycolate (2PG) is produced (Andersson et al., 2008). The recycling of 2PG back into the 3PGA via the photorespiration pathway requires consumption of ATP and NADPH and releases a molecule of previously fixed CO_2 (Andersson et al., 2008; Bauwe, 2010). In plants, oxygenation reactions occur at a high frequency of twice per every five-carboxylation

reactions (Walker et al., 2016). It was estimated that the photorespiration pathway leads to approximately 30% loss in photosynthetic productivity (Zhu et al., 2010). To compensate for its inefficiencies, Rubisco is often highly expressed in plants and can account for up to 50% of total soluble proteins from leaves (Parry et al., 2013). Thus, Rubisco is considered to be the most abundant protein on earth (Ellis, 1979).

The poor substrate discrimination between CO₂ and O₂ is due to similarities in their size and electrostatic potential, and the disproportionate abundance of these two gases in the atmosphere (Whitney et al., 2011). Rubisco is an ancient enzyme that evolved more than 3.5 billion years ago under a non-oxygenic environment. The ancient atmosphere was CO₂ rich and O₂ free (Whitney et al., 2011). This is very different from the current atmosphere, which contains ~20% O₂ and 0.04% CO₂. Therefore, to minimize the impact of photorespiration, nature has developed two general solutions. The solutions involved either evolving mechanisms to concentrate CO₂ around Rubisco and/or to increase the enzyme's CO₂/O₂ specificity ($S_{C/O}$). The $S_{C/O}$ is defined as the ability of Rubisco to distinguish CO₂ from O₂ and this will be elaborated in section 1.5.

Cyanobacteria and many algae evolved carbon-concentrating mechanism (CCM) in the form of specialized compartments, carboxysome in cyanobacteria (Rae et al., 2013) and pyrenoids (Meyer et al., 2013) in algae. Bicarbonate pumps transport inorganic carbon into the cells and subsequently into these compartments. Carbonic anhydrase that are localized with Rubisco in these compartments, interconvert bicarbonate and CO₂ (Badger et al., 1998). Therefore, there is a higher local concentration of CO₂ surrounding Rubisco which reduces wasteful oxygenation reaction and photorespiration pathway (Giessen et al., 2017).

Some plants have evolved pathways to concentrate CO₂, namely C₄ and crassulacean acid metabolism (CAM) photosynthesis (Kromdijk et al., 2014; Leegood, 2002). In these plants, CO₂ is incorporated into 4-carbon compounds such as malate and oxaloacetate, which occurs in mesophyll cells. In C₄ plants, the malate is shuttled into the bundle-sheath cells where Rubisco is localized and isolated from atmospheric oxygen. The decarboxylation of the organic acid saturates Rubisco with CO₂. In CAM plants, the formation of the 4-carbon compound occurs in the night and stored in vacuoles. During the day, the organic acid is decarboxylated to CO₂ to Rubisco for carbon fixation.

1.3 Structure of Rubisco

Phylogenetic analysis showed that there are four forms of Rubisco enzymes found in nature, forms I, II, III and IV (Figure 1-2). Out of these four, only forms I, II and III catalyze RuBP-dependent carbon fixation (Badger and Bek, 2008; Tabita et al., 2008). Despite sharing little sequence identity (25-30%), all the different forms of Rubisco have highly conserved tertiary structures. This is due to the conservation of the antiparallel dimer of Rubisco large subunit (RbcL₂) as the fundamental functional unit.

1.3.1 The different forms of Rubisco

The form I Rubisco is the predominant form and the only form that contains small subunits. It forms a hexadecameric (L₈S₈) complex that consists of eight large subunits (RbcL, ~50-55 kDa) and eight small subunits (RbcS, ~13-16 kDa). The core of the complex is a tetramer of L₂ dimers and both ends are capped by two groups of four small subunits. The form I Rubisco can be phylogenetically divided into green-type (Form IA, IB) and red-type (Form IC, ID) (Figure 1-3) (Tabita et al., 2008a).

The green-type Rubisco (Form IA, IB) are found in higher plants, green algae, cyanobacteria, and proteobacteria. Form IA Rubiscos are found in prokaryotic organisms such as proteobacteria and Form IA can be subdivided into form IA_q or IA_c based on their genome associations with CbbQ and α -carboxysome genes respectively (Badger et al., 2008). Form IB Rubiscos are found in higher plant, green algae and beta-cyanobacteria (β -cyanobacteria). The Rubisco from β -cyanobacteria is associated with beta-carboxysome (β -carboxysome), thus it is classified as IB_c. In plants and green algae (form IB), the genes encoding for RbcL (*rbcL*) are located in the chloroplast genome while multiple RbcS genes (*rbcS*) are encoded in the nucleus.

The red-type Rubisco (form IC and ID) are present mainly in red algae, chemoautotrophic bacteria, phototrophic bacteria and phytoplankton (Andersson et al., 2008). The large and small subunit (*cbbL* and *CbbS*) genes in red-type Rubisco are encoded in a bicistronic operon and are often associated with other structural genes in the CBB pathway (Tabita et al., 2008a). These red-type Rubiscos have higher $S_{C/O}$ values as compared to green-type enzymes (Parry et al., 2013; Read et al., 1994). The RbcL sequences of form I Rubisco are highly conserved (~60% amino acid identity)(Parry et al., 2003) while RbcS sequences are more diverse (~30% amino acid

identity) and appear to have essential accessory functions (Hauser et al., 2015b; Spreitzer, 2003).

The form II Rubiscos can be found in different type of proteobacteria and dinoflagellates (Gibson et al., 1977; Liu et al., 2017; Tabita, 1988). It is made up of dimers of RbcL units $(L_2)_n$ and lacks the small subunit (Figure 1-2)(Gibson et al., 1977; Tabita, 1995). The first reported crystal structure of Rubisco was from the form II Rubisco of *Rhodospirillum rubrum*, comprising of a single RbcL dimer (Schneider et al., 1990). Recently, form II Rubisco from *Rhodopseudomonas palustris* and *Acidithiobacillus ferrooxidans* have been reported to form hexameric oligomers (Satagopan et al., 2014; Tsai et al., 2015). Many phototrophic and chemoautotrophic organisms such as *Rhodobacter sphaeroides*, *R. palustris* and *A. ferrooxidans* contain operons for both form I and II Rubisco (Dangel et al., 2015; Tsai et al., 2015). These two forms of Rubisco function to fix CO₂ under different CO₂ levels with the form I functioning at low CO₂ levels and form II functioning at high CO₂ level (Dangel et al., 2015). Thus, the form I Rubiscos are often synthesized during limiting CO₂ level as they have a higher affinity for CO₂ as compared to the form II Rubisco (Tabita, 1999).

Form III Rubiscos are found in archaea and bacteria where they form either a dimer (RbcL₂) or pentamer of RbcL₂ (RbcL₁₀, Figure 1-2) (Wrighton et al., 2016). Form III Rubisco have a high affinity for RuBP but low carboxylation rate and low S_{C/O} (Alonso et al., 2009; Nishitani et al., 2010; Tabita et al., 2008b). In addition, form III Rubisco have a high affinity for O₂. Mutagenic studies on Rubisco from *Archaeoglobus fulgidus* and *Thermococcus kodakaraensis* had shown that residues, Met295 and Ser363 play an important role in interacting with oxygen (Kreel et al., 2007; Kreel et al., 2015). Although these Rubisco are able to catalyze carboxylation reaction, they do not possess a CBB cycle (Tabita et al., 2008a). Instead, form III Rubisco metabolize RuBP generated from nucleoside degradation in the pentose bisphosphate pathway (Aono et al., 2015). The non-photosynthetic role of these form III Rubisco suggests that they have undergone different selective pressures during evolution as compared to Rubisco of photosynthetic organisms (Wilson et al., 2016).

Form IV Rubisco are Rubisco-like proteins (RLPs) and they lack conserved active site residues. Thus, they are unable catalyzed RuBP-dependent CO₂ fixation (Ashida et al., 2008; Badger et al., 2008; Hanson et al., 2001). RLPs can be found in proteobacteria,

Archaea, cyanobacteria, and algae. They form an antiparallel dimers of Rubisco large subunit and can be phylogenetically separated into at least seven clades (Erb et al., 2017; Tabita et al., 2008a). The function of some of these RLPs subfamilies have been characterized. The RLP from *Bacillus subtilis*, belonging to the tautomerase subfamily was the first characterized RLP and it functions in the methionine salvage pathway (Ashida et al., 2003). RLP from *R. rubrum* has also been shown to be involved in a novel isoprenoid shunt for 5'-methylthio-adenosine metabolism (Erb et al., 2012). Recently, a subfamily of decarboxylase was identified and the RLPs are involved in the degradation pathway of four carbon sugar acids (Zhang et al., 2016).

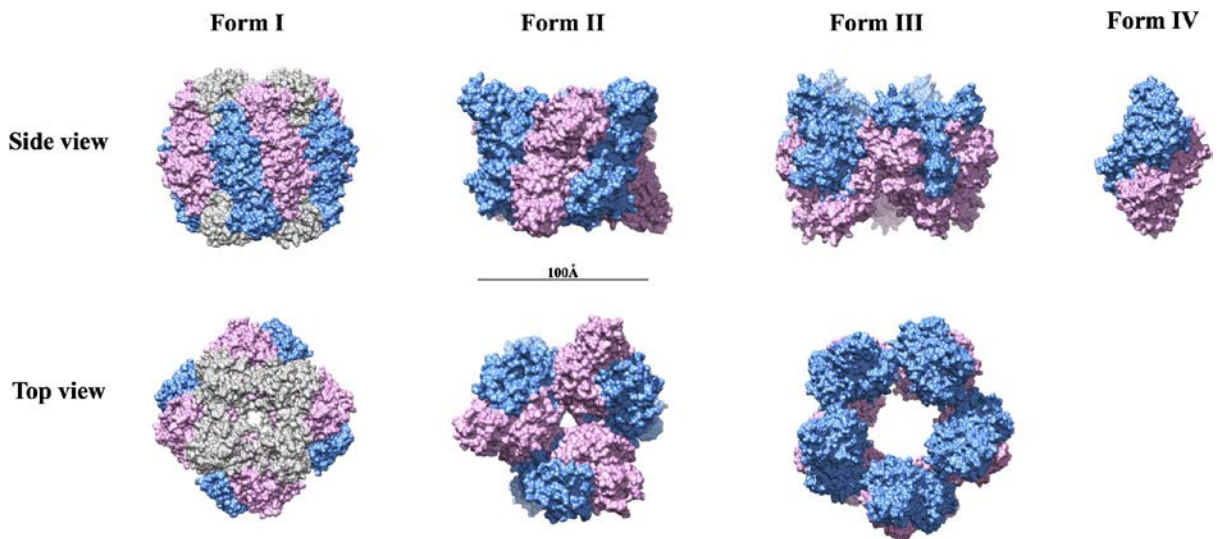


Figure 1-2: Structures of the four different forms of Rubisco.

All forms of Rubisco are made from functional units of antiparallel dimers of large subunit. Form I Rubisco forms from tetramers of dimer and crowned by four small subunits on the top and bottom (from *Oryza sativa*, PDB:1WDD). Form II Rubisco can form oligomers of dimer ($(L_2)_n$), the hexameric form is shown here (from *Rhodospseudomonas palustris*, PDB:4LF1). Form III Rubisco can form either tetramers or pentamers of dimers, the decameric form is shown here (from *Thermococcus kodakarensis*, PDB: 1GEH). Form IV Rubisco is made of a single dimer (from *Chlorobium tepidum*, PDB: 1YKW). Side and top views of Rubisco are shown in surface representation. The antiparallel dimer of large subunits is colored in purple and blue while the small subunits are colored in grey. Figure was made using chimera.



Figure 1-3: Phylogenetic tree of form I Rubisco large subunit sequences.

The form I Rubisco can be subdivided into either the green-type (form IA and IB) or the red-type (form IC and ID). Rubisco activase and molecular chaperones associated with folding and assembly for the different clade of Rubiscos are indicated. Rubisco activase from species that had been characterized are indicated in bold. The phylogenetic tree was constructed using multiple sequence alignment by CLUSTALW alignment and produced by neighborhood joining using the Mega 7 software suite (Kumar et al., 2016). The lineage of green-type Rubisco are in green while the lineage of red-type Rubisco are in red. List of the species name and accession code are listed in table A-1 in the appendix. Figure was modified from (Bhat et al., 2017b; Bracher et al., 2017).

1.3.2 Structure of Rubisco large and small subunit

RbcL consists of two distinct domains, N-terminal terminal $\alpha+\beta$ domain of about 150 amino acids long and a C-terminal domain comprising of about 325 amino acids. The C-terminal domain contains a $\beta 8\alpha 8$ triosephosphate isomerase (TIM)-barrel domain of about 310 residues and a flexible C-terminal tail (~15 residues) (Figure 1-4A, (Bhat et al., 2017a; Bracher et al., 2017)). In red-type Rubisco of red algae and some photosynthetic bacteria such as *R. sphaeroides*, the C-terminal tail is extended to ~24 residues (Bhat et al., 2017a). The large subunits are associated in a head to tail

arrangement with the active site being formed by residues in the N-domain of one subunit and C-domain of adjacent RbcL (Whitney et al., 2011). The majority of active site residues are contributed by the C-terminal domain with the remainder coming from the N-terminal domain. Side chains of these residues are shown in their respective domains in Figure 1-4A, B (Taylor et al., 1996a). Thus, in an antiparallel RbcL dimer, there will be two active sites. In all four forms of Rubisco, the secondary structure of RbcL is highly conserved despite the low sequence identity between them and the differences in structure are localized to a few of the loops (Andersson et al., 2008).

The RbcS core structure is comprised of a four-stranded anti-parallel β sheet with two α helices on the N-terminal side. The differences among different species' RbcS structure occur at the loop between strands β A and β B and at the C terminus (Figure 1-4C). The loop is longer in green-type Rubisco as compared to red-type Rubisco. In addition, the red-type Rubisco RbcS has an extension in C-terminal that forms a β hairpin (strands β E and β F) and this additional hairpin structure is absent in green-type RbcS (Joshi et al., 2015; Spreitzer, 2003). The β hairpin of four RbcS subunits assembles to form an eight stranded barrel structure which lines the pore on top and bottom of the octameric core complex (Bracher et al., 2017). In diatom Rubisco (red type), there is an additional helix after the β hairpin in the small subunit (Valegård et al., 2018).

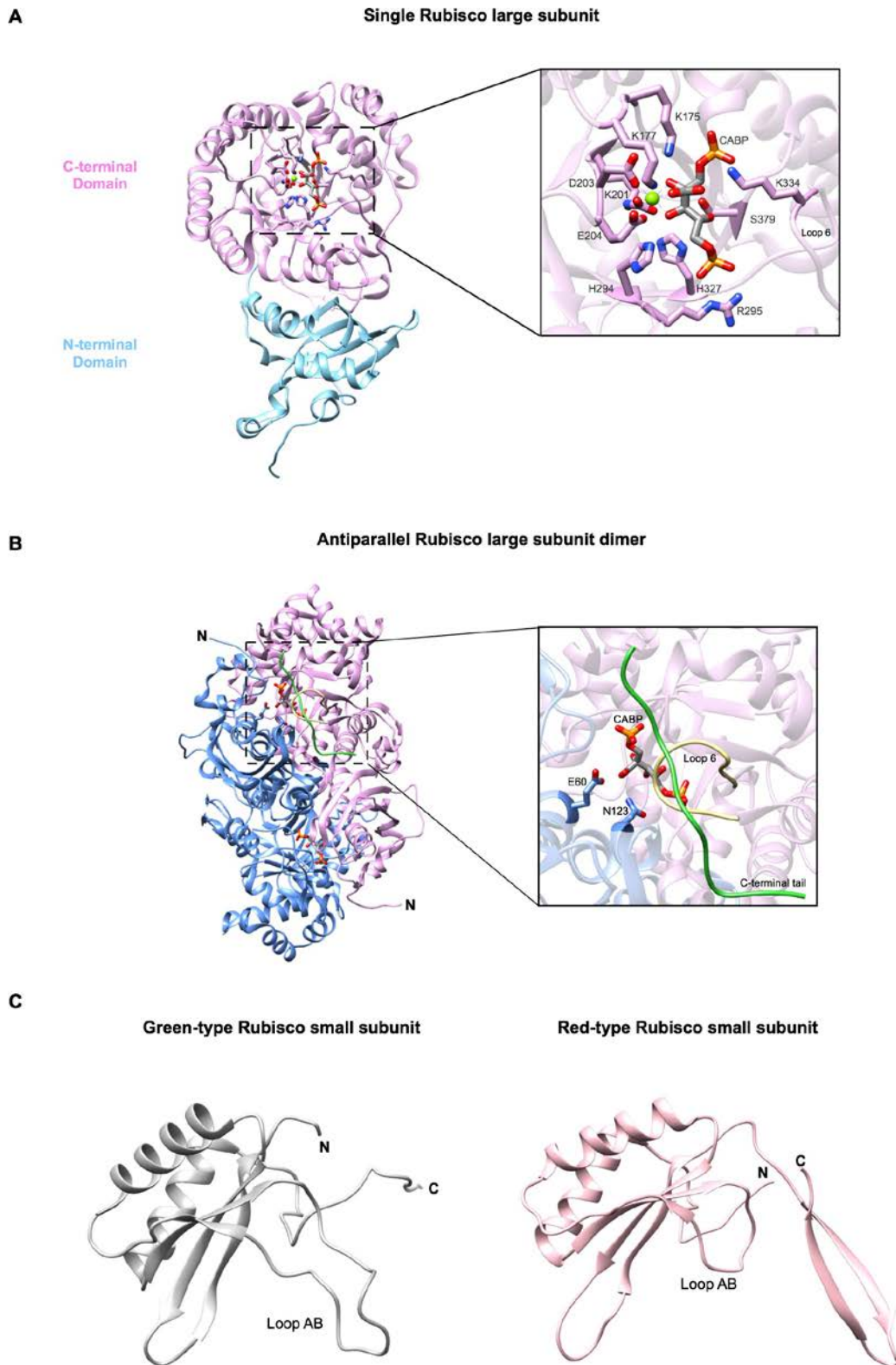


Figure 1-4: Structure of Rubisco large and small subunits.

(A) Structural details of a single Rubisco large subunit with CABP bound (from *Oryza sativa*, PDB: 1WDD). The N- and C-terminal domains are colored in light blue and respectively. The active site is formed from the N-terminal domain of one subunit and the C-terminal domain of the adjacent subunit. In the boxed area, CABP and conserved active site residues of the C-

terminal domain are labeled and shown in stick representation. The catalytic Mg^{2+} ion is shown as a green sphere. (B) Structural details of an antiparallel Rubisco large subunit dimer (from *O. sativa*, PDB: 1WDD). The dimers are colored in purple and blue. In the boxed area, conserved active site residues of the N-terminal domains are labeled and shown in stick representation. The loop 6 and C-terminal tail are colored in yellow and green respectively. The numbering of RbcL is according to spinach Rubisco. (D) Structural detail of green-type and red-type Rubisco small subunit from *O. sativa* (PDB: 1WDD) and *Rhodobacter sphaeroides* (PDB: 5NV3).

1.4 Reaction mechanism of Rubisco

Rubisco's complex reaction mechanism has been well studied over the years. Despite the variation in the quaternary structure of different forms of bona fide Rubisco, the reaction mechanism and the associated catalytic residues are completely conserved.

1.4.1 Activation of Rubisco

Rubisco must undergo an activation process prior to becoming catalytically competent. During the activation process, the essential active site residue Lys201 (Spinach Rubisco numbering) is carbamylated at its ϵ -amino group by a non-substrate CO_2 molecule and this results in the formation of a negatively charged carboxyl group (Figure 1-5A). The carbamylated Lys201 is subsequently co-ordinated to a magnesium ion (Mg^{2+}) to form an aci-carbamate. The Mg^{2+} is also coordinated to Asp203 and Glu204 with the remaining co-ordination sites being occupied by either C-2 and C-3 hydroxyl groups of bound sugar-phosphate or water molecules (Figure 1-5A) (Knight et al., 1990). This activation process is essential for the positioning of the ligand, RuBP, and this will enable efficient electrophilic attack by CO_2 or O_2 .

1.4.2 Conformational changes during catalysis

Following the activation, RuBP can then bind to Rubisco's active site. Rubisco undergoes a conformational change, from an open state to a closed state where the active site is being sequestered from the bulk solvent. The closing of the active site is achieved by folding of Loop 6 of the C-terminal TIM barrel over the active site. Loop 6 is subsequently fixed in position by the flexible C-terminal strand of RbcL. C-terminal strand will fold over loop 6 and it is secured by a single domain point in a manner of a latch via Asp473 (Figure 1-4B) (Duff et al., 2000). In the N-terminal domain, the binding of ligand triggers a two-degree rotation of the N-terminal domain of the neighboring RbcL. This movement brings the N-terminal domain active site residues to be closer to the active site, resulting in the ligand to be within a binding distance of these

residues (Taylor et al., 1996a).

1.4.3 The carboxylation reaction

The carboxylation reaction of RuBP by Rubisco consists of five steps (Figure 1-5B). The activated Rubisco catalyzes an enolization where it acts as a base to deprotonate C-3 of RuBP to form the highly reactive enediol intermediate. This enediol intermediate then attaches either CO₂ or O₂ in an irreversible reaction. These gaseous substrates interact with an essential active residue, Lys334 that is located at the apex of loop 6 (Andersson, 2008). During the carboxylation reaction, the ϵ -amino-group of K334 facilitates a nucleophilic attack of the enediol by withdrawing electron density from the carbon atom of CO₂. Subsequently, CO₂ and water can be added to form hydrated carboxyketone. The hydrated carboxyketone intermediate undergoes cleavage to form one molecule of 3PGA and one molecule of *aci*-acid. The *aci*-acid is subsequently protonated by Lys175 to produce the second 3PGA (Cleland et al., 1998; King et al., 1998; Knight et al., 1990).

1.4.4 Rubisco's side reactions

Due to the complexity of the multistep reaction, Rubisco catalyzes multiple unwanted side reactions. In the oxygenation side reaction, the enediol intermediate reacts with oxygen to form peroxyketone which undergoes bond cleavage to produce one molecule of 3PGA and one molecule of 2PG. However, the peroxyketone intermediate can also eliminate H₂O₂ to produce D-Glycero-2,3-pentodiulose-1,5-bisphosphate (PDBP). PDBP can be subsequently rearranged to form 2-carboxytetritol-1,4-bisphosphate (CTBP) (Andersson, 2008).

In addition to the oxygenation reaction, the enediol intermediate can undergo another two unwanted side reactions. It can be reprotonated to form Xylulose 1,5-bisphosphate (XuBP) which is a stereoisomer of RuBP. In plants, this side reaction occurs at a frequency of approximately one per 400 carboxylation reactions (Pearce, 2006). The enediol intermediate can also undergo β -elimination of inorganic phosphate to form 1-deoxy-D-glycero-2,3-pentodiulose-5-phosphate and pyruvate.

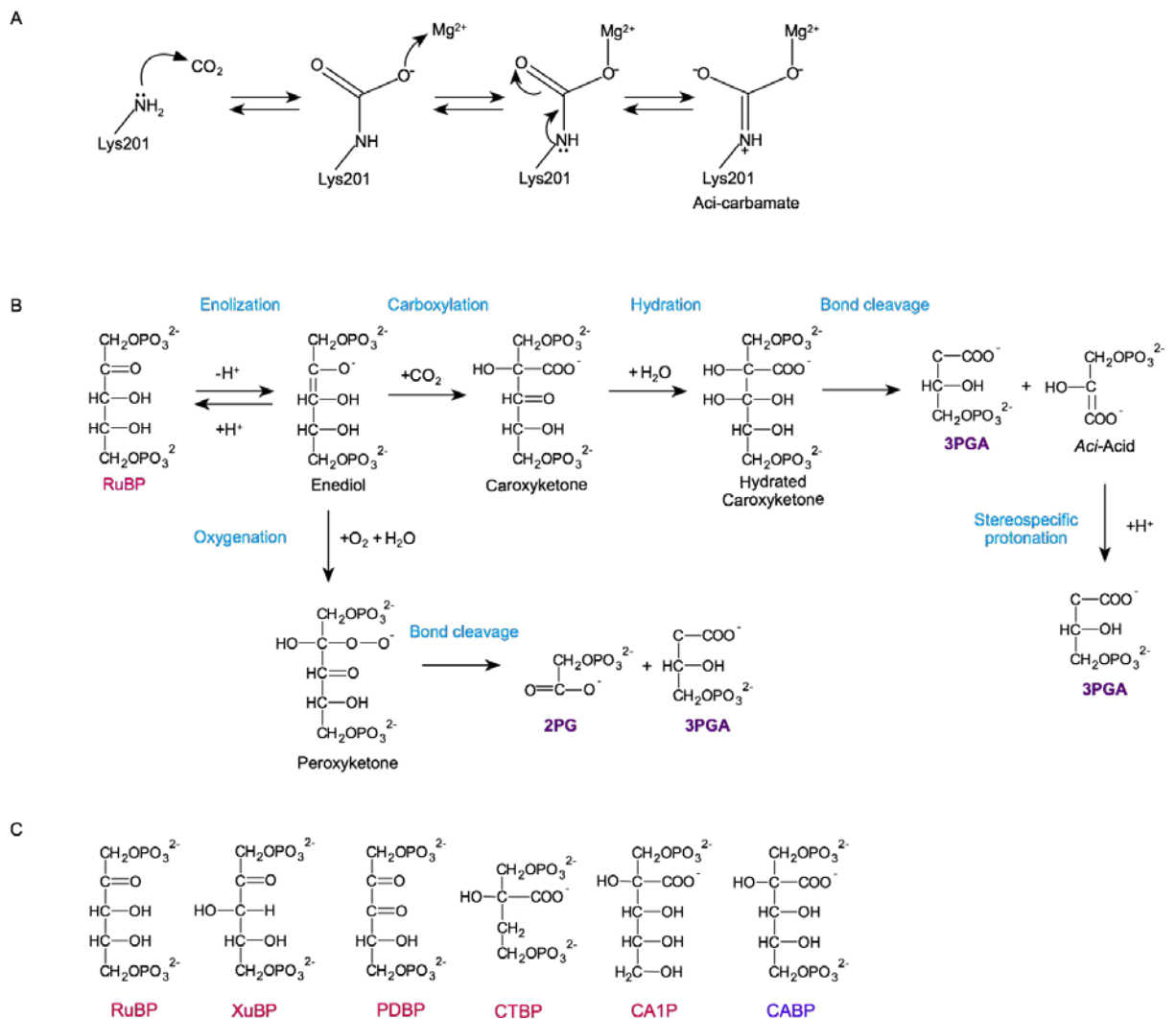


Figure 1-5: Rubisco's reaction mechanism and inhibitors.

(A) Activation of Rubisco by carbamylation of non-substrate CO_2 and binding of Mg^{2+} . (B). Carboxylation and oxygenation reaction of Rubisco. (C) Naturally occurring and artificial inhibitor of Rubiscos color coded in red and blue respectively. Abbreviations: RuBP, ribulose-1,5-bisphosphate; 3PGA, 3-phosphoglycerate; 2PG, 2-phosphoglycolate; PDBP, D-glycero-2,3-pentodiulose-1,5-bisphosphate; KABP, 3-ketoarabinitol-1,5-bisphosphate; CTBP, 2-carboxytetritol-1,4-bisphosphate; XuBP, D-xylulose-1,5-bisphosphate; CA1P, 2-carboxy-D-arabinitol-1-phosphate; CABP, 2-carboxyarabinitol-1,5-bisphosphate. Figure made with modification from (Andersson, 2008) and (Mueller-Cajar, 2017).

1.4.5 Sugar phosphate inhibitors

Naturally occurring sugar phosphate inhibitors of Rubiscos include its own substrate RuBP, and misfired products such as XuBP, PDBP, CTBP. Plants also synthesize another inhibitor, 2-carboxy-arabinitol 1-phosphate (CA1P) during low light or shade to regulate Rubisco's activity (Moore et al., 1992). The structure of these inhibitors are analogs to either the substrate or intermediate compounds from catalysis (Figure 1-5C)(Parry et al., 2008). Based on the structure of the carboxyketone, an intermediate of the carboxylation reaction, an artificial inhibitor 2-carboxyarabinitol-1, 5-bisphosphate (CABP) was synthesized. CABP is commonly used as an inhibitor for *in-vitro* characterization of Rubisco's activity (McCurry et al., 1981).

These sugar phosphates are able to bind to either the apo Rubisco (termed E) that has not undergone carbamylation and subsequent binding of the Mg^{2+} co-factor or the activated form of Rubisco (termed ECM) to form inhibited complexes, EI and ECMI respectively (Figure 1-6A). RuBP and XuBP are slow, tight-binding inhibitors and they can be bound to the apo Rubisco (Pearce, 2006; Pearce et al., 2003). Slow, tight-binding inhibitions involve two steps. Using RuBP as an example, the initial binding of RuBP to the apo Rubisco is a rapid equilibrium step to form an enzyme-inhibitor complex where RuBP could rapidly associate and dissociate from the enzyme. Over time, binding of RuBP induces slow conformational changes to the active site that cause tight-binding, forming a tightly bound enzyme-inhibitor complex (E:RuBP*) which eventually leads to dead-end inhibition (Figure 1-6B)(Morrison, 1982). Other sugar phosphates that bind to the ECM complex are also slow and tight-inhibitors. The release of these inhibitors from the active site is slow and limited by the rate of spontaneous opening of the active site. To accelerate the release of the inhibitor, Rubisco activases (Rca) are recruited. Rca will recognize the inhibited complexes (EI, ECMI) and facilitate the removal of these inhibitors in an ATP dependent manner (Bhat et al., 2017a). Rubisco can then be activated and subsequently fix CO_2 .

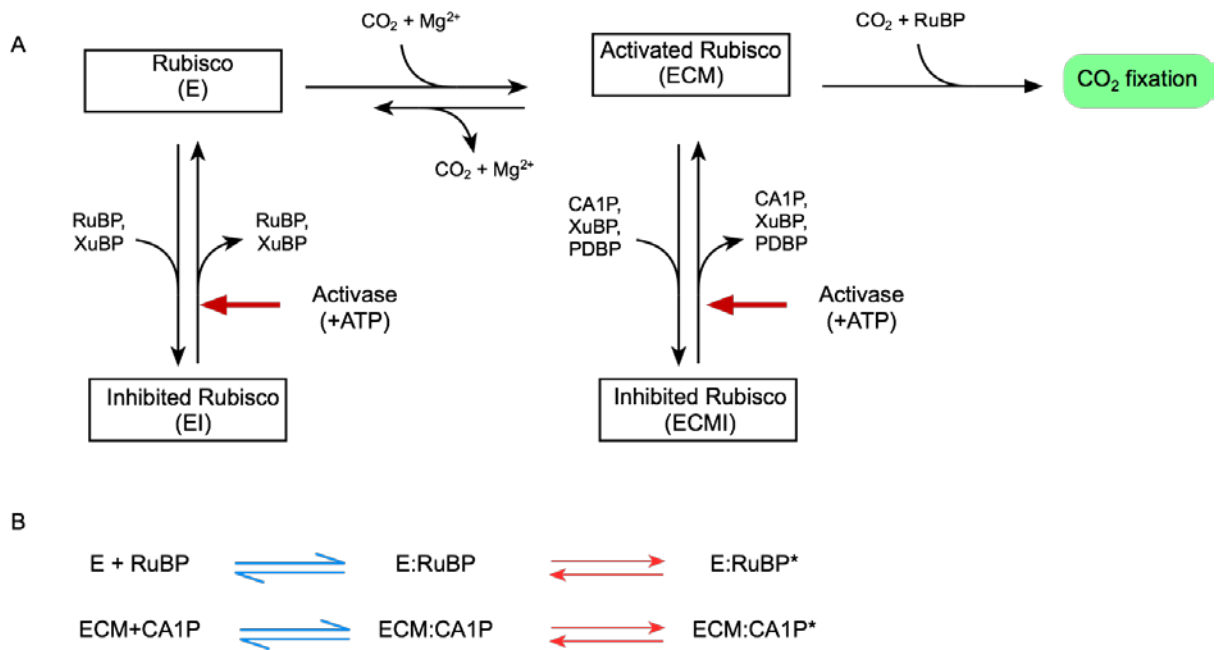


Figure 1-6: Regulation of Rubisco's activation state.

The activation state of Rubisco is controlled by Rubisco activase and inhibitors (RuBP, XuBP, CA1P, and PDBP). (A). The apo Rubisco (E) undergoes carbamylation with a non-substrate CO_2 and binding of a Mg^{2+} to be a catalytically competent holoenzyme ECM. Binding of inhibitors to the E and ECM will form inhibited Rubisco complexes, EI and ECMI respectively. Rubisco activase (Rca) recognizes these inhibited complexes and uses energy from ATP hydrolysis to induce conformational changes in the active site which leads to the release of inhibitor. (B). Examples of slow and tight-binding inhibitors of Rubisco. The initial binding was a rapid equilibrium step as denoted by blue arrows. This is followed by a slow conformational change in the active site of the Rubisco to form a tightly-bound enzyme-inhibitor complex where the inhibitor release is slow (denoted by an asterisk).

1.5 Diverse kinetic properties of Rubisco

Rubisco from diverse sources displays significant natural variation in the kinetic properties despite sharing the same catalytic chemistry (Table 1). The enzymatic properties of Rubisco are quantified using various parameters; the maximum carboxylation rate (k_{cat}^c), CO_2/O_2 specificity ($S_{C/O}$), Michaelis constant for CO_2 (Rubisco's affinity towards CO_2 , K_C) and the carboxylation efficiency (k_{cat}^c / K_C). The k_{cat}^c of Rubisco ranges between organisms with those from plants having lower k_{cat}^c of $\sim 3-5 \text{ s}^{-1}$ while those from cyanobacteria have the highest k_{cat}^c of $\sim 11-14 \text{ s}^{-1}$. The $S_{C/O}$ is a measure of Rubisco to distinguish CO_2 from O_2 and is described in Equation 1.1. Rubisco from red algae, *Galdiera sulfuraria*, have the highest $S_{C/O}$ (~ 166) which is twice the value as compared to plants and it also has high affinity towards for CO_2 (low K_C).

The carboxylation efficiency is another measure of Rubisco's enzymatic efficiency. The carboxylation efficiency is determined by the ratio of k_{cat}^c to the K_C and it reflects the ability of Rubisco to perform carboxylation reaction during limiting CO₂ levels (Badger et al., 1998). The most efficient Rubisco has a carboxylation efficiency of $\sim 4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and this is several orders of magnitude slower than catalytic perfection (10^8 - $10^9 \text{ M}^{-1} \text{ s}^{-1}$) where the catalysis is limited by second-order diffusion rate constant (Hanson, 2016)(Table 1-1). Thus, Rubisco is highly expressed in plants and often known as the most abundant protein on earth.

$$S_{C/O} = \frac{k_{cat}^c}{K_C} \frac{K_O}{k_{cat}^o}$$

Equation 1.1: The CO₂/O₂ specificity ($S_{C/O}$) of Rubisco.

k_{cat}^c and k_{cat}^o are the maximum carboxylation and oxygenation rate, respectively, and K_C and K_O are the Michaelis constants for CO₂ and O₂, respectively. The $S_{C/O}$ value represents the ratio of carboxylation to oxygenation at equal concentrations of CO₂ and O₂.

Comparisons of the kinetic properties of Rubisco have revealed a trade-off between the k_{cat}^c , $S_{C/O}$ and K_C (Savir et al., 2010; Tcherkez et al., 2006). Rubiscos with low k_{cat}^c have high $S_{C/O}$ and low K_C while those with high k_{cat}^c have low $S_{C/O}$. Those with high k_{cat}^c also tend to have higher K_C and are often found in organisms with CCM. Tcherkez et al. (2006) proposed the transition state theory which predicts that Rubisco with high $S_{C/O}$ favors tight binding of transition state intermediate from the carboxylation reaction as compared to intermediates from oxygenation reaction. The tight binding is achieved by the rigidification of the active site. The active site has reduced flexibility to allow stabilization and tighter binding of carboxylation transition state intermediates. The tighter binding results in slower carboxylation rate and a higher rate of misfire product formation (Tcherkez et al., 2006). Therefore, resulting in the trade-off between k_{cat}^c with $S_{C/O}$ and K_C .

The trade-off between the k_{cat}^c with $S_{C/O}$ and K_C are not linear and may vary with different photosynthetic taxa (Heureux et al., 2017). Form ID Rubisco from *G. sulfuraria* has the highest $S_{C/O}$ among the Rubisco but similar k_{cat}^c and K_C to plants Rubisco. This high $S_{C/O}$, however, is not observed in diatoms and haptophyte which also contain form ID Rubisco. Recently, Rubisco from 11 diatom species was characterised and they show a large degree of variation in K_C , $S_{C/O}$ and less variation in its k_{cat}^c (Table

1)(Young et al., 2016). Overall, there was no observed trade-off between the k_{cat}^c and K_C among these diatom Rubisco. However, there is a negative correlation between K_C and cellular Rubisco content. Resource are allocated to either increase CCM or increase cellular Rubisco content to counterbalance the CO₂ affinity of these Rubiscos, which is similar to the resource allocation towards Rubisco production in plants (Young et al., 2016). In comparison to C₃ plants, diatom Rubiscos have lower carboxylation efficiency. Likewise, cyanobacterial Rubisco which has some of the fastest carboxylation rate among Rubisco, also has lower carboxylation efficiency when compared to plant Rubisco.

The significant variation in the kinetic parameters among the different Rubisco suggests that there is an adaptation to selective pressures in the different gaseous and thermal environments. Tcherkez et al. (2006) hypothesized that most Rubisco may be nearly optimized for their environment despite their sluggish kinetics. Therefore, engineering a kinetically superior Rubisco variant into plants may not be feasible. Alternatively, an artificially generated Rubisco variant could be a more promising choice.

	k_{cat}^c (s ⁻¹)	K _C (μM)	k_{cat}^c/K_C (M ⁻¹ s ⁻¹)	S _{C/O}
C₃ plants				
<i>Arabidopsis thaliana</i>	4.1	10	4.1 x 10 ⁵	
<i>Spnacia oleracea</i>	3.7	14	2.6 x 10 ⁵	80
<i>Nicotiana tabacum</i>	3.1	10	3.1 x 10 ⁵	83
C₄ plants				
<i>Amaranthus edulis</i>	4.1	18	2.3 x 10 ⁵	77
<i>Flaveria australasica</i>	3.8	22	1.7 x 10 ⁵	77
<i>Zea mays</i>	5.5	19	2.9 x 10 ⁵	88
Green algae				
<i>Chlamydomonas reinhardtii</i>	5.8	29	2.0 x 10 ⁵	61
<i>Euglena gracilis</i>		25		54
<i>Scenedesmus obliquus</i>		38		63
Cyanobacteria				
<i>Synechococcus 6301</i>	11.8	200	5.9 x 10 ⁴	42
<i>Synechococcus 7002</i>	13.4	246	5.5 x 10 ⁴	52
<i>Prochlorococcus marinus MIT9313</i>	6.5	309	2.1 x 10 ⁴	59.9
Diatoms				
<i>Chaetoceros calcitrans</i>	3.4	31	1.1 x 10 ⁵	75
<i>Phaeodactylum tricornutum</i>	3.3	41.1	8.0 x 10 ⁴	116
<i>Thalassiosira weissflogii</i>	3.2	65	4.9 x 10 ⁴	79
Red algae				
<i>Griffithsia monilis</i>	2.6	9.0	2.9 x 10 ⁵	167
<i>Galdiera sulfuraria</i>	1.2	3.0	4.0 x 10 ⁵	166
<i>Porphyridium cruentum</i>	1.6	22	7.3 x 10 ⁴	129
Haptophytes				
<i>Pleurochrysis carterae</i>	3.3	17.7	1.9 x 10 ⁵	102
<i>Pavlova lutheri</i>	2.5	14.5	1.7 x 10 ⁵	125
<i>Tisochrysis lutea</i>	2.2	24.1	9.1 x 10 ⁴	89
Proteobacteria bacteria				
<i>Bradyrhizobium japonicum</i>	3.2	66	4.9 x 10 ⁴	74.8
<i>Xanthobacter flavus</i>	2.0	100	2.0 x 10 ⁴	44.4
Photosynthetic bacteria				
<i>Chromatium vinosum</i>	6.7	37	1.8 x 10 ⁵	41
<i>Rhodospirillum rubrum</i>	7.3	67	1.1 x 10 ⁵	12
Archaea				
<i>Methanococcus burtonii</i>	0.6	56.9	1.1 x 10 ⁴	1.3
<i>Thermococcus kodakaraensis</i>	0.3	52	5.8 x 10 ³	11

Table 1-1: Kinetic properties of the Rubisco from various species.

Kinetic data adopted from (Badger et al., 2008), (Young et al., 2016), (Whitney et al., 2011), (Heureux et al., 2017), and (Shih et al., 2016). Kinetics of various Rubisco were performed at 25 °C.

1.6 Maintenance of Rubisco's activity

Rubisco's activity is regulated by the binding of inhibitory sugar phosphates and the removal of sugar phosphates by Rca. In addition to Rca, specific sugar phosphatases have also been recently discovered in photosynthetic organisms and they play a role in the metabolic regulation of Rubisco activity by degradation of sugar phosphate. These machinery ensure that there is a timely removal of inhibitory sugar phosphates and allowing Rubisco to carry its catalysis (Bracher et al., 2017).

1.6.1 Rubisco activase

Rca proteins all belong to the ATPase associated with various cellular activities (AAA+) protein superfamily. These proteins are characterized by AAA+ modules and utilize energy from ATP hydrolysis to carry out molecular remodeling events (Ogura et al., 2001). Three distinct classes of Rca (green-, red- and CbbQO type) have been identified in photosynthetic and chemoautotrophic organisms containing red or green-type Rubiscos (Loganathan et al., 2016; Mueller-Cajar et al., 2011; Salvucci et al., 1985; Stotz et al., 2011; Tsai et al., 2015). These activases share a common feature of having a core AAA+ module, which consists of an N-terminal α/β nucleotide binding subdomain linked to a C-terminal α -helical subdomain (Figure 1-8)(Henderson et al., 2011; Mueller-Cajar et al., 2011; Stotz et al., 2011). The N-terminal domain contains the characteristic Walker A and B nucleotide binding motifs. Like other AAA+ proteins, these activases form a hexameric structure with a central pore (Mueller-Cajar, 2017). The sequence identity between the different activase are low and suggest that the different activase have emerged as a result of convergent evolution.

1.6.1.1 The Red-type Rubisco activase remodels red-type Form I Rubisco

The mechanistically best studied Rca is the red-type activase RsRca (or RsCbbX) from the alpha-proteobacteria *R. sphaeroides* which contains a red-type Rubisco (Form IC). RsRca is encoded by the *cbbX* gene which is in an operon containing Rubisco genes (*cbbL* and *cbbS*) (Gibson et al., 1997). The RsRca subunit is approximately 35 kDa and is composed of a core AAA+ module with an alpha helical N-terminal extension (Figure 1-7A). RsRca requires both ATP and allosteric binding of RuBP to form an active ring-shaped hexamer. The hexameric RsRca has a narrow central pore (~ 25 Å) and contains classical pore-loop sequences, which is present in many AAA+ proteins (Hanson et al.,

2005). In the absence of RuBP, RsRca is inactive and forms a long fibrillary structure (Mueller-Cajar et al., 2011). This structure was suggested to be a storage form which regulates RsRca activity during low photosynthetic activity (Mueller-Cajar et al., 2011).

RsRca interacts with the inhibited Rubisco by docking onto Rubisco side on and over one active site at a time. It subsequently engages the flexible C-terminal strand of RbcL and induced conformational remodeling of Rubisco by transient pulling of the C-terminal strand through the central pore. The pulling causes the detachment of the C-terminal strand from Rubisco's active site and results in retraction of loop 6. This will facilitate the opening of the active site and allow the release of the inhibitory sugar phosphate. The remodeling event by Rca is precise and limited to the active site (Bhat et al., 2017a; Mueller-Cajar et al., 2011).

Another red-type Rca has been recently characterized in the red algae *Cyanidioschyzon merolae*. *C. merolae* contains a red-type form ID Rubisco and two *cbbX* genes where one is nuclear-encoded and the other is plastid-encoded. Individual subunits expressed from either nuclear- or plastid-encoded genes alone are inactive. The functional CmRca requires both subunits and is a heterohexamer complex comprising of 1:1 ratio of nuclear and plastid-encoded subunits. These two Rca subunits share 60–70% identity with RsRca and the CmRca is able to remodel Rubisco from *R. sphaeroides* using the same conserved C-terminal strand threading mechanism (Loganathan et al., 2016). In both RcAs, RuBP functions as an allosteric regulator where its binding induces conformational changes and leads to higher ATPase rate (Loganathan et al., 2016). However, CmRca does not require RuBP for hexamer formation as compared to RsRca.

1.6.1.2 Green-type Rubisco activase acts on green-type form IB Rubisco

Rca of plants (green-type form IB) have been known since the 1980s and only recently have the first crystal structures of green-type Rca become available (Figure 1-7C) (Hasse et al., 2015; Henderson et al., 2011; Salvucci et al., 1985; Stotz et al., 2011). These green-type RcAs have longer sequences as compared to red-type Rca. One NtRca subunit is approximately 42 kDa and contains an AAA+ core module. Green-type Rca has an additional small domain (N-domain) at the N-terminus and a C-terminal extension, which has not been resolved in their crystal structure (Stotz et al., 2011). Green-type Rca recognizes its substrate using two features: the α -helical subdomain also referred to

as the specificity helix and the N-domain which is absent in red-type Rca (Portis et al., 2008). The specificity helix of *N. tabacum* (H9) interacts with residues Arg89 and His94 (*N. tabacum* numbering) on the inhibited *N. tabacum* Rubisco large subunit, allowing Rca to differentiate between solanaceous (eg. *N. tabacum*) and non-solanaceous (eg. *Oryza sativa*, Rice) Rubisco (Portis et al., 2008; Wachter et al., 2013).

In higher plants, there are two quasi-identical isoforms of Rca. They are the α - and β -isoforms and the α -isoform have a slightly longer C-terminal extension. These isoforms are either expressed individually from two genes or by alternative splicing of a single gene (Portis et al., 2008). Green-type Rca is regulated by the intracellular ATP/ADP-ratio and also by the reversible formation of a disulfide-bond in the conserved cysteine residues on C-terminal extensions of the α -isoforms in some plants (Portis et al., 2008).

Unlike red-type Rca, the mechanism of green-type Rca is less well understood. The central pore of NtRca is wider (~ 36 Å) and does not contain the classical pore-loop residues (Stotz et al., 2011). Green-type form IB Rubisco also lacks the extended C-terminus in its large subunit, which is present in red-type Rubisco for red-type Rca to act on (Bhat et al., 2017b). In addition, green-type Rca does not have a binding pocket for RuBP and its ATPase activity is constitutively active (Portis et al., 2008). All these suggest that green-type Rca uses a different mechanism for remodeling inhibited Rubisco and a different mode of regulation.

The thermolability of green-type Rca has been implicated to be responsible for the inhibition of photosynthesis at moderate heat stress (Salvucci and Crafts-Brandner 2004). Studies have shown that by transforming a thermostable Rca variant into *Arabidopsis rca* mutants, these plants are able to better tolerate elevated temperatures. They also have a higher photosynthetic rate when compared to wild type Rca (Kumar et al., 2009; Kurek et al., 2007). Recent work by Shivhare et al. (2017) has characterized a thermostable Rca from CAM plant, *Agave tequilana* (AtRca), which is $\sim 10^{\circ}\text{C}$ more thermostable than Rca from *O. sativa* (OsRca). In addition, AtRca is able to activate inhibited *O. sativa* Rubisco and has higher activity than OsRca. Thus, an improvement to the thermolability of green-type Rca by either using thermostable variant or those from CAM plants may contribute towards higher photosynthetic rates in plants.

1.6.1.3 CbbQO Rubisco activase is associated with green-type form IA Rubisco

The most recently discovered class of activase is the CbbQO activase from the chemoautotrophic bacteria *A. ferrooxidans* and *Halothiobacillus neapolitanus* (Sutter et al., 2015; Tsai et al., 2015). They contain green-type Rubisco form IA and the activase is encoded by the *cbbQ* and *cbbO* genes, which are associated with rubisco genes in an operon. *cbbQ* gene encodes for CbbQ protein (AfRca1/AfCbbQ1 and HnRca/HnCbbQ) which belongs to the MoxR group of prokaryotic AAA+ proteins. (Sutter et al., 2015; Tsai et al., 2015). MoxR AAA+ proteins are often associated in an operon with genes encoding for proteins that contain a von Willebrand Factor Type A (VWA) domain (Wong et al., 2012). *cbbO* gene is found downstream of *cbbQ* and it encodes the CbbO protein which contains a VWA domain. The VWA consist of a metal ion-dependent adhesion site (MIDAS) which mediates protein-protein interaction via single divalent cation such as magnesium ions (Whittaker et al., 2002). This MIDAS motif of CbbO likely interacts with RbcL of *A. ferrooxidans* via Asp82 (Tsai et al., 2015).

The CbbQO activase complex is comprised of a CbbQ hexamer and a single adaptor protein CbbO. This complex formation is essential for its activase function and is the only class of activase that requires an adaptor. A single CbbQ subunit is ~30 kDa while a single CbbO protein is ~86 kDa. The hexameric CbbQ complex has a small central pore of ~19 Å. The activation of inhibited form IA Rubisco requires the C-terminal RbcL similar to red-type Rca (Tsai et al., 2015). However, the CbbQ complex lacks the lacks the classical pore-loop residues. Thus, the mechanism of CbbQO activase is yet to be elucidated.

A. ferrooxidans contains multiple operons with encode for both Rubisco and CbbQO. One of these operons encodes for a form II Rubisco (AfCbbM) and another set of *cbbQ2/cbbO2* gene. Like other Form II Rubiscos, AfCbbM forms a hexameric structure composed of a trimer of RbcL₂ (Satagopan et al., 2014; Varaljay et al., 2016). AfCbbQ2O2 activase is able to remodel inhibited AfCbbM complexes and this is the first example of an activase acting on form II Rubisco (Tsai et al., 2015).

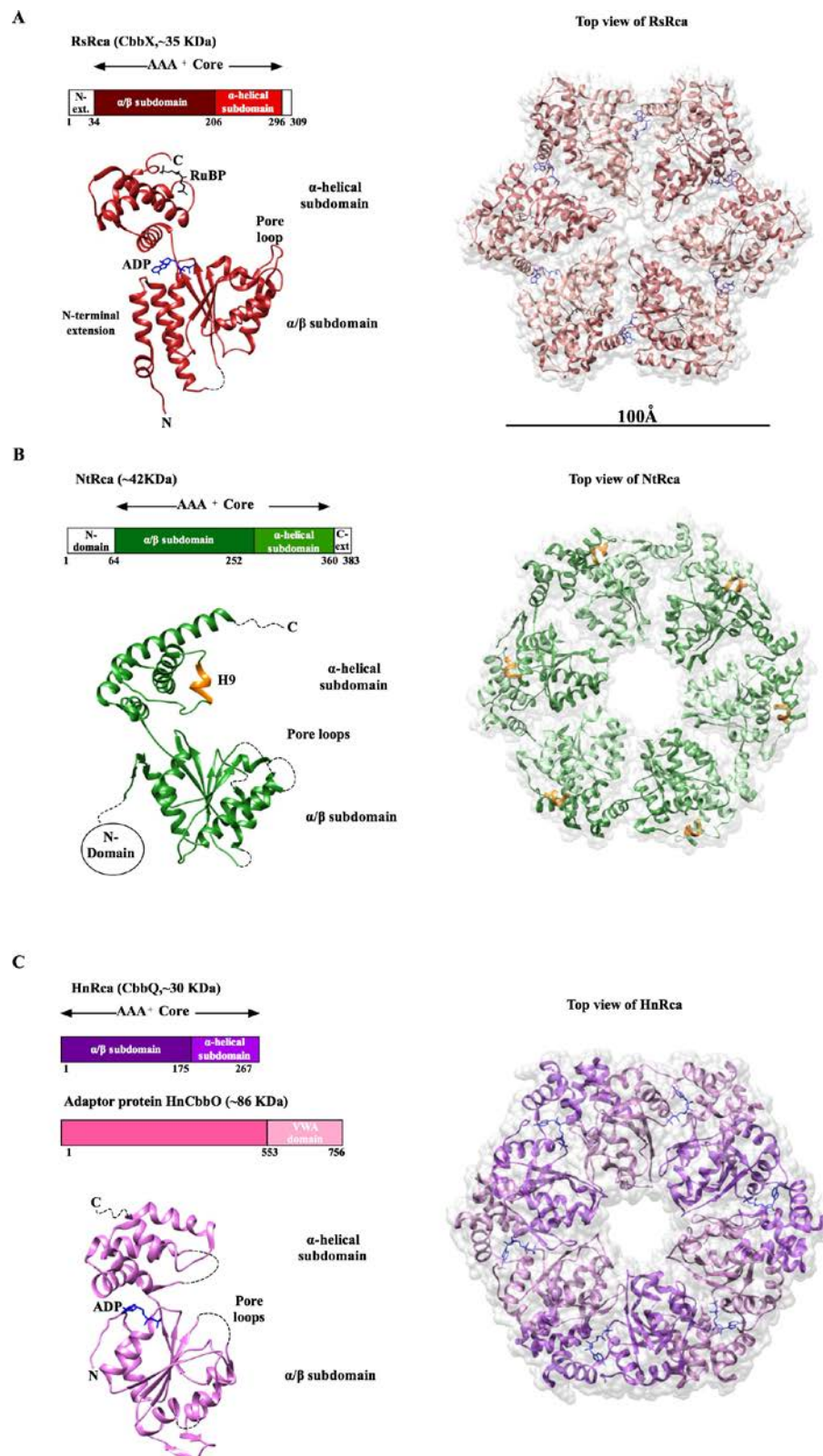


Figure 1-7: Structures of the different Rubisco activase classes.

(A) Red-type Rubisco activase (RsRca/RsCbbX) of form IC Rubisco from *R. sphaeroides* (PDB of RsRca: 3ZUH). (B) Green-type Rubisco activase (NtRca) of form IB Rubisco from *N. tabacum* (PDB of NtRca: 3ZW6) (C) CbbQO Rubisco activase of form IA Rubisco from *H. neapolitanus* (PDB of HnRca :5C3C). For each activase, there is a schematic representation of the domain structure of activase and its adaptor. Crystal structure of a single activase subunit is shown in ribbon presentation. The α/β and α -helical subdomains of AAA+ core, the positions of the pore loops, ADP (blue) and the allosteric regulator, RuBP (black) are also indicated. The C-terminal is denoted by C and N-terminal is denoted as N. In NtRca, the structure of N-terminal domain and C-terminal extension (C-ext.) are not resolved and their position are denoted with dash. The top view of each hexameric activase and their central pore is shown with alternating subunit been a different shade. Figure was made with modification from (Bhat et al., 2017b) and (Bracher et al., 2017).

1.7 Biogenesis and assembly of form I Rubisco

The biogenesis of form I Rubisco requires folding of the large subunit by chaperonin followed by assembly with small subunits, which is assisted by the various assembly factors to form the RbcL₈RbcS₈ complex (Figure 1-3)(Bracher et al., 2017). The folding of RbcL is mediated by the chaperonin system (Brinker et al., 2001; Liu et al., 2010). Chaperonins are large cylindrical complexes consisting of two heptameric rings. These subunits each form a central cavity for binding of non-native protein substrate (Farr et al., 2000). GroEL and Cnp60 is the chaperonin for bacteria and chloroplast, respectively (Vitlin Gruber et al., 2013). They function together with their respective co-factors, GroES in bacteria and Cnp10 and Cnp20 in chloroplast (Hauser et al., 2015b). Rubisco large subunit is prone to forming aggregates and requires encapsulation in the chaperonin cavity for folding. This was shown in previous *in vitro* study where active Form II Rubisco was obtained in the presence of GroEL-ES and ATP (Goloubinoff et al., 1989). However, Form I RbcL has pronounced tendency to rebind with GroEL after folding. Thus, additional assembly factors are required for assembly of RbcL and RbcS into functional RbcL₈RbcS₈ complex after folding.

Three assembly factors have been identified for green-type Rubisco; RbcX, Rubisco accumulation factor 1 (Raf1), and Raf2 (Bracher et al., 2015; Feiz et al., 2014; Feiz et al., 2012; Hauser et al., 2015a; Kolesinski et al., 2014; Li et al., 1997; Liu et al., 2010). RbcX and Raf1 are found in organism with green-type form IB Rubisco, such as plants, green algae and cyanobacteria. Whereas, Raf2 are found in organisms containing green type form IA Rubisco and eukaryotic organisms harboring form IB. RbcX and Raf1 interacts with RbcL downstream of chaperonin-associated RbcL folding. They function

to stabilized RbcL₂ but bind using different interaction sites (Bracher et al., 2017). Both RbcX and Raf1 exist as functional dimer and it binds to single RbcL₂ to form either RbcL₂RbcX₂ or RbcL₂Raf1₂ complex. This complex subsequently assembles into a higher oligomeric complex RbcL₈RbcX₈ or RbcL₈Raf1₈. Subsequently, the RbcS will displace the Raf1 or RbcX from the RbcL₈RbcS₈ complex (Bracher et al., 2017). The role of Raf2, on the other hand, has not been well characterized as compared to RbcX and Raf1. Raf2 has been shown to interact with RbcS in plants. Raf2 could play a role in chaperoning RbcS for imported into chloroplast and also assembly with RbcL (Feiz et al., 2014). Recently, Aigner et al. (2017) have successfully expressed *A. thaliana* Rubisco in *E. coli* by co-expressing chaperonins Cpn60/Cpn20 and assembly factors, Raf1, Raf2, RbcX, and BSD2.

The assembly of prokaryotic red-type Rubisco (form IC) is not dependent on the assembly factors used by green-type Rubisco and there are no sequence homologs of these assembly factors found in the organism containing red-type Rubisco. *R. sphaeroides* Rubisco can effectively assemble in *E. coli* as the assembly is mediated by Rubisco small subunit (RbcS) using the red-type RbcS exclusive C-terminal β -hairpin sequence (Joshi et al., 2015). In contrast, the assembly requirements of eukaryotic red-type Rubisco (form ID) has not yet been uncovered.

1.8 Directed evolution of Rubisco

Improving Rubisco kinetics has been predicted to increase the photosynthetic efficiency of plants. However, structure-based engineering of Rubisco has had limited success due to difficulties in pinpointing features that contribute to kinetic improvement. It was proposed that most natural Rubisco may be catalytically optimized due to the necessary trade-offs between $S_{C/O}$ and catalytic turnover (Tcherkez et al., 2006). Therefore, structure-based engineering could only improve one or more kinetic properties at the expense of others (Parry et al., 2003).

It was later hypothesized that multiple complementary mutations may be required to subtly reposition critical residues in the active site to lift Rubisco out of its local optimum (Spreitzer, 1993). This could be achieved by using a high-throughput direct evolution system which can examine Rubisco's sequence space and kinetic variants (Mueller-Cajar et al., 2008a).

1.8.1 Directed evolution of Rubisco using Rubisco deficient *R. capsulatus* and *R. eutropha* strains

The first reported directed evolution system for Rubisco used a Rubisco-deletion mutant of the metabolically versatile photosynthetic bacterium, *Rhodobacter capsulatus* (SBI-II⁻) (Paoli et al., 1998; Smith et al., 2003). This SBI-II⁻ strain could support CO₂-dependent growth when complemented with either native or heterologous Rubisco genes. When the SBI-II⁻ was *trans*-complemented *Synechococcus* PCC6301 Rubisco gene (SynRbcLS) and Syn-Rubisco was only able to complement photoautotrophic growth of SBI-II⁻ at high CO₂ (5%) but not at low CO₂ (1.5%) (Smith et al., 2003). Syn-Rubisco was chosen as it possesses a high K_C and selection under low CO₂ could lead to the selection of variants with lower K_C and thus enhanced kinetics. Therefore, a library of randomly mutated SynRbcLS genes was generated and it was screened using SBI-II⁻ for photoautotrophic growth at 1.5% CO₂. Three positive Rubisco variants were identified to be able to complement photoautotrophic growth of SBI-II⁻ at 1.5% CO₂ and they contained single substitution, F342V, F342I and M259T (Syn-Rubisco numbering) (Smith et al., 2003). These variants had similar kinetics properties (k_{cat}^c , K_C , $K_m(RuBP)$, $S_{C/O}$) as compared to the wild type with exception of the F345V mutant where it had an increase in affinity towards RuBP. The library was also screened using negative selection where Rubisco variants that were unable to support photoautotrophic growth under high CO₂ were chosen. The variants containing either D103V or G176D were isolated as negative mutants, and both exhibited higher K_C values as compared to the wild type (Smith et al., 2003, 2004). In a subsequent study, Satagopan et al. (2009) identified A378V substitution as a suppressor mutation for D103V using a D103V-RbcLS mutant library screened with the SBI-II⁻ host. The A378V substitution was able to suppress the negative phenotype caused by D103V but not G176D. Both A332V and D103V/A375V mutants were able to support anaerobic phototrophic growth of SBI-II⁻ strain at high CO₂. In addition, the A375V mutant was able to support anaerobic phototrophic growth of SBI-II⁻ at low CO₂ and this would indicate that this substitution is a positive mutation. The ability for A378V to sustain growth at lower CO₂ was due to the higher K_O/K_C ratio whereby the mutant enzyme had reduced oxygen sensitivity (Satagopan et al., 2009).

The functional role of two residues (Ile165 and Met331, *R. palustris* numbering) Form

II Rubisco from *Rhodopseudomonas palustris* was examined using the SBI-II⁻ host (Satagopan et al., 2014). These two residues were located near the active site and numerous mutants (I165T, I165A, M331L, M331A, and I165T/M331L) were created to mimic form I- and III-like sequences. Among the various mutants, only the wild type and I165V mutant were able to support autotrophic CO₂-dependent growth of SBI-II⁻. These mutants were subsequently subjected to random mutagenesis and suppressor mutations were selected using the SBI-II⁻ system. The A47V suppressor mutation was isolated for the negative mutant M331A and the double mutant of A47V/M331A was able to barely support the growth of the host. In contrast, the A47V mutant was able to complement the photoautotrophic growth of the host to a near wild-type level (Satagopan et al., 2014). The Rubisco deficient *R. capsulatus* strain was also used to study two form IA marine Rubisco (B15 and 4N23) from an oceanic metagenomic library (Witte et al., 2010). Both Rubisco were able to support the phototropic and photoheterotrophic growth of SBI-II⁻ host. This study highlights the feasibility and proof of concept that the SBI-II⁻ strain can function as a screening system for investigating the properties of diverse Rubisco genes isolated from the environment.

In a recent study, a Rubisco deficient *Ralstonia eutropha* strain (H16ΔLS) was constructed and the CO₂-dependent growth can be complement using heterologous enzyme with functionally compatible properties similar to the SBI-II⁻ strain (Satagopan et al., 2016). However, *R. eutropha* grows faster as compared to *R. capsulatus* and its Rubiscos has evolved to catalyze near optimal CO₂-fixation rate at the presence of high O₂ level. To assess the potential of this H16ΔLS strain for artificial selection, it was complemented with three *R. eutropha* Rubisco mutants containing F347V, A380V or F347V/A380V substitutions (*R. eutropha* numbering, equivalent to F342V and A375V in *Syn*-Rubisco). These two substitutions were identified from a previous study involving the directed evolution of *Syn*-Rubisco using SBI-II⁻ strain. Only the A380V mutant was not able to complement aerobic CO₂-dependent growth and random mutagenesis was carried out on the A380V mutant Rubisco gene followed by suppresser selection. Two suppressor mutations, T330A and T348C, were identified and they were able to support CO₂-dependent growth of H16ΔLS either in isolation or with the A380V substitution. Therefore, this Rubisco-deletion strain of *R. Eutropha* could be utilized in directing evolution Rubisco.

1.8.2 Directed evolution of Rubisco using Rubisco-dependent *E. coli*

E. coli is a preferred host for direct evolution applications due to its high transformation efficiency (up to 10^{10} colony forming units per microgram of plasmid DNA). Morell et al. (1992) proposed the concept of metabolically engineering *E. coli* towards Rubisco dependency. This involved deleting the glyceraldehyde 3-phosphate dehydrogenase (*gapA*) gene in *E. coli*, causing a glycolytic blockage. This mutant *E. coli* would then require upstream (eg. Glycerol) and downstream carbon source (eg. casamino acid, malate) for survival. The blockage can be bridged by the introduction of a two-step metabolic shunt comprising of PRK and Rubisco (Figure 1-6). PRK production is toxic to *E. coli* due to the accumulation of its product, RuBP which is inhibitory to cell growth and cannot be metabolized (Hudson et al., 1992). The understanding that PRK's lethality towards *E. coli* can be alleviated by expression of functional Rubisco led to the development of Rubisco-dependent selection systems using *E. coli* (Mueller-Cajar et al., 2007; Parikh et al., 2006). Parikh et al. (2006) used a wild type *E. coli* expressing the *Synechococcus* PCC7942 PRK to screen for *Synechococcus* PCC6301 Rubisco with improved fitness. The screenings yielded three *Synechococcus* Rubisco variants, all of which contained M259T at the Rubisco large subunit. This mutation conferred fitness by improving the functional assembly of the Rubisco in *E. coli* and allowed survival by increasing RuBP turnover (Greene et al., 2007).

A comparable Rubisco-dependent *E. coli* (RDE) system using the concept of Morell et al. (1992) was developed by Mueller-Cajar et al. (2007) (Figure 1-6). In RDE, the host $\Delta gapA$ *E. coli* strain, MM1, express *Synechococcus* PCC7942 PRK (pAC^{BAD}PRK) under the tight regulation of an arabinose inducible BAD promoter (Guzman et al., 1995). The use of MM1-*prk* for the screening of mutant Rubisco was more stringent compared to using wild type *E. coli*, as there was no false positive production. This was due to the increased susceptibility of MM1-*prk* to RuBP toxicity (Mueller-Cajar et al., 2007). MM1-*prk* RDE system has been used for evolving three form of Rubisco, form I (*Syn*-Rubisco), form II (*R. rubrum* Rubisco) and form III (*Methanococcoides burtonii*) (Durao et al., 2015; Mueller-Cajar et al., 2007; Mueller-Cajar et al., 2008b).

1.8.2.1 Directed evolution of Form II Rubisco using RDE

Directed evolution of Form II *R. rubrum* Rubisco using the RDE system generated variants with improved fitness. These variants contain one of four large subunit point mutations (H44Q/N or D117H/V) that were repeatedly isolated (Mueller-Cajar et al., 2007). These mutants did not affect the Rubisco expression and had quasi-identical changes in their kinetics with reduced $S_{C/O}$ and RuBP carboxylation turnover. The lowered $S_{C/O}$ will favor oxygenation reaction of RuBP and this results in increased production of 2-PG relative to 3-PGA. It was hypothesized that 2-PG may be metabolically favored over 3-PGA by the cells and this would serve as a selection pressure for these mutants to be selected. This provides providing a possible alternative fitness solution in this RDE system other than increasing RuBP turnover (Mueller-Cajar et al., 2007).

1.8.2.2 Directed evolution of Form I Bc cyanobacteria Rubisco using RDE

Synechococcus PCC6301 Rubisco was also evolved using the RDE system, but with different selection pressures by varying the level of L-arabinose and CO₂ level in growth media (Mueller-Cajar et al., 2008b). Multiple Rubisco large subunit variants were obtained (F345L, I174V, Q211L) including ones that were previously identified (M262T, F345I) from Δrbc *R. capsulatus*. These mutants were able to improve the folding survival by increasing the RuBP turnover (Mueller-Cajar et al., 2008b).

More recently, the *Synechococcus* PCC6301 Rubisco was evolved using the RDE system both in the absence and presence of chaperones (GroEL/ES or RbcX) (Duraio et al., 2015). The overexpression of the nonspecific chaperone GroEL/ES was able to enhance the rate of evolution by buffering destabilizing mutations that affect the folding of RbcL. Therefore, increasing the number of variants that can form the functional L₈S₈. This was previously observed in other directed evolution studies, where the overexpression of GroEL/ES was able to promote folding of variants of monomeric enzymes carrying mutations in their protein core (Tokuriki et al., 2009a; Wyganowski et al., 2013). In contrast, the Rubisco assembly chaperone RbcX specifically recognized surface properties of folded RbcL (Liu et al., 2010) and this restricts the range of permissible mutations. Mutations that affect the interaction surface of RbcL with RbcX will not assemble into an L₈S₈ complex. Thus, RbcX limits the evolvable sequence space

of Rubisco and appear to reinforce the evolutionarily trapped state of Rubisco (Durao et al., 2015). It is probable that other assembly chaperones would have the same effect (Durao et al., 2015) and these chaperones may have to co-evolved in order to express functionally improved Rubisco (Hauser et al., 2015b). Fifty unique Rubisco variants were identified in this screening and the F140I mutation in the RbcL identified in the absence of RbcX was the most interesting (Durao et al., 2015). This mutant had a 2.9-fold improvement in carboxylation efficiency but a slight reduction of 9% in CO₂ specificity. It also retained similar solubility as wild type *Synechococcus* Rubisco (Durao et al., 2015). Notably, isoleucine or leucine conserved at this position in Rubisco from plants, green and red algae while phenylalanine was conserved in cyanobacteria. The F140I mutant was transformed into a Δrbc strain of *Synechocystis* sp. PCC6803 (Syn6803) (Amichay et al., 1993) and it was able to support photoautotrophic growth at ambient CO₂ concentration. In addition, the F140I mutant was able to improve the photosynthetic rate of *Synechocystis* by ~55% (Durao et al., 2015).

1.8.2.3 Directed evolution of Form III Rubisco using RDE

The MM1-PRK RDE system was also recently used for directing the evolution of form III Rubisco from the Archaea *M. burtonii* (Wilson et al., 2016). Three mutant libraries of *M. burtonii* Rubisco (MbR) were generated and screening under RDE system generated 80 mutants with improved fitness compared with wild type. These mutants were transformed back into MM1-PRK and screened under increase stringent selective condition with higher PRK expression. Seven mutants out of the 80 mutants shown distinct growth advantage and these were transformed into XL1-Blue *E. coli* where the cellular content and the catalytic properties of each mutant is characterized. The three MbR mutants (MbR-K332E, MbR-E138V, MbR-T421A) showed significant improvements of 40-90% in the k_{cat}^C and 10-25% in $S_{C/O}$. The wild type *mbR* gene and the genes of two mutants (MbR-K332E and MbR-E138V) were synthesized with codon used for *N. tabacum* RbcL and there is an additional modification in the N-terminal sequence for post-translation modification which confers protection from proteolysis (Wilson et al., 2016). These modified genes were subsequently transformed individually into an *N. tabacum* transplastomic line, ^{cm}trL, to form produce three transplastomic lines *tob*^{MbR}, *tob*^{MbRK332E}, and *tob*^{MbRE138V} (Wilson et al., 2016).

Rubiscos from the T₀ progenies of *tob*^{MbR}, *tob*^{MbRK332E}, and *tob*^{MbRE138V} lines were

characterized. The two mutants (MbR-K332E and MbR-E138V) have higher k_{cat}^C and carboxylation efficiencies ($k_{cat}^C/K_C^{21\%O_2}$) as compared to those from wild type. These catalytic improvements did not come with any trade-off and the affinity of MbR to RuBP was maintain. In the subsequent T₁ progenies, the $tob^{MBRK332E}$ and $tob^{MbRE138V}$ plants show improvement in growth and healthier phenotype compared to tob^{MbR} lines.

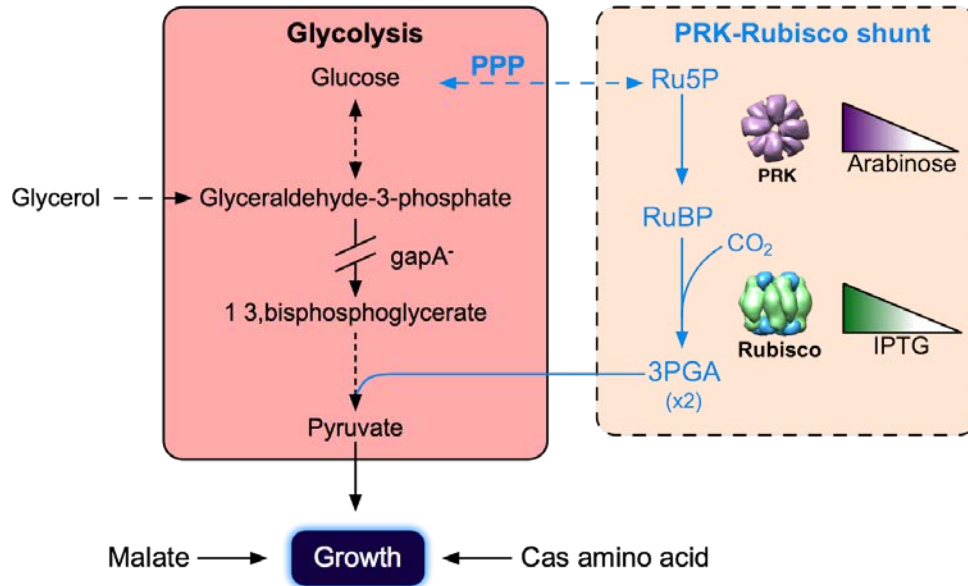


Figure 1-8: Screening of Rubisco using the Rubisco-dependent *Escherichia coli* system

In the mutant *E. coli* MM1 strain, there is blockage of both glycolysis and gluconeogenesis pathway due to the deletion of glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*⁻). This mutant strain requires upstream (eg. glycerol) and downstream carbon source (eg. Casamino acid and malate) for its survival. A metabolic bypass shunt can bridge this blockage by expression of Rubisco and phosphoribulokinase (PRK). PRK produces RuBP from Ru5P via the pentose phosphate pathway (PPP) and RuBP accumulation is toxic to the cell. This toxicity can be alleviated by Rubisco catalyzing reaction with RuBP to produce 3PGA. 3PGA can be utilized to produce pyruvate and be used for growth of the cell. The expression of Rubisco and PRK is regulated by varying the level of their respective inducers, IPTG and arabinose respectively. The mutant MM1 strain still requires trace levels of upstream and downstream carbon source to facilitate the early cell division. The figure is made with modifications from (Mueller-Cajar et al., 2008b) and (Wilson et al., 2016).

1.9 Aim and scope of the study

Rubisco is the key enzyme that catalyzes the rate limiting carbon fixation step in photosynthesis. In spite of its essential role, Rubisco is an inefficient enzyme with sluggish kinetics and poor substrate discrimination. In addition, Rubisco from many species are prone to inhibition by sugar phosphates including their substrate, RuBP. The

inhibited Rubisco complexes require remodeling by Rubisco activase to release the inhibitors from the active site. Therefore, Rubisco has been targeted for improvement using directed evolution. The RDE system has been successful in directing the evolution of various Rubisco and provide valuable insights into the fitness landscape of these Rubiscos. In this current study, we aim utilize this RDE system to direct the artificial evolution of green type form I Rubisco from *A. ferrooxidans* and red-type form I Rubisco from *R. sphaeroides* (AfLS and RsLS respectively). Both AfLS and RsLS form stable inhibited complexes and are associated with activases, AfCbbQO and RsRca respectively (Gibson et al., 1979; Mueller-Cajar et al., 2011; Tsai et al., 2015). Mutants obtained from RDE screening will provide insights into the how Rubisco can escape from activase dependency. In addition, chimeric (*Synechococcus/O. sativa*) Rubiscos were also generated and expressed in *E. coli*. Thus, using a combination of rational and evolutionary methods to modify Rubisco, we are obtaining insights into the underlying principles governing the shortcomings of the enzyme and its protein-protein interaction with accessory proteins.

2 MATERIAL AND METHODS

2.1 Bacteriological Methods

Four *E. coli* strains (BL21, DH5 α , DH10B, and MM1) were used in this study. For BL21, DH5 α , and DH10B, Luria-Bertani (LB) media and LB agar plate were used for competent cell preparation and transformation. For MM1 cells, rich permissive media (permissive media supplemented with 0.4% malate) was used for competent cell preparation and transformation, while permissive agar plate was used for plating.

	Composition
Luria-Bertani media	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar for solid medium
M9 minimal media	M9 salt (6% (w/v) Na ₂ HPO ₄ , 3% (w/v) KH ₂ PO ₄ , 0.5% (w/v) NaCl, 1% (w/v) NH ₄ Cl), 1 mM thiamine-HCl, 0.1 mM CaCl ₂ , 2 mM MgSO ₄ , 0.4% (v/v) glycerol
Permissive media	M9 minimal media supplemented with 0.4% (w/v) casamino acid, 1.5% (w/v) bactoagar for solid medium
TFBI	30 mM CH ₃ CO ₂ K, 100 mM RbCl, 10 mM CaCl ₂ , 50 mM MnCl ₂ ·4H ₂ O, 15% (v/v) glycerol, pH 5.8
TFBII	10 mM MOPS, 10 mM RbCl, 75 mM CaCl ₂ , 15% (v/v) glycerol, pH 6.5

Table 2-1: Composition of media and buffer used for the preparation of competent cells.

2.1.1 Preparation of Chemical Competent *E. coli* cells

Chemically competent *E. coli* cells (BL21 and DH5 α) were prepared by the calcium chloride (CaCl₂) method (Sambrook et al., 2001). A single colony was inoculated from glycerol stock or streak plate and grown overnight in INNOVA 44R incubator shaker (New Brunswick Scientific) at 37 °C, 220 rpm. The overnight culture was used to inoculate one liter of media and cells were grown to mid-log phase (OD₆₀₀=0.5) at 37 °C, 220 rpm. The culture was chilled on ice for 20 minutes and then centrifuged at 4000 rpm, 4 °C for 20 minutes. The supernatant was removed and cells were resuspended gently with 20 mL of chilled 0.1 M CaCl₂, 15% glycerol (autoclaved). The cell suspension was then incubated on ice for 20 minutes, and thereafter 100 μ L was transferred into sterile 1.5 mL microcentrifuge tubes. The tubes were frozen with either dry ice or liquid nitrogen and stored at -80 °C.

MM1 cells were prepared using the rubidium chloride (RbCl) method (Sambrook et al., 2001). The growth of MM1 cell culture used rich permissive media supplemented with

kanamycin (Kan, 30 mg/μL) and follows the same methodology as the CaCl₂ method. After obtaining the cell pellet, it was resuspended with TFB I buffer (40% of initial culture volume, Table 2-1). Resuspended cells were chilled on ice for 5 minutes and subsequently centrifuged at 4 °C for 15 minutes. The supernatant was discarded and the pellet was resuspended with TFB II buffer (4% of initial culture volume). Resuspended culture was chilled on ice for 15 minutes and thereafter, 100 μL was aliquot into sterile 1.5 mL microcentrifuge tubes. The tubes were frozen with either dry ice or liquid nitrogen and stored at -80 °C.

2.1.1.1 Chemical Transformation

100 μL of chemical competent cells were thawed on ice and 100 ng of plasmid was added to these cells. The mixture was incubated on ice for 30 minutes. After incubation, the cells were heat-shocked at 42 °C for 90 seconds and placed back on ice for 2 minutes. 400 μL of recovery media (LB or rich permissive media) was added and the cells were incubated in a shaking incubator at 37 °C, 1200 rpm for one hour. Cells were plated onto selection plates which were agar plates (LB or permissive) supplemented with respective antibiotics and incubated overnight at 37 °C.

2.1.2 Preparation of Electrocompetent MM1 cells

As described in 2.1.1, electrocompetent MM1 cells were grown to mid-log phase ($OD_{600}=0.5$) in 1 L of rich permissive media supplemented with kanamycin seeded from the overnight culture. The culture was chilled on ice for 20 minutes and harvested by centrifugation (20 minutes, 4000 rpm, 4 °C). The supernatant was removed and the pellet was washed with one liter of Milli-Q water (cold and autoclaved). The cells were then centrifuged and the supernatant removed. This washing step was repeated two more times, with the second wash using 500 mL of cold Milli-Q water and the third wash using 20 mL of 10% v/v glycerol (cold and autoclaved). After the washing, the cell pellet was resuspended with 10% v/v glycerol to achieve a final OD_{600} of 300. 50 μL of the electrocompetent cells were aliquoted into sterile 1.5 mL microcentrifuge tubes. The tubes were frozen with either dry ice or liquid nitrogen and stored at -80 °C.

2.1.2.1 Electroporation

100 ng of plasmid was added to 25-50 μ L of electrocompetent cells (DH10B and MM1). The mixture was transferred into 0.1 cm electroporation cuvettes (Bio-Rad, Singapore) and it was electroporated (1.8 kV, 25 μ F, and 200 Ω). Immediately after electroporation, pre-warm recovery media was added to the cells and the culture was transferred into a 1.5 mL microcentrifuge tube. The cells were incubated at 37 °C for one hour with shaking at 1200 rpm and cells were spread onto respective selection plates and incubate overnight at 37 °C. The electrocompetent DH10B competent cells (ElectroMAX™ DH10B™) were commercially purchased from Life Technologies (Singapore).

2.2 Molecular Biology

2.2.1 Polymerase Chain Reaction (PCR)

A high-fidelity polymerase with proofreading activities, Pfu polymerase (Thermo Scientific), was used for both DNA amplification and site-directed mutagenesis. The PCR reaction mix was as manufacturer's instruction. Cycling conditions were one cycle of 95 °C (2 minutes) then 35 cycles at 95 °C (2 minutes), T_m -5 °C (1 minute) and 72 °C (1 minute /kb). T_m is the annealing temperature of the primers.

2.2.2 pGemT Vector Cloning

The PCR product was polyadenylated at both the 5' and 3' ends to allow cloning of PCR product into the pGem T vector (TV; Promega). Polyadenylation was done by adding 10 U of Taq polymerase and 200 μ M dATP to the amplified DNA and incubated at 72 °C for 30 minutes. The polyadenylated PCR product was mixed with 5X DNA loading dye prior to loading into 1% TAE-agarose gel (40 mL) supplemented with 4 μ L of SYBR Safe DNA Gel Stain (Invitrogen). GeneRuler 1 kb DNA Ladder (Thermo Scientific) was loaded to serve as marker. Agarose gel electrophoresis was performed in TAE buffer at 100 volts for 30 minutes. DNA fragments were purified from the gel using the Wizard SV Gel and PCR Clean-Up system (Promega) according to the manufacturer's instructions.

Purpose of primer		Primer	Sequence
Remove <i>NdeI</i> restriction site from pTrcHisB backbone		NdeIpTrcBBFor	GTATTTACACCGCAGATGGTGCACTC
		NdeIpTrcBBRev	GAGTGCACCATCTGCGGTGTGAAATAC
Remove <i>PvuI</i> restriction site from pTrcHisB backbone		PvuIpTrcBBFor	TCTGACAACGATTGGAGGACCGAAG
		PvuIpTrcBBRev	CTTCGGTCCTCCAATCGTTGTCAGA
To generate RsΔ2		RsCbbLDelta2For	ACCGCCTCGGTCTAATAATAAGGAGACCGG
		RsCbbLDelta2Rev	CCGGTCTCCTTATTATTAGACCGAGGCGGT
To generate RsΔ4		RsCbbLDelta4For	GTGCCGACCGCCTAATAAGCGATGTAAGGA
		RsCbbLDelta4Rev	TCCTTACATCGCTTATTAGGCGGTCGGCAC
To mutate <i>NcoI</i> restriction site to <i>SacII</i> in pTrcHisB		NcoIpTrcSacIIFor	GAGGAATAAACCGCGGCCGGGGTGAA
		NcoIpTrcSacIIRev	TTACCCCCGGCCGCGGTTTATTCCTC
To mutate <i>SacII</i> restriction site to <i>NcoI</i> in pTrcHisB		SacIIpTrcNcoIFor	GAGGAATAAACCATGGCCGACGCGGCAA
		SacIIpTrcNcoIRev	TTGCCGCGTCGGCCATGGTTTATTCCTC
To remove internal <i>NcoI</i> restriction site in RsCbbX		NcoIRsCbbXFor	TCTCAAGCGCGCAATGGGCGGCGTG
		NcoIRsCbbXRev	CACGCCGCCCATTGCGCGCTTGAGA
To mutate <i>NcoI</i> restriction site to <i>NdeI</i> in pTrcHisB		NcoIpTrcNdeIFor	TAAGGAGGAATAACATATGGCCGACGCGG
		NcoIpTrcNdeIRev	CCGCGTCGGCCATATGTTATTCCTCCTTA

To mutate <i>HindIII</i> restriction site to <i>SpeI</i> in pTrcHisB	HindIIIpTrcSpeIFor	CTCTGGCCAGATGACTAGTGGCTGTTTTG
	HindIIIpTrcSpeIRev	CAAAACAGCCACTAGTCATCTGGCCAGAG
To generate RCbbX ^{K123A}	RCbbXK123AFor	GCGCCCAAGACGGCAGAGGTTCTCAA
	RCbbXK123ARev	TTGAGAACCTCTGCCGTCTTGGGCGC
To generate RCbbX ^{E138Q}	RCbbXE138QFor	GCTCTTCATCGACCAGGCCTATTACCTC
	RCbbXE138QRev	GAGGTAATAGGCCTGGTCGATGAAGAGC
To remove internal <i>SphI</i> restriction site in RsCbbLS	SphIRsSFor	GAAATGTGGGGAATGCCGATGTT
	SphIRsSRev	AACATCGGCATTCCCCACATTTTC
To remove internal <i>SphI</i> restriction site in SynLS	SphISeRbcLFor	CGCTAAAGAACTCGGAATGCCGATCATCATGC
	SphISeRbcLRev	GCATGATGATCGGCATTCCGAGTTCTTTAGCG
To mutate <i>HindIII</i> restriction site to <i>PstI</i> in pTrcHisB	HindIIIpTrcPstIFor	CCGATACTAAGCTGCAGCTGTTTTGGCG
	HindIIIpTrcPstIRev	CGCCAAAACAGCTGCAGCTTAGTATCGG
To mutate 3' <i>NdeI</i> restriction site to <i>SpeI</i> in pAC ^{BAD} PRK ^{DM}	NdeIPRKSpeIFor	CGGATCCTCGAGACTAGTGATTGAAACCTGA
	NdeIPRKSpeIRev	TCAGGTTTCAATCACTAGTCTCGAGGATCCG

Table 2-2 Primers used for Site-directed mutagenesis.

Ligation of the polyadenylated PCR product into the TV was carried out according to manufacturer's instruction. The ligation mixture was transformed into DH5 α and plated on ampicillin (Amp, 300 mg/mL) LB plate supplemented with X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside). X-gal enables blue white screening of the TV

and only white colonies containing insert will be inoculated, which will be grown overnight in a shaking incubator at 37 °C, 220 rpm. The plasmids were purified using Wizard Plus SV Miniprep DNA Purification System (Promega) and their concentrations were measured using NanoDrop 2000 Spectrophotometer (Thermo-scientific) at 260 nm. Protein-encoding sequences were verified by DNA sequencing (First base).

2.2.3 Site-directed Mutagenesis

Site-directed mutagenesis was used to introduce point mutations and insertion in existing plasmids. It was done using the QuikChange protocol (Stratagene) according to the manufacturer's instructions. Primers used were designed to have 12 complementary nucleotides to the plasmid at the 3' and 5' end with the center of the primer containing the desired nucleotide changes. The forward and reverse primers were reverse complementary to each other and the list of primers used for mutagenesis are listed in Table 2-2. *DpnI* treatment was used to remove the original plasmids, 10 U of *DpnI* was added to the mutagenesis PCR product followed by incubation at 37 °C for one hour. The treated PCR product will be subsequently transformed into DH5 α , colonies would inoculate and plasmid purified sequenced to check for the mutation.

2.2.4 Restriction digestion and ligation

0.5- 3 μ g of DNA were digested using 20-60 U of restriction enzymes (single or double digestion) in 10-50 μ L reactions with buffer according to manufacturer's instructions and carried out at 37°C for either 1-3 hours or overnight. For single digestion, 5 U of calf intestinal phosphatase (CIP) was added and then 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.

Ligation of purified DNA fragments (insert and vector) was carried out at room temperature for 1 hour or overnight at 16 °C. 20 μ L ligation reaction mixture consisting of 1X T4 DNA ligase buffer, 400 U of T4 Ligase, 100-300 ng of vector and insert (calculated based on 1:3 molar ratio of vector to insert). Ligation reaction could either be directly transformed into DH5 α or underwent DNA precipitation to obtain remove the salt.

2.2.5 DNA Precipitation

The ligation reaction was topped up to 100 μ L using PCR grade water. 10 μ L of sodium acetate (3M, pH=5) and 300 μ L of chilled 100% ethanol was added, followed by incubation on ice for 30 minutes. The mixture was centrifuged (4°C, 16000 rpm, 40 minutes) and the supernatant was discarded. The pellet was washed with 700 μ L of chilled 70% ethanol and centrifuged (4 °C, 16000 rpm, 30 minutes). The supernatant was pipetted away and the pellet was dried by leaving the cap open for 10 minutes. 10 μ L of PCR grade water was used to dissolve the pellet and the concentration was measured using NanoDrop.

2.3 Direct Evolution Methodology

2.3.1 pTrcHisB Plasmids

The expression vector pTrcHisB was used for the heterologous expression of Rubisco and activase in the various *E. coli* strains. Restriction enzyme sites, *Nde*I and *Pvu*I, were removed from the vector backbone using site-directed mutagenesis with appropriate primers (Table 2-2). All the pTrcHisB vector backbone originated from pTrcSynLS (Mueller-Cajar et al., 2008b), which was flanked by *Nco*I and *Hind*III sites and contains the Amp resistance gene. pTrcSynLS contains wild type *Synechococcus* PCC 6301 Rubisco and was obtained from Dr. S. Whitney (ANU, Australia).

Rhodobacter sphaeroides 2.4.1 Rubisco was amplified via PCR using primers, 5'*Nco*IRsL and T7 Term (CTCGAGCCATGGATACCAAGACCA and GCTAGTTATTGCTCAGCGG), from existing plasmid, pET30bRsCbbLS (Mueller-Cajar et al., 2011). The resulting amplicon was cloned into the TV and subsequently digested with *Nco*I and *Hind*III to obtain the RsCbbLS fragment flanked by *Nco*I and *Hind*III. This fragment was ligated into the pTrcHisB vector via *Nco*I and *Hind*III sites. Site-directed mutagenesis was used to obtain mutants Rs Δ 2 and Rs Δ 4 (Table 2-2), which have two and four amino acid deletions at C-terminus of the large subunit respectively.

To facilitate the ligation of RsCbbX into pTrcHisB vector, the *Nco*I restriction site was mutated to *Sac*II (Table 2-4). RsCbbX fragment was digested from pHUERsCbbX using *Sac*II and *Hind*III and cloned into pTrcHisB. Subsequently, the *Sac*II restriction site was mutated to *Nde*I, while the *Hind*III restriction site was mutated into *Spe*I (Table 2-1).

The internal *NcoI* restriction site in RsCbbX was also removed via mutagenesis. Mutants of RsCbbX (RsCbbX^{K123A}, RsCbbX^{E138Q}) were obtained via site-directed mutagenesis (Table 2-2). The list of pTrcHisB plasmid was listed below in Table 2-3 and a representation of the vector is shown in Figure 2-1A.

2.3.2 pTrcDuet Plasmid

To create a pTrcDuet plasmid containing both the activase and Rubisco, internal *SphI* restriction sites in Rubisco genes were removed using appropriate primers via site-directed mutagenesis (Table 2-1). The cassette Trc-RsCbbX was amplified from pTrcRsCbbX using primers 5'*SphI*Trc and 3'*SphI*Trc (AAGCATGCTTGACAATTAATCATCCG and ATGAGACAATAACCCTGATAGCATGCAG) and it contains the Trc promoter and terminator. The amplified product was polyadenylated and ligated into the TV (as described in 2.2.2). The Tv-TrcRsCbbX was digested with *SphI* to obtain the cassette fragment was flanked with *SphI* sites and was ligated directly into *SphI* site in the vector backbone of pTrcRsLS. The final pTrcDuet plasmid, pTrcRsLsRsX was obtained with both genes under the control of Trc promoter. Different types of pTrcDuet plasmid were generated containing either mutants of RsLS or RsCbbX and they are listed in Table 2-3. A representation of the pTrcDuet plasmid is shown in Figure 2-1B.

pTrcHisB plasmid	pTrcDuet plasmid
pTrcSynLS	pTrcRsLSRsX
pTrcRsLS	pTrcRsLSRsX ^{K123A}
pTrcRsΔ2	pTrcRsLSRsX ^{E138Q}
pTrcRsΔ4	pTrcRsΔ2RsX
pTrcRsCbbX	pTrcRsΔ4RsX
pTrcRsCbbX ^{K123A}	pTrcAfLSQ1O1
pTrcRsCbbX ^{E138Q}	pTrcAfMQ2O2
pTrcAfLS	
pTrcAfQ1O1	
pTrcAfM	
pTrcAfQ2O2	
pTrcOsRbcLS	
pTrcAtRcaβ	

Table 2-3: List of pTrcHisB and pTrcDuet plasmids used in this study.

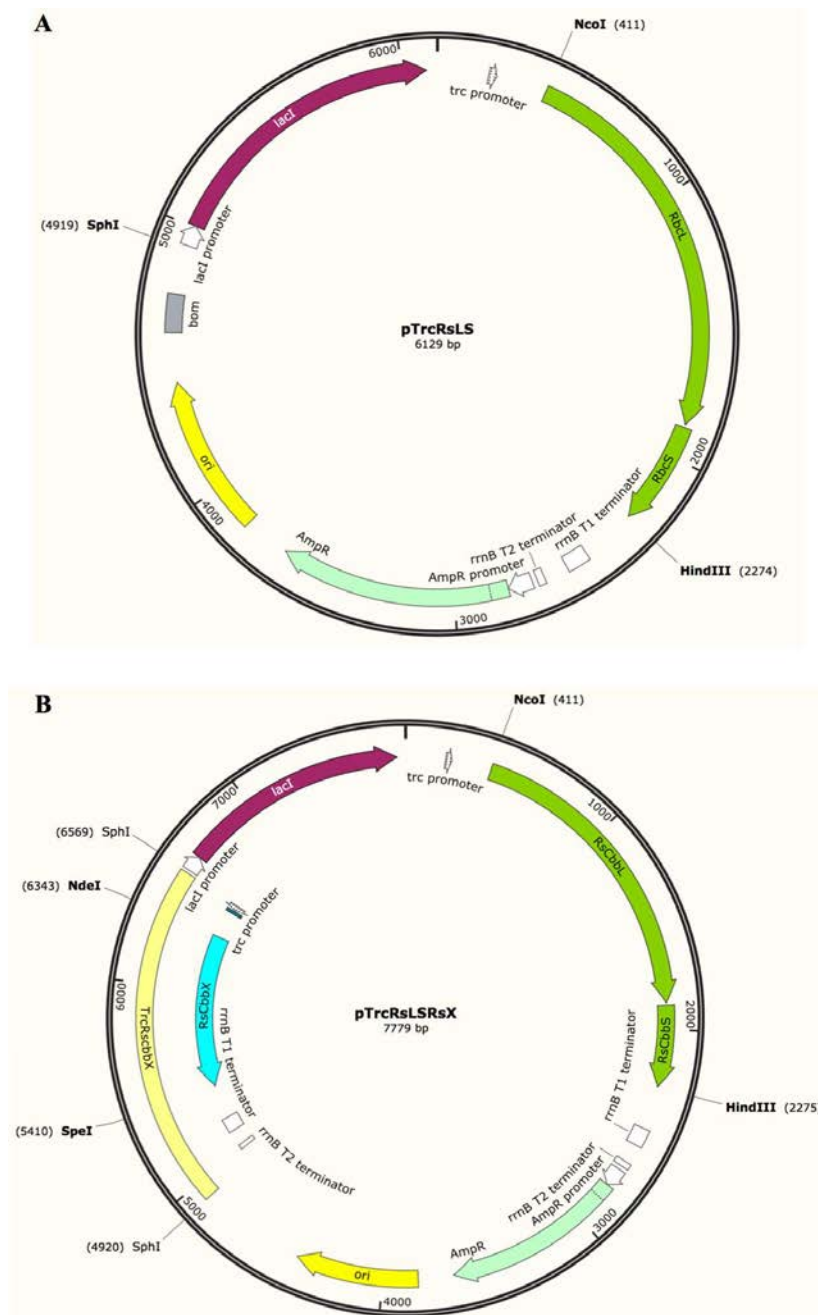


Figure 2-1: Representative vector map for pTrcHisB and pTrcDuet.

A. pTrcRsCbbLS. B. pTrcRsLSRsX

2.3.3 Reconstruction of PCC7942PRK expression plasmid

Plasmid pA^{CBAD}PRK encodes for the *Synechococcus* PCC7942 *prkA* gene is under the BAD promoter (Mueller-Cajar et al., 2008b). The laboratories' copy of pA^{CBAD}PRK plasmid was found to contain a Tn1000 transposon within the PRK gene and it was

removed via PCR using primers 5'PRK and 3'PRK+P (CATCTCTTCGAGGTTCTCGAACTGACCG and p-ATCTACGTCGAGGGCCACCTCAGCAAGA). The 3'PRK+P was designed to contain a 5' phosphate group to allow ligation of the PCR product using T4 Ligase after PCR clean up. After ligation, the plasmid was transformed into DH5 α and plated onto LB plate supplemented with chloramphenicol (Chl, 30mg/ μ L). Colonies were inoculated into LB supplemented with Chl and the plasmid was purified and verified by DNA sequencing. The *Synechococcus* PCC7942 *prkA* obtained from the PCR contained two mutations, E253G and G303D, and the final plasmid was designated as pAC^{BAD}PRK^{DM}.

2.3.4 Construction of pAC^{BAD}PRK⁶⁸⁰³ and pAC^{BAD}PRK⁷⁰⁰² vectors

Genomic DNA (gDNA) of *Synechocystis* sp. PCC6803 and *Synechococcus* sp. 7002 were obtained from 100 μ L of culture using cyanobacteria extraction buffer (Paz-Yepes et al., 2013) and Promega genomic DNA purification system. These cyanobacterial cultures were obtained from Dr. Tiago Toscano Selao (NTU, Singapore). The culture was pelleted down and the supernatant was discarded. The pellet was resuspended in 360 μ L of cyanobacteria extraction buffer (50 mg/mL lysozyme, 2 mM EDTA, 1.2% (v/v) Triton X-100, and 20 mM Tris, pH 8.0) and it was incubated at 37°C for 1 hour. Subsequently, 40 μ L of Proteinase K (20 mg/mL) was added and it was incubated at 56°C for 3hr with mild shaking. The subsequent steps utilized the gDNA extraction kit, Promega genomic DNA purification system and followed the manufacturer's instruction.

The *prkA* gene was amplified from the gDNA of the two cyanobacterial strains using their respective primer and it is flanked by restriction sites *Nde*I and *Spe*I. *Synechocystis* PCC6803 *prkA* (encoding for PRK⁶⁸⁰³) was amplified from its gDNA using primers 5'*Nde*I6301PRK and 3'*Spe*I6301PRK (ATATCATATGAGCAAGCCAGATCGT, ATATACTAGATTGAAACCTGAGCAAC). *Synechococcus* PCC 7002 *prkA* (encoding for PRK⁷⁰⁰²) was amplified using primers 5'*Nde*I7002PRK and 3'*Spe*I7002PRK (CGATCATATGTCTAAAAACATCCTAT, ATATACTAGTCTAACGGATTTTCCGGG).

The *Synechococcus* PCC 7942 *prkA* in pAC^{BAD}PRK was flanked by two *Nde*I restriction site and the 3'*Nde*I mutated to *Spe*I to facilitate cloning of other *prkA* into the vector. The amplified PCR products were digested with *Nde*I and *Spe*I. The fragment was

ligated to the pAC^{BAD} vector via the *NdeI* and *SpeI* restriction site. The ligation reaction was transformed into MM1 cells and plasmids were subsequently purified and sequenced. The wild type *Synechocystis* PCC6803 *prkA* and *Synechococcus* PCC 7002 *prkA* were obtained and denoted as pAC^{BAD}PRK⁶⁸⁰³ and pAC^{BAD}PRK⁷⁰⁰² respectively.

2.3.5 Establishment of RDE and RADE system

To establish the RDE and RADE system, various plasmids listed in Table 2-3, were chemically transformed into MM1 cells harboring the different PRK plasmids. The transformants were plated onto two sets of plates, permissive plate and selective plate. RDE system was previously described as using RDE selective media containing M9 minimal media supplemented with 0.05% casamino acid, 0.5 mM isopropyl-beta-D-thiogalactopyranoside (IPTG), 0.05% arabinose and antibiotics (Amp, Kan, and Chl) (Mueller-Cajar et al., 2008b). RADE system is a modification of a previously established RDE system (Duraio et al., 2015; Mueller-Cajar et al., 2008b). To establish the RADE system, multiple combinations of the M9 minimal selective media with different concentration of the components were attempted (Table 2-4). In addition, the level of CO₂ in the CO₂ incubator was also varied from 0.05% to 2%. Each plasmid was plated on half a section of the plates. The permissive plate was incubated for 24 hours at 37 °C, while the selective plates were incubated for 10 days. They were placed either in the CO₂ incubator at 23 °C with controlled atmosphere containing supplement of 0.5%-2% CO₂ or on bench top under ambient atmosphere (0.04% CO₂). The selective plates were checked every 24 hours and the day of first colony appearance was noted. At the end of the screening, the number of colonies was either counted manually or counted using GeneSys gel documentation and its software GeneTools (Synoptics Ltd). Images of the plates were also taken with GeneSys gel documentation. The RADE system was achieved when the condition that provided growth advantage towards the presence of wild type Rubisco and activase.

Components of selective media	Final concentration
M9 minimal media	1X
Casamino acid	0.05%- 0.2% (w/v)
Malate	0.005%- 0.05% (w/v)
IPTG	0- 0.5 mM
Arabinose	0.025%- 0.20% (w/v)

Table2-4: Components for optimization of RADE selective media.

2.3.6 Mutant Library construction

The construction of Rubisco mutant libraries using error-prone PCR (epPCR) as previously described (Mueller-Cajar et al., 2008b), with slight modifications. In brief, the Rubisco operon was amplified from their plasmid using primer 5'EpTrc and 3'EpTrc (CTGTGTGGGCACTCGACCGGAA, CCACCTGACCCCATGCCGAAC) using Taq polymerase (Invitrogen) in a 50 μ L epPCR supplemented with 0.25 μ L 10X mutagenic buffer (8 mM dTTP, 8 mM dCTP, 48 mM MgCl₂ and 5 mM MnCl₂). The epPCR cycling conditions were one cycle of 95°C (2 minutes) then 36 cycles at 95°C (1 minute), 55°C (1 minute) and 72°C (2.5 minutes).

The epPCR product was digested overnight with 50 U of *DpnI* at 37 °C and purified using Wizard SV Gel and PCR Clean-Up System. It was digested with *NcoI* and *HindIII* overnight at 37 °C and cloned into pTrcHisB or into the pTrcDuet followed by DNA precipitation as described in 2.2.4 and 2.2.5. The precipitated DNA was electroporated into DH10B as described in 2.1.4 and a 100-1000 dilution of the culture was plated on ampicillin LB agar plate to calculate the size of the mutant library. The transformants were topped up to 10 mL using LB media with Amp and grown overnight at 37 °C, 220 rpm. Mini-prep was performed on the overnight culture to obtain supercoiled plasmid libraries. The mutant plasmid libraries are denoted by "Ep" (for example, a mutant library of RsLS would be known as EpRsLS).

The construction of mutant RsCbbX library was performed as described above, with slight modifications. The epPCR cycling conditions were one cycle of 95°C (2 minutes)

then 36 cycles at 95°C (1 minute), 55°C (1 minute) and 72°C (1.5 minutes). The clean-up PCR products were digested with *NdeI* and *SpeI* instead. Thereafter, it was cloned into the pTrcDuet plasmid via the *NdeI* and *SpeI* site in the second Trc cassette.

2.3.7 Gene design of *A. ferrooxidans* Rubisco and activase

A. ferrooxidans form I and II Rubisco, AfLS and AfM, and their respective activases, AfQ1O1 and AfQ2O2 were synthesized (GenScript, USA). These genes were designed with all internal restriction sites removed. The Af-Rubiscos was flanked by *NcoI* and *PstI* while the Af-activases, were flanked by *NdeI* and *SpeI* (Figure 2-2). These synthesized genes were cloned in pUC57 simple vector and they were digested with the respective restriction enzyme to obtain the fragment of Rubisco or activase. The fragments were ligated into the pTrcHisB vector to obtain pTrcAfLS, pTrcAfM, pTrcAfQ1O1 and pTrcAfQ2O2. To obtain the duet of pTrcAfLSQ1O1, the AfLS fragment was ligated into the vector pTrc(*NcoI/PstI*)RsX and this was followed by a replacement of RsCbbX by AfQ1O1 via digestion and ligation. The pTrc(*NcoI/PstI*)RsX vector was obtained from restriction digestion of RsLS with *NcoI* and *PstI*. The duet of pTrcAfMQ2O2 was also obtained in a similar fashion as pTrcAfLSAfQ1O1. Mutant library of pTrcAfLS was generated as described in section 2.3.6 with modification in the restriction enzyme used. The clean-up PCR product of AfLS was digested with *NcoI* and *PstI* instead.

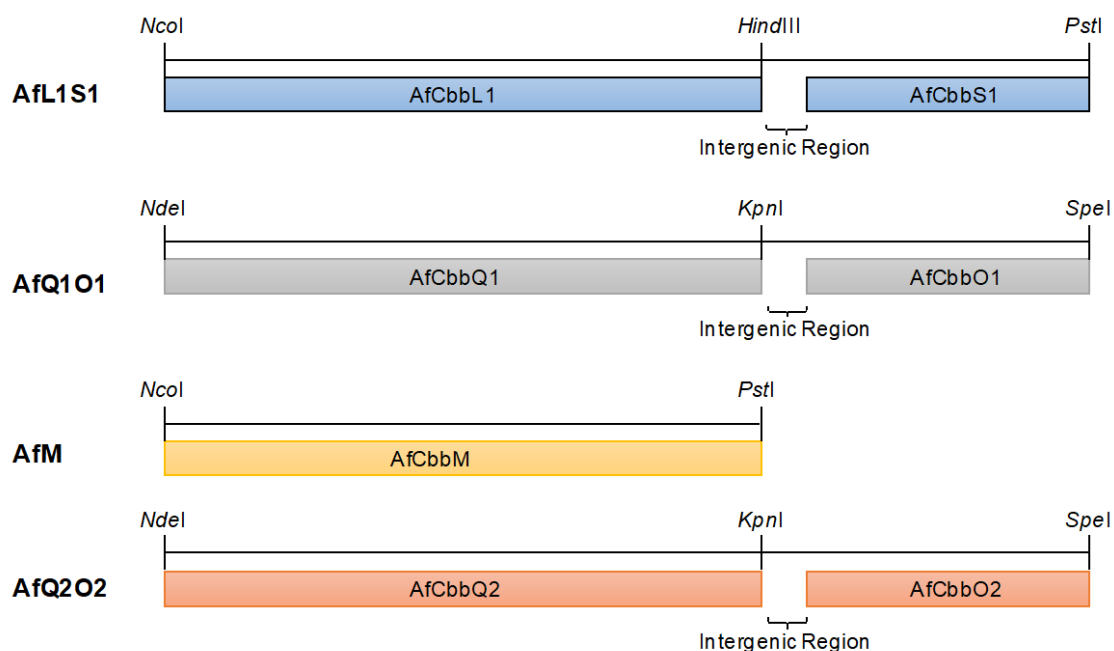


Figure 2-2: Gene design of *A. ferrooxidans* Rubisco and activase.

2.3.8 Selection in RDE and RADE

Mutant libraries were screened using MM1- pAC^{BAD}6803PRK cells under the RDE or RADE with wild type counterpart plasmid as controls. The plasmids and mutant libraries were electroplated into MM1-pAC^{BAD}6803PRK. 1% of the transformants were plated onto the permissive plates and the remaining (99%) were plated onto the selective plates. The permissive plates were placed in at 37 °C incubator for 24 hours while the selective plates were placed either in an incubator at 23 °C with 1% supplemented CO₂ or on bench top under ambient conditions for 14-30 days. Screening in RDE system uses RDE selective media supplemented with arabinose (0.05% (w/v)). RDE selective media composed of M9 minimal media supplemented with 0.04% (w/v) casamino acid, 0.5 mM IPTG and antibiotics (Kan, Chl, Amp). Whereas, screening RADE system uses RADE selective media supplemented with 0.05%-0.15% (w/v) arabinose. RADE selective media is M9 minimal media supplemented with 0.04% (w/v) casamino acid and antibiotics.

Colonies of mutant libraries that appeared earlier than wild type were inoculated into rich permissive media supplemented with antibiotics (Amp, Chl). These were grown overnight and the plasmids were purified. These plasmids were chemically transformed

into DH5 α and plated onto ampicillin LB plate. Colonies inoculated from the ampicillin LB agar plate would contain the mutant plasmid and the purified plasmid was sequenced.

2.4 Chimeric Rubisco Methodology

2.4.1 Gene Design

O. sativa and *Synechococcus* PCC 6301 Rubisco containing an F345I mutation in large subunit were synthesized (GenScript, USA) and were denoted as OsRbcL^{F345I}S and SynL^{F345I}S. The Rubisco large and small subunit were designed with all internal restriction enzyme sites removed and flanked by *NcoI* and *PstI*. The synthesized genes were cloned in a pUC57 simple plasmid which contains ampicillin resistance gene.

2.4.1.1 Rubisco Large subunit Gene Design

The UniProt accession codes for the large subunit of *O. sativa* and *Synechococcus* PCC6301 are: P0C512 and P00880, expressed amino acids 3-477 and 1-472 respectively. The large subunit was designed to be flanked with *NcoI* and *HindIII* restriction enzyme sites and contain an F345I mutation (numbering is based on *Nicotiana tabacum* Rubisco large subunit). Sequence alignment between *O. sativa* and *Synechococcus* PC C6301Rubisco reveal multiple regions with conserved amino acid sequences (Figure 7.1). Four restriction enzyme sites were introduced within these conserved amino acid sequences to produce five fragments. They were *BamHI* (G111, S112), *XbaI* (G196, L197, D198), *PvuI* (I301, D302, R303) and *BglII* (E392, I393, F394). The amino acid sequences denoted were from *O. sativa* and these five fragments were labeled A to E respectively. The codon for conserved amino acid residues of *O. sativa* was also changed to those from *Synechococcus*.

2.4.1.2 Rubisco Small Subunit Gene Design

The UniProt accession codes for the Rubisco small subunit of *O. sativa* and *Synechococcus* PCC 6301 are: Q0INY7 and P04716, expressed amino acids 48-175 and 2-111 respectively. The small subunit of both genes was designed to be flanked by *HindIII* and *PstI* restriction enzyme sites and contain an intergenic sequence from *Synechococcus* PCC 6301 Rubisco operon.

2.4.2 Cloning of Synthesized Genes into pTrcHisB

The large subunits of *Synechococcus* PCC 6301 (SynL^{F345I}) were cloned into pTrcHisB via *NcoI* and *HindIII* from the pUC57 Simple vector to form pTrcSynL^{F345I}. Rubisco small subunit of *Synechococcus* PCC 6301 (SynS) was amplified using PCR using primers, 5'*HindIII*SeRbcS and 3'*HindIII*SeRbcS (AAGCTTCCTCTGACTATCGCTGGGGGAGTGAGCGTTGCTGCGTAAT, AAGCTTAGTATCGGCCGGGACGATGAACGATG), and ligated into the TV. The SynS fragment was cloned into pTrcSynL^{F345I} via *HindIII*. The *HindIII* site at the 3' end of the SynS was mutated to *PstI* site using primers (Table2-4). Subsequent digestion with *HindIII* and *PstI* removes the SynS and allow ligations of the SynS and OsRbcS to produce pTrcSynL^{F345I}S and pTrcSynL^{F345I}OsS. pTrcSynL^{F345I}OsS was digested with *NcoI* and *HindIII* to allow cloning OsRbcL^{F345I} to form pTrcOsRbcL^{F345I}S.

2.4.3 Chimeric Rubisco

To obtain chimeric *O. sativa* and *Synechococcus* PCC 6301 Rubisco, both plasmids (pTrcOsRbcL^{F345I}S and pTrcSynL^{F345I}S) were digested with the same restriction enzymes that flanked the fragments. Each fragment of SynRbcL^{F345I} was exchanged with the respective counterpart from OsRbcL^{F345I} to generate the chimeric Rubisco of SynRbcLS containing single OsRbcL fragments. The small subunit of both Rubisco was exchanged as well. The plasmids of chimeric Rubisco were checked via sequencing. Chimeric Rubisco along with the wild types and SynLOsS where be expressed in BL21 and BL21 harboring pBADESL to test protein expression. Samples were analyzed using western blot under both denaturing and native conditions.

2.4.3.1 Restriction-free cloning of chimeric Rubisco

Chimeric *O. sativa*/*Synechococcus* Rubisco (SynL(OsL^N)S) contains the N-terminal domain of *O. sativa* (OsL^N) and the C-terminal domain of *Synechococcus* RbcL. This SynL(OsL^N)S chimeric Rubisco was generated using restriction-free cloning. Restriction-free cloning involves two rounds of PCR, the primary PCR is the amplification of the mega-primer which is used as primer in the secondary PCR. In the primary PCR, pTrcOsRbcL^{F345I}S was used as template, primers were OsL(N)For3 and OsL(N)Rev3 (CTGTTTGAAGAAGGATCCGTCACCAACATGTTTACCTCGATAGTCCGT,

GGATACCGTGGGGAGGACCTTGGAAGGTTTTTGAATAAGTAGGGGG). The insert (OsL^N) sequence in the primer is denoted in italics. The primary PCR uses Pfu polymerase as described in section 2.2.1 and the PCR product was separated by 1% agarose. The band corresponding to the product size was extracted and gel clean-up was performed. In the secondary PCR, Phusion polymerase was used and a 50 µL PCR reaction was carried out using 121 ng of mega-primer and 73 ng of vector (pTrcSynL^{F345IS}). The secondary PCR cycling conditions were one cycle of 95 °C (5 minutes) then 25 cycles at 95 °C (30 seconds), 55 °C (30 seconds) and 72 °C (2 minutes 10 seconds). After the PCR cycle, 10 U of *DpnI* was added to the PCR product and incubated at 37 °C for 1 hour. The *DpnI* treated PCR product was chemically transformed into DH5α and plated onto Amp LB plate.

Colony PCR was performed on the colonies to check for insertion of OsL^N. Colonies were inoculated using a pipette tip and it was swirl into 20 µL of stabilizing buffer (20 mM Tris-HCl, pH8 and 10% glycerol). 5 µL of this mixture was used for colony PCR as template while the remaining is store away at 4 °C. The PCR reaction mix was as manufacturer's instruction. Cycling conditions for colony PCR were one cycle of 95 °C (5 minutes) then 20 cycles at 95 °C (30 seconds), 55 °C (20 seconds) and 72 °C (90 seconds /kb). PCR product was analyzed using 1% agarose gel and samples with correct size bands were noted. The stabilizing mixture of the correct colony sample was used for growing an overnight culture in LB supplemented with Amp. Miniprep was performed on the overnight culture and plasmids were sent for sequencing.

2.5 Protein Biochemistry

2.5.1 Protein expression test and SDS-PAGE

Rubisco and activase plasmids were chemically transformed into BL21 or BL21 harboring pBADESL (Ewalt et al., 1997). The pBADESL plasmid encodes for chaperonin GroEL/ES under the control of arabinose-inducible pBAD promoter (Ewalt et al., 1997). A single colony was inoculated in 2 mL of LB supplemented with respective antibiotics and grown overnight in an incubator shaker at 37 °C, 220 rpm. Two falcon tubes containing 5 mL of LB and respective antibiotics were prepared and 50 µL of the overnight culture was transferred to each tube. The tubes were grown at 37 °C, 220 rpm until the OD₆₀₀ reached 0.6. After achieving the desired OD₆₀₀, the tubes were shifted to 23 °C and one tube was induced with 0.5 mM IPTG. For co-expression

with pBADESL, 0.2% (w/v) L-arabinose was added 30 minutes prior to the addition of IPTG. The culture was grown overnight at 23 °C, 220 rpm. The OD₆₀₀ of the overnight culture was measured and cells were harvested by centrifugation (4000 g, 10 minutes).

To lyse the cells, cells were resuspended with 1mL of lysis buffer (Table 2-5) and incubated on ice for 30 minutes. 1 mM of phenylmethylsulfonyl fluoride (PMSF) was added to the cells prior to sonication. Sonication was done at 30% amplitude for 5 minutes using pulses of 10 seconds alternated with 50 seconds of rest.

Protein expressions were analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). For analysis of total protein expression, 80 µL of the whole cell lysate was added to 20 µL of 5X SDS loading dye (Table 2-5). 100 µL of the lysate was centrifuged (16000 rpm, 10 minutes, 4 °C) and 80 µL of the supernatant was collected. 80µL of the supernatant was added to 20 µL of 5X SDS loading buffer and this was for analysis of soluble protein expression. The remaining supernatant in the centrifuged sample was discarded and the pellet was resuspended with 125 µL of 1X SDS loading buffer. The pellet was used for analysis of insoluble protein expression. The samples were denatured by incubation at 95 °C on a heat block and centrifuged for 2 minutes prior to loading on 12.5% or 16% SDS polyacrylamide gel. The samples loaded was 6×10^7 *E. coli* cells and this was calculated based on the OD₆₀₀ of the overnight culture where OD₆₀₀=1 is approximated to be 8×10^8 cells/ml. In addition, 5 µL of pre-stained marker or 5 µL unstained marker was loaded along with the samples (PageRuler™ Plus Prestained Protein Ladder 10-250 KDa, PageRuler™ Unstained Protein Ladder; Thermo Scientific) was loaded alongside the samples. Gel electrophoresis was performed using SDS running buffer and at a starting voltage of 120 volts followed by 160 volts for the rest of the run. The separated proteins were visualized by Coomassie staining solution (Table 2-5) and stained for an hour at room temperature on a rocking platform. The gel was destained using Coomassie destaining solution (Table 2-5) before being imaged using Epson V700 photo scanner.

SDS samples of *E. coli* lysate and supernatant were also loaded onto 4-12% Bis-Tris buffered NuPAGE SDS-gels (Invitrogen) and separated using 2-[N-morpholino] ethanesulfonic acid (MES) running buffer at 60 volts, 15 hours at 4 °C using a Novex electrophoresis cell (Invitrogen). The samples loaded was based on ¹⁴C-CABP binding as described below in 2.6.3.3. The gel was fixed using the fixing solution for 1 hour and

subsequently washed with repeated changes of H₂O. The separated proteins were visualized by Coomassie staining using Gelcode Blue Stain Reagent (Thermo Scientific). After 1 hour at room temperature, the gels were destained with repeated changes of Milli-Q H₂O before being imaged.

2.5.2 Western blot and Immunodetection

After SDS-PAGE, proteins were transferred to a nitrocellulose membrane using transfer buffer (Table 2-5) in a semi-dry western blot apparatus (Bio-rad). This transfer was performed at 10 volts for 30 minutes. The membrane would be stained with Ponceau S staining solution (Table 2-2) to check the efficiency of transfer and the staining solution was rinsed off with Milli-Q water. The membrane would be blocked with skimmed milk for one hour at RT. Primary antibody; anti-Rubisco (Rabbit, polyclonal) or anti-RsCbbX (Rabbit, polyclonal) were diluted to 1:5000 in TBS (Table 2-5). The primary antibody was incubated with the membrane for one hour at RT and subsequently washed three times with TBST (Table 2-5) with 10 minutes incubation time between each wash. The secondary antibody, horseradish peroxidase conjugated, anti-rabbit, was diluted to 1:0000 in TBS and incubated with the membrane for 1 hour at room temperature followed by another three times of wash. The HRP coupled antibody was detected by enhanced chemiluminescence (ECL) staining with a 1:1 mix of ECL detection reagent 1 and 2 (Thermo scientific) on a LAS-3000 Imaging system (Fuji).

For the pTrcDuet of *R. sphaeroides* plasmid, the membrane went through two rounds of detection. The first round of detection is as described above, thereafter, the membrane was washed three times with TBST. The membrane was then incubated with the other primary antibody and the remaining steps were the same as described above.

2.5.3 Native-PAGE

The supernatant of induced samples from the protein expression test were mixed with 5X native loading buffer (Table 2-5) and loaded into 6% native polyacrylamide gel. Pure Rubisco was loaded as a control along with the samples and these were separated at 180 volts, 2 hours at 4 °C. 6% Native polyacrylamide gel was prepared as per standard protocol and gel electrophoresis was performed using Native PAGE running buffer. After Native-PAGE, the gel was stained using Coomassie staining solution (Table 2-2).

The background staining of the gel was subsequently removed using coomassie destaining solution and gel was imaged.

The supernatant of *E. coli* extract was also loaded onto 4-12% Novex Tris-glycine buffered PAGE gels (Invitrogen) and separated using NATIVE PAGE running buffer at 60 volts, 4 °C for approximately 16 hours. The separated proteins were visualized by Gelcode Blue Stain Reagent coomassie staining as described in 2.5.1.

Western blot was also carried out for the Native-PAGE, the native polyacrylamide gel was soaked in transfer buffer supplemented with 0.1% SDS for 30 minutes, followed by transferring onto a nitrocellulose membrane using the same buffer. The subsequent steps were as previously described and the primary antibody used was anti-RbcL.

	Composition
5X Native loading buffer	62.5 mM Tris-HCl, 40% (v/v) glycerol, 0.1% (w/v) Bromophenol Blue
5X SDS loading buffer	225 mM Tris-HCl, pH 6.8, 50% (v/v) glycerol, 5% (w/v) SDS, 5% β-mercaptoethanol, 0.01% (w/v) Bromophenol blue
Coomassie destaining solution	10% (v/v) ethanol, 7% (v/v) acetic acid
Coomassie staining solution	0.1% (w/v) Serva Coomassie blue R-250, 40% (v/v) ethanol, 8% (v/v) acetic acid
Fixing solution	50% (v/v) methanol, 5% (v/v) glacial acetic acid
Lysis buffer	50 mM Tris-HCl, pH 8.0, 50 mM NaCl
MES running buffer	50 mM MES , 50 mM Tris Base, 3.47 mM SDS, 1.25 mM EDTA
Native PAGE running buffer	190 mM glycine 25 mM Tris
Ponceau S staining solution	0.1% (w/v) Ponceau S, 5% (v/v) acetic acid
SDS Running buffer	120 mM Tris-HCl, 134 mM glycine, 0.1% (w/v) SDS
TBS	50 mM Tris-HCl pH8.0, 137 mM NaCl, 2.7 mM KCl
TBS-T	TBS, 0.1% (v/v) TWEEN 2.0
Transfer buffer	50 mM Tris-HCl, pH 8.0, 20% (v/v) methanol, 192 mM glycine

Table 2-5: Buffer for protein biochemistry

2.5.4 Rubisco purification

Mutants and wild type Rubisco from *A. ferrooxidans* and *R. sphaeroides* were purified as previously described by Tsai et al. (2015). pTrcHisB plasmids encoding for these

Rubiscos were transformed into BL21 and subsequently grown in 1 L of LB supplemented with Amp. The culture was grown at 37 °C for 3 hours followed by induction with 0.5 mM IPTG at 23°C and the culture continue to grow overnight at 23 °C. The pellet of the overnight culture was obtained from centrifugation (6000 *g*, 20 minutes, and 4 °C). The pellet was lysed with lysis buffer and incubated with 0.3 mg/mL of lysozyme on ice for 30 minutes. 1 mM of PMSF was added to the sample prior to sonication for 5 minutes with 15 seconds pulse on and 30 seconds pulse off interval. The supernatant was obtained from centrifugation (40000 *g*, 45 minutes, and 4 °C) and it was applied to an equilibrated Source 30Q column (GE Healthcare). The Source 30Q column was pre-equilibrated with buffer A (20 mM Tris-HCl, 50 mM NaCl, pH 8.0) and the proteins were eluted with a linear salt gradient to 0.5M NaCl. Fractions were analyzed using 16% SDS-PAGE and fractions containing Rubisco were combined and concentrated to 5 mL using Amicon Ultra-15 Centrifugal Filter Units (50 kDa cut-off, Merck, USA). The combined fraction was applied to a Superdex 200 16/60 size-exclusion column and proteins were eluted with buffer A. Fractions from the size exclusion column were analyzed using 16% SDS-PAGE. The purest Rubisco fractions were pooled and concentrated using Amicon Ultra-15 Centrifugal Filter Units (100 kDa cut-off). The concentrated Rubisco was supplemented with 5% (v/v) glycerol followed by flash-frozen in liquid nitrogen and stored in aliquots at -80 °C. *R. sphaeroides* wild type and mutants were purified using the same methodology as *A. ferrooxidans* Rubisco.

Chimeric Rubisco of *O. sativa*/*Synechococcus* and *Syn*-Rubisco were purified using the same methodology as *Af*-Rubisco with additional anion exchange step after the size exclusion column. The concentrated sample was applied to an anion exchange column, MonoQ 10/100 GL column (GE Healthcare) which was pre-equilibrated with buffer A. Protein was eluted with a linear salt gradient up to 0.5 M NaCl and the fractions were analyzed using SDS-PAGE. Fractions of pure Rubisco were pooled and concentrated using Amicon Ultra-4 (100 kDa cut-off). 5% (v/v) of glycerol was added to the concentrated Rubisco and it was stored at -80 °C.

2.5.5 Activase purification

Rubisco activase from *A. tequila* was purified as previously described by Shivhare et al. (2017).

2.5.6 Protein concentration determination by absorbance at 280 nm

Purified proteins were quantified by their absorbance at 280 nm measured by NanoDrop and their extinction coefficient. The extinction coefficient was calculated using the program ProtParam(Gasteiger et al., 2005). The final concentration of the protein was calculated according to the method of Pace et al. (1995) using the following equation:

$$\text{Concentration (mg/mL)} = \left(\frac{\text{Abs}}{\epsilon} \right) \times \text{Mw} \quad \text{Equation 2.1}$$

Where Abs is the absorbance measured by NanoDrop at 280 nm, ϵ is the extinction coefficient $\text{M}^{-1}\text{cm}^{-1}$ and Mw is the molecular weight.

2.5.7 Protein concentration determination by Bradford assay

The soluble protein contents from *E. coli* were measured using Pierce Coomassie Plus (Bradford) Protein Assay Kit (Thermo Scientific) as per the manufacturer's recommendations. This was compared against standards of known amounts of Bovine Serum Albumin (BSA). The assays were performed in a 96 well plate with 150 μL of standards containing 0, 0.75, 1.5 or 2.25 μg BSA. Protein sample was prepared by using 149 μL H_2O and 1 μL of protein. 150 μL of Bradford reagent was added to each well and incubated at room temperature for 5 minutes before measuring the absorbance at 600 nm using a Multiscan RC microtitre plate reader (Labsystems).

2.6 Enzymatic assay

Enzymatic activity of Rubisco can be determined using either spectrophotometric coupled-enzyme assay or radioisotopic activity assay. In the spectrophotometric assay, the production of 3PGA by Rubisco is coupled with the oxidization of NADH (Figure 2-3) and this is monitor by reduction at 340 nm using a UV-1800 spectrophotometer (Shimadzu). Alternatively, radioisotopic labeled $^{14}\text{CO}_2$ and ^{14}C -CABP are used in the measurement of the Rubisco content and catalysis from the total cellular protein extracts from *E. coli*. The radioisotopic Rubisco assay measurements were carried out in the laboratory of Associate Professor Spencer Whitney in Australia National University (ANU, Australia).

2.6.1 RuBP and CABP synthesis

Commercially available RuBP contains sugar phosphate inhibitors which affect Rubisco catalysis. Thus, RuBP was synthesized from R5P using PRK as previously described by Kane et al. (1998). Unlabeled and ^{14}C labeled carboxypentitol-1,5-bisphosphate (CPBP) was synthesized by reacting unlabeled and Carboxy- ^{14}C -labeled KCN was reacted with RuBP as described by (Pierce et al., 1980). The purified CPBP is an isomeric mixture of 2-carboxy-D-ribitol-1, 5-bisphosphate (CRBP) and CABP. CABP binds essentially irreversibly to Rubisco while CRBP does not bind tightly (Pierce et al., 1980; Schloss, 1988). The specific activity of ^{14}C -CABP was equivalent to that of the K^{14}CN used (2.04 GBq/mmol).

2.6.2 Spectrophotometric Rubisco assay

Rubisco activation assay for purified Rubisco was performed spectrophotometrically at 25°C as previously described by (Tsai et al., 2015). The assay was performed in a 100 μL reaction containing 100 mM Tricine-NaOH, pH 8.0, 5 mM MgCl_2 , with 3 μL of coupling enzymes mixture (creatine P-kinase (2.5 U/mL), glyceraldehyde-3-P dehydrogenase (2.5 units/mL), 3-phosphoglycerate kinase and triose-P isomerase/glycerol-3-P dehydrogenase (20/2 units/mL)), 20mM NaHCO_3 , 0.5mM NADH, 1mM ATP/DTT/RuBP, 10mM creatine-P, and Rubisco (ECM, ER, EX) as stated in the figure legend. The oxidation of NADH is coupled with the generation of 3PGA and this is monitor by the reduction in absorbance at 340 nm on a spectrophotometer.

Activated Rubisco complex (ECM) was obtained by incubating Rubisco (10-20 μM of active sites) with buffer A supplemented with 40 mM NaHCO_3 and 10 mM MgCl_2 in for 10 minutes at room 25 °C. Inhibited Rubisco complexes with RuBP (ER) or XuBP (EX) was obtained by incubating Rubisco with buffer A supplemented with 4 mM EDTA for 10 minutes followed by addition of either 0.8 mM RuBP or 2.4 mM XuBP. Rubisco activation assay using soluble *E. coli* extract was carried out in a similar fashion as purified Rubisco and the amount of Rubisco used for formation of ECM and ER complexes was calculated based on the OD of the *E. coli* culture (Refer to section 4.2).

The ECM and ER activity can be calculated from 3PGA production versus time curve. The 3PGA produced at any time interval is calculated using the following equation:

$$3PGA(\mu M) = -\frac{(dAbs(T)-dAbs(0))/0.00622}{2} \quad \text{Equation 2.2}$$

Where dAbs(T) and dAbs(0) are the absorbance values at time T and zero, 0.00622 is the extinction coefficient of NADH in $\mu M^{-1} cm^{-1}$. The carboxylation of RuBP produces two molecules of 3PGA and the coupled enzyme assay consumes four molecules of NADH per molecule of RuBP consumed. Each molecule of 3PGA produced is coupled to two molecules of NADH consume. The carboxylation rate of Rubisco can be calculated from using RuBP consumption rate where one molecule of RuBP consumed is half of 3PGA produced (Equation 2-2).

$$ECM \text{ rate}(s^{-1}) = \frac{RuBP \text{ consumption rate}}{Concentration \text{ of } ECM} \quad \text{Equation 2.3}$$

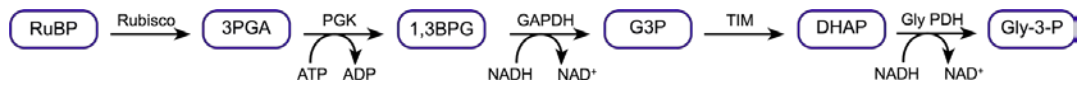


Figure 2-3: Spectrophotometric Rubisco assay.

Rubisco catalyzed carboxylation of RuBP to produce two molecules of 3-Phosphoglycerate (3PGA). 3PGA was phosphorylated to 1, 3 Bisphosphoglycerate (1,3BPG) by 3PGA kinase (PGK). Glyceraldehyde-3-P dehydrogenase (GAPDH) reduces the 1,3BPG to glyceraldehyde 3-phosphate (G3P). Triose-phosphate isomerase (TIM) produces dihydroxyacetone (DHAP) from G3P. DHAP was reduced by glycerol-P dehydrogenase (Gly PDH) to Glycerol 3-phosphate (Gly-3-P). Rubisco produces two molecules of 3PGA for each RuBP carboxylated; four NADH are oxidized per RuBP and oxidation of NADH leads to a decrease in absorbance at 340 nm.

2.6.2.1 Kinetics of RuBP-inhibited Rubisco complexes

The rate of inhibitor release from RuBP-inhibited Rubisco complexes can be estimated using the product accumulation over time which is calculated using RuBP consumed divided by the concentration of Rubisco. This product accumulation versus time was fitted to Equation 2-3 estimate values for initial velocity (v_i), final velocity (v_f) and first-order rate constant (k_{obs}) using non-linear regression software (OriginLab). The k_{obs} value provides an estimate on the rate of RuBP release from tightly bound inhibited complex (Pearce, 2006; Pearce et al., 2003).

$$[product] = v_f \cdot t + ((v_i - v_f)(1 - e^{-k_{obs} \cdot t}))/k_{obs} \quad \text{Equation 2.4}$$

2.6.3 Radioisotopic Rubisco activity assay

For Rubisco measurement using the radioisotopic methodology, the pTrcHisB plasmids encoding for wild type and Rubisco variants from *A. ferrooxidans* and *R. sphaeroides* Rubisco were transformed into *E. coli* XL1-Blue. Rubisco expression was induced with 1 mM IPTG at 28 °C and cells were harvested after 6 hours. Cell pellets were snap freeze with liquid nitrogen prior to storage at -80 °C. Soluble protein was obtained following lysis by French pressure cell (140 mPa) using extraction buffer (50 mM EPPEs-NaOH, pH 8.0, 15 mM MgCl₂, 1 mM EDTA, 1 mM PMSF, 5 mM DTT).

2.6.3.1 Rubisco activation by CO₂-Mg²⁺

Total protein extract was obtained following lysis by French pressure cell (140 mPa) using extraction buffer (50 mM EPPEs-NaOH, pH 8.0, 15 mM MgCl₂, 1 mM EDTA, 1 mM PMSF, 5 mM DTT). The total protein extract was centrifuged for 3 minutes, 4 °C at 15000 g and the supernatant was mixed with an equal volume of assay buffer (50 mM EPPEs-NaOH, pH 8.0, 15 mM MgCl₂, 1 mM EDTA) supplemented with 40mM NaH¹⁴CO₃. The samples were incubated at 25 °C for 10 minutes to activate Rubisco which will be used for assaying Rubisco activity and measuring the active site content by ¹⁴CABP binding.

2.6.3.2 Rubisco activity measurement using ¹⁴CO₂ incorporation

The specific activity of Rubisco was determined from the rate of RuBP-dependent ¹⁴CO₂ fixation into acid-stable products. The ¹⁴CO₂ fixation assays (0.5 mL total volume) were performed in 7.7 mL septum-capped glass scintillation vials with assay buffer containing 10 µg/mL carbonic anhydrase and saturating amount of RuBP (0.4 mM). All the assay components were equilibrated in CO₂-free air (*i.e.* 20.9% (v/v) O₂ in N₂) prior to adding a series of five different amounts of NaH¹⁴CO₃. The final ¹⁴CO₂ concentrations were 0 to 200 µM for both *A. ferrooxidans* and *R. sphaeroides* Rubisco. The assays were initiated by adding 20µL of ¹⁴CO₂-activated *E. coli* soluble protein into the vial. The significant carryover of O₂ in the Rubisco sample was ignored while the carryover CO₂ was considered when calculating the final CO₂ and changes in the specific activity of the NaH¹⁴CO₃. RuBP-independent ¹⁴CO₂ fixation control assays were run for each protein sample and contained H₂O in place of RuBP. The assays were stopped with 200 µL of 20% (v/v) formic acid after 2 minutes and dried at 90°C using a heat block. 500

μL of H₂O was added to each vial followed by 1 mL of Ultima-Gold scintillant (PerkinElmer) before quantification using a Tri-carb 4910TR scintillation counter (PerkinElmer). The carboxylase activity between the technical repeats varying by <2% confirmed that the extracted Rubisco was fully activated and stable.

The specific activity of each NaH¹⁴CO₃ stock was measured in assays (n=4) containing the highest ¹⁴CO₂ concentration and 5 nmol of RuBP. These assays were allowed to react for 90 minutes to ensure complete RuBP fixation. After acid treatment and drying, the scintillation counter ¹⁴C values were divided by 5 to derive the specific activity value – which varied between 1500 to 1800 cpm per nmol CO₂ fixed.

The CO₂ concentration in the assays were calculated using the Henderson-Hasselbalch derived equation

$$[CO_2] = \frac{(C_t)}{1 + \frac{V}{vqRT} + 10^{(pH-pK_1)} + 10^{(2pH-pK_1-pK_2)}} \quad \text{Equation 2.5}$$

where C_t is the total NaHCO₃ concentration (including the 0.4 μmol NaH¹⁴CO₃ in the *E. coli* soluble protein), V/v is the ratio of reaction vial headspace (V) to assay volume (v), q is the CO₂ solubility at 1 atm (0.0329 Mol.L⁻¹.atm⁻¹ at 25°C), R is the universal gas constant (0.082057 L.atm.K⁻¹.mol⁻¹), T is the assay temperature (25°C = 298K), pK₁ and pK₂ are the dissociation constants for carbonic acid (6.251 and 10.329).

The kinetic parameters were obtained by fitting the data to the Michaelis-Menten equation.

$$V_c = \frac{V_c^{max}C}{C + K_c} \quad \text{Equation 2.6}$$

Where V_c is the carboxylation rate, V_c^{max} is the maximal rate of carboxylation and C is the CO₂ concentration and K_c is the apparent Michaelis-Menten constant (K_m) for CO₂ under air levels of O₂ (K_c^{21%O₂}).

The carboxylation rate for each Rubisco (k_{cat}^c) was calculated by dividing V_c^{max} by the number of Rubisco-catalytic sites in the *E. coli* protein quantified by ¹⁴C-CABP.

2.6.3.3 Quantification of Rubisco content using ^{14}C CABP binding

CABP is able to bind tightly and specifically to the active site of Rubisco which allows for quantification of active sites in cellular protein extract. 1 μL of 2.2 mM ^{14}C -CPBP (containing both ^{14}C -CABP and ^{14}C -CRBP) was added to the soluble extracts containing activated Rubisco and the samples were incubated at room temperature for 10 minutes to allow binding of ^{14}C CABP to the active sites. The ^{14}C CABP bound Rubiscos were separated by size exclusion chromatography through Sephadex G50 resin (fine) which was pre-equilibrated with Sephadex column buffer (20mM EPPS-NaOH, 75mM NaCl, pH 8.0). The fractions were eluted using Sephadex column buffer and collected into 6 mL plastic scintillation vials. Two volumes of Ultima-Gold scintillant were added each vial and the Rubisco active site content in the ^{14}C CABP bound Rubisco fractions were then quantified using the scintillation counter.

To determine Rubisco expression as a percentage of the cellular soluble protein (% t.s.p.) in *E. coli* extracts, the mass of Rubisco was calculated by multiplying the molar active site concentration determined by ^{14}C -CABP binding by the molecular weight of a RbcL-RbcS pair for the *A. ferrooxidans* or *R. sphaeroides* enzyme (65 kDa) and then dividing the result by the total protein concentration in the extract. The soluble protein concentration in each soluble *E. coli* extract was quantified using a Coomassie dye binding assay against BSA.

2.6.3.4 Determination of $K_m(\text{RuBP})$

The apparent Michaelis constant for RuBP was measured using $^{14}\text{CO}_2$ -incorporation assays with a series of five RuBP concentrations (0-80 μM). Soluble *E. coli* extract of Rubisco was activated for 10 minutes in assay buffer containing 40mM $\text{NaH}^{14}\text{CO}_3$. Each reaction vial contains assay buffer with 20mM $\text{NaH}^{14}\text{CO}_3$ and different RuBP concentration. The reaction was initiated with the addition of 200 μL of $^{14}\text{CO}_2$ -activated soluble *E. coli* extract to the reaction vials. The reactions were stopped with 200 μL of 20% formic acid after 30 seconds and dried. 500 μL of H_2O was added to each vial followed by 1 mL of Ultima Gold-XR scintillant before quantification using a scintillation counter. The apparent $K_m(\text{RuBP})$ was determined by fitting the obtained data to Equation 2.3 whereby the C stands for the concentration of RuBP in the assay and K_C stands for the $K_m(\text{RuBP})$.

2.6.3.5 Rubisco CO₂/O₂ specificity ($S_{C/O}$)

Rubisco CO₂/O₂ specificity ($S_{C/O}$) measurements were made as described by Kane et al. (1994) using recombinant purified Rubisco from anion exchange column, Bio-Scale Mini Macro-Prep High Q column (Bio-Rad), followed by size exclusion column, Superdex 200 10/300 (GE Healthcare). The measurements were made using 20 mL septum-capped glass scintillation vials containing 1 mL of specificity buffer (30 mM Triethanolamine pH 8.3, 10 mM MgSO₄), 10 µg/mL of carbonic anhydrase and purified Rubisco. The vial is moved into a 25 °C shaking water bath and it was equilibrated for 1 hour with a gas mixture of 0.05% (v/v) CO₂ and 99.6 - 99.95% (v/v) O₂ accurately mixed using Wostoff gas mixing pumps. After 1 hour, the reaction was initiated by addition of 1 nmol of [1-³H]-RuBP (10 MBq/nmol), then 10U of alkaline phosphatase (Sigma) was added 30 to 45 minutes later. The alkaline phosphatase will dephosphorylate the RuBP products, ³H-2-phosphoglycolate and ³H-2-phosphoglycerate, to form ³H-glycolate and ³H-glycerate respectively. The reactions were passed through 0.4 mL bed volume of anion exchange AG1-X8 (Formate) resin and washed with 3 mL of milliQ-water. ³H-glycolate and ³H-glycerate will be bound to the resin and were eluted by 0.5 mL of 10% (v/v) H₂SO₄. Proteins were removed from these products using a 10 kDa cut-off UltraFree-MC centrifugal filter device (Millipore) by centrifugation for 2 hours at 5000 g. 50-100 µL of flow through was applied to an equilibrated HPX-87H column (Bio-Rad) to separate the glycerate and glycolate. 0.015N H₂SO₄ was used as the mobile phase, with a flow rate of 0.4 mL/minute and at a temperature of 65 °C. Fractions were collected onto a 96 well plate containing 200 µL of Ultima-Gold scintillant and the relative levels of ³H in the glycerate and glycolate peaks measured in a scintillation counter. Two HPLC runs were performed per sample. The specificity factor, $S_{C/O}$, was calculated using the following equation:

$$S_{C/O} = (R_{glycerate/glycolate} \times \frac{M_{O_2}}{M_{CO_2}}) \times 0.037 \quad \text{Equation 2.7}$$

Where M_{O_2} and M_{CO_2} are the mole fractions of O₂ and CO₂, respectively, in the gas mixture and 0.037 is the ratio between the aqueous solubilities of O₂ and CO₂ at 25 °C and $R_{glycerate/glycolate}$ is the ratio of ³H counts in the separated glycerate to glycolate peak.

2.6.3.6 Statistical Analysis

The statistical significance of observed differences between biological samples under the defined experimental conditions was evaluated by a 2-tailed student's T-test.

3 DIRECTING THE EVOLUTION OF RUBISCO IN THE PRESENCE AND ABSENCE OF ITS ACTIVASE IN MM1 *E. COLI*

Rubisco catalyzes the rate limiting carbon fixation step in the CBB cycle which is the major pathway for inorganic carbon entry into the biosphere. Despite its essential role, Rubisco is infamous for its sluggish kinetics and poor substrate discrimination. This makes Rubisco a long-standing target for improvements to enhance the photosynthetic efficiency and crop yield in plants. From the half-century of Rubisco research, numerous structures of different Rubisco from diverse sources show that all bona fide Rubiscos are comprised of large subunit dimers that are assembled in varying stoichiometries. These dimers have highly conserved overall three-dimensional structure with strictly conserved active site residues. Despite the understanding of Rubisco's structure and catalysis mechanism, we are still unable to improve Rubisco via rational design. This has led to the development of artificial directed evolution methodologies to screen for Rubisco with improvements properties such as catalysis. Among the different directed evolution of Rubisco system, the high throughput MM1-*prk* Rubisco-dependent E. *coli* (RDE) system has been applied to artificially evolve of three forms of Rubisco with some mutants obtained showing improvements in kinetic properties (Duraio et al., 2015; Mueller-Cajar et al., 2007; Wilson et al., 2016). The MM1 *E. coli* is a genetically modified strain that contains a *gapA* deletion and this results in the blockage of the glycolysis/glycogenesis pathway due to the deletion of the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH). This blockage can be bridged using a metabolic shunt consisting of Rubisco and phosphoribulokinase (PRK). PRK catalyze the production of RuBP, however, RuBP accumulation is toxic to *E. coli* (Hudson et al., 1992; Parikh et al., 2006). The presence of a functional Rubisco is required to alleviate the RuBP toxicity by carboxylation of RuBP to form 3-phosphoglycerate (3PGA), therefore, the giving rise to Rubisco-dependency. In addition, the production of 3PGA can be utilized for production of pyruvate thereby circumventing the blockage and allowing cell growth (Mueller-Cajar et al., 2007) (Figure 3-1A, blue arrows).

One aspect that has not been explored in the previous directed evolution of Rubisco studies concerning the topography of the evolutionary landscape for Rubiscos that form inhibited complexes with sugar phosphates. In the previous MM1-*prk* RDE studies, the

Rubiscos evolved were not prone to inhibition by sugar phosphate such as *R. rubrum* Rubisco and *Synechococcus* PCC 6301 Rubisco (Andrews et al., 1981; Durao et al., 2015; Mueller-Cajar et al., 2007; Mueller-Cajar et al., 2008b; Whitman et al., 1979). In addition, these Rubiscos are also not known to be associated with any activase. However, many Rubiscos, including those from higher plants, are known to be prone to inhibition by sugar phosphates and these Rubiscos require assistance from their Rubisco activase to facilitate removal of the inhibitors. It appeared likely that the expression of these Rubiscos alone in MM1-*prk* will not be able to support growth as they will become dead-end inhibited by RuBP. The co-expression of an activase with these Rubiscos could potentially promote survival as the activase can facilitate inhibitor release which would allow Rubisco to carry out RuBP catalysis (Figure 3-1A). Therefore, it is possible for the MM1-*prk* selection system to screen for both Rubisco and activase function, resulting in a Rubisco and activase dependent E. coli (RADE). The screening conditions may need to be optimized to achieve both Rubisco and activase dependency in the RADE system. Although the two different selection systems may require different selection conditions, the experimental setup for both RDE and RADE screening is similar (Figure 3-1B). In brief, randomly mutated Rubisco genes were generated using error-prone PCR and these were ligated into either empty pTrcHisB vector or pTrcDuet containing an additional Rubisco activase gene. These vectors were subsequently transformed into MM1-*prk* cells and screened using RDE or RADE selection systems. Faster growing colonies were inoculated and their plasmids were sequenced. Reoccurring mutations acquired from these screenings were characterized as these were likely to be true positives and thus conferring increased fitness to the colonies.

In this study, two class of Rubisco and activase were chosen to be evolved in the RDE and RADE systems. They are the red-type form IC Rubisco from *R. sphaeroides* (Rs-Rubisco, RsLS) and its activase (RsCbbX) and Rubisco from *A. ferrooxidans* (both green-type form IAq and form II, Af-Rubisco) and its activase (AfCbbQO). These Rubiscos have been shown to form inhibited complexes with RuBP and their activases are able to facilitate RuBP removal (Mueller-Cajar et al., 2011; Tsai et al., 2015). These Rubiscos can be also evolved in the absence of their activase using the RDE selection system and this will provide insights into the evolutionary pathways that Rubisco can undergo to lose its dependency on its activase. This information will be valuable in the future directed evolution of higher plant Rubisco which is also known to form inhibited

complexes with sugar phosphates. Higher plants also contain Rubisco activase and it has been shown that the activase regulate Rubisco's activity according to the photosynthetic requirements of the plants. The *Rs*-Rubisco and *Af*-Rubisco can also be evolved in the presence of their activase using the RADE selection system. Findings from the RADE screening could provide insight into the Rubisco and activase protein-protein interactions and success of these screenings could guide further studies towards generating a compatibility between different types of Rubisco and activase. Currently, there are three known types of Rubisco activase, green type (plant's Rubisco activase), red-type (*RsCbbX*) and *CbbQO* (*AfCbbQO*). These activases have all emerged independently via convergent evolution and carry out the same function of remodeling of inhibited Rubisco complexes via ATP hydrolysis (Mueller-Cajar, 2017). Despite the highly conserved structure of different Rubisco and the different activase performing the same function, the different Rubisco and activase are not compatible with each other. It is possible that the incompatible pairs of Rubisco and activase can be evolved using the RADE selection system to become compatible (Figure 3-1C). This will allow Rubisco to harness the diverse properties of different activases such as thermostability. Plant Rubisco is thermolabile and it is believed to be responsible for the loss of photosynthesis in plants caused by the inhibition of Rubisco at elevated temperatures (Salvucci et al., 2004). If plant Rubiscos were compatible with thermostable Rubisco activase from the red alga, *C. merolae*, the transplanting of these activases into plants would possibly increase the heat tolerance of photosynthesis (Loganathan et al., 2016). In addition, the generation of compatibility between different Rubisco and activase will also provide insights into their interaction surfaces and potentially discover new protein-protein interactions.

In this chapter, we outline the experiments performed towards establishing Rubisco and activase dependent *E. coli* growth conditions. This includes substantial difficulties that had been identified and overcome before robust selection systems were obtained.

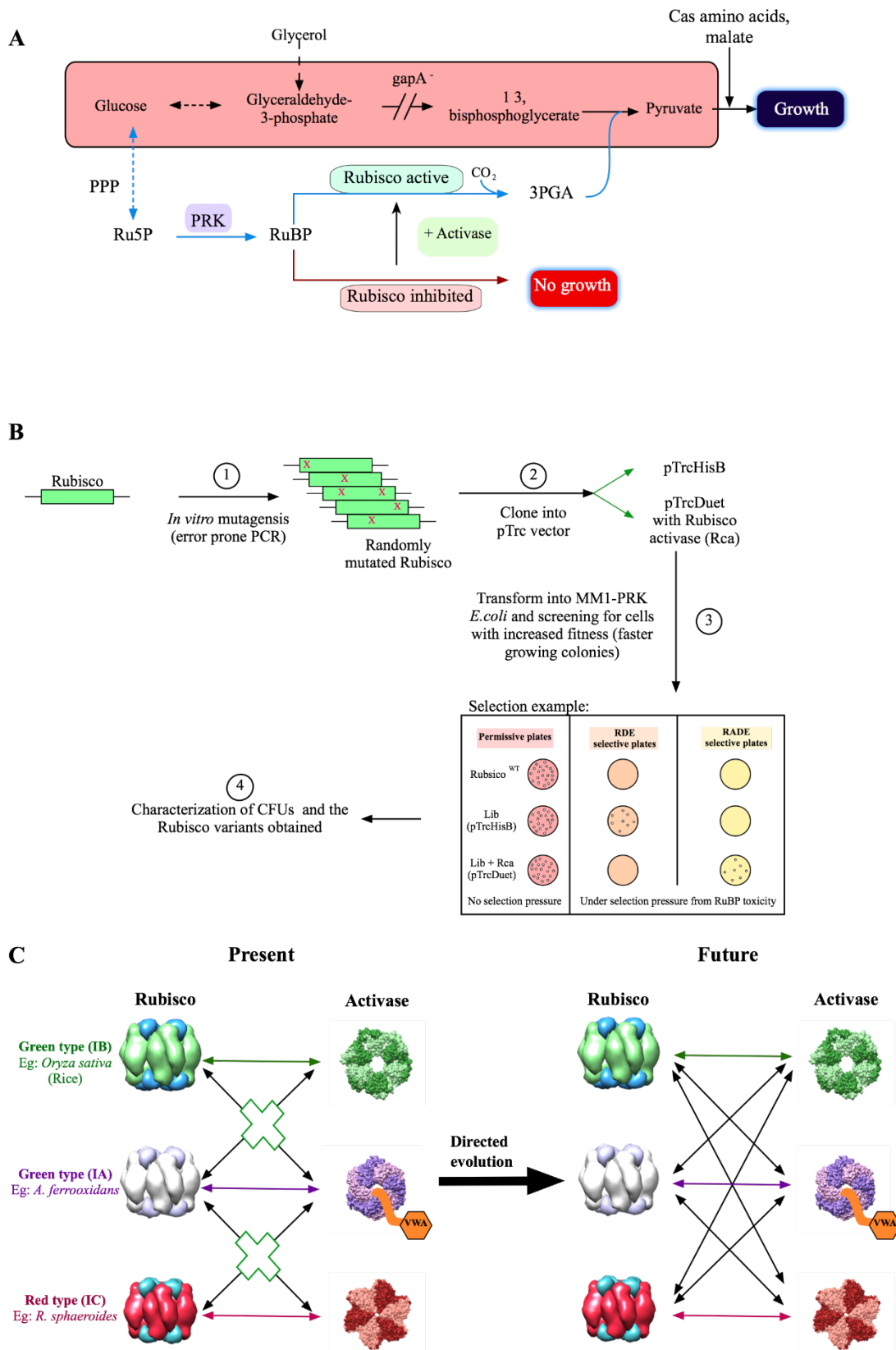


Figure 3-1: Overview of MM1 *E. coli* screening system and compatibility between different Rubisco and activase.

(A) Directed evolution screen of Rubisco using MM1 *E. coli* cells. The concept is as described in the legend of Figure 1-8. Rubiscos which are prone to inhibition by sugar phosphate inhibitors, such as RuBP, can be remodeled by Rubisco activase and allow Rubisco to become functional for RuBP catalysis (dotted box). In this metabolic shunt, the expression of PRK is regulated using arabinose while Rubisco and activase are regulated using IPTG. (+): RuBP turnover, (-): No RuBP turnover. This diagram was modified from Wilson et al. (2017b). (B) Experimental setup for directing the evolution of Rubisco in the presence and absence of activase. The target Rubisco genes (large and small subunit) were randomly mutated using error-prone PCR (step 1). These mutated Rubisco genes were cloned into either pTrcHisB vector or pTrcDuet vector containing an activase gene (step 2). These mutant library vectors were transformed into MM1-pAC^{BAD}PRK and screened using RDE and RADE selective plates (step 3). The survival in RDE requires the presence of a functional Rubisco while survival in RADE requires the presence of both Rubisco and activase which are functional and compatible. These constructs were also plated onto the permissive plate where there is no selective pressure (PRK is not induced) and the media is rich. Faster growing colonies are associated with increased fitness and these colonies were inoculated and the plasmids were sequenced (step 4). Reoccurring Rubisco variants obtained were characterized. This figure was modified from Durao et al. (2015). (C) The potential outcomes of RADE screening system whereby compatibility between different Rubisco and activase can be achieved. Currently, the different types of Rubisco are only compatible with their own activase. It is possible to artificially evolve the different Rubisco and activase to become compatible with each other using RADE screening system where both Rubisco and activase are required for survival.

3.1 Three different PRK enzymes were used in establishing RDE and RADE selection system

In this study, three different PRK-encoding genes were used in the establishment of the RDE and RADE selection systems. The first gene was from *Synechococcus* sp. PCC 7942 PRK (PRK⁷⁹⁴²) which was used in previous RDE studies and it was cloned in a pAC^{BAD} vector where the PRK expression was under the tightly controlled BAD promoter (Mueller-Cajar et al., 2007; Wilson et al., 2017b). The other two genes were from the common laboratory cyanobacterial strains, *Synechocystis* PCC 6803 (PRK⁶⁸⁰³) and *Synechococcus* PCC 7002 (PRK⁷⁰⁰²). These three PRK genes shared ~80% sequence identity between them (Figure 3-2). The three PRK-encoding genes were used in the different stages of the development of the RDE and RADE in this study, as discussed in detail below.

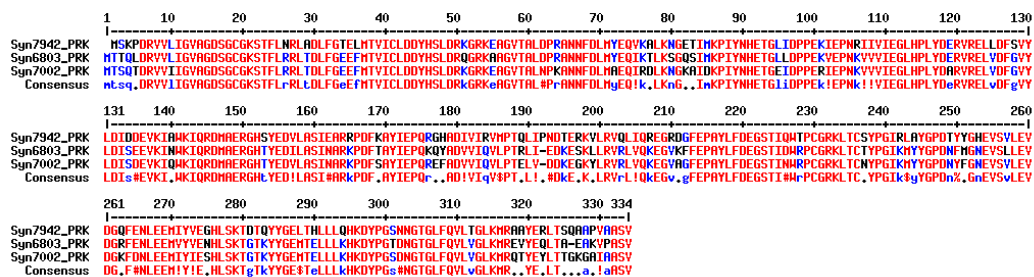


Figure 3-2: Sequence alignment of the PRK from three different cyanobacteria, *Synechococcus* PCC 7942/6301, *Synechocystis* PCC 6803 and *Synechococcus* PCC 7002.

3.1.1 Reconstruction of the PRK⁷⁹⁴² containing plasmid

While re-establishing the RDE system, we found that the existing copy of the PRK-encoding plasmid pAC^{BAD}PRK in the laboratory contained an *E. coli* transposon gene (Tn1000) within the *prkA* gene. The Tn1000 was inserted at position M290 which led a frameshift and resulted in a 29 amino acid truncation in the C-terminus of PRK. The truncated PRK is functional, as it is routinely used by our laboratory to produce RuBP enzymatically *in vitro* from ribulose-5-phosphate. However, it is likely to have lower catalytic activity and hence toxicity in *E. coli*, which we assume is the reason it was selected during plasmid propagation. Indeed, during the initial development of the RDE system, Tn1000, IS2 and IS10 insertions into *prkA* genes were frequently observed in an *E. coli* XL1-Blue background (Mueller-Cajar, 2006; Wilson et al., 2017b). Hence, when working with this and related systems, it is important to periodically monitor the integrity of the *prkA* encoding plasmid.

Despite having activity, the presence of Tn1000 within PRK might interfere with future screenings and was undesirable. Thus, Tn1000 was removed via PCR, amplifying all the elements within the plasmid except for Tn1000. The primer, 3'PRK+P (described in 2.2.3), was designed to contain a 5' phosphate group to allow ligation of the PCR product to form a circular plasmid. Multiple attempts failed to recover the type *prkA* sequence. Finally, we settled for a double mutant *prkA* (PRK^{DM}) which contained mutation resulting in E253G and G303D substitutions and the final plasmid was designated as pAC^{BAD}PRK^{DM}. We reasoned that due to the high specific activity of PRK (230 U/mg), even if there was a decrease in activity in the double mutant, it would still be sufficient for screening purposes since *E. coli* appears unable to metabolize RuBP

without Rubisco (Kobayashi, 2003). Therefore, the pAC^{BAD}PRK^{DM} was used in the initial re-establishment of RDE.

3.1.2 Cloning of PRK⁶⁸⁰³ and PRK⁷⁰⁰² into the pAC^{BAD} vector

The *prkA* genes for PRK⁶⁸⁰³ and PRK⁷⁰⁰² were cloned from their respective genomic DNA and directly into the pAC^{BAD} vector using DH5 α *E. coli* competent cells. This approach was unsuccessful as the plasmids obtained contained various mutation in the *prkA* sequence. It is likely that these mutant PRKs were selected as they have a lower catalytic activity which allows the cells to alleviate RuBP toxicity. This result was similar to the reconstruction of wild type PRK⁷⁹⁴² where only mutants of PRK were obtained and the attempts to revert the mutations back to wild type using mutagenesis were not successful. MM1 *E. coli* cells were subsequently used for cloning instead of DH5 α cells and wild type *prkA* for both *Synechocystis* PCC 6803 and *Synechococcus* PCC 7002 were successfully cloned into pACBAD vector giving rise to pAC^{BAD}PRK⁶⁸⁰³ and pAC^{BAD}PRK⁷⁰⁰² respectively. It appears DH5 α is especially sensitive to RuBP toxicity for unknown reasons.

3.2 Protein expression of Rubisco and activase in pTrcHisB and pTrcDuet vector

In the establishment of RDE and RADE, various Rubisco and activase constructs were created and the expression of the proteins from these constructs were analyzed using *E. coli* BL21 cells. The Rubisco and activase were cloned either individually into pTrcHisB vector or co-expressed using the pTrcDuet vector. The expression of protein in both pTrcHisB and pTrcDuet are under the Trc promoter. The Rubisco and activase used in this study include form I *Synechococcus* PCC 6301 Rubisco (*Syn*-Rubisco), form I *Rs*-Rubisco (*RsLS*) and its activase (*RsCbbX*), form I *Af*-Rubisco (*AfLS*) and its activase (*AfQ1O1*) and lastly, form II *Af*-Rubisco (*AfM*) and its activase (*AfQ2O2*). The soluble expression tests of all these Rubisco and activase were carried out as shown in Figure 3-3. The expression of *SynLS* was not carried out as it has been demonstrated in a previous RDE study (Mueller-Cajar et al., 2008b).

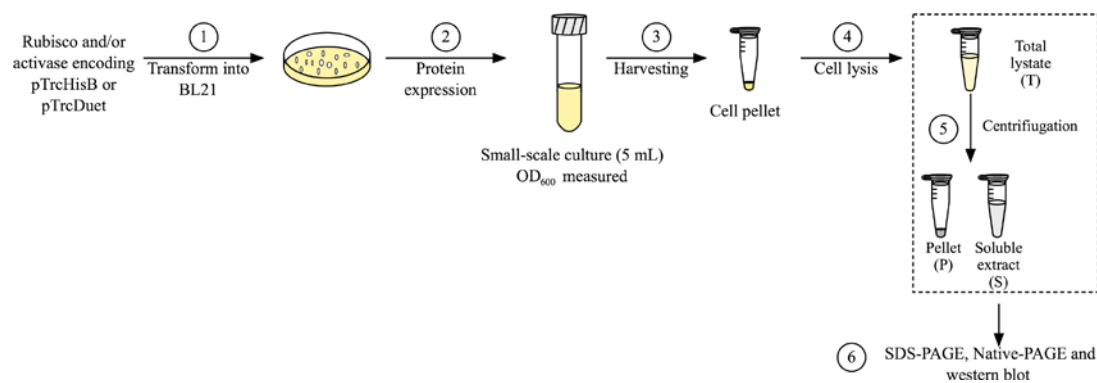


Figure 3-3: Overview of small-scale protein expression test for Rubisco and/or activase in pTrcHisB and pTrcDuet.

The plasmids encoding for Rubisco and/or activase in either pTrcHisB or pTrcDuet were transformed into BL21 *E. coli* cells and subsequently, a small-scale culture was grown (step 1). Expression of Rubisco and activase were induced using IPTG and the OD_{600} was measured prior to harvesting (step 2, 3). The harvested cell pellet can be either stored away or lysed using sonication to obtain total cell lysate (step 4). Some of the cell lysates were centrifuged to obtain the pellet and soluble extract (supernatant) (step 5). The total lysate, pellet, and soluble extract can be subsequently analyzed using SDS-PAGE, Native-PAGE in combination with immunoblotting (step 6). The amount of sample loaded will be calculated using the original OD_{600} of the small-scale culture, where $OD_{600}=1$ is approximately equivalent to 8×10^8 *E. coli* cells/ml.

3.2.1 Soluble protein expression of *Rs*-Rubisco and activase in pTrcHisB and pTrcDuet

Rs-Rubisco and its activase were cloned individually into pTrcHisB to obtain pTrcRsLS and pTrcRsCbbX and they were also cloned together into pTrcDuet to produce pTrcRsLSRsX (Figure 3-4A). The protein expression analysis using SDS-PAGE and western blot showed that all the various plasmids were able to produce soluble Rubisco and/or activase (Figure 3-4 B, C, D). It was also observed that there were also significant basal expressions of all the different protein expressed in the pTrcHisB or pTrcDuet. This basal expression was attributed to the leakiness of the Trc promoter.

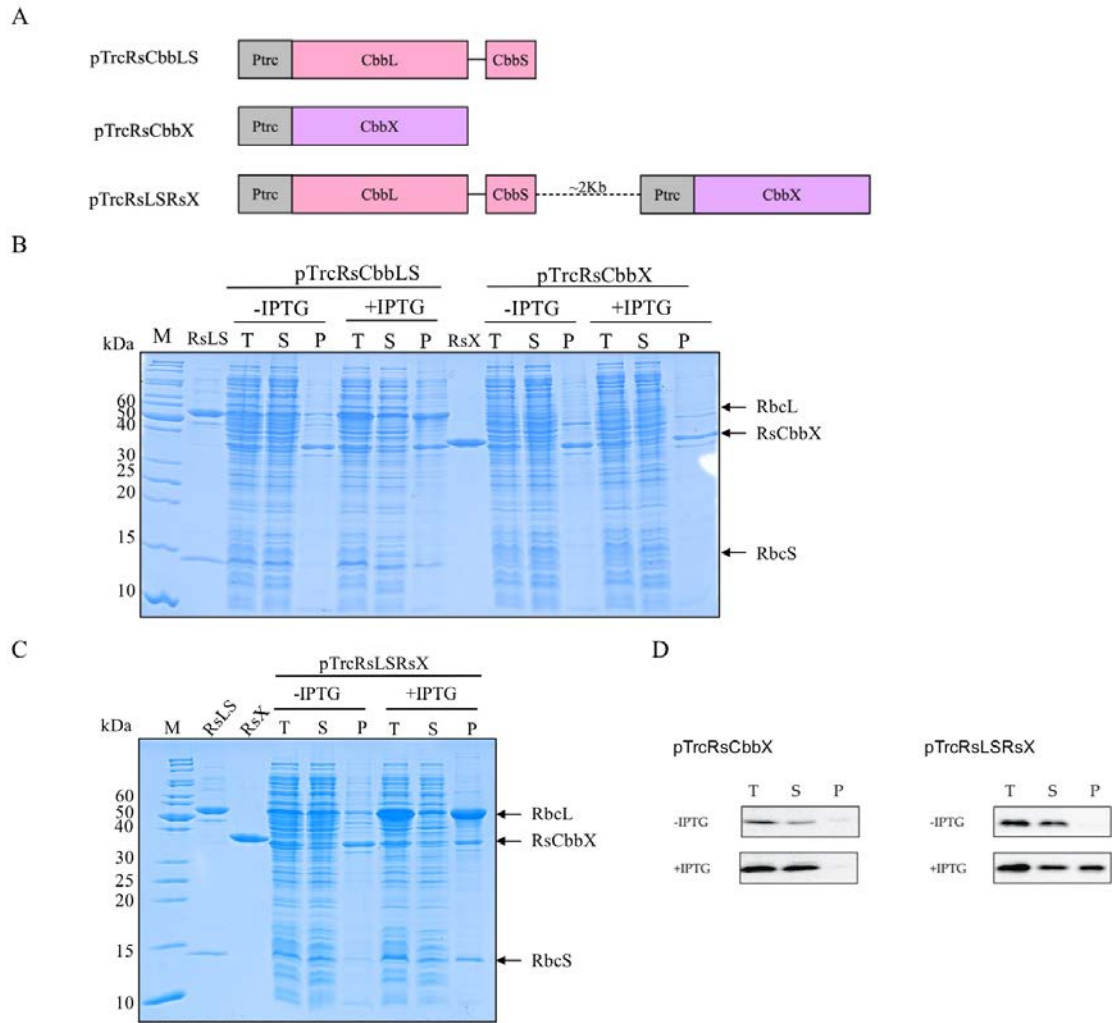


Figure 3-4: Soluble expression of *Rs*-Rubisco and its activase in pTrcHisB and pTrcDuet.

(A) *Rs*-Rubisco and activase gene organization in pTrcHisB and pTrcDuet. (B) Protein expression of *Rs*-Rubisco or activase from pTrcRsCbbLS and pTrcRsCbbX respectively. (C) Protein expression of *Rs*-Rubisco and its activase in pTrcDuet, pTrcRsLSRsX. (D) α -RsCbbX Western blot for detecting RsCbbX expression in pTrcRsCbbX and pTrcRsLSRsX. This western blot was performed on a different experiment as compared to the samples in the SDS-PAGE in C and B. $\sim 1 \mu\text{g}$ of purified *Rs*-Rubisco (RsLS) and activase (RsX) was loaded as controls along with the samples on a 16% SDS-PAGE. Expression of both Rubisco and activase was induced by 0.5 mM IPTG. The amount of *E. coli* extract loaded per lane was from 6.0×10^7 of *E. coli* BL21 cells. T: Total, S: Soluble, P: Pellet. +: Induced, -: Non-induced. The legend for all the diagrams will be consistent for the rest of this chapter.

3.2.2 Soluble expression of *Af*-Rubisco and activase in pTrcHisB and pTrcDuet

A. ferrooxidans Rubiscos and activases were cloned individually into the pTrcHisB vectors to obtain pTrcAfLS, pTrcAfM, pTrcAfQ1O1, pTrcAfQ2O2 and the respective Rubisco and activase pair were also cloned into the pTrcDuet to produce pTrcAfLSAfQ1O1 and pTrcAfMAfQ2O2 (Figure 3-5A, Figure 3-6A). Their protein expressions in these vectors were validated using small-scale protein expression tests. Soluble expression of Rubisco and activase was achieved in both the induced and uninduced conditions (Figure 3-5B, C, Figure 3-6B, C). The expression of AfCbbO proteins did not result in prominent SDS-PAGE bands in comparison to AfCbbQ proteins. This was likely due to the stoichiometric ratio of CbbQ to CbbO of 6:1 in the AfCbbQO activase complexes (Tsai et al., 2015).

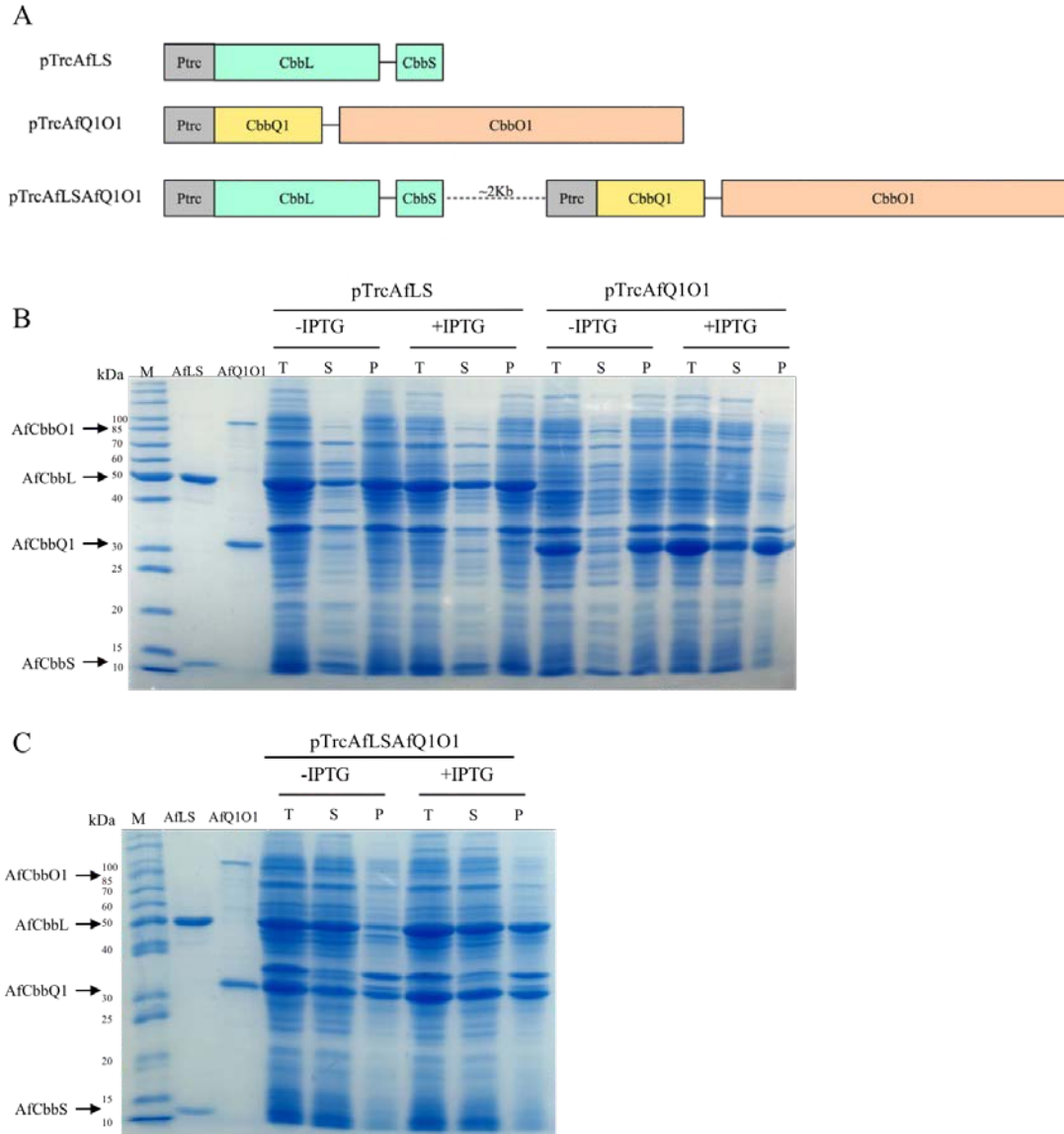


Figure 3-5: Protein expression of form I *Af-Rubisco* (AfLS) and its activase (AfQ1O1) in pTrcHisB and pTrcDuet vectors.

(A) *A. ferrooxidans* form I Rubisco and its activase gene organization in pTrcHisB and pTrcDuet. (B) Protein expression test of *A. ferrooxidans* form I Rubisco and its activase in pTrcAfLS and pTrcAfQ1O1 respectively. (C) Protein expression test of *A. ferrooxidans* form I Rubisco and its activase in pTrcAfLSAfQ1O1. ~1 µg of purified AfLS and AfQ1O1 were loaded as controls along with the samples on a 12.5% SDS-PAGE. Expression of both Rubisco and activase was induced by 0.5 mM IPTG. The amount of *E. coli* extract loaded per lane was from 6.0×10^7 of *E. coli* BL21 cells.

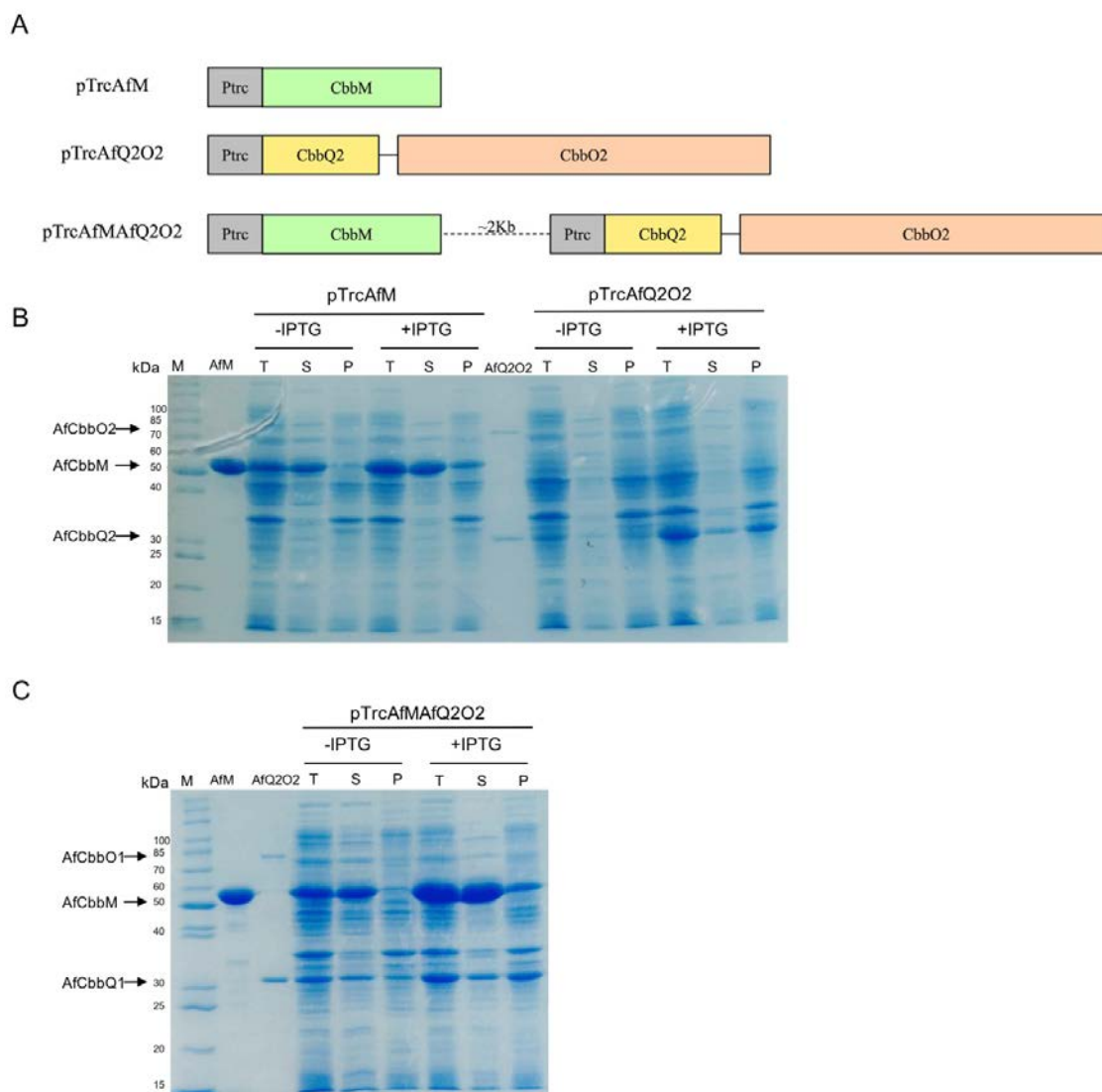


Figure 3-6: Protein expression of form II *Af*-Rubisco (AfM) and its activase (AfQ2O2) in pTrcHisB and pTrcDuet vector.

(A) *A. ferrooxidans* form II Rubisco and its activase gene organization in pTrcHisB and pTrcDuet. (B) Protein expression test of *A. ferrooxidans* form II Rubisco and its activase in pTrcAfM and pTrcAfQ2O2 respectively. (C) Protein expression test of *A. ferrooxidans* form II Rubisco and its activase in pTrcAfMAfQ2O2. ~1 µg of purified AfM and AfQ2O2 were loaded as controls along with the samples on a 12.5% SDS-PAGE. Expression of both Rubisco and activase was induced by 0.5 mM IPTG. The amount of *E. coli* extract loaded per lane was from 6.0×10^7 of *E. coli* BL21 cells.

3.3 Development of Rubisco and activase dependent *Escherichia coli* system

3.3.1 pAC^{BAD}PRK^{DM} had limited success in the development of RDE and RADE for the different Rubisco and activase

The RDE had to be re-established with the new pAC^{BAD}PRK^{DM} plasmid prior to the development of RADE. MM1-pAC^{BAD}PRK^{DM} was transformed with pTrcHisB or pTrcDuet plasmids encoding either a Rubisco and/or an activase. The *Syn*-Rubisco encoding plasmid, pTrcSynLS, functioned as a positive control as it had been previously used in RDE (Mueller-Cajar et al., 2008b). The plasmids encoding *Rs*-Rubisco, pTrcRsCbbLS, and pTrcDuet encoding both *Rs*-Rubisco and its activase, pTrcRsLSRsX, were examined for their fitness in RDE on selective media. The *R. sphaeroides* activase encoding plasmid, pTrcRsCbbX, served as a negative control as it does not catalyze carboxylation of RuBP. The transformants were plated on two types of media, permissive media and selective media. The permissive media enabled survival of MM1-pAC^{BAD}PRK^{DM} cells independent of Rubisco, as an additional downstream carbon source (casamino acid) was supplied and PRK was not induced, thus there was no selection pressure and little RuBP toxicity. In contrast, the selective media should only allow the survival of cells that expressed a functional Rubisco as PRK been induced and a limiting concentration of casamino acid was supplied (Figure 3-1B). The limiting amount of casamino acid was required in the selective media as the MM1 cells are a proline/leucine auxotrophic *E. coli* strain (Mueller-Cajar et al., 2007). Based on previous studies, selective media (M9 minimal media supplemented with 0.4% (v/v) glycerol, 0.05% (w/v) casamino acid, 0.5 mM IPTG) containing 0.05-0.075% (w/v) arabinose and an elevated CO₂ of 2% were required to achieve Rubisco-dependency (Duraio et al., 2015; Mueller-Cajar et al., 2008b). Therefore, these media conditions were used and the cells were grown either in a CO₂ incubator in air supplemented with 2% (v/v) CO₂ at 23°C or in air (0.04% CO₂) at room temperature.

The various plasmids were successfully transformed into MM1-pAC^{BAD}PRK^{DM} cells as shown on the corresponding permissive media plates (Figure 3-7A). At 2% CO₂, the selective media supplemented with 0.075% (w/v) arabinose had growth of MM1-pAC^{BAD}PRK^{DM} cells expressing either *Syn*-Rubisco or *Rs*-Rubisco (Figure 3-7B). The *Syn*-Rubisco had higher fitness when compared to *Rs*-Rubisco since more colonies had been established. *R. sphaeroides* activase alone was not able to support the growth of

the cells. A similar situation was observed when *R. sphaeroides* activase was co-expressed with its Rubisco (pTrcRsLSRsX), even though *Rs*-Rubisco alone was able to sustain the growth of MM1-pAC^{BAD}PRK^{DM} cells. This difference could be due to the increased metabolic burden on the MM1-pAC^{BAD}PRK^{DM} cells to overexpression three proteins (PRK, RsLS, and RsCbbX). Under atmospheric conditions (0.04% CO₂, will be referred to as air), there was no growth on the selective media. Similar results were observed on selective media containing 0.05% (w/v) arabinose. In summary, the Rubisco Dependent *E. coli* (RDE) system was re-established using pAC^{BAD}PRK^{DM}.

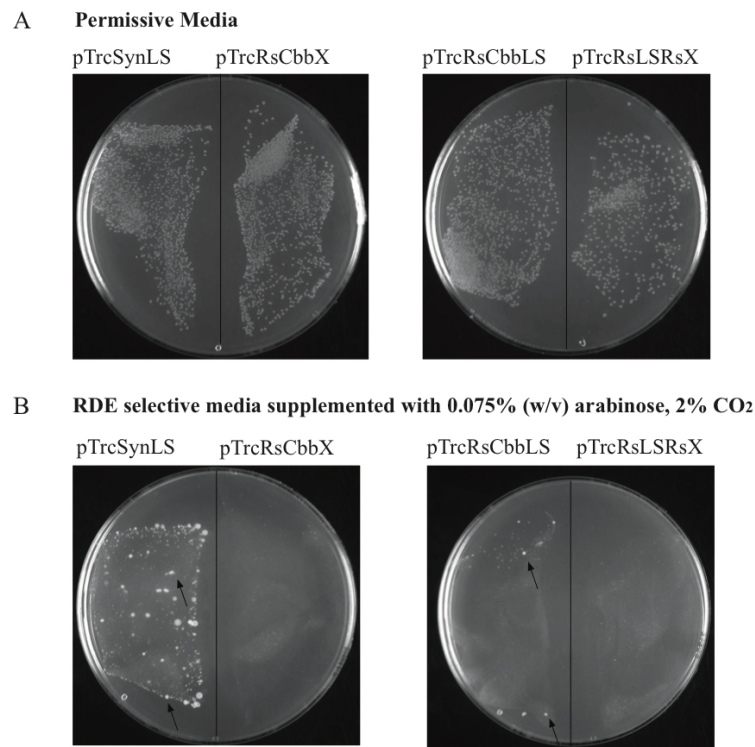


Figure 3-7: Representation of Rubisco-dependent *E. coli* selection using MM1-pAC^{BAD}PRK^{DM}

The different plasmids were transformed into MM1-pAC^{BAD}PRK^{DM} and plated either permissive or selective media. (A) Permissive media supplemented antibiotics (Amp, Chl, and Kan). Plates were incubated at 37 °C for 24 hours. (B) RDE selective media supplemented with 0.075% (w/v) arabinose and antibiotics (Amp, Chl, and Kan). Plates were incubated at 23 °C, in air supplemented with 2% CO₂ for 10 days. There were no colonies on pTrcRsCbbX and pTrcRsLSRsX. Arrows indicate examples of colonies on selective media.

3.3.1.1 Presence of functional and compatible pair of Rubisco and activase are required for survival in RADE selection

Towards achieving Rubisco and activase dependent conditions, a titration of various components of the selective media and CO₂ level was performed. Different combinations of selective media were prepared using varying concentrations of casamino acids, malate, IPTG and arabinose (Table 2-4). Out of the various combinations of selective media, the combinations that did not contain IPTG conferred survival advantage towards pTrcDuet containing *Rs*-Rubisco and activase (Figure 3-8, Appendix Figure A-4 to A-7). In the absence of IPTG, the addition of malate or higher percentage of casamino acid (> 0.05% (w/v)) allowed the growth of MM1-pAC^{BAD}PRK^{DM} independent of Rubisco (Appendix Figure A-4 to A-7). Therefore, the condition of Rubisco and activase dependency was found to be optimal on selective media composed of M9 minimal media supplemented with 0.4% (v/v) glycerol, 0.05% (w/v) casamino acid, antibiotics (Amp, Chl, and Kan) and a minimum of 0.05% (w/v) L-arabinose. This selective media will be termed as RADE selective media where there is a lack of IPTG.

The MM1-pAC^{BAD}PRK^{DM} co-expressing both wild type *Rs*-Rubisco and activase were able to grow in both atmospheric and elevated CO₂ (Figure 3-8, pTrcRsLSRsX). The differences in colony count in the two different RADE selective media (0.05% and 0.10% (w/v) arabinose) for cells harboring pTrcRsLSRsX were due to the differences in transformation efficiency between the batches of competent cells used. The growth of colonies under atmospheric condition was not achieved in previous studies, likely due to the kinetic properties (very high *K_C* values) of the Rubisco enzymes used (Wilson et al., 2017b). The leakiness of Trc promoter enabled basal protein expression of both Rubisco and activase and this would have alleviated the burden of overexpressing multiple proteins for MM1 cells. The expression of *Rs*-Rubisco and activase in MM1-pAC^{BAD}PRK^{DM} cells were designed to mimic the physiological condition in *R. sphaeroides*. The activase facilitates the removal of RuBP that bound to the apo-Rubisco and thus allows *Rs*-Rubisco to catalyze carboxylation of RuBP, thus permitting growth in MM1-*prk* cells (Figure 3-1A). In the absence of activase, RuBP inhibited *Rs*-Rubisco complexes were unable to catalyze RuBP turnover which leads to accumulation of toxic levels of RuBP in the cell. Thus, there was no colony growth for cells expressing *Rs*-

Rubisco (Figure 3-8A, pTrcRsLS). Similarly, there was also no growth for cells co-expressing *Rs*-Rubisco and nonfunctional activase (*RsCbbX*^{E138Q}). *RsCbbX*^{E138Q} is a Walker B mutant and it is unable to carry out ATP hydrolysis which is required during the remodeling of inhibited Rubisco complex (Mueller-Cajar et al., 2011). This result demonstrated the presence of a functional activase is required for MM1-pAC^{BAD}PRK^{DM} cells expressing *Rs*-Rubisco to survive in the RADE system.

It was also observed in the RADE screening at 2% CO₂, MM1-pAC^{BAD}PRK^{DM} cells expressing *Syn*-Rubisco were able to support growth in contrast to cells expressing *Rs*-Rubisco. This difference in fitness could be the different propensity of these two Rubiscos in forming inhibited complexes with RuBP. *Syn*-Rubisco does not form inhibited complexes with RuBP and therefore, the basal expression of *Syn*-Rubisco was still able to support growth as these Rubiscos remained functional in comparison to the inhibited *Rs*-Rubisco.

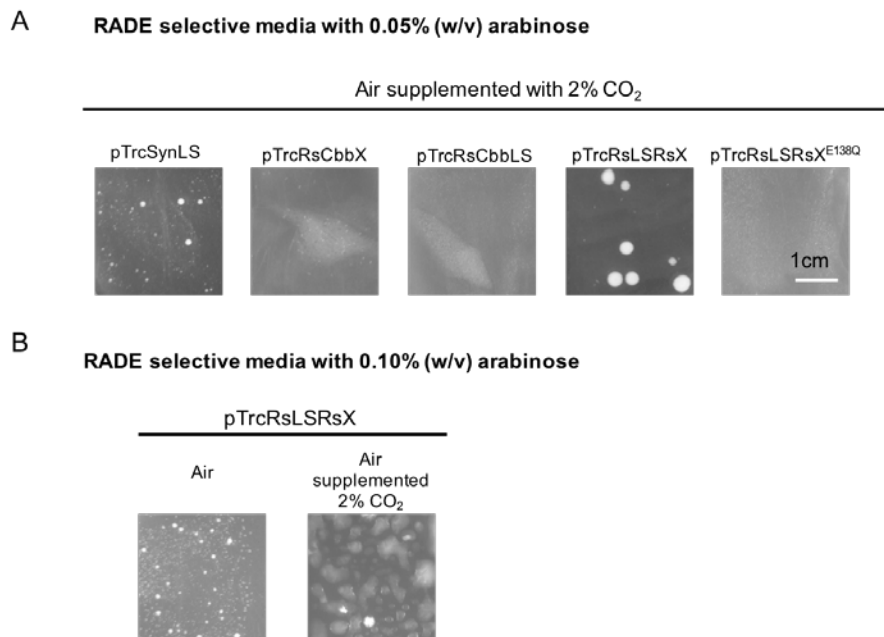


Figure 3-8: Representation of Rubisco and activase dependent *E. coli* selection using MM1-pAC^{BAD}PRK^{DM}.

(A) RADE selective media supplemented with 0.05% (w/v) arabinose incubated in air supplemented with 2% CO₂. There was colony growth of MM1-pAC^{BAD}PRK^{DM} cells harboring pTrcSynLS and pTrcRsLSRsX but not pTrcRsCbbX, pTrcRsCbbLS or pTrcRsLSRsX^{E138Q}. (B) RADE selective media supplemented with 0.10% (w/v) arabinose incubated under atmospheric condition with 0.04% CO₂ (air) and air supplemented with 2% CO₂. There was colony growth of MM1-pAC^{BAD}PRK^{DM} cells harboring pTrcRsLSRsX in both atmospheric and elevated levels

of 2% CO₂. These colonies started fusing due to their close proximity as they appear at day 5 while the plate images were taken at day 10. The RADE selective media is composed of M9 minimal media supplemented with 0.05% (w/v) casamino acid, antibiotics (Amp, Chl, Kan) and with 0.05% or 0.10% (w/v) of arabinose. Images of plates were taken at the end of ten days.

To further characterize the Rubisco and activase requirement in the RADE system, two known Rubisco and activase mutants were used as negative controls. They were RsΔ4 and RsCbbX^{K123A} and two duet plasmids were created pTrcRsΔ4RsX and pTrcRsLSRsX^{K123A}. RsΔ4 is a *Rs*-Rubisco mutant with a four-amino acid truncation at the C-terminus of the Rubisco large subunit. This truncation prevents the remodeling of inhibited RsΔ4 complexes by its activase as the mode of activation involves the threading the C-terminus through activase's central pore (Bhat et al., 2017a; Mueller-Cajar et al., 2011). RsCbbX^{K123A} is an activase mutant that has only ~18% of the wild type activase activity and this decrease in activity is due to the mutation being localized at the Rubisco-activase interaction surface (Bhat et al., 2017a; Mueller-Cajar et al., 2011). In RADE selective media containing 0.05% (w/v) arabinose and grown at 2% CO₂, there was no growth of MM1-pAC^{BAD}PRK^{DM} cells harboring pTrcRsΔ4RsX (Table 3-1). This finding was expected as RsCbbX will not be able to remodel the inhibited RsΔ4 complex. In contrast, cells harboring pTrcRsLSRsX^{K123A} had a lower survival rate as compared to those harboring wild type duet plasmid (pTrcRsLSRsX). This suggests that the screening condition used was not stringent enough as an impaired activase was sufficient for survival of cells harboring pTrcRsLSRsX^{K123A}.

The screening was repeated under more stringent conditions with increased PRK levels via increasing arabinose concentration and decreasing the CO₂ levels used. The screening condition changed to RADE selective media supplemented with 0.2% (w/v) arabinose and air supplemented with 0.5% CO₂. Under atmospheric conditions (0.04% CO₂), there was no growth in the RADE selective media (Table 3-1). When supplemented with 0.5% CO₂, there was only growth of MM1-pAC^{BAD}PRK^{DM} cells co-expressing both wild type *Rs*-Rubisco and activase. The increased stringency of the selection conditions led to lower survival rate when compared to previous screening condition. This led to a delay in the appearance of the first colony from day 5 to day 9. Overall, these screening results using different mutants of Rubisco and activase result highlighted the requirement for the presence of both compatible and functional Rubisco and activase for survival in the RADE system.

Rubisco and/or activase	MM1-pAC ^{BAD} PRK ^{DM}			
	RADE selective media supplemented with 0.05% (w/v) arabinose		RADE selective media supplemented with 0.2% (w/v) arabinose	
	Air (0.04% CO ₂)	Air supplemented with 2% CO ₂	Air (0.04% CO ₂)	Air supplemented with 0.5% CO ₂
pTrcSynLS	-	+	-	-
pTrcRsCbbX	-	-	-	-
pTrcRsCbbLS	-	-	-	-
pTrcRsLSRsX	+	+++	-	+
pTrcRsLSRsX ^{K123A}	-	+	-	-
pTrcRsA4RsX	-	-	-	-

Table 3-1: Relative fitness of MM1-pAC^{BAD}PRK^{DM} expressing Rubisco and/or activase in the RADE system

The RADE selective media supplemented with either 0.05% or 0.2% (w/v) arabinose under different CO₂ conditions. The screening was performed for 10 days, in atmospheric condition (air, 0.04% CO₂) at room temperature or in a CO₂ incubator with air supplemented with 0.5% or 2% CO₂ at 23 °C. – no growth, + minimum growth, ++ moderate growth, +++ maximum growth.

3.3.1.2 RADE and RDE screening condition of *Af-Rubisco* and activase were not achieved using MM1-pAC^{BAD}PRK^{DM}

The initial success of establishing screening conditions for both RDE and RADE using *Rs*-Rubisco and activase motivated us to set up the same system using *Af*-Rubisco and its activase, AfCbbQO. Similar to *R. sphaeroides*, *Af*-Rubisco had been shown to form inhibited complexes with RuBP and required remodeling by its activase for removal of the inhibitory sugar phosphates. *A. ferrooxidans* has three operons encoding for three different forms of Rubisco and each Rubisco was accompany by its activase (Tsai et al., 2015). In this study, only the form I Rubisco (AfLS) and the form II Rubisco (AfM) and their respective activase (AfQ1O1 and AfQ2O2) were chosen as they had been recently biochemically characterized.

Hence, I succeeded in establishing RADE screening condition for the form I *Af*-Rubisco and its activase. Using RADE selective media supplemented with 0.15% (w/v) arabinose at elevated CO₂, MM1-pAC^{BAD}PRK^{DM} cells were able to survive by expressing both Rubisco and activase. However, these cells expressing Rubisco alone were also able to survive but grew slower as compared to those co-expressing both Rubisco and activase (pTrcAfLSQ1O1) (Figure 3-9). In contrast to the form I Rubisco, complete Rubisco and activase dependency for form II *Af*-Rubisco, AfM, and its

activase, AfQ2O2, was achieved. On RADE selective media supplemented with 0.15% (w/v) arabinose at elevated CO₂, only cells expressing both AfM and AfQ2O2 were able to survive. This suggested that AfM formed inhibited complexes with RuBP and required the presence of its activase to alleviate the inhibition. However, the pTrcAfMAfQ2O2 duet harboring cells had low transformation efficiency ($\sim 10^2$ cfu/ μ g of plasmid DNA) compared to the other plasmids ($\sim 10^5$ cfu/ μ g of plasmid DNA), including plasmid expressing either form II Rubisco or its associated activase. Due to the lower transformation efficiency of the pTrcAfMAfQ2O2, there was a lower colony count in the RADE selective media. Attempts to improve the transformation efficiency by increasing the volume of MM1-pAC^{BAD}PRK^{DM} competent cells or plasmid were not successful. This lower transformation efficiency may be attributed to toxicity from the expression of both AfM and AfQ2O2.

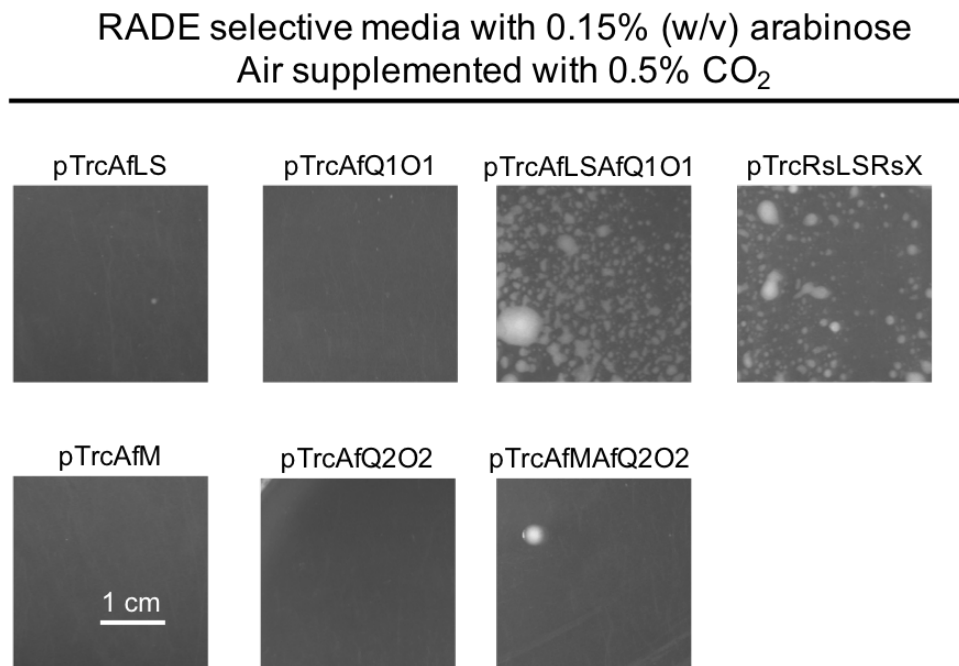


Figure 3-9: Screening of *Af*-Rubisco and activase in RADE selective media using MM1-pAC^{BAD}PRK^{DM} cells.

The screening conditions used were RADE selective media was supplemented with 0.15%(w/v) arabinose, grown in air supplemented with 0.5% CO₂ and at 23 °C. There was growth of colonies in MM1-pAC^{BAD}PRK^{DM} expressing the following plasmids, pTrcAfLS, pTrcAfLSAfQ1O1, pTrcRsLSRsX, and pTrcAfMAfQ2O2. pTrcRsLSRsX was used as positive control. Images of these plates were taken after 10 days.

3.3.1.3 MM1-pAC^{BAD}PRK^{DM} proved ineffective in the screening of mutant Rubisco libraries

In parallel to the setting of RDE and RADE system for *Af*-Rubisco and its activase, mutant libraries of RsLS in the pTrcHisB and pTrcDuet (EpRsLS and EpRsLSRsX) were generated. EpRsLS and EpRsLSRsX libraries were screened in under RDE and RADE condition respectively (Table 3-1). The experiments using wild type Rubisco and activase plasmids indicated a solid selection was in place. Disappointingly, the MM1-pAC^{BAD}PRK^{DM} cells were found ineffective in discriminating for fitter mutants and many instances of wild type sequences were isolated from these libraries. This undesirable screening results coupled with the failure to establish RDE and RADE using *Af*-Rubisco and its activase suggested that the double mutant of *Synechococcus* PCC 7942 PRK was not stringent enough for screening purpose. Previous RDE studies used wild type *Synechococcus* 7942 PRK (PRK⁷⁹⁴²) for screening which would have higher catalytic efficiency as compared to the double mutant (Mueller-Cajar et al., 2008b; Parikh et al., 2006). Therefore, we reasoned a wild type PRK should be ideally used for screening.

3.4 A high throughput selection screen for Rubisco in the presence and absence of activase using MM1-pAC^{BAD}PRK⁶⁸⁰³

The inefficiency of PRK^{DM} prompted the search for alternative PRK enzymes, PRK⁶⁸⁰³ and PRK⁷⁰⁰² were chosen due to their high sequence identity with the original PRK⁷⁹⁴². The wild type *prkA* genes were successfully cloned in pAC^{BAD} vector and transform into MM1 cells. The screening stringency of these two constructs was compared using two rounds of screening. In both sets of screening, the wild type *Syn*-Rubisco was no longer able to support the growth of the cells. In comparison, the F345I mutated *Syn*-Rubisco (*SynL*^{F345I}S) obtained from a previous RDE study, was able to support cell growth. However, MM1-pAC^{BAD}PRK⁷⁰⁰² expressing either the negative control RsCbbX or RsCbbLS were also able to support growth in the RDE selective media. This suggested that PRK⁷⁰⁰² was not stringent enough to generate sufficient RuBP toxicity. In contrast, the MM1-pAC^{BAD}PRK⁶⁸⁰³ cells were able to grow only when expressing the F345I mutant Rubisco and this suggested the plasmid is more suitable for screening compared to pAC^{BAD}PRK⁷⁰⁰².

Rubisco and/or activase	RDE selective media supplemented with 0.05% (w/v) arabinose Air supplemented with 0.5% CO ₂	
	MM1-pAC ^{BAD} PRK ⁶⁸⁰³	MM1-pAC ^{BAD} PRK ⁷⁰⁰²
pTrcSynLS	-	-
pTrcSynL ^{F345I} S	+++	++
pTrcRsCbbX	-	+
pTrcRslS	-	+
pTrcRslSRsX	-	-

Table 3-2: Relative fitness of Rubisco and activase in MM1 cells expressing either PRK⁷⁰⁰² or PRK⁶⁸⁰³.

The RDE screening was carried out in a CO₂ incubator with air supplemented with 0.5% CO₂ at 23°C for a duration of 10 days. The RDE selective media consists of M9 minimal media supplemented with 0.05% (w/v) casamino acid, 0.5mM IPTG, antibiotics (Amp, Chl, Kan) and 0.05% (w/v) arabinose. – no growth, + minimum growth, ++ moderate growth, +++ maximum growth.

3.4.1 RDE and RADE screening using MM1-pAC^{BAD}PRK⁶⁸⁰³

Rubisco dependency (RDE), as well as Rubisco and activase dependency (RADE), were achieved with MM1-pAC^{BAD}PRK⁶⁸⁰³ cells. Therefore, these cells were selected for subsequent screenings. Both RDE and RADE screening conditions were optimized using selective media either containing IPTG (termed as RDE selective media) or lacking IPTG (termed as RADE selective media). IPTG was used as an inducer for Rubisco and activase expression under the Trc promoter. In the absence of IPTG, there was still basal levels of protein expression as the Trc promoter is leaky (Figure 3-4, Figure 3-5). pTrcSynL^{F345I}S was used as a positive control for Rubisco dependency as cells expressing this F345I mutant was able to sustain growth in RDE screening conditions. In contrast, MM1-*prk* cells expressing either *Rs*-Rubisco or *Af*-Rubisco did not survive in RDE screening as these Rubiscos likely formed inhibited complexes with RuBP and there was insufficient RuBP turnover. Similarly, MM1-*prk* cells expressing activases (*RsCbbX* or *AfQ1O1*) also did not survive as they do not catalyze RuBP turnover reaction. The co-expression of both Rubisco and activase (pTrcRslSRsX or pTrcAfLSQ1O1) in MM1-*prk* cells was not able to sustain growth in the RDE screening system despite the presence of activase to facilitate RuBP removal from inhibited Rubisco complex. This was likely due to the increased metabolic burden in

overexpressing three proteins (PRK, Rubisco, and activase). Therefore, the directed evolution of *Rs*-Rubisco and *Af*-Rubisco alone (in absence of activase) was performed using selective media containing IPTG. Mutant libraries of *Rs*-Rubisco and *Af*-Rubisco were generated using error-prone PCR and the PCR product was ligated back into the pTrcHisB vector to give rise to the libraries EpRsLS and EpAfLS (Table 3-4). These two Rubisco libraries were then screened using the RDE selection system.

The Rubisco and activase dependency were achieved for *Rs*-Rubisco, *Af*-Rubisco and their respective activase using selective media lacking IPTG (RADE selective media) (Figure 3-11). Using RADE selective media supplemented with 0.05% (w/v) arabinose, MM1-pAC^{BAD}RK⁶⁸⁰³ cells expressing either *Rs*-Rubisco and *Af*-Rubisco were unable to support growth as they were likely forming inhibited complexes with RuBP. Similarly, cells expressing the only activase were not able to sustain growth as activase does not carry out RuBP catalysis. However, when these Rubiscos were co-expressed with their activase using the pTrcDuet plasmids (pTrcAfLSQ1O1, pTrcRsLSRsX), the MM1-*prk* cells were able to grow (Figure 3-11A, B). In addition, the co-expression of Rubisco and activase was able to support growth in both air and elevated CO₂ levels with those grown in elevated CO₂ levels having higher fitness as compared to those in air. This difference in fitness was due to the different rates of carboxylation whereby, at elevated CO₂ levels, the carboxylation rate is close to the maximum carboxylation velocity as compared to those in air.

To evolve these Rubiscos in the presence of their activase, mutant libraries of Rubisco were generated and ligated back into the pTrcDuet backbone containing wild type activase. Only the mutant library of *Rs*-Rubisco (EpRsLSRsX) was created successfully (Table 3-4). The construction of mutant libraries for duet containing randomly mutated AfLS and wild type AfQ1O1 were apparently hindered by the large size of the activase gene, AfQ1O1 was ~3200bp in comparison to RsCbbX which was only ~900bp. In future, the mutant libraries of AfLS and its activase and potentially other combinations can be commercially purchased as the cost of such services are becoming highly affordable. Therefore, in this current study, only the EpRsLSRsX library was screened using RADE selective media supplemented with 0.15% (w/v) arabinose in air. This condition was chosen as the MM1-*prk* cells expressing both wild-type *Rs*-Rubisco and activase were unable to sustain growth (Table 3-3, Figure 3-11C).

MM1-pAC ^{BADPRK} ⁶⁸⁰³			
Rubisco and/or activase	RDE selective media supplemented with 0.05% (w/v) arabinose	RADE selective media supplemented with 0.05% (w/v) arabinose	
	Air supplemented with 1% CO ₂	Air (0.04% CO ₂)	Air supplemented with 1% CO ₂
pTrcSynL ^{F345IS}	+++	n.d.	n.d.
pTrcRsCbbX	-	-	-
pTrcRsLS	-	-	-
pTrcRsLSRsX	-	++	+++
pTrcAfLS	-	-	-
pTrcAfQ1O1	-	-	-
pTrcAfLSQ1O1	-	++	+++

Table 3-3: Relative fitness of Rubisco and activase in MM1-pAC^{BADPRK}⁶⁸⁰³ cells screened on selective media containing (RDE) and lacking (RADE) 0.5 mM IPTG.

RDE and RADE screenings were carried out in either in a CO₂ incubator with air supplemented with 1% CO₂ at 23°C or under atmospheric condition of 0.04% CO₂ (Air) at room temperature for a duration of 14 days. The selective media consists of M9 minimal media supplemented with 0.05% (w/v) casamino acid, ± 0.5mM IPTG, antibiotics (Amp, Chl, Kan) and 0.05% (w/v) arabinose. The RADE selective media consists of M9 minimal media supplemented with 0.05% (w/v) casamino acid, antibiotics (Amp, Chl, Kan) and 0.05% (w/v) arabinose. – no growth, + minimum growth, ++ moderate growth, +++ maximum growth, n.d not determined.

Figure 3-10: Representative plate images for RDE screening using MM1-pAC^{BAD}PRK⁶⁸⁰³.

(A) Permissive plate images of successful transformation of plasmids into MM1-pAC^{BAD}PRK⁶⁸⁰³. (B) Screening of Rubisco and/or activase and mutant libraries using RDE screening system. *Af*-B13 is one of the *Af*-Rubisco variants obtained and it was transformed back into MM1-pAC^{BAD}PRK⁶⁸⁰³ to demonstrate that this variant was able to support growth in RDE. Marking on the mutant libraries images corresponds to faster growing colonies that were inoculated during screening. (-): No growth of colonies observed, (+) growth of colonies observed.

Figure 3-11: Screening of Rubisco and/or activase under RADE system using MM1-pAC^{BAD}PRK⁶⁸⁰³.

Screening of Rubisco and/or activase using RADE selective media with 0.05% arabinose grown in (A) elevated CO₂ or (B) in air. (C) Screening of EpRsLSRsX using RADE selective media with 0.15% arabinose and grown in air at room temperature. RADE selective consisted of M9 minimal media supplemented with 0.5% casamino acid and antibiotics (Amp, Chl, Kan) and a variable amount of arabinose. RADE selective media does not contain any IPTG.

Mutant libraries	Size of mutant libraries	Screening conditions
EpRsl.S-1	3,000	RDE selective media supplemented with 0.05% (w/v) arabinose in air supplemented with 1% CO ₂ at 23 °C
EpRsl.S-2	7,600	
EpAfl.S-1	162,000	
EpAfl.S-2	48,000	
EpRsl.SRsX	27,000	RADE selective media supplemented with 0.15% (w/v) arabinose under atmospheric condition (0.04% CO ₂) at room temperature

Table 3-4: Mutant libraries of Rubisco in pTrcHisB and pTrcDuet and the screening conditions for each library.

The mutant library sizes for *Rs*-Rubisco in both pTrcHisB and pTrcDuet was much smaller than for *Af*-Rubisco. The methodology for making both libraries was identical and differences in sizes may have been attributed to batch variation in the DH10B competent cells used. The selective media consists of M9 minimal media supplemented with 0.05% (w/v) casamino acid, $\pm$ 0.5mM IPTG, antibiotics (Amp, Chl, Kan) and 0.05% (w/v) arabinose. RDE selective media contains IPTG while RADE selective media does not.

3.4.2 Repeatedly mutated residues were obtained from RDE and RADE screening

The screening of the various libraries, under their respective screening conditions, was successful in yielding many variants with repeating mutations (Table 3-5). These variants were acquired from fast-growing colonies on the various screening media (Figure 3-10B, Figure 3-11C). The number of mutants isolated for the RDE screening was similar for both *R. sphaeroides* (59 variants) and *A. ferrooxidans* (56 variants) Rubisco. In contrast, there was a smaller number of variants (20 variants) obtained from RADE screening. In the different screenings, reoccurring substitutions were found, which is a strong statistical indicator that mutations that confer survival fitness were repeatedly selected for. This shows that our system was robust, the screening was successful and the variants will be described in detail in chapter 4 and 5. However, there were of instances of false positives whereby colonies encoding for wild type plasmids were able to support cell growth (Table 3-5). This may have arisen from a large number of transformants (95%) that were plated onto the selective plate as compared to the 5% transformant on the permissive plate where there was already log phase growth (Figure 3-10A). Cells probably overcome RuBP toxicity via other mechanisms such as PRK mutation. Hence, a PRK-antibiotic fusion construct as recently introduced by Wilson et al. (2017a) may be good to adopt for future investigations.

The mutation frequency of the different mutant libraries was similar as the amount of mutagenic buffer used in error-prone PCR was identical. This methodology of using error-prone PCR had its drawback with mutation bias towards transition for EpAfLS and EpRsLSRsX as shown in Table 3-5. Future work on generating different libraries could use commercially available kits for which do not have mutation bias and this could provide a better representation of potential solutions that can be acquired from directed evolution.

	Rubisco gene mutated					
	RsLS (EpRsLS)		RsLS (EpRsLSRsX)		AfLS (EpAfLS)	
Number of colonies inoculated	82		31		63	
Number of variants identified	59		20		56	
Number of variants with repeatedly mutated residues	28		8		23	
Number of wild type (false positive, %)	23 (28%)		11 (35%)		7 (11%)	
Mutation frequency (mutations per kb)	1.1		1.2		1.1	
Type of mutation						
Transitions	Number	%	Number	%	Number	%
A→G; T→C	32	27.8%	18	32.7%	70	49.3%
G→A; C→T	31	27.0%	19	34.5%	24	16.9%
Tranversions						
A→T; T→A	31	27.0%	11	20.0%	33	23.3%
A→C; T→G	3	2.6%	5	9.1%	8	5.6%
G→C; C→G	5	4.3%	0	0.0%	3	2.1%
G→; C→A	13	11.3%	2	3.6%	4	2.8%

Table 3-5: Summary of mutations from various mutant library screened.

The mutation frequency for each library was calculated from ten randomly chosen mutants. In the calculation of the type of mutations, silent mutations were also considered. The percentage for each type of mutation was calculated from the number of the individual mutation as a percentage of the total of mutation.

3.5 Discussion

In this chapter, the RDE and RADE selection systems were finally established using the construct pAC^{BAD}PRK⁶⁸⁰³ which contains a wild type *prkA* from *Synechocystis* PCC 6803. Mutant libraries of *R*s-Rubisco and *A*f-Rubisco were screened under RDE selection system which consists of RDE selective media containing 0.5% (w/v) arabinose in air supplemented with 1% CO₂ at 23 °C. The RADE selection system was also set up and the presence of both functional and compatible Rubisco and activase were required to support growth in this system. Mutant library of *R*s-Rubisco was screened in the presence of its activase in the RADE selective media supplemented with 0.15% (w/v) arabinose under atmospheric conditions with 0.04% CO₂ at room temperature. Re-occurring Rubisco variants were obtained from the various screenings of the mutant libraries. Their characterization will be described in detail in chapter 4 and chapter 5.

In this study, there were instances of false positives whereby colonies encoding for wild type plasmids were able to support cell growth. In comparison to other RDE systems which utilized *E. coli* transformed with a PRK plasmid, the MM1-PRK system has a low false positive rate and variants with repeatedly residues been obtained (Mueller-Cajar et al., 2008a, 2008b; Wilson et al., 2017b). The main disadvantage of using MM1-pAC^{BAD}PRK is the slow growth rate of the MM1 cells and the requirement for a specialized growth media consisting of M9 salt and different carbon source (glycerol, casamino acid, and malate). These factors translated into prolonged screening duration of 14-30 days. The use of wild type *E. coli* could overcome these limitations of MM1 strain, however, previous RDE studies using wild type *E. coli* transformed with a PRK plasmid had high false positive. A recent study by Wilson et al. (2017a) demonstrated that these false positives can be eliminated by using a PRK linked to a kanamycin resistant gene fusion protein (PRK-NTPII). In this RDE system, there is a dual selection of RuBP-toxicity and antibiotic resistance which eliminates PRK-toxicity escape mutants. In addition, this RDE system has a shorter screening duration of six days and uses LB as growth media. Directed evolution of form II Rubisco from *Thermosynechococcus elongatus* BP1 using this RDE system had yield mutants improved carboxylation rate, efficiency, and specificity (Wilson et al., 2017a). Alternatively, Antonovsky et al. (2016) had evolved a metabolically engineered *E. coli*

to carry out CO₂-dependent sugar synthesis using the same non-native CBB-cycle consisting of PRK and Rubisco. Although the feasibility of this *E. coli* strain in evolving Rubisco had yet to be demonstrated.

In this study, we were able to establish the Rubisco and activase dependency in the MM1-*prk* cells. This selection system has significant potential in creating compatibility between different type of Rubisco and activase which will allow beneficial properties of activase to be harnessed. Among the different activases, the red-type Rubisco activase from red-type Rubisco has been well characterized. Red-type Rubisco activase remodels the inhibited Rubisco complexes by threading the C-terminus of Rubisco through its central pore. Green-type rubiscos are currently not compatible with the red-type activases and the engineering of an extended C-terminus to these green-type Rubiscos could aid in improving compatibility between these two pairs. This modified green-type Rubisco can be subsequently evolved with the red-type activase using the RADE system. Therefore, the combination of rational design and directed evolution can work together in creating novel compatibility to allow superior properties of Rubisco and activase to be recruited towards the improvement of photosynthetic capacity in future plants.

4 DIRECTED EVOLUTION OF ACTIVASE-REQUIRING RUBISCO REVEALS FACILE ESCAPE FROM SUGAR PHOSPHATE INHIBITION

The directed evolution of *R. sphaeroides* Rubisco (*Rs*-Rubisco) and *A. ferrooxidans* Rubisco (*Af*-Rubisco) using RDE or RADE screening with MM1-pAC^{BAD}PRK⁶⁸⁰³ were successful in yielding multiple Rubisco variants that encoded repeatedly mutated residues. *Rs*-Rubisco and *Af*-Rubisco were both evolved in the absence of their activase using the RDE screening system while *Rs*-Rubisco was also evolved in the presence of its activase using the RADE system. Both Rubiscos are known to be prone to inhibition by RuBP and require assistance from their activase to facilitate inhibitor release (Mueller-Cajar et al., 2011; Tsai et al., 2015). In addition, both Rubiscos form stable inhibited complexes where the rate of spontaneous release of inhibitor is low. In this chapter, we describe an initial characterization of the evolved variants, in particular assessing their tendency to form stably inhibited complexes.

4.1 Directed evolution of Rubisco using MM1-pAC^{BAD}PRK⁶⁸⁰³ resulted in Rubisco variants with repeatedly mutated residues

The directed evolution of *Rs* *Rs*-Rubisco and *Af*-Rubisco resulted in multiple variants containing repeatedly mutated residues as well as variants containing mutations that were only observed once. These mutations were mapped onto the Rubisco large and small subunit gene as shown in Figure 4-1 and Figure 4-2. In addition, the repeatedly mutated large subunit residues were also mapped on the Rubisco large subunit as shown in Figure 4-3. In the RDE screening of *Rs*-Rubisco and *Af*-Rubisco, there were clusters of reoccurring mutations observed at the 60s loop of the N-terminal domain (residues 61 to 82, Spinach numbering) and C-terminus of the C-terminal domain (residues 462 to end) respectively. Both regions are involved in active site closure and mutations at these loops could influence the stability of the closed active site as discussed further in Chapter 5 (Duff et al., 2000; Taylor et al., 1996b). A destabilization of the closed active site would favor inhibitor release and thus survival in the RDE screening system. In contrast, the RADE screening of *Rs*-Rubisco displayed no clustering of mutations at particular regions of Rubisco.

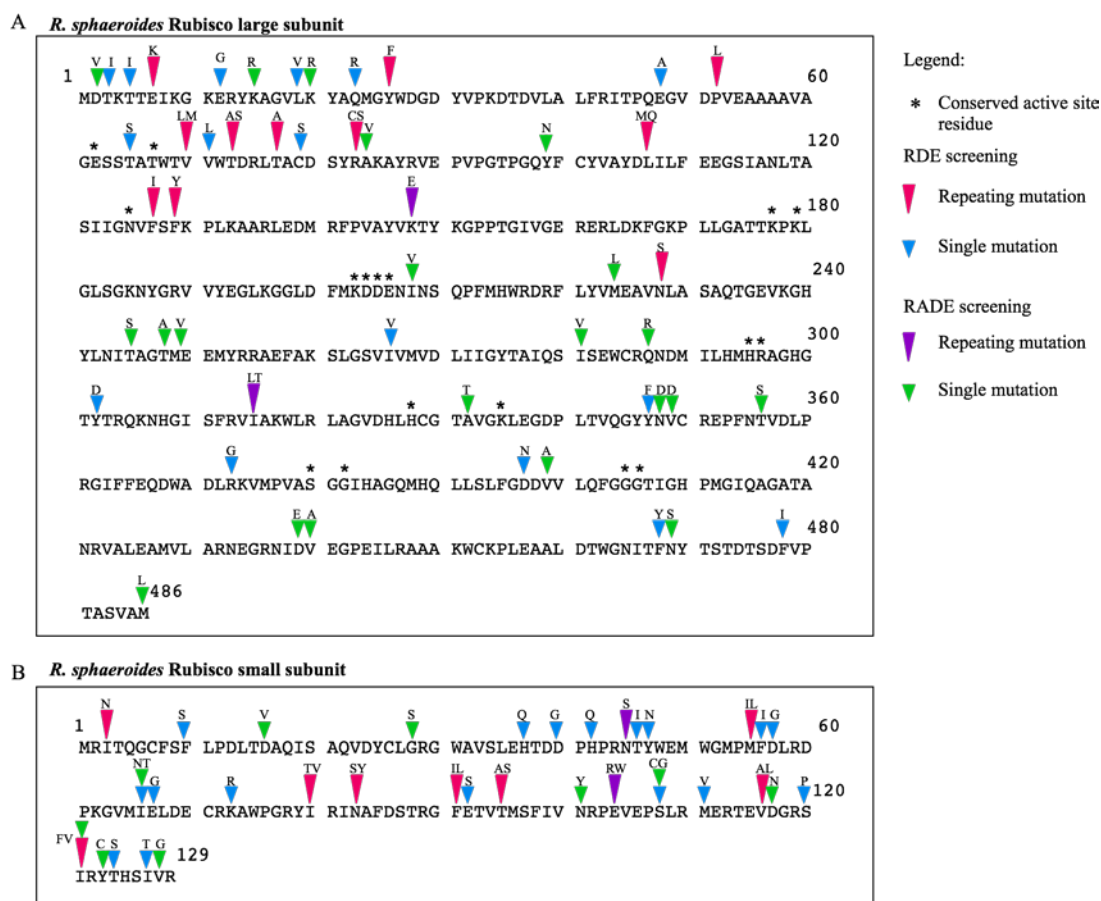


Figure 4-1: Mutations in *Rs*-Rubisco obtained from RDE and RADE screening.

(A) Mutations in *Rs*-Rubisco large subunit. (B) Mutations in *Rs*-Rubisco small subunit. The directed evolution of *Rs*-Rubisco was carried out in the presence and absence of its activase in RADE (purple and green arrow) and RDE (red and blue arrow) screening systems respectively. The substitution(s) for each mutated residue were denoted on top on the arrows. The asterisk (*) on top of residues corresponds to conserved active site residues. The numbering of *Rs*-Rubisco large and small subunit residues is according to the *R. sphaeroides* sequence.

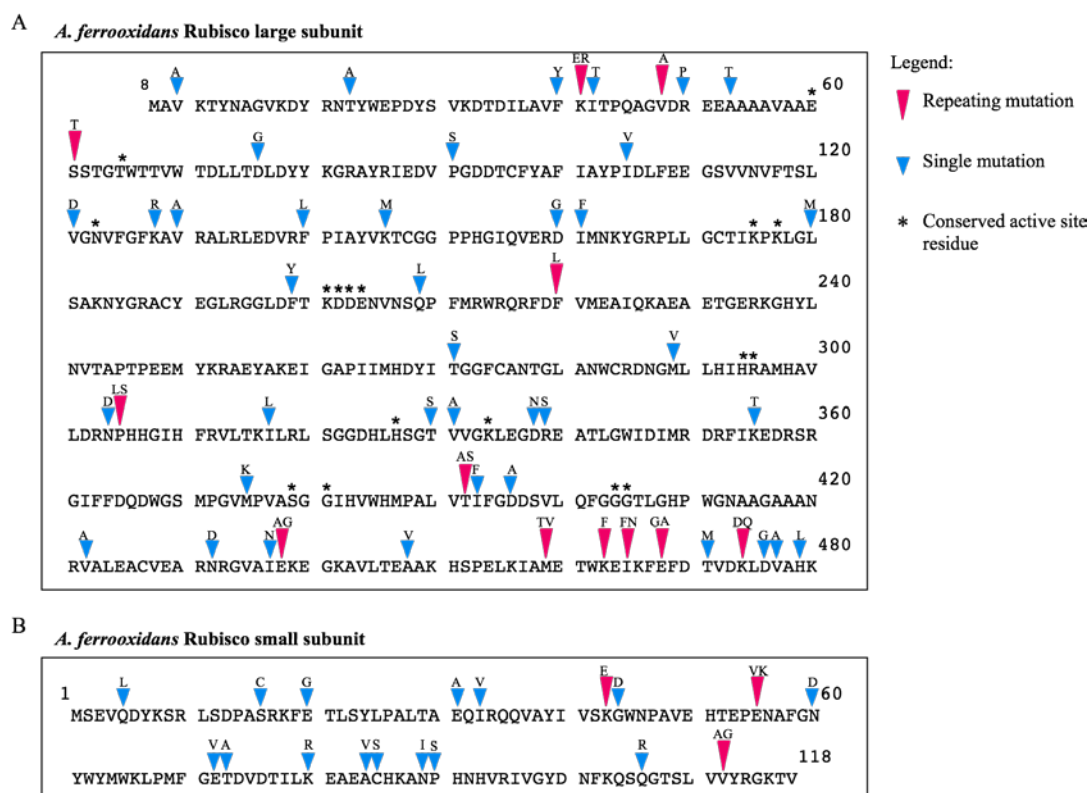


Figure 4-2: Mutations in *Af*-Rubisco obtained from RDE screening.

(A) Mutations in *Af*-Rubisco large subunit. (B) Mutations in *Af*-Rubisco small subunit. The directed evolution of *Af*-Rubisco was carried out in absence of its activase using RDE screening systems. The substitution(s) for each mutated residue were denoted on top on the arrows. The asterisk (*) on top of residues corresponds to conserved active site residues. The numbering of *Af*-Rubisco large subunit numbering is according to the Spinach Rubisco sequence while the small subunit is according to *A. ferrooxidans* sequence.

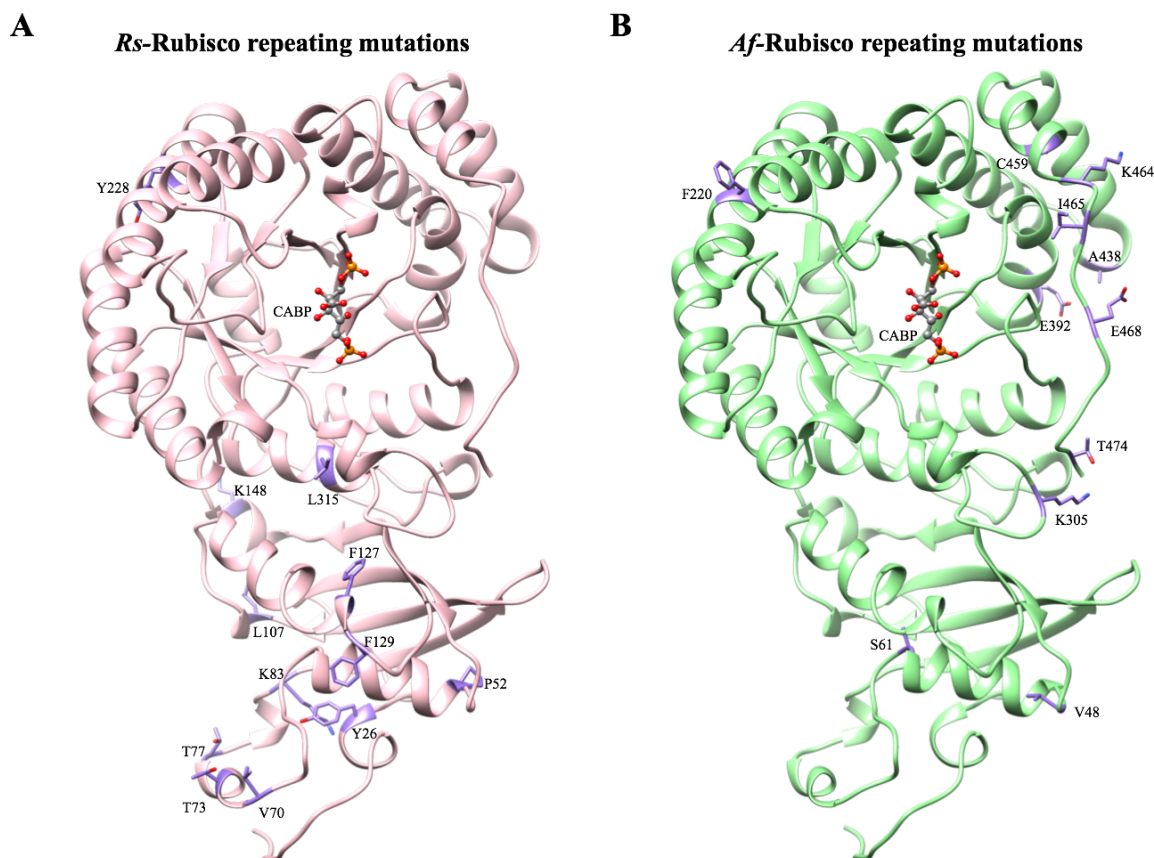


Figure 4-3: Position of repeatedly mutated residues of *Rs*- and *Af*-Rubisco variants.

(A) Repeatedly mutated residues in the large subunit of *Rs*-Rubisco variants. (B) Repeatedly mutated residues in the large subunit of *Af*-Rubisco variants. The structure of Spinach Rubisco (8RUC, closed state with CABP bound to the active site) was used to illustrate the position of these repeatedly mutated residues. The backbone of the large subunits was shown in ribbon representation while the mutated residues were shown in stick representation. In addition, the large subunit backbone coloured pink for *Rs*-Rubisco and green for *Af*-Rubisco while the mutated residues were coloured purple. The numberings of the residues were accordingly to the respective Rubisco while the residue was accordingly to the Spinach Rubisco.

4.1.1 Directed evolution of *Rs*-Rubisco using the RDE system

The directed evolution of *Rs*-Rubisco in the RDE screening resulted in 59 variants and 34 of these contained substitutions at repeating residues. These variants are listed in Table 4-1 and termed *Rs*-R'*X*' variants where '*X*' is an integer. The remaining variants that contained unique mutations are listed in Table 4-2 and they were termed *Rs*-T'*X*'. Nineteen of these *Rs*-R'*X*' variants were characterized using soluble *E. coli* extract with

priority placed on large subunit mutations. The non-repeating *Rs*-T'X' variants were not characterized due to resource and time restraints.

Mutant Rubisco		Amino acid changes(s)		Repeating mutated residue(s)
		RbcL	RbcS	
<i>Rs</i> -R1	RsL ^{E7K} S	E7K	nil	E7K ^a
<i>Rs</i> -R2	RsL ^{V70M} S	V70M	nil	V70M/L ^a
<i>Rs</i> -R3	RsL ^{E48A/V70L} S ^{S120P}	E48A, V70L	S120P	
<i>Rs</i> -R4	RsL ^{T73A} S	T73A	nil	T73A/S ^a
<i>Rs</i> -R5	RsL ^{T73S} S	T73S	nil	
<i>Rs</i> -R6	RsL ^{R83C} S	R83C	nil	R83C/S ^a
<i>Rs</i> -R7	RsL ^{R83S} S ^{K73R}	R83S	K73R	
<i>Rs</i> -R7.1 ^c	RsL ^{R83S} S ^{F10S}	R83S	F10S	
<i>Rs</i> -R8	RsL ^{F127I} S	F127I	nil	F127I ^a
<i>Rs</i> -R8.1 ^c	RsL ^{F127I/Y347F} S	F127I, Y347F	nil	
<i>Rs</i> -R9	RsL ^{F129Y} S	F129Y	nil	F129Y ^a
<i>Rs</i> -R10	RsLS ^{F91I/T95S/S108C}	nil	F91I, T95S, S108C	F91I/L ^b , T95A/S ^b
<i>Rs</i> -R11	RsLS ^{F91L}	nil	F91L	
<i>Rs</i> -R12	RsLS ^{T95A}	nil	T95A	
<i>Rs</i> -R13	RsLS ^{I3N}	nil	I3N	I3N ^b
<i>Rs</i> -R13.1 ^c	RsL ^{L19V} S ^{I3N}	L19V	I3N	
<i>Rs</i> -R14	RsL ^{Y26F} S	Y26F	nil	Y26F ^a
<i>Rs</i> -R14.1 ^c	RsL ^{Y26F} S ^{I121V}	Y26F	I121V	
<i>Rs</i> -R15	RsL ^{P52L} S	P52L	nil	P52L ^a
<i>Rs</i> -R16	RsL ^{T77A} S	T77A	nil	T77A ^a
<i>Rs</i> -R16.1 ^c	RsL ^{T77A} S ^{I127T}	T77A	I127T	
<i>Rs</i> -R17	RsL ^{N228S} S	N228S	nil	N228S ^a
<i>Rs</i> -R17.1 ^c	RsL ^{N228S} S ^{E67G}	N228S	E67G	
<i>Rs</i> -R18	RsL ^{L107M} S	L107M	nil	L107M/Q ^a
<i>Rs</i> -R19	RsL ^{L107Q} S	L107Q	nil	
<i>Rs</i> -R20 ^c	RsL ^{T31D397N} S ^{M55I}	T31, D397N	M55I	M55I/L ^b
<i>Rs</i> -R21 ^c	RsLS ^{M55L}	nil	M55L	
<i>Rs</i> -R22 ^c	RsLS ^{I121F}	nil	I121F	I121F/V ^b
<i>Rs</i> -R14.1 ^c	RsL ^{Y26F} S ^{I121V}	Y26F	I121V	
<i>Rs</i> -R23 ^c	RsLS ^{N83S/E92A/M111V}	nil	N83S, E92A, M111V	N83S/Y ^b
<i>Rs</i> -R24 ^c	RsLS ^{D57G/N83Y}	nil	D57G, N83Y	
<i>Rs</i> -R25 ^c	RsL ^{I266V} S ^{V116A}	I266V	V116A	V116A/L ^b
<i>Rs</i> -R26 ^c	RsLS ^{Y47N/V116I}	nil	Y47N, V116I	
<i>Rs</i> -R27 ^c	RsLS ^{T46I/I80V}	nil	T46I, I80V	I80V/T ^b
<i>Rs</i> -R28 ^c	RsL ^{T51I/T65S} S ^{I80T}	T51I, T65S	I80T	

Table 4-1: *Rs*-Rubisco variants with repeatedly mutated residues obtained from RDE screening.

^a: Repeating mutated residue is located in the Rubisco large subunit. ^b: Repeating mutated residue is located in the Rubisco small subunit. ^c: These *Rs*-Rubisco variants were not characterized further.

Mutant Rubisco		Amino acid change(s)	
		RbcL	RbcS
<i>R_S</i> -T1	RsL ^{R373G} S	R373G	nil
<i>R_S</i> -T2	RsL ^{C79S} S	C79S	nil
<i>R_S</i> -T3	RsL ^{Q23R} S	Q23R	nil
<i>R_S</i> -T4	RsLS ^{F56I}	nil	F56I
<i>R_S</i> -T5	RsL ^{F467Y} S	F468Y	nil
<i>R_S</i> -T6	RsL ^{V71L} S	V71L	nil
<i>R_S</i> -T7	RsL ^{H37Q} S	nil	H37Q
<i>R_S</i> -T8	RsL ^{E12G} S	E12G	nil
<i>R_S</i> -T9	RsL ^{F478I} S ^{I66N}	F478I	I66N
<i>R_S</i> -T10	RsLS ^{H42Q} ^{T124S}	nil	H42Q, T124S
<i>R_S</i> -T11	RsLS ^{D86G}	nil	D86G
<i>R_S</i> -T12	RsL ^{Y302D} S	Y302D	nil
<i>R_S</i> -T13	RsLS ^{D40G/+8aa}	nil	D40G, Stop130W (+8aa) [#] (WKLGCFFGG)

Table 4-2: *R_S*-Rubisco variants with unique mutation(s) obtained from RDE screening.

#: Mutation of the stop codon in *R_S*-T13 variant resulted in an eight amino acid extension of the Rubisco small subunit.

4.1.2 Directed evolution of *Af*-Rubisco using RDE system

The directed evolution of *Af*-Rubisco in the RDE screening yielded 56 variants and 26 of these contained substitutions at repeating residues. These variants are listed in Table 4-3 and termed *Af*-B'X' variants where 'X' is an integer. The remaining variants that contained unique mutations are listed in Table 4-4 and they were termed *Af*-B'X'. Twenty-one of these *Af*-B'X' variants were characterized using soluble extract with a priority placed on large subunit mutations. *Af*-Rubisco variants harboring mutations that were only observed once were also not characterized due to time restraints.

Mutant Rubisco		Amino acid change(s)		Repeating mutated residue(s)
		RbcL	RbcS	
<i>Af</i> -B1	AfL ^{K41R/I105V} S	K41R, I105V	nil	K41E/R ^a
<i>Af</i> -B2	AfL ^{K41E/A448V} S	K41E, A448V	nil	
<i>Af</i> -B3	AfL ^{V48A S^{K43E/N89I}}	V48A	K43E, N89I	V48A ^a
<i>Af</i> -B4	AfL ^{V48A/M375K S^{E55V}}	V48A, M375K	E55V	
<i>Af</i> -B5	AfL ^{S61T} S	S61T	nil	S61T ^a
<i>Af</i> -B6	AfL ^{F220L} S	F220L	nil	F220L ^a
<i>Af</i> -B7	AfL ^{P305L} S	P305L	nil	P305L/S ^a
<i>Af</i> -B8	AfL ^{P305S} S	P305S	nil	
<i>Af</i> -B9	AfL ^{T392A} S	T392A	nil	T392A/S ^a
<i>Af</i> -B10	AfL ^{T392S} S	T392S	nil	
<i>Af</i> -B11	AfL ^{E438A} S	E438A	nil	E438A/G ^a
<i>Af</i> -B12	AfL ^{E438G} S	E438G	nil	
<i>Af</i> -B13	AfL ^{I42T/K463E} S	I42T, K463E	nil	K463E ^a
<i>Af</i> -B14	AfL ^{F199Y/K463E} S	F199Y, K463E	nil	
<i>Af</i> -B15	AfL ^{T330S/K463E S^{N60D}}	T330S, K463E	N60D	I465F/N ^a
<i>Af</i> -B16	AfL ^{I465F} S	I465F	nil	
<i>Af</i> -B17	AfL ^{I465N} S	I465N	nil	M459T/V ^a
<i>Af</i> -B18	AfL ^{V10A/D160G/M459V} S	V10A, D160G, M459V	nil	
<i>Af</i> -B19	AfL ^{V422A/M459T S^{A84V}}	V422A, M459T	A84V	K43E ^b
<i>Af</i> -B20	AfL ^{T23A S^{K43E/K80R}}	T23A	K43E, K80R	
<i>Af</i> -B3	AfL ^{V48A S^{K43E/N89I}}	V48A	K43E, N89I	V112A/G ^b
<i>Af</i> -B21	AfL ^{K146M S^{V112A}}	K146M	V112A	
<i>Af</i> -B22	AfL ^{F140L S^{V112G}}	F140L	V112G	E55V/K ^b
<i>Af</i> -B23 ^c	AfL ^{K474D S^{E20G/E55K+16aa}}	K474D	E20G, E55K	
<i>Af</i> -B4	AfL ^{V48A/M375K S^{E55V}}	V48V, M375K	E55V	E468G/V ^a
<i>Af</i> -B24 ^c	AfL ^{D396A/E481G} S	D396A, E468G	nil	
<i>Af</i> -B25 ^c	AfL ^{F40Y/Q209L/E468V} S	F40Y, Q209L, E468V	nil	K474D/Q ^a
<i>Af</i> -B26 ^c	AfL ^{K474Q} S	K474Q	nil	
<i>Af</i> -B23 ^c	AfL ^{K474D S^{E20G/E55K+16aa}}	K474D	E20G, E55K	

Table 4-3: *Af*-Rubisco variants with repeatedly mutated residues obtained from RDE screening.

a: Repeating mutated residue is located in the Rubisco large subunit. b: Repeating mutated residue is located in the Rubisco small subunit. c: These *Af*-Rubisco variants were not characterized further.

Mutant Rubisco		Amino acid change(s)	
		RbcL	RbcS
<i>Af</i> -A1	AfL ^{V331A} S ^{G44D}	V331A	G44D
<i>Af</i> -A2	AfL ^{N304D} S	N304D	nil
<i>Af</i> -A3	AfL ^{I161F/R339S/V477A} S	I161F, R339S, V477A	nil
<i>Af</i> -A4	AfL ^{I393F} S	I393F	nil
<i>Af</i> -A5	AfL ^{P91S} S	P91S	nil
<i>Af</i> -A6	AfL ^{S^{P30S/Q106R}}	nil	P90S, Q106R
<i>Af</i> -A7	AfL ^{V121D} S	V121D	nil
<i>Af</i> -A8	AfL ^{S^{S359I}} S	S359I	I33V, T73A
<i>Af</i> -A9	AfL ^{R50P/L180M} S	R50P, L180M	nil
<i>Af</i> -A10	AfL ^{D76G} S	D76G	nil
<i>Af</i> -A11	AfL ^{H479L} S	H479L	nil
<i>Af</i> -A12	AfL ^{A53T} S	A53T	nil
<i>Af</i> -A13	AfL ^{K355T} S	K355T	nil
<i>Af</i> -A14	AfL ^{M289V} S	M289V	nil
<i>Af</i> -A15	AfL ^{T471M} S	T471M	nil
<i>Af</i> -A16	AfL ^{S^{Q5L}}	nil	Q5L
<i>Af</i> -A17	AfL ^{I317L} S	I317L	nil
<i>Af</i> -A18	AfL ^{S^{S16C/E72V}}	nil	S16C, E72V
<i>Af</i> -A19	AfL ^{K128R} S	K128R	nil
<i>Af</i> -A20	AfL ^{V130A} S	V130A	nil
<i>Af</i> -A21	AfL ^{I437N} S	I437N	nil
<i>Af</i> -A22	AfL ^{V384A} S	V384A	nil
<i>Af</i> -A23	AfL ^{T271S/N432D/D476G} S	T271S, N432D, D476G	nil
<i>Af</i> -A24	AfL ^{S^{E31A/C85S}}	nil	E31A, C85S
<i>Af</i> -A25	AfL ^{D338N} S	D338N	nil

Table 4-4: *Af*-Rubisco variants containing unique mutation(s) from RDE screening.

4.1.3 Directed evolution of *Rs*-Rubisco in the presence of its activase using the RADE system

The directed evolution of *Rs*-Rubisco in the presence of its activase using RADE screening resulted in the isolation of 20 variants with eight of them containing repeatedly mutated residues (Table 4-5). These variants were termed as *Rs*-S'X' and one of these repeating variants, *Rs*-S1, was characterized using purified protein. Whereas, the non-repeating Rubisco variants were labeled as *Rs*-T'X'. The RADE screening used the pTrcDuet plasmid which contains both Rubisco and activase. Therefore, performing the Rubisco activation assay using soluble *E. coli* extract of the duet plasmid would result in a complicated readout as there is activase present which could remodel the inhibited ER complexes. To analyze the properties of *Rs*-S1, the gene was cloned back into the pTrcHisB vector backbone and the protein subsequently purified.

Mutant Rubisco		Amino acid change(s)		Repeating mutated residue(s)
		RbcL	RbcS	
<i>Rs-S1</i>	RsL ^{B15T} S	I315T	nil	I315L/T ^a
<i>Rs-S2^c</i>	RsL ^{B15L} S ^{G28S}	I315L	G28S	
<i>Rs-S3.1^c</i>	RsL ^{A84V/K148E/V349D} S	A84V, K148E, V349D	nil	K148E ^a
<i>Rs-S3.2^c</i>	RsL ^{K148E/N348D} S ^{I121F}	K148E, N348D	I121F	
<i>Rs-S4^c</i>	RsL ^{S^{N45S}}	nil	N45S	N45S ^b
<i>Rs-S4.1^c</i>	RsL ^{M224L} S ^{N45S}	M224L	N45S	
<i>Rs-S5^c</i>	RsL ^{S^{E104R}}	nil	E104R	E104R/W ^b
<i>Rs-S6^c</i>	RsL ^{V399A} S ^{E104W}	V399A	E104W	

Table 4-5: *Rs*-Rubisco variants with repeatedly mutated residues obtained from RADE screening.

^a: Repeating mutated residue is located in the Rubisco large subunit. ^b: Repeating mutated residue is located in the Rubisco small subunit. ^c: These *Rs*-Rubisco variants were not characterized.

Mutant Rubisco		Amino acid change(s)	
		RbcL	RbcS
<i>Rs-U1</i>	RsL ^{Q287R} S ^{+8aa}	Q287R	Stop 130R (+8aa) [#] (RKLGCFFG)
<i>Rs-U2</i>	RsLS ^{S108G}	nil	S108G
<i>Rs-U3</i>	RsL ^{I208V/T245S} D ^{I117N}	I208V, T245S	D117N
<i>Rs-U4</i>	RsL ^{D2V/D439E/V440A} S ^{V128G}	D2V, D439E, V440A	V128G
<i>Rs-U5</i>	RsL ^{K15R/I281V} S ^{Y123C}	K15R, I281V	Y123C
<i>Rs-U6</i>	RsL ^{N469S} S	N469S	nil
<i>Rs-U7</i>	RsL ^{M486L} S	M486L	nil
<i>Rs-U8</i>	RsL ^{Y99N/A332T} S	Y99N, A332T	nil
<i>Rs-U9</i>	RsL ^{T356S} S ^{D16V/I66T/N101Y}	T356S	D16V, I66T, N101Y
<i>Rs-U10</i>	RsL ^{T248A} S	T248A	nil
<i>Rs-U11</i>	RsL ^{K20R} S	K20R	nil
<i>Rs-U12</i>	RsL ^{E250V} S	E250V	nil

Table 4-6: *Rs*-Rubisco variants with unique mutations obtained from RADE screening.

[#]: Mutation of the stop codon in *Rs-U1* variant resulted in an eight amino acid extension of the Rubisco small subunit.

4.2 Preliminary biochemical characterization of Rubisco variants using soluble *E. coli* extract and spectrophotometric assay

The biochemical characterization of the Rubisco variants with repeatedly mutated residues was carried out using the spectrophotometric Rubisco activation assay (Figure 4-4). Rubisco variants were expressed in *E. coli* strain BL21 and soluble *E. coli* extracts were obtained from small-scale protein expression (Figure 3-3, Figure 4-4A). Based on SDS-PAGE, there is an equivalent amount of Rubisco loaded for the variants for each species with *Af*-Rubisco variants having a higher expression level as compared to *Rs*-Rubisco variants. The amount of soluble *E. coli* extract loaded was 3×10^7 *E. coli* cells and this was calculated based on the OD₆₀₀ of the small-scale initial culture (Figure 4-5). The OD₆₀₀ measurement was necessary as a normalization step to ensure that the amount of lysate added in each reaction was comparable. Attempts to precisely quantify the amount of Rubisco via immunoblotting and densitometry were unsatisfactory due to the semiquantitative nature of the method. Therefore, these assays using the soluble *E. coli* extract should be considered a preliminary characterization methodology used to identify variants of interest for more detailed evaluation.

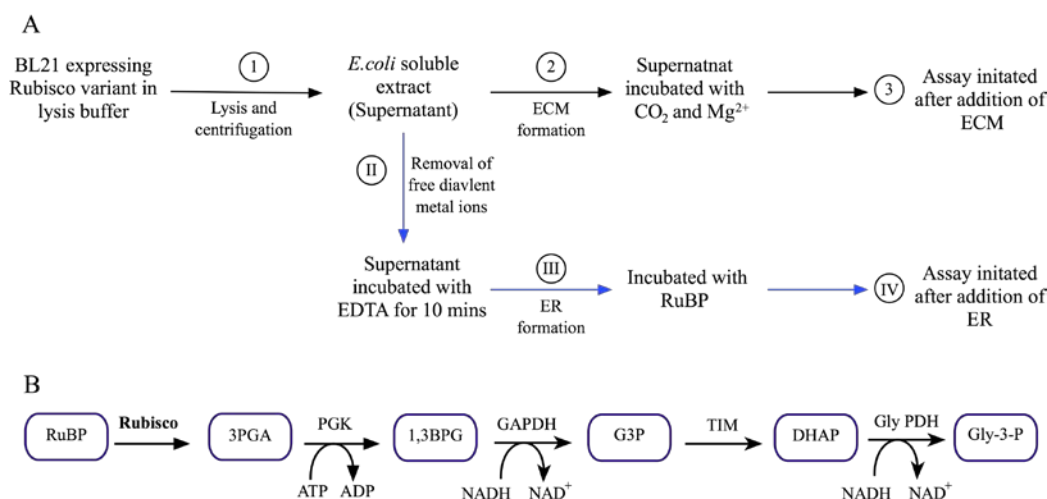


Figure 4-4: Representation of Rubisco activation assay using soluble *E. coli* extract.

(A) Rubisco activation assay performed using soluble extract from BL21 *E. coli* cells expressing Rubisco variant encoded by pTrcHisB vectors. These soluble extracts were obtained from BL21 *E. coli* cells following lysis via sonication and centrifugation (step 1, also refer to Figure 3-2). The amount of supernatant used for the preparation of ECM and ER was derived from 6×10^7 *E. coli* cells and this amount was calculated based on the OD₆₀₀ of the original small-scale culture (refer to Figure 3-2). To form fully activated Rubisco complex (ECM), supernatant from 6×10^7

E. coli cells was incubated for 10 minutes with 40 mM NaHCO₃, 10 mM MgCl₂ and top up to 10 µl with buffer A (20 mM Tris-HCl, 50 mM NaCl, pH 8.0) (step 2). The assay was initiated using 2.5 µl of ECM complex which corresponds to the soluble extract from 1.5x10⁷ *E. coli* cells (step 3). To form ER complex, the supernatant from 6x10⁷ *E. coli* cells was incubated with 4 mM EDTA for 10 minutes to chelate any free metal divalent ions and it was top up to 10 µl with buffer A (step II, blue arrows). Subsequently, 0.8 mM RuBP was added and incubated for 10 minutes to obtain ER complex (step III). The assay was initiated after addition of 5 µl of ECM complex which corresponds to soluble extract from 3x10⁷ *E. coli* cells (step IVs). (B) Representation of the spectrophotometric Rubisco assay, taken from Figure 2-3.

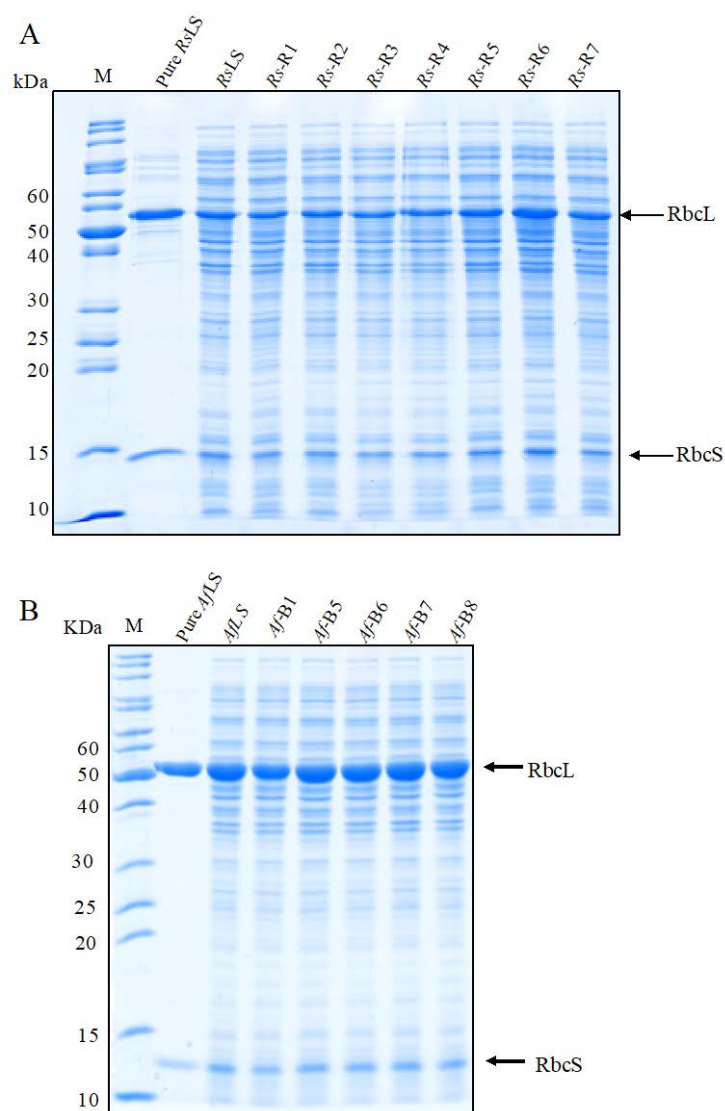


Figure 4-5: Representative SDS-PAGE analysis of soluble *E. coli* extract from *Rs*-Rubisco and *Af*-Rubisco variants.

(A) 16% SDS-PAGE of soluble *E. coli* extract from *Rs*-Rubisco variants (B) 16% SDS-PAGE of soluble *E. coli* extract from *Af*-Rubisco variants. The amount of soluble *E. coli* extract loaded for wild type Rubisco and variants were from 3x 10⁷ BL21 *E. coli* cells. ~1 µg of purified Rubisco (Pure RsLS, AfLS) was also loaded as control.

4.2.1 Characterization of *Rs*-Rubisco variants using soluble *E. coli* extract

Rs-Rubisco variants (*Rs*-R1 to *Rs*-R19) contained repeatedly mutated residues and these were characterized using soluble *E. coli* extract prepared from small-scale protein expression. The soluble *E. coli* extracts of these extract were used to prepare the fully activated Rubisco (ECM) complexes and RuBP-inhibited Rubisco (ER) complexes as shown in Figure 4-4A. The amount of soluble *E. coli* extract used for ECM and ER activity of *Rs*-Rubisco were from 1.5×10^7 and 3.0×10^7 *E. coli* cells respectively. In addition, the amount of Rubisco in the soluble extract from 1.5×10^7 *E. coli* cells was optimized to be equivalent to approximately 0.05 μ M active site of purified Rubisco (Figure 4-6A). Therefore, the Rubisco activity for ECM and ER complexes of the wild type *Rs*-Rubisco (*RsLS*) from soluble extract and purified source were similar (Figure 4-6, Figure 4-7). Most importantly, the *RsLS* from soluble *E. coli* extract was able to form the stable inhibited complex with a similar low rate of spontaneous activation as purified *RsLS* (Figure 4-7A, E). The similar rate of spontaneous activation suggested that the influence of other low abundance proteins in the soluble extract was not significant over the course of the assay duration. Thus, the measurement of ECM and ER activity with soluble extract using the spectrophotometric assay was feasible. Rubisco activity of ECM and ER were displayed as product (3PGA) accumulation over time (Section 2.6.2). The activity of the ECM form of the Rubisco variants was calculated by fitting the data using a linear fit and displayed as a percentage of wild type *RsLS*. ER activity of variants was calculated using the first-order derivative at two time points (five and seven minutes) and this data was displayed as a percentage of its ECM activity.

The majority of the activated (ECM) *Rs*-Rubisco variants exhibited a lower or similar carboxylase activity as compared to the wild type (Figure 4-6). *Rs*-R16 and *Rs*-R18 were exceptions as they had higher carboxylase activity with approximately 150% and 120% of the wild type ECM activity respectively. While *Rs*-R4, *Rs*-R5, *Rs*-R6, *Rs*-R14, *Rs*-R15, and *Rs*-R17 had similar ECM activity as *RsLS*. The remaining variants; *Rs*-R1 to *Rs*-R3, *Rs*-R7 to *Rs*-R12, and *Rs*-R19; had lower ECM activity than wild type. In contrast, the majority of these *Rs*-Rubisco variants' ER complexes exhibited higher Rubisco activity than wild-type after five and seven minutes (Figure 4-7D, E). This suggests that they experienced increased spontaneous activation of their ER complexes

indicating a reduced ER complex stability compared to wild type. It is likely that the rate of RuBP dissociation from the active site has increased, which allow the active site to be free and available for activation by CO₂ and Mg²⁺ in the reaction buffer. These activated Rubiscos can subsequently carry out carboxylation reaction and this is manifested in the rapid increase in 3PGA production over time (Figure 4-7A, B, C).

Based on the percentage ECM activity of the ER complexes, the *Rs*-Rubisco variants could be categorized into three groups based on the relative rate of spontaneous activation; low, moderate and high (Figure 4-7E). For wild type *RsLS*, the rate of spontaneous activation was considered low exhibiting 10% activation after seven minutes. The second group with moderate spontaneous activation will be those variants that were 10% to 20% activated after seven minutes, and they were *Rs*-R1, *Rs*-R2, *Rs*-R11, *Rs*-R12, *Rs*-R17, and *Rs*-R18. Lastly, those variants that are considered to have a high spontaneous rate of activation with greater than 20% activation after seven minutes as this was at least double the rate of wild type Rubisco. These were the remaining 12 variants; *Rs*-R3 to *Rs*-R10, *Rs*-14 to *Rs*-16, and *Rs*-19. These variants with a high spontaneous rate of activation were of interest for further characterization, however, some of these had significantly lower ECM activity (less than 50% of wild type) such as *Rs*-R3, *Rs*-R8 and *Rs*-R9 were therefore excluded. There were also two other interesting variants, *Rs*-R17 and *Rs*-R18, which displayed moderate rates of spontaneous activation rate but possessed a higher ECM activity than wild type. The *Rs*-R19 variant was also chosen as it was mutated at the same residue (L107) as *Rs*-R18 but exhibited a different activity profile. Therefore, it would be interesting to study how the two mutations influence the activity of the enzyme. Overall, the *Rs*-Rubisco variants chosen for further characterization were *Rs*-R4, *Rs*-R5, *Rs*-R6, *Rs*-R7, *Rs*-R10, *Rs*-14, *Rs*-R15, *Rs*-16, *Rs*-R17, *Rs*-R18 and *Rs*-R19. These proteins were purified and characterized in detailed using spectrophotometric and radioisotopic Rubisco activity assays (discussed in chapter 5).

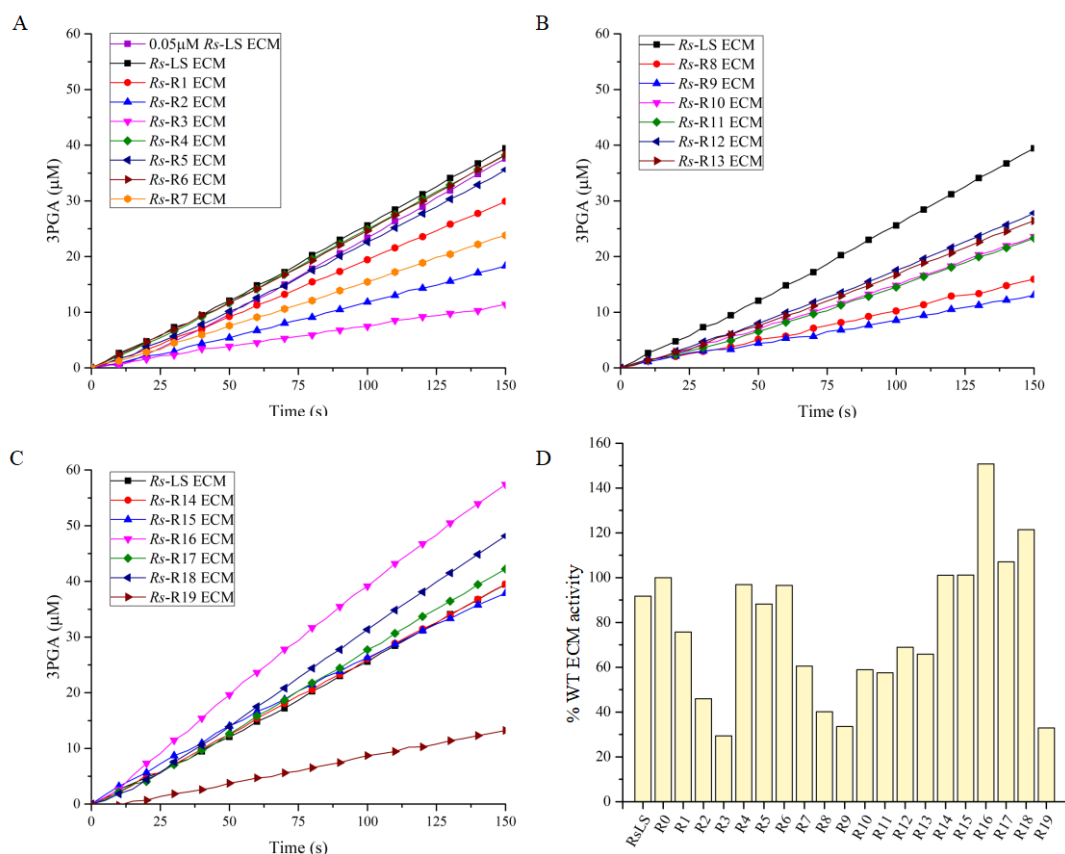


Figure 4-6: Rubisco activity of fully activated Rubisco complexes (ECM) of *Rs*-Rubisco variants assayed using soluble *E. coli* extract.

(A) ECM activity of wild type *Rs*LS (purified, 0.05 μM *Rs*LS and soluble *E. coli* extract, *Rs*LS) and variants *Rs*-R1 to *Rs*-R7. (B) ECM activity of variants *Rs*-R8 to *Rs*-R13. (C) ECM activity of variants *Rs*-R14 to *Rs*-R19. (D) Relative ECM activity of Rubisco variants expressed as a percentage of wild type *Rs*LS (soluble extract). The ECM activities of Rubisco variants were calculated by fitting the plot using a linear fit. The amount of Rubisco (ECM) assayed was from 1.5×10^7 of *E. coli* cells which were equivalent to approximately 0.05 μM of purified Rubisco. The fully activated Rubisco complex (ECM) was formed by incubating supernatant for 10 minutes with 40 mM NaHCO_3 and 10mM MgCl_2 . The assay was carried out in 20 mM NaHCO_3 at 25°C. *Rs*LS is the purified *Rs*-Rubisco while R0 is the *Rs*-Rubisco from soluble *E. coli* extract. The Rubisco variants were labeled with their respective short derivative (R1-R19). The labeling of these *Rs*-Rubisco variants in figures will be consistent for the rest of this chapter. The 3PGA production versus time and ECM rate is calculated using Equation 2.2 and Equation 2.3 respectively.

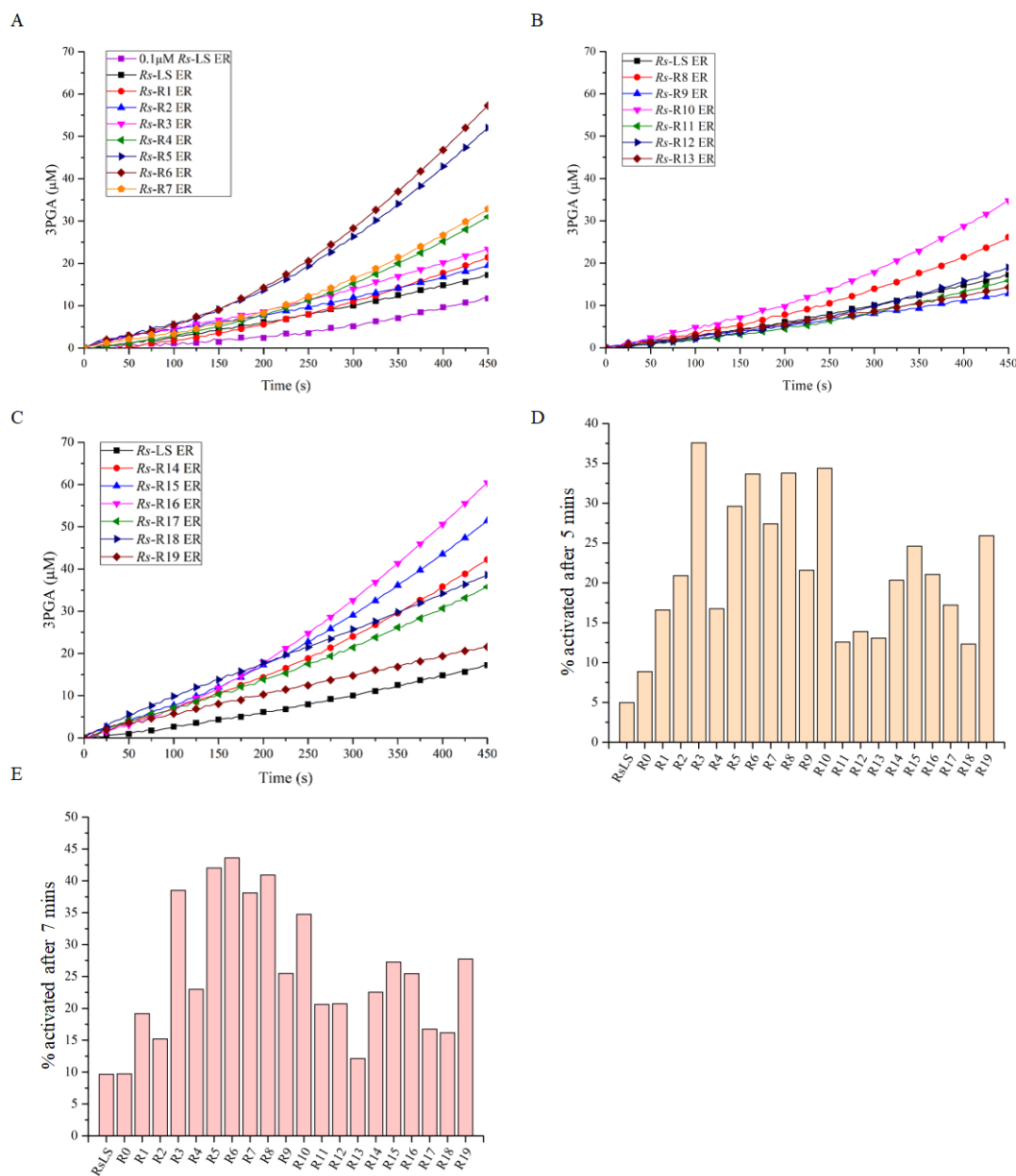


Figure 4-7: Time course of spontaneous activation of RuBP-inhibited Rubisco complexes (ER) of *Rs*-Rubisco variants assayed with soluble *E. coli* extract.

(A) ER activity of wild type RsLS (purified, 0.1 μM Rs-LS and soluble extract, Rs-LS) and its variants *Rs*-R1 to *Rs*-R7. (B) ER activity of variants *Rs*-R8 to *Rs*-R13. (C) ER activity of variants *Rs*-R14 to *Rs*-R19. (D, E) Relative ECM activity of the ER complexes at five or seven minutes expressed as a percentage of their fully activated ECM activity. The overall trend of the percentage ECM activity was similar between five and seven minutes. The ECM activity of ER at five minutes and seven minutes were obtained using first-order derivative and value was divided by two as the amount of soluble extract loaded for ER used was double of the ECM. This ER's ECM activity was subsequently expressed as a percentage of their respective fully activated Rubisco complex. The amount of Rubisco in the soluble extract assayed for ER activity was from 3×10^7 *E. coli* cells. The RuBP inhibited Rubisco complexes (ER) were formed

by pre-incubating either the purified or the soluble extract with 4 mM EDTA for 10 minutes at room temperature followed by incubation with 0.8mM RuBP. The assay was carried out in 20 mM NaHCO₃ at 25°C.

4.2.2 Characterizing *Af*-Rubisco variants using soluble *E. coli* extract

The characterization of *Af*-Rubisco variants, *Af*-B1 to *Af*-B22 was carried out using soluble *E. coli* extract in the same fashion as *Rs*-Rubisco variants. The amount of soluble *E. coli* extract used for ECM and ER activity of *Af*-Rubisco were also from 1.5x 10⁷ and 3.0 x 10⁷ *E. coli* cells respectively. The ECM activity of Rubisco in the soluble extracts from 1.5x 10⁷ *E. coli* cells was optimized to be equivalent to approximately 0.1 µM purified *Af*-Rubisco (Figure 4-8). Similar to *Rs*-Rubisco, the *Af*-Rubisco from the soluble extract and purified source had similar ECM and ER activity.

Most *Af*-Rubisco variants had lower ECM activity than the wild type with the exception of *Af*-B20. *Af*-B17 and *Af*-B21 which had a similar ECM rate as wild type. *Af*-B2, *Af*-B18, and *Af*-B19 had the lowest ECM activities with less than 15% activity of wild type and this may be attributable to lower protein expression of these variants.

All the *Af*-Rubisco variants exhibited an increased rate of spontaneous activation where they had at least 4% activation after seven minutes as compared to the 2% in wild type *Af*LS (Figure 4-9E). This could suggest that these *Af*-Rubisco variants form a less stable ER complex as compared to the wild type. These variants with a high spontaneous rate of activation were of interest and among these variants, those exhibiting a higher than 50% wild type ECM activity were chosen for further analysis. The variants were *Af*-B1, *Af*-B5, *Af*-B6, *Af*-B7, *Af*-B8, *Af*-B9, *Af*-B12, *Af*-B17, and *Af*-B21. *Af*-B20 also fulfilled the criteria but was not further characterized as it contained multiple mutated residues and therefore, the structural analysis could be impossible. *Af*-B13 was also chosen despite falling short of the criteria with approximately 45% ECM activity as wild type as it contained the K463E substitution. This substitution was repeatedly selected for in three different variants (*Af*-B13, *Af*-B14, and *Af*-B15) which suggest that this substitution was under genuine selection. *Af*-B13 was selected out of the three variants as it had only two substitutions in comparison to three substitutions in *Af*-B15 and had a higher spontaneous rate of activation as compared to *Af*-B14.

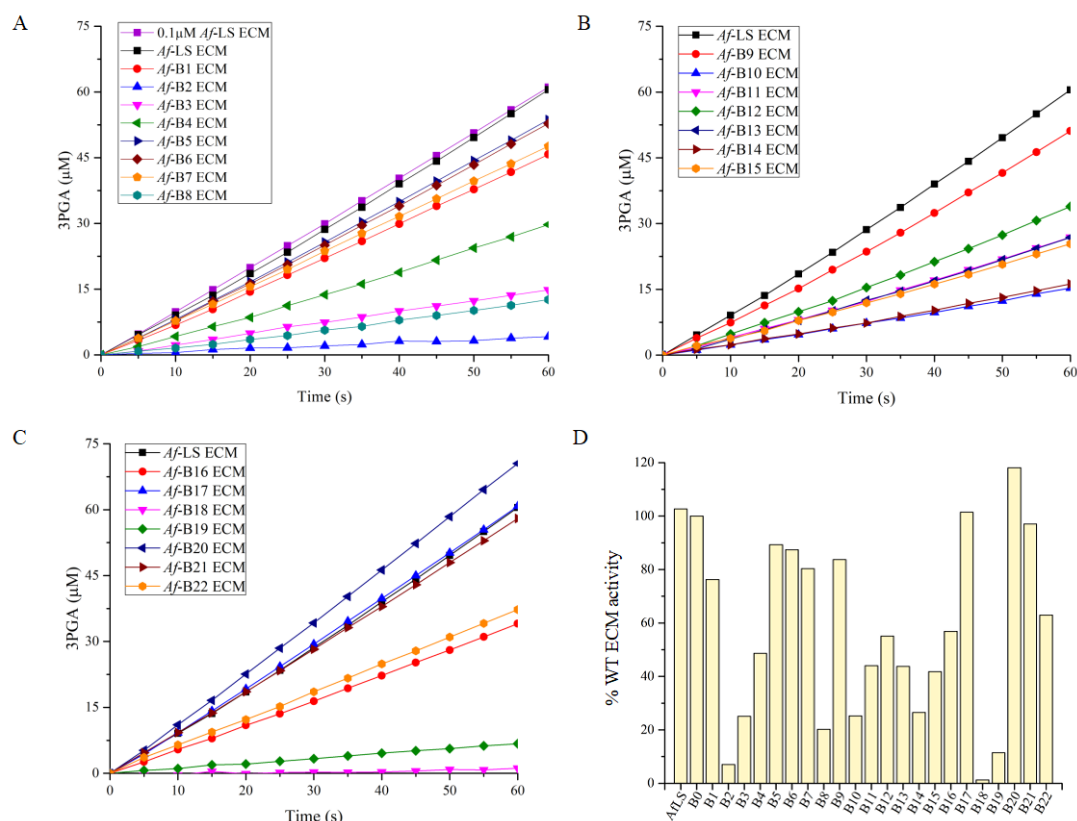


Figure 4-8: Rubisco activity of fully activated ECM complex of *Af*-Rubisco variants assayed using soluble *E. coli* extract

(A) ECM activity of wild type *Af*LS (purified, 0.1 μM *Af*LS and soluble *E. coli* extract, *Af*LS) and variants *Af*-B1 to *Af*-B8. (B) ECM activity of variants *Af*-B9 to *Af*-B15. (C) ECM activity of variants *Af*-B16 to *Af*-B22. (D) Relative ECM activity of *Af*-Rubisco variants as a percentage of wild type *Af*LS Rubisco (soluble extract, B0). The ECM activities of Rubisco variants were calculated by fitting the plot using a linear fit. The amount of Rubisco assayed was from 1.5×10^7 of *E. coli* cells which is which was equivalent to approximately 0.01 μM of purified Rubisco. The assay was carried out in 20 mM NaHCO_3 at 25°C. *Af*LS is the purified *Af*-Rubisco and B0 is *Af*-Rubisco from the soluble extract. The mutant Rubiscos were labeled with their respective short derivative (B1-B22). The labeling of these *Af*-Rubisco variants in figures will be consistent for the rest of this chapter.

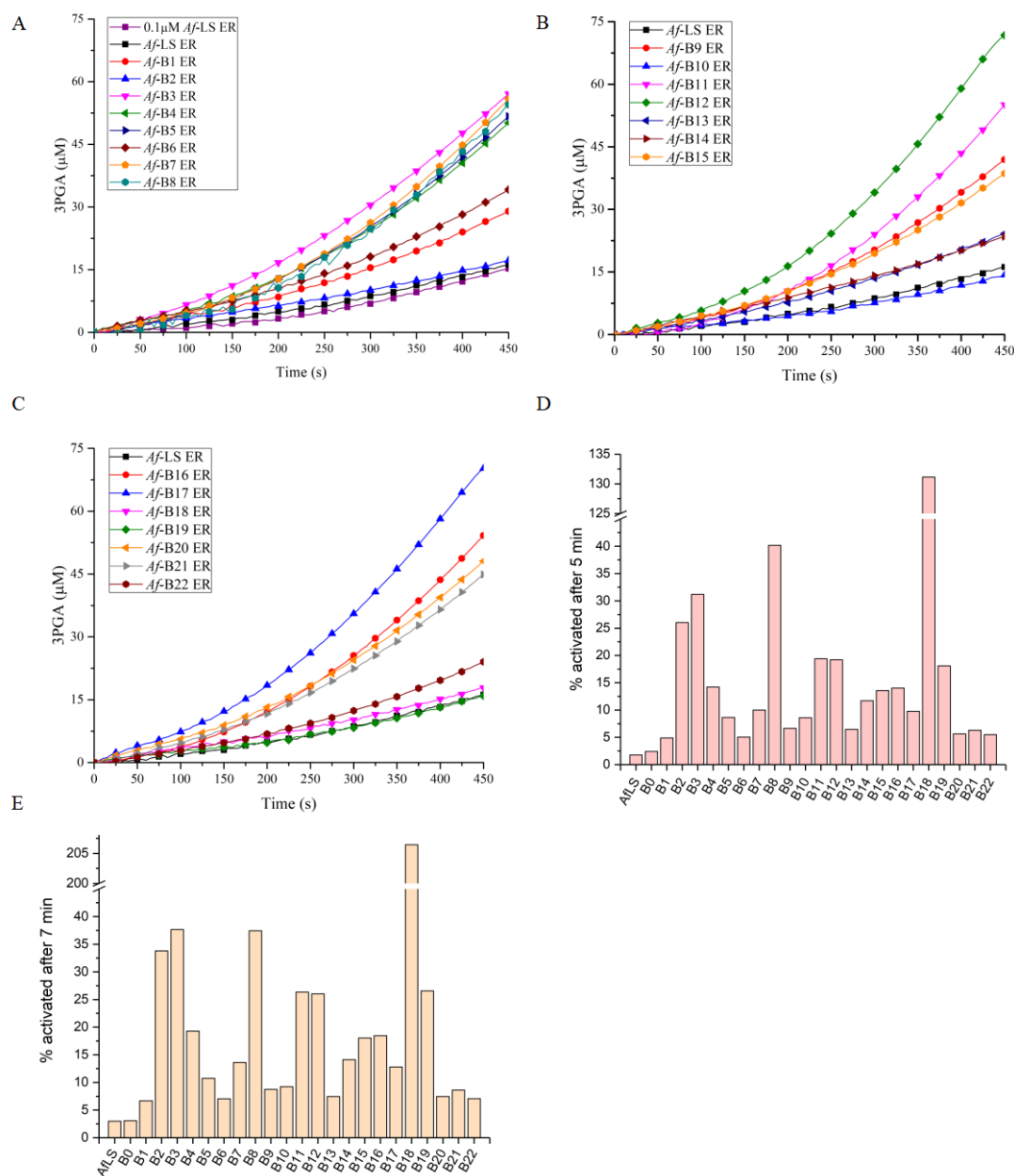


Figure 4-9: Time course of spontaneous activation of RuBP-inhibited Rubisco complexes (ER) *Af*-Rubisco variants assayed using soluble *E. coli* extract.

(A) ER activity of wild type *Af*LS (purified, 0.1 μM *Af*LS and soluble *E. coli* extract, *Af*LS) and its variants *Af*-B1 to *Af*-B8. (B) ER activity of variants *Af*-B9 to *Af*-B15. (C) ER activity of variants *Af*-B16 to *Af*-B22. (D, E) Relative ECM activity of the ER complexes after five or seven minutes expressed as a percentage of their fully activated ECM activity. The amount of Rubisco assayed was from 3.0×10^7 of *E. coli* cells and these cells contained approximately 0.01 μM of purified Rubisco. Rubisco assay was carried out in 20 mM NaHCO_3 at 25°C.

4.3 Purification of selected *R*s-Rubisco and *A*f-Rubisco variants

The initial characterization of the *R*s-Rubisco and *A*f-Rubisco using *E. coli* soluble extracts provided initial insights into the selected variant's activity and guided further analysis. Large-scale *E. coli* cultures of 0.5-1L were grown and soluble extracts were generated. The Rubisco variants were purified from the supernatant using anion exchange column (Source 30Q) followed by a size-exclusion column (Superdex 200 16/600) as shown in Figure 4-10 (Gibson et al., 1977; Mueller-Cajar et al., 2011; Tsai et al., 2015). From the SDS-PAGE and the Native-PAGE, *R*s-Rubisco and *A*f-Rubisco variants were pure and migrated as monodisperse L₈S₈ Rubisco complexes (Figure 4-10E-H). These purified Rubisco variants were also characterized using the spectrophotometric Rubisco assay.

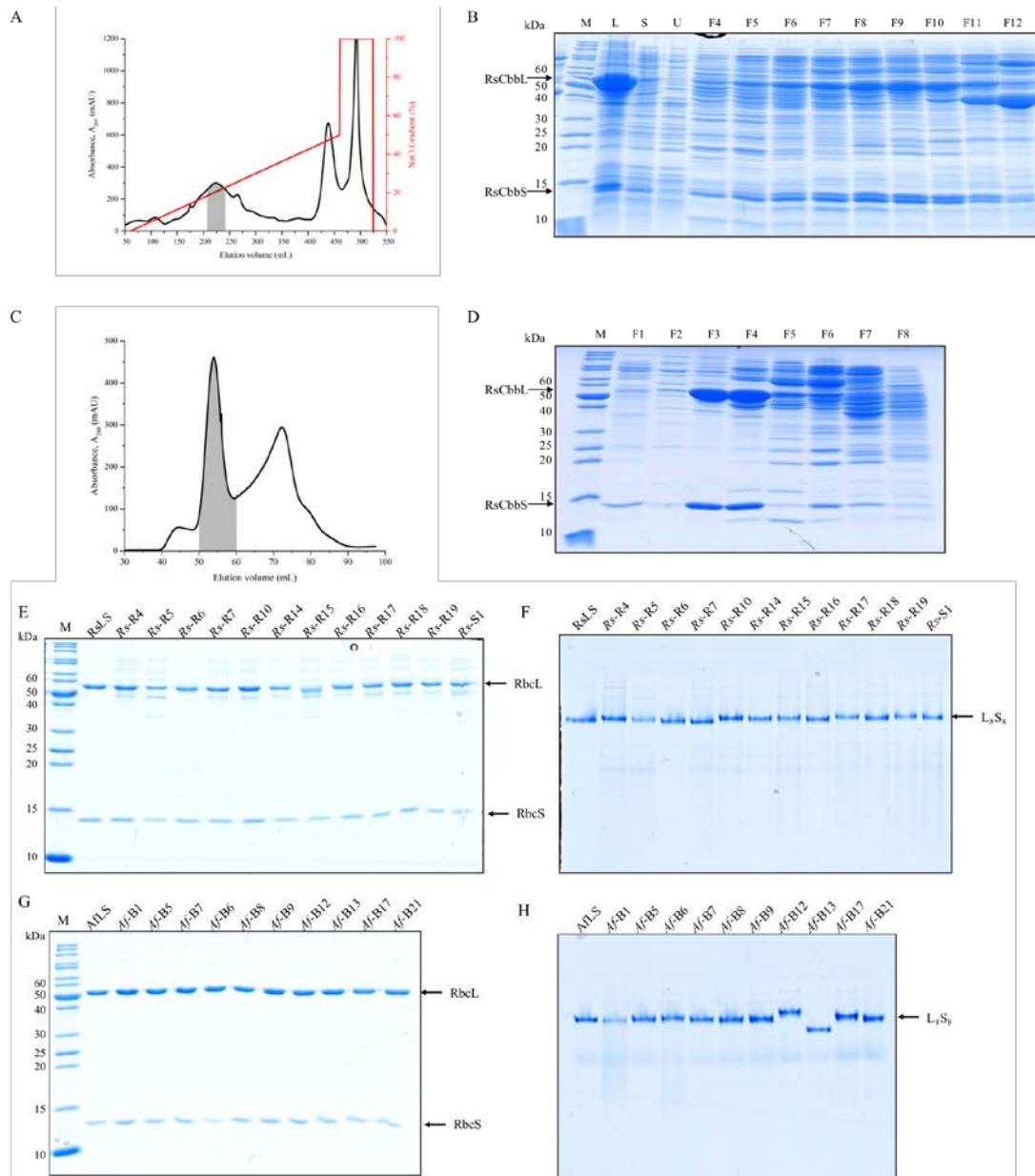


Figure 4-10: Representative purification of *Rs*-Rubisco and *Af*-Rubisco.

(A, B) Purification of *Rs*-Rubisco using a linear gradient of NaCl (0- 0.5 M NaCl) by anion exchange chromatography using a Source 30Q column. Eluted proteins were analyzed using 16% SDS-PAGE and Rubisco started to elute at 160 mM of NaCl (F5). The area shaded in grey indicates the fractions (F7-F10) that contained high Rubisco concentrations. These fractions were combined and concentrated prior to further purification by size exclusion chromatography. (C, D) Size exclusion purification of *Rs*-Rubisco using a Superdex 200 16/600 column. Proteins eluted from the column was analyzed by SDS-PAGE and Rubisco rich fractions were found to be F3 and F4 (area shaded in grey in chromatogram). These two fractions were pooled and concentrated using Amicon Ultra-15 Centrifugal Filter Units (100 kDa cut-off). (E, F) SDS-PAGE and Native-PAGE of purified wild type and variants of *Rs*-Rubisco. (G, H) SDS-PAGE and Native-PAGE of purified wild type and variants of *Af*-Rubisco. ~1 μ g of purified Rubisco was loaded and analyzed using 16% SDS-page under denaturing conditions. ~2 μ g of purified Rubisco was loaded and analyzed using 6% Native-page.

4.4 Characterization of selected purified Rs-Rubisco and Af-Rubisco variants

The purified enzymes were assayed using the spectrophotometric Rubisco assay to determine the carboxylation activity and rates of inhibitor release. The carboxylase activity of Rubisco is determined using a fully activated Rubisco (ECM) whereby the active site is carbamylated with a non-substrate CO₂ and bound by a Mg²⁺ ion co-factor (Lorimer et al., 1976). Whereas, the rate of RuBP release was determined using RuBP-inhibited complexes (ER). The ER complex was prepared by pre-incubating the apo-Rubisco with RuBP and subsequently the recovery of the activity was measured using a reaction mixture with saturating bicarbonate, Mg²⁺, and RuBP. The activity increased from an initial zero to a final steady state with a low initial velocity (v_i) and a high final velocity (v_f) (Figure 4-11C). The v_i for the ER complexes are slow ($\sim 0 \text{ s}^{-1}$) as RuBP is bound to the active site and Rubisco were not able to catalysis RuBP carboxylation. Over time, the inhibitors spontaneously release from the active site which free up the active site for activation and subsequently carrying out RuBP carboxylation. Therefore, the v_f of ER Rubisco complexes will be higher than v_i .

The product accumulation over time progression curve of ER complexes can be fitted into Equation 2.4 using non-linear regression software in OriginLab to estimate the v_i , v_f and first-order rate constant (k_{obs}) (Figure 4-11B)(Morrison, 1982; Pearce, 2006; Pearce et al., 2003). The v_f describe the percentage of Rubisco population that has become activated following inhibitor release and a v_f value similar to the carboxylation rate indicates that the majority of the Rubisco complexes were no longer inhibited. We thus take v_f to indicate the activation state at equilibrium under our assay conditions. k_{obs} describes the transition from the v_i to v_f and it provides an estimate on the rate of RuBP release from ER complex (Pearce, 2006; Pearce et al., 2003).

RuBP and other sugar phosphate inhibitor of Rubisco are considered slow-tight binding inhibitor where the initial formation of inhibited complex (EI) is rapid followed by a slower isomerization phase in which a much tighter bound inhibited complex is formed (EI*). The spontaneous release of inhibitor from EI* complex varies between species of Rubisco and type of sugar phosphate. For instance, in tobacco, the k_{obs} for XuBP ($0.6 \times 10^{-3} \text{ s}^{-1}$) is lower than that of RuBP ($2.09 \times 10^{-3} \text{ s}^{-1}$) which suggest that XuBP are bound tighter compared to RuBP (Pearce, 2006). In comparison, the k_{obs} for XuBP ($28.8 \times 10^{-3} \text{ s}^{-1}$) and RuBP ($28.9 \times 10^{-3} \text{ s}^{-1}$) release from inhibited complexes of Syn-Rubisco were

similar which suggest that *Syn*-Rubisco has low affinity for these two inhibitors. Therefore, the ER complex of *Syn*-Rubisco was able to regain nearly fully carboxylation activity after a minute. On the other hand, *G. sulphuraria* Rubisco has high affinity towards RuBP and XuBP as k_{obs} for both sugar phosphate is low ($0.13 \times 10^{-3} \text{ s}^{-1}$, $0.07 \times 10^{-3} \text{ s}^{-1}$) (Pearce, 2006). The estimated k_{obs} for *Rs*-Rubisco and *Af*-Rubisco were $8.26 \times 10^{-4} \text{ s}^{-1}$ and $2.67 \times 10^{-4} \text{ s}^{-1}$ (Table 4-7) which was similar to that of *G. sulphuraria*. This indicates RuBP release is very slow from the inhibited complex. It suggests the active site of *Rs*-Rubisco and *Af*-Rubisco are rigid resulting in tightly bound RuBP.

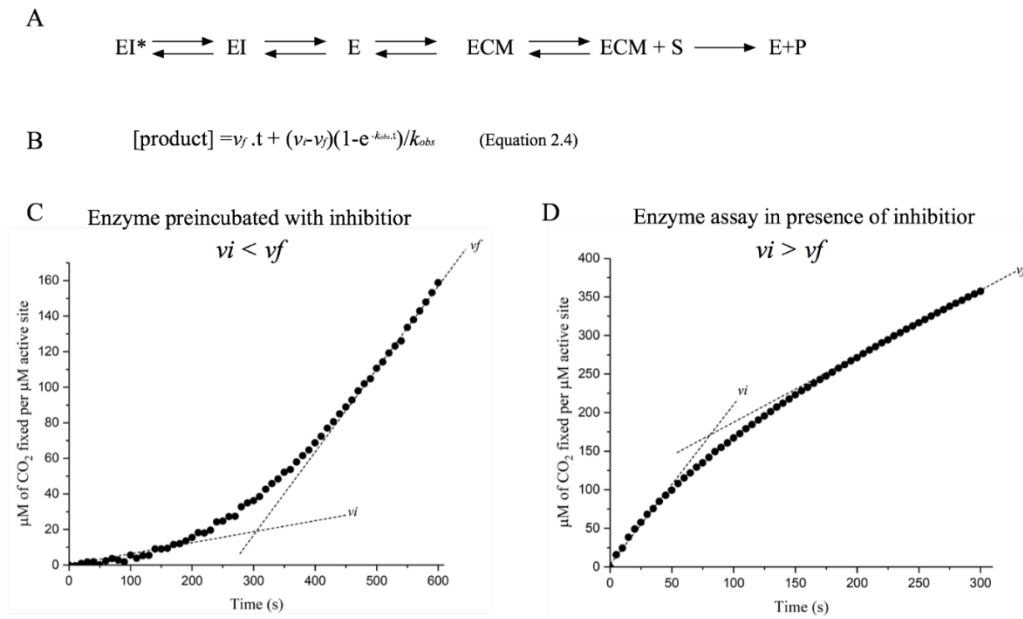


Figure 4-11: Kinetics of slow-binding inhibitors of Rubisco.

(A) Mechanism of Rubisco activation and inhibition. The binding of inhibitors to Rubisco use the slow-tight binding mechanism where the initial formation of inhibited complex (EI) was rapid followed by a slow isomerization phase to form a tightly bounded inhibited complex (EI^*). (B) Equation 2.4 that is used for fitting of progression curve of ER complexes to obtained v_i , v_f and k_{obs} . (C) Representative progression curve of the activity of RuBP-inhibited Rubisco complex. (D) Representative progression curve of the activity of Rubisco assayed in presence of inhibitor. This progression curve is also representative of the fallover effect whereby Rubisco activity declines over time due to an accumulation of misfired product. The initial (v_i) and final (v_f) velocities are defined by the slope values in the early and late stage of the curve as shown by the dashed lines. E: apo-Rubisco, ECM: Fully activated Rubisco, I: inhibitor, EI: Rubisco bound with inhibitor, EI^* : Rubisco tightly bound with inhibitor. S: substrate, P: product. v_i : initial velocity, v_f : final steady-state velocity, k_{obs} : apparent first-order rate constant characterizing the transient.

4.4.1 Characterization of purified *Rs*-Rubisco variants

The overall trend of the carboxylation rate of activated Rubisco complex of purified *Rs*-Rubisco was similar to those using soluble extract (Figure 4-12, Table 4-7). The carboxylation rate of wild type *Rs*-Rubisco was calculated from the ECM activity using a linear fit to be 2.8 s^{-1} . This result was similar to the previously reported carboxylation rate of *Rs*-Rubisco (Bhat et al., 2017a). The carboxylation rate of the majority of the variants was lower than wild type, with the exception of *Rs*-R16, *Rs*-R17, and *Rs*-R18 (Figure 4-12F, Table 4-7). *Rs*-R16, *Rs*-R17 and *Rs*-R18 had higher carboxylation rate of 3.5 s^{-1} , 3.3 s^{-1} , and 3.6 s^{-1} respectively. The ECM activity of *Rs*-R15 showed that *Rs*-R15 exhibited a fallover effect where there was a decline in Rubisco activity over time due to inhibition by misfire products (Edmondson et al., 1990). It is possible that substitution in *Rs*-R15 may have reduced the rigidity of the active site such that side reactions are more common and misfired product such as XuBP and PDBD are produced. The fallover effect is similar to the progression curve of an enzyme assayed in presence of their inhibitor where v_f is lower than v_i (Figure 4-11D). *Rs*-R14, *Rs*-R16, and *Rs*-R19 also show slight curvature in their ECM activity over time which suggests that these variants may also be exhibiting the fallover effect. Therefore, the carboxylation rates of these Rubiscos were calculated using a linear fit from the first 80 seconds where the activity was linear and yet to not declined due to inhibition.

The activity of ER complexes of *Rs*-Rubisco variants was assayed and the v_f and k_{obs} for *Rs*LS were estimated to be 0.41 s^{-1} and $8.3 \times 10^{-4} \text{ s}^{-1}$ (Table 4-7, Figure 4-13). The overall trend of ER activity was slightly different between soluble extract and purified protein assay. *Rs*-R4, *Rs*-R6, *Rs*-R7, *Rs*-R14, *Rs*-R15, and *Rs*-R16 all had a high spontaneous rate of activation based on percentage activation of the ER complex and this was consistent with the high k_{obs} observed for these variants where their k_{obs} were at least two-fold of the wild type. Therefore, these variants were likely to form unstable inhibited complexes where RuBP rapidly dissociate from their active site. There were two other variants which had higher slightly higher k_{obs} rate than wild type and they were *Rs*-R5 and *Rs*-R19. All these variants with higher k_{obs} were likely to provide fitness by increasing the rate of inhibitor release which subsequently frees up the active site for RuBP turnover. It is likely that the substitution in these variants had cause conformational changes such that the closed active site becomes loosen and this

facilitated the inhibitor release. This loosening of the active site may also affect other catalytic properties of Rubisco and it was observed that majority of these *Rs*-Rubisco variants with increased in their k_{obs} had a trade-off in their carboxylation rate. However, *Rs*-R16 was an exception as it increased in k_{obs} while having higher ECM activity than wild type. In addition, many of these variants with increased k_{obs} also had higher v_f than wild type and they were *Rs*-R4, *Rs*-R5, *Rs*-R6, and *Rs*-R7. The high v_f meant that there was an increased tendency towards forming the activated state (ECM) compared to the wild-type enzyme. The equilibrium is shown in Fig. 4.10A thus favors the ECM state compared to the wild-type, and there will thus be an increased RuBP turnover.

In contrast to the previously described *Rs*-Rubisco variants, *Rs*-R10, *Rs*-R17, and *Rs*-R18 had similar k_{obs} as wild type and this was different from the previous assay using soluble extract where their ER activity had greater percentage activation as compared to than the wild type. This similar value in k_{obs} suggests that the rate of spontaneous inhibitor release from these variants ER complexes was similar to wild type and these variants were likely to form stable inhibited complexes. However, a comparing the v_f of the variants to wild type Rubisco show that *Rs*-R10 had high v_f while *Rs*-R17 and *Rs*-R18 had similar v_f as wild type. The high v_f suggests that *Rs*-R10 (like *Rs*-R4-7) favors the activated state (ECM). It is possible that *Rs*-R17 and *Rs*-R18 may have evolved other solutions dependent on the *E. coli* intracellular environment and this would warrant further investigation.

The *Rs*-S1 variant was obtained from RADE screening whereby *Rs*-Rubisco was evolved in the presence of its activase. *Rs*-S1 had lower carboxylation rate, lower k_{obs} , and higher v_f as compared to the wild type (Figure 4-12E, F). The low k_{obs} would suggest that the active site remain rigid and RuBP is still bound tightly to the active site. This stable inhibited complex formation is necessary to maintain compatibility with the activase. In addition, *Rs*-S1 has a higher v_f and in the presence of its activase, the equilibrium would be pushed towards the formation of an activated state where all the free Rubisco (E) would become activated and carry out RuBP catalysis. The decrease in carboxylation rate observed in *Rs*-S1 could be trade-off due to conformational changes in the active site. Given that only one variant was characterized for the RADE selection, it is currently premature to speculate about the properties selected for in this system.

Rubisco	Amino acid change(s)		ECM		ER		r^2
	RbcL	RbcS	Carboxylation rate (s^{-1})	v_i (s^{-1})	v_f (s^{-1})	k_{obs} (s^{-1})	
RsLS	nil	nil	3.0 ± 0.060	$4.7 \pm 0.14 \times 10^{-2}$	0.41 ± 0.052	$8.3 \pm 0.53 \times 10^{-4}$	0.9987
Rs-R4	T73A	nil	2.2 ± 0.11	$9.6 \pm 0.11 \times 10^{-13}$	1.5 ± 0.062	$28 \pm 7.8 \times 10^{-4}$	0.9998
Rs-R5	T73S	nil	1.8 ± 0.18	$2.9 \pm 3.8 \times 10^{-15}$	1.9 ± 0.12	$16 \pm 0.21 \times 10^{-4}$	0.9995
Rs-R6	R83C	nil	1.8 ± 0.42	$2.9 \pm 0.11 \times 10^{-17}$	1.7 ± 0.14	$19 \pm 0.21 \times 10^{-4}$	0.9996
Rs-R7	R83S	K73R	2.1 ± 0.25	$4.1 \pm 5.2 \times 10^{-15}$	1.9 ± 0.0020	$22 \pm 1.4 \times 10^{-4}$	0.9997
Rs-R10	nil	F91I, T95S, S108C	2.5 ± 0.25	$2.8 \pm 0.0 \times 10^{-14}$	3.3 ± 0.0090	$11 \pm 0.28 \times 10^{-4}$	0.9996
Rs-R14	Y26F	nil	2.7 ± 0.13	$5.9 \pm 8.3 \times 10^{-3}$	0.45 ± 0.042	$28 \pm 0.14 \times 10^{-4}$	0.9995
Rs-R15	P52L	nil	2.0 ± 0.030	$5.3 \pm 4.1 \times 10^{-2}$	0.44 ± 0.013	$36 \pm 3.5 \times 10^{-4}$	0.9998
Rs-R16	T77A	nil	3.6 ± 0.15	$5.2 \pm 0.86 \times 10^{-16}$	0.81 ± 0.018	$36 \pm 0.71 \times 10^{-4}$	0.9999
Rs-R17	N228S	nil	3.34 ± 0.34	$7.1 \pm 0.22 \times 10^{-2}$	0.55 ± 0.022	$9.4 \pm 0.31 \times 10^{-4}$	0.9995
Rs-R18	L107M	nil	3.84 ± 0.11	$5.0 \pm 2.5 \times 10^{-2}$	0.55 ± 0.052	$7.2 \pm 0.89 \times 10^{-4}$	0.9992
Rs-R19	L107Q	nil	2.93 ± 0.22	$6.3 \pm 7.3 \times 10^{-3}$	0.31 ± 0.050	$14 \pm 4.4 \times 10^{-4}$	0.9986
Rs-S1	I315T	nil	2.23 ± 0.16	$0.78 \pm 1.1 \times 10^{-2}$	1.0 ± 0.056	$4.5 \pm 1.3 \times 10^{-4}$	0.9987

Table 4-7: Summary of kinetic parameters purified *Rs*-Rubisco variants assayed using spectrophotometric assay

Data for this table was taken from Figure 4-12 and Figure 4-13.

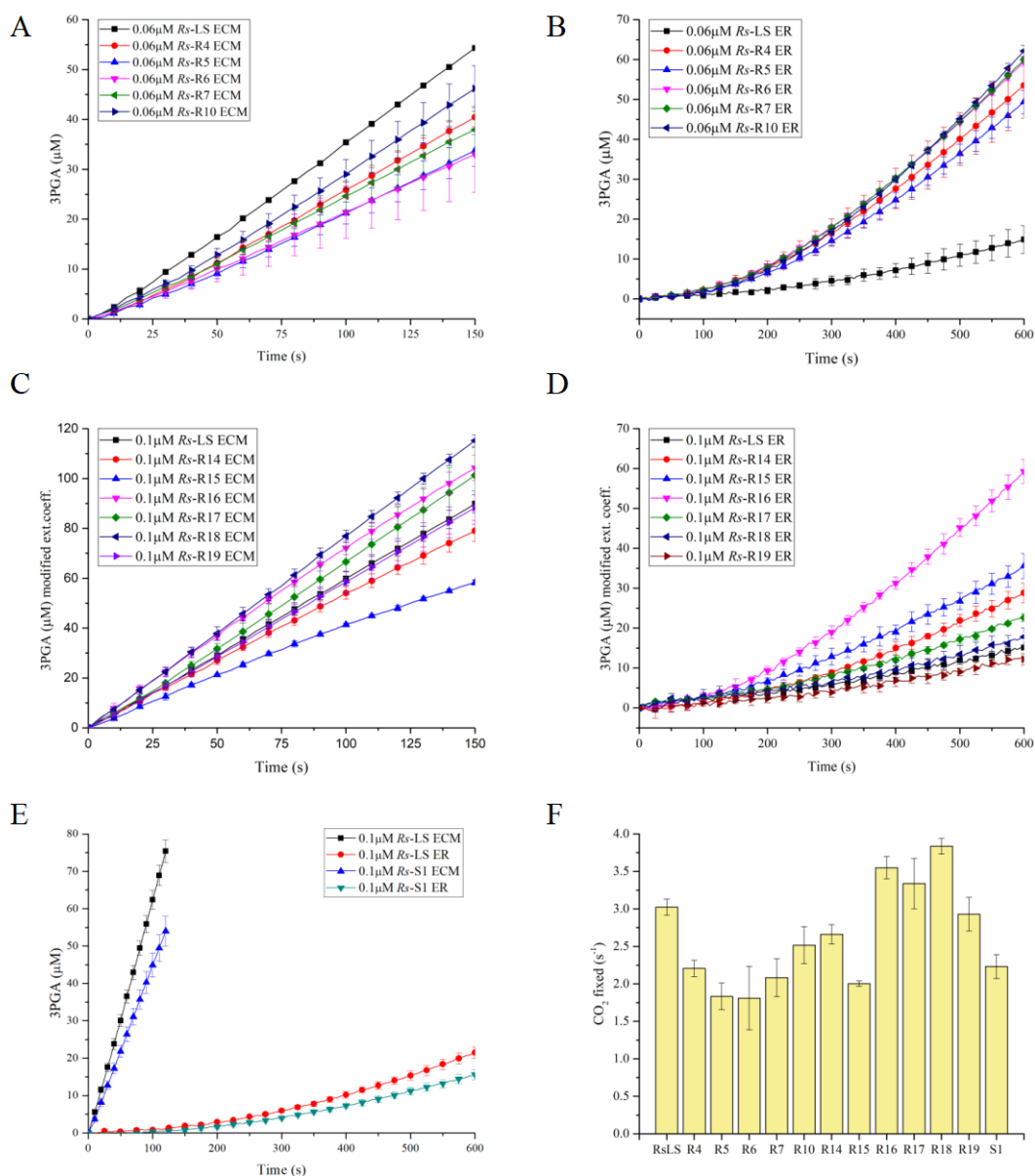


Figure 4-12: Rubisco activity of ECM and ER complexes assayed with purified *Rs*-Rubisco variants.

(A) ECM activity of 0.06 μM of *Rs*-Rubisco variants (*Rs*-R4, *Rs*-R5, *Rs*-R6, *Rs*-R7, and *Rs*-R10). (B) ER activity of 0.06 μM of *Rs*-Rubisco variants (*Rs*-R4, *Rs*-R5, *Rs*-R6, *Rs*-R7, and *Rs*-R10). (C) ECM activity of 0.1 μM of *Rs*-Rubisco variants (*Rs*-R14 to *Rs*-R19). (D) ER activity of 0.1 μM of *Rs*-Rubisco variants (*Rs*-R14 to *Rs*-R19). (E) ECM and ER activity of 0.1 μM *Rs*-S1 Rubisco variant. (F) Carboxylation rate of the *Rs*-Rubisco variants. The extinction coefficient NADH for activities measures in (C) and (D) was modified from 0.00622 to 0.00346, this was to account for the shift in wavelength due to deterioration of light source. The calibration of the wavelength was performed using the carboxylation rate of wild type *Rs*-Rubisco. The ECM and ER complex preparation was same as those for soluble extract. The assay was carried out in 20

mM NaHCO₃ at 25 °C. Error bars indicate the mean and s.d. of at least three independent experiments.

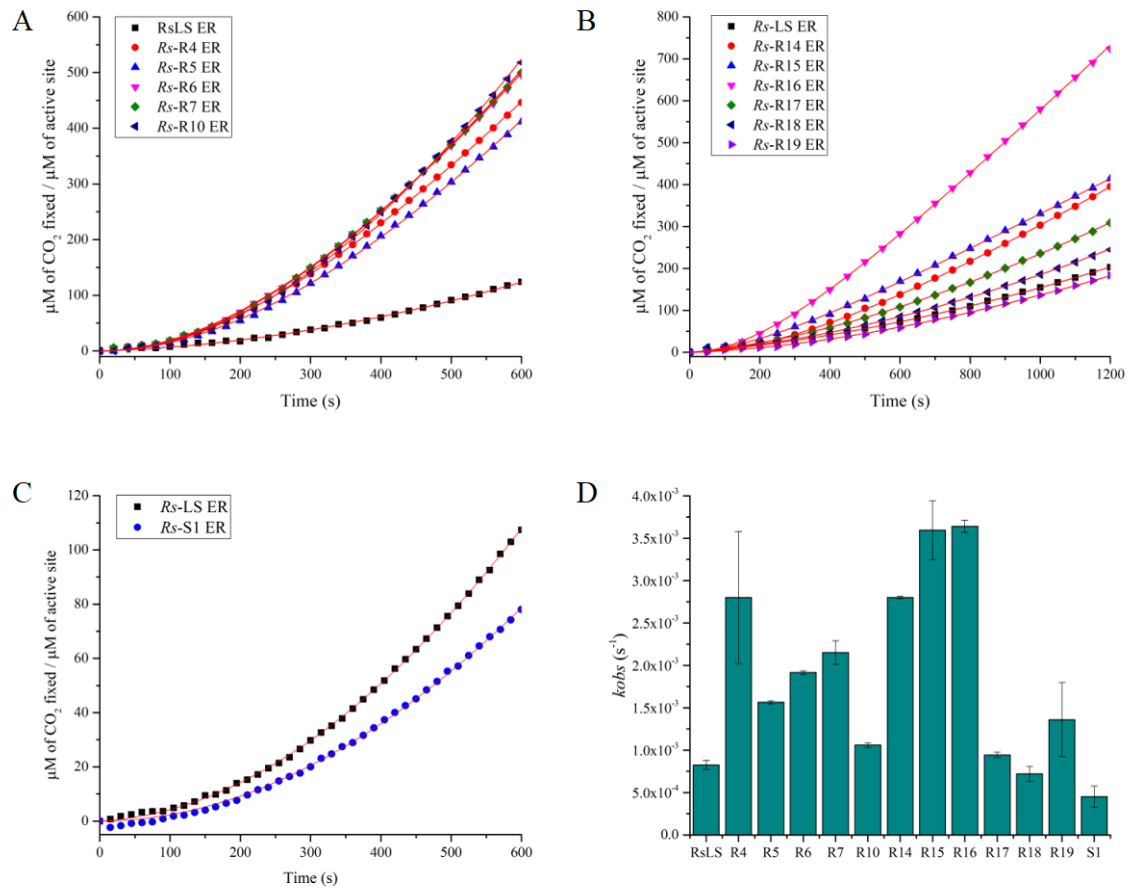


Figure 4-13: Rate of RuBP-inhibitor (k_{obs}) release for *Rs*-Rubisco variant.

(A), (B), (C) The red line in the graph is the fitting of Equation 2.4 to the data set. The data set for product accumulation over time shown was from a single reaction. (D) The rate of inhibitor release (k_{obs}) for the *Rs*-Rubisco variants. Error bars indicate the mean and s.d. of at least two sets of data fitting.

4.4.2 Characterization of mutant *A. ferrooxidans* Rubisco using purified Rubisco

Af-Rubiscos were purified using anion exchange chromatography followed by size exclusion column and example of the purification was shown in Figure 4-10. The ECM and ER activity of these purified *Af*-Rubisco variants were assayed and carboxylation rate of wild type *Af*LS was calculated based on the linear activity of ECM to be 3.9 s^{-1} (Table 4-8 and Figure 4-14). This calculated ECM rate was slower than the report rate of 5.2 s^{-1} (Tsai et al., 2015) however, this value was consistently obtained in the assay.

Therefore, this value will be used as a baseline for comparison with the ECM activity of *Af*-Rubisco variants.

All the *Af*-Rubisco variants had lower carboxylation rate and higher k_{obs} as compared to the wild type (Figure 4-14, Figure 4-15). This finding was consistent with the results obtained from using soluble extract where these Rubisco variants had carboxylation rate and high rate of spontaneous release. It is possible that the substitutions in these *Af*-Rubisco variants alter the local conformation near that influence that loosens the active site such that the inhibitors are no longer form a tightly bound inhibited complexes. This potential changes in closed active site conformation were likely to affect the Rubisco's catalytic activity as shown by the decreased in carboxylation rate. This strategy of increasing inhibitor rate of release would be similar to those undertaken by *Rs*-Rubisco variants.

The wild type *Af*-Rubisco had a v_f of 2.9 s^{-1} and this suggests that *Af*-Rubisco tends to eventually form the activated state as compared to the mostly inhibited state observed in *Rs*-Rubisco which has a lower v_f of 0.41 s^{-1} (Table 4-8). These variants selected for *Af*-Rubisco had a higher tendency towards activated state as compared to those *Rs*-Rubisco variants with the *Af*-Rubisco variants having higher v_f . Based on the analysis of ER complexes of both *Rs*-Rubisco and *Af*-Rubisco variants from RDE screening, it was observed that the Rubisco variants selected could be broadly categorized into three categories; higher v_f , higher k_{obs} or both higher v_f and k_{obs} . This meant that variants selected had changes in their tendency to become activated (high v_f), and/or exhibit an increased rate of inhibitor release (high k_{obs}).

Rubisco	Amino acid change(s)		ECM		ER		
	RbcL	RbcS	Carboxylation rate (s ⁻¹)	v_i (s ⁻¹)	v_f (s ⁻¹)	k_{obs}	r ²
AfLS	nil	nil	3.9 ± 0.30	7.7 ± 0.11 x 10 ⁻³	2.9 ± 0.87	2.7 ± 0.25 x 10 ⁻⁴	0.9992
Af-B1	K41R, I105V	nil	3.1 ± 0.44	5.1 ± 7.2 x 10 ⁻⁴	2.3 ± 0.48	9.4 ± 0.65 x 10 ⁻⁴	0.9998
Af-B5	S61T	nil	2.5 ± 0.28	5.4 ± 7.2 x 10 ⁻¹⁶	3.4 ± 0.32	5.9 ± 0.63 x 10 ⁻⁴	0.9998
Af-B6	F220L	nil	1.3 ± 0.10	0.72 ± 1.0 x 10 ⁻²	1.1 ± 0.010	8.6 ± 0.20 x 10 ⁻⁴	0.9997
Af-B7	P305L	nil	2.1 ± 0.03	2.8 ± 0.90 x 10 ⁻¹⁶	2.0 ± 0.010	17 ± 5.8 x 10 ⁻⁴	0.9999
Af-B8	P305S	nil	2.8 ± 0.10	2.8 x 10 ⁻¹⁶ ± 1.2 x 10 ⁻²⁰	2.3 ± 0.27	9.4 ± 0.73 x 10 ⁻⁴	0.9985
Af-B9	T392A	nil	3.5 ± 0.29	3.0 ± 3.8 x 10 ⁻¹⁵	3.1 ± 0.050	5.5 ± 1.5 x 10 ⁻⁴	0.9993
Af-B12	E438G	nil	2.5 ± 0.17	1.4 ± 2.0 x 10 ⁻¹⁴	2.3 ± 0.82	18 ± 2.5 x 10 ⁻⁴	0.9949
Af-B13	I42T, K463E	nil	3.3 ± 0.27	9.5 ± 9.4 x 10 ⁻¹⁶	3.4 ± 0.16	5.0 ± 0.86 x 10 ⁻⁴	0.9985
Af-B17	I465N	nil	2.7 ± 0.52	2.8 x 10 ⁻¹⁶ ± 2.1 x 10 ⁻²¹	1.8 ± 0.15	13 ± 3.1 x 10 ⁻⁴	0.9952
Af-B21	K146M	V112A	3.1 ± 0.080	2.8 x 10 ⁻¹⁶ ± 7.1 x 10 ⁻²¹	1.7 ± 0.070	15 ± 4.2 x 10 ⁻⁴	0.9951

Table 4-8: Summary of kinetic properties of Af-Rubisco variants assayed using purified Rubisco.

Data from this table was calculated from Figure 4-14 and Figure 4-15.

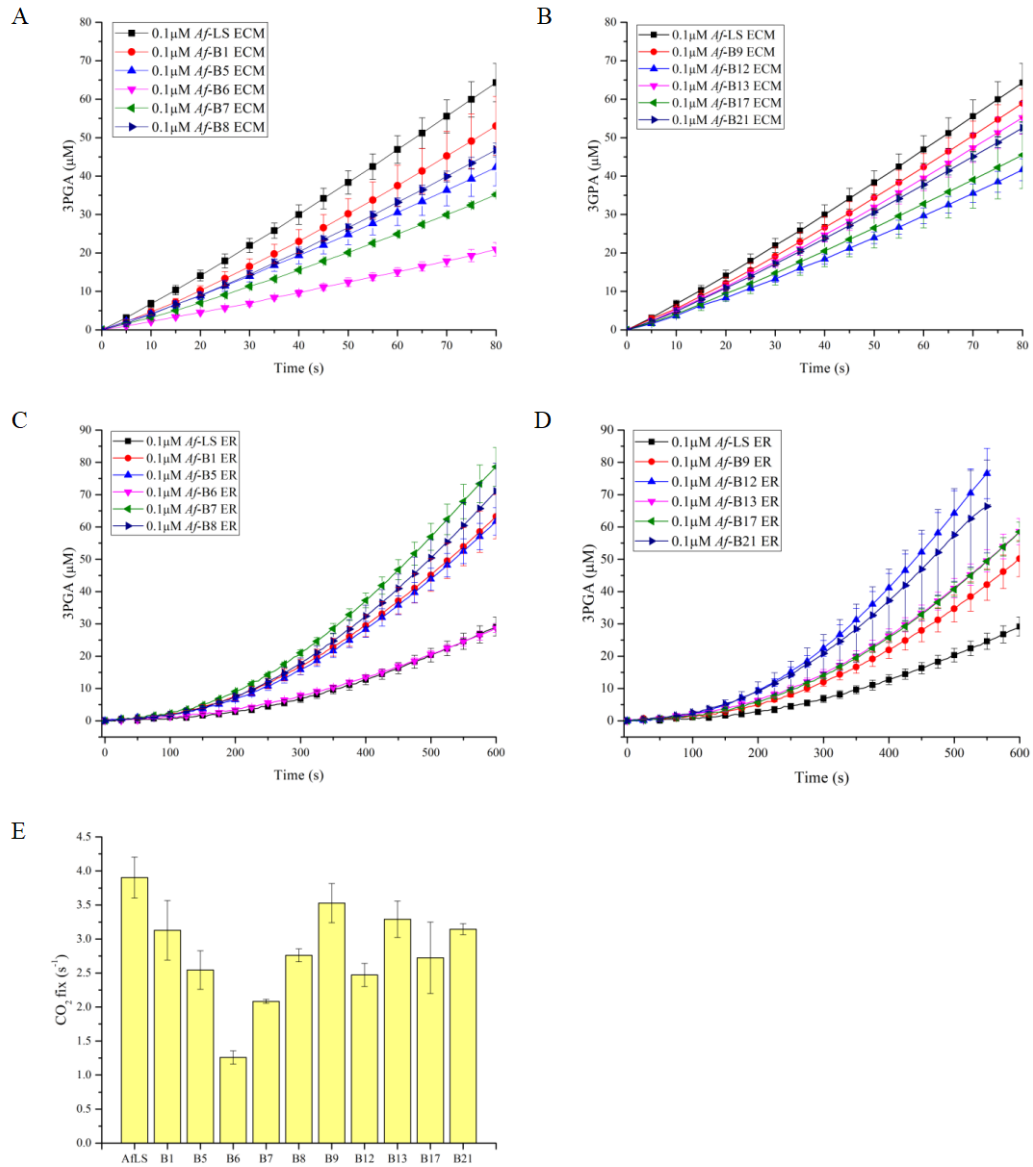


Figure 4-14: Rubisco activity ECM complex and time course of spontaneous activation of ER complex of *Af*-Rubisco variants assayed using purified protein.

(A, B) ECM activity of 0.1 μM of *Af*-Rubisco variants. (C, D) ER activity of 0.1 μM of *Af*-Rubisco variants. (E) Carboxylation rate of the *Af*-Rubisco variant. The assay was carried out in 20 mM NaHCO₃ at 25°C. Error bars indicate the mean and s.d. of at least three independent experiments.

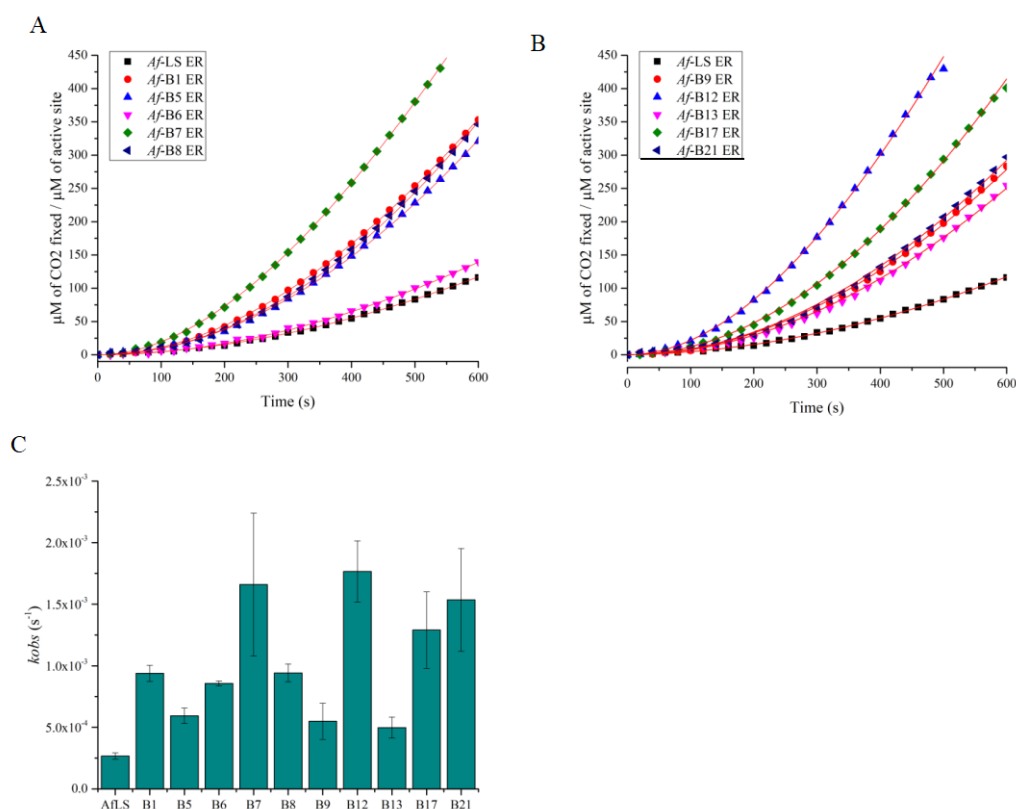


Figure 4-15: Rate of inhibitor release for *Af*-Rubisco variants.

(A), (B) The rate of CO₂ fixed per μM of active site data was fitted to Equation 2.4 (red line). (C) The rate of inhibitor release for *Af*-Rubisco variants. Error bars indicate the mean and s.d. of at least two sets of data fitting.

4.5 Summary of characterizing Rubisco variants using spectrophotometric assay

In this chapter, the Rubisco variants obtained from directed evolution of *R. sphaeroides* and *A. ferrooxidans* Rubisco using RDE and RADE screenings were characterized. The initial screening of variants using soluble *E. coli* extracts to perform spectrophotometric assay provided insights and helped to identify variants of interest to be purified. The majority of Rubisco variants from RDE screening of *R. sphaeroides* and *A. ferrooxidans* form ER complex with increased in k_{obs} and/or increase in v_f as well as a trade-off in the carboxylation rate. This was mostly consistent with the preliminary studies using soluble extract. The high k_{obs} meant that inhibitors are rapidly dissociated from the active site which allows Rubisco to become activated for RuBP carboxylation. The high v_f suggests that the equilibrium favors ECM complexes and free Rubisco could form activated state to carry out RuBP carboxylation. However, there were variants with high

k_{obs} but their v_f was low. It is likely that the equilibrium favors ER complexes over ECM in these variants.

The characterization of the rate of inhibitor release showed that Rubisco can be easily evolved away from activase dependency as many of the variants characterized only had single mutation. In contrast, in the presence of activase, the selected Rubisco variants were likely to form the inhibited complex as there was no selection pressure against this property. This was shown using *Rs*-S1 variant from RADE screening where it had lower k_{obs} than wild type. Therefore, the presence of activase may restrict the evolutionary space of Rubisco (possibly analogous to that observed for other chaperones (Duraio et al., 2015)) and deter Rubisco from acquiring mutations that increase its k_{obs} . Possible reasons on why Rubisco had evolved to require an activase and maintain the relationship with its activase will be discussed in detail in chapter 5. All the purified Rubisco variants will also be further characterized in detail on their kinetic properties using radioisotopic Rubisco assay and structural analysis of these substitutions will also be described in chapter 5.

In summary, the findings in this chapter reveal that single mutations were sufficient to allow Rubisco variants to gain new phenotype of an increase rate of inhibitor release. This unprecedented phenotype free *Rs*- and *Af*-Rubisco from activase dependency. These Rubisco variants were obtained from RDE screening where Rubiscos were evolved in the absence of activase. In contrast, the newly established RADE system allows the directed evolution of Rubisco and activase simultaneously. The RADE system has potential to create novel compatibilities between different Rubiscos and activases which allow beneficial characteristics of activase to be harnessed.

5 DIFFERENT MOLECULAR ADAPTION OF RED- AND GREEN- TYPE RUBISCO

The previous chapter described an initial characterization of *R. sphaeroides* and *A. ferrooxidans* Rubisco variants with increased fitness in the *E. coli* evolution system. Rubisco variants of interest were purified and subsequently were assayed spectrophotometrically for the activity of activated and RuBP-inhibited (ER) complexes. It was observed that there were changes in the stability of these ER complexes and this was often accompanied by a lower carboxylation rate. The work in this chapter will focus on further characterization of these Rubisco variants using radiolabeled ^{14}C -CO₂ incorporation assays and 2'-carboxyarabinitol-1, 5-bisphosphate (^{14}C CABP) active site determinations to more accurately quantify the kinetic parameters of these Rubisco such as the carboxylation rate (k_{cat}), Rubisco's affinity for its substrates, CO₂ and RuBP, and its ability to discriminate between CO₂ and O₂. In addition, structural analysis of the mutated residues will provide insights into how the different substitutions may affect both catalysis and stability of the inhibited complexes. Experiments described in this chapter were performed under the guidance of Associate Professor Spencer Whitney at the Australia National University (Australia) and Ms. Sally Buck assisted with data collection.

5.1 Soluble expression levels of selected variants are variable

In previous RDE study of *Synechococcus* PCC 6301 Rubisco (*Syn*-Rubisco), *Syn*-Rubisco variants selected showed increase Rubisco expression of 5- to 17-fold in XL1-Blue cells (Duraó et al., 2015; Mueller-Cajar et al., 2008b). These *Syn*-Rubisco variants were able to confer fitness by dramatically increasing the amount of Rubisco which in turn increased the rate of RuBP turnover. Therefore, to investigate if the fitness of *Rs*-Rubisco and *Af*-Rubisco variants were contributed by increased solubility, the expression of these Rubiscos in *E. coli* XL1-Blue was quantified and examined by ^{14}C CABP binding and SDS-PAGE. The expression level of both wild type *Rs*-Rubisco and *Af*-Rubisco were calculated to be approximately 5.7% and 15.3% of the soluble *E. coli* extract respectively (Table 5-1, Table 5-2, Figure 5-1).

The majority of *Rs*-Rubisco variants selected from RDE screening exhibited increased levels of expression compared to the wild type, with the exception of *Rs*-R7 which had

similar levels as the wild type (Figure 5-1A, 5-2). The percentage increase in expression level compared to wild type ranged between 14% to 46% but was not statistically significant in most cases. Prominently, *Rs*-R16 and *Rs*-R17 had a significant increase in expression level of approximately 40% compared to wild type. This suggested that the increase in solubility may confer fitness to these variants by either increasing RuBP turnover rate or compensate for the trade-off in kinetics which was measured in Chapter 4.

In contrast to RDE selected *Rs*-Rubisco variants, *Rs*-S1 from RADE screening had a significantly reduced Rubisco expression level of ~3.3% of the total soluble protein, which was almost half the amount compared to the wild type *Rs*-LS (Figure 5-1A, Figure 5-2). This was observed by the lower amount of Rubisco protein in both Native-PAGE and SDS-PAGE of the soluble extract (Figure 5-1A, C-S1). The decrease in Rubisco content may be selected to reduce the metabolic burden of the MM1 *E. coli* cells. In RADE screening, there were three proteins expressed; PRK, Rubisco, and activase. Although there was only basal expression of Rubisco and activase, PRK's expression was highly induced by 0.15% (w/v) of arabinose which serves as a mean to elevate RuBP content and stringency in RADE screening. *Rs*-Rubisco will be inhibited by RuBP and activase will constantly remodel these inhibited Rubisco complexes via ATP hydrolysis to maintain the activation state of Rubisco (Gibson et al., 1979; Mueller-Cajar et al., 2011). The combination of higher expression of PRK and usage of ATP by activase contributed toward increasing metabolic burden to the cells. Thus, the production of less Rubisco could be a solution to counter the metabolic burden and as such a fitter *Rs*-Rubisco variant from RADE screening could be one with reduced expression.

Notably, the second RDE screening involving *Af*-Rubisco, had the majority of *Af*-Rubisco variants exhibit a similar level of expression compared to wild type (Figure 5-1B, Figure 5-3). *Af*-B7, *Af*-B17, and *Af*-B13 were exceptions, with ~29% increased, ~25% increased and 63% decreased expression levels measured respectively. The increase in Rubisco content for *Af*-B7 and *Af*-B17 may increase RuBP turnover rate or compensate kinetic trade-offs similar to that of the *Rs*-Rubisco variant. The reduction in *Af*-B13's expression was significant and *Af*-B13's Rubisco content accounted for 5.7% of total soluble protein which was similar to wild type *Rs*-Rubisco. This reduction in *Af*-

B13 may confer fitness via reducing metabolic burden similar to that of *Rs*-S1. In comparison to the majority of the *Rs*-Rubisco variants, the majority of *Af*-Rubisco variants did not increase their Rubisco content. This may be attributed to the three-fold differences in the expression level between wild type *Rs*-Rubisco (~5.7%) and *Af*-Rubisco (~15.3%).

Although *Rs*-Rubisco and *Af*-Rubisco variants had an increase in total Rubisco content, this increase was marginal compared to the multiple fold changes observed earlier in *Syn*-Rubisco variants. The probable cause could be credited to the differences in the expression of the wild type Rubisco for these three species which stems from the different requirement for Rubisco assembly. Unlike *Rs*-Rubisco and *Af*-Rubisco, *Syn*-Rubisco's biogenesis involves the assembly chaperone, RbcX, to assemble into L₈S₈ holoenzyme (Emlyn-Jones et al., 2006; Saschenbrecker et al., 2007). The expression of wild-type *Syn*-Rubisco alone in *E. coli* had low yields with the Rubisco content only accounting for ~2% of the total soluble protein (Emlyn-Jones et al., 2006). Therefore, improvements contributing to increasing *Syn*-Rubisco content was selected for. However, *Rs*-Rubisco and *Af*-Rubisco were observed to possess high solubility when expressed in *E. coli*, therefore similar improvements in soluble expression are probably not feasible.

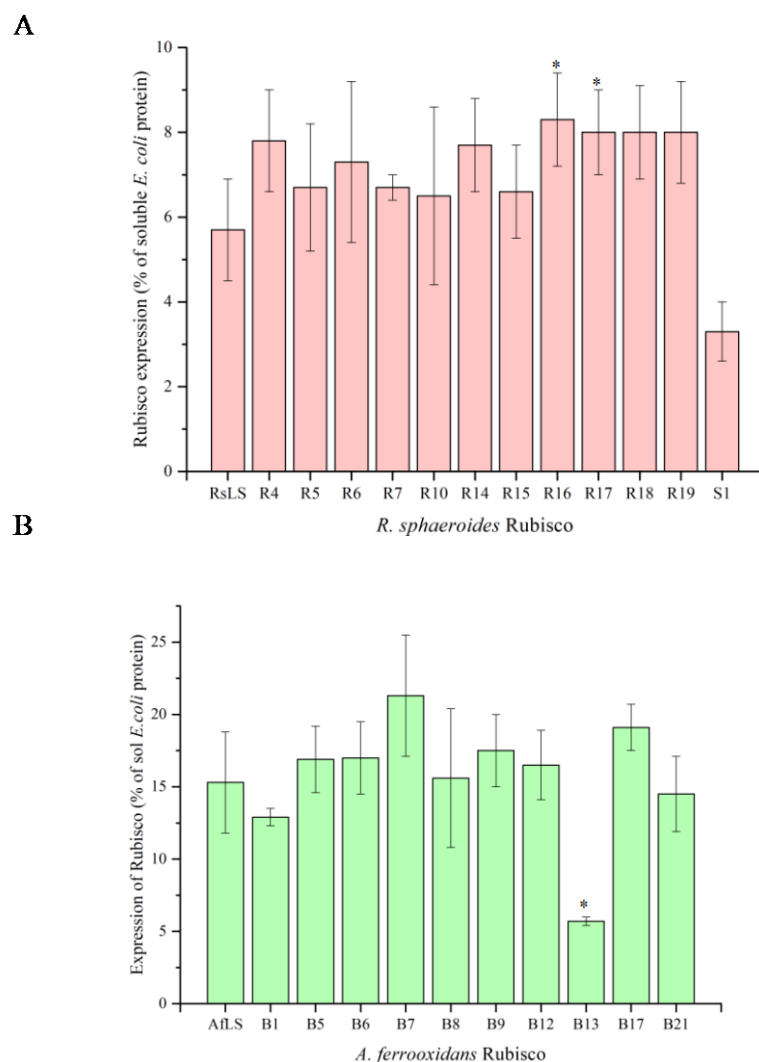


Figure 5-1: The expression level of *Rs*-Rubisco and *Af*-Rubisco as a percentage of soluble *E. coli* protein using *E. coli* XL1-Blue cells.

(A, B) The expression level of soluble Rubisco as a percentage of total soluble *E. coli* protein for *Rs*-Rubisco and *Af*-Rubisco variants. The expression of the Rubisco variants was quantified by ^{14} CABP binding. The data represent mean $\pm$ standard deviation (s.d.) from at least three independent experiments and statistical significance of variation relative to their respectively wild type Rubisco was determined by T-test (*= $p < 0.05$; ** = $p < 0.01$). Bar graphs corresponding to *Rs*-Rubisco and *Af*-Rubisco were colored coral and light green respectively and this color scheme will be used throughout this chapter.

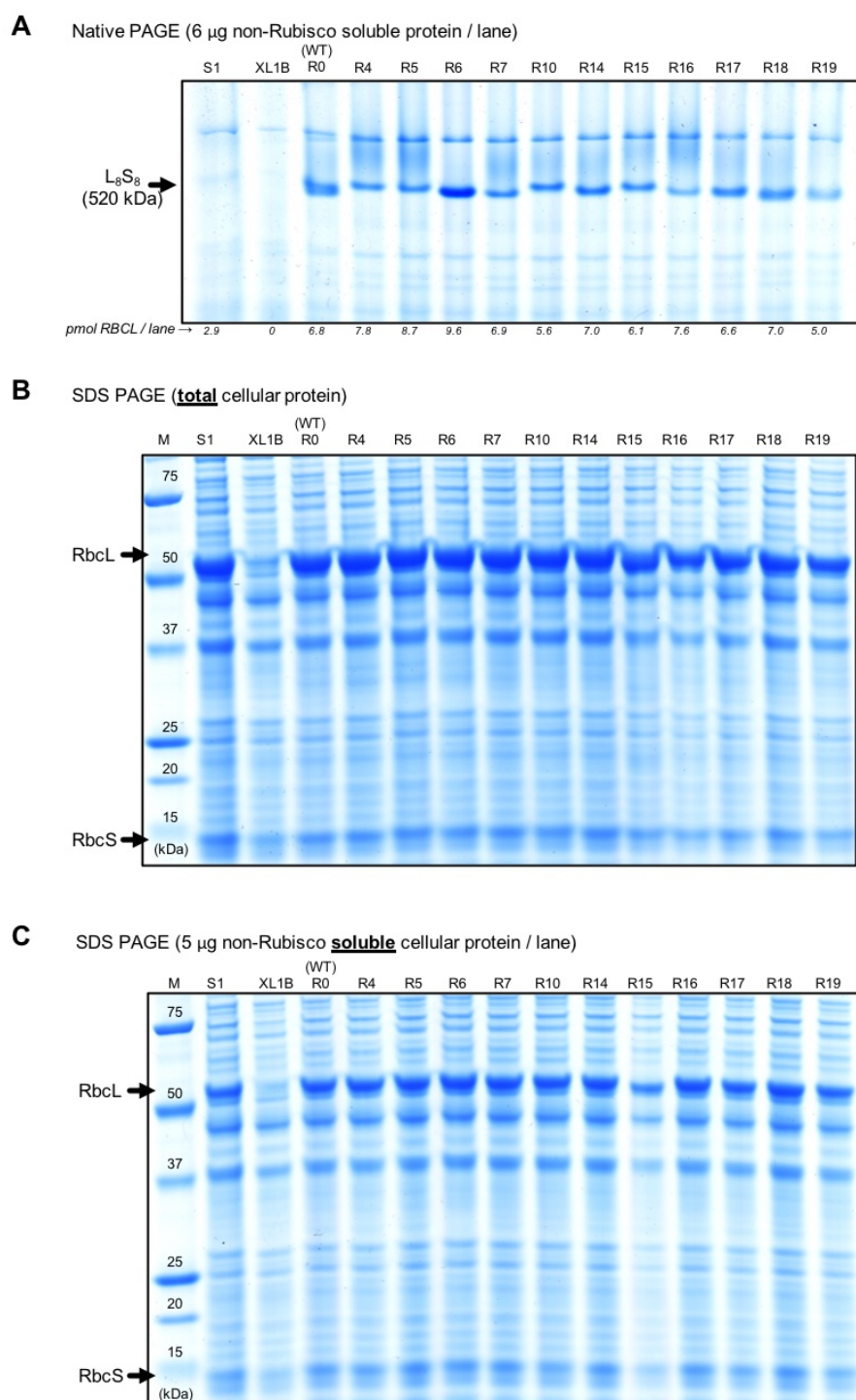


Figure 5-2: Representative gel of *Rs*-Rubisco variants' expression in *E. coli* XL1-Blue

(A) 4-12% Native-PAGE gel of 6 μ g of non-Rubisco soluble protein loaded per lane and pmol of Rubisco in each lane were indicated. (B) 4-12% SDS-PAGE of total cellular protein in *E. coli* extract. (C) 4-12% SDS-PAGE of 5 μ g of non-Rubisco soluble cellular protein per lane. The expression of *Rs*-Rubisco variants was induced with 1mM IPTG at 28 $^{\circ}$ C for 6 hours and the Rubisco content was quantified by ^{14} CABP binding. R0: Wild type *Rs*-LS Rubisco, R'X': *Rs*-Rubisco variants (X is an integer), XLB: XL1-Blue.

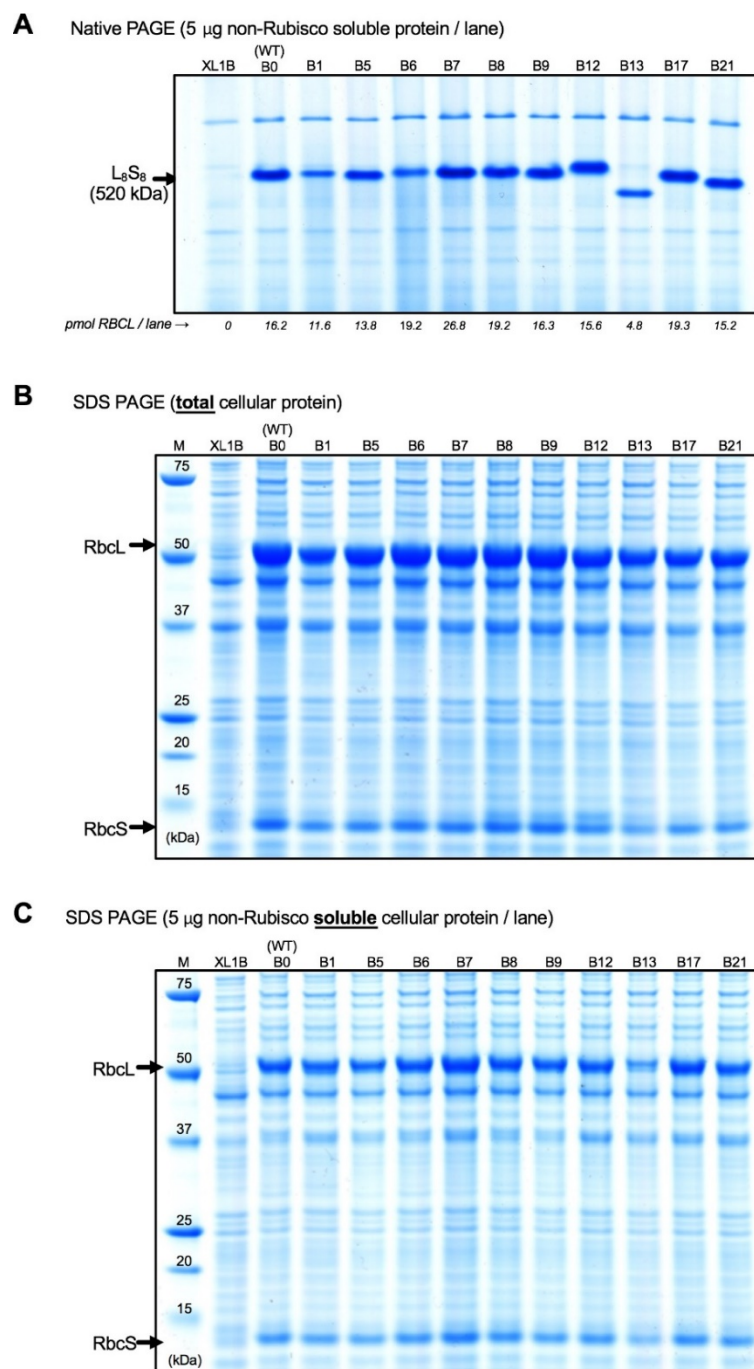


Figure 5-3: Representative gel of *Af*-Rubisco variants' expression in *E. coli* XL1-Blue.

(A) 4-12% Native-PAGE gel of 6 μ g of non-Rubisco soluble protein loaded per lane and pmol of Rubisco in each lane were indicated. (B) 4-12% SDS-PAGE of total cellular protein in extract. (C) 4-12% SDS-PAGE of 5 μ g of non-Rubisco soluble cellular protein per lane. The expression of *Af*-Rubisco variants was induced with 1mM IPTG at 28 °C for 6 hours and the Rubisco content was quantified by ¹⁴CABP binding. B0: wild type *Af*-LS Rubisco, B'X' (X is an integer): *Af*-Rubisco variants, XLB: XL1-Blue.

5.2 *Rs*-Rubisco and *Af*-Rubisco variants showed diversity in their kinetic properties

Rs-Rubisco and *Af*-Rubisco variants were assayed on the various kinetic properties using radiolabeled substrates and the results were summarized in Table 5-1 and Table 5-2 respectively. The pTrcHisB plasmids encoding for the various Rubisco variants and wild type Rubisco were transformed into *E. coli* strain XL-1Blue for expression of Rubisco. Soluble *E. coli* extract were used in all the kinetic measurements, except CO₂/O₂ specificity where purified proteins were required. The amount of Rubisco active site in the soluble *E. coli* extract and pure proteins were quantified using ¹⁴CABP binding as CABP is a transition state analogue that binds almost irreversibly to carbamylated Rubisco's active site. Therefore, the stoichiometric binding of ¹⁴CABP allows for quantification of Rubisco's content.

The Michaelis constant for CO₂ (K_C) of these Rubisco variants was measured using ¹⁴CO₂ incorporation assay with varying concentration of bicarbonate (equivalent to 0 to 190 μ M of CO₂). To exclude CO₂ from the atmosphere, all the reagents were bubbled with nitrogen and the assays were carried out in septum capped reaction vials. Activated Rubisco complexes (ECM) was prepared by incubating the soluble extract with 20 mM MgCl₂ and 40 mM bicarbonate for ten minutes. The assay was initiated by addition of ECM to vials containing assay components and the reaction was stopped after 2 minutes by the addition of 20% formic acid and then dried down at 90 °C. Next, water and scintillant fluid were mixed with the dried residue and the radioactivity was measured using a scintillation counter. The kinetic parameters were obtained by fitting the data into the Michaelis-Menten equation (Equation 2.6).

$$V_c = \frac{V_c^{max} C}{C + K_C} \quad \text{(Equation 2. 6)}$$

where V_c is the carboxylation rate, V_c^{max} is the maximal rate of carboxylation and C is the CO₂ concentration and K_C is the apparent Michaelis-Menten constant (K_m) for CO₂ under atmospheric levels of O₂ ($K_c^{21\%O_2}$). The carboxylation turnover rate (k_{cat}^c) of Rubisco was calculated by dividing V_c^{max} by the number of active sites in the soluble *E. coli* extract which was quantified by ¹⁴CABP binding.

Rubisco	Expression (% (w/v) soluble <i>E. coli</i> protein)		k_{cat}^c (s ⁻¹)	$K_c^{21\% O_2}$ (μM)	$k_{cat}^c / K_c^{21\% O_2}$ (mM ⁻¹ .s ⁻¹)	$S_{C/O}$ (mol.mol ⁻¹)	$Km(RuBP)$ (μM)	
	Amino acid change(s)							
	RbcL	RbcS						
<i>Rs</i> -LS (wt)	nil	nil	5.7 ± 1.2	5.2 ± 0.2	73.4 ± 3.7	71.1 ± 2.1	61.0 ± 3.4	13.9 ± 2.4
<i>Rs</i> -R4	T73A	nil	7.8 ± 1.2	4.8 ± 0.3	98.8 ± 8.7*	48.3 ± 2.7	59.3 ± 1.2	9.6 ± 0.8
<i>Rs</i> -R5	T73S	nil	6.7 ± 1.5	4.3 ± 0.4*	62.3 ± 4.9*	68.8 ± 4.1	61.1 ± 2.4	19.6 ± 1.7
<i>Rs</i> -R6	R83C	nil	7.3 ± 1.9	5.5 ± 0.1	99.9 ± 9.8*	55.0 ± 5.2*	55.5 ± 0.7*	13.6 ± 0.9
<i>Rs</i> -R7	R83S	K73R	6.7 ± 0.3	5.4 ± 0.3	100.1 ± 4.8**	53.7 ± 2.7**	48.7 ± 0.4**	11.0 ± 0.6
<i>Rs</i> -R10	nil	F91I, T95S, S108C	6.5 ± 2.1	4.6 ± 0.2*	102.1 ± 5.7**	44.6 ± 1.9**	59.8 ± 2.2	17.2 ± 1.2
<i>Rs</i> -R14	Y26F	nil	7.7 ± 1.1	5.6 ± 0.3	86.9 ± 3.3*	64.1 ± 3.0	59.5 ± 2.1	9.8 ± 0.4
<i>Rs</i> -R15	P52L	nil	6.6 ± 1.1	5.1 ± 0.1	76.5 ± 2.5	66.5 ± 3.3	58.2 ± 2.1	21.4 ± 1.6
<i>Rs</i> -R16	T77A	nil	8.3 ± 1.1*	5.5 ± 0.1	138.0 ± 13.9*	39.9 ± 3.2**	60.2 ± 0.8	8.0 ± 1.1
<i>Rs</i> -R17	N228S	nil	8.0 ± 1.0*	4.6 ± 1.0	86.5 ± 22.5	57.5 ± 23.7	64.1 ± 3.9	10.2 ± 0.8
<i>Rs</i> -R18	L107M	nil	8.0 ± 1.1	5.5 ± 0.0	80.5 ± 1.2*	67.9 ± 0.8	63.1 ± 1.9	11.5 ± 0.6
<i>Rs</i> -R19	L107Q	nil	8.0 ± 1.2	5.3 ± 0.2	82.6 ± 10.0	74.7 ± 10.1	60.1 ± 2.6	12.1 ± 0.4
<i>Rs</i> -S1	I315T	nil	3.3 ± 0.7	4.3 ± 0.8	64.3 ± 2.3*	66.4 ± 13.3	64.0 ± 1.8	10.0 ± 0.7

Table 5-1: Summary of the kinetic properties and expression level of *Rs*-Rubisco variants.

The data for expression of *Rs*-Rubisco as percentage of soluble *E. coli* protein was taken from Figure 5-1. Maximum carboxylation turnover rate (k_{cat}^c) and Michaelis Menten constant for CO₂ ($K_c^{21\% O_2}$ or K_c) under atmospheric level of oxygen were obtained from Figure 5-4. Carboxylation efficiency of the Rubisco variants were obtained by dividing k_{cat}^c with K_c . CO₂/O₂ Specificity ($S_{C/O}$) of *Rs*-Rubisco variants were from Figure 5-5 while Michaelis Menten constant for RuBP ($Km(RuBP)$) was from Figure 5-6. The data represent mean ± s.d. from at least three independent experiments unless stated in their respective figures. The statistical significance of variation relative to their respectively wild type Rubisco was determined by T-test (*= p <0.05; ** = p <0.01).

Rubisco	Amino acid change(s)		Expression (% (w/v) soluble <i>E. coli</i> protein)	k_{cat}^c (s ⁻¹)	$K_C^{21\% O_2}$ (μ M)	$k_{cat}^c / K_C^{21\% O_2}$ (mM ⁻¹ .s ⁻¹)	$S_{C/O}$ (mol.mol ⁻¹)	$Km(RuBP)$ (μ M)
	RbcL	RbcS						
<i>Af</i> -LS (wt)	nil	nil	15.3 ± 3.5	6.8 ± 0.1	81.4 ± 2.8	84.1 ± 4.1	53.5 ± 1.5	7.4 ± 0.6
<i>Af</i> -B1	K41R, I105V	nil	12.9 ± 0.6	5.2 ± 0.3**	83.4 ± 0.7	62.2 ± 3.3**	51.1 ± 3.0	7.9 ± 0.5
<i>Af</i> -B5	S61T	nil	16.9 ± 2.3	5.0 ± 0.3**	93.3 ± 1.6**	53.5 ± 3.7**	51.8 ± 6.82	6.7 ± 0.7
<i>Af</i> -B6	F220L	nil	17.0 ± 2.5	5.5 ± 0.2**	93.1 ± 5.5*	58.6 ± 1.3**	53.0 ± 2.6	9.3 ± 1.2
<i>Af</i> -B7	P305L	nil	21.3 ± 4.2	3.4 ± 0.2**	88.0 ± 3.5	39.2 ± 3.7**	49.2 ± 2.1*	11.1 ± 0.8
<i>Af</i> -B8	P305S	nil	15.6 ± 4.8	4.7 ± 0.8*	90.6 ± 2.6*	51.5 ± 10.1*	51.3 ± 2.6	10.2 ± 1.3
<i>Af</i> -B9	T392A	nil	17.5 ± 2.5	6.3 ± 0.2*	88.3 ± 5.8	71.9 ± 4.5*	51.6 ± 1.6	8.1 ± 0.6
<i>Af</i> -B12	E438G	nil	16.5 ± 2.4	4.2 ± 0.2**	103.1 ± 6.6*	40.7 ± 2.4**	50.8 ± 3.2	9.1 ± 0.6
<i>Af</i> -B13	I42T, K463E	nil	5.7 ± 0.3*	5.7 ± 0.2**	79.3 ± 1.2	71.6 ± 2.1*	52.9 ± 1.7	20 ± 0.9
<i>Af</i> -B17	I465N	nil	19.1 ± 1.6	4.4 ± 0.1**	70.5 ± 3.1*	62.6 ± 2.1**	51.3 ± 2.3	13.8 ± 1
<i>Af</i> -B21	K146M	V112A	14.5 ± 2.6	6.80 ± 0.0	97.1 ± 5.7*	70.4 ± 3.8*	51.2 ± 2.9	7.7 ± 0.4

Table 5-2: Summary of the kinetic properties and expression level of *Af*-Rubisco variant.

The data for expression of *Af*-Rubisco variants as percentage of soluble *E. coli* protein was taken from Figure 5-1. Maximum carboxylation turnover rate (k_{cat}^c) and Michaelis constant for CO₂ ($K_C^{21\% O_2}$ or K_C) under atmospheric level of oxygen were obtained from Figure 5-4. Carboxylation efficiency of the Rubisco variants were obtained by dividing k_{cat}^c with K_C . CO₂/O₂ Specificity ($S_{C/O}$) of *Af*-Rubisco variants were from Figure 5-5 while Michaelis constant for RuBP ($Km(RuBP)$) was from Figure 5-6. The data represent mean ± s.d. from at least three independent experiments unless stated in their respective figures. The statistical significance of variation relative to their respectively wild type Rubisco was determined by T-test (*= p <0.05; ** = p <0.01).

5.2.1 Rubisco variants had diverse carboxylation turnover rate

The maximum carboxylation turnover rate (k_{cat}^c) for wild type *Rs*-Rubisco and *Af*-Rubisco was calculated to be 5.2 s^{-1} and 6.8 s^{-1} (Table 5-1, Table 5-2, Figure 5-4A, B). *Rs*-Rubisco variants showed diversity in their k_{cat}^c with half of the variants exhibiting either similar or slight increases in k_{cat}^c while the other half of the variants exhibit reduction in k_{cat}^c compared to the wild type (Figure 5-4). *Rs*-R15, *Rs*-R18, and *Rs*-R19 maintained similar k_{cat}^c as wild type while *Rs*-R6, *Rs*-R7, *Rs*-14, and *Rs*-R16 had a moderate increase of $\sim 6\%$ in k_{cat}^c . In contrast, *Rs*-R4, *Rs*-R5, *Rs*-R10, *Rs*-R17, and *Rs*-S1 had a reduction in k_{cat}^c and this reduction was significant in *Rs*-R5 and *Rs*-R17. The expression levels of these variants with lower k_{cat}^c were increased as compared to the wild type. This was observed for these variants except *Rs*-S1 (Table 5-1). This suggests that the increased expression level could be necessary to compensate for the reduction in k_{cat}^c .

The majority of *Af*-Rubisco variants, except *Af*-B21, had reduced k_{cat}^c values ranging between $\sim 8\%$ to $\sim 50\%$ as compared to the wild type (Figure 5-4, Table 5-2). The reductions in k_{cat}^c were significant for all these variants. In particular, *Af*-B7 had the greatest decrease of $\sim 50\%$. In contrast to *Rs*-Rubisco variants with decreased k_{cat}^c , the majority of these *Af*-Rubisco variants did not alter the Rubisco expression levels to compensate for the reductions in k_{cat}^c . This was probably due to *Af*-Rubisco variants possessing higher Rubisco content as compared to *Rs*-Rubisco variants. Therefore, metabolic management of the RuBP turnover rate was possible despite having lower carboxylation rate. *Af*-B21 was the only *Af*-Rubisco variant that maintained a similar k_{cat}^c as the wild type and it had a slight decrease in the protein expression. Overall, the lower k_{cat}^c of the *Af*-Rubisco and *Rs*-Rubisco variants meant that the improvement in fitness was not contributed by improving carboxylation rates. In addition, the trade-off in k_{cat}^c was not observed in all the evolved variants that had high k_{obs} and this was consistent with the results presented in chapter 4 (Table 5.1).

Figure 5-4: The carboxylation turnover rate (k_{cat}^c) and Michaelis constant for CO₂ (K_C) of *Rs*-Rubisco and *Af*-Rubisco variants.

(A, B) Representative plots of carboxylation rate against CO₂ concentration for wild type *Rs*-Rubisco and *Af*-Rubisco. The Michaelis constant for CO₂ ($K_C^{21\%O_2}$, K_C) and maximum carboxylation turnover rate (k_{cat}^c) were determined by fitting data to Michaelis Menten equation (Equation 2.6). (C, D) The k_{cat}^c rate of *Rs*-Rubisco variants and *Af*-Rubisco variants. (E, F) K_C of *Rs*-Rubisco and *Af*-Rubisco variants. The data represent mean $\pm$ s.d. from three independent experiments except for *Rs*-R4 and *Rs*-18 which were from two independent experiments. The s.d. for k_{cat}^c of *Rs*-R18 and

Af-B21 were less than 0.0, therefore it could not be visualized in the chart. Statistical differences between variants and their respective wild type Rubisco were determined by T-test and the asterisk denoted differences were statistically significant (* = $P < 0.05$; ** = $P < 0.01$).

5.2.2 Majority of Rubisco variants had reduced affinity towards CO₂

The Michaelis constant for CO₂ ($K_C^{21\%O_2}$, K_C) for wild type *Rs*-Rubisco and *Af*-Rubisco were 73.4 μ M and 81.4 μ M respectively (Figure 5-4C, D). K_C served as an indication of the affinity of CO₂ as a substrate for the enediol form of RuBP as CO₂ and O₂ does not bind directly to the active site Rubisco (Pierce et al., 1986; Tcherkez et al., 2006). The majority of the *Rs*-Rubisco variants had an increase in K_C which meant that these variants had a decreased affinity for the substrate, CO₂ (Figure 5-4C). *Rs*-R5 and *Rs*-S1 were exceptions with significant decreases in K_C which translates to them having an increased affinity towards CO₂. *Rs*-R15 was the only variant to maintain similar K_C as the wild type. For *Rs*-S1, the increase in affinity towards CO₂ (lower K_C) could be beneficial as the screening for RADE was performed in air as compared to 1% CO₂ used for the RDE screening. It was often observed by comparing Rubiscos derived from different species that the improvement in CO₂ fixation will come at expense of affinity for CO₂ (Tcherkez et al., 2006). This was consistent with some of the *Rs*-Rubisco variants, *Rs*-R6, *Rs*-R14, *Rs*-R16, where the improvement in k_{cat}^c was accompanied with a reduction in affinity towards CO₂.

A similar trend of increased in K_C (decreased affinity for CO₂) was also observed in the majority of *Af*-Rubisco variants (Figure 5-4D). The reduction in affinity for CO₂ was significant in *Af*-B5, *Af*-B6, *Af*-B8, *Af*-B12 and *Af*-B21. Whereas, *Af*-B1 and *Af*-B13 maintained similar K_C as wild type. *Af*-B17 was the only variant that had a significantly increased affinity towards CO₂ with a reduction of 14% in K_C compared to wild type. Overall, the decreased affinity for CO₂ was probably tolerated in *Rs*-Rubisco and *Af*-Rubisco variants obtained from RDE screening as they were screened under elevated CO₂ condition.

5.2.3 Most selected Rubisco variants exhibited a reduced carboxylation efficiency

One of the key indicators for the efficiency of Rubisco in catalyzing the carboxylation reaction is via the carboxylation efficiency. The carboxylation efficiency (k_{cat}^c / K_C) of Rubisco was determined using the ratio of the substrate saturating catalytic rate (k_{cat}^c) and to the K_m (K_C) of its substrate, CO_2 . Wild type *Rs*-Rubisco had carboxylation efficiency of $71.1 \text{ mM}^{-1}\text{s}^{-1}$ while wild type *Af*-Rubisco had $84.1 \text{ mM}^{-1}\text{s}^{-1}$. This suggests that *Af*-Rubisco was more efficient in performing carboxylation reaction compared to *Rs*-Rubisco (Table 5-1, 5-2). Majority of *Rs*-Rubisco and *Af*-Rubisco variants, except *Rs*-R19, had decreased carboxylation efficiency. *Rs*-R19 had a slight increase of 5% in the carboxylation efficiency as its k_{cat}^c was similar to wild type and K_C modestly reduced. This decrease in carboxylation efficiency was significant in all *Af*-Rubisco and some *Rs*-Rubisco variants (*Rs*-R6, *Rs*-R7, *Rs*-R10 and *Rs*-R16). Thus, suggesting that these evolved Rubisco variants are inferior carboxylases compared to the wild type.

5.2.4 The CO_2/O_2 specificity of *Rs*-Rubisco and *Af*-Rubisco variants were similar to wild type

The CO_2/O_2 specificity factor ($S_{C/O}$) is a diagnostic parameter that had been widely used to determine the overall efficiency of Rubisco. $S_{C/O}$ describes the ability of Rubisco to discriminate between CO_2 and O_2 and its value will thus determine the extent of the metabolically wasteful oxygenation reaction. The measurement of $S_{C/O}$ was carried out using purified Rubisco and ^3H -labelled RuBP (^3H -RuBP) was used as a substrate. The ratio of ^3H -RuBP incorporation into the glycerate and glycolate peaks was used to calculate $S_{C/O}$ using Equation 2.7 (Kane et al., 1994).

The $S_{C/O}$ of wild type *Rs*-Rubisco and *Af*-Rubisco were determined to be $61.0 \text{ mol.mol}^{-1}$ and $53.5 \text{ mol.mol}^{-1}$ respectively. The majority of both *Rs*-Rubisco and *Af*-Rubisco variants had similar $S_{C/O}$ as their wild type (Figure 5-5 C, D). This suggested that the majority of the evolved variants were able to distinguish between CO_2 and O_2 with similar efficiency as their wild type counterpart. The exceptions to this trend were *Rs*-R6, *Rs*-R7, and *Af*-B7. These three variants had significant reductions in their $S_{C/O}$ with *Rs*-R6 having the greatest

reduction of ~20% as compared to wild type. Therefore, suggesting that these three variants had impaired ability in differentiating between CO₂ and O₂.

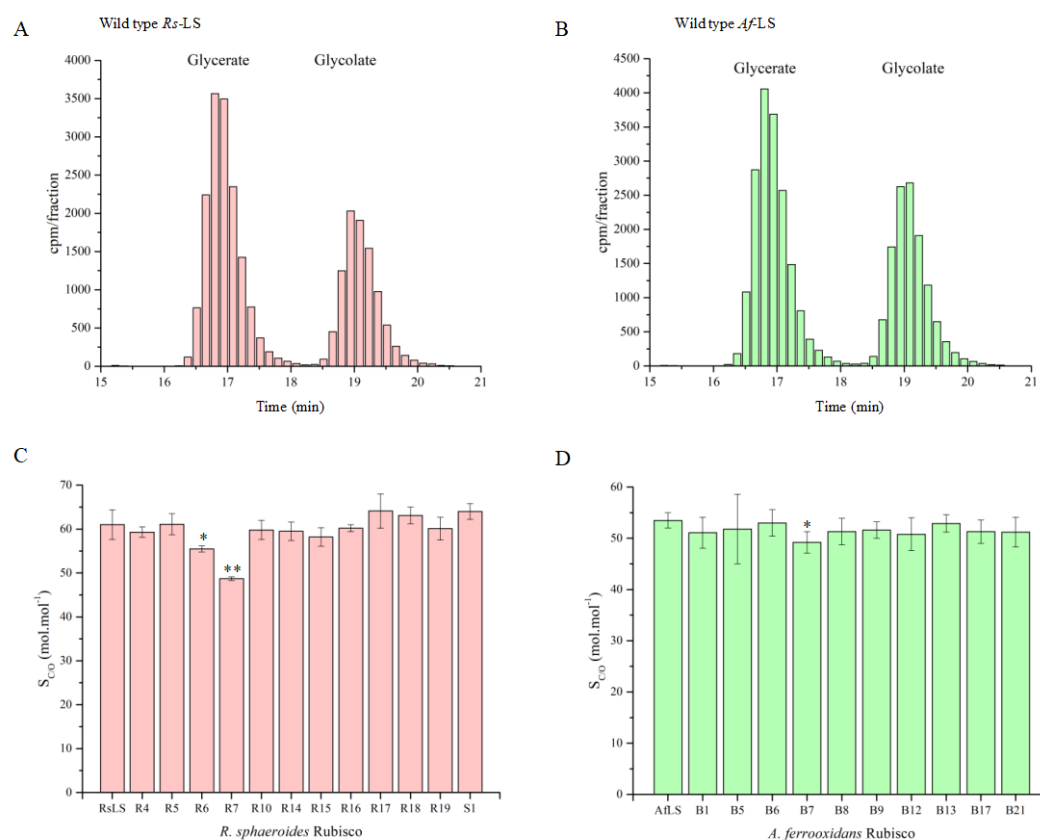


Figure 5-5: The CO₂/O₂ specificity ($S_{C/O}$) of *Rs*-Rubisco and *Af*-Rubisco variants.

(A, B) Representative chromatograms of the ³H incorporation into the glycerate and glycolate peaks following separation from HPLC. (C) $S_{C/O}$ for *Rs*-Rubisco variants. (D) $S_{C/O}$ for *Af*-Rubisco variants. The chromatogram of both wild type Rubisco for *R. sphaeroides* and *A. ferrooxidans* were shown as examples. Recombinantly purified Rubiscos were used in the determination of $S_{C/O}$ and they were purified using anion exchange column followed by size exclusion. ³H-RuBP was used as a substrate for Rubisco using a gas mixture of 0.05% (v/v) CO₂ and 99.6 - 99.95% (v/v) O₂ as described in section 2.6.3.5. The glycerate/glycolate ratio was used to calculate the $S_{C/O}$ of Rubisco using Equation 2.7.

5.2.5 Determination of $K_m(RuBP)$ for variant Rubisco

Measurement for Michaelis constant of RuBP ($K_m(RuBP)$) was performed to determine the changes in RuBP binding behaviors and to ensure that the amount of RuBP used in the other kinetic assays were sufficiently saturating. $K_m(RuBP)$ of wild type *Rs*-Rubisco and *Af*-

Rubisco were 13.9 μM and 7.4 μM respectively. Hence, the amount of RuBP used in other kinetic assays (0.4 mM) and spectrophotometric assay (1 mM) was indeed sufficiently saturating.

The majority of *Rs*-Rubisco variants exhibited reductions in $K_m(\text{RuBP})$ as compared to wild type and this meant that affinity of the activated ECM Rubisco towards RuBP as a substrate was increased. *Rs*-Rubisco variants with an increased affinity towards RuBP were *Rs*-R4, *Rs*-R7, *Rs*-R14, *Rs*-R16, *Rs*-R17, *Rs*-R18, *Rs*-R19, and *Rs*-S1. Among these variants, *Rs*-R16 had the greatest reduction of $\sim 41\%$ in $K_m(\text{RuBP})$. On the other hand, *Rs*-R5, *Rs*-R10, and *Rs*-R15 had increased $K_m(\text{RuBP})$ of 20-50% as compared to the wild type and their affinity towards RuBP was decreased. *Rs*-R6 was the only variant that maintained a similar $K_m(\text{RuBP})$ as the wild type.

In contrast to *Rs*-Rubisco variants, the majority of *Af*-Rubisco variants had increased $K_m(\text{RuBP})$. These *Af*-Rubisco variants were *Af*-B1, *Af*-B6, *Af*-B7, *Af*-B8, *Af*-B9, *Af*-B13, *Af*-B12, *Af*-B17, and they had lowered affinity towards RuBP as a substrate. *Af*-B13 had the highest increased in $K_m(\text{RuBP})$ where the affinity was decreased by ~ 2 fold as compared to the wild type. *Af*-B5 was the only variant with a slight decrease of 9% in its $K_m(\text{RuBP})$ while *Af*-B21 had similar $K_m(\text{RuBP})$ as wild type.

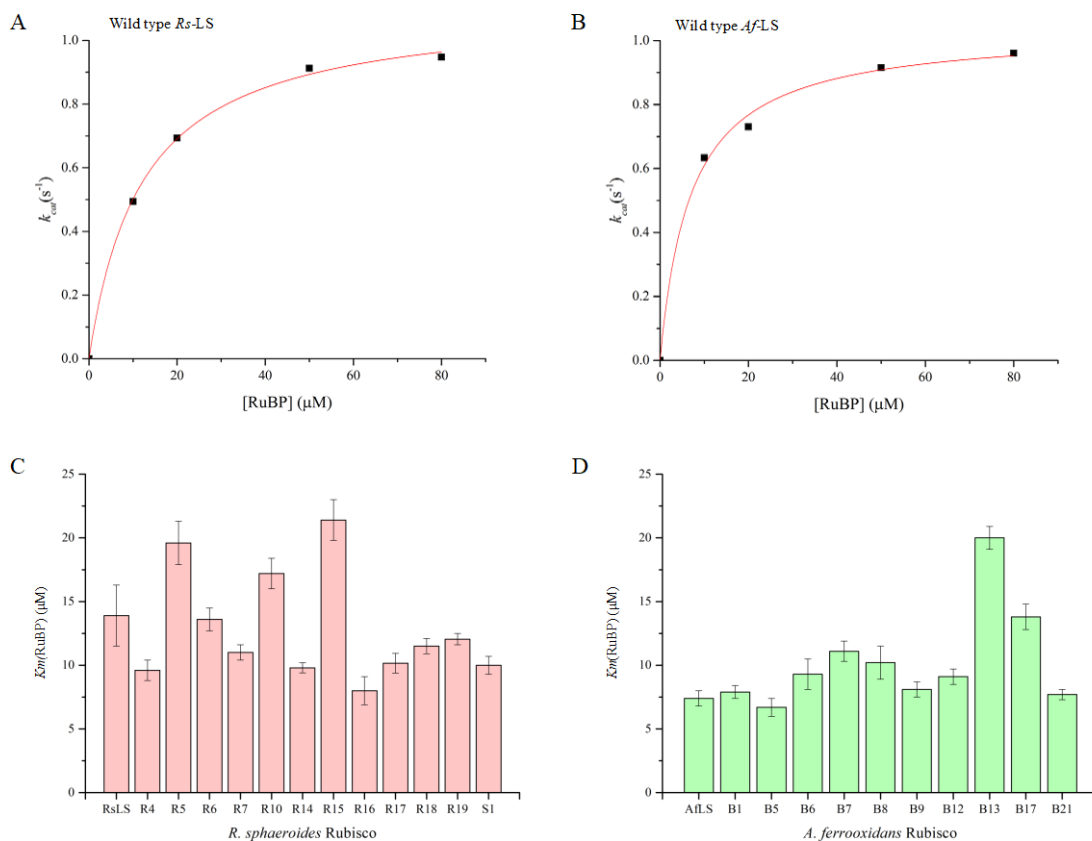


Figure 5-6: The Michaelis constant for RuBP, $K_m(RuBP)$, of *Rs*-Rubisco and *Af*-Rubisco variants.

(A, B) Representative plots of carboxylation rate against RuBP concentration for wild type *Rs*-Rubisco and *Af*-Rubisco. (C, D) $K_m(RuBP)$ of *Rs*-Rubisco variants and *Af*-Rubisco variants. The measurement for $K_m(RuBP)$ was performed using 0 to 80 μM of RuBP as described in section 2.6.3.4 and data obtained was fitted into the Michaelis-Menten equation (Equation 2.6) to obtain the $K_m(RuBP)$. Error bar represents the s.d. of duplicate from a single set of samples.

5.3 Evolved Rubisco variants target different regions of Rubisco

Multiple evolved Rubisco variants released RuBP more rapidly from the ER complex. The residues mutated in variants were mapped onto structural models to investigate the effects of their substitution on the structure. Clusters of selected residues were located in regions of the large subunit; the 60s loop, C-terminal tail (C-terminus), and dimer-dimer interface (Figure 5-7)(Table 5-4). In the small subunit, the selected residues were located near β -strand D (Table 5-4).

Due to a lack of high-resolution crystal structures for both *Rs*-Rubisco and *Af*-Rubisco, the crystal structure of other form I Rubiscos were used as Rubiscos are highly conserved in both sequence identity and overall structure. For *Rs*-Rubisco variants, the crystal structure of red-type Rubisco from *Galderia partia* (1BWV, closed) was used (Sugawara et al., 1999). *Af*-Rubisco variants were mapped onto the structure from *Halothiobacillus neapolitanus* (1SVD, open) and Spinach (8RUC, closed) (Andersson, 1996). The open and closed state of the Rubisco is discussed below at section 5.3.1. The secondary structure was aligned according to Knight et al. (1990). For consistency, the numbering of *Rs*-Rubisco variants (large subunit) will be adjusted according to Spinach Rubisco similar to *Af*-Rubisco variants. For the *Rs*-Rubisco variants, the original numbering will be indicated in the bracket. For small subunits, the numbering will be according to their respective organism. This numbering modification will be applied for the rest of this chapter. By systematically analyzing the location of amino acid substitutions and integrating this information with the current understanding of Rubisco structure and function, we can hypothesize mechanistic explanations for the observed biochemical phenotypes.

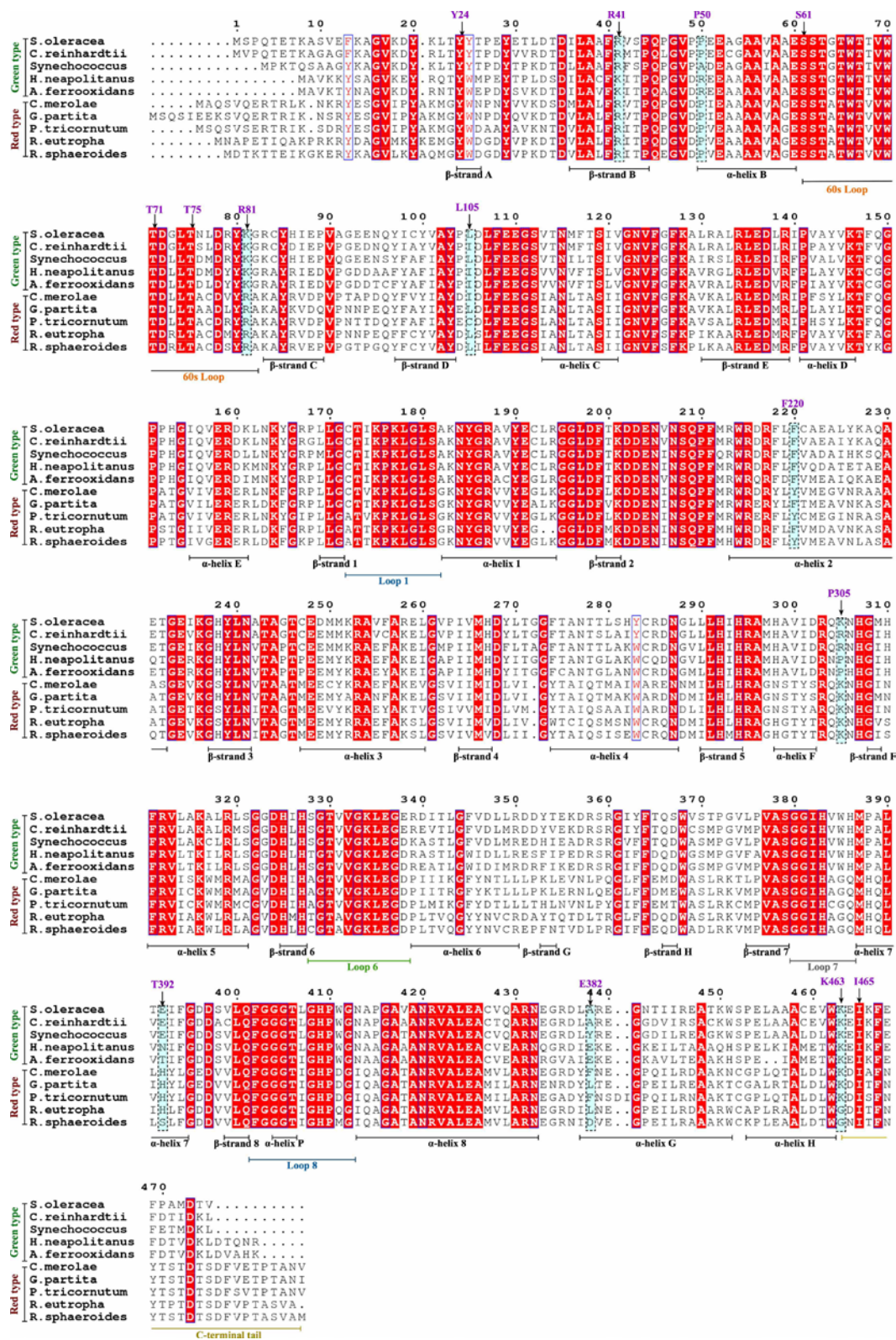


Figure 5-7: Sequence alignments of selected green-type and red-type Rubisco large subunits.

Sequence alignment of Rubisco large subunit of green-type Rubisco; *S. oleracea* (Spinach, P00875), *C. reinhardtii* (P00877), *Synechococcus* PCC6301 (P00880), *A. ferrooxidans* (B7JA24), *H. neapolitanus* (O85040); and red-type Rubisco; *C. merolae* (Q85G81), *G. partita* (O98949), *Phaeodactylum tricornutum* (*P. tricornutum*; E9PAI6), *R. eutropha* (P0C2C2), *R. sphaeroides* (Q3IX55). Residues that are conserved are shown in white fonts with red background while residues with similar properties are shown in red font. Residues selected from screening were indicated with purple labels and residues that are not conserved has a light blue background. The numbering of the residues and secondary structure of α -helices and β -strands were according to Spinach Rubisco as denoted by Knight et al. (1990). The 60s loop in the N-terminal domain, loop 6 and C-terminal tail are indicated in orange, light green and yellow respectively. Loop 1 and 8 of the C-terminal domain that interacts extensively with the 60s loop were also labeled in dark blue. Loop 7 which contributes two residues to the P1 phosphate site was labeled in grey. The sequence alignments were performed using MultAlin (<http://multalin.toulouse.inra.fr/multalin/>) and alignment image was constructed with ESPript (Robert et al., 2014).

Interacting region	Rubisco variant with increase k_{obs}	Substitution		Repeatedly mutated residue
		RbcL	RbcS	
60s loop of large subunit	<i>Rs</i> -R15	P50L	nil	P50
	<i>Af</i> -B5	S61T	nil	S61
	<i>Rs</i> -R14	Y24F	nil	Y24
	<i>Rs</i> -R4	T71A	nil	T71
	<i>Rs</i> -R5	T71S	nil	
	<i>Rs</i> -R16	T75A	nil	T75
	<i>Rs</i> -R6	R81C	nil	R81
	<i>Rs</i> -R7	<u>R81S</u>	K73R	
C-terminus of large subunit	<i>Af</i> -B13	I42T, <u>K463E</u>	nil	K463
	<i>Af</i> -B17	I465N	nil	I465
	<i>Af</i> -B1	K41R, I105V	nil	K41
	<i>Af</i> -B7	P305L	nil	
	<i>Af</i> -B8	P305S	nil	P305
C-terminal α -helical subdomain of large subunit	<i>Af</i> -B9	T392A	nil	T392
	<i>Af</i> -B12	E438G	nil	E438
Dimer-dimer interface	<i>Rs</i> -R18	L105M	nil	
	<i>Rs</i> -R19	L105Q	nil	L105
	<i>Af</i> -B1	<u>K41R</u> , I105V	nil	K41
	<i>Af</i> -B6	F220L	nil	F220
β -strand D of small subunit	<i>Rs</i> -R10	nil	<u>F91L</u> , <u>T95S</u> , S108C	F91, T95
	<i>Af</i> -B21	K146M	<u>V112A</u>	V112

Table 5-3: Interacting region of selected Rubisco variants with the phenotype of increased k_{obs} .

5.3.1 Conformational changes during transition between open-to-closed state in Rubisco

The structure of Rubisco and the conformational changes that occur during transition from open and closed state have been well studied for the Form IB enzymes from tobacco and spinach (Andrews, 1996; Duff et al., 2000; Schneider et al., 1992; Schreuder et al., 1993; Taylor et al., 1996b). The form I Rubisco is comprised of four dimers of large subunits with four small subunits capping the top and bottom of the large subunit octamer. A single functional unit of Rubisco is comprised of a dimer of anti-parallel large subunits with two active sites located at the interface between these two large subunits. The loops of the α/β barrel of the C-terminal domain contributes the majority of the active site residues while the N-terminal domain of the adjacent large subunit contributes additional residues (Table 5-4)(Cleland et al., 1998; Knight et al., 1990; Schneider et al., 1992). The substrate or inhibitor binds in an extended bent conformation in the active site. The substrate binding site can be divided into three distinct sites consisting of the two phosphate-binding sites (P1- and P2-phosphate binding sites) which are located at the opposite end of the α/β barrel and a metal binding site in the middle (Figure 5-8A). Residues involved in binding of substrate are from the various loops of the C-terminal and N-terminal domain (Figure 5-8, Table 5-4).

Previous studies using three-dimensional superpositions of structures of Rubiscos show that Rubisco can be divided into two distinct states, either open or closed state. The difference between these two states depends on the degree of solvent exposure in the active site (Duff et al., 2000; Schneider et al., 1992; Schreuder et al., 1993; Taylor et al., 1996b). In the open state, the active site is exposed to the solvent and has either no ligand bound (empty active site) or loosely bound ligands (substrate, inhibitor or products). The initial binding of substrate or inhibitor to the active site in the open state could be rapidly associated and dissociated. Over time, the binding of substrate or inhibitor-induced conformational changes leading to active site closure to form a closed tightly-bound complex (Pearce et al., 2003). In this closed state, the catalysis occurs using a mechanism involving electron and proton movements and no major structural transitions occur during catalysis (Cleland et al., 1998; Taylor et al., 1997). Therefore, in the closed state, the active site is occupied with a

tightly bound ligand (substrate or inhibitor) and shielded from the bulk solvent (Duff et al., 2000; Taylor et al., 1996b)(Figure 5-8). In many of the closed state Rubisco structures, the transition state analog CABP was used as it binds almost irreversibly to Rubisco.

During the open-to-closed state transition, there are several significant structural changes in both the large and small subunit (Schreuder et al., 1993). Following the binding of ligand to the active site, there is a two degree rigid-body rotation of the N-terminal domain towards the neighboring C-terminal domain as well as small subunit of one large-small pair around an axis parallel to the four-fold axis of Rubisco (Schreuder et al., 1993; Taylor et al., 1996b). This rotation brings the residues of the N-terminal domain closer to the active site and within binding distance to the ligand. Moreover, the rotation also partially blocks the active site from the bulk solvent (Taylor et al., 1996b). There are also movements of four loops from both the N-terminal and the C-terminal domain to facilitate active site closure (Schreuder et al., 1993). Mutations to residues of these loops would likely affect the rigidity of the closed state and allow premature opening of the active site. As such, with the significant structural changes during the transition from open-to-closed state, the structural analysis using Rubisco in a closed state would, therefore, provide a more accurate reflection on the positions of residues.

Residue	Role in catalysis	Location
N-terminal domain		
Glu60 (E60)	Forms intersubunit salt bridge with K334 which holds it in place; assists in loop 6 closure; involved in binding	60s loop
Thr65 (T65)	Binding of P1 phosphate	60s loop
Asn123 (N123)	Binding of C2-carboxyl (O-2) of the hydrated carboxyketone intermediate	α C- β E loop
C-terminal domain		
Lys175 (K175)	Accepts proton after enolization; stereospecific protonation of <i>aci</i> -acid in the last step of carboxylation reaction	Loop 1
Lys177 (K177)	Controls the pK of K175	Loop 1
Lys201 (K201)	Forms a carbamate and coordinated to a Mg ²⁺ ion; base that abstract the C-3 proton from RuBP to initiate after enolization	Loop 2
Asp202 (D202)	Carbonyl oxygen accepts proton from carbamate amino group (K201-NH) via hydrogen bonding which increase the positive charge at the nitrogen and stabilize the carbamate.	Loop 2
Asp203 (D203)	Coordinates Mg ²⁺	Loop 2
Glu204 (E204)	Coordinates Mg ²⁺	Loop 2
His294 (H294)	General base for nucleophilic attack at C-3	Loop 5
Arg295 (R295)	Binding of P2 phosphate	Loop 5
His327 (H327)	Binding of P2 phosphate	β -strand 6
Lys334 (K334)	Stabilizes the carboxylate during carboxylation; binding of P1 phosphate; intersubunit salt bridge with E60	Loop 6
Ser379 (S379)	Binding of C4-carboxyl (O-4)	Loop 7
Gly381 (G381)	Binding of P1 phosphate	Loop 7
Gly403 (G403)	Binding of P1 phosphate	Loop 8
Gly 404 (G404)	Binding of P1 phosphate	Loop 8

Table 5-4: Role and location of active site residues from N-terminal and C-terminal domain.

The role and position of the various active site residues are taken from (Cleland et al., 1998; Knight et al., 1990).

5.3.1.1 Ordering of two loops from the N-terminal domain during active site closure

Two loops from the N-terminal domain undergo structural changes during active site closure. The first is the loop at the N-terminus consisting of residues 9 to 20 which is disordered in the open state and becomes ordered in the closed state (Schreuder et al., 1993). The second loop known as the 60s loop (residues 61 to 82) resides next to the conserved active site residue, Glu60 (Duff et al., 2000). Glu60 is a completely conserved and catalytically essential residue which is involved in loop 6 closure and substrate binding via salt bridge formation with Lys334 (K334) of the adjacent large subunit (Cleland et al., 1998).

The 60s loop is also the loop connecting α -helix B and β -strand C of the N-terminal domain (Figure 5-7). During the transition from open-to-closed state, there are different structural changes at different regions of this loop (Duff et al., 2000). Residues (61 to 65) near the beginning of the 60s loop participate in the quasi-body rotation. While residues (66 to 75) at the center of the 60s loop remained anchored against the C-terminal domain and they maintain a fixed position relative to the C-terminal domain in both open and closed state. Lastly, there is a gradual transition between open and closed state for residues (76 to 81) near the end of the 60s loop (Duff et al., 2000).

Many residues within this 60s loop are highly conserved, including active site residues, Thr65 and Trp66 (Figure 5-9B). Thr65 forms part of the P1 phosphate binding site (P1-binding site) where it forms a hydrogen bond with the P1 phosphate group of the incoming sugar phosphate (Table 5-4). Other residues from the P1-binding sites include G381 from loop 7 and Gly403 and Gly404 from loop 8 (Figure 5-8A). During the transition to the closed state, the rotation of the N-terminal domain is coupled with rotation of the backbone ψ torsion angle of Thr65 which optimally positions the side chain of Thr65 for substrate binding. The preceding glycine residue (Gly64) provides the flexibility for this transition. In addition, there is an 8° rotation in Trp66's side chain which could be important for the rotation of Thr65 (Duff et al., 2000). Trp66 sits adjacent to Thr65 and is also part of the P1-binding site where its indole ring is in van der Waals contact with two of the phosphate oxygen atoms (Knight et al., 1990).

Similar to the two distinct states of the active site, Thr65 also exists in two states depending on the state of the active site. In the open state, the side chain of Thr65 forms a hydrogen bond with the carbonyl oxygen of Thr67 and Thr67 also forms a hydrogen bond with Trp462 via their side chain. Trp462 is the first residue of the C-terminus and its interaction with Thr67 restricts the latch residue, Asp473, of C-terminus from interacting at the latch site. Therefore, these interactions contribute to stabilizing the open state (Duff et al., 2000). In the closed state, the hydrogen bond between Thr67 and Thr65 is broken as the side chain of Thr65 is rotated away to interact with the P1-phosphate. The loss of hydrogen bond affects the position of Trp462 where its C α move 1-1.5Å. This, in turn, alters the conformational state of C-terminus where it is not restricted and can be attached to the latch

site via Asp473. Therefore, Thr65's rotation could contribute towards ordering of the C-terminus in the closed state (Duff et al., 2000).

Previous mutagenic study on Thr65 using *Synechococcus* PCC6301 Rubisco (*Syn*-Rubisco) by Morell et al. (1994) demonstrated that maintaining the hydrogen bond at position 65 is critical for stabilizing both the open and closed state. Three *Syn*-Rubisco mutants were generated with serine, alanine or valine present at position 65 in the large subunit. These mutants had reductions in k_{cat}^C , carboxylation efficiency, and $S_{C/O}$, as well as increases in $K_m(RuBP)$. These changes were due to the perturbations to the P1-binding site and this affects the transition from open-to-closed state whereby the active site is pre-maturely open. The reduction in $S_{C/O}$ could be due to the indirect perturbation to the weak hydrogen bond between Lys334 of loop 6 with Thr65 (Duff et al., 2000). The changes in kinetic properties were the least in T65S compared to the other two mutants as serine contains a hydroxyl group which allows hydrogen bonding with Thr67 and P1-phosphate (Morell et al., 1994).

5.3.1.2 Loop 6 functions as a lid that closes over the active site

Loop 6 (residues 328 to 338) of the C-terminal domain is the third loop that undergoes extensive conformational changes during the open-to-closed transition. It is the loop connecting α -helix 6 and β -strand 6 and many residues within this loop are well conserved (Figure 5-7). In the open state, loop 6 is retracted away from the active site and the active site is accessible to the bulk solvent. In the closed state, loop 6 becomes extended and folds over the active site, functioning as a lid. This closes off the active site from the bulk solvent (Schreuder et al., 1993; Taylor et al., 1996b). This movement of loop 6 is achieved by the changes in the ϕ and ψ angle of Val332 and Glu338 as well as the two glycine residues (Gly333 and Gly337) residing adjacent to them, which provides the flexibility for this transition (Duff et al., 2000). In addition, Val331 has also been shown to be part of the hinge for loop 6's movement (Newman et al., 1993). These residues are highly conserved in Rubisco along with the strictly conserved active site residue, Lys334. Lys334 sits at the apex of loop 6 and in the closed state, its side chain extends into the active site. Lys334 plays an important role in stabilizing and orientating the intermediate as it interacts with one of the two oxygens of CABP, which is equivalent to those of the canonical substrate

CO₂. During the carboxylation reaction, Lys334 polarizes the oxygens atom in CO₂ and this promotes the electrophilic attack on C₂ of the RUBP-enediol, giving the rise to the carboxyketone intermediate (Cleland et al., 1998). In addition, Lys334 also forms an inter-subunit salt bridge with Glu60 and hydrogen bonds with Thr65 from the N-terminal domain of the adjacent large subunit (Duff et al., 2000; Knight et al., 1990; Lorimer et al., 1993).

In addition to functioning as a lid of the active site, loop 6 also plays an important role in influencing the ability of Rubisco to discriminate between its gaseous substrates. It was first demonstrated by Chen et al. (1989) that a V331A substitution in *Chlamydomonas* Rubisco led to a ~40% decrease in specificity. Subsequent mutagenetic studies on Val331, Lys334, and Leu335 from different species of Rubisco all resulted in decreases in specificity (Gutteridge et al., 1993; Lee et al., 1993; Parry et al., 1992; Whitney et al., 1999). This decrease is likely mediated through Lys334 as it is the only residue in loop 6 that interacts with the enediol intermediate.

5.3.1.3 The C-terminus folds over loop 6 during active site closing

The last structural element that undergoes conformational changes during the open-to-closed transition is the C-terminus (residues 462 to variable terminus). In the open state, the C-terminus residues 462 to 487 add two turns to the C-terminal helix (α -helix G and H) while the rest of the residues (488 onwards) are disordered (Schreuder et al., 1993). In contrast, the C-terminus becomes ordered and extends towards the active site where it folds over loop 6 as a nearly straight polypeptide chain (Duff et al., 2000; Schreuder et al., 1993). This stabilizes the closed state and excludes the active site from the bulk solvent. The C-terminus is held in place via a single residue latch residue, Asp473. Asp473 is a strictly conserved residue and there is an extensive network of interactions involving Asp473 and nearby residues that contribute towards securing the C-terminus in place.

Asp473 forms salt bridges with Arg134 of the N-terminal domain and His310 of the C-terminal domain (Satagopan et al., 2004). In addition, the main chain amide of Asp473 and its adjacent residue, Ile472 also forms a hydrogen bond with the side chain of the Glu336 of loop 6 which secures the C-terminus over loop 6. Surrounding the latch site, there are multiple positively charged residues (Arg41, Lys305, and Arg312) and this creates a local

positive charge. This also suggests that the salt bridge formed by Asp473 is stronger than an ordinary salt bridge and that the C-terminus needs to be maintained at a high tension for it to remain secure via the single latch. This tension is contributed by two sets of lysine and phenylalanine residues; Phe467 and Lys334 (loop 6), Phe469 and Lys128 (α C- β B loop) (Duff et al., 2000). All in all, the interactions of the C-terminal residues with the body of the enzyme via loop 6 residues and other N-terminal residues contribute towards securing the C-terminus in place.

As the C-terminus plays a role in securing loop 6 in the closed state, studies have also demonstrated that C-terminus also influences the specificity of Rubisco. The truncation of at least 10 amino acids (including residue Phe467) from C-terminus in *Syn*-Rubisco was shown to cause a significant reduction in specificity (Gutteridge et al., 1993). While substitution of Asp473 to either alanine or aspartate in *Chlamydomonas* Rubisco caused ~14% decreased in specificity and ~86% decreased in carboxylation efficiency. This reduction in specificity is due to the disruption in interactions between C-terminus and loop 6 and likely mediated through Lys334 (Satagopan et al., 2004). These studies also demonstrated that the C-terminus is not essential for catalysis but is required for maximal activity and stability (Andersson, 2008).

The C-terminus also plays an important role in activation of inhibited red-type Rubisco complexes where the active site is in the closed site with a tightly bound inhibitor (Bhat et al., 2017a; Mueller-Cajar et al., 2011). Red-type Rubisco has elongated C-terminus and the latching of C-terminus over the active site involves a triad of three highly conserved residues; Asp473, Asp476, and Phe477. The red-type Rubisco activase recognizes the inhibited complex and docks onto Rubisco side-on. It subsequently threads the C-terminus through its central pore and this induced conformational changes at the active site which leads to inhibitor release (Bhat et al., 2017a; Mueller-Cajar et al., 2011). The transient pulling of the C-terminus by the activase will cause the anchor points to release and the C-terminus will detach from the enzyme body. At the same time, the activase shifts from a side-on to a tilted position and this movement destabilizes the residues near the end of the 60s loop. Therefore, the release of C-terminus and destabilization of the 60s loop results in

the retraction of loop 6. This leads to opening of the active site and eventually release of the inhibitor (Bhat et al., 2017a).

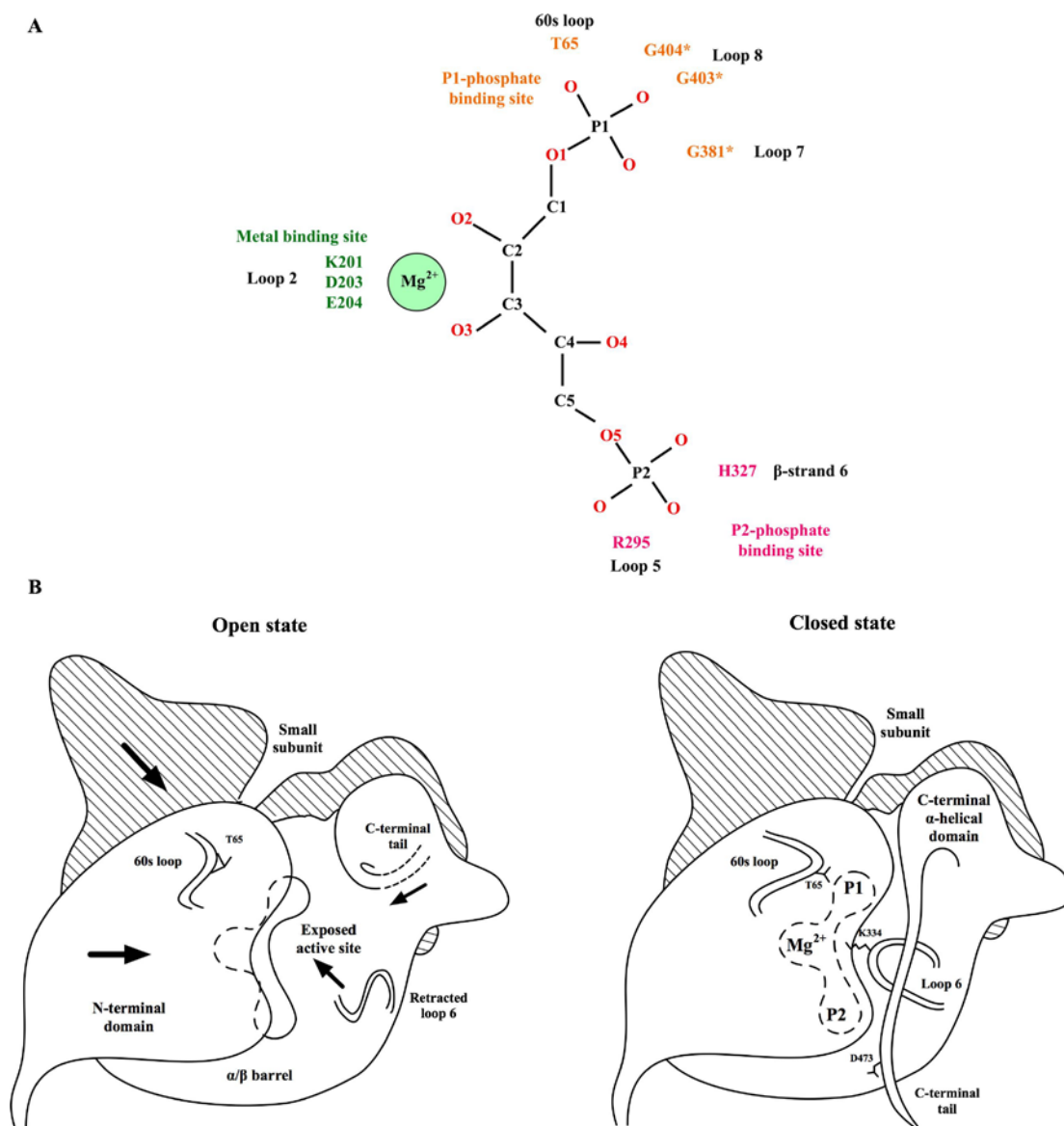


Figure 5-8: Substrate-binding sites and conformational changes during the transition between open and closed state in form I Rubisco

(A) There are three distinct parts in the substrate binding site where the ligand (substrate or inhibitor) binds in a bent conformation. The three sites are P1-phosphate, P2-phosphate, and metal (Mg^{2+}) binding sites. Residues that form these sites are shown and those marked with an asterisk used their main chain amide group for interaction with ligand while the rest used their side chain. In the

absence of Mg^{2+} metal ion, the side chains of residues at the metal binding site will directly interact with the ligand. In this diagram, the ligand is RuBP. Other active residues that interact with RuBP after enolization and CO_2 additions are not shown in this figure. (B) Simplified scheme of the conformational changes in Rubisco during the transition between the open and closed states. In the open state, the active site is exposed to the solvent whereas, in the closed state, the active site is shielded from the solvent. The arrows in the open state represent the conformational changes that occur during transition towards the closed state. Following ligand binding, there is a rotation of the N-terminal domain and small subunit relative to the α/β barrel. The 60s loop, which contained Thr65 rotates towards the C-terminus and Thr65 come in contact with the P1-binding site. In the C-terminal domain, loop 6 extends from its retracted form, folding over the active site. This allows the side chain of Lys334 of loop 6 to be extended into the active site where it interacts with the ligand. The C-terminus fold over loop 6 and further secure it while Asp473 of C-terminus holds the C-terminus in place. This figure was made with modification from Duff et al. (2000).

5.3.2 Mutations found in the 60s loop caused increased inhibitor release rates and displayed trade-offs in kinetic properties

The first region where clustering of selected residues was observed in this study was at the 60s loop which contains two active site residues, Glu60 (E60) and Thr65 (T65). Glu60 is involved in loop 6 closure and substrate binding while Thr65 is involved in binding of P1-phosphate of sugar phosphate at P1-binding site (Table 5-4). The 60s loop is also a major interface between the N-terminal domain of one large subunit and the C-terminal domain of the neighboring large subunit where residues on the 60s loop interact extensively with loop 1 and loop 8 of the adjacent C-terminal domain. Loop 1 contains conserved catalytic residues, Lys 175 and Lys177 while loop 8 contains conserved P1-binding site residues, Gly403 and Gly404 (Cleland et al., 1998)(Table 5-4).

Multiple Rubisco variants had mutated residues which were either located on the 60s loop or interact with this loop. Rubisco variants located within the 60s loop were *Af*-B5 with S61T substitution, *Rs*-R4 and *Rs*-R5 with T71A/S (T73A/S) substitution, *Rs*-R6 and *Rs*-R7 with R81C/S (R83C/S) substitution and *Rs*-R16 with T75A (T77A) substitution. Rubisco variants that interact with the 60s loop residues were *Rs*-R14 with Y24F (Y26F) substitution and *Rs*-R15 with P50L (P52L) substitution. Most of these residues were strictly conserved except, Pro50 (P50) and Arg81 (R81) (Figure 5-9).

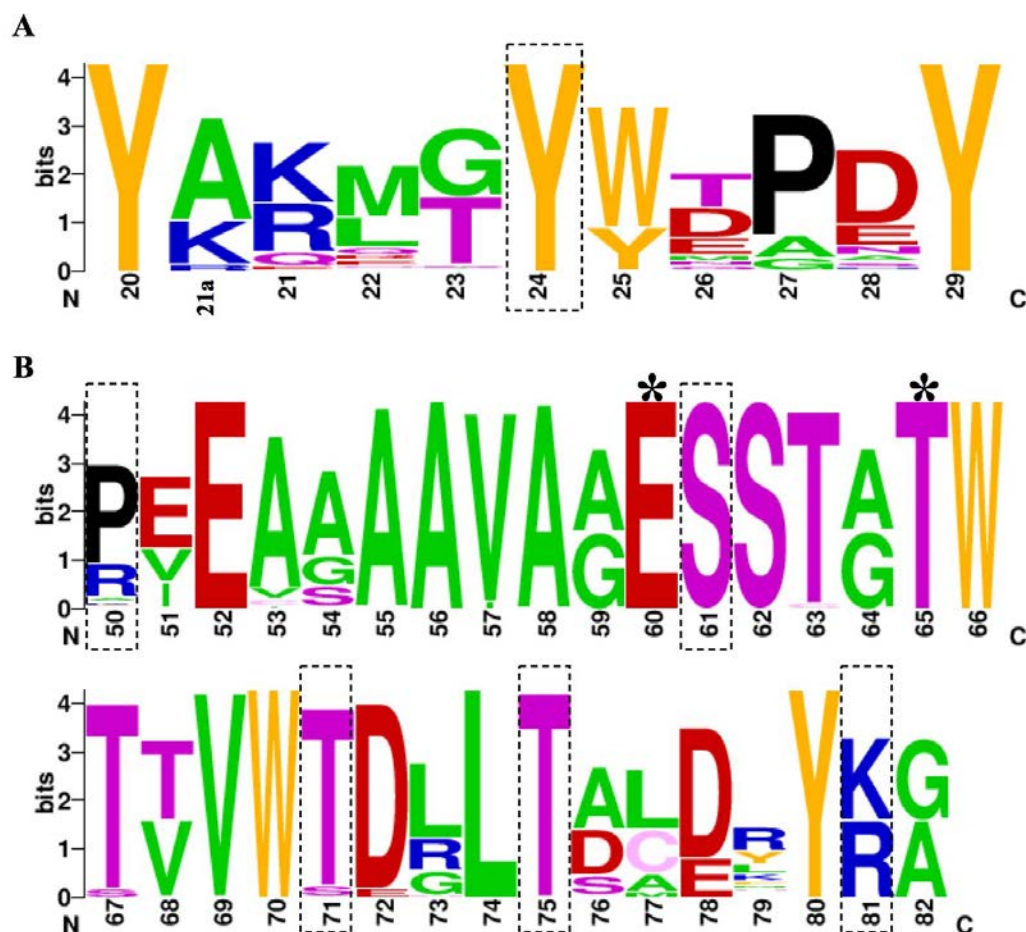


Figure 5-9: Sequence conservation of residues from two selected regions of the N-terminal domain of Rubisco large subunit.

(A) Sequence conservation of residues (20 to 29) near the N-terminus where Tyr24 is highly conserved. The residue at position 21a is only present in a few species of Rubisco. (B) Sequence conservation of residues from α -helix B (residues 50-60) and the 60s loop (residues 58-82). Selected residues from Rubisco variants are highlighted using dotted box while conserved active site residues are indicated with an asterisk. The sequence logos of these residues were generated using WebLogo (Crooks et al., 2004) where it represents the sequence conservation. This sequence logo used the sequence alignment of 120 form I Rubisco sequence where there is equal representation of each sub-type of form I Rubisco (Form IA, IB, IC, and ID). The overall height of each stack of letter(s) represents the sequence conservation at that position. While, the height of the individual letter within the stack indicates the relative frequency of the residue at that position. Positively charged residues are in red, negatively charged residues are in blue, polar non-charged residues are in purple, hydrophobic residues are in green, aromatic residues are in yellow, proline is in black while cysteine is in pink. The numbering of Rubisco large subunit is according to Spinach Rubisco and residues that are not commonly found in the different species are labeled alphabetically labeled with the

numbering of the subsequent conserved residue. This color scheme and the numbering of residues will be consistent for all the sequence logos in this thesis.

5.3.2.1 Pro50 substitution may influence the position of Thr65 in the active site

Pro50 (P50) is located on the surface of Rubisco near the equator of the holoenzyme and specific at the first position of α -helix B which also contains Glu60 at the end. As proline has one fixed dihedral angle, the side chain ring was position towards the surface where it does not affect the orientation of the helix. Therefore, the presence of Pro50 is likely to provide rigidity for correct orientation of α -helix B. Pro50 is also in van der Waals' contact with neighboring residues, Val48 (V48), Asp49 (D49) and Tyr97 (Y97) which are also located on the surface. The P50L substitution could disrupt these contacts and affect the positioning of the helix as well as the following 60s loop. In comparison to proline, leucine has a bulkier and hydrophobic side chain. This hydrophobic side chain should be oriented away from the surface towards the interior which may cause a steric clash with nearby residues. In addition, leucine can form hydrogen bond via its backbone to residues in the helix. Therefore, to accommodate these changes, the orientation of α -helix B and 60s loop are probably affected in P50L variants. It is likely the P50L substitution affects the conserved residues (Glu60 and Thr65) in these two places.

The kinetic analysis revealed that P50L variant (Rs-R15, P52L) demonstrated relative similar carboxylation rate, K_C and specificity as wild type (Table 5-1). This suggested that the final position of catalytic residues, Glu60 was not severely affected as the carboxylation reaction proceeded similarly to wild type. Instead, the positioning of Thr65 of the P1-binding site was likely altered as reflected in the low affinity towards RuBP with the high $K_m(\text{RuBP})$ value among the variants assayed. In addition, there was a ~4-fold increase in the rate of inhibitor released (k_{obs}) compared to wild type (Figure 5-10A). Therefore, the binding of RuBP as both substrate and inhibitor were both affected in P50L variants and likely due to perturbation in Thr65's position.

5.3.2.2 Ser61 substitutions negatively influence catalysis via Glu60

In contrast to Pro50, Ser61 (S61) is a highly conserved residue (Figure 5-9B). Ser61 resides next to conserved active site residue, Glu60, and is located on the 60s loop. The side chain of Ser61 forms hydrogen bonds with both the carbonyl oxygen of Val57 (V57) as well as the side chain of Tyr103 (Y103) of the same large subunit (Figure 5-10D). Ser61 is also in van der Waals contact with Ala58, Ser119, and Val124. All these interacting residues are highly conserved at their respective positions which suggest that these interactions are important and contribute towards the structural rigidity of the active site. Similar to Ser61, Val124 also resides next to an active site residue, Asn123, which interacts with the carboxylate-intermediate during catalysis.

The S61T substitution in *Af-B5* is likely to maintain hydrogen bonding with Val57 and Tyr103 as well as maintaining the van der Waals contact with Ala59, Ser119, and Val124. These interactions remain as threonine is structurally similar to serine with a hydroxyl group (Betts et al., 2003). However, threonine contains an additional methyl group which adds bulk to the local environment. This would cause a change in local conformation and affect the positions of the 60s loop residues including Glu60 and Thr65. As Ser61 resides close to the active site, the S61T mutation caused changes to the kinetics with decrease carboxylation rate, reduced affinity for CO₂ and RuBP and increased k_{obs} . It is likely that the effect of S61T's substitution is mediated through the two active site residues of the 60s loop, Glu60 and Thr65, and potentially Asn123. The increased k_{obs} was likely due to alteration in the positioning of Thr65 which reduce the rigidity of the active to favor inhibitor release.

5.3.2.3 Tyrosine 24 mutation increased the flexibility of the 60's loop and N-terminus

Tyr24 (Y24) resides near the N-terminus loop where it forms a hydrogen bond via its side chain with Thr63 (T63) of the 60s loop of the same large subunit (Figure 5-10E). Both Tyr24 and Thr63 are highly conserved which suggested that their interaction is important for the packing around the active site. In addition, Thr63 also forms a hydrogen bond with the conserved active site residue Lys177 (K177) of the adjacent large subunit via their backbone amide and carbonyl oxygen respectively (Figure 5-10E).

The Y24F substitution would disrupt the hydrogen bond formation between Tyr24 and Thr63 due to the absence of a hydroxyl group in phenylalanine. Other interactions such as van der Waal contacts that Tyr24 may have via its aromatic ring would be likely maintained when substituted with phenylalanine. The loss of a hydrogen bond between Tyr24 and Thr63 is likely to reduce the rigidity of the active site via increasing flexibility of the 60s loop and N-terminus loop which favors active site opening to allow inhibitor or substrate release. This is reflected in the change in affinity for CO₂ and RuBP as well as the ~3.5-fold increase in the rate of inhibitor release (Figure 5-10A).

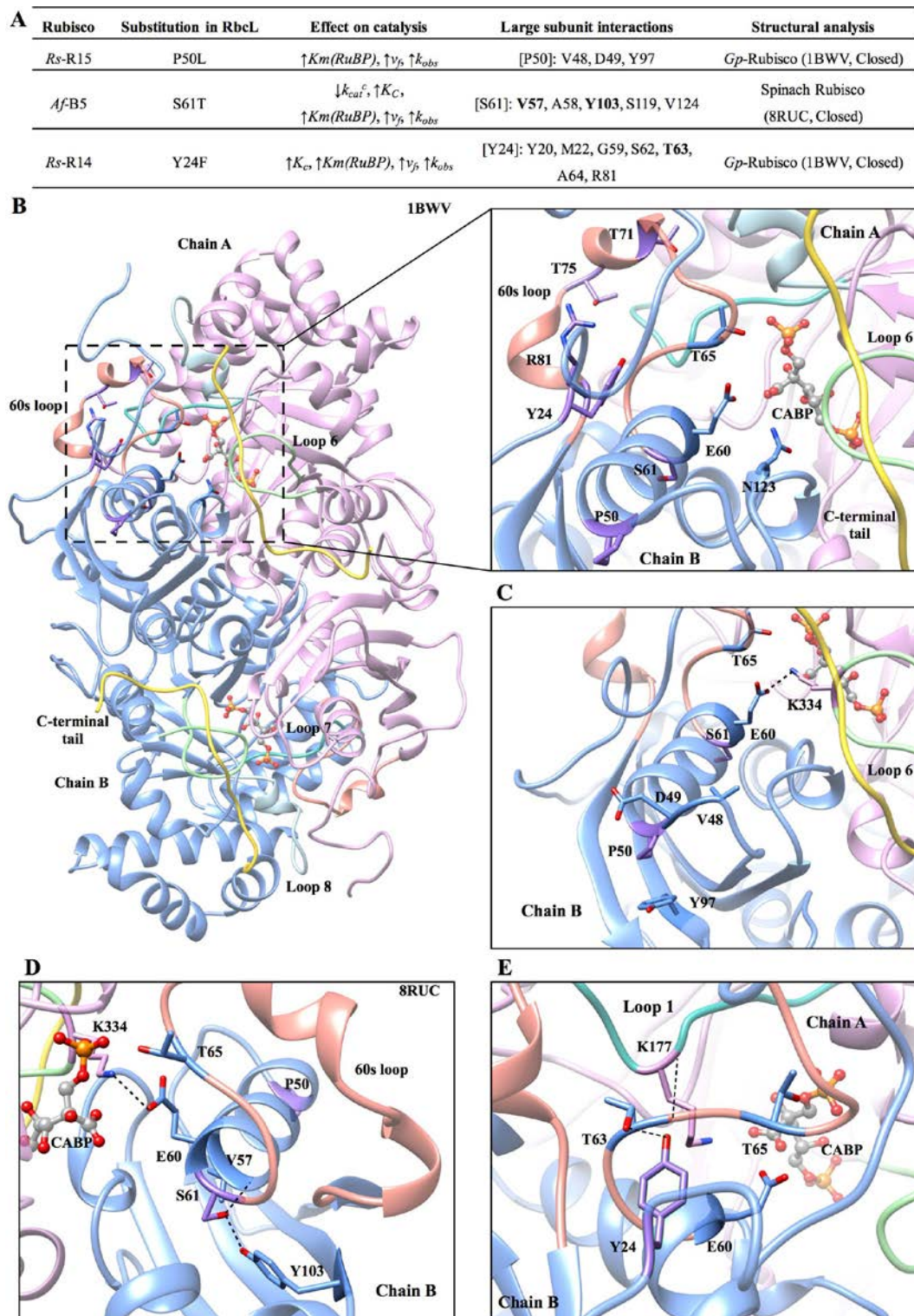


Figure 5-10: Position and interaction of selected residues residing or interacting with the 60s loop.

(A) Summary of the effects on catalysis of Rubisco variants that interact with 60s loop residues. The name of the Rubisco variant and the substitutions it contained are stated. Repeating mutated residues are underline for variants that contained multiple mutations. The effect on catalysis for each variant are relative to their wild type Rubisco. Under the section of interactions, the residues used in structural analysis are indicated in a bracket as the residue at the certain positions may not be conserved in the Rubisco used for structural analysis. In additions, the interactions are from side chain of selected residue to both main chain and side chain of interacting partners. Interacting residues that are involved in polar contacts (hydrogen bond or salt bridge) are in bold while the rest of are in van der Waals contact. The species, PDB code and state of the active site (open or closed) of Rubisco used for structural analysis are also stated. k_{cat}^c is maximum carboxylation rate, K_C and $K_m(RuBP)$ is the Michaelis constant for CO₂ and RuBP as a substrate respectively. While k_{obs} is the relative rate of inhibitor release. The data for these kinetic constants were obtained from Table 5-1, 5-2 and 4-7. (B) An overall view of the 60s loop and active site from a single large subunit dimer. The close-up view of the position of the selected residues in the 60s loop is shown on the right panel. Conserved N-terminal active site residues Glu60, Thr65 and Asn123 are also shown. (C) Pro50 resides at the amino-end of α -helix B which also contain Glu60 at the carboxyl end. (D) Ser61 resides next to conserved active site residue, Glu60 which formed a salt bridge (3.15Å) with Lys334 of loop 6 of the adjacent large subunit. The side chain of Ser61 forms hydrogen bonds with V57 and Y103 with a distance of 2.58Å, 2.82Å. In this panel, the view was from the bottom (carbonyl end) of α -helix B where Glu60 resides in. (E) Tyr24 is located near the N-terminus and it interacts with Thr63 of the 60s loop via hydrogen bonding with a distance of 2.91Å. Thr63 also form hydrogen bond (3.10Å) with Lys177 of the adjacent chain via the main chain amide and carbonyl groups. Structural analysis for these selected residues was performed using *G. partita* Rubisco (1BWV) except Ser61 where Spinach Rubisco (8RUC) was used instead as shown in section (C) of this Figure. In both 1BWV and 8RUC structures, the Rubisco's active site was in the closed state with transition state analog CABP bound. The Rubisco structure was shown in ribbon representation with the side chain of residues of interested in ball and stick representation. Residues that were selected during screening were colored shown in dark purple while other residues were colored according to their large subunit chain, chain A in plum and chain B in cornflower blue. The 60s loop, loop1, loop 6, loop 8 and C-terminal tail were in orange, cyan, light green, light blue and yellow respectively. Hydrogen bond and salt bridge between residues were represented using black dash. The nomenclature for numbering the large subunits is alphabetical (E.g. Chain A) while small subunits are labeled as numerically as Chain S'X' where X is a number (E.g. Chain S1). All the structural images in this thesis were created using Chimera and color scheme for the different structural elements will be consistent in the rest of the chapter.

5.3.2.4 Thr71, Thr75 and Arg81 interact with the 3₁₀ helix residues of the 60s loop

Thr71 (T71), Thr75 (T75) and Arg81 (R81) are located near the end of the 60s loop where there is a short 3₁₀ helix (residues 70-74) (Figure 5-11B). The residues from the 3₁₀ helix interact with residues of loop 1 and loop 8 of the adjacent large subunit and small subunit (Genkov et al., 2009). Loop 1 contains two active site residues, Lys175 and Lys177 while

loop 8 contains two P1-binding site residues, Gly403 and Gly404 (Cleland et al., 1998)(Table 5-4). The residues within the 3_{10} helix do not rotate during the transition between the open-to-closed state in contrast to other residues of the N-terminal domain (Duff et al., 2000). Sequence alignment of large subunits show that Thr71 and Thr75 are highly conserved while only positively charged residues are conserved at position 81 (Figure 5-9B). Arg81 and Lys81 are present in red-type and green-type Rubisco respectively (Figure 5-7).

5.3.2.4.1 The effect of Thr71 substitutions is likely mediated through the 60s loop and loop 8

Thr71 resides in the 3_{10} helix within the 60s loop and forms a structural hydrogen bond with Lys175 (K175) of the adjacent large subunit. Lys175's carbonyl oxygen atom is the acceptor and a proton of the hydroxyl side chain of Thr71 is the donor (Figure 5-11B). Lys175 is a strictly conserved active site residue located on loop 1 of the C-terminal domain where it protonates the *aci*-acid during the last step of catalysis to produce the second 3PGA molecule (Cleland et al., 1998)(Table 5-4). Thr71 is also in van der Waals contact with Val69 of the 60s loop, Pro176 of loop 1 and Ile407 of loop 8. Pro176 is highly conserved while leucine or isoleucine is conserved at position 407. Pro176 sits between two active site residues, Lys175 and Lys177 while Ile407 is located upstream of P1-binding site residues, Gly403 and Gly404 in loop 8 (Figure 5-11B). The conserved nature of Thr71 and its interacting partners suggested that these interactions contribute towards the conformational organization and the correct positioning of these active site residues.

The T71A substitution would lead to a loss of the hydrogen bond with Lys175 as alanine lacks polar side chains that can participate in hydrogen bonding. In addition, alanine is a smaller amino acid with a single methyl side chain as compared to threonine which has an additional branched side chain at the C β with methyl and hydroxyl group. This substitution could lead to closer packing around Ala71 and affect the positioning of Lys175, Pro176, and Ile407 as well as residues within 60s loop. In the T71S mutant, serine is likely to maintain hydrogen bond with Lys175 via its hydroxyl side chain. However, serine lacks the additional methyl group and this may affect interactions with other residues and affect the

local conformation. Both T71A and T71S substitutions led to a decrease in k_{cat}^c and an increased rate of inhibitor release but maintained similar $S_{C/O}$ as wild type. The T71A variant also had a decreased affinity for CO₂ (increase K_C) coupled with an increase in affinity for RuBP (decrease $Km(RuBP)$) while T71S had opposite changes in affinities for CO₂ and RuBP (Figure 5-11A). The differences in affinity towards the substrates might be attributed to the different changes in their local structural conformation brought by the substitution.

In both T71A and T71S variants, there was an increased rate of inhibitor release. As Thr71 is located on the 60s loop, substitutions of Thr71 could potentially affect the position of Thr65. In addition, Thr71 interacts with Ile407 of loop 8 and Thr71 substitution may influence positions of Gly403 and Gly404 of the same loop. Thr65 of the 60s loop, as well as Gly403 and Gly404 of loop8, contribute to the P1-binding site and the changes in local conformation at the P1-binding site as a result of Thr71 substitution could have facilitated the opening of the active site to allow faster inhibitor release.

5.3.2.4.2 T75A variant had the lowest affinity towards CO₂

Thr75 is located immediately after the 3₁₀ helix of the 60s loop where it forms a hydrogen bond with Asp72. Asp72 is highly conserved and resides at the center of this helix (Figure 5-11A). The proton of the hydroxyl side chain of Thr75 is the hydrogen bond donor while the carbonyl oxygen of Asp72 is the acceptor. Thr75 is in van der Waals contact with multiple residues. For instance, the hydroxyl side chain is in van der Waals contact with Ala77. While the methyl group at C β is in van der Waals contact with three highly conserved residues, Pro176, Gly179, and Leu180 of loop 1 of the neighboring large subunit (Figure 5-11C). The conservation of residues in these interactions suggests that they play an important role in ensuring the correct position of active site residues for catalysis. In addition, Thr75's hydrogen bond partner Asp72 forms extensive interactions with neighboring residues. This includes a salt bridge with Arg81 via its carboxyl side chain as well as Asp72 is in van der Waals contact with Tyr13, Val69, and Ala77 of the same subunit and Pro176 of loop 1 in the neighboring subunit (Figure 5-11D). Both Thr75 and Asp72 are in van der Waals contact with Pro176 via their C β and C γ atoms respectively.

The substitution of a smaller and hydrophobic alanine at position 75 would disrupt hydrogen bond formation with Asp72 and its interactions with loop 1 residues; Pro176, Gly179, and Leu180. The position of Asp72, loop 1 residues and other residues within the 60s loop may be altered due to conformational changes brought about by the T75A substitution. This led to T75A variant having increased affinity towards RuBP coupled with significant decreases for CO₂ (high K_C) while maintaining similar $S_{C/O}$ as wild type similar to T71A variant. The delineating of the K_C was most prominent observed among the variants assayed which resulted in significant reduction in its carboxylation efficiency as the k_{cat}^c of T75A variant had a slight increase. Unlike T71A, T75A variant's increased rate of inhibitor release is likely due to influence on Thr65's position in the 60s loop as Thr75 and Asp72 does not interact with loop 8.

5.3.2.4.3 Arg81 substitutions may influence positions of N-terminus loop and 60s loop

Arg81 resides near the end of the 60s loop where it interacts with the 3₁₀ helix via Asp72. The charged side chains of Arg81 and Asp72 form a salt bridge and this likely contributes towards the stability of this 3₁₀ helix as both residues are highly conserved (Figure 5-11E). Arg81 is also in van der Waals contact with multiple N-terminus residues (Tyr13, Met22, and Tyr24) and 60s loop residue (Thr63 and Ala77).

The substitution of Arg81 to either cysteine or serine would disrupt the salt bridge formation with Asp72 as both residues lack positively charge side chains. The shorter side chains of cysteine and serine would also likely affect the interactions with the residues from the N-terminus and 60s loop. In addition, the loss of the salt bridge could destabilize the 3₁₀ helix and this may alter interactions of 3₁₀ helix residues (70-74) with loop 1, loop 8, 60s loop and small subunit (Figure 5-11G). These substitutions are likely to alter the local conformation such that the rigidity of the active site is reduced. Therefore, it is possible that opening of the active site is facilitated and inhibitors are rapidly released as shown by the increased k_{obs} and decreased specificity. This decrease in specificity could be mediated through Thr65 which forms a weak hydrogen bond with Lys334 of loop 6 (Duff et al., 2000).

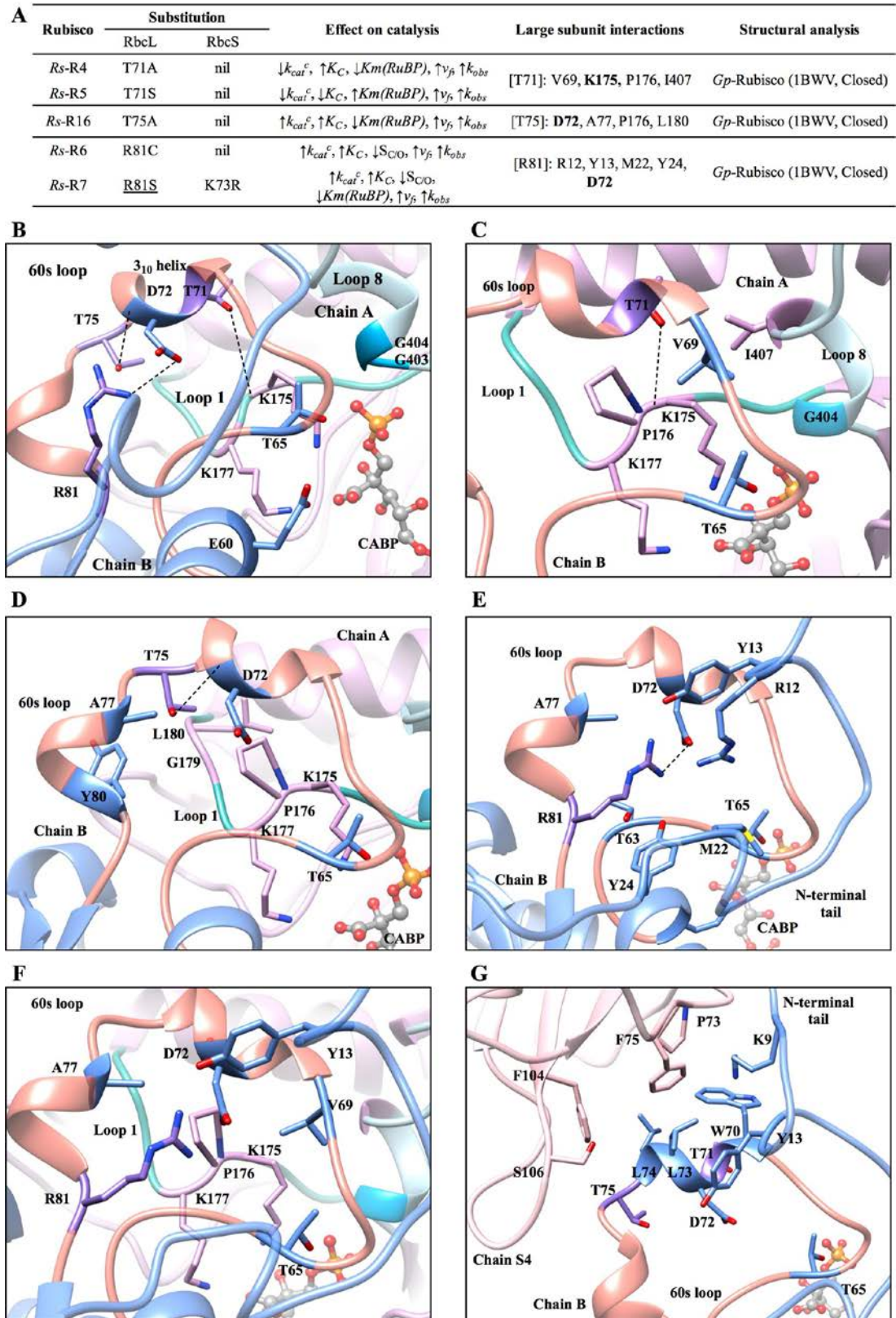


Figure 5-11: Interaction of selected residues with 3₁₀ helix residues of the 60s loop.

(A) Summary of the effect on catalysis of substitutions of the variants that interact with 3_{10} helix residues. $S_{C/O}$ is the specificity factor of Rubisco. (B) Overview of the interactions of selected residues with 3_{10} helix residues in the 60s loop. Thr71 resides within the 3_{10} helix and its hydroxyl side chain forms a hydrogen bond to the amide group of Lys175 with a distance of 2.84Å. Thr75 and Arg81 both form polar contacts with Asp72 of the 3_{10} helix with a distance of 2.82Å, 3.10Å respectively. Lys175 and Lys177 of loop 1 and Gly403 and Gly404 of loop 8 (deep blue) were indicated. (C) Interactions of Thr71 from the 3_{10} helix. Thr71 is in van der Waals contact with neighboring residues in the 60s loop (Val69), loop 1(Pro176) and loop 8 (Ile407). (D) Thr75 is positioned right behind 3_{10} helix and its side chain is oriented towards the 60s loop and loop 1 of the adjacent large subunit. Thr75 is in van der Waals contact with Ala77 and Tyr80 of the 60s loop, Gly179 and Leu180 of loop 1. (E) Asp72's intra- and inter- large subunit interactions. Asp72 is in van der Waals contact with Val69 and Ala77 of the 60s loop of the same large subunit. In the adjacent large subunit (Chain A), Asp72 is in van der Waals contact with Tyr13 and Pro176. (F) Arg81's interaction with residues from the N-terminus and 60s loop of the same large subunit. (G) The interface between the 60s loop of the large subunit (Chain B) with its adjacent small subunit (Chain S1). Trp70, Leu73, and Leu74 of the 3_{10} helix are in van der Waals contact with Pro73, Phe75, Phe104, and Ser106. Trp70 is in van der Waals contact with Lys9 and Pro73 (small subunit), while Leu73 is van der Waals contact with Tyr13 of the large subunit, Phe76 and Phe106 from the small subunit. Lastly, Leu74 is in van der Waals contact with Phe104 and Ser106 of the large subunit. Pro73 and Phe75 are two highly conserved small subunit residues located on the loop connecting β -B and α -B. Structural analysis for these residues was performed using *G. partita* Rubisco (1BWV) which is in a closed state with transition state analog CABP bound. The small subunit residues from chain S4 is colored in light pink while loop 1 and loop 8 are in light blue and yellow respectively. The color scheme of other structural elements is as stated in Figure 5-10.

5.3.3 Mutations targeting the C-terminus and C-terminal α -helical subdomain resulted in increased rate of inhibitor release

In this set of variants, the substitutions of residues were found at the C-terminus, C-terminal α -helical domain or interfaces near these two regions. There are three sets of mutations at these interfaces, the first set is Lys463 and Ile465 which reside at the beginning of the C-terminus. The second set consists of Arg41 and Pro305 which are located near the latch site of the C-terminus which is also near the end of C-tail. The last set comprised of Glu438 from α -helix G of the C-terminal α -helical domain and Thr392 from α -helix 7 which is close in space to Glu438. Substitutions of these residues resulted in an increased rate of inhibitor release compared to wild type.

The C-terminus plays an important role in active site closure by folding over loop 6 and it is held in place via the highly conserved latch residue Asp473 (Duff et al., 2000). The C-terminus is capped by the α -helical domain which is part of the C-terminal domain along

with the α/β -barrel. In contrast to the C-terminus, this α -helical subdomain has not been widely discussed in Rubisco research. This α -helical subdomain is comprised of α -helix G and α -helix H which do not contribute any residues towards substrate binding or catalysis. In addition, the majority of the residues from these two helices are not conserved, similar to the C-terminus. The C-terminal α -helical domain is close in space and in van der Waals contact with residues from α -helix 8, α -helix 7, loop 7, loop 8 and the adjacent small subunit.

5.3.3.1 Substitutions at K463 and I465 of the C-terminal strand may influence inhibitor release

Lys463 (K463) and Ile465 (I465) are both located at the beginning of the C-terminal strand and on the surface of Rubisco. Sequence alignments show that Lys463 and Ile465 are highly conserved among the different Rubiscos (Figure 5-12). Although Gly463 and Val465 can also be found at these two positions in some Rubisco. Both Lys463 and Ile465 had also been encountered in previous RDE screenings of *Syn*-Rubisco and only the I456V mutant was characterized (Duraó et al., 2015). Structural analysis of these two residues was performed using Spinach Rubisco (8RUC) where the active site is in the closed state.

Due to the proximity of Lys463 and Ile465, they have common interacting partners. They are the immediate neighboring residues, Trp462 and Glu464. Trp462 is a highly conserved residue which marks the beginning of the C-terminus. It forms extensive contacts with the loops containing substrate binding residues, including the 60s loop, loop 7 and loop 8. These interactions contribute towards the tight packing surrounding the active site during the closed state (Duff et al., 2000). In contrast to the highly conserved Trp462, negatively charged residues are commonly found at position 464. Glu464 forms a polar contact with a water molecule and its side chain is in van der Waals contact to the neighboring Lys463.

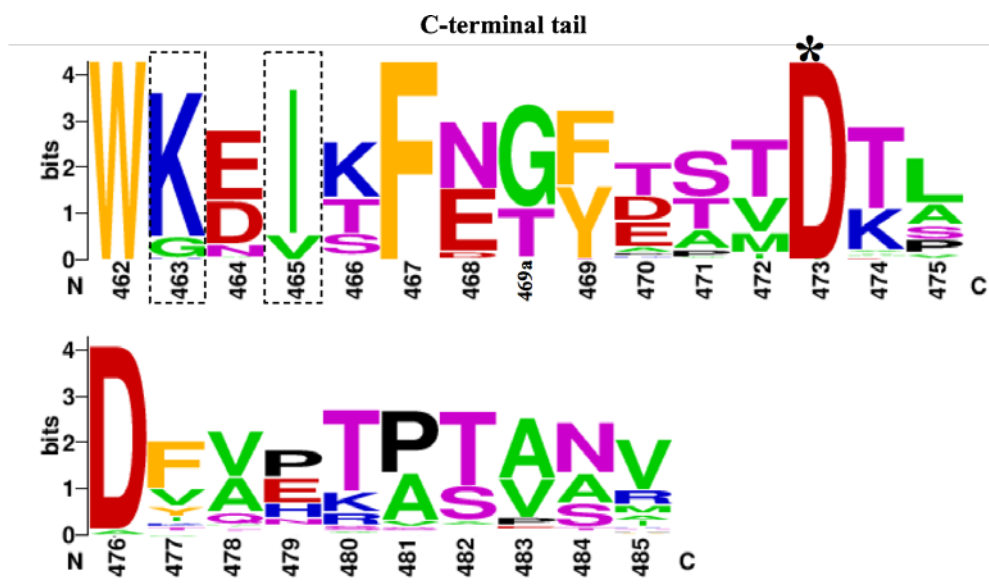


Figure 5-12: Sequence conservation of the C-terminal residues.

The C-terminus in Rubisco starts from the conserved Trp462 and its length varies from 13 to 23 residues. Lys463 and Ile465 are commonly found at their respective positions (dotted box). The strictly conserved latch residue, Asp473 is indicated with an asterisk. Residues at position 469a are found in C4 plants only. The color scheme and numbering of sequence logo are as stated in Figure 5-9.

5.3.3.1.1 Substitution at Lys463 alone may not be sufficient to allow rapid inhibitor release

Lys463 is located between Trp462 and Glu464 and its side chain is in van der Waals contact with Glu464 and Cys459 of α -helix G. The main chain of Lys463 is in van der Waals contact with Trp385 of loop 7 which also forms a hydrogen bond via its side chain to the carbonyl oxygen of Cys459. Trp385 is highly conserved while methionine and leucine are commonly found along with cysteine at position 459. In *Af*-Rubisco, Met459 is present and the hydrogen bond with Trp385 is likely to be maintained as it is involved in the main chain carbonyl oxygen. Interestingly, Cys459 had been previously studied using *Chlamydomonas* Rubisco by Marin-Navarro et al. (2006). Cys459 and Cys449 along with other cysteine residues were proposed to be part of the Rubisco catabolism network where the translation of redox signals subsequently results in conformational changes. However, *Af*-Rubisco lack

of some of these cysteine residues including Cys459 and Cys449. This suggests that the regulation of Rubisco catabolism may be different in photosynthetic and non-photosynthetic organisms.

In the RDE screening of *Af*-Rubisco, three variants (*Af*-B13, *Af*-B14, and *Af*-B15) containing K463E substitutions were isolated and they also had additional mutations in large and small subunits. The K463E substitution is likely to affect the position of the C-terminus as the side chain of both mutant Glu463 and Glu464 need to be orientated away from each other to avoid repulsive force. This may cause local conformational changes that could cause perturbation to the entire C-terminus and facilitate faster inhibitor release. However, it is also possible that K463E substitution alone may not be sufficient to cause such a perturbation as the latch residue, Asp473, was present to hold C-terminus in place. K463E alone will need to be characterized to clarify this point.

5.3.3.1.2 A single substitution at Ile465 can increase the rate of inhibitor release

Ile465 resides on the C-terminus where it is in van der Waals contact with residues from the N-terminus (Val17), 60s loop (Trp66), N-terminus (Val17), loop7 (His383) and C-terminus (Glu464 and Trp462) (Figure 5-13A, E). Ile465 participates in hydrophobic interactions with Val17, Trp66, and Phe467. In addition, the side chain of His383 forms a hydrogen bond to the carbonyl oxygen of Ile465 and His383 is close to Gly381 which is part of the P1-binding site (Figure 5-8A). These interacting residues are from either the same large subunit (Glu464, Trp462, and His383) as Ile465 or the adjacent large subunit (Trp66 and Val17). As these interacting residues are located on loops that are involved in active site closure, the majority of them are strictly conserved at their position except Glu464 (Figure 5-7).

Mutation at position 465 was selected twice with I456F and I456N substitution found in *Af*-B16 and *Af*-B17 respectively. Only the I465N mutant was characterized in detail and the substitution to a polar residue may disrupt the van der Waals contacts and hydrophobic interaction with the various loop residues. In addition, it is possible that the substitution would affect the position of Phe467 which interacts with Lys334 to maintain the high tension of C-terminus for latching securely via Asp473 (Duff et al., 2000). Therefore, I465N

substitution causes a local conformational change whereby these loops are slightly displaced such that the active site becomes less rigid and allow a premature release of inhibitor. This was shown to be a ~4.8-fold increase in the rate of inhibitor release and a trade-off with a reduction in k_{cat}^c and K_C .

The I465N substitution alone was able to bring about an increased rate of inhibitor release, probably the due to disruption of interactions with multiple loop residues that are involved in active site opening and closing. In addition, isoleucine and valine are commonly found at position 465, suggesting that hydrophobic interactions contribute to packing near the top of the active site.

In the previous directed evolutionary screening of *Syn*-Rubisco by Durao et al. (2015), a substitution of Ile465 to valine was observed. The I465V mutant had a slight increased k_{cat}^c as well as approximately twofold increased in solubility in *E. coli* while maintaining similar specificity as wild type *Syn*-Rubisco. This I465V mutation provided a survival advantage to the *Syn*-Rubisco by increasing Rubisco content which translated to higher RuBP turnover rate. It is likely that this substitution helps promote stable holoenzyme assembly in *E. coli*. However, the I465V substitution was not observed in the screening of *Rs*-Rubisco and *Af*-Rubisco which may indicate that potential improvement in assembly via this mutation does not confer fitness to those Rubiscos. In contrast to *Rs*-Rubisco and *Af*-Rubisco, *Syn*-Rubisco intrinsically possesses high k_{obs} rate constant for inhibitors (Pearce, 2006). Therefore, the I465V substitution may not be able to increase k_{obs} to provide fitness for *Rs*-Rubisco and *Af*-Rubisco to survive in RDE screening.

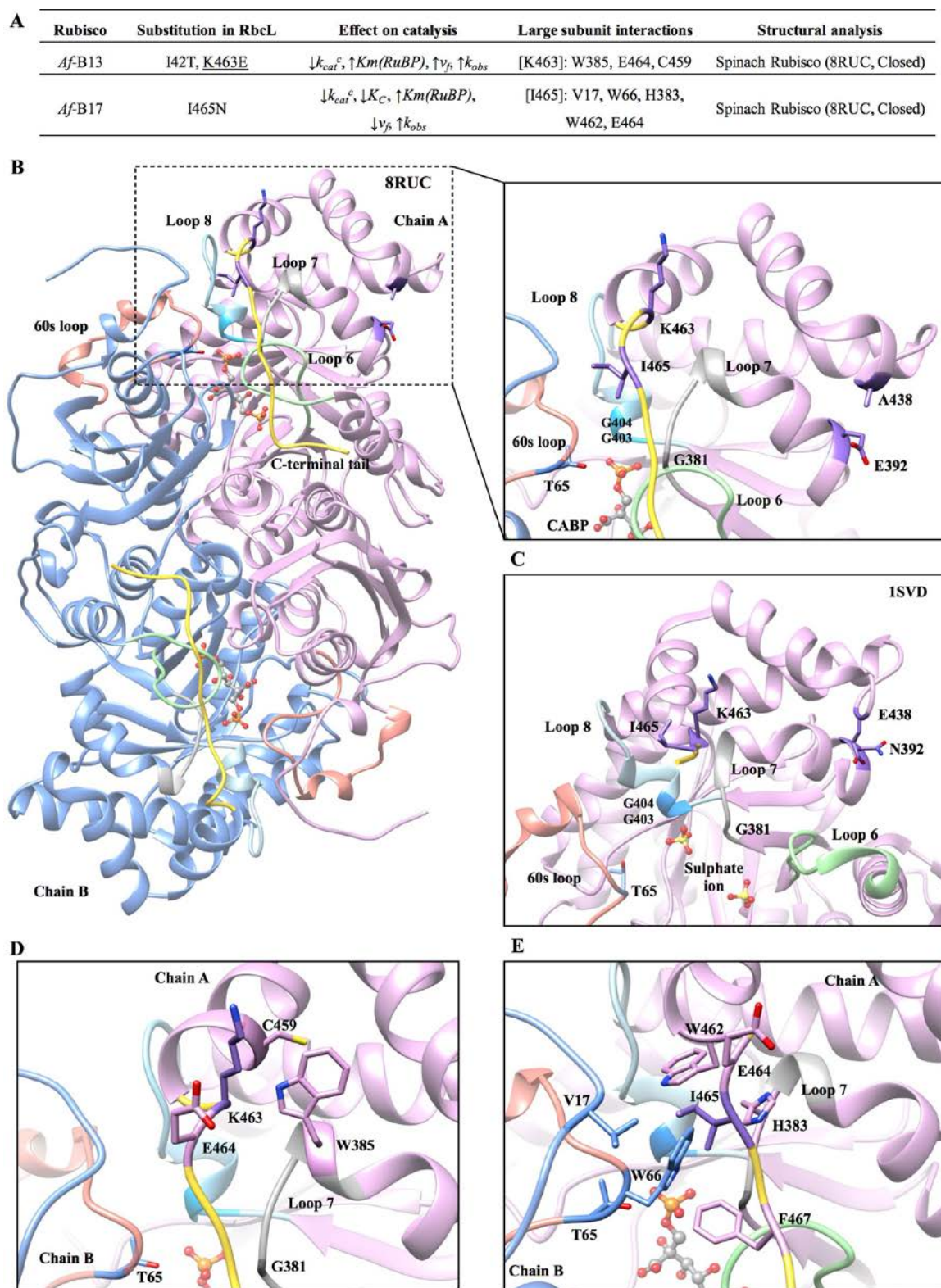


Figure 5-13: Substitution of C-terminal residues increased the rate of inhibitor release.

(A) Summary of effect of substitution of residues at the beginning of C-terminus. (B) Overview of the position of selected residues at the C-terminus and the C-terminal α -helical domain in Spinach Rubisco (8RUC, closed). Lys463 and Ile465 are located at the amino end of the C-terminal tail (right panel). The C-terminal α -helical domain (α -helix H and G) which connects the C-terminus to the α/β -barrel is also shown. (C) The position of selected residues in *Hn*-Rubisco which is in the open state. Residue 392 (Asn392 or Glu392) is located close in space with residue 438 (Ala438 or Glu438) (D) Lys463 of the C-terminal tail is in van der Waals contact with Glu464 and Cys459 in the closed state. (E) Interaction of Ile465 with residues of same and adjacent large subunit. Ile465 is in van der Waals contact with Val17, Trp66, His383, Trp462, and Glu464.

5.3.3.2 K41 and P305 are a pair of interacting residues located near the latch site of the C-terminus

Lys41 (K41) and Pro305 (P305) are a pair of interacting residues that were found independently from RDE screening of *Af*-Rubisco in *Af*-B1, *Af*-B7, and *Af*-B8. Although the residues are located far apart in the primary sequence, they are found close in space at the interface between N-terminal and C-terminal domain of a single large subunit (Figure 5-15, 5-11). Structural analysis of *H. neapolitanus* Rubisco (*Hn*-Rubisco) shows Lys41 forms a hydrogen bond with Pro305 where the proton from amide side chain of Lys41 is the donor while carbonyl oxygen of Pro305 is the acceptor (Figure 5-15A). Both residues are also located on the surface and near the latch site of the C-terminus. Arg41 (in *Hn*-Rubisco) and two positively charged other residues Arg312 and Lys305 create a local positive charge around the latch site where Asp473 forms a salt bridge with Arg134 of the N-terminal domain (Figure 5-16A). Both Asp473 and Arg134 are highly conserved in Rubisco and Asp473 is often referred to as the latch residue as it secures the C-terminus in place in the closed state (Duff et al., 2000; Satagopan et al., 2004).

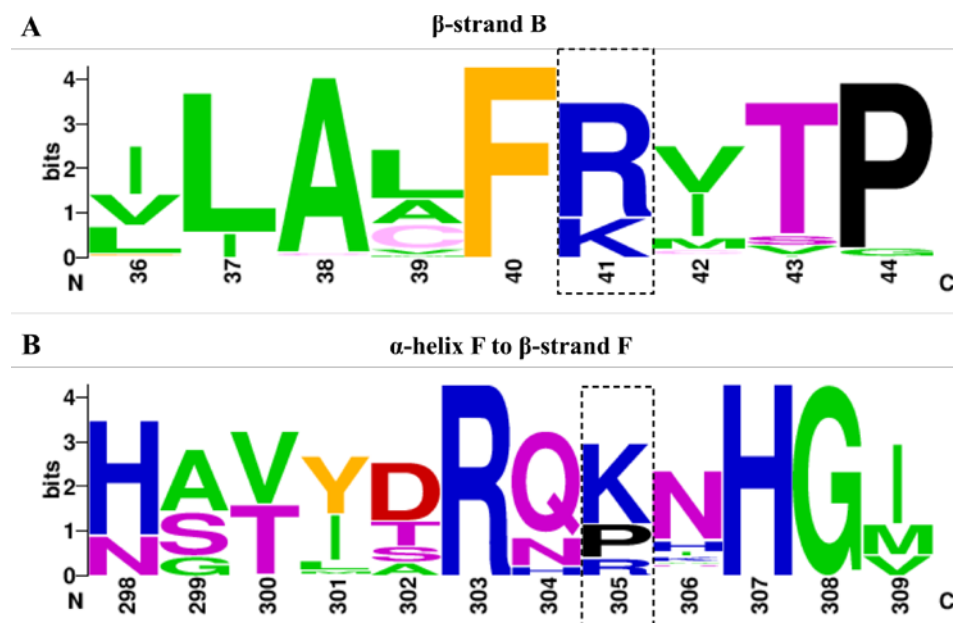


Figure 5-14: Sequence conservation of residues at position 41 and 305 of Rubisco large subunit.

(A) Sequence conservation of the β -strand B where positively charged residues are commonly found at position 41. (B) Sequence conservation of residues from α -helix F to β -strand F where residues at position 305 are not conserved. Lysine, proline, and arginine are found at position 305 with lysine been most prevalent compared to the other two residues.

5.3.3.2.1 Substitutions of Lys41 may affect active site closure via C-terminal tail

Lys41 is located in β -strand B and the positively charged residues (lysine and arginine) are always found at this position (Figure 5-14A). In addition to the hydrogen bonding with Pro305, Lys41 is in van der Waals contact with Val90 (V90), Ala96 (A96), Tyr98 (Y98) and Glu136 (E136) (*Hn*-Rubisco, Figure 5-15B). Lys41 and Glu136 are close in space with the conserved Arg134 where Lys41 is in the adjacent β -strand and Glu136 is on the same β -strand E as Arg134. All the interacting residue of Lys41 are conserved in *Af*-Rubisco except at position 96 where cysteine is present instead of alanine. These residues are all from the adjacent anti-parallel β -strands (Figure 5-15D).

Substitutions at Lys41 were found in two variants, *Af*-B1 and *Af*-B2 where both contained an additional mutation in the large subunit. Only *Af*-B1 was characterized in detail and it had K41R and I105V substitutions while *Af*-B2 had K41E and A448V substitutions. The

second I105V substitution in *Af*-B1 will be discussed below under the section 5.3.4 as mutations at position 105 were also encountered during a screening of *Rs*-Rubisco.

In nature, Arg41 has been selected for other form I Rubisco as positively charged residues are conserved at this position. Structural analysis of Spinach Rubisco has shown that Arg41's guanidinium side chain forms polar contacts (hydrogen bond or salt bridges) with Gln96 (Q96) and Glu136 via their side chains (Figure 5-16A). Arg41 also form additional polar contact with Glu467 as shown using *Gp*-Rubisco which had a longer C-terminal tail than Spinach Rubisco (Figure 5-16B). Glu467 is highly conserved among Rubisco with elongated C-terminus and this interaction is likely part of a polar contact network holding the C-terminus in place during catalysis. In addition to polar contacts, Arg41 is in van der Waals contact with Val90, Ile98 (I98), Arg134, Lys305 (K305) (Figure 5-16C). The positions of residues interacting with Arg41 are identical with Lys41 but the nature of some residues is different compared to *Af*-Rubisco and *Hn*-Rubisco. Among these interacting residues only Val90, Arg134 and Glu136 are highly conserved. For the remaining residues, Cys96, Tyr98, and Pro305 are present at *Af*-Rubisco instead.

The K41R substitution in *Af*-B1 likely maintains the hydrogen bond with Pro305 via arginine's guanidinium side chain. In addition, Arg41's side chain may form multiple new polar contacts with side chains of Cys96 and Glu136, similar to those observed in Spinach Rubisco. The formation of these new polar interactions would affect the local conformation around the N-terminal and C-terminal interface as well as with the C-terminus. In particular, the position of Arg134 is likely displaced as it is close in space with Arg41 and Glu136. This would disrupt the interaction between Asp473 and Arg134 and increase the flexibility near the end of C-terminus. Without the latch residue to hold C-terminus in place, the additional polar contact between Arg41 and Asp476 would likely be affected as well. Therefore, the packing of C-terminus to loop 6 would weaken and this would enable inhibitors to be rapidly released from the active site.

5.3.3.2.2 Pro305 substitution negatively influence C-terminal tail and loop 6 placement

Pro305 is a surface residue located on the loop between α -helix F and β -strand F (α F- β F loop, residues 303-307). This loop is a crossroad for multiple interfaces including the N-terminal and C-terminal domain interface of a single large subunit, interface near the end of C-terminus as well as the interface between a dimer pair of large subunits (Figure 5-15, 5-11). Pro305's rigid side chain aids in the turning of α F- β F loop but it does not interact with residues from the adjacent large subunit (Figure 5-15A). Proline is not conserved at position 305, arginine and lysine are more commonly present (Figure 5-7).

In addition to forming a hydrogen bond with Lys41 as described previously, Pro305 is also in van der Waals contact to Arg134, Arg303, and Asn304 (Figure 5-15C). The two arginine residues are highly conserved while asparagine and glutamine are conserved at position 304 (Figure 5-7). Both arginine residues are involved in active site closure via interacting with C-terminus and loop 6 residues. Arg134 forms salt bridge with latch residue Asp473 of the C-terminus while Arg303 forms polar contacts with Glu336 of loop 6 as well as van der Waals contact with Asp473 (Figure 5-16D)(Duff et al., 2000; Satagopan et al., 2004). Glu336 of loop 6 also form a hydrogen bond with the amide backbone of Asp473 (Figure 5-16D). These polar contacts stabilized the C-terminus during the closing of active site and contribute to the rigidity of the active site. In addition to polar contacts with loop 6, Arg303 forms extensive van der Waals contact with residues from the adjacent N-terminal domain from the neighboring large subunit (Chain B). These residues are from the β -strand E and the loop between α -helix C and β -strand E (α C- β E loop, residues 121-130) which also contains active site residue Asn123.

Two variants containing single substitution at Pro305 were obtained from screening of *Af*-Rubisco and they were *Af*-B7 and *Af*-B8 containing P305L and P305S substitutions respectively. Leu305 and Ser305 were not observed to be selected in nature and instead, positively charged residues like lysine and arginine are present in other Rubiscos. Structural analysis of Spinach Rubisco showed that Lys305 forms a salt bridge with Thr474 of the C-

terminus and contribute towards the extensive polar contact to stabilize C-terminus in the closed state.

Both P305L and P305S substitutions would lead to loss of rigid proline side chain which alters the positioning of α F- β F loop. This may lead to the loss of hydrogen bonding with Lys41 as result of shifting Ser305/Leu305's position and potentially destabilize the interface between the N-terminal and C-terminal domain. Arg303 of the α F- β F loop may be shifted such that it is no longer within polar contact distance with Glu336 of loop 6. This would also affect Arg303's interactions with the adjacent large subunit and potentially affect Asn123. In addition, Arg134's position may be shifted to accommodate changes in α F- β F loop and this would affect the salt bridge formation with Asp473 of C-terminus. The combination of these changes would consequently reduce the rigidity of the active site such that inhibitors could be rapidly dissociated. Indeed, the increased in k_{obs} was in both variants. However, this was coupled with a trade-off in carboxylation efficiency (Table 5-2).

Although similar trends in kinetics were observed in both variants, the magnitude of these kinetic changes was different. P305L variant had a 50% reduction in k_{cat}^c , 6-fold increase in k_{obs} and a significant reduction in specificity. In contrast, kinetic changes observed for P305S were half (25% reduction in k_{cat}^c , a 3-fold increase in k_{obs}) as compared to the P305L mutant and similar $S_{C/O}$ as wild type. This difference in kinetics was due to the nature of the substituted amino acid at position 305 (P305L, P305S). Leucine is a hydrophobic residue with a branched hydrocarbon side chain while serine is a small polar residue with a hydroxyl side chain. Ser305's polar side chain could potentially remain on the surface, thus participating in the extensive polar contacts near the C-terminus interface. In contrast, Leu305's hydrophobic side chain will be oriented away from the surface, distancing away from the polar contacts with the nearby C-terminus. In addition, the bulkier side chain of Leu305 may have a steric clash with nearby residues causing more residues to shift to accommodate it. Therefore, the conformational changes induced by the P305L mutation would be greater compared to P305S variants.

In the P305L substitution, there was a significant reduction of 10% in $S_{C/O}$ which could be mediated by either Arg134 or Arg303 with their interactions with Glu473 of C-terminus Glu336 of loop 6 respectively. Correspondingly, previous mutagenesis studies on Glu473 and loop 6 residues have reported that mutation of these residues resulted in a reduction in $S_{C/O}$ (Parry et al., 2003; Satagopan et al., 2004). Although the kinetic properties of the P305L variant were significantly impaired compared to other variants and wild type, a higher level of Rubisco expression was observed, likely compensating for the kinetics.

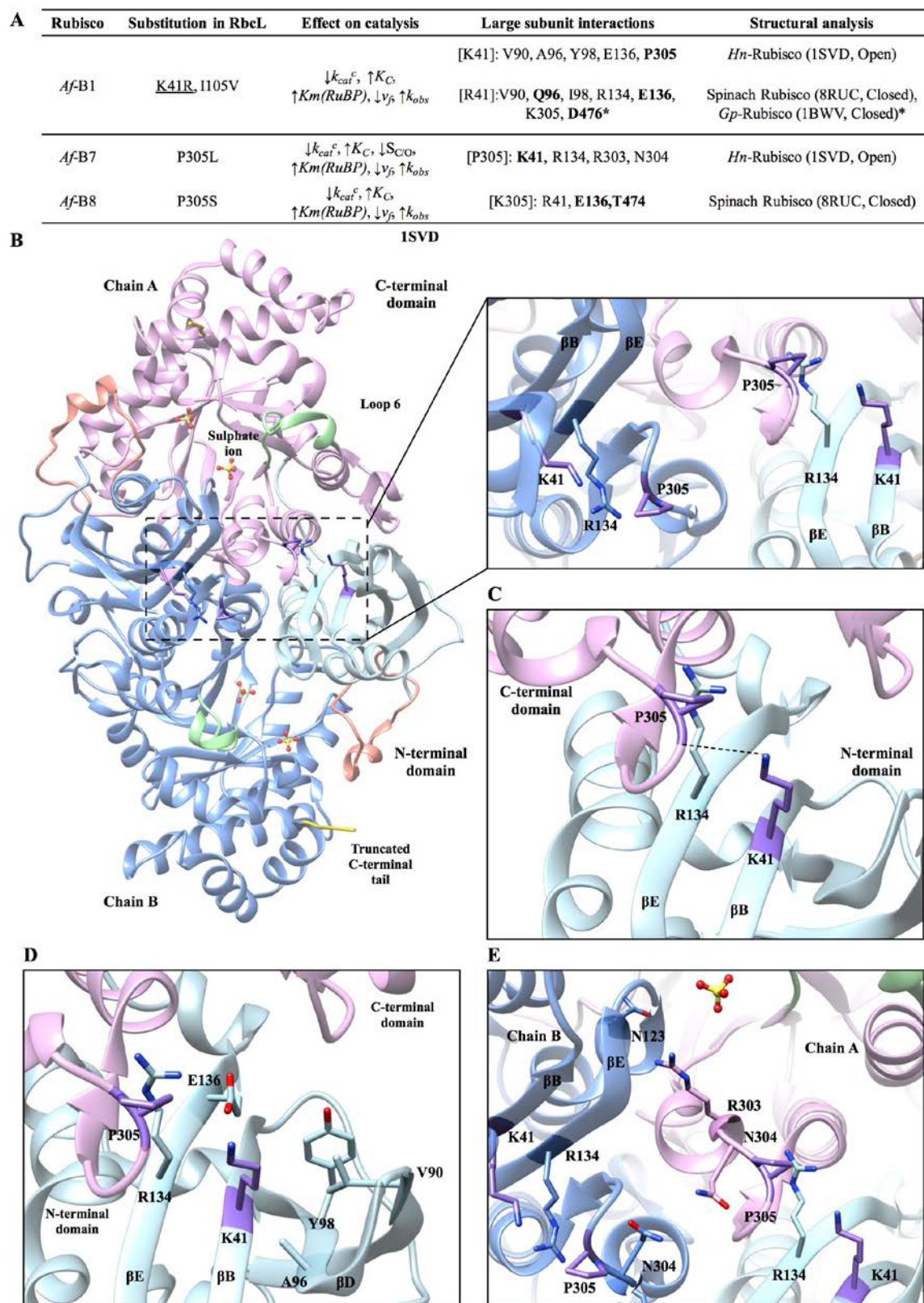


Figure 5-15: Position and interactions of Lys41 and Pro305 in *Hn*-Rubisco.

(A) Summary of the effect of substitution of Lys41 and Pro305 and their interactions with other large subunit residues. At position 41, lysine and arginine are present in Spinach and *Gp*-Rubisco respectively while at position 305, arginine and proline are found in Spinach and *Gp*-Rubisco respectively. Structural analysis of these residues is demonstrated with *Gp*-Rubisco (this figure) and Spinach Rubisco (Figure 5-16). (B) Overview of a single large subunit dimer from *Hn*-Rubisco (1SVD) where the active site is in the open state. In the close-up view, Lys41 and Pro305 are close in space to each other and Arg134 which resides in the space between these two residues (right panel). Pro305 is also located near the intra-dimer interface where the two large subunits are in contact. (C) Lys41 and Pro305 are located at the interface between the N-terminal and C-terminal domain which are colored sky blue and plum respectively. Carbonyl oxygen of Pro305 forms a hydrogen bond to the amine side chain of K41 with a distance of 2.60Å. (D) Lys41 is in van der Waals contact with Arg134 and Glu136 of β -strand E, V90, A96 and Tyr98 of β -strand D. (E) Pro305 is located on the loop connecting α -helix F and β -strand F where it is in van der Waals contact with Arg134, Arg303 and Asn304. Arg303 is close in space to the amino-end β -strand E of the N-terminal domain from the adjacent large subunit (Chain B) where the conserved active site residue, Asn123 resides in the preceding loop. As *Hn*-Rubisco is in the open state, the C-terminal tail was not resolved and it is also truncated.

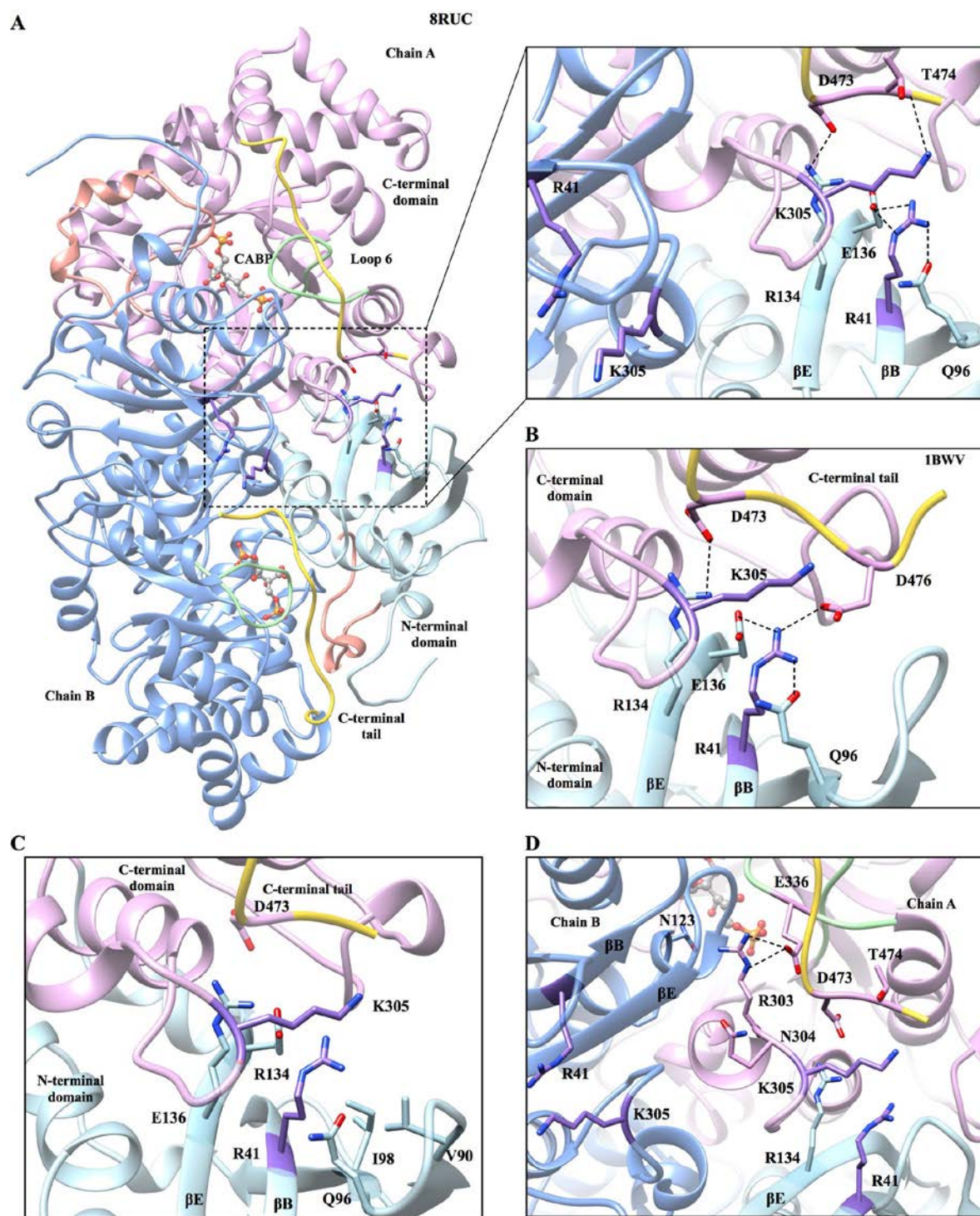


Figure 5-16: Position and interactions of Arg41 and Lys305 in Spinach Rubisco.

(A) An overview of position and interactions of Arg41 and Lys305 in large subunit dimer from Spinach Rubisco (8RUC) where the active site is in the closed state with CABP bound. Arg41 and

Lys305 are located at the interface between the N-terminal and C-terminal domain of a single large subunit and close in space to the end of the C-terminal tail. The polar contacts (Salt bridge or hydrogen bond) formed by Arg41 and Lys305 are shown in the zoom-in view (right panel). Arg41's side chain forms polar contacts with side chains from Gln96 and Glu136 (two contacts) at a distance of 3.06Å, 2.90Å, and 2.92Å respectively. Lys305 forms a hydrogen bond using its amino side chain to the carbonyl oxygen of Thr474 (Å). Thr474 is located near the end of C-terminus and sit next to the highly conserved Asp473. Asp473 forms a salt bridge with Arg134 of β -strand E (2.95Å) and this interaction is via their charged side chains. (B) Arg41's side chain form multiple polar contacts with side chains of Gln96, Asp136, and Glu476 with a distance of 2.94Å, 2.89Å, and 2.85Å respectively. This was shown using *G. partita* Rubisco (1BWV, closed state) which has a longer C-terminal tail compared to Spinach and *Af*-Rubisco and more residues on the C-terminus are resolved. The salt bridge between Arg134 and Glu473 is also shown. (C) Zoom in view of Arg41 and Lys305's interactions with neighboring residues and C-terminal tail. Arg41 is in van der Waals contact with Val90, Ile98, and Arg134 of the N-terminal domain as well as Lys305 of C-terminal domain. Whereas, Lys305 is van der Waals contact with Glu136 and Arg41. (D) Arg303's side chain forms two polar contacts with the side chain of Glu336 of loop 6 with a distance of 3.15Å and 2.95Å respectively. Glu336's side chain also forms a hydrogen bond with the main chain amide group of Asp473 with a distance of 2.95Å.

5.3.3.3 Thr392 and Glu438 are potentially a pair of interacting residues

Thr392 (T392) and Glu438 (E438) are a potential pair of interacting residues that are located near the surface of Rubisco. They are also close in space with each other at the interface between α -helix G and α -helix 7. Thr392 resides near the end of α -helix 7 of α/β barrel while Glu438 resides within the α -helix G in the C-terminal α -helical domain. Sequence conservation of these two residues at position 392 and 438 are low among different species (Figure 5-17A, B).

5.3.3.3.1 Thr392 substitution could affect inhibitor binding via loop 7

Thr392 is located on α -helix 7 where the preceding loop 7 contains P1-binding site residue, Gly381. Residue 392 is poorly conserved and multiple residues are found in this position with glutamate and aspartate being the most common. In *Af*-Rubisco, Spinach Rubisco and *Hn*-Rubisco, threonine, glutamate, and asparagine are present respectively (Figure 5-7, Figure 5-17A). Structural analysis of both Spinach and *Hn*-Rubisco show that both Asn392 and Glu392 are in van der Waals contact with the same residues from α -helix 7 (Pro338) and α -helix G (Ile473/ Leu473). However, Asn392 and Glu392 form polar contacts with different residues via their side chains. Asn392 forms hydrogen bond with Glu438 of α -

helix G while Glu392 forms salt bridge with Arg339 of α -helix 6 (Figure 5-15C, D, E). As α -helix G does not participate in the movements during the transition from open-to-closed state, the interactions from structural analysis of the open state *Hn*-Rubisco is likely maintained in the closed state as well. Arg339 is the first residue of α -helix 6 and it is also located immediately after Glu338 of loop 6 which rotates to allow loop 6 closure. Three residues are commonly found at position 339 with arginine and lysine being present in green-type Rubisco while proline is conserved in red-type Rubisco (Figure 5-7). Previous study by Karkehabadi et al. (2007) reported that the main chain C α atom of Arg339 is in van der Waals contact with the side chain of Val331. In addition, structural analysis using Spinach Rubisco (8RUC) have shown that Arg339's side chain is in van der Waals contact with Ile393 while Arg339's main chain amide forms a hydrogen bond with the carboxyl group of Leu343 which contributes to the hydrogen bond network in α -helix 6.

In *Af*-Rubisco, Thr392 is present instead of Glu392 or Asn392. Thr392's side chain contains a hydroxyl group which can donate and accept protons for hydrogen bonding simultaneously (Betts et al., 2003). Since the side chains of Arg339 and Glu438 are extended towards residue 392, it is possible that Thr392 forms hydrogen bond with both Arg339 and Glu438 via their side chains. Substitutions of Thr392 were found in two *Af*-Rubisco variants, *Af*-B9 and *Af*-B10, where threonine was substituted to either alanine or serine. Alanine and serine are present at this position in other species of Rubisco, with Ala392 found in the carboxysomal form of *Af*-Rubisco (form IAc Rubisco) and Ser392 present in *Rs*-Rubisco. The substitution to serine at position 392 may conserve the hydrogen bonds as serine also contains a hydroxyl group. In contrast, the substitution to alanine would disrupt the hydrogen bond formation with Arg339 and Glu438 as it lacks hydroxyl side chain. This substitution may result in changes to the local conformational around these few α -helix (6, 7, and G).

The T392A mutant characterization displayed an increased rate of inhibitor release with trade-off with decreased carboxylation rate and affinity for substrates, CO₂ and RuBP (Figure 5-17C). The effect of this substitution is likely mediated through the Gly391 of the preceding loop 7 and potentially loop 6 and α -helix 6. However, the specificity of T392A mutants was similar to wild type and this suggested that Lys334 of loop 6 has not been

displaced. Instead, subtle changes to the local conformation to loop 6 or α -helix 6 may contribute to the changes in kinetics. In addition, the overall effects on catalysis for T392A mutant was similar with that observed in the E438G mutant. Therefore, suggesting the importance of interaction between Thr392 and Glu438 as it contributes to the rigidity surrounding the active site.

5.3.3.3.2 Glu438 substitution may affect the position of loop 7 and loop 8

Glu438 resides near the amino-end of α -helix G which is part of the C-terminal α -helical domain. Sequence conservation is low at position 438 with glutamate, alanine, and valine being the most commonly present (Figure 5-17B). Structural analysis of *Hn*-Rubisco shows that Glu438 forms hydrogen bond with Asn392 of α -helix 7 via their side chain and Glu438 is also in van der Waals contact with Ala389 of the same α -helix 7. Thr392 is present in *Af*-Rubisco instead of asparagine and the hydrogen bond with Glu438 is likely conserved as Thr392 contains a hydroxyl group.

Substitution of Glu438 to either alanine or glycine was observed in *Af*-B11 and *Af*-B12 respectively. Both substitutions would disrupt the hydrogen bond with Thr392 as both alanine and glycine lack polar side for hydrogen bonding. The loss of the hydrogen bond with Thr392 may displace the preceding loop 7. In addition, the substitution to glycine may increase the flexibility of α -helix G and potentially the previous α -helix 8 as glycine lacks a side chain. The E438G mutant was characterized and the substitution had a greater effect on the carboxylation rate and k_{obs} as compared to the T392A mutant. There was a ~7-fold increase in the rate of inhibitor released observed in E438G mutant compared to the ~2-fold increase in T392A mutant. This difference in the rate of inhibitor release could be attributed to the increased in the flexibility of α -helix 8 which may alter the position of loop 8 residues. Loop 8 also contained two P1-binding site residues, Gly403 and Gly404. Therefore, the effects E438G substitution is likely mediated through three of the P1-binding site residues from loop 7 and loop 8.

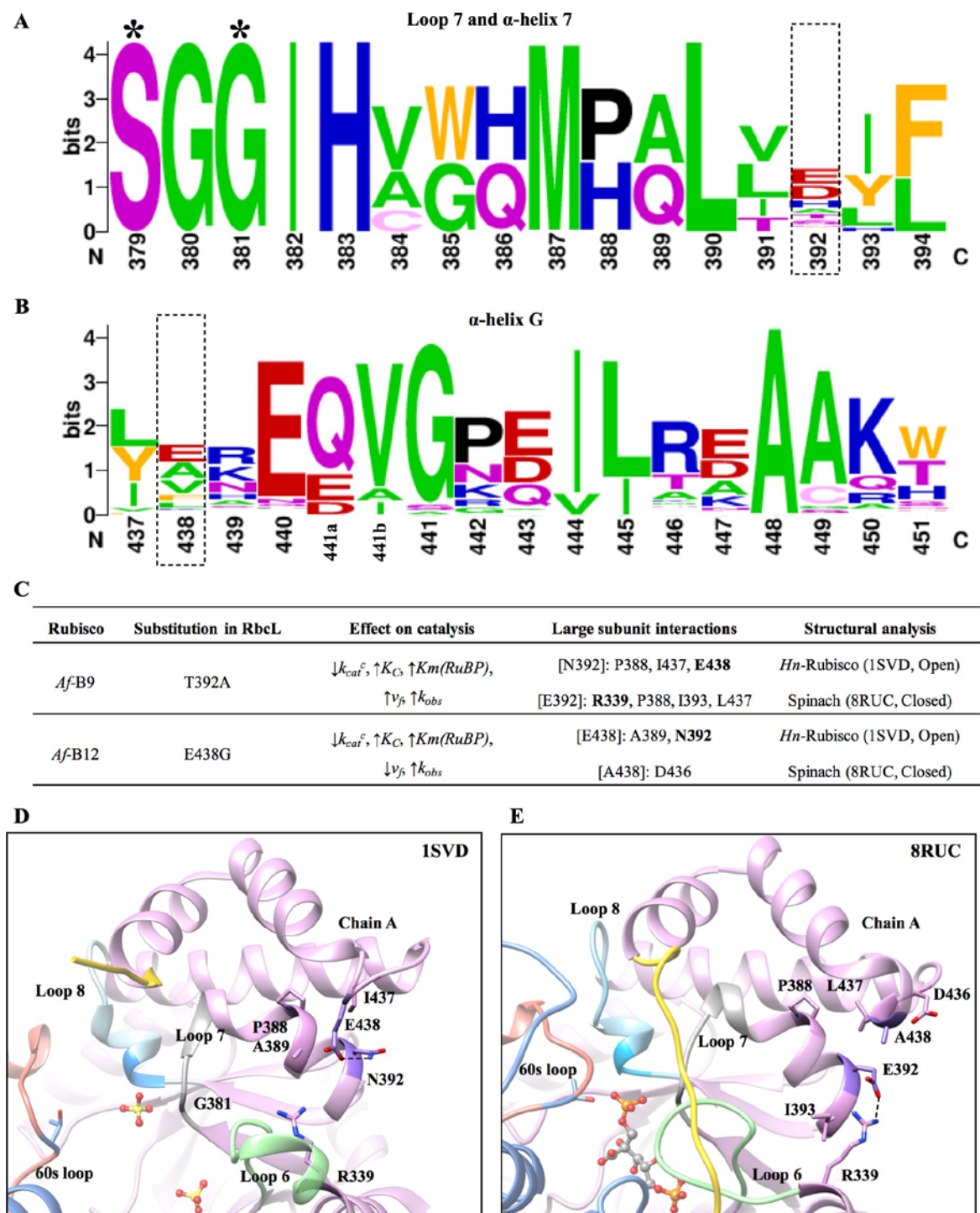


Figure 5-17: Structural analysis of residues at position 392 and 438 of the C-terminal domain.

(A) Sequence conservation of residues from loop 7 (residues 379- 386) to α -helix 7 (residue 387- 394). Ser379 and Gly381 are strictly conserved active site residues that bind to substrate while residues at position 392 are not conserved as shown by the short stack. (B) Sequence conservation

of residues from α -helix G of the C-terminal α -helical domain. Residues at position 438 are not conserved while residues at position 441a and 441b are not commonly found in most Rubisco sequences. (C) Summary of the changes in catalytic properties of T392A and E438G substitution in *Af*-Rubisco and their interacting partners. Thr392 is not conserved in the Rubiscos used for analysis and the residues interacting that at position 392 for the different Rubisco are shown in the brackets. (D) Zoom in view of the interactions of residues at position 392 and 438 in *Hn*-Rubisco which is in the open state. Glu438 and Asn392 form a polar contact (2.77Å) with their side chains. Both residues are in van der Waals contact with Ala389 of α -helix 7. Asn392 is also in van der Waals contact with Pro388 and Ile437 which resides next to Glu438. Arg339 of loop 6 is also shown and loop 6 is retracted away from the active site. (F) Interactions of Glu392 and Ala438 in Spinach Rubisco where the active site is in the closed state. Glu329 forms a salt bridge with Arg339 of loop 6 via their side chains with a distance of 2.80Å. In addition to the salt bridge, Glu329 is also in van der Waals contact with Ile393 and Pro338 of the same α -helix 7. Ala438 of α -helix G is in van der Waals contact with Asp436 and Leu437.

5.3.4 Mutations affecting the dimer-dimer interface

Another cluster of mutations was found at the dimer-dimer interface where there are multiple inter-subunit salt bridges and a major hydrophobic patch consisting of Leu105, Val142, Ala143 and Val369 (Knight et al., 1990) (Figure 5-19). The mutations were Phe220 (F220) from *Af*-B6, Leu105 (L105) from *Rs*-R18 and *Rs*-R19. In contrast to many of the previously selected residues, these residues at the dimer-dimer interface are not highly conserved at their respective positions (Figure 5-18B, C, D).



Figure 5-18: The conservation and effect of substituting residues from the dimer-dimer interface.

(A) Summary of the changes in kinetic properties of variants with residues found at the dimer-dimer interface. (B) Sequence conservation of residues from the loop between β -strand C and α -helix C (β C- α C loop). Leucine and isoleucine are two of the commonly present residues at this position. (C) Sequence conservation of residues from the beginning of α -helix 2 where phenylalanine is most commonly found at position 220. ^ The fold increase in k_{obs}^E for L105Q is inconclusive as the standard deviation is large and requires further studies to confirm. * The interaction between Phe220 and Arg65 is a π -cation interaction.

5.3.4.1 Substitution at residue 105 may influence inhibitor release indirectly through the β C- α C loop residues

Leu105 resides at the loop between β -strand C and α -helix C (β C- α C loop). Leu105 is not highly conserved at position 105 with leucine and isoleucine being the two more commonly present residues (Figure 5-18B). Leu105 is part of the hydrophobic patch along with other three hydrophobic residues at the dimer-dimer interface near the surface of Rubisco (Figure 5-19C). In this hydrophobic patch, two sets of these four residues are contributed by large subunit from adjacent dimers and these residues are all close in space to each other. Structural analysis of Spinach Rubisco (8RUC, closed state) showed that Leu105 is also in van der Waals contact with Thr34, Tyr144 and Ala143 of the same large subunit as well as Lys146 and Val369 of the adjacent large subunit. All these residues, except Thr34, are also involved in hydrophobic interactions with Leu105 via their methyl groups or hydrocarbon chain (Figure 5-19D).

Substitution of residue 105 had been selected in both *Af*-Rubisco and *Rs*-Rubisco, isoleucine and leucine are positioned in these Rubiscos respectively. Isoleucine and leucine are both hydrophobic residues with the same length of hydrocarbon chain. It is highly likely that isoleucine would maintain the interactions Leu105 has and also be part of the hydrophobic patch. The I105V mutation was found in *Af*-B1 variant along with K41R substitution which has been discussed above at section 5.3.3.2.1. The substitution to valine may affect the packing at the dimer-dimer interface as valine is a smaller amino acid with a shorter side chain compared to isoleucine.

The substitution of Leu105 to either methionine or glutamine was found in *Rs*-R18 and *Rs*-R19 respectively. Among these two mutants, only L105Q substitution had an observable phenotype of an increased in the rate of inhibitor release. However, the standard deviation for L105Q's k_{obs} value is large, suggesting that further studies would be required to validate the effect of L105Q substitutions on inhibitor release. It is likely that L105M substitution could maintain the hydrophobic interactions as methionine also contains a long hydrocarbon chain. Nonetheless, the presence of the sulfur atom increased the length of the methionine's side chain which would add bulk towards the dimer-dimer interface. This

L105M substitution may result in changes to the local conformational and indirectly affecting catalysis as shown by the increase in k_{cat}^c and K_C . In contrast, the L105Q mutant overall presented with similar kinetics as the wild type except there was an increase in k_{obs} . Glutamine has a polar amide group which may disturb the hydrophobic interactions at the dimer-dimer interface and is likely to affect the local conformation.

The effect of substitutions to Leu105 or Ile105 is potentially mediated through other residues of the β C- α C loop. Residues 107 to 112 of the β C- α C loop contains many highly conserved residues which participate in inter-subunit interaction in a single L₂ dimer. Many of these interactions involve residues near the end of loop 2 (residues 207 to 213) which also includes a salt bridge between Glu110 and Arg213 (Knight et al., 1990) (Figure 5-19E). Loop 2 contains many catalytically essential residues which also contribute to the metal binding site (Table 5-4). Leu107 of the β C/ α C loop is in van der Waals contact with two highly conserved 60s loop residue (Ser62 and Tyr80) and Leu178 of loop 1. The changes in the local conformation brought about by Leu105 or Ile105 substitution may influence the position of Leu107 or Glu110 such that there could be a domino effect whereby positions of residues from the 60s loop, loop 1 or loop 2 are also affected. As discussed earlier in section 5.3.1, residues from 60s loop and loop 2 contribute towards the substrate binding pocket. The changes to position of these residues would affect the affinity of inhibitor binding and influence the rate of inhibitor release. Therefore, substitutions at residue 105 are likely to mediate inhibitor release in an indirect manner via the β C- α C loop.

5.3.4.2 The effect for F220L substitution might be mediated through long distance to the active site

Phe220 resides near the middle of α -helix 2 of the α/β barrel domain of the large subunit and α -helix 2 is preceded by loop 2 which contains multiple conserved active site residue. Phe220 is located between two interfaces, the first near the top and bottom interface between two large subunit dimers and secondly between large and small subunit (Figure 5-19A, B). Phenylalanine and asparagine are commonly found at position 220 (Figure 5-18D). Structural analysis with Spinach Rubisco revealed that Phe220 is in van der Waals

contact with multiple residues from both large and small subunits (Figure 5-18A, Figure 5-19F). The side chain of Phe220 also participates in hydrophobic interaction with the hydrocarbon side chains of Asp160, Lys161, Ala182, and Lys183. Phe220 is also in van der Waals contact with the hydroxyl group of Tyr66 (Tyr61 in *Af*-Rubisco) of the small subunit. In addition, Phe220's aromatic ring makes π -cation interactions with the positively charged side chain of Arg65 (Asn60 in *Af*-Rubisco) from the small subunit. It is possible for Phe220 to make π -cation interactions with asparagine as it contains an amide side chain.

The interacting partners of Phe220 also form polar contacts with each other. For instance, the hydrogen bond between the hydroxyl side chain of Tyr66 with the carbonyl oxygen of Lys183 as well as an inter-dimer salt bridge between Lys161 and Glu216 (Knight et al., 1990). The side chain of Arg65 also forms multiple polar contacts. These polar contacts include with the carbonyl oxygen of Gly60 from the small subunit, carbonyl oxygen of Lys161 from the large subunit and side chain of Glu223 from the large subunit (Knight et al., 1990). Among these interacting residues of Phe220, only the two small subunit residues, Arg65 and Tyr66 have been previously characterized using site-directed mutagenesis in *Chlamydomonas* Rubisco. Both residues are located near the carbonyl end of the highly variable β A- β B loop (Figure 5-20). *Chlamydomonas* Rubisco mutants containing either R65A or Y66A had exhibited similar trends with reductions in various catalytic properties and the holoenzyme thermal stability (Esquivel et al., 2006; Spreitzer et al., 2001). These findings and many others have all demonstrated β A- β B loop's role in influencing Rubisco's catalytic activity and thermal stability (Esquivel et al., 2006; Karkehabadi et al., 2005; Spreitzer, 2003; Spreitzer et al., 2001).

Rubisco variants with F220L substitution were repeatedly isolated during screening. Although leucine has a bulky and non-polar side chain like phenylalanine, it lacks an aromatic ring (Betts et al., 2003). Some of the interactions may be affected such as the π -cation interactions with Arg65 and there could be changes in the local conformational around the dimer-dimer interface and interface between large and small subunit. It is possible that lack of aromatic ring could lead to an alteration in the positions of Arg65, Tyr66, and potentially the β A- β B loop. In addition, the preceding loop 2 may be influenced by this F220L substitution. Loop 2 contains active site residues (Lys201, Asp203, and

Glu204) that form the metal binding site for substrate binding and alteration in loop 2 could affect the binding of the incoming sugar phosphate as well as the catalytic properties. Therefore, the effect of F220L substitution is potentially mediated to the active site via loop 2 and β A- β B loop which lead to the kinetic changes in k_{cat}^C , K_C and k_{obs} .

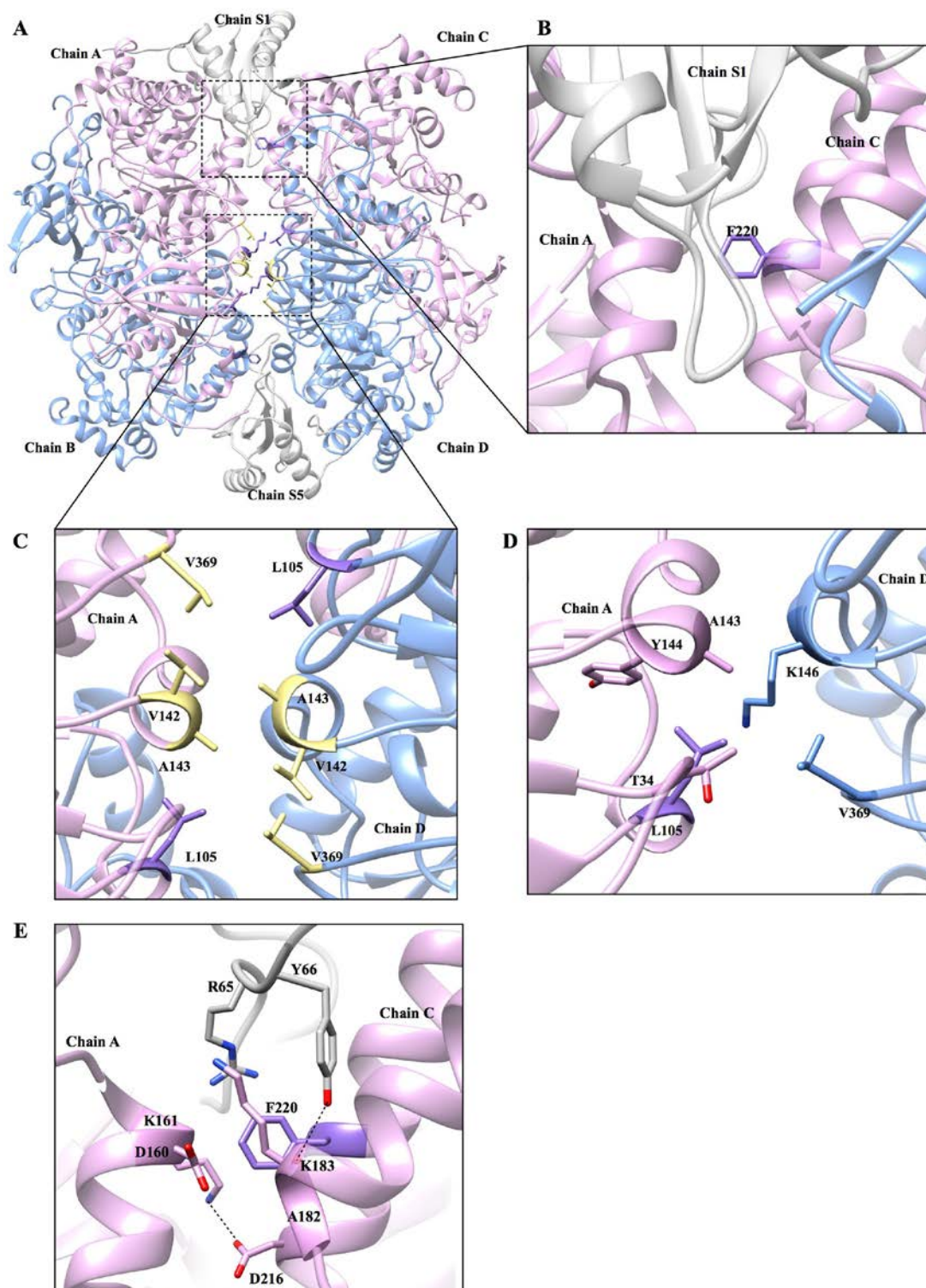


Figure 5-19: Interaction and position of selected residues at the large subunit dimer-dimer interface.

(A) Overview of the selected residues located at the dimer-dimer interface of Spinach Rubisco (8RUC). Small subunits (chain S1 and S5) interacting with the large subunits (Chain A-D) at the interface were also shown. (B) A close-up view of the interface between large subunit and small subunit where Phe220 resides at the carbonyl end of α -helix 2. (C) A close-up view of the hydrophobic patch within the dimer-dimer dyad. This hydrophobic patch consisted of two sets of four different residues; Leu105, Val142, Ala143, and Val369, which were contributed by large subunit of each dimer. (D) Leu105 of the hydrophobic path is in van der Waals contact with multiple residues. It is in contact with Thr34, Tyr144, and Ala143 of the same large subunit and Lys146 and Val369 of the adjacent large subunit. (E) Phe220 interacts with multiple residues from different subunit. Phe220 is in van der Waals contact with residues from same large subunit (Ala182, Lys183, and Asp233), adjacent large subunit (Asp160 and Lys161) and small subunit (Tyr66). Phe220 is also interacting with Arg65 of small subunit via π -cation interaction. Tyr66 of small subunit also forms a hydrogen bond via its hydroxyl side chain to the carbonyl oxygen of Lys183 with a distance of 2.66Å. Lys161 and Glu216 forms inter-dimer salt bridge via their charge side chains with a distance of 2.91Å. Residues involved in the hydrophobic patch are color in yellow as shown in panel A and C. The central solvent channel of Rubisco is behind the displayed dimer-dimer interface.

5.3.5 Small subunit mutation clustering around β -strand D affects the rate of inhibitor release

The Rubisco small subunit has lower sequence identity as compared to the large subunit and only 10 highly conserved residues in the small subunit. Despite the low sequence homology, the overall structure feature of small subunit is conserved (Spreitzer, 2003). The green-type and red-type Rubisco small subunit can be differentiated by the presence and absence of the additional β EF hairpin (β -strand E and β -strand F). Only red-type Rubisco small subunit contains this additional β EF hairpin and this hairpin has been demonstrated to be critical for red-type Rubisco assembly (Joshi et al., 2015). During the RDE screening, there were multiple small subunit mutations obtained. Among the variants that were characterized, those with small subunit mutations were clustered around β -strand D. In this section, the numbering of small subunit residues is according to their respective wild type Rubisco.

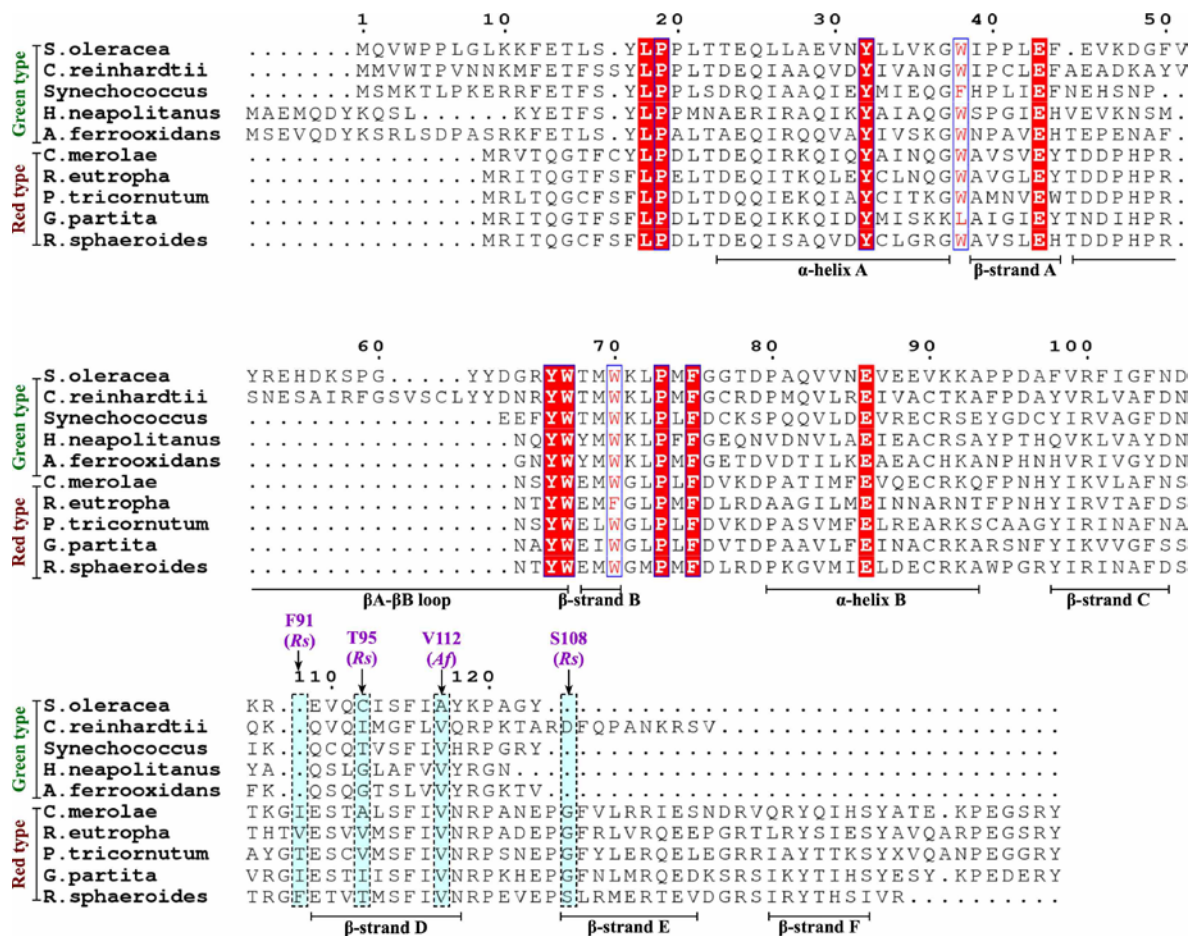


Figure 5-20: Sequence alignment of selected Rubisco small subunits.

Sequence alignment of selected Rubisco small subunit from green-type Rubisco; *S. oleracea* (Spinach), *C. reinhardtii* (P00873), *Synechococcus* PCC6301 (P04716), *H. neapolitanus* (P45686) and *A. ferrooxidans* (B7JA25); and red-type Rubisco; *C. merolae* (O22024), *G. partita* (P23756), *P. tricornutum*; (Q5PYA1), *R. eutropha* (P09658) and *R. sphaeroides* (P27998). Residues that conserved were in white fonts with red background while residues with similar properties were in red font. Residues selected from screening were indicated with purple labels and residues that are not conserved had light blue background. The numbering of the residues and secondary structure of α -helices and β -strands were according to Spinach Rubisco as denoted by Knight et al. (1990). The additional β -strand E and F were denoted as previously shown by Genkov et al. (2009). The sequence alignments were performed using MultAlin (<http://multalin.toulouse.inra.fr/multalin/>) and alignment image was constructed with ESPript (Robert et al., 2014).

5.3.5.1 Substitutions of Phe91 and Thr95 in *Rs*-Rubisco may influence rate of inhibitor release via the 60s loop

Mutation of two small subunit residues, Phe91 and Thr95, around the β -strand D was observed in three *Rs*-Rubisco variants. Phe91 is located at the carboxyl end of the loop connecting β -strand C and β -strand D (β C- β D loop) while Thr95 resides in the middle of β -strand D. The β C- β D loop sits at the dimer-dimer interface and in red-type Rubisco, this loop is slightly longer than green-type counterpart due to the presence of two additional residues (residues 90 and 91) (Spreitzer, 2003). Sequence conservation of these two residues (91 and 95) are low and in *Gp*-Rubisco, isoleucine is present at these two positions instead (Figure 5-20).

Structural analysis of *Gp*-Rubisco shows that Ile91 and Ile95 are in van der Waals contacts with each other as well as multiple residues from the β C- β D loop and β -strand D (Figure 5-21C). Ile91 is also in van der Waals contact with large subunit residue, Lys373 which is located on the loop connecting β -strand H and β -strand 7 (β H- β 7 loop, Figure 5-7). In *Rs*-Rubisco, Phe91 is present instead of isoleucine and the aromatic ring of phenylamine could possibly form π -cation interaction with Lys373. Ile95 is also van der Waals contact with residues from the α -helix A of small subunit. Many residues from α -helix A also interact extensively with α -helix 8 of the adjacent large subunit (Chain A)(Knight et al., 1990). In *Rs*-Rubisco, Thr95 is present instead of isoleucine and threonine could possibly, form polar contacts with nearby residues.

Glu92 is another residue in the β -strand D and it is situated next to Ile91. Glu92 forms a hydrogen bond with Tyr80 of the 60s loop via their side chains (Figure 5-21C). Both Glu92 of the small subunit and Tyr80 of the large subunit are highly conserved in red-type Rubisco. The conservation of Glu92 suggests that these interactions may play a role in Rubisco function and assembly (Figure 5-9B, Figure 5-20).

Mutations of Phe91 and Thr95 were found either individually or together in the different *Rs*-Rubisco variants. *Rs*-R11 and *Rs*-R12 contained F91L and T95A substitutions respectively, while *Rs*-R10 contained a combination of F91I, T95S, and S108C mutations.

The presence of variants containing only single substitution suggests that mutations of F91L or T95A are likely to cause more changes to the local conformation as compared to F91I or T95S respectively. Substitution of Phe91 to either isoleucine or leucine could potentially affect the interactions involving phenylalanine's aromatic ring, in particular, the π -cation interaction with Lys373 from the adjacent large subunit. The loss of inter-subunit interaction could cause conformational changes that affect Rubisco's catalytic ability, possibly via loop 7 which is located right after β -strand 7 and close to Lys373. In the case of substitutions involving Thr95 to either serine or alanine, the impact of T95S is likely less than T95A. As serine has hydroxyl side chain and therefore the capacity to maintain hydrogen bonds that Thr95 may have. In contrast, alanine lacks the hydroxyl group which would disturb hydrogen bond formation. In addition, the shorter side chain of alanine would disrupt interactions with residues 28, 30 and 91 as these van der Waals contact involved the C γ atoms in Ile95. The substitutions of Phe91 and Thr95 are likely to affect the local conformation such that the position of Glu92 might be influenced and this would in turn affect its interaction with Tyr80 of the 60s loop. Therefore, the effects of substitutions, Phe91 and Thr95, are likely to affect the active site in an indirect manner via the 60s loop and loop 7. The influence of these substitutions on 60s loop and loop 7 may alter the binding affinity of substrate which is reflected in the high k_{obs} and high $K_m(RuBP)$.

5.3.5.2 V112 substitution may influence β C- β D loop's position

Val112 is at the core of the small subunit and situated near the end of β -strand D. It is in van der Waals contact with Val88 of the adjacent β -strand C. Substitution of Val112 to either alanine or glycine was observed in *Af*-B21 and *Af*-B22 variant respectively. Alanine has a shorter side chain while glycine does not have a side chain, therefore the substitution may alter the local conformation such that position of β -strand D and the preceding β C- β D loop could be affected. It is possible that the effect of Val112 substitution may influence catalysis in a similar mechanism as those described for Phe91 and Thr95.

In the Rubisco variants containing Val112 substitutions, additional mutations were observed in the large subunit. And in the future, the mutation should, therefore, be characterized in isolation.

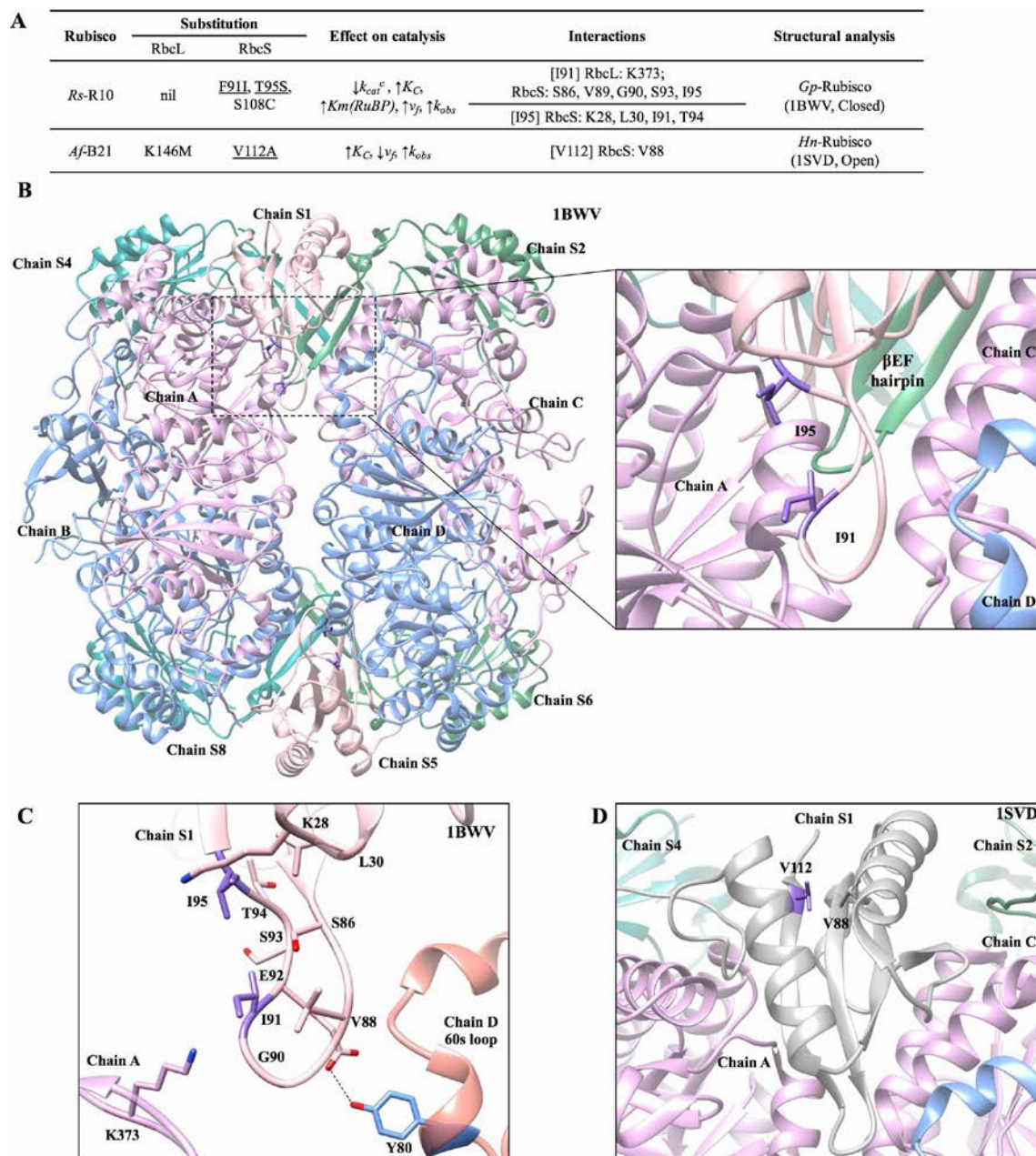


Figure 5-21: Position and interaction of small subunit residues.

(A) Summary of effect of substitutions of small subunit residues and interacting partners of these residues. (B) An overview of two large subunit dimers with small subunits from *Gp*-Rubisco. The close-up view of Ile91 and Ile95 from the small subunit (Chain S1) which resides in the interface between two dimers (right panel). Ile91 and Ile95 are in van der Waals contacts with each other. (C) Van der Waals contact of Ile91 and Ile95 with neighboring residues from small subunit and adjacent large subunit. Glu92 of the β -strand D forms a hydrogen bond with Tyr80 from the 60s loop of the

adjacent large subunit with a distance of 2.97Å. (D) The close-up view of two large subunit dimer with small subunits from *Hn*-Rubisco (1SVD) where Val112 resides in the β -strand D of small subunit. Val112 is van der Waals contact with Val88. The color scheme for the small subunits is as follow: Chain S1 and S5 are colored either pink for red-type Rubisco or grey for green-type, Chain S2 and S4 are colored cyan and green respectively. The central solvent channel is behind the dimer-dimer interface where the β EF hairpin is positioned.

5.4 Discussion

Our directed evolutionary screens of two bacterial form I Rubisco (*Rs*-Rubisco and *Af*-Rubisco) from both red and green lineage show that Rubisco could be evolved away from its dependency on activase. Wild type *Rs*-Rubisco and *Af*-Rubisco form stable inhibited complexes with sugar phosphate and subsequently require activase to remodel the inhibited complex to allow inhibitor release (Gibson et al., 1979; Mueller-Cajar et al., 2011; Tsai et al., 2015). When these two Rubisco were evolved in the absence of their activase in the RDE system, numerous mutations were obtained at different clusters on the large and small subunits. The biochemical characterization revealed that these mutant Rubiscos were functional and assembled successfully into L_8S_8 complexes. The presences of functional Rubisco complexes suggest that the mutations acquired were within a neutral range where Rubisco's functional state was maintained and nonsynonymous mutations that could destabilize Rubisco's assembly or impair Rubisco's function have been subjected to negative selection (Durao et al., 2015; Tokuriki et al., 2009b). Many evolved Rubisco variants also gained a new phenotype whereby they form no longer form a stable inhibited complex as demonstrated by the increase in k_{obs} . The reduced tendency to stable inhibited complexes was often coupled with a trade-off in kinetics. However, it was unexpected that the specificity of the majority of the variants remains relatively similar to their wild type counterparts. The similar specificity suggested that the mutant Rubisco were able to distinguish between CO_2 and O_2 with a similar level of competency as its wild type. It is likely that maintaining the specificity may confer survival advantage whereby there is sufficient RuBP turnover rate to alleviate the RuBP-toxicity. In addition, the solubility of these rubisco variants was slightly increased by up to ~40% and this could be happening to counter the reduced catalytic efficiency of these Rubisco variants.

The directed evolution of *Rs*-Rubisco in the presence of its activase was also performed and there was a lower yield of mutants and few reoccurring mutations compared to screening in the absence of activase. The difference in screening results suggest that the presence of activase limits the protein sequence space for Rubisco's evolution as there is a need to maintaining compatibility in between Rubisco and its activase. This is also consistent with the findings of a previous directed evolution study where *Syn*-Rubisco's evolvability was restricted in the presence of its assembly chaperone, RbcX (Durao et al., 2015). Only the *Rs*-Rubisco variant containing the I315T mutation in the large subunit was characterized to still form a stable inhibited complex. The I315T mutant had a modest reduction in its kinetics compared to wild type but there was a lower investment with ~43% in Rubisco content to support growth in RADE system. The lower Rubisco content suggests that the presence of activase reduced the amount of Rubisco required as the activase can constantly remodel the inhibited Rubisco complexes and allow Rubisco to catalyze RuBP carboxylation.

5.4.1 Loops involved in substrate binding and active site closure were influenced in Rubisco variants

The structural analysis of the different mutations which confer the phenotype of unstable inhibited complexes shows that these mutations were influencing loops that are either involved in substrate binding or loops that undergo movements during active site closure. As described in section 5.3.1, the substrate binding site contains residues from various loop of the α/β barrel (loop 2, 5, 7 and 8) and the 60s loop of the N-terminal domain. While the loops involved in active site closure are the loop near the N-terminus, 60s loop, loop 6 and C-terminus (Taylor et al., 1996b). The correct placement of these loops ensures that the substrate is orientated correctly for catalysis and that the active site is tightly closed and shield from the bulk solvent. Therefore, changes to the local conformation of these loops would allow a premature opening of the active site and substrates to rapidly dissociate.

In many of the variants, the mutated residues either influence the loops directly or indirectly via other residues that could interact with the loop. For instance, the mutations clustered around the 60s loop. On one hand, direct effects occur with residues such as Ser61, Thr71,

Thr75, and Arg81 as they are located on the 60s loop. In addition, mutation such as Tyr24 has a similar direct effect as it interacts with Thr63 of the 60s loop. On the other hand, the indirect effect happens with Phe91 and Thr95 of the small subunit. Both Phe91 and Thr95 could alter the position of neighbouring residue, Glu92. Glu92 interacts with Tyr80 of the 60s loop and mutations affecting the position of Glu92 could have an impact of the 60s loop as well. The effect of the various substitutions to the 60s loop is likely mediated through Thr65 as it is part of the P1-binding site. Other examples include mutation of C-terminus residues, Lys463 and Ile465, where a direct effect would occur to the C-terminus closure. While mutations of Lys41 and Pro305 which are located near the latch site of C-terminus were indirectly influencing C-terminus.

The structural analysis provided detailed insights into the mechanism that are affected by the respective mutation and allow future studies to be focus on characterizing the inhibitor binding properties of RuBP and the side reactions of these mutants. RuBP as an inhibitor is considered slow, tight-binding where the initial binding of RuBP to the active site is rapid and over time, there is conformational change to the active site and RuBP's binding becomes tight (Morrison, 1982). The inhibitor kinetics might be able to pinpoint the stage of RuBP's binding that these mutations are affecting. It is possible that mutations targeting the C-terminus will influence the latter stage of RuBP binding whereby the conformational changes were not complete such that RuBP can rapidly dissociate from the active site. Another aspect for future studies is the side reaction of Rubisco as Rubisco catalysis contains many misfired side reactions which could lead to self-inhibition. One of the misfired reactions produced pyruvate which could be utilized as a carbon source and confer survival fitness. It is possible that mutants that form unstable inhibited complexes would have also lessened sensitivity to these misfired products. Similar studies focusing on the inhibitor kinetics and side reactions have been previously carried out for tobacco and *Syn-Rubisco*. Thus, similar methodology can be used to study our mutants (Pearce, 2006; Pearce et al., 2003).

5.4.2 Why did nature evolve Rubiscos to form a tightly inhibited complex?

Our directed evolutionary screen shows that many of the Rubisco variants that no longer form tightly inhibited complexes contain only a single mutation arising from a single nucleotide change. We can hence speculate that in other Rubisco with activase, such as plant Rubisco, they can be easily evolved to form an unstable inhibited complex and relieve dependency from Rubisco activase. However, such plant Rubiscos have not been selected in nature. There are two possible reasons on why the tendency to form a stable inhibited complex appears to be under positive selection. Firstly, the requirement for activase's assistance in the removal of bound inhibitor from Rubisco has allowed activase to regulate Rubisco's activity (Bracher et al., 2017). Secondly, the inhibited Rubisco complex also protect Rubisco from proteolytic degradation which allows the regulation of Rubisco's degradation (Feller et al., 2008).

As the of Rubisco's activation state is maintained by its activase, the regulation of the activase is directly linked to the regulation of Rubisco. There are currently three types of activase known and they all require ATP for performing remodeling of inhibited Rubisco complexes (Bhat et al., 2017a). The divergence in the primary sequence of these three types of activases has suggested that they have emerged independently via convergent evolution. The regulation of green-type activases from plants have been well studied among the three types of activase. Higher plants express two quasi-identical isoforms of Rca (α and β), with the α -isoform possess a long C-terminal extension. These green-type Rcas are regulated by the intracellular ATP/ADP-ratio and also by the reversible formation of a disulfide-bond in the conserved cysteine residues on C-terminal extensions of the α -isoforms in some plants (Portis et al., 2008). During the daytime, photosynthesis is occurring, the activase would be functioning as the ATP/ADP ratio is high and the stroma would have a reducing environment. Rca could then facilitate removal of inhibitors such as RuBP and misfired products like XuBP and PDBP. This removal of inhibitor would ensure Rubisco's active site to be available for activation and subsequently carrying out carbon fixation using ATP and NADPH from the light reaction. In contrast, during the night time, Rca would be inactive due to low ATP/ADP ratio and the stroma would have an oxidizing environment. Rubisco is also inhibited by a night-time inhibitor, CA1P which is synthesized by the plants.

Therefore, in the absence of functional activase, these Rubiscos remained inhibited during night time and does not consume ATP or NADPH (Parry et al., 2008; Portis et al., 2008). As such, plant Rubisco's activity could be regulated by its activase in a light- and redox-dependent manner (Bhat et al., 2017a).

In contrast, the red-type Rubisco activase from *Rs*-Rubisco, RsRca, is regulated by RuBP. RsRca requires the presence of both ATP and RuBP to form its active hexameric complex. In the absence of RuBP, RsRca is inactive and forms a long fibrillary structure. This structure was suggested to be a storage form which regulates RsRca activity during low photosynthetic activity (Mueller-Cajar et al., 2011). Therefore, the production of RuBP during photosynthesis could enable the formation of functional activase which can subsequently regulate Rubisco activity. The third class of activase is the CbbQO activase which has been characterized in *Af*-Rubisco (Tsai et al., 2015). Although it is currently not known how CbbQO's activity is regulated, the ATPase activity of CbbQO has been shown to be stimulated in the presence of inhibited *Af*-Rubisco complex. It is likely that CbbQO's regulation is linked to the functional state of the Calvin cycle similar to that in the other two activase. The regulation of Rubisco by their activase ensures that Rubisco is only functional when carbon fixation is required. If Rubisco loses its propensity to form a stable inhibited complex, it could then carry out carbon fixation without regulation, leading to depletion of RuBP. Maintaining a high level of RuBP and other intermediates may be important to maintain flux through this autocatalytic cycle (Stitt et al., 2010). If intermediates drain from the Calvin cycle, it may be difficult to rapidly achieve a high flux state once ATP and NADPH become available again as a result of the light reactions.

Therefore, the loss of regulation of could have been selected against in organisms that express activase. In addition, the presence of activase may act as an evolutionary force to select against Rubisco evolution towards form an unstable inhibited complex. As demonstrated using the *Rs*-Rubisco variant containing an I315T mutation that has evolved in the presence of activase, there was lower investment in Rubisco content compared to those evolved to become independent of activase. The lower investment in Rubisco meant that resources can be better used in other areas which could confer additional survival

advantages. Therefore, nature has evolved Rubisco form stable inhibited complex to allow regulation of Rubisco activity via activase.

The second reason Rubisco have evolved to maintain a tight inhibited complex with sugar phosphate is that the inhibited Rubisco complex is protected against proteolytic degradation. Previous study by Khan et al. (1999) show that when Rubisco is bound by CA1P, it is protected from proteolytic degradation by both endogenous and exogenous proteases. In addition, activated Rubisco complexes with substrate bound under catalytic conditions have been shown to be protected from degradation (Houtz et al., 1991). In both cases, the Rubiscos are in the closed state where the flexible loops such as C-terminus and N-terminus are packed against the enzyme body. The isolation of large subunit that has missing parts of either C-terminus or N-terminus from chloroplast under different conditions suggested that Rubisco may have a different mechanism for degradation. The protection against degradation by the formation of a closed state would allow fine-tuning of Rubisco's degradation during periods of stress (Feller et al., 2008). In comparison, Rubiscos that do not form inhibited complexes and lack activase are often found in specialized micro compartments such as the carboxysome. *Syn*-Rubisco is a cyanobacterial Rubisco and localized in the carboxysome where CO₂ is concentrated for Rubisco to carry out CO₂ fixation. In the carboxysome, Rubisco is protected from performing oxygenation reaction with O₂ and protection is provided against degradation from protease (Codd et al., 1984). Therefore, with these protections provided in microcompartments, these Rubiscos would tend to not evolve to form stable inhibited complexes and subsequently not requiring activase as well.

5.5 Conclusions

Our direction evolutionary screens have shown that Rubisco can easily escape activase dependency. However, this escape from activase dependency did not occur during the course of evolution for many Rubiscos. The formation of an inhibited complex had advantages as it allows regulation of Rubisco's activity and degradation. This study also highlighted the various loops that are involved in active site closure and substrate binding. All of these structural components play a crucial role in ensuring the maintenance of a tight

and rigid active site in the closed state. These detailed insights about Rubisco structure-function relationship will be valuable for the engineering of more efficient Rubisco in the future. In addition, the establishment of the Rubisco and activase dependent (RADE) system has much potential to create compatibility between different types of Rubisco and activase. This would allow us to harness the benefits of superior properties of different activase. In the future, the carbon fixation machinery would likely consist of carefully selected modules that can maximize their performance at their local environment and these modules could also be altered to meet the challenging changes in the external environment (Mueller-Cajar et al., 2014).

6 DIRECTED EVOLUTION OF CHIMERIC PLANT/CYANOBACTERIAL RUBISCO IN *E. COLI*

Rubisco has been a long-standing target for improvement in attempts to enhance agricultural crop yield as its sluggish enzymatic activity is often the rate limiting step of the CBB cycle. Attempts to improve the higher plant enzyme's activity has been hindered by the inability to express these Rubiscos in the bacterial host such as *E. coli* where they tend to form insoluble aggregates (Cloney et al., 1993; Whitney et al., 2011). In contrast, other Rubiscos such as those from cyanobacteria can be expressed in *E. coli* and the presence of additional chaperonins like GroEL/ES and assembly factor, RbcX, can further enhance their expression. Plants and cyanobacteria both contain the form IB Rubisco and the difference in their solubility in *E. coli* has been suggested due to the requirement of extensive auxiliary factors for plant Rubisco's biogenesis (Bracher et al., 2017). In fact, in the past year, a seminal breakthrough has been achieved, where it was shown that the presence of specific molecular chaperones permits functional expression of higher plant Rubisco in *E. coli* (Aigner et al., 2017). In addition, the large and small subunits of plant Rubisco are also encoded in plastid and nucleus respectively. In comparison, the genes encoding for cyanobacterial Rubisco large and small are both in the same an operon. Despite having differences in the requirement for biogenesis, both plant and cyanobacteria Rubisco have high sequence identity with highly conserved active site residues as well conserved structure.

Many higher plants such as *Oryza sativa* (rice) has been extensively studied as it is an agriculturally and commercially important crop. *O. sativa* also serves as a primary source of calories for more than one-third of the world's population (Goff et al., 2002; Yu et al., 2002). *Os-Rubisco* contains Form IB Rubisco and sequence alignment of *Os-Rubisco* large subunit with *Synechococcus* PCC6301 Rubisco large subunit reveals 80% sequence identity with multiple regions of conserved amino acid (Figure 6-1A). In this chapter, we plan to investigate which region of the *Os-Rubisco* is responsible for its insolubility in *E. coli* by generating chimeric cyanobacterial/plant Rubisco. These chimeric Rubiscos will have higher sequence identity when compared to *Os-Rubisco* and these can be interpreted as a move in protein sequence space towards *Os-Rubisco* (Maynard Smith, 1970).

6.1 Chimeric *Syn*/*Os*-Rubisco

6.1.1 Synthesizing *Syn*-Rubisco and *Os*-Rubisco gene

To understand whether specific regions of the large subunit contribute towards the insolubility of higher plant Rubisco, *Syn*-Rubisco and *Os*-Rubisco genes were synthesized by Genscript (U.S.A.). For the large subunits, the codons of *Syn*-Rubisco were used for the conserved amino acids to minimize problems associated with eukaryotic codons. The synthesized Rubisco genes also had a number of internal restriction enzyme sites removed to facilitate cloning via the newly introduced restriction sites at the conserved regions (Figure 6-1A, C). The F345I mutation was introduced in the large subunits of both Rubisco as it was previously found to dramatically enhance the solubility of *Syn*-Rubisco in *E. coli* (Mueller-Cajar et al., 2008b). The large and small subunit genes of *Os*-Rubisco were positioned in an artificial operon using the 90-nucleotide intergenic region found in the cyanobacterial genome. The need for this sequence was due to the fact that the genes for large and small subunits were encoded in different genomes in the organism. As in all plants, the small subunit of *Os*-Rubisco is encoded by the nuclear genome while the large subunit is encoded by the plastid genome (Spreitzer, 2003; Whitney et al., 2011).

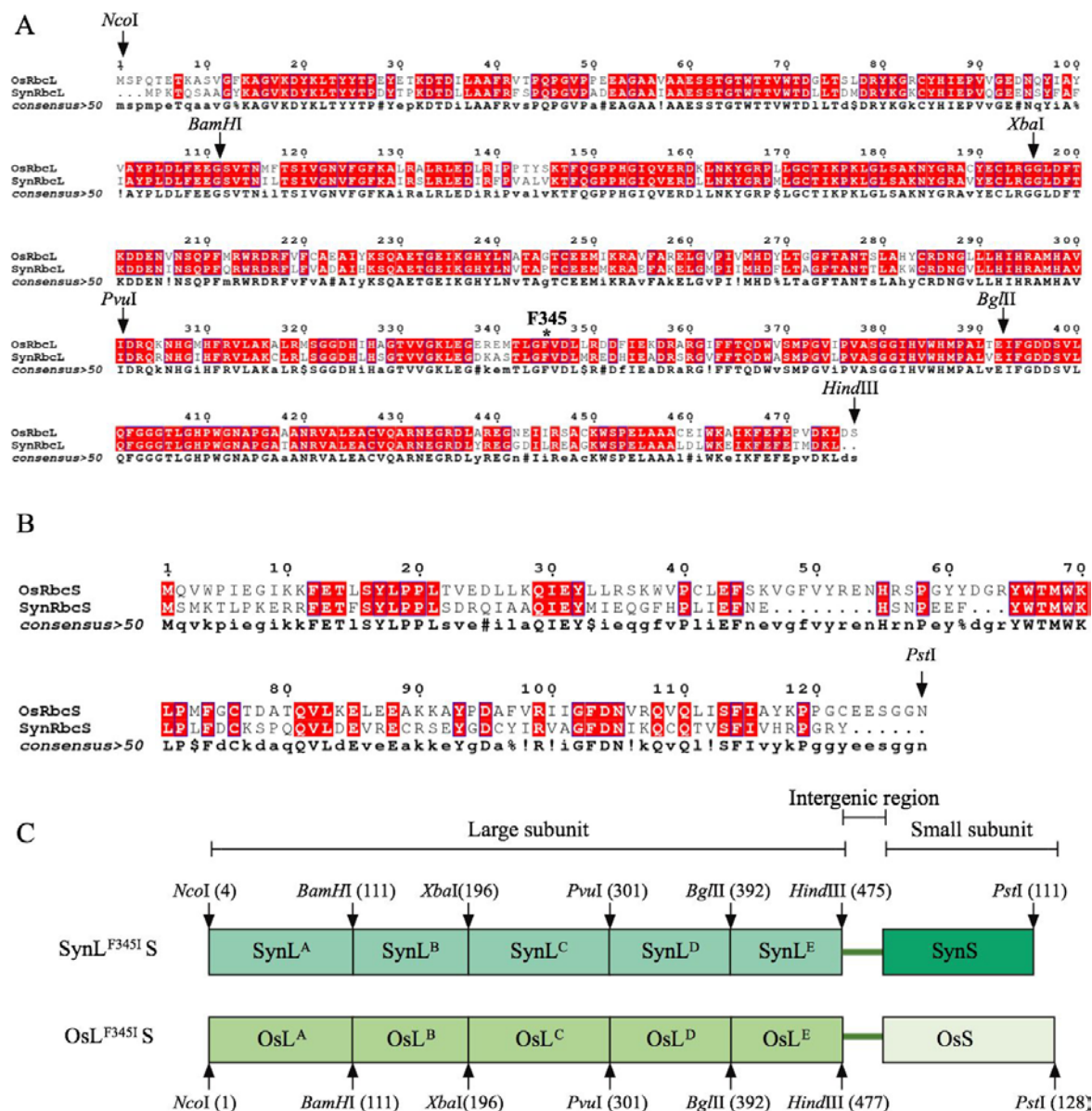


Figure 6-1: Sequence alignment and synthetic constructs encoding *Synechococcus* PCC6301 and *Oryza sativa* Rubisco.

(A) Sequence alignment of the large subunit of Syn-Rubisco and Os-Rubisco. (B) Sequence alignment of small subunit of Syn-Rubisco and Os-Rubisco. Residues that are conserved in these two Rubiscos are in white letters with a red background while those that are not conserved are in black letter with a white background. Symbols below letters in the consensus>50 such as #, %, !, meant that these residues from both species have similar chemical properties. The newly introduced internal restriction enzyme cut sites and the position of Phe345 are indicated with an arrow and an asterisk respectively. (C) A representative diagram of the synthesized Syn-Rubisco and Os-Rubisco constructs with an F345I mutation in the large subunit. The large subunit was divided into five fragments via six restriction enzyme sites with the internal restriction enzyme sites introduced at

conserved regions. The amino acid positions of the large subunit were labeled along with the restriction sites and the numbering is according to Spinach Rubisco. The small subunit numbering is according to the respective Rubisco. The sequence alignments were performed using MultAlin with ENDscript 3 (Corpet, 1988; Robert et al., 2014).

6.1.2 *Syn-Rubisco and Os-Rubisco presented with different solubilities in E. coli*

The synthesized Rubisco genes were cloned into the pTrcHisB vector and expressed to check for the expression of soluble protein. This was carried out in *E. coli* BL21 and in the presence and absence of overexpressed chaperonin GroEL/ES. GroEL/ES is a promiscuous folding chaperonin that assists in the folding of proteins in prokaryotes (Hartl et al., 2002). Previous studies revealed that folding of the Rubisco large subunit obligatorily requires the chaperonin GroEL/ES and the overexpression of GroEL/ES could improve the yield of the folded cyanobacterial Rubisco large subunit in *E. coli* (Emlyn-Jones et al., 2006; Goloubinoff et al., 1989; Liu et al., 2010). The expression vector, pTrcHisB contained the Trc promoter and protein expression was induced by IPTG. The GroEL/ES genes were under the regulation of BAD promoter in the vector pBAD33 (Ewalt et al., 1997; Guzman et al., 1995) and gene expression of GroEL/ES was induced by L-arabinose.

A representative protein expression analysis using *Syn-Rubisco* and *Os-Rubisco* is shown in Figure 6-2. Purified Rubiscos wild type *Synechococcus* PCC6301 and *O. sativa* were loaded along with the samples as control and marker proteins. This was confirmed with the pure Rubisco and degradation of the large subunit leads to the formation of smaller bands below the 52 kDa. Bands corresponding to each condition (induced and non-induced) were cropped from the western blot images and compiled into the final figure. The same approach was used for the chimeric Rubisco and for co-expression with GroEL/ES (Figure 6-2D). Western blot analysis showed that *Syn-Rubisco* with the F345I mutation ($\text{SynL}^{\text{F345I}}\text{S}$) was soluble while the corresponding *Os-Rubisco* mutant ($\text{OsL}^{\text{F345I}}\text{S}$) was not soluble. Both Rubisco exhibited some basal protein expression due to the leakiness of the Trc promoter. There was a slight faint band observed from the soluble extract for the induced sample of $\text{OsL}^{\text{F345I}}\text{S}$ and this likely correspond to unfolded large subunit being encapsulated in the GroEL/ES cavity. This phenomenon has been observed previously (Cloney et al., 1993) and can be verified by western blotting of PAGE run under non-denaturing conditions.

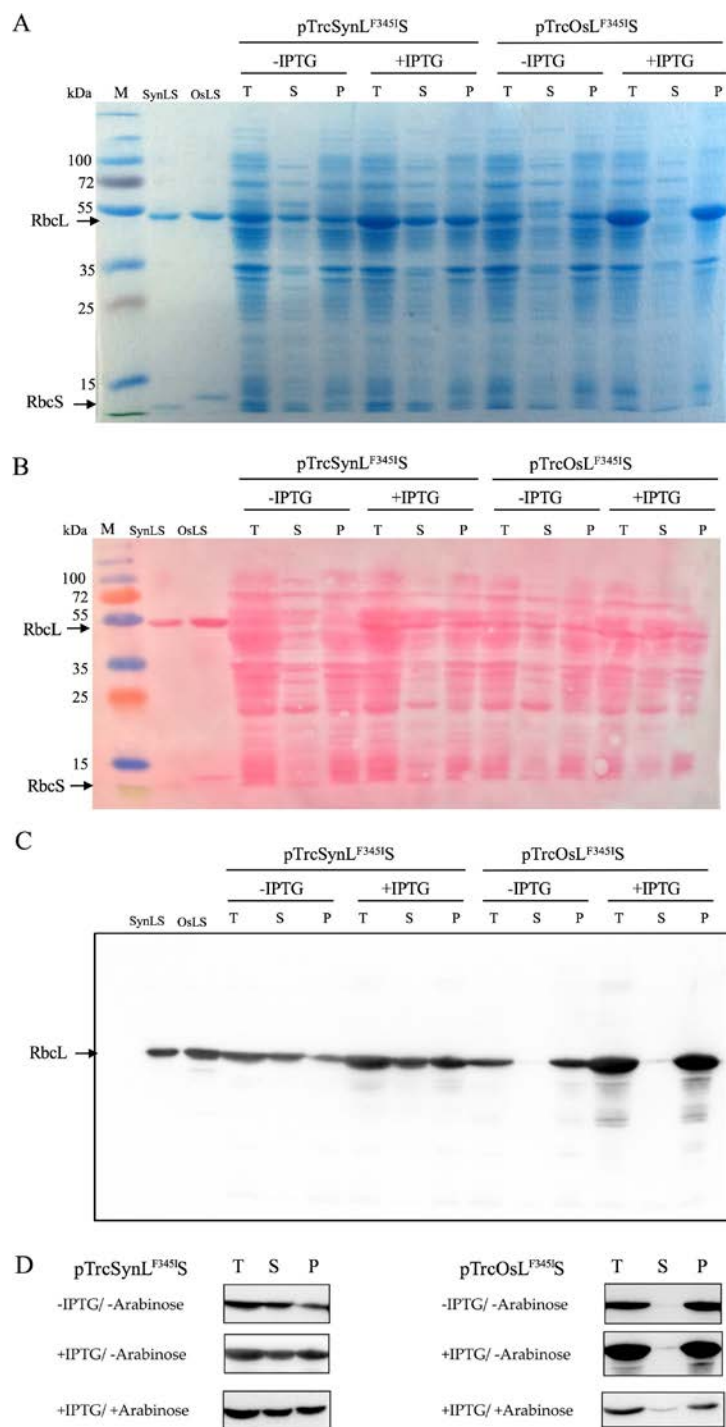


Figure 6-2: Protein expression and western blot of pTrcSynL^{F345I}S and pTrcOsL^{F345I}S in *E. coli*.

(A) The 12.5% SDS-PAGE showed the relative loading of each sample. (B) Ponceau S stain of membrane showed that the transfer was successful and the relative loading of each sample was similar to that in the 12.5% SDS-PAGE. (C) α -RbcL western blot showed that all the bands

corresponding to the large subunit would migrate at the same position (~52 kDa). (D) Western blot of protein expression of Rubisco in presence and absence of GroEL/ES supplement. The amount of sample loaded per lane was from 6.0×10^7 *E. coli* BL21 cells. ~1 µg of purified wild type *Syn*-Rubisco (*SynLS*) and ~ 1.2 µg of wild type *Os*-Rubisco (*OsLS*) were loaded as controls. The expression of Rubisco was induced by IPTG while GroEL/ES was induced by expression of arabinose. These two pure Rubisco were from the common laboratory stock and pure *Os*-Rubisco was purified from source. M: Marker, T: Total lysate, S: Soluble extract, P: Pellet, +: Induced, -: Non-induced. The amount of purified Rubisco loaded and the labeling of the diagrams are consistent in all the figures in this chapter.

6.1.3 Chimeric Rubisco had different stability in their holoenzyme complex

To investigate which regions of the *Os*-Rubisco large subunit were responsible for its insolubility in *E. coli*, a series of chimeric Rubisco large subunit genes was generated with each fragment of the *Syn*-Rubisco large subunit being replaced with its counterpart from *Os*-Rubisco. The small subunit of *Syn*-Rubisco was also replaced with that of *O. sativa* to investigate if the small subunit plays a role in the *O. sativa* insolubility in *E. coli*. This was carried out by double digestion with the respective restriction enzymes followed by ligation. The chimeric Rubisco generated in this study were shown in Figure 6-3. Protein expression test and western blot analysis were carried out to assess the solubility and assembly of these chimeric Rubiscos. The large subunit of the chimeric Rubisco was named as *SynL*(*OsL*^{'x'}) where the region replaced with those from *Os*-Rubisco will be specified by '*OsL*^x', in which 'x' represents the fragment replaced.

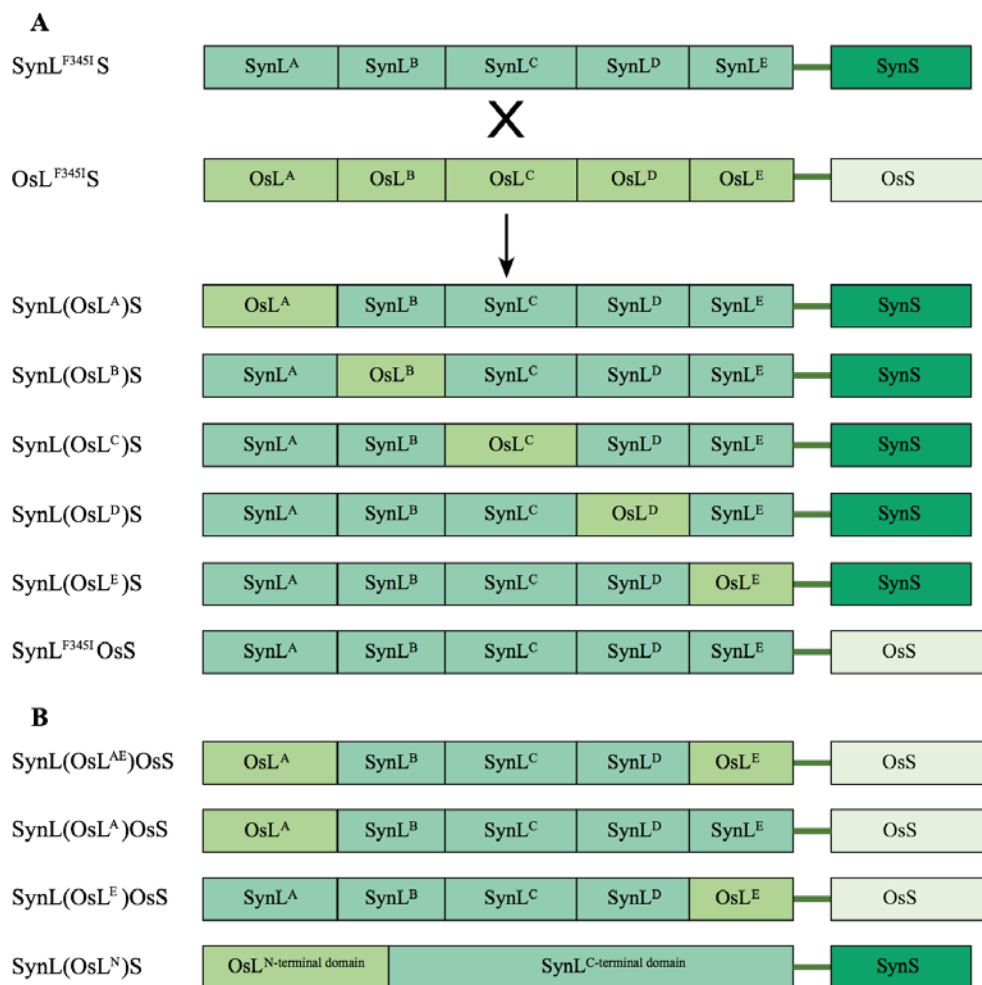


Figure 6-3: Representation of the different chimeric Rubiscos generated.

(A) Chimeric Rubiscos that had either a section of the large subunit or the small subunit replaced with the corresponding sequence from *O. sativa*. (B) Chimeric Rubiscos with more than one fragment and/or small subunit replaced with those from *Os*-Rubisco. The ‘OsL^N’ fragment refers to the N-terminal domain of *Os*-Rubisco which consists of residues 1 to 150.

6.1.3.1 Varying solubility of the chimeric Rubisco with single section replaced with *Os*-Rubisco

The chimeric Rubiscos with a single region replaced with a counterpart of *Os*-Rubisco were expressed in *E. coli* BL21 in the presence and absence of GroEL/ES supplementation (Figure 6-3A). Protein expression tests followed by western blot analyzes were performed to evaluate the solubility of these chimeric Rubiscos. The western blot analysis shows that the different chimeric Rubiscos varied in solubility when expressed in *E. coli* (Figure 6-4).

Chimeric Rubiscos containing either OsL^A or OsL^E fragment in SynL(OsL^A)S and SynL(OsL^E)S were soluble (Figure 6-4A, E). In addition, the hetero-oligomer SynL^{F345I}OsS Rubisco, which is comprised of large subunits from *Syn*-Rubisco and small subunits of *Os*-Rubisco was also soluble. Similar to the protein expression test of *Rs*-Rubisco and *Af*-Rubisco in the pTrcHisB vector, there was also basal expression of soluble SynL(OsL^A)S and SynL^{F345I}OsS. In addition, the supplementation of GroEL/ES and induction of IPTG also increased the amount of soluble protein for these two chimeric Rubiscos (Figure 6-4A, F). In contrast, there was minimal basal expression of SynL(OsL^E)S and the expression of SynL(OsL^E)S required IPTG induction.

The remaining three chimeric Rubiscos containing the fragment OsL^B, OsL^C or OsL^E were not soluble and the overexpression of GroEL/ES also did not improve the solubility of these Rubisco (Figure 6-4B, C, D). This suggested that these chimeric Rubiscos form aggregates where they are found in the extract's pellet. SynL(OsL^B)S contained *O. sativa* large subunit fragments OsL^B had a faint band in the soluble extract lane after induction with IPTG or together with arabinose (Figure 6-4B). This was likely due to the large subunit being bound to the GroEL/ES cage and not soluble expression of the chimeric Rubisco.

To further investigate the assembly of chimeric Rubisco, the induced soluble extracts from protein expression of chimeric Rubisco were analyzed on a 6% Native-PAGE followed by western blot. Soluble extracts from these Rubisco co-expressed with GroEL/ES were also analyzed. Purified *Syn*-Rubisco and *Os*-Rubisco, as well as the soluble extract of these two Rubiscos, were loaded as controls. In Native-PAGE, the proteins were separated under non-denaturing conditions and this maintains the quaternary structure of the protein. The presence of band in the Native-PAGE would correspond to folded and assembled soluble protein and suggested that the Rubisco was successfully assembled into the L₈S₈ holoenzyme. After Native-PAGE, the gel was soaked in transfer buffer supplemented with 0.1% SDS followed by transfer onto a nitrocellous membrane using the same buffer. Western blot was carried out as previously described. The additional SDS in the transfer buffer would denature the GroEL/ES such that any encapsulated Rubisco large subunit would be freed to be allow binding by primary antibody and subsequently detected. In the

western blot, the presence of bands above the Rubisco correspond to Rubisco large subunits that were encapsulated within the GroEL/ES.

From the Native-PAGE and western blot analysis, the migration of purified Rubisco was different, wild type *Os*-Rubisco (~542.4 kDa, OsLS) migrated slower as compared to *Syn*-Rubisco (~525.84 kDa, SynLS). This was expected as the two Rubisco had different molecular weight, but in this method may also reflect protein charge. Similar to western blot analysis of soluble extracts under denaturing conditions, SynL^{F345I}S was soluble, while Os^{F345I}LS was not soluble. The SynL^{F345I}S was successfully assembled and migrated in a similar fashion to the purified SynLS. Whereas, Os^{F345I}LS was encapsulated within the GroEL/ES as seen by the faint band above the Rubisco bands (Figure 6-5).

The soluble chimeric Rubiscos, SynL(OsL^A)S, SynL(OsL^E)S, and SynL^{F345I}OsS, also successfully assembled into the L₈S₈ structure. This was shown by the presence of a predominant band corresponding to the Rubisco large subunit on Native-PAGE and presence of a single band on western blot (Figure 6-5A, B). In addition, the co-expression of GroEL/ES also increased the amount of these chimeric Rubiscos. The assembled SynL(OsL^A)S (~528.0kDa) and SynL^{F345I}OsS (~538.72 kDa) migrated similarly to wild type *Syn*-Rubisco while SynL(OsL^E)S (~526.4 kDa) migrated similarly to the wild type *Os*-Rubisco (Figure 6-5A, C). The Rubisco large subunit of SynL(OsL^A)S and SynL(OsL^E)S had increased identity of 87% and 84% respectively when compared to wild type *Os*-Rubisco large subunit. This reflects a significant move towards the *O. sativa* sequence in protein sequence space (Maynard Smith, 1970), while maintaining the ability to assemble in *E. coli*.

The insoluble chimeric Rubiscos, SynL(OsL^B)S, SynL(OsL^C)S and SynL(OsL^D)S, did not assemble into an L₈S₈ structure as expected. These chimeric Rubiscos were likely to form aggregates similar to the Rice enzyme (Figure 6-5). Therefore, certain as yet unidentified residues present in *Os*-Rubisco large subunit fragments, OsL^B, OsL^C and OsL^D, all contribute to the insolubility of *Os*-Rubisco in *E. coli*.

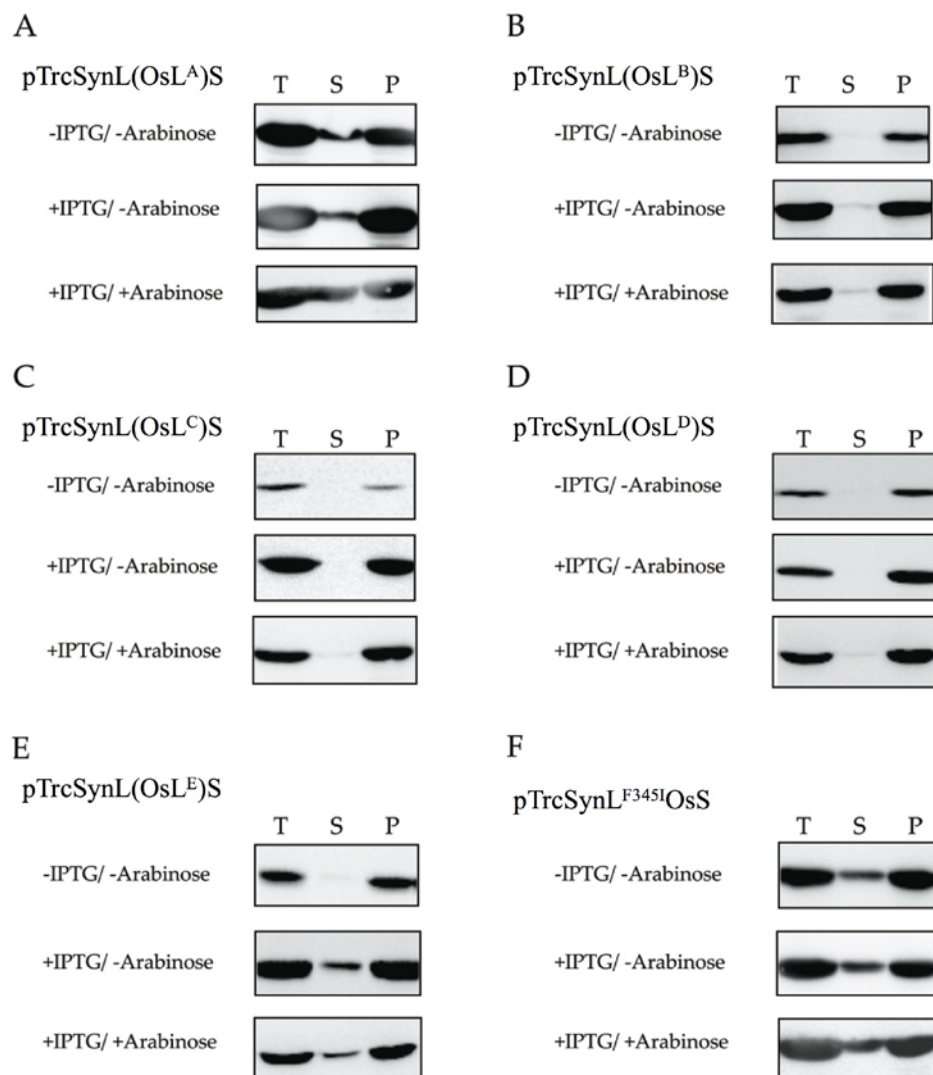


Figure 6-4: Western blot analysis of the expression of chimeric Rubisco with single region substitution with Os-Rubisco with and without the supplement of GroEL/ES in *E. coli*.

Expression of Rubisco was induced by IPTG, while GroEL/ES was induced by arabinose. The amount of sample loaded per lane was from 6.0×10^7 *E. coli* BL21 cells. Western blot analysis was performed with α -RbcL.

6.1.4 Chimeric Rubisco consisting of more than one fragment of Os-Rubisco had different solubility in *E. coli*

The analysis of chimeric Rubiscos containing a single fragment of Os-Rubisco shows that fragments containing either OsL^A or OsL^E as well as Os-Rubisco small subunit did not affect the chimeric Rubisco's solubility. Therefore, four additional chimeric Rubisco were generated to contain multiple fragments Os-Rubisco (Figure 6-6A, Figure 6-7A).

We hypothesized that the SynL(OsL^{AE})OsS chimeric Rubisco contains the two fragment OsL^A and OsL^E of the large subunit and the small subunit of Os-Rubisco could be soluble in *E. coli* as the individual regions did not affect the chimeric Rubisco's solubility. However, the protein expression and western blot analysis show that SynL(OsL^{AE})OsS was not soluble and the Rubisco was likely forming aggregates. The presence of chaperonin, GroEL/ES, was also not able to improve its solubility (Figure 6-6B, D). Therefore, the substitution of 44 amino acids in this chimeric Rubisco was deleterious regarding its solubility.

The next two chimeric Rubiscos were SynL(OsL^A)OsS and SynL(OsL^E)OsS where the small subunit is from Os-Rubisco (Figure 6-7A). The western blot analysis of the SynL(OsL^A)S and SynL(OsL^E)S had demonstrated that the chimeric large subunit was soluble. Since the folding of the two subunits is independent of each other, the substitution of small subunit is unlikely to influence the solubility of the large subunit. The last chimeric Rubisco is SynL(OsL^N)S where the large subunit consisted of the N-terminal domain (residue 1-150) from Os-Rubisco and the C-terminal domain of Syn-Rubisco (Figure 6-7A). The gene construct for SynL(OsL^N)S was obtained via restriction-free cloning as described in section 2.4.3.1. Since the OsL^A fragment which consists of residues 1 to 111 did not affect the solubility of the chimeric large subunit, it is possible that the substitution of the entire N-terminal domain could have a similar effect. Therefore, these three chimeric Rubiscos were directly purified.

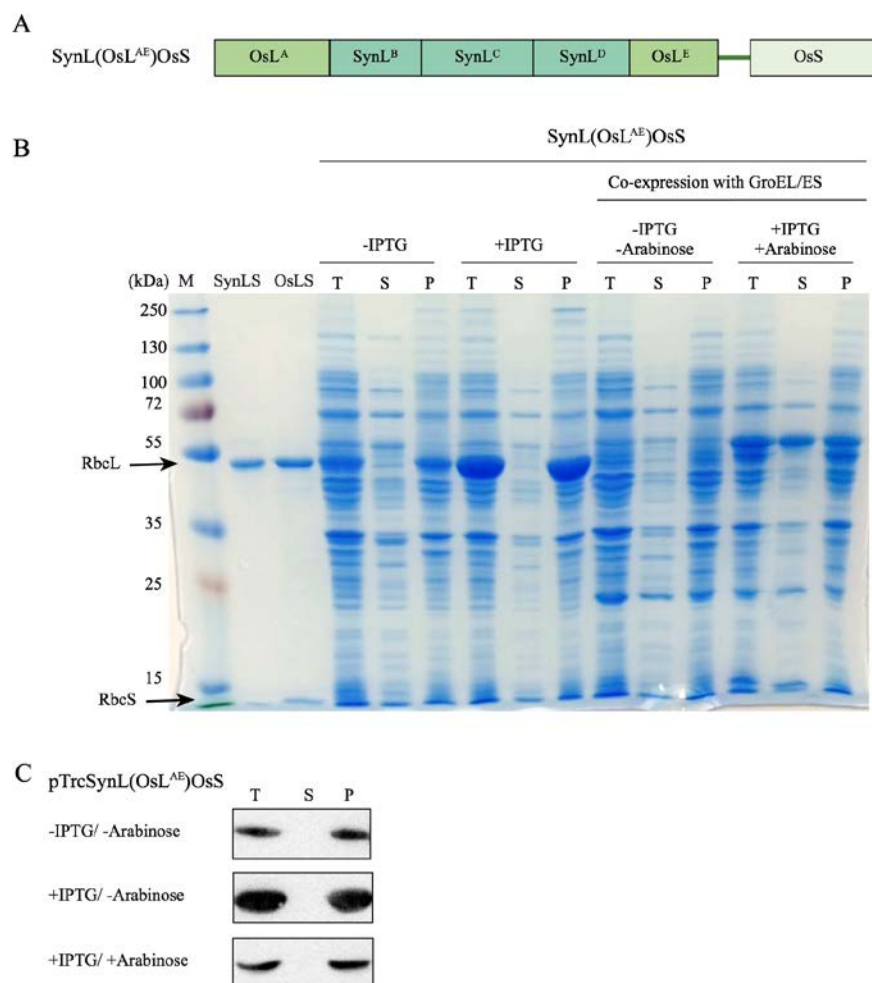


Figure 6-6: Protein expression of chimeric Rubisco, SynL(OsL^{AE})OsS in presence and absence of GroEL/ES.

(A) Representation of the chimeric Rubisco, SynL(OsL^{AE})OsS's construct where the large subunit is comprised of two fragments (OsL^A and OsL^E) and the small subunit from Os-Rubisco. (B) 12.5% SDS-PAGE of the protein expression of SynL(OsL^{AE})OsS in presence and absence of GroEL/ES. (C) Western blot analysis of the expression of SynL(OsL^{AE})OsS.

6.1.5 Assembly of soluble chimeric Rubisco were different

The various soluble chimeric Rubiscos were purified from *E. coli* BL21 using a combination of anion-exchange chromatography and size exclusion in a similar fashion as previously described for the purification of Rs-Rubisco and Af-Rubisco. As expected SynL(OsL^A)S and SynL(OsL^E)S were soluble and formed stable L₈S₈ complex similar to

the previous Native-PAGE analysis of their soluble extract. However, SynL^{F345I}OsS was soluble but it did not form a stable holoenzyme complex.

During the purification process, it became evident that chimeric Rubisco containing Os-RbcS does not form stable L₈S₈ holoenzyme as the small subunit often dissociated from the large subunit. From the SDS-PAGE of the purified Rubisco with Os-RbcS, SynL(OsL^A)OsS, SynL(OsL^E)OsS, SynL^{F345I}OsS, it was clear that there was only the large subunit was present while small subunit was absent. This was not surprising as the previous study by Wang et al. (2001) had shown using various chimeric Rubisco containing Syn-RbcL with different plant small subunit including Os-RbcS that the small subunit rapidly dissociates from the octamer core during the purification process. It was suggested that the interaction between Syn-RbcL and the plant small subunit were weak. In the absence of the small subunits, these purified chimeric Rubiscos will likely not have any Rubisco activity. Therefore, future characterization studies of these chimeric Rubiscos could be performed using soluble extract or by reconstitution of L₈ octamer core with purified small subunits.

Similar to the chimeric Rubisco with Os-RbcS, the SynL(OsL^N)S Rubisco also lost its small subunit during the purification process. In contrast to the previously chimeric Rubisco, there were multiple bands in the SDS-PAGE and the western blot. It is likely that this chimeric large subunit is unstable and degraded into smaller fragments when incubated with SDS. In the absence of stable octamer core, the small subunit is also unable to associate to form the holoenzyme. Although SynL(OsL^N)S is soluble, the Rubisco complex is unstable when purified from *E. coli* and it will not have any Rubisco activity. The difference between SynL(OsL^N)S and SynL(OsL^A)S were residues 112 to 150 and many of these residues are conserved as they are involved in inter-domain interactions as well as inter-subunit interaction (Knight et al., 1990). There were 10 amino acid substitutions between SynL(OsL^N)S and SynL(OsL^A)S and these substitutions may alter the position of critical residues needed to maintain interaction between domains and adjacent large subunit. Although these chimeric Rubiscos do not form the stable Rubisco holoenzyme, it is possible that using directed evolution these Rubiscos can be evolved to form a stable complex.

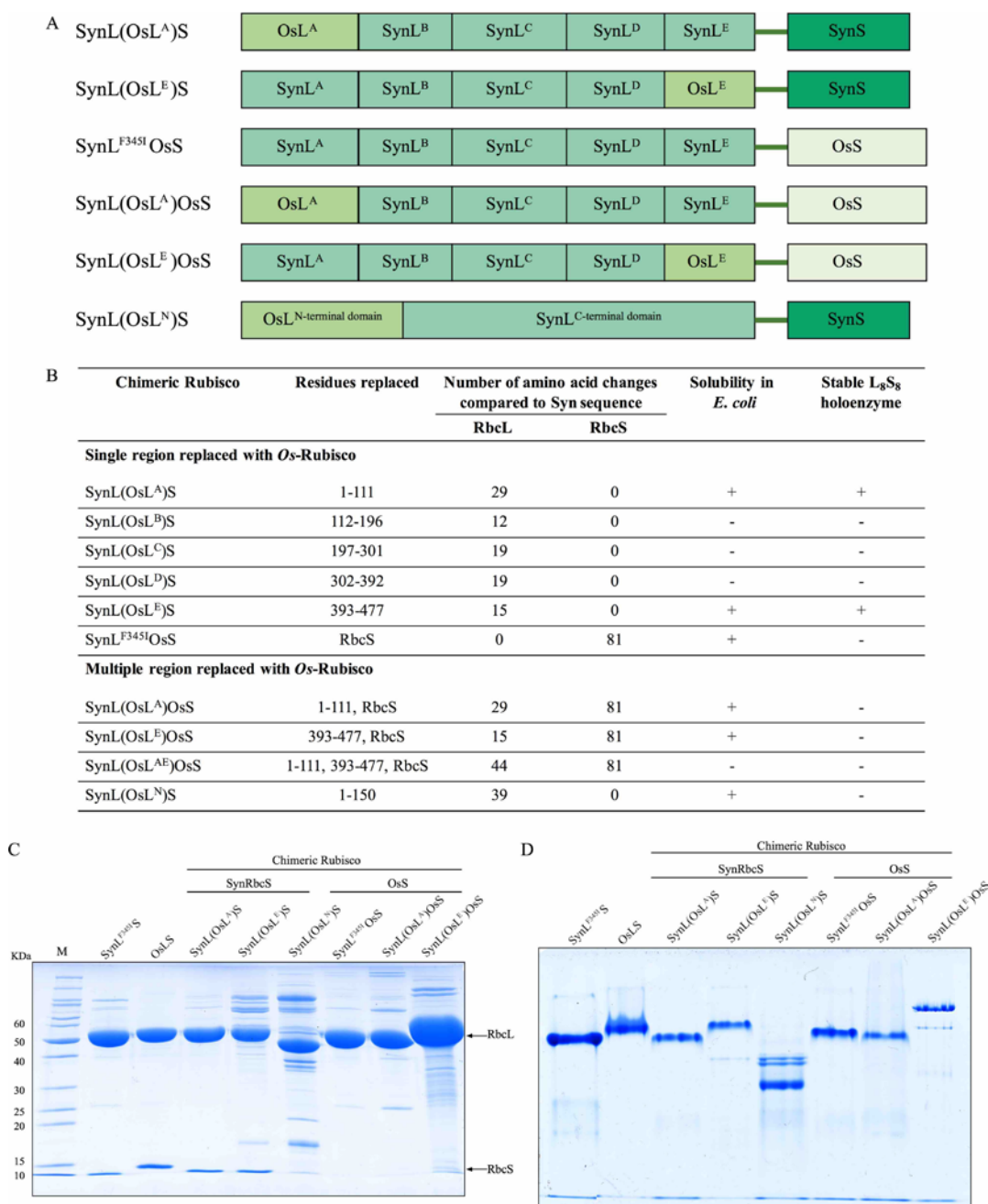


Figure 6-7: Summary of soluble chimeric Rubisco.

(A) Construct of soluble chimeric Rubisco. (B) Summary of the solubility and stability of chimeric Rubisco in *E. coli*. Representative gels of purified chimeric Rubisco under denaturing condition, (C) 12.5% SDS PAGE, and non-denaturing conditions, (D) 6% Native PAGE. (C) Summary of the solubility and stability of chimeric Rubisco in *E. coli*. ~5μg of pure Rubisco was for the SDS-PAGE and Native-PAGE except for SynL(OsL^E)OsS where ~10μg was loaded instead.

6.2 Biochemical characterization of purified chimeric Rubisco

Higher plant Rubisco such as those from *Os*-Rubisco have formed inhibited complexes similar to that of *Rs*-Rubisco and *Af*-Rubisco and they required assistance from its Rubisco activase (*Os*-Rca) to facilitate the inhibitor release. Similar to most plants, *O. sativa* contains two isoforms of *Os*-Rca (*Os*-Rca α and *Os*-Rca β) and they differ in length at the C-terminus. The shorter isoform, *Os*-Rca β , has higher ATPase activity and more abundant compared to the longer isoform, *Os*-Rca α (Scafaro et al., 2016; Shivhare et al., 2017). In contrast, *Syn*-Rubisco has a high rate of inhibitor release and does not form a stable inhibited complex (Lee et al., 1991; Pearce, 2006). Thus, *Syn*-Rubisco did not evolve the need for an activase.

It is possible that the chimeric Rubisco which contains some elements from *Os*-Rubisco may gain characteristic similar to *Os*-Rubisco. Therefore, the chimeric Rubiscos were characterized using spectrophotometric Rubisco assay to determine their Rubisco activity and their ability to form inhibited complexes. The various soluble Rubiscos were assayed for their carboxylation rate and carbamylated Rubisco complexes (ECM) were prepared by pre-incubating Rubisco with bicarbonate and Mg^{2+} . The reaction was initiated by adding ECM to reaction mixtures containing the substrates 10 mM $NaHCO_3$ and 1mM RuBP.

6.2.1 Soluble chimeric Rubiscos show diversity in their Rubisco activity

The ECM activity of chimeric Rubisco, *SynL*(*OsL*^A)*S* and *SynL*(*OsL*^E)*S*, were assayed using 0.1 μ M of Rubisco. *SynL*(*OsL*^E)*S* was shown to have carboxylation activity with a rate of $2.1s^{-1}$ which was slightly slower compared to $2.5s^{-1}$ in *SynL*^{F345I}*S*. The lower carboxylation rate of *SynL*^{F345I}*S* observed here was due to the use of low non-saturating CO_2 concentrations (10 mM). In contrast, there was little to no activity observed for *SynL*(*OsL*^A)*S* under the same reaction (Figure 6-8A). This was initially puzzling as *SynL*(*OsL*^A)*S* was able to form stable L_8S_8 complex similar to *SynL*(*OsL*^E)*S* and therefore, it should have the ability to catalyze the reaction. This would also suggest that the current experimental condition may not be suitable for characterizing *SynL*(*OsL*^A)*S* and was subsequently optimized to better capture *SynL*(*OsL*^A)*S*'s activity.

6.2.1.1 SynL(OsL^A)S exhibits characteristic similar to plant Rubisco

The initial characterization of SynL(OsL^A)S were not successful and the experimental condition was adjusted. Previous studies on plant activase and CbbQO activase used molecular crowding agents such as PEG to demonstrate the activase activity and it was shown that the presence of PEG enhances the Rubisco and activase interaction (Salvucci, 1992; Tsai et al., 2015). Therefore, 5% PEG was added into the reaction mix and higher concentrations of 1 μ M of Rubisco was used for the Rubisco activation assay.

The ECM activity of SynL(OsL^A)S has a concave downwards slope as compared to the linear ECM activity of SynL^{F345I}S and SynL(OsL^E)S. This concave downwards slope suggests that SynL(OsL^A)S exhibits ‘fallover’ which meant that the Rubisco activity declined over time due to inhibition by misfired products produced during catalysis. This fallover inhibition is an in-vitro phenomenon that is also observed in higher-plant Rubiscos (Parry et al., 2008). Therefore, to calculate the carboxylation activity of SynL(OsL^A)S, the production accumulation over time plot was fitted with Equation 2.4 as previously described in chapter 4. The initial rate (v_i) obtained for SynL(OsL^A)S ECM with PEG was 0.27s⁻¹ and this was higher than the final rate (v_f) of 0.11s⁻¹ (Table 6-3). This result is consistent with Rubiscos exhibiting fallover effect where the initial activity is higher than the final activity as the amount of active site is reduced over time due to inhibition and activity will decline with time.

The ability to exhibit fallover suggest that SynL(OsL^A)S was likely forming stable inhibited complexes with inhibitors, which is another characteristic of plant Rubisco. To form the inhibited complex, SynL(OsL^A)S was incubated EDTA followed by with RuBP or XuBP to form RuBP-inhibited (ER) complex and XuBP-inhibited (EX) complex respectively. XuBP is a misfire product from the carboxylation reaction and it binds to the apo-Rubisco complex with a higher affinity as compared to RuBP (Pearce et al., 2003). The assay results show that SynL(OsL^A)S was able to form inhibited complexes with both RuBP and XuBP (Figure 6-8B). In addition, the presence of PEG did not influence affect the stability of the inhibited complex (Figure 6-8C). The rate of spontaneous release of RuBP from ER complex was also higher as compared to XuBP release from the EX complex as shown by

the higher k_{obs} and higher value of 3PGA produced over time (Figure 6-8B, C, Table 6-3). This was consistent with the nature of these two inhibitors where XuBP binds more tightly to the apo-Rubisco, which translates to a slower rate of spontaneous release from the active site of the inhibited complex.

These inhibited complexes of SynL(OsL^A)S were subsequently assayed with the addition of Rubisco activase from plants to investigate if these activases were able to recognize and facilitate removal of inhibitor. Os-Rubisco's activase OsRca β and a CAM plant's activase, AtRca β , were chosen to carry out the remodeling to these inhibited complexes. AtRca β is the short isoform of Rubisco activase from *Agave tequilana* and it was recently shown to be able to activate inhibited Os-Rubisco with higher efficiency as compared to OsRca β (Shivhare et al., 2017). Both activase were able to recognize the ER and EX complexes and facilitate inhibitor removal as shown by the increased in Rubisco activity over time (Figure 6-8B, D). In addition, AtRca β was able to remodel the inhibited complexes (ER and EX) with higher efficiency as compared to OsRca β as demonstrated by the higher Rubisco activity observed in inhibited complexes with AtRca β . AtRca β was also added to the activated complex (ECM) and it was able to increase the Rubisco activity of ECM. This further support the hypothesis that SynL(OsL^A)S was exhibiting fallover effect as AtRca β was able to facilitate inhibitor release and maintain the pool of active site available for catalysis (Figure 6-8C).

The biochemical characterization of SynL(OsL^A)S indicated that this chimeric Rubisco exhibits the characteristics of plant Rubisco whereby it demonstrated in-vitro fallover inhibition and formed RuBP- and XuBP-inhibited complexes which can be recognized and remodeled by plant Rubisco activase (Robinson et al., 1989). These characteristics are not found in Syn-Rubisco which does not form inhibited complexes and does not exhibit fallover effect (Horken et al., 1999; Lee et al., 1991). SynL(OsL^A)S contains residues 1 to 110 from Os-Rubisco and among these residues there are 29 amino acid substitutions while the remaining 81 residues were conserved. Some of these changes, therefore, encode the tendency of SynL(OsL^A)S Rubisco to form a tightly closed complex when bound by an inhibitor. In addition, it likely that the Rubisco activase (AtRca β or OsRca β) interacts with the SynL(OsL^A)S via residues Pro89 and Glu94 on the OsL^A fragment which has been

shown to interact with residues Lys311 and Val314 of Rubisco activase (Li et al., 2006; Portis et al., 2008). Although SynL(OsL^A)S has many characteristics of plant Rubisco, its carboxylase activity is low in comparison. Therefore, directed evolution methodology could be used to evolved SynL(OsL^A)S towards a more functional enzyme.

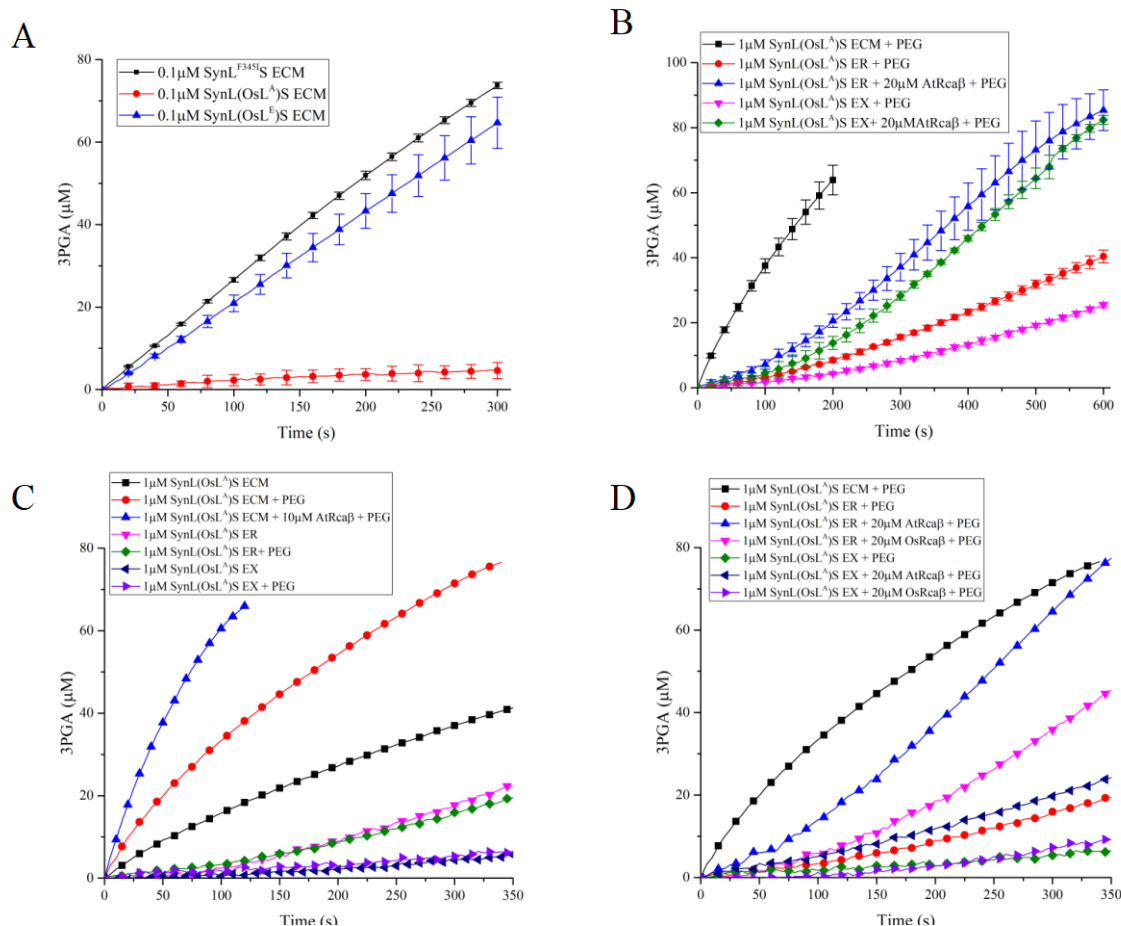


Figure 6-8: Rubisco activation assay of chimeric Rubisco.

(A) Rubisco activation assay of chimeric Rubisco that forms stable holoenzyme complex. SynL(OsL^E)S has similar ECM activity as SynL^{F345I}S while SynL(OsL^A)S had very little activity. (B) Rubisco activation assay of SynL(OsL^A)S with PEG where the ECM activity exhibits fallover. SynL(OsL^A)S also formed ER and EX complex which could be recognized and remodeled by AtRcaβ (C) Rubisco activation assay of activated (ECM), RuBP- inhibited (ER) and XuBP- inhibited (EX) complexes of SynL(OsL^A)S in presence and absence of PEG. (D) Rubisco activation assay of ER and EX complexes of SynL(OsL^A)S with the addition of activase, AtRcaβ and OsRcaβ. The presence of an error bar in (A) and (B) indicates that the result is from three independent experiments. The results from (C) and (D) were from a single experiment. Rubisco activation assay was carried out using 10 mM NaHCO₃ and 1 mM RuBP at 25°C.

6.3 Directed evolution of chimeric Rubisco

Among the different chimeric Rubiscos generated, only SynL(OsL^A)S and SynL(OsL^E)S were soluble and assembled into stable L₈S₈ holoenzyme complex. Mutant libraries of these two chimeric Rubiscos were generated as well as libraries for two other chimeric Rubiscos of interest, SynL(OsL^N)S and SynL(OsL^{AE})OsS (Table 6-1). These libraries were generated using error-prone PCR with the same reaction conditions as the libraries for *Rs*-Rubisco and *Af*-Rubisco, therefore, the mutagenic rates of the libraries should be similar as well. These libraries were screened using the MM1-pAC^{BAD}PRK⁶⁸⁰³ system with RDE selective media. The SynL(OsLA)S Rubisco had exhibited many characteristics of plant Rubisco whereby it required an activase. Therefore, the directed evolution of SynL(OsLA)S was carried out alone and also in the presence of an activase. The control pTrcDuet construct (pTrcSynL(OsL^A)S_ AtRcaβ) containing both SynL(OsL^A)S and AtRcaβ was generated in a similar fashion as pTrcRsLSRsX (Figure 2-1B). The mutant library containing randomly generated mutations of SynL(OsL^A)S with wild type AtRcaβ was generated in the same manner. Unlike higher plant Rubisco, Rubisco activase can be expressed recombinantly in a bacterial host. Therefore, the presence of AtRcaβ can be present to alleviate inhibition by RuBP or misfired product and allow SynL(OsLA)S to carried out RuBP catalysis to sustain MM1-cell growth.

Mutant libraries	Size of mutant libraries	Screening conditions
EpSynL(OsL ^A)S-1	2,000	RDE selective media supplemented with 0.05% (w/v) arabinose in air supplemented with 1% CO ₂ at 23 °C
EpSynL(OsL ^A)S-2	151,000	
EpSynL(OsL ^E)S-1	3,000	
EpSynL(OsL ^E)S-2	235,000	
EpSynL(OsL ^N)S	27,000	
EpSynL(OsL ^{AE})OsS	22,500	
EpSynL(OsL ^A)S_AtRcaβ	40,000	1. RDE selective media and RADE selective media supplemented with 0.05% (w/v) arabinose in air supplemented with 1% CO₂ at 23 °C 2. RADE selective media supplemented with 0.05% (w/v) arabinose under atmospheric condition (0.04% CO₂) at room temperature

Table 6-1: The chimeric Rubisco libraries generated and the screening conditions.

6.3.1 Small numbers of variants were obtained from directed evolution of chimeric Rubisco

The directed evolution of the various chimeric Rubisco was carried out using RDE and RADE selective media. Under the various screening conditions, all the various constructs containing chimeric Rubisco (pTrcSynL(OsL^A)S, pTrcSynL(OsL^E)S, pTrcSynL(OsL^N)S and pTrcSynL(OsL^A)S_AtRcaβ) were unable to sustain the growth of the MM1-cells. In contrast to the directed evolution of *Rs*-Rubisco and *Af*-Rubisco, the directed evolution of these chimeric Rubiscos yielded only a handful of variants (Table 6-2). This could be due to the lower fitness of these chimeric Rubiscos as compared to the two bacterial wild type Rubisco and therefore, it could be more difficult to evolve them in comparison. In addition, there were no variants obtained from the screenings of EpSynL(OsL^{AE})OsS where its wild type counterpart was insoluble in *E. coli*. This could suggest that the mutagenic rate of the library may not be sufficient to provide sufficient advantageous mutation that enables solubility and functionality of this chimeric Rubisco to allow survival in RDE screening. Among the various variants obtained from screening, there were only two variants obtained and these were repeatedly found from the directed evolution of SynL(OsL^A)S in the absence of activase. These two variants were SO-CA1 and SO-CA2 and they were characterized using purified Rubisco.

Rubisco variants	Amino acid mutation(s)	
	RbcL	RbcS
EpSynL(OsL^A)S		
SO-CA1*	H86L, R213S	S67G
SO-CA2*	Q209L, I465N	
EpSynL(OsL^A)S_AtRcaβ		
SO-CAT1	V69E	K66E, R107H
SO-CAT2	-	I42T, H106R, G109D
SO-CAT3	-	Q25R
SO-CAT4	K236N	P20H, P50L
EpSynL(OsL^E)S		
SO-CE1	-	Y32N
EpSynL(OsL^N)S		
SO-CN1	I98N, G442S	L6P

Table 6-2: Chimeric Rubisco variants obtained from screening.

*Among the various Rubisco variants obtained, only *SO-CA1* and *SO-CA2* were repeatedly selected for. For the *SO-CAT1* -3 variants were obtained from RDE screening media while *SO-CAT4* was obtained from RADE screening media and they were all grown under elevated CO₂ conditions.

6.3.2 *SO-CA1* and *SO-CA2* may have evolved away from plant Rubisco

The two evolved variants of SynL(*OsL^A*)S, *SO-CA1* and *SO-CA2*, were purified and characterized for their carboxylation activity and nature to form inhibited complexes with RuBP and XuBP using spectrophotometric Rubisco activation assay. From the SDS-PAGE gels of the purified *SO-CA1* and *SO-CA2*, it can be observed that both variants contain large and small subunit. However, on the Native-PAGE, *SO-CA1* ran as a smear which suggests that this chimeric Rubisco was unstable after purification and further characterization must be interpreted with caution. In comparison, *SO-CA2* ran as a single band on the Native-PAGE similar to the control Rubisco samples which suggest it forms a stable holoenzyme complex and maintain this complex after purification.

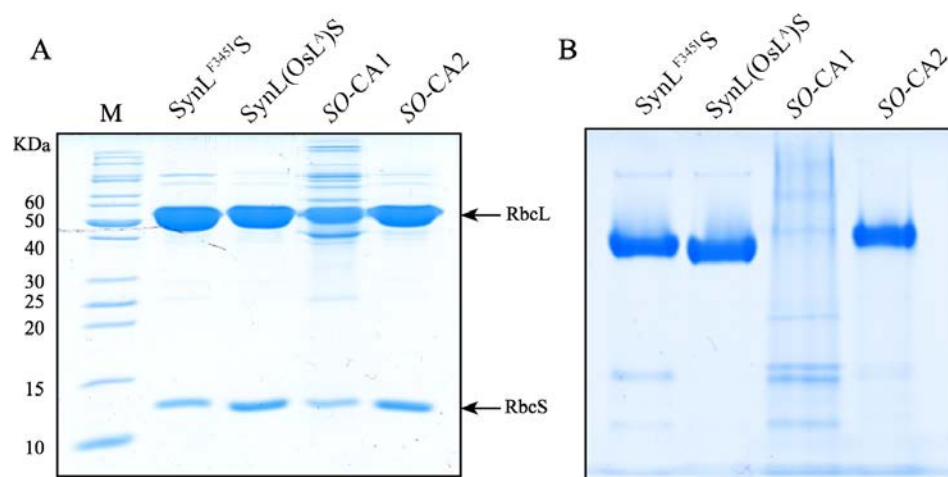


Figure 6-9: Representative gel of the purified chimeric SynL(*OsL^A*)S and its evolved Rubisco variant.

(A) 16% SDS-PAGE (B) 6% Native-PAGE of the purified chimeric rubisco. ~5 µg of purified Rubisco was loaded on each lane.

Biochemical characterization of the two variants was carried out using the same experimental condition as SynL(*OsL^A*)S. Both variants exhibited linear ECM activity, which suggests that they no longer exhibit fallover (Figure 6-10C). In addition, it also meant that the variants were likely not forming inhibited complexes with misfired products. *SO-*

CA2 also had a higher CO₂ turnover rate as compared to SynL(OsL^A)S and SO-CA1. This higher CO₂ turnover rate would translate to a higher rate of RuBP turnover which could have provided survival fitness for this variant. In contrast, SO-CA1 ECM had the lowest CO₂ turnover rate which may be attributable to the instability of the Rubisco complex.

The comparison of the activity of ER and EX complexes show that both variants no longer form inhibited complexes with RuBP and XuBP. Although the ER and EX activity of SO-CA1 may appear similar to SynL(OsL^A)S, the comparison of the relative CO₂ turnover rate for the inhibited complexes of SO-CA1 was similar to the ECM. The similar ECM rate suggests that inhibitors are rapidly released from the active site of SO-CA1 thereby allow it to catalyze CO₂ fixation with rates similar to ECM. Further studies are required to confirm this finding and the characterization of SO-CA1 could be performed using soluble extracts as purification appears problematic. In contrast to SO-CA1, the results of the biochemical characterization of SO-CA2 were more concrete. It was evident that SO-CA2 no longer form inhibited complexes with ER and EX as both activities had higher 3PGA production rates as well as a two-fold increase in the k_{obs} as compared to SynL(OsL^A)S (Figure 6-10C, D, Table 6-3).

Structural analysis for the SO-CA2 was carried out using the structure of *Syn*-Rubisco in the closed state which was bounded by CABP (PDB ID: 1RBL). SO-CA2 contained two substitutions, Q209L and I465N. Q209 resides in loop 2 which also contains several essential active site residues that are involved in the Mg²⁺ binding site (Figure 6-11B). It is likely that Q209L substitution alters the local conformation such that the active site becomes loosen and allow a premature release of inhibitors. The second substitution is I465N which has been selected for in *Af*-B17 variant and it was characterized in detail (Section 5.3.3.1.2). Ile465 is located on the C-terminus where it interacts with residues from loops involved in active site closure. Therefore, the I465N mutation is likely affecting inhibitor binding by reducing the rigidity of the active site. This reduction allows increased rate of inhibitor release as supported by the four-fold increase k_{obs} observed in *Af*-B17.

Overall the results from the biochemical characterization of the two evolved SynL(OsL^A)S variants showed similar results to the directed evolution of *Rs*-Rubisco and *Af*-Rubisco in

the absence of their activase where they no longer formed inhibited complexes with RuBP. The *SO-CA1* and *SO-CA2* were also evolved Rubisco forming unstable complexes with XuBP. This loss in the ability to form an inhibited complex also meant that the fallover effect was also not observed in these variants. The variants appeared to evolve towards the phenotype of wild-type *Syn*-Rubisco which does not form the stable inhibited complex.

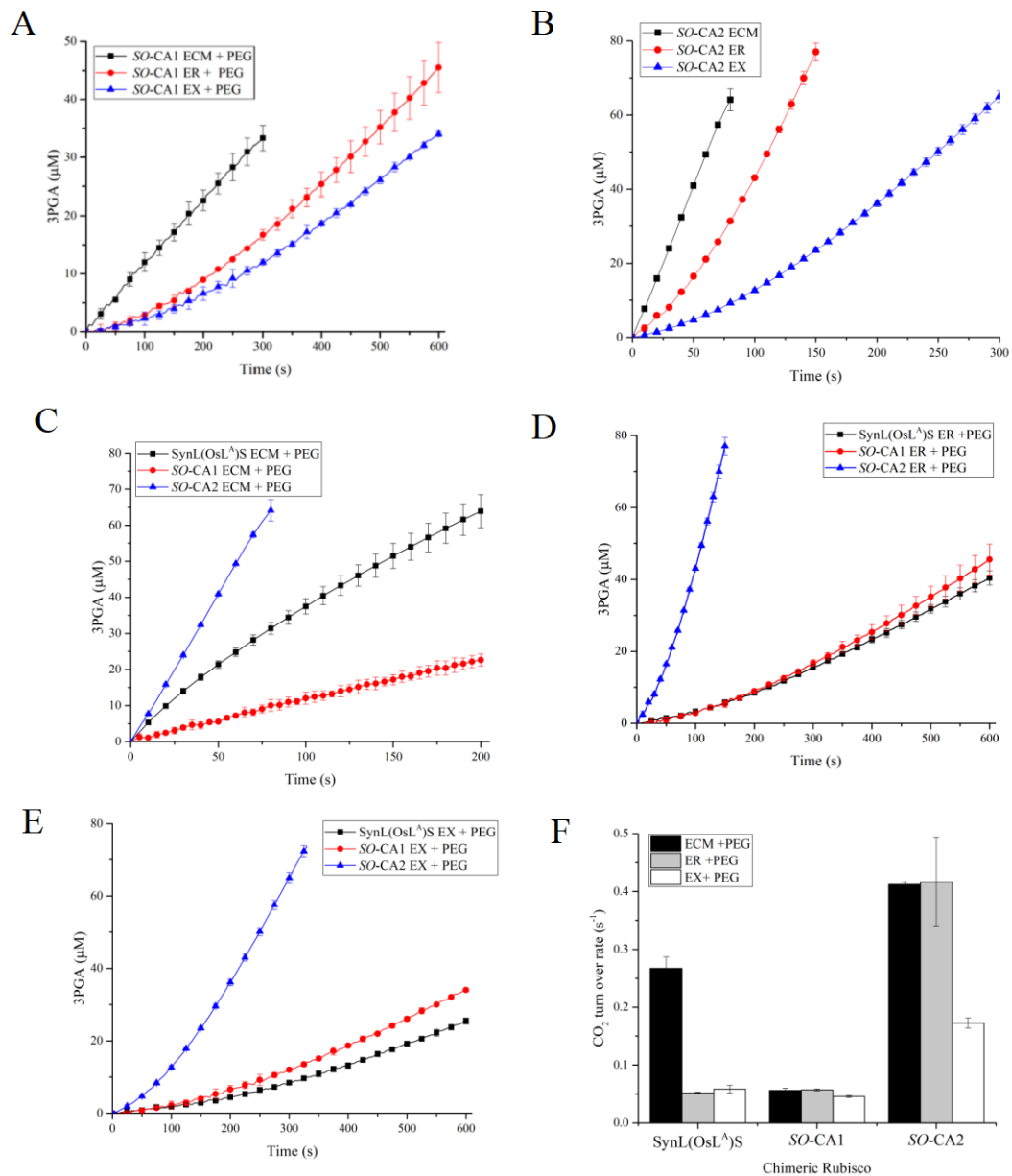


Figure 6-10: Rubisco activation assay of the evolved *SO-CA1* and *SO-CA2* Rubisco.

(A) Rubisco activation assay of SO-CA1. (B) Rubisco activation assay of SO-CA2. (C) The ECM activity of SynL(OsL^A)S and the evolved variants where SO-CA2 had the highest ECM activity. (D) Rubisco activation assay of ER complexes of SynL(OsL^A)S and the evolved variants where SO-CA2 no longer form ER complex. (E) Rubisco activation assay of EX complexes of SynL(OsL^A)S and the evolved variants where SO-CA2 no longer form inhibited complex with XuBP (F) Summary of the CO₂ turnover rate of the SynL(OsL^A)S and the evolved variants. For the CO₂ turnover rate of the ECM, the v_i value was taken. While for the ER and EX, the v_f value was taken as the initial rate would be low as Rubisco requires time to release the inhibitors and subsequently carbamylated prior to carry out catalysis. The values for this figure was obtained Table 6-3 and the error bars represent the standard deviation of data from at least two independent experiments. The Rubisco activation assay was carried out using 1 μ M of Rubisco complex with a reaction mixture containing 10 mM NaHCO₃, 1 mM RuBP and 5% PEG at 25°C.

	v_i (s ⁻¹)	v_f (s ⁻¹)	k_{obs} (s ⁻¹)	r^2
SynL(OsL^A)S				
ECM + PEG	$2.67 \pm 0.20 \times 10^{-1}$	$1.12 \pm 0.05 \times 10^{-1}$	$1.98 \pm 0.21 \times 10^{-2}$	0.9957
ER + PEG	$8.71 \pm 2.68 \times 10^{-3}$	$5.21 \pm 0.12 \times 10^{-3}$	$3.50 \pm 0.39 \times 10^{-3}$	0.9998
EX + PEG	$5.92 \pm 2.56 \times 10^{-3}$	$5.85 \pm 0.65 \times 10^{-2}$	$1.33 \pm 0.27 \times 10^{-3}$	0.9990
SO-CA1 Substitutions: RbcL-H86L,R213S;RbcS-S67G				
ECM + PEG	$5.62 \pm 0.38 \times 10^{-2}$	-	-	-
ER + PEG	$9.54 \pm 2.75 \times 10^{-3}$	$5.71 \pm 0.16 \times 10^{-2}$	$1.98 \pm 0.21 \times 10^{-3}$	0.9997
EX + PEG	$9.21 \pm 1.05 \times 10^{-3}$	$4.58 \pm 0.16 \times 10^{-2}$	$1.44 \pm 0.39 \times 10^{-3}$	0.9995
SO-CA2 Substitutions: RbcL-Q209L,I465N				
ECM + PEG	$4.12 \pm 0.05 \times 10^{-1}$	-	-	-
ER + PEG	$1.10 \pm 0.02 \times 10^{-1}$	$4.17 \pm 0.76 \times 10^{-1}$	$8.04 \pm 0.96 \times 10^{-3}$	0.9999
EX + PEG	$2.40 \pm 0.27 \times 10^{-2}$	$1.73 \pm 0.09 \times 10^{-1}$	$6.66 \pm 0.52 \times 10^{-3}$	0.9999

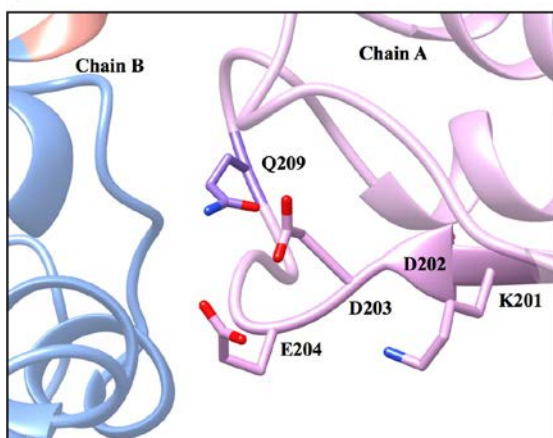
Table 6-3: Summary of the kinetic properties of SynL(OsL^A)S and its evolved variants.

The values of initial rate (v_i), final rate (v_f) and rate of inhibitor release (k_{obs}) were calculated by fitting of the curves from Figure 6-8B, Figure 6-10A, B Equation 2.4. For ECM activity of SO-CA1 and SO-CA2, the rate of activity is calculated using linear fit. The data in this table represent results from at least three independent experiments.

A

Rubisco	Substitution		Large subunit interactions	Structural analysis
	RbcL	RbcS		
SO-CA2	Q209L, I465N	nil	[Q209]: L107, F108, E109, L178, N205, N207, P210, F211, R213	Syn-Rubisco (1RBL, Closed)

B



C

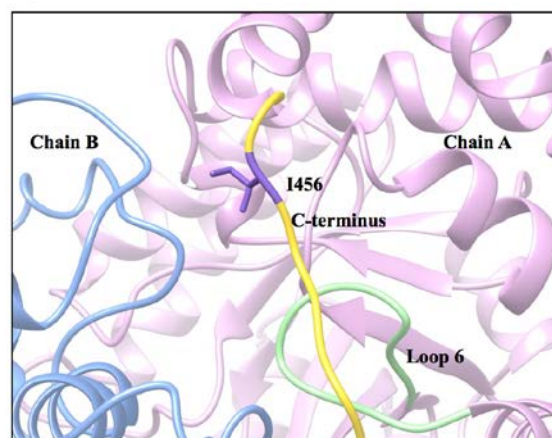


Figure 6-11: Structural analysis of the mutated residues in SO-CA2.

(A) Summary of the interactions of Q209 in SO-CA2 Rubisco variant. (B) Position of Gln209 and essential residues of loop 2 and strand 2. (C) Position of Ile456 on the C-terminus.

6.4 Discussion

6.4.1 Higher plant Rubisco's insolubility in *E. coli* may be caused by a small number of residues

In this study, we demonstrated two *Synechococcus/O. sativa* chimeric Rubisco (SynL(OsL^A)S and SynL(OsL^E)S) that were soluble and assembled successfully into an L₈S₈ complex in *E. coli*. These two variants contained hybrid Rubisco large subunit with amino acid sequences from cyanobacteria, *Synechococcus*, and plant, *O. sativa*. Many of the previous studies of chimeric Rubisco focused on the exchange of large subunit or small subunit between two species which was similar to SynL^{F345I}OsS (Genkov et al., 2010; Sharwood et al., 2008; Wang et al., 2001). A recent study by Koay et al. (2016) generated chimeric Rubisco with large subunit sequences from *Synechococcus* and *Chlamydomonas*. The large subunit was divided into 50 amino acid segments and chimeric Rubisco was generated in a similar manner by swapping those from *Chlamydomonas* into Syn-Rubisco. Immunoblot analysis showed similar results whereby the residues 1 to 98 and residues 448

to 472 did not affect the solubility of these chimeric large subunits. These residues corresponded to the fragments that were exchanged in SynL(OsL^A)S and SynL(OsL^E)S. Koay et al. (2016) also showed that residues 198 to 247 also did not affect the solubility of the chimeric Rubisco. These residues belong to the first half of OsL^C. However, the activity of the different soluble chimeric Rubisco was not measured and there is a possibility that these Rubiscos may not be functional.

Taking all these together, the different amino acids (11) between residues 248 to 301 were likely to contribute towards the insolubility of SynL(OsL^C)S. These finding all suggested that there is only a handful of residues that may contribute towards plant Rubisco's insolubility in *E. coli*. Therefore, it is possible that if these residues were identified and mutated, plant Rubisco could potentially be recombinantly expressed and purified using bacterial host without necessitating the additional requirement of specific folding and assembly chaperones (Aigner et al., 2017).

6.4.2 Rubisco can be easily evolved to have a higher rate of inhibitor release

Syn-Rubisco and Os-Rubisco have very different kinetic properties which reflected the divergence in their evolutionary path. Syn-Rubisco evolved in the high CO₂ environment in the carboxysome where it has low specificity, a high K_C , and does not form inhibited complexes with sugar phosphates. Whereas Os-Rubisco, evolved under atmospheric CO₂ levels, has high specificity factors, a very low K_C and is also prone to inhibition by sugar phosphates. Os-Rubiscos require assistance from activase which in term regulates Os-Rubisco's activity according to the photosynthetic activity of the plant. Therefore, we hypothesized that when chimeric Rubiscos are generated from Syn-Rubisco and Os-Rubisco, it could potentially inherit some of the characteristics of plant Rubisco. Indeed, SynL(OsL^A)S has plant Rubisco traits where it could form RuBP-inhibited complexes, exhibited fallover, and could be subsequently recognized and remodeled by plant Rubisco activase.

The directed evolution of SynL(OsL^A)S in the absence of activase yielded two repeating variants which had evolved away from plant Rubisco and towards Syn-Rubisco. These Rubiscos lost their propensity to form inhibited complexes which is a trait similar to that of

Syn-Rubisco. This result was consistent with the findings of chapter 5 in which Rubisco can be easily evolved to form unstable inhibited complexes in the absence of their activase.

SynL(*OsL*^A)*S* was also evolved in the presence of *AtRcaβ* and the variants obtained (*SO*-CAT) were not yet characterized due to time restraints. However, we speculate that these *SO*-CAT Rubisco were likely to maintain the characteristics of *Os*-Rubisco and still form inhibited complexes with inhibitor. As shown in chapter 5 where *Rs*-Rubisco was also evolved in the presence of its activase, the *Rs*-S1 variant was shown to form a stable inhibited complex with RuBP. Therefore, future studies can be focused on characterizing these *SO*-CAT Rubisco variants and any improvements observed in these variants could provide valuable insights into potential sites that can be translated into plant Rubisco.

6.4.3 Future prospects of directing the evolution of Rubisco

A major milestone of Rubisco research has recently been achieved by Aigner et al. (2017) where they successfully expressed *Arabidopsis thaliana* Rubisco (*At*-Rubisco) in *E. coli*. This was achieved by co-expressing *At*-Rubisco with five different chaperonins, Cpn60/Cpn20, Rubisco assembly factors 1 and 2 (Raf1, Raf2), RbcX and bundle-sheath-defective 2 (BSD2). In particular, the recent understanding of various assembly factors (Raf1, 2, RbcX and BSD2) has all contributed towards expediting this long-standing goal in Rubisco research (Bracher et al., 2017). The ability to express plant Rubiscos recombinantly in *E. coli* will allow rapid and easy access to many plant Rubisco as the current protocol for purification of these Rubiscos are restricted to physical plant leaves. This process generally limit access to the wildtype enzyme, as mutants can only be made in tobacco (Whitney et al., 2008).

Although it is tempting to co-expressed all these chaperonins and evolve plant Rubisco, the presence of these chaperonins could potentially to limit the protein sequence space available for evolution. It was previously shown that when *Syn*-Rubisco was evolved in presence of the assembly chaperone RbcX, RbcX limited the number of available mutations likely due to the need to maintain interaction between RbcX and *Syn*-Rubisco (Durao et al., 2015). It is likely that if these chaperonins were to be co-expressed in *E. coli*, they should be integrated into the genome as multiple plasmids could complicate the process to obtain the

evolved Rubisco plasmid. One good host for such an experiment could be the *E. coli* strain created by Antonovsky et al. (2016). This artificially evolved *E. coli* contains a non-native CBB cycle with Rubisco and PRK which can perform sugar synthesis using the CO₂ fixed by Rubisco. Therefore, it has great potential for directing the evolution of Rubisco. In addition, Rubisco activase should also be co-expressed to ensure that plant Rubisco maintain their propensity to form inhibited complexes which are important for the regulation of Rubisco's activity in plant. As shown by our directed evolution studies in this chapter and the previous chapter, Rubisco can easily be evolved away from forming an inhibited complex in the absence of their activase. It is likely that over the course of evolution, activase has restricted Rubisco's evolutionary space such that Rubisco continues to form inhibited complexes. The recent findings on Rubisco biogenesis and availability of new tools such as the artificially evolved *E. coli* would transcend the current of barriers Rubisco research and bring mankind another step closer towards improving the kinetics of plant Rubisco and potentially another green revolution.

7 SUMMARY AND CONCLUSION

Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) is the key enzyme that catalyzes the rate limiting carbon fixation step in photosynthesis. In spite of its essential role, Rubisco is an inefficient enzyme with sluggish kinetics and poor substrate discrimination. In addition, Rubisco from many species (including plants) are prone to inhibition by sugar phosphates which include its substrate, Ribulose 1,5-bisphosphate (RuBP). Inhibited Rubisco complexes require remodeling by Rubisco activase to release the inhibitors from the active site. Despite the availability of Rubisco structure from diverse sources and understanding of the reaction mechanism, attempts to improve Rubisco via rational design have been fruitless. Therefore, a directed evolution approach using Rubisco-dependent *E. coli* (RDE) system has been developed to artificially direct the evolution of Rubisco.

In this thesis, the RDE system was re-established with a metabolically engineered MM1 *E. coli* which harbors a phosphoribulokinase (PRK) plasmid (pAC^{BAD}PRK⁶⁸⁰³). PRK catalyzes the production of RuBP which is toxic to the cells. Thus, the presence of functional Rubisco is required for survival of RDE. In addition, Rubisco and activase dependency (RADE) was established using the MM1-*prk* where the presence of both functional and compatible Rubisco and activase were required to support growth in this system. *Rhodobacter sphaeroides*, red-type form IC (*Rs*-Rubisco) and *Acidithiobacillus ferrooxidans*, green-type form IAq Rubisco (*Af*-Rubisco) were evolved in the absence of their activase using the RDE system. Both Rubiscos are known to form inhibited complexes with sugar phosphate and thus associated with activases. In the RDE screening system, these Rubiscos formed inhibited complexes and were unable to confer survival on the host cells. Thus, the directed evolution of these Rubiscos would evolve them away from activase dependency. In addition, *Rs*-Rubisco was evolved in the presence of its activase (*RsCbbX*) using the RADE system. Mutant libraries of these Rubiscos were generated using error-prone PCR and 135 Rubisco variants with improved fitness were isolated from the various screenings. Hence, both RDE and RADE systems were successful in directing the evolution of Rubisco.

41 Rubisco variants from RDE screening with repeated mutated residues were initially characterized using soluble *E. coli* extract. Subsequently, 11 *R*s-Rubisco variants and 10 *A*f-Rubisco variants were selected for detailed characterization using radioactive and spectrophotometric Rubisco assay as their RuBP-inhibited (ER) complexes exhibited a higher rate of activation over time as compared to wild type. In addition, one of the Rubisco variants (*R*s-S1) from RADE screening was also characterized. The Rubisco activation assay revealed that the majority of the variants from RDE screening had gained a new phenotype where they lost the tendency to form stably inhibited complexes with RuBP. In general, this new phenotype came with trade-offs in kinetics where carboxylation activity and carboxylation efficiency are lowered. In addition, affinity towards substrate CO₂ and RuBP were also affected. However, the majority of these variants were able to retain similar CO₂/O₂ specificity values as compared to their wild type. This meant that they were able to discriminate between CO₂ and O₂ with similar capacity as their wild type. Nevertheless, these mutations reduced the enzyme's dependency on their activase. In contrast, the *R*s-S1 variant evolved in presence of its activase, still forms a stable inhibited complex with RuBP and had modest reductions in kinetics. This suggests that the presence of activase would likely restrict Rubisco's evolution from forming an unstable inhibited complex.

The structural analysis of the different mutations that confer the phenotype of forming unstable inhibited complexes show that these mutations were influencing loops that were either involved in substrate binding or active site closure. These mutations were clustered at the 60s loop, the C-terminus and the dimer-dimer interface. The 60s loop contained Thr65 which forms part of the P1-binding site while the C-terminus folds over loop 6 during active site closure. Mutations clustered at dimer-dimer interface were likely influencing substrate binding residues indirectly via common residues or loops. These mutations would likely change the local conformation of these loops such that the active site becomes less rigid and this would facilitate a more rapid release of inhibitors from the active site. Many of these variants also contained single mutations which suggest that Rubisco can be easily evolved away from activase dependency. However, in nature, these mutations have not been selected for. It is likely that the advantages of forming a stable inhibited complex allow for regulation by activase according to the photosynthetic cycle as well as protection from

protease degradation. Thus, selection for mutations that allow Rubisco to escape activase dependency would likely not occur in nature.

Chimeric plant/cyanobacteria (*O. sativa*/*Synechococcus*) Rubisco were generated and expressed in *E. coli*. Although *O. sativa* Rubisco and *Synechococcus* Rubisco (*Syn*-Rubisco) are both form I Rubisco, *Syn*-Rubisco does not form stable inhibited complexes and is also not associated with activase. The *SynL*(*OsL^A*)*S* chimeric Rubisco exhibited kinetic similar to *Os*-Rubisco where it exhibited the fallover effect and formed stable inhibited complexes with RuBP and XuBP. In addition, *SynL*(*OsL^A*)*S* Rubisco can be recognized and remodeled by both *OsRca* and *AtRca*. The directed evolution of this chimeric Rubisco using the RDE system yielded two variants with repeatedly mutated residues. Both variants gained a new phenotype similar to that of the *Rs*-Rubisco and *Af*-Rubisco variants where the Rubiscos no longer form stable inhibited complexes with sugar phosphate. This finding further illustrated that Rubisco that forms inhibited complexes can be easily evolved away from activase dependency.

In this study, we have successfully evolved three different Rubiscos using the RDE and RADE screening systems and shown that Rubisco can easily escape from the Rubisco activase dependency. Rubiscos were artificially evolved using the synthetic biological method of directed evolution where the protein sequence space of Rubisco was explored for variants of improved fitness. Although, the kinetics of the Rubisco variants were not drastically improved, many gain a new phenotype of an increased rate of inhibitor release. The directed evolution system can be used in the future for evolving plant Rubisco in *E. coli* as the requirement for these Rubisco's expression in *E. coli* has been resolved (Aigner et al., 2017). This artificial evolution of plant Rubisco can improve their kinetics and potentially improve carbon dioxide fixation when these evolved variants are transplanted back into plants. In addition, the newly established RADE screening system has great potential to evolve novel compatibility between diverse Rubiscos and activases so as to harness superior properties for improvement of photosynthetic capacity in future plants. Therefore, improvements in Rubisco would bring about increased productivity in the photosynthetic carbon metabolism.

Rubisco has long been targeted for modification either through improving photosynthetic carbon metabolism or reducing the propensity of oxygenation. The oxygenation reaction of Rubisco produces 2PG which can be recycled back to 3PGA via the photorespiratory pathway. However, this photorespiratory pathway is detrimental to photosynthetic carbon metabolism as it releases a previously fixed CO_2 and consumes ATP and NADPH. Numerous studies have shown that the canonical photorespiratory pathway can be bypassed by introducing synthetic glycolate metabolism pathways and some of these pathways were also able to abolish the CO_2 release (Weber et al., 2019). Recent advances by South et al. (2019) demonstrated the engineering of an enhanced photorespiratory pathway by introducing malate synthase and glycolate dehydrogenase into the tobacco chloroplast. At the same time, RNA interference module was also introduced to inhibit the native chloroplast glycolate transporter. Instead of the native photorespiratory pathway, inhibition of the glycolate transporter would drive the flux of glycolate towards the artificial pathway. This alternative pathway increased the local availability of CO_2 concentration in the chloroplast resulted in increased maximum carboxylation rate of Rubisco. Consequently, improving photosynthetic efficiency and biomass. Taken together by coupling a kinetically superior Rubisco with bypassing of photorespiration, photosynthetic efficiency of plants could be potentially improved.

Correspondingly, other recent radical approaches to improve photosynthetic carbon metabolism have been demonstrated. This includes the exploration of a novel CO_2 fixing pathway or redesigning a synthetic novel CO_2 fixation pathway such as the CETCH cycle (Erb et al., 2017; Schwander et al., 2016). The CETCH cycle was designed to bypass the kinetic limitations of Rubisco by utilizing a combination of 17 known and engineered enzymes to carried CO_2 fixation into organic compounds. Moreover, the CETCH cycle turns at a rate comparable to the CBB cycle. However, the feasibility of implementing the CETCH cycle in a cell has yet to be explored as the metabolic networks are highly integrated and further adaption may be required (Erb et al., 2017). Overall, these efforts are targeted at improvements in carbon metabolism either by improving Nature's CBB cycle or implementation of novel pathways that are vital in ensuring the future sustainability in terms of food security, biofuel, and commodity chemicals.

In conclusion, the utilization of rational and evolutionary methods to improve Rubisco has provided insightful information on the fundamental principles that govern the shortcoming of Rubisco and its protein-protein interactions with accessory proteins. The strategies and systems described here will contribute to the comprehensive Rubisco engineering toolbox that is emerging and will be applied towards the synthetic biology of CO₂ fixation (Conlan et al., 2018). It is thus obvious that our food and bioenergy crops will soon be utilizing carefully tuned carbon fixation modules to overcome the constraints inherent in this process, which will lead to enhanced photosynthetic efficiency (Ort et al., 2015; Zhu et al., 2010).

8 APPENDIX

8.1 Appendix I: List form I Rubisco from the phylogenetic tree

Accession code (NCBI or Uniprot) for phylogenetic tree (Figure 1-3) of the various form I Rubisco listed in the table below.

Species	Accession code
Form IA (Prokaryote):	
<i>M. purpuratum</i> , <i>Marichromatium purpuratum</i>	W0E6T9
<i>H. marinus</i> , <i>Hydrogenovibrio marinus</i>	Q59458
<i>T. crunogena</i> , <i>Thiomicrospira crunogena</i>	Q31IK0
<i>T. denitrificans</i> , <i>Thiobacillus denitrificans</i>	Q56259
<i>A. ferrooxidans</i> , <i>Acidithiobacillus ferrooxidans</i>	B7JA24
<i>N. europaea</i> , <i>Nitrosomonas europaea</i>	Q82TG6
<i>H. neapolitanus</i> , <i>Halothiobacillus neapolitanus</i>	O85040
<i>T. intermedia</i> , <i>Thiomonas intermedia</i>	Q9ZH71
<i>T. mobilis</i> , <i>Thioflavicoccus mobilis</i>	L0GWD2
<i>A. vinosum</i> , <i>Allochromatium vinosum</i>	P22849
<i>T. marina</i> , <i>Thiocapsa marina</i>	F9U578
Form IB (Eukaryote):	
<i>P. vulgaris</i> , <i>Phaseolus vulgaris</i>	A4GG89
<i>L. japonicus</i> , <i>Lotus japonicus</i>	Q9BBU1
<i>S. oleracea</i> , <i>Spinacia oleracea</i>	P00875
<i>G. hirsutum</i> , <i>Gossypium hirsutum</i>	P14958
<i>Z. mays</i> , <i>Zea mays</i>	P00874
<i>T. aestivum</i> , <i>Triticum aestivum</i>	P11383
<i>O. sativa</i> , <i>Oryza sativa</i>	P0C512
<i>N. tabacum</i> , <i>Nicotiana tabacum</i>	P00876
<i>B. oleracea</i> , <i>Brassica oleracea</i>	P48686
<i>A. thaliana</i> , <i>Arabidopsis thaliana</i>	O03042
<i>C. reinhardtii</i> , <i>Chlamydomonas reinhardtii</i>	P00877
Form IB (Prokaryote):	
<i>Syn. PCC6803</i> , <i>Synechocystis PCC6803</i>	P54205
<i>Syn. PCC7002</i> , <i>Synechococcus PCC7002</i>	Q44176
<i>Syn. PCC6301</i> , <i>Synechococcus PCC6301</i>	P00880
<i>Syn. PCC7502</i> , <i>Synechococcus sp. PCC 7502</i>	K9SU87
<i>F. contorta</i> , <i>Fortiea contorta</i>	WP_017655457.1
<i>N. punctiforme</i> , <i>Nostoc punctiforme</i>	B2J8T2
<i>C. stagnale</i> , <i>Cylindrospermum stagnale</i>	K9WVP9
Form IC (Prokaryote):	
<i>R. pickettii</i> , <i>Ralstonia pickettii</i>	U3QSC3
<i>R. eutropha</i> , <i>Ralstonia eutropha</i>	P0C2C2
<i>A. methanolica</i> , <i>Acidomonas methanolica</i>	A0A023D5D1
<i>R. sphaeroides</i> , <i>Rhodobacter sphaeroides</i>	Q3IX55

Table A-1: List of form I Rubisco and their accession code.

8.2 Appendix II: Combination of selective media for optimizing RADE screening

Plasmids in Table 2-3 were transformed into MM1-pAC^{BAD}PRK^{DM} and screened in the various selective media plates to optimized the media composition for RADE dependency. Selective media plates were placed either in ambient atmospheric conditions (Air- 0.04% CO₂) and grown at room temperature or in CO₂ incubator elevated CO₂ (Air supplemented with 2% CO₂) at 23°C. Images of screening with selective media in the appendix were taken after 10 days. Absence of image in screening under air will indicate no growth.

	M9 minimal media with antibiotics and supplemented with the following:
Selective media A	0.05% (w/v) arabinose, 0.05% (w/v) casamino acid, 0.25 mM IPTG, 0.01% (w/v) malate
Selective media B	0.05% (w/v) arabinose, 0.2% (w/v) casamino acid, 0.5 mM IPTG
Selective media C	0.05% (w/v) arabinose, 0.05% (w/v) casamino acid, 0.25 mM IPTG, 0.005% (w/v) malate
Selective media D	0.05% (w/v) arabinose, 0.1% (w/v) casamino acid
Selective media E	0.10% (w/v) arabinose, 0.1% (w/v) casamino acid
Selective media F	0.15% (w/v) arabinose, 0.1% (w/v) casamino acid
Selective media G	0.05% (w/v) arabinose, 0.05% (w/v) casamino acid, 0.005% (w/v) malate

Table A-2: Composition of different selective media.

M9 minimal media is composed of M9 salt (6% (w/v) Na₂HPO₄, 3% (w/v) KH₂PO₄, 0.5% (w/v) NaCl, 1% (w/v) NH₄Cl), 1 mM thiamine-HCl, 0.1 mM CaCl₂, 2 mM MgSO₄, 0.4% (v/v) glycerol. The various selective media is composed of M9 minimal media supplemented with antibiotics (Amp, Kan, and Chl) and the respective amount of arabinose, casamino acid, IPTG and malate as indicated in the Table A-1.

Selective media A

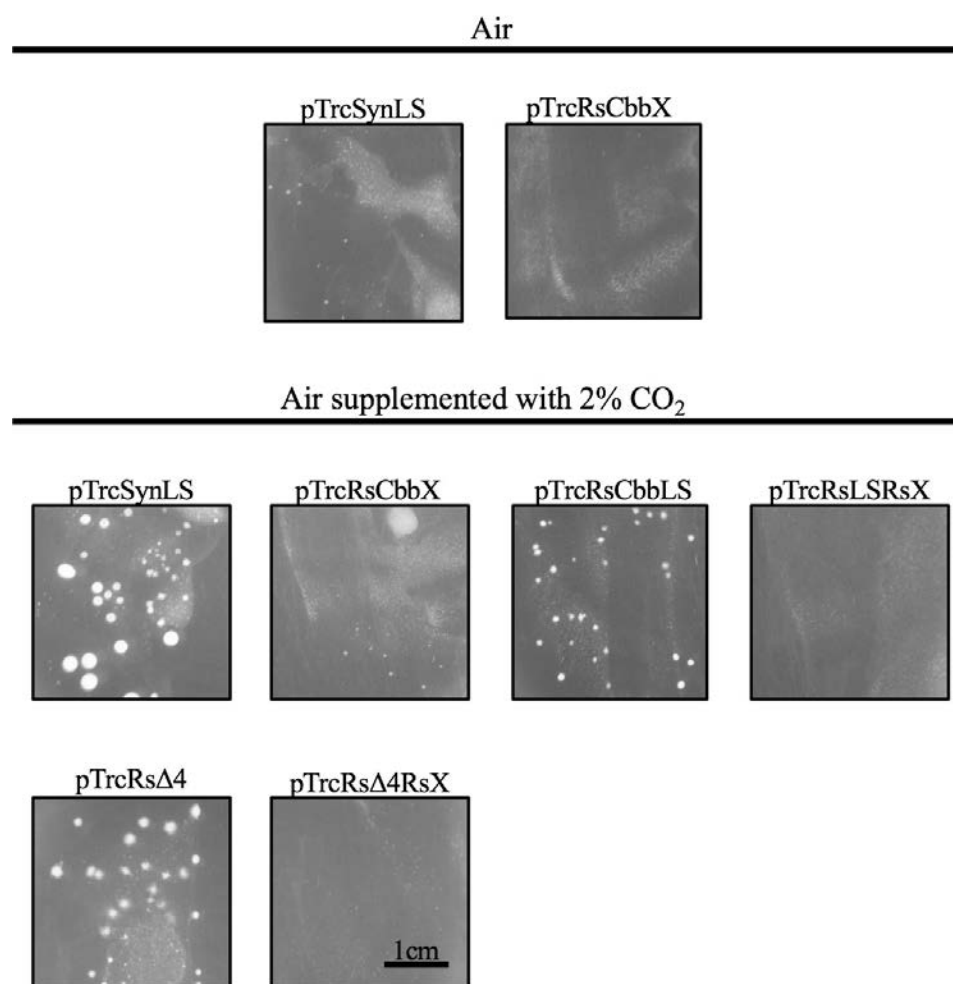


Figure A-1: Selective media A.

Selective media A is M9 minimal media with antibiotics and supplemented 0.05% (w/v) arabinose, 0.05% (w/v) casamino acid, 0.25 mM IPTG, 0.01% (w/v) malate.

Selective media B

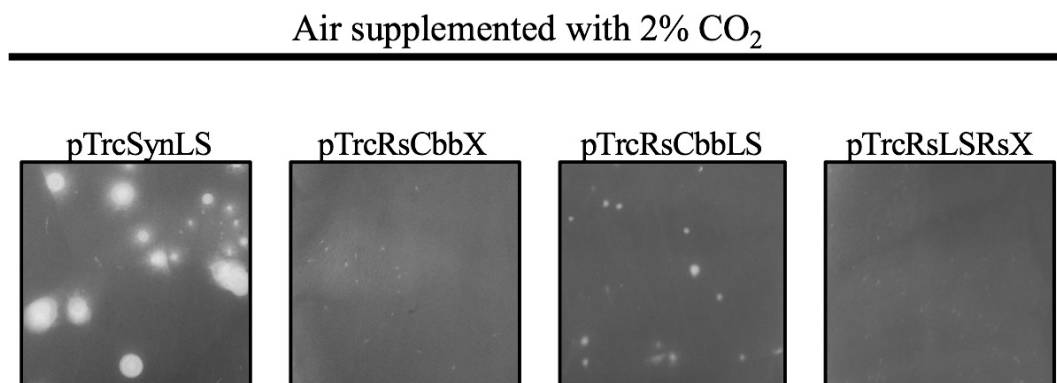


Figure A-2: Selective media B.

Selective media B is M9 minimal media with antibiotics and supplemented with 0.05% (w/v) arabinose, 0.2% (w/v) casamino acid, 0.5 mM IPTG.

Selective media C

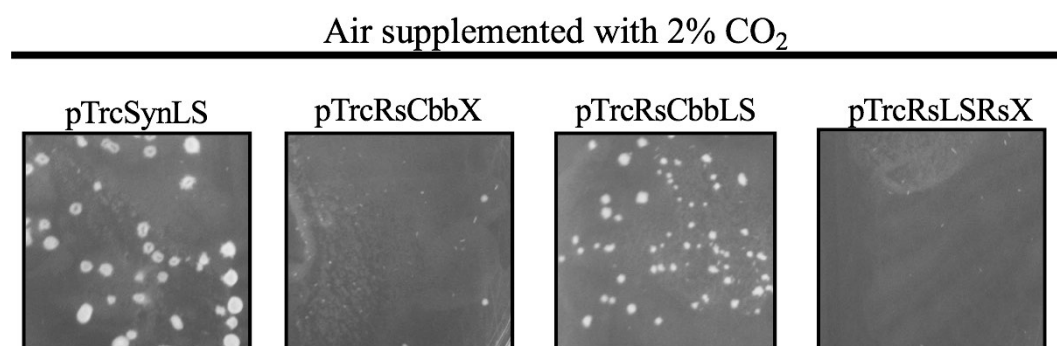


Figure A-3: Selective media C.

Selective media C is M9 minimal media with antibiotics and supplemented with 0.05% (w/v) arabinose, 0.05% (w/v) casamino acid, 0.25 mM IPTG, 0.005% (w/v) malate

Selective media D

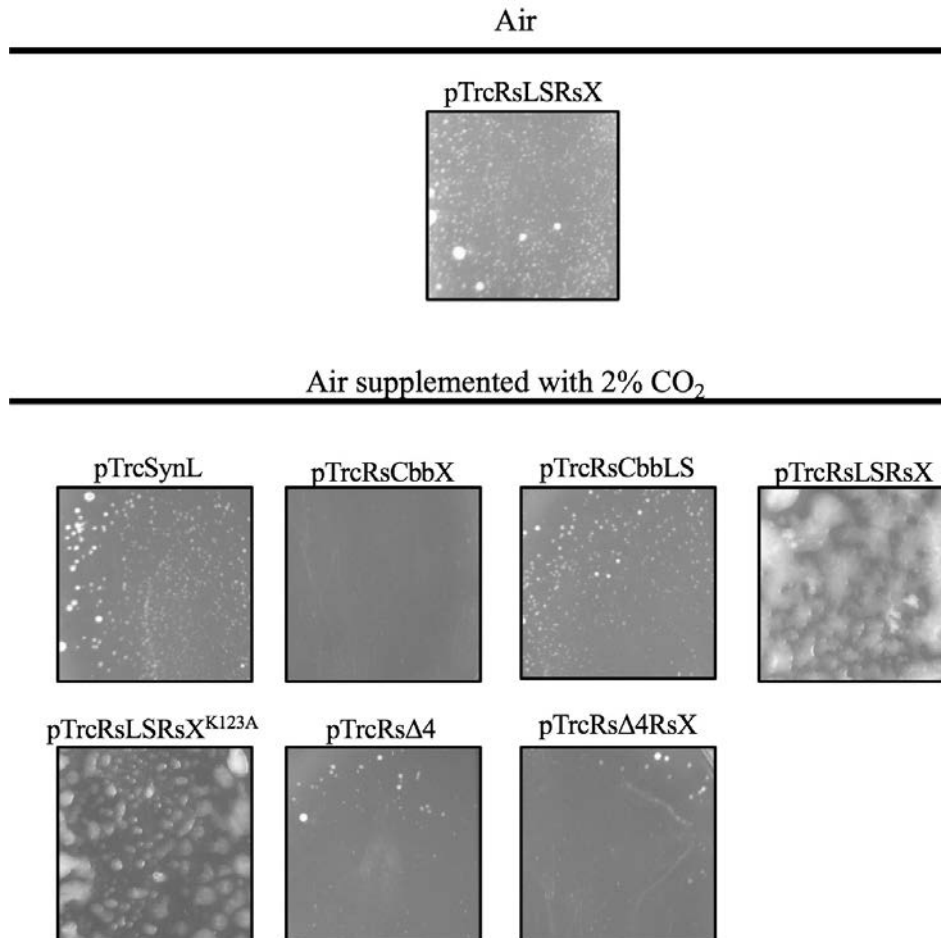


Figure A-4: Selective media D.

Selective media D is M9 minimal media with antibiotics and supplemented with 0.05% (w/v) arabinose, 0.1% (w/v) casamino acid.

Selective media E

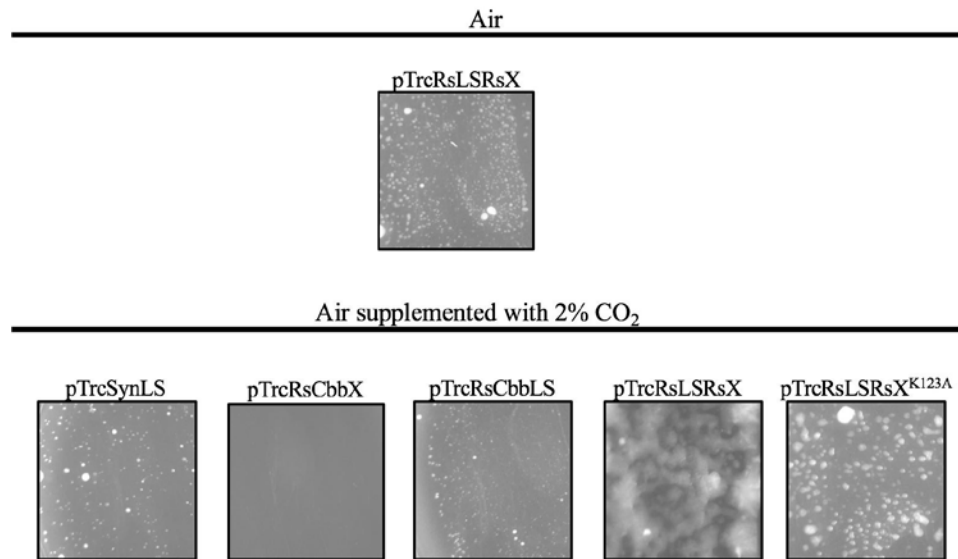


Figure A-5: Selective media E

Selective media E is M9 minimal media with antibiotics and supplemented with 0.10% (w/v) arabinose, 0.1% (w/v) casamino acid.

Selective media F

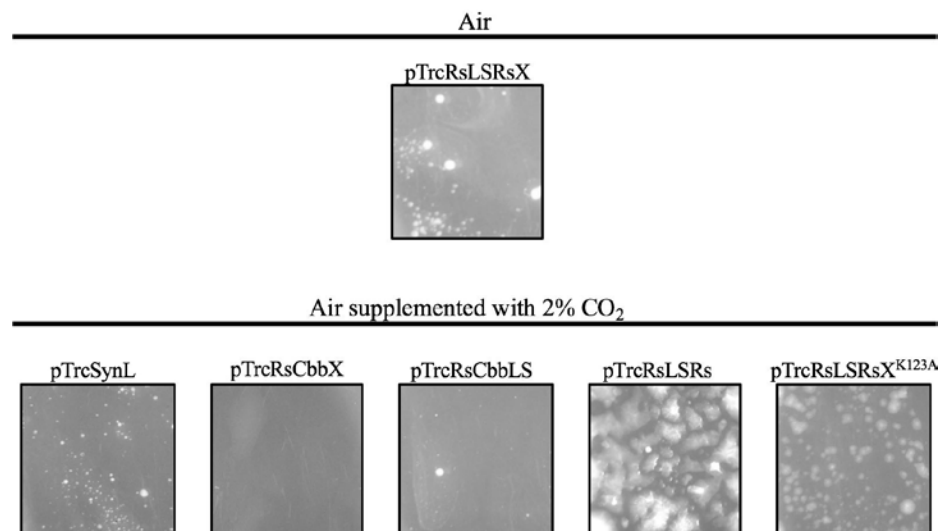


Figure A-6: Selective media F

Selective media F is M9 minimal media with antibiotics and supplemented with 0.15% (w/v) arabinose, 0.1% (w/v) casamino acid.

Selective media G

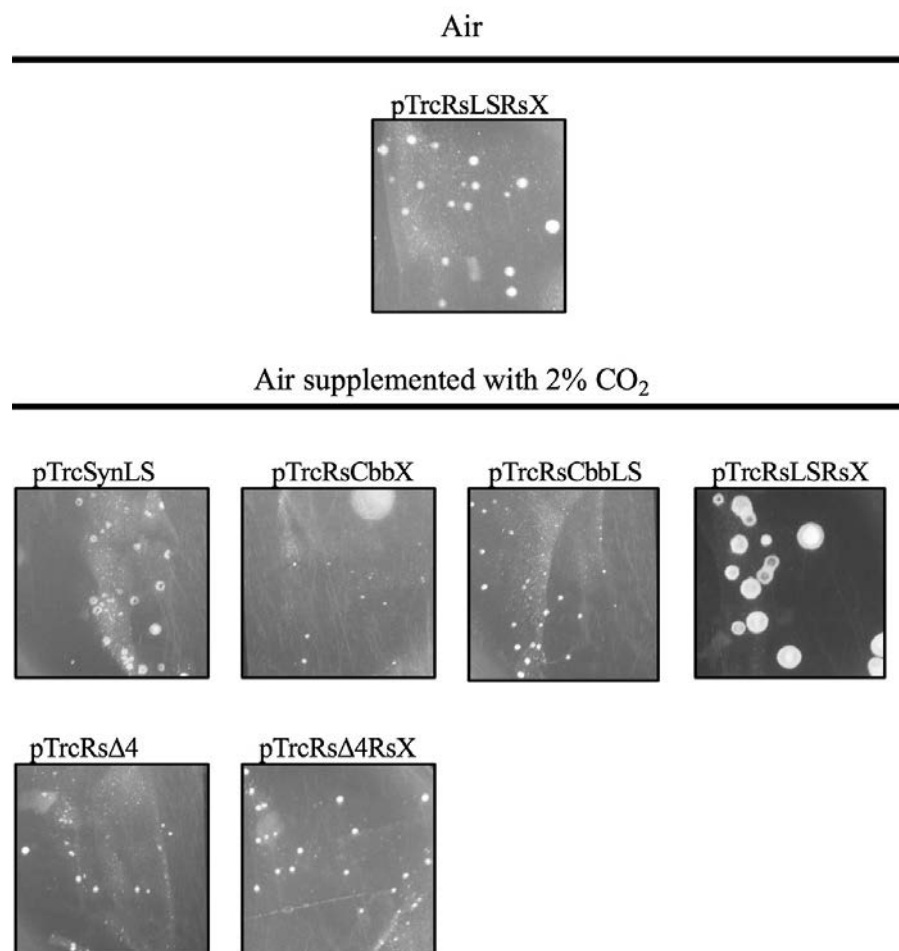


Figure A-7: Selective media G

Selective media G is M9 minimal media with antibiotics and supplemented with 0.05% (w/v) arabinose, 0.05% (w/v) casamino acid, 0.005% (w/v) malate. The unusual morphology is of pTrcRsLRSX is related to advanced stage of colony growth, and is not due to contamination.

8.3 Appendix III: Reoccurring mutations between *Rs*-Rubisco and *Af*-Rubisco variants

The evolved Rubisco variant of *Rs*-Rubisco and *Af*-Rubisco from RDE and RADE screenings had several common mutated residues (Table A-3). The numbering of the rubisco large subunit of these two Rubisco are different and their equivalent numbering are shown in the table. Firstly, from the RDE screenings of both Rubisco, there were three repeatedly mutated large subunit residues (Table A-2, green section). This suggest that different type of Rubisco (red-type and green-type) could be evolved to use similar solutions to improve fitness for surviving RDE screening system and these non-characterized variants are potential subjects for future studies. Secondly, the directed evolution of *Rs*-Rubisco from RDE and RADE screening, had three repeatedly mutated small subunit residues (Table A-3, blue section). Lastly, *Af*-Rubisco variants from RDE screening and *Rs*-Rubisco variants from RADE screening had five repeatedly mutated residues from the large subunit (Table A-3, grey section).

Mutant Rubisco		Amino acid change(s)		Repeating mutated residue(s)
		RbcL	RbcS	
Common mutations between <i>Rs</i> -Rubisco and <i>Af</i> -Rubisco variants from RDE screening				
<i>Rs</i> -R1*	RsL ^{P52L} S	P52L	nil	P52L ^a
<i>Af</i> -A9	AfL ^{R50P/L180M} S	R50P, L180M	nil	R50P ^a
<i>Rs</i> -R18*	RsL ^{L107M} S	L107M	nil	L107M/Q ^a
<i>Rs</i> -R19*	RsL ^{L107Q} S	L107Q	nil	
<i>Af</i> -B1*	AfL ^{K41R/I105V} S	K41R, I105V	nil	I105V ^a
<i>Rs</i> -R20	RsL ^{T31/D397N} S ^{M55I}	T3I, D397N	M55I	D397N ^a
<i>Af</i> -B24	AfL ^{D396A/E481G} S	D396A, E468G	nil	D396A ^a
Common mutations between <i>Rs</i> -Rubisco variants from RDE and RADE screening				
<i>Rs</i> -T9	RsL ^{F478I} S ^{I66N}	F478I	I66N	I66N/T ^b
<i>Rs</i> -U9	RsL ^{T356S} S ^{D16V/I66T/N101Y}	T356S	D16V, I66T, N101Y	
<i>Rs</i> -R10*	RsLS ^{F91I/T95S/S108C}	nil	F91I, T95S, S108C	S108C/G ^b
<i>Rs</i> -U2	RsLS ^{S108G}	nil	S108G	
<i>Rs</i> -R22	RsLS ^{I121F}	nil	I121F	I121F/V ^b
<i>Rs</i> -R14.1	RsL ^{Y26F} S ^{I121V}	Y26F	I121V	
<i>Rs</i> -S3.2	RsL ^{K148E/N348D} S ^{I121F}	K148E, N348D	I121F	
Common mutations between <i>Rs</i> -Rubisco variants from RADE screening and <i>Af</i> -Rubisco variants from RDE screening				
<i>Rs</i> -S3.1	RsL ^{A84V/K148E/V349D} S	A84V, K148E, V349D	nil	K148E ^a
<i>Rs</i> -S3.2	RsL ^{K148E/N348D} S ^{I121F}	K148E, N348D	I121F	
<i>Af</i> -B21*	AfL ^{K146M} S ^{V112A}	K146M	V112A	K146M ^a
<i>Rs</i> -U8	RsL ^{Y99N/A332T} S	Y99N, A332T	nil	A332T ^a
<i>Af</i> -A1	AfL ^{V331A} S ^{G44D}	V331A	G44D	V331A ^a
<i>Rs</i> -U9	RsL ^{T356S} S ^{D16V/I66T/N101Y}	T356S	D16V, I66T, N101Y	T356S ^a
<i>Af</i> -A13	AfL ^{K355T} S	K355T	nil	K355T ^a
<i>Rs</i> -U4	RsL ^{D2V/D439E/V440A} S ^{V128G}	D2V, D439E, V440A	V128G	D439E ^a
<i>Af</i> -B11*	AfL ^{E438A} S	E438A	nil	E438A/G ^a
<i>Af</i> -B12*	AfL ^{E438G} S	E438G	nil	
<i>Rs</i> -U6	RsL ^{N469S} S	N469S	nil	N469S ^a
<i>Af</i> -B24	AfL ^{D396A/E481G} S	D396A, E468G	nil	E468G/V ^a
<i>Af</i> -B25	AfL ^{F40Y/Q209L/E468V} S	F40Y, Q209L, E468V	nil	

Table A-3: Repeatedly mutated residues between *Rs*-Rubisco and *Af*-Rubisco variants from RDE and RADE screening.

Rs-R'X', *Rs*-T'X' are *Rs*-Rubisco variants were obtained RDE screening while *Rs*-S'X' and *Rs*-U'X' are *Rs*-Rubisco variants were obtained RADE screening. *Af*-A'X' and *Af*-B'X' are *Af*-Rubisco variants obtained from RDE screening. ^a: Repeating mutated residue is located in the Rubisco large subunit. ^b: Repeating mutated residue is located in the Rubisco small subunit. *: These Rubisco variants were characterized.

8.4 Appendix IV: List of publication

- Guo, ZJ., Whitney, S., Mueller-cajar, O., (2019). 'Evolutionary escape from activase requirement of form I Rubisco', Manuscript in preparation.

9 REFERENCES

- Aigner, H., Wilson, R. H., Bracher, A., Calisse, L., Bhat, J. Y., Hartl, F. U., & Hayer-Hartl, M. (2017). Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science*, 358(6368), 1272-1278. doi:10.1126/science.aap9221
- Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO₂ enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO₂. *New Phytol*, 165(2), 351-371. doi:10.1111/j.1469-8137.2004.01224.x
- Alonso, H., Blayney, M. J., Beck, J. L., & Whitney, S. M. (2009). Substrate-induced assembly of *Methanococcoides burtonii* D-ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. *J Biol Chem*, 284(49), 33876-33882. doi:10.1074/jbc.M109.050989
- Amichay, D., Levitz, R., & Gurevitz, M. (1993). Construction of a *Synechocystis* PCC6803 mutant suitable for the study of variant hexadecameric ribulose bisphosphate carboxylase/oxygenase enzymes. *Plant Mol Biol*, 23(3), 465-476.
- Andersson, I. (1996). Large structures at high resolution: the 1.6 Å crystal structure of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate. *J Mol Biol*, 259(1), 160-174. doi:10.1006/jmbi.1996.0310
- Andersson, I. (2008). Catalysis and regulation in Rubisco. *J Exp Bot*, 59(7), 1555-1568. doi:10.1093/jxb/ern091
- Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiol Biochem*, 46(3), 275-291. doi:10.1016/j.plaphy.2008.01.001
- Andrews, T. J. (1996). The bait in the Rubisco mousetrap. *Nat Struct Biol*, 3(1), 3-7.
- Andrews, T. J., & Abel, K. M. (1981). Kinetics and subunit interactions of ribulose bisphosphate carboxylase-oxygenase from the cyanobacterium, *Synechococcus* sp. *J Biol Chem*, 256(16), 8445-8451.
- Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., . . . Milo, R. (2016). Sugar Synthesis from CO₂ in *Escherichia coli*. *Cell*, 166(1), 115-125. doi:10.1016/j.cell.2016.05.064
- Aono, R., Sato, T., Imanaka, T., & Atomi, H. (2015). A pentose bisphosphate pathway for nucleoside degradation in Archaea. *Nat Chem Biol*, 11(5), 355-360. doi:10.1038/nchembio.1786

- Ashida, H., Saito, Y., Kojima, C., Kobayashi, K., Ogasawara, N., & Yokota, A. (2003). A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science*, 302(5643), 286-290. doi:10.1126/science.1086997
- Ashida, H., Saito, Y., Nakano, T., Tandeau de Marsac, N., 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. *J Exp Bot*, 59(7), 1543-1554. doi:10.1093/jxb/ern104
- Badger, M. R., Andrews, T. J., Whitney, S. M., Ludwig, M., Yellowlees, D. C., Leggat, W., & Price, G. D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO₂-concentrating mechanisms in algae. *Canadian Journal of Botany*, 76(6), 1052-1071. doi:10.1139/b98-074
- Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J Exp Bot*, 59(7), 1525-1541. doi:10.1093/jxb/ern297
- Bauwe, H. (2010). Recent developments in photorespiration research. *Biochem Soc Trans*, 38(2), 677-682. doi:10.1042/BST0380677
- Betts, M. J., & Russell, R. B. (2003). Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists*, 317, 289.
- Bhat, J. Y., Milicic, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., . . . Hayer-Hartl, M. (2017a). Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Mol Cell*. doi:10.1016/j.molcel.2017.07.004
- Bhat, J. Y., Thieulin-Pardo, G., Hartl, F. U., & Hayer-Hartl, M. (2017b). Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Front Mol Biosci*, 4, 20. doi:10.3389/fmolb.2017.00020
- Bracher, A., Hauser, T., Liu, C., Hartl, F. U., & Hayer-Hartl, M. (2015). Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from *Chlamydomonas reinhardtii*. *PLoS One*, 10(8), e0135448. doi:10.1371/journal.pone.0135448
- Bracher, A., Whitney, S. M., Hartl, F. U., & Hayer-Hartl, M. (2017). Biogenesis and Metabolic Maintenance of Rubisco. *Annu Rev Plant Biol*, 68, 29-60. doi:10.1146/annurev-arplant-043015-111633
- Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U., & Hayer-Hartl, M. (2001). Dual Function of Protein Confinement in Chaperonin-Assisted Protein Folding. *Cell*, 107(2), 223-233. doi:10.1016/s0092-8674(01)00517-7
- Chen, Z. X., & Spreitzer, R. J. (1989). Chloroplast intragenic suppression enhances the low CO₂/O₂ specificity of mutant ribulose-bisphosphate carboxylase/oxygenase. *Journal of Biological Chemistry*, 264(6), 3051-3053.

- Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C., & Lorimer, G. H. (1998). Mechanism of Rubisco: The Carbamate as General Base. *Chem Rev*, 98(2), 549-562.
- Cloney, L. P., Bekkaoui, D. R., & Hemmingsen, S. M. (1993). Co-expression of plastid chaperonin genes and a synthetic plant Rubisco operon in *Escherichia coli*. *Plant Mol Biol*, 23(6), 1285-1290.
- Codd, G. A., & Marsden, W. J. N. (1984). The carboxysomes (polyhedral bodies) of autotrophic prokaryotes. *Biological Reviews*, 59(3), 389-422. doi:doi:10.1111/j.1469-185X.1984.tb00710.x
- Conlan, B., & Whitney, S. (2018). Preparing Rubisco for a tune up. *Nature plants*, 4(1), 12.
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, 16(22), 10881-10890. doi:10.1093/nar/16.22.10881
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res*, 14(6), 1188-1190. doi:10.1101/gr.849004
- Dangel, A. W., & Tabita, F. R. (2015). CbbR, the Master Regulator for Microbial Carbon Dioxide Fixation. *J Bacteriol*, 197(22), 3488-3498. doi:10.1128/JB.00442-15
- Duff, A. P., Andrews, T. J., & Curmi, P. M. (2000). The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J Mol Biol*, 298(5), 903-916. doi:10.1006/jmbi.2000.3724
- Durao, P., Aigner, H., Nagy, P., Mueller-Cajar, O., Hartl, F. U., & Hayer-Hartl, M. (2015). Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat Chem Biol*, 11(2), 148-155. doi:10.1038/nchembio.1715
- Edgerton, M. D. (2009). Increasing crop productivity to meet global needs for feed, food, and fuel. *Plant Physiol*, 149(1), 7-13. doi:10.1104/pp.108.130195
- Edmondson, D. L., Badger, M. R., & Andrews, T. J. (1990). Slow Inactivation of Ribulosebisphosphate Carboxylase during Catalysis Is Caused by Accumulation of a Slow, Tight-Binding Inhibitor at the Catalytic Site. *Plant Physiol*, 93(4), 1390-1397. doi:10.1104/pp.93.4.1390
- Ellis, R. J. (1979). The most abundant protein in the world. *Trends in Biochemical Sciences*, 4(11), 241-244. doi:10.1016/0968-0004(79)90212-3
- Emlyn-Jones, D., Woodger, F. J., Price, G. D., & Whitney, S. M. (2006). RbcX can function as a rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol*, 47(12), 1630-1640. doi:10.1093/pcp/pcl028

- Erb, T. J., Evans, B. S., Cho, K., Warlick, B. P., Sriram, J., Wood, B. M., . . . Gerlt, J. A. (2012). A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nat Chem Biol*, 8(11), 926-932. doi:10.1038/nchembio.1087
- Erb, T. J., & Zarzycki, J. (2017). A short history of RubisCO: the rise and fall (?) of Nature's predominant CO₂ fixing enzyme. *Curr Opin Biotechnol*, 49, 100-107. doi:10.1016/j.copbio.2017.07.017
- Esquivel, M. G., Pinto, T. S., Marin-Navarro, J., & Moreno, J. (2006). Substitution of tyrosine residues at the aromatic cluster around the betaA-betaB loop of rubisco small subunit affects the structural stability of the enzyme and the in vivo degradation under stress conditions. *Biochemistry*, 45(18), 5745-5753. doi:10.1021/bi052588y
- Ewalt, K. L., Hendrick, J. P., Houry, W. A., & Hartl, F. U. (1997). In Vivo Observation of Polypeptide Flux through the Bacterial Chaperonin System. *Cell*, 90(3), 491-500. doi:10.1016/s0092-8674(00)80509-7
- Farr, G. W., Furtak, K., Rowland, M. B., Ranson, N. A., Saibil, H. R., Kirchhausen, T., & Horwich, A. L. (2000). Multivalent Binding of Nonnative Substrate Proteins by the Chaperonin GroEL. *Cell*, 100(5), 561-573. doi:10.1016/s0092-8674(00)80692-3
- Feiz, L., Williams-Carrier, R., Belcher, S., Montano, M., Barkan, A., & Stern, D. B. (2014). A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J*, 80(5), 862-869. doi:10.1111/tpj.12686
- Feiz, L., Williams-Carrier, R., Wostrikoff, K., Belcher, S., Barkan, A., & Stern, D. B. (2012). Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell*, 24(8), 3435-3446. doi:10.1105/tpc.112.102012
- Feller, U., Anders, I., & Mae, T. (2008). Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. *J Exp Bot*, 59(7), 1615-1624. doi:10.1093/jxb/erm242
- Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 281(5374), 237-240.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server*: Springer.
- Genkov, T., Meyer, M., Griffiths, H., & Spreitzer, R. J. (2010). Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS

- cDNA for expression in chlamydomonas. *J Biol Chem*, 285(26), 19833-19841. doi:10.1074/jbc.M110.124230
- Genkov, T., & Spreitzer, R. J. (2009). Highly conserved small subunit residues influence rubisco large subunit catalysis. *J Biol Chem*, 284(44), 30105-30112. doi:10.1074/jbc.M109.044081
- Gibson, J. L., & Tabita, F. R. (1977). Different molecular forms of D-ribulose-1, 5-bisphosphate carboxylase from *Rhodospseudomonas sphaeroides*. *Journal of Biological Chemistry*, 252(3), 943-949.
- Gibson, J. L., & Tabita, F. R. (1979). Activation of ribulose 1,5-bisphosphate carboxylase from *Rhodospseudomonas sphaeroides*: probable role of the small subunit. *J Bacteriol*, 140(3), 1023-1027.
- Gibson, J. L., & Tabita, F. R. (1997). Analysis of the cbbXYZ operon in *Rhodobacter sphaeroides*. *J Bacteriol*, 179(3), 663-669.
- Giessen, T. W., & Silver, P. A. (2017). Engineering carbon fixation with artificial protein organelles. *Curr Opin Biotechnol*, 46, 42-50. doi:10.1016/j.copbio.2017.01.004
- Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., . . . Varma, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science*, 296(5565), 92-100.
- Goloubinoff, P., Christeller, J. T., Gatenby, A. A., & Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature*, 342(6252), 884-889. doi:10.1038/342884a0
- Greene, D. N., Whitney, S. M., & Matsumura, I. (2007). Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. *Biochem J*, 404(3), 517-524. doi:10.1042/BJ20070071
- Gutteridge, S., Rhoades, D. F., & Herrmann, C. (1993). Site-specific mutations in a loop region of the C-terminal domain of the large subunit of ribulose bisphosphate carboxylase/oxygenase that influence substrate partitioning. *Journal of Biological Chemistry*, 268(11), 7818-7824.
- Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol*, 177(14), 4121-4130.
- Hanson, D. T. (2016). Breaking the rules of Rubisco catalysis. *J Exp Bot*, 67(11), 3180-3182. doi:10.1093/jxb/erw197
- Hanson, P. I., & Whiteheart, S. W. (2005). AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol*, 6(7), 519-529. doi:10.1038/nrm1684

- Hanson, T. E., & Tabita, F. R. (2001). A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *Proc Natl Acad Sci U S A*, 98(8), 4397-4402. doi:10.1073/pnas.081610398
- Hartl, F. U., & Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, 295(5561), 1852-1858. doi:10.1126/science.1068408
- Hasse, D., Larsson, A. M., & Andersson, I. (2015). Structure of *Arabidopsis thaliana* Rubisco activase. *Acta Crystallogr D Biol Crystallogr*, 71(Pt 4), 800-808. doi:10.1107/S1399004715001182
- Hauser, T., Bhat, J. Y., Milicic, G., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2015a). Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat Struct Mol Biol*, 22(9), 720-728. doi:10.1038/nsmb.3062
- Hauser, T., Popilka, L., Hartl, F. U., & Hayer-Hartl, M. (2015b). Role of auxiliary proteins in Rubisco biogenesis and function. *Nat Plants*, 1, 15065. doi:10.1038/nplants.2015.65
- Henderson, J. N., Kuriata, A. M., Fromme, R., Salvucci, M. E., & Wachter, R. M. (2011). Atomic resolution x-ray structure of the substrate recognition domain of higher plant ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase. *J Biol Chem*, 286(41), 35683-35688. doi:10.1074/jbc.C111.289595
- Heureux, A. M. C., Young, J. N., Whitney, S. M., Eason-Hubbard, M. R., Lee, R. B. Y., Sharwood, R. E., & Rickaby, R. E. M. (2017). The role of Rubisco kinetics and pyrenoid morphology in shaping the CCM of haptophyte microalgae. *J Exp Bot*. doi:10.1093/jxb/erx179
- Horken, K. M., & Tabita, F. R. (1999). The "green" form I ribulose 1,5-bisphosphate carboxylase/oxygenase from the nonsulfur purple bacterium *Rhodospirillum rubrum*. *J Bacteriol*, 181(13), 3935-3941.
- Houtz, R. L., & Mulligan, R. M. (1991). Protection of Tryptic-Sensitive Sites in the Large Subunit of Ribulosebisphosphate Carboxylase/Oxygenase by Catalysis. *Plant Physiol*, 96(1), 335-339. doi:10.1104/pp.96.1.335
- Hudson, G. S., Morell, M. K., Arvidsson, Y. B. C., & Andrews, T. J. (1992). Synthesis of Spinach Phosphoribulokinase and Ribulose 1,5-Bisphosphate in *Escherichia coli*. *Australian Journal of Plant Physiology*, 19(3), 213. doi:10.1071/pp9920213
- Joshi, J., Mueller-Cajar, O., Tsai, Y. C., Hartl, F. U., & Hayer-Hartl, M. (2015). Role of small subunit in mediating assembly of red-type form I Rubisco. *J Biol Chem*, 290(2), 1066-1074. doi:10.1074/jbc.M114.613091

- Kane, H. J., Viil, J., Entsch, B., Paul, K., Morell, M. K., & Andrews, T. J. (1994). An Improved Method for Measuring the CO₂/O₂ Specificity of Ribulosebisphosphate Carboxylase-Oxygenase. *Australian Journal of Plant Physiology*, 21(4). doi:10.1071/pp9940449
- Kane, H. J., Wilkin, J.-M., Portis, A. R., & John Andrews, T. (1998). Potent Inhibition of Ribulose-Bisphosphate Carboxylase by an Oxidized Impurity in Ribulose-1,5-Bisphosphate. *Plant Physiol*, 117(3), 1059-1069. doi:10.1104/pp.117.3.1059
- Karkehabadi, S., Peddi, S. R., Anwaruzzaman, M., Taylor, T. C., Cederlund, A., Genkov, T., . . . Spreitzer, R. J. (2005). Chimeric small subunits influence catalysis without causing global conformational changes in the crystal structure of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry*, 44(29), 9851-9861. doi:10.1021/bi050537v
- Karkehabadi, S., Satagopan, S., Taylor, T. C., Spreitzer, R. J., & Andersson, I. (2007). Structural analysis of altered large-subunit loop-6/carboxy-terminus interactions that influence catalytic efficiency and CO₂/O₂ specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry*, 46(39), 11080-11089. doi:10.1021/bi701063f
- Khan, S., Andralojc, P. J., Lea, P. J., & Parry, M. A. J. (1999). 2'-Carboxy-D-arabinitol 1-phosphate protects ribulose 1,5-bisphosphate carboxylase/oxygenase against proteolytic breakdown. *European Journal of Biochemistry*, 266(3), 840-847. doi:10.1046/j.1432-1327.1999.00913.x
- King, W. A., Gready, J. E., & Andrews, T. J. (1998). Quantum chemical analysis of the enolization of ribulose bisphosphate: the first hurdle in the fixation of CO₂ by Rubisco. *Biochemistry*, 37(44), 15414-15422. doi:10.1021/bi981598e
- Knight, S., Andersson, I., & Brändén, C.-I. (1990). Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. *J Mol Biol*, 215(1), 113-160. doi:10.1016/s0022-2836(05)80100-7
- Koay, T. W., Wong, H. L., & Lim, B. H. (2016). Engineering of chimeric eukaryotic/bacterial Rubisco large subunits in Escherichia coli. *Genes Genet Syst*, 91(3), 139-150. doi:10.1266/ggs.15-00054
- Kobayashi, D. (2003). Molecular Characterization and Redox Regulation of Phosphoribulokinase from the Cyanobacterium Synechococcus sp. PCC 7942. *Plant and Cell Physiology*, 44(3), 269-276. doi:10.1093/pcp/pcg048
- Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M., & Szczepaniak, A. (2014). Rubisco Accumulation Factor 1 from Thermosynechococcus elongatus participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in Escherichia coli cells and in vitro. *FEBS J*, 281(17), 3920-3932. doi:10.1111/febs.12928

- Kreel, N. E., & Tabita, F. R. (2007). Substitutions at methionine 295 of *Archaeoglobus fulgidus* ribulose-1,5-bisphosphate carboxylase/oxygenase affect oxygen binding and CO₂/O₂ specificity. *J Biol Chem*, 282(2), 1341-1351. doi:10.1074/jbc.M609399200
- Kreel, N. E., & Tabita, F. R. (2015). Serine 363 of a Hydrophobic Region of Archaeal Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase from *Archaeoglobus fulgidus* and *Thermococcus kodakaraensis* Affects CO₂/O₂ Substrate Specificity and Oxygen Sensitivity. *PLoS One*, 10(9), e0138351. doi:10.1371/journal.pone.0138351
- Kromdijk, J., Ubierna, N., Cousins, A. B., & Griffiths, H. (2014). Bundle-sheath leakiness in C₄ photosynthesis: a careful balancing act between CO₂ concentration and assimilation. *J Exp Bot*, 65(13), 3443-3457. doi:10.1093/jxb/eru157
- Kumar, A., Li, C., & Portis, A. R., Jr. (2009). Arabidopsis thaliana expressing a thermostable chimeric Rubisco activase exhibits enhanced growth and higher rates of photosynthesis at moderately high temperatures. *Photosynth Res*, 100(3), 143-153. doi:10.1007/s11120-009-9438-y
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*, 33(7), 1870-1874. doi:10.1093/molbev/msw054
- Kurek, I., Chang, T. K., Bertain, S. M., Madrigal, A., Liu, L., Lassner, M. W., & Zhu, G. (2007). Enhanced Thermostability of Arabidopsis Rubisco activase improves photosynthesis and growth rates under moderate heat stress. *Plant Cell*, 19(10), 3230-3241. doi:10.1105/tpc.107.054171
- Lee, B., Read, B. A., & Tabita, F. R. (1991). Catalytic properties of recombinant octameric, hexadecameric, and heterologous cyanobacterial/ bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys*, 291(2), 263-269. doi:10.1016/0003-9861(91)90133-4
- Lee, G. J., McDonald, K. A., & McFadden, B. A. (1993). Leucine 332 influences the CO₂/O₂ specificity factor of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Anacystis nidulans*. *Protein Sci*, 2(7), 1147-1154. doi:10.1002/pro.5560020709
- Leegood, R. C. (2002). C₄ photosynthesis: principles of CO₂ concentration and prospects for its introduction into C₃ plants. *J Exp Bot*, 53(369), 581-590. doi:10.1093/jexbot/53.369.581
- Li, C., Wang, D., & Portis, A. R., Jr. (2006). Identification of critical arginine residues in the functioning of Rubisco activase. *Arch Biochem Biophys*, 450(2), 176-182. doi:10.1016/j.abb.2006.04.002

- Li, L. A., & Tabita, F. R. (1997). Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcX* gene. *J Bacteriol*, 179(11), 3793-3796.
- Liu, C., Young, A. L., Starling-Windhof, A., Bracher, A., Saschenbrecker, S., Rao, B. V., . . . Hayer-Hartl, M. (2010). Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature*, 463(7278), 197-202. doi:10.1038/nature08651
- Liu, D., Ramya, R. C. S., & Mueller-Cajar, O. (2017). Surveying the expanding prokaryotic Rubisco multiverse. *FEMS Microbiol Lett*, 364(16). doi:10.1093/femsle/fnx156
- Loganathan, N., Tsai, Y. C., & Mueller-Cajar, O. (2016). Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proc Natl Acad Sci U S A*, 113(49), 14019-14024. doi:10.1073/pnas.1610758113
- Lorimer, G. H., Badger, M. R., & Andrews, T. J. (1976). The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry*, 15(3), 529-536. doi:10.1021/bi00648a012
- Lorimer, G. H., Chen, Y. R., & Hartman, F. C. (1993). A role for the epsilon-amino group of lysine-334 of ribulose-1,5-bisphosphate carboxylase in the addition of carbon dioxide to the 2,3-enediol(ate) of ribulose 1,5-bisphosphonate. *Biochemistry*, 32(35), 9018-9024. doi:10.1021/bi00086a006
- Marin-Navarro, J., & Moreno, J. (2006). Cysteines 449 and 459 modulate the reduction-oxidation conformational changes of ribulose 1.5-bisphosphate carboxylase/oxygenase and the translocation of the enzyme to membranes during stress. *Plant Cell Environ*, 29(5), 898-908.
- Maynard Smith, J. (1970). Natural Selection and the Concept of a Protein Space. *Nature*, 225(5232), 563-564. doi:10.1038/225563a0
- McCurry, S., Pierce, J., Tolbert, N., & Orme-Johnson, W. (1981). On the mechanism of effector-mediated activation of ribulose bisphosphate carboxylase/oxygenase. *Journal of Biological Chemistry*, 256(13), 6623-6628.
- Meyer, M., & Griffiths, H. (2013). Origins and diversity of eukaryotic CO₂-concentrating mechanisms: lessons for the future. *J Exp Bot*, 64(3), 769-786. doi:10.1093/jxb/ers390
- Moore, B. d., & Seemann, J. R. (1992). Metabolism of 2'-Carboxyarabinitol in Leaves. *Plant Physiol*, 99(4), 1551-1555. doi:10.1104/pp.99.4.1551
- Morell, M. K., Paul, K., Kane, H. J., & Andrews, T. J. (1992). Rubisco: Maladapted or Misunderstood. *Australian Journal of Botany*, 40(5), 431. doi:10.1071/bt9920431

- Morell, M. K., Paul, K., O'Shea, N. J., Kane, H. J., & Andrews, T. J. (1994). Mutations of an active site threonyl residue promote beta elimination and other side reactions of the enediol intermediate of the ribulosebiphosphate carboxylase reaction. *Journal of Biological Chemistry*, 269(11), 8091-8098.
- Morrison, J. F. (1982). The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. *Trends in Biochemical Sciences*, 7(3), 102-105. doi:10.1016/0968-0004(82)90157-8
- Mueller-Cajar, O. (2006). *Directed evolution of Rubisco in Escherichia coli* (PHD Doctoral dissertation), Australian National University.
- Mueller-Cajar, O. (2017). The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol Biosci*, 4, 31. doi:10.3389/fmolb.2017.00031
- Mueller-Cajar, O., Morell, M., & Whitney, S. M. (2007). Directed evolution of rubisco in Escherichia coli reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochemistry*, 46(49), 14067-14074. doi:10.1021/bi700820a
- Mueller-Cajar, O., Stotz, M., & Bracher, A. (2014). Maintaining photosynthetic CO₂ fixation via protein remodelling: the Rubisco activases. *Photosynth Res*, 119(1-2), 191-201. doi:10.1007/s11120-013-9819-0
- Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature*, 479(7372), 194-199. doi:10.1038/nature10568
- Mueller-Cajar, O., & Whitney, S. M. (2008a). Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynth Res*, 98(1-3), 667-675. doi:10.1007/s11120-008-9324-z
- Mueller-Cajar, O., & Whitney, S. M. (2008b). Evolving improved Synechococcus Rubisco functional expression in Escherichia coli. *Biochem J*, 414(2), 205-214. doi:10.1042/BJ20080668
- Newman, J., & Gutteridge, S. (1993). The X-ray structure of Synechococcus ribulose-biphosphate carboxylase/oxygenase-activated quaternary complex at 2.2-A resolution. *Journal of Biological Chemistry*, 268(34), 25876-25886.
- Nishitani, Y., Yoshida, S., Fujihashi, M., Kitagawa, K., Doi, T., Atomi, H., . . . Miki, K. (2010). Structure-based catalytic optimization of a type III Rubisco from a hyperthermophile. *J Biol Chem*, 285(50), 39339-39347. doi:10.1074/jbc.M110.147587
- Ogura, T., & Wilkinson, A. J. (2001). AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells*, 6(7), 575-597.

- Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., . . . Long, S. P. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the national academy of sciences*, 112(28), 8529-8536.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*, 4(11), 2411-2423. doi:10.1002/pro.5560041120
- Paoli, G. C., & Tabita, F. R. (1998). Aerobic chemolithoautotrophic growth and RubisCO function in *Rhodobacter capsulatus* and a spontaneous gain of function mutant of *Rhodobacter sphaeroides*. *Arch Microbiol*, 170(1), 8-17.
- Parikh, M. R., Greene, D. N., Woods, K. K., & Matsumura, I. (2006). Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E.coli*. *Protein Eng Des Sel*, 19(3), 113-119. doi:10.1093/protein/gzj010
- Parry, M. A., Andralojc, P. J., Mitchell, R. A., Madgwick, P. J., & Keys, A. J. (2003). Manipulation of Rubisco: the amount, activity, function and regulation. *J Exp Bot*, 54(386), 1321-1333.
- Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., & Whitney, S. M. (2013). Rubisco activity and regulation as targets for crop improvement. *J Exp Bot*, 64(3), 717-730. doi:10.1093/jxb/ers336
- Parry, M. A., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E., & Andralojc, P. J. (2008). Rubisco regulation: a role for inhibitors. *J Exp Bot*, 59(7), 1569-1580. doi:10.1093/jxb/ern084
- Parry, M. A., Madgwick, P., Parmar, S., Cornelius, M. J., & Keys, A. J. (1992). Mutations in loop six of the large subunit of ribulose-1,5-bisphosphate carboxylase affect substrate specificity. *Planta*, 187(1), 109-112. doi:10.1007/BF00201631
- Paz-Yepes, J., Brahamsha, B., & Palenik, B. (2013). Role of a microcin-C-like biosynthetic gene cluster in allelopathic interactions in marine *Synechococcus*. *Proc Natl Acad Sci U S A*, 110(29), 12030-12035. doi:10.1073/pnas.1306260110
- Pearce, F. G. (2006). Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem J*, 399(3), 525-534. doi:10.1042/BJ20060430
- Pearce, F. G., & Andrews, T. J. (2003). The relationship between side reactions and slow inhibition of ribulose-bisphosphate carboxylase revealed by a loop 6 mutant of the tobacco enzyme. *J Biol Chem*, 278(35), 32526-32536. doi:10.1074/jbc.M305493200

- Pierce, J., Lorimer, G. H., & Reddy, G. S. (1986). Kinetic mechanism of ribulosebisphosphate carboxylase: evidence for an ordered, sequential reaction. *Biochemistry*, 25(7), 1636-1644. doi:10.1021/bi00355a029
- Pierce, J., Tolbert, N. E., & Barker, R. (1980). Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogs. *Biochemistry*, 19(5), 934-942. doi:10.1021/bi00546a018
- Portis, A. R., Jr., Li, C., Wang, D., & Salvucci, M. E. (2008). Regulation of Rubisco activase and its interaction with Rubisco. *J Exp Bot*, 59(7), 1597-1604. doi:10.1093/jxb/erm240
- Rae, B. D., Long, B. M., Badger, M. R., & Price, G. D. (2013). Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO₂ fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev*, 77(3), 357-379. doi:10.1128/MMBR.00061-12
- Raines, C. A. (2003). The Calvin cycle revisited. *Photosynth Res*, 75(1), 1-10. doi:10.1023/A:1022421515027
- Read, B. A., & Tabita, F. R. (1994). High substrate specificity factor ribulose bisphosphate carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant cyanobacterial RubiSCO containing "algal" residue modifications. *Arch Biochem Biophys*, 312(1), 210-218. doi:10.1006/abbi.1994.1301
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, 42(Web Server issue), W320-324. doi:10.1093/nar/gku316
- Robinson, S. P., & Portis, A. R., Jr. (1989). Adenosine triphosphate hydrolysis by purified rubisco activase. *Arch Biochem Biophys*, 268(1), 93-99.
- Salvucci, M. E. (1992). Subunit interactions of Rubisco activase: polyethylene glycol promotes self-association, stimulates ATPase and activation activities, and enhances interactions with Rubisco. *Arch Biochem Biophys*, 298(2), 688-696.
- Salvucci, M. E., & Crafts-Brandner, S. J. (2004). Inhibition of photosynthesis by heat stress: the activation state of Rubisco as a limiting factor in photosynthesis. *Physiol Plant*, 120(2), 179-186. doi:10.1111/j.0031-9317.2004.0173.x
- Salvucci, M. E., Portis, A. R., & Ogren, W. L. (1985). A soluble chloroplast protein catalyzes ribulosebisphosphate carboxylase/oxygenase activation in vivo. *Photosynth Res*, 7(2), 193-201. doi:10.1007/bf00037012
- Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor Laboratory Press.

- Saschenbrecker, S., Bracher, A., Rao, K. V., Rao, B. V., Hartl, F. U., & Hayer-Hartl, M. (2007). Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell*, 129(6), 1189-1200. doi:10.1016/j.cell.2007.04.025
- Satagopan, S., Chan, S., Perry, L. J., & Tabita, F. R. (2014). Structure-function studies with the unique hexameric form II ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from *Rhodospseudomonas palustris*. *J Biol Chem*, 289(31), 21433-21450. doi:10.1074/jbc.M114.578625
- Satagopan, S., Scott, S. S., Smith, T. G., & Tabita, F. R. (2009). A Rubisco mutant that confers growth under a normally "inhibitory" oxygen concentration. *Biochemistry*, 48(38), 9076-9083. doi:10.1021/bi9006385
- Satagopan, S., & Spreitzer, R. J. (2004). Substitutions at the Asp-473 latch residue of *Chlamydomonas* ribulosebisphosphate carboxylase/oxygenase cause decreases in carboxylation efficiency and CO(2)/O(2) specificity. *J Biol Chem*, 279(14), 14240-14244. doi:10.1074/jbc.M313215200
- Satagopan, S., & Tabita, F. R. (2016). RubisCO selection using the vigorously aerobic and metabolically versatile bacterium *Ralstonia eutropha*. *FEBS J*, 283(15), 2869-2880. doi:10.1111/febs.13774
- Savir, Y., Noor, E., Milo, R., & Tlusty, T. (2010). Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc Natl Acad Sci U S A*, 107(8), 3475-3480. doi:10.1073/pnas.0911663107
- Scafaro, A. P., Galle, A., Van Rie, J., Carmo-Silva, E., Salvucci, M. E., & Atwell, B. J. (2016). Heat tolerance in a wild *Oryza* species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. *New Phytol*, 211(3), 899-911. doi:10.1111/nph.13963
- Schloss, J. V. (1988). Comparative affinities of the epimeric reaction-intermediate analogs 2- and 4-carboxy-D-arabinitol 1,5-bisphosphate for spinach ribulose 1,5-bisphosphate carboxylase. *Journal of Biological Chemistry*, 263(9), 4145-4150.
- Schneider, G., Lindqvist, Y., & Branden, C. I. (1992). RUBISCO: structure and mechanism. *Annu Rev Biophys Biomol Struct*, 21, 119-143. doi:10.1146/annurev.bb.21.060192.001003
- 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. *J Mol Biol*, 211(4), 989-1008. doi:10.1016/0022-2836(90)90088-4
- Schreuder, H. A., Knight, S., Curmi, P. M., Andersson, I., Cascio, D., Branden, C. I., & Eisenberg, D. (1993). Formation of the active site of ribulose-1,5-bisphosphate carboxylase/oxygenase by a disorder-order transition from the unactivated to the

- activated form. *Proceedings of the national academy of sciences*, 90(21), 9968-9972. doi:10.1073/pnas.90.21.9968
- Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., & Erb, T. J. (2016). A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*, 354(6314), 900-904. doi:10.1126/science.aah5237
- Sharwood, R. E., von Caemmerer, S., Maliga, P., & Whitney, S. M. (2008). The catalytic properties of hybrid Rubisco comprising tobacco small and sunflower large subunits mirror the kinetically equivalent source Rubiscos and can support tobacco growth. *Plant Physiol*, 146(1), 83-96. doi:10.1104/pp.107.109058
- Shih, P. M., Occhialini, A., Cameron, J. C., Andralojc, P. J., Parry, M. A., & Kerfeld, C. A. (2016). Biochemical characterization of predicted Precambrian RuBisCO. *Nat Commun*, 7, 10382. doi:10.1038/ncomms10382
- Shivhare, D., & Mueller-Cajar, O. (2017). In Vitro Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop. *Plant Physiol*, 174(3), 1505-1516. doi:10.1104/pp.17.00554
- Smith, S. A., & Tabita, F. R. (2003). Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Mol Biol*, 331(3), 557-569. doi:10.1016/S0022-2836(03)00786-1
- Smith, S. A., & Tabita, F. R. (2004). Glycine 176 affects catalytic properties and stability of the *Synechococcus* sp. strain PCC6301 ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Biol Chem*, 279(24), 25632-25637. doi:10.1074/jbc.M401360200
- South, P. F., Cavanagh, A. P., Liu, H. W., & Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*, 363(6422), eaat9077. doi:10.1126/science.aat9077
- Spreitzer, R. J. (1993). Genetic dissection of Rubisco structure and function. *Annu Rev Plant Biol*, 44(1), 411-434.
- Spreitzer, R. J. (2003). Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys*, 414(2), 141-149. doi:10.1016/s0003-9861(03)00171-1
- Spreitzer, R. J., Esquivel, M. G., Du, Y.-C., & McLaughlin, P. D. (2001). Alanine-Scanning Mutagenesis of the Small-Subunit β A- β B Loop of Chloroplast Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase: Substitution at Arg-71 Affects Thermal Stability and CO₂/O₂ Specificity†. *Biochemistry*, 40(19), 5615-5621. doi:10.1021/bi002943e

- Stitt, M., Lunn, J., & Usadel, B. (2010). Arabidopsis and primary photosynthetic metabolism - more than the icing on the cake. *Plant J*, 61(6), 1067-1091. doi:10.1111/j.1365-313X.2010.04142.x
- Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Structure of green-type Rubisco activase from tobacco. *Nat Struct Mol Biol*, 18(12), 1366-1370. doi:10.1038/nsmb.2171
- Sugawara, H., Yamamoto, H., Shibata, N., Inoue, T., Okada, S., Miyake, C., . . . Kai, Y. (1999). Crystal Structure of Carboxylase Reaction-oriented Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase from a Thermophilic Red Alga, *Galdieria partita*. *Journal of Biological Chemistry*, 274(22), 15655-15661. doi:10.1074/jbc.274.22.15655
- Sutter, M., Roberts, E. W., Gonzalez, R. C., Bates, C., Dawoud, S., Landry, K., . . . Kerfeld, C. A. (2015). Structural Characterization of a Newly Identified Component of alpha-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Sci Rep*, 5, 16243. doi:10.1038/srep16243
- Tabita, F. R. (1988). Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. *Microbiol Rev*, 52(2), 155-189.
- Tabita, F. R. (1995). The Biochemistry and Metabolic Regulation of Carbon Metabolism and CO₂ Fixation in Purple Bacteria. In R. E. Blankenship, M. T. Madigan, & C. E. Bauer (Eds.), *Anoxygenic Photosynthetic Bacteria* (pp. 885-914). Dordrecht: Springer Netherlands.
- Tabita, F. R. (1999). Microbial ribulose 1, 5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth Res*, 60(1), 1-28.
- Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H., & Kreel, N. E. (2008a). Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philos Trans R Soc Lond B Biol Sci*, 363(1504), 2629-2640. doi:10.1098/rstb.2008.0023
- Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., & Scott, S. S. (2008b). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot*, 59(7), 1515-1524. doi:10.1093/jxb/erm361
- Taylor, T. C., & Andersson, I. (1996a). Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the activated unliganded spinach enzyme. *Nat Struct Biol*, 3(1), 95-101. doi:doi:10.1038/nsb0196-95
- Taylor, T. C., & Andersson, I. (1996b). Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the

- activated unliganded spinach enzyme. *Nat Struct Biol*, 3(1), 95-101.
doi:10.1038/nsb0196-95
- Taylor, T. C., & Andersson, I. (1997). The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate. *J Mol Biol*, 265(4), 432-444.
doi:10.1006/jmbi.1996.0738
- Tcherkez, G. G., Farquhar, G. D., & Andrews, T. J. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U S A*, 103(19), 7246-7251.
doi:10.1073/pnas.0600605103
- Tokuriki, N., & Tawfik, D. S. (2009a). Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459(7247), 668-673.
doi:10.1038/nature08009
- Tokuriki, N., & Tawfik, D. S. (2009b). Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol*, 19(5), 596-604. doi:10.1016/j.sbi.2009.08.003
- Tsai, Y. C., Lapina, M. C., Bhushan, S., & Mueller-Cajar, O. (2015). Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat Commun*, 6, 8883. doi:10.1038/ncomms9883
- Valegård, K., Andralojc, P. J., Haslam, R. P., Pearce, F. G., Eriksen, G. K., Madgwick, P. J., . . . Andersson, I. (2018). Structural and functional analyses of Rubisco from arctic diatom species reveal unusual posttranslational modifications. *Journal of Biological Chemistry*. doi:10.1074/jbc.RA118.003518
- Varaljay, V. A., Satagopan, S., North, J. A., Witte, B., Dourado, M. N., Anantharaman, K., . . . Tabita, F. R. (2016). Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environ Microbiol*, 18(4), 1187-1199. doi:10.1111/1462-2920.13138
- Vitlin Gruber, A., Nisemlat, S., Azem, A., & Weiss, C. (2013). The complexity of chloroplast chaperonins. *Trends Plant Sci*, 18(12), 688-694.
doi:10.1016/j.tplants.2013.08.001
- Wachter, R. M., Salvucci, M. E., Carmo-Silva, A. E., Barta, C., Genkov, T., & Spreitzer, R. J. (2013). Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco-Rubisco activase interaction. *Photosynth Res*, 117(1-3), 557-566. doi:10.1007/s11120-013-9827-0
- Walker, B. J., VanLoocke, A., Bernacchi, C. J., & Ort, D. R. (2016). The Costs of Photorespiration to Food Production Now and in the Future. *Annu Rev Plant Biol*, 67(1), 107-129. doi:10.1146/annurev-arplant-043015-111709
- Wang, Y. L., Zhou, J. H., Wang, Y. F., Bao, J. S., & Chen, H. B. (2001). Properties of hybrid enzymes between *Synechococcus* large subunits and higher plant small

- subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Escherichia coli*. *Arch Biochem Biophys*, 396(1), 35-42. doi:10.1006/abbi.2001.2555
- Weber, A. P. M., & Bar-Even, A. (2019). Improving the efficiency of photosynthetic carbon reactions. *Plant Physiol*, pp.01521.02018. doi:10.1104/pp.18.01521
- Whitman, W. B., Martin, M. N., & Tabita, F. R. (1979). Activation and regulation of ribulose biphosphate carboxylase-oxygenase in the absence of small subunits. *J Biol Chem*, 254(20), 10184-10189.
- Whitney, S. M., Houtz, R. L., & Alonso, H. (2011). Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiol*, 155(1), 27-35. doi:10.1104/pp.110.164814
- Whitney, S. M., & Sharwood, R. E. (2008). Construction of a tobacco master line to improve Rubisco engineering in chloroplasts. *J Exp Bot*, 59(7), 1909-1921. doi:10.1093/jxb/erm311
- Whitney, S. M., von Caemmerer, S., Hudson, G. S., & Andrews, T. J. (1999). Directed mutation of the Rubisco large subunit of tobacco influences photorespiration and growth. *Plant Physiol*, 121(2), 579-588.
- Whittaker, C. A., & Hynes, R. O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell*, 13(10), 3369-3387. doi:10.1091/mbc.E02-05-0259
- Wilson, R. H., Alonso, H., & Whitney, S. M. (2016). Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci Rep*, 6, 22284. doi:10.1038/srep22284
- Wilson, R. H., Martin-Avila, E., Conlan, C., & Whitney, S. M. (2017a). An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO₂-fixation kinetics. *J Biol Chem*. doi:10.1074/jbc.M117.810861
- Wilson, R. H., & Whitney, S. M. (2017b). Improving CO₂ Fixation by Enhancing Rubisco Performance. *Directed Enzyme Evolution: Advances and Applications*, 101-126. doi:10.1007/978-3-319-50413-1_4
- Witte, B., John, D., Wawrik, B., Paul, J. H., Dayan, D., & Tabita, F. R. (2010). Functional prokaryotic RubisCO from an oceanic metagenomic library. *Appl Environ Microbiol*, 76(9), 2997-3003. doi:10.1128/AEM.02661-09
- Wong, K. S., & Houry, W. A. (2012). Novel structural and functional insights into the MoxR family of AAA+ ATPases. *J Struct Biol*, 179(2), 211-221. doi:10.1016/j.jsb.2012.03.010

- Wrighton, K. C., Castelle, C. J., Varaljay, V. A., Satagopan, S., Brown, C. T., Wilkins, M. J., . . . Banfield, J. F. (2016). RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J*, 10(11), 2702-2714. doi:10.1038/ismej.2016.53
- Wyganowski, K. T., Kaltenbach, M., & Tokuriki, N. (2013). GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J Mol Biol*, 425(18), 3403-3414. doi:10.1016/j.jmb.2013.06.028
- Young, J. N., Heures, A. M., Sharwood, R. E., Rickaby, R. E., Morel, F. M., & Whitney, S. M. (2016). Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *J Exp Bot*, 67(11), 3445-3456. doi:10.1093/jxb/erw163
- Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., . . . Zhang, X. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 296(5565), 79-92.
- Zhang, X., Carter, M. S., Vetting, M. W., San Francisco, B., Zhao, S., Al-Obaidi, N. F., . . . Gerlt, J. A. (2016). Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars. *Proc Natl Acad Sci U S A*, 113(29), E4161-4169. doi:10.1073/pnas.1605546113
- Zhu, X. G., Long, S. P., & Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol*, 61(1), 235-261. doi:10.1146/annurev-arplant-042809-112206
- Ziska, L. H., Bunce, J. A., Shimono, H., Gealy, D. R., Baker, J. T., Newton, P. C., . . . Wilson, L. T. (2012). Food security and climate change: on the potential to adapt global crop production by active selection to rising atmospheric carbon dioxide. *Proc Biol Sci*, 279(1745), 4097-4105. doi:10.1098/rspb.2012.1005