Structural and functional studies of 1) Actin interacting protein 5, a novel actin assembly regulator in *Saccharomyces cerevisiae* 2) CbbX, a red-type Rubisco activase in *Cyanidioschyzon merolae*

Sun, Jialin

2018

Sun, J. (2018). Structural and functional studies of 1) Actin interacting protein 5, a novel actin assembly regulator in *Saccharomyces cerevisiae* 2) CbbX, a red-type Rubisco activase in *Cyanidioschyzon merolae*. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/86103>

<https://doi.org/10.32657/10220/46705>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

Institute of
Molecular and
Cell Biology

Structural and functional studies of

- 1) Actin interacting protein 5, a novel actin assembly regulator in *Saccharomyces cerevisiae***
- 2) CbbX, a red-type Rubisco activase in *Cyanidioschyzon merolae***

Sun Jialin

**School of Biological Sciences, Nanyang Technological
University, Singapore
Institute of Cell and Molecular Biology (IMCB), A*STAR,
Singapore**

2018

Structural and functional studies of

- 1) Actin interacting protein 5, a novel actin assembly regulator in *Saccharomyces cerevisiae***
- 2) CbbX, a red-type Rubisco activase in *Cyanidioschyzon merolae***

Sun Jialin

**School of Biological Sciences, Nanyang Technological
University, Singapore
Institute of Cell and Molecular Biology (IMCB), A*STAR,
Singapore**

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

2018

ACKNOWLEDGEMENT

Four years of PhD study coming to an end, it has been a fruitful journey with many valuable lessons learnt. This journey would not be possible without the guidance and care from my supervisors, colleagues, friends, and family.

First and foremost, I would like to express my deepest heartfelt appreciation to my supervisor, Associate Professor Gao Yonggui, who has given me immense amount of guidance and advice during the past four years. I first met Prof. Gao during a lecture about ribosomes when I was just in my first year of undergraduate study. I was fascinated by the structure of ribosomes and this got me interested in his research. I had minimal wet lab experience prior to joining the lab. However, that did not prevent Prof. Gao from taking me in as his PhD student. Not only that he has taught me research skills but also given me life lessons and constant encouragement throughout this journey. It is his enthusiasm in science that has motivated me to go further. His willingness to explore problems together with me whenever I meet bottlenecks during my research has greatly encouraged me and has led me to where I am today.

Secondly I would like to express my sincere gratitude to my co-supervisor, Professor Hong Wanjin, for his warm support and guidance since the start of my PhD journey. Being the Executive Director of IMCB with a packed schedule, he has always made sure to take time to meet me for discussions on my research progress, to attend each of my TAC meetings and Qualifying Exam. Each time we met, he has always been encouraging, and giving me valuable advice on my research topics. His warm personality and constant support are the major factors that made my PhD journey a much smoother one.

Thirdly, I would like to thank my collaborator Assistant Professor Miao Yansong and Ms Xie Ying. Prof. Miao offered me the opportunity to take up a very interesting and challenging project. He had given me much valuable advice and inspired me to think more dynamically and to think deeper when I am met with a research problem. Ms. Xie Ying rendered me significant amount of help during this project for the past two years. I truly enjoyed the times that we did experiment together.

I would also like to thank Dr. Ye Fuzhou and Dr. Chen Yun for their kind guidance. Dr. Ye was my very first mentor when I first joined Prof. Gao's lab as an attachment student. He taught me the basics of research techniques and cultivated many good habits of wet lab skills in me, which laid a strong foundation in my subsequent research during PhD. Dr. Chen Yun instilled the thought of flexibility in my approach to a research problem, often honing my problem-solving skills whenever I encounter difficulties.

Next I would like to extend my sincere thanks to Dr. Joel Toh. I am greatly appreciative that Prof Gao assigned Dr. Toh to be my mentor at the start of my PhD. Dr. Toh provided me with huge amount of guidance throughout my PhD years. Often I felt frustrated with my experimental results, and he was always the one who patiently goes through the problem with me, aiding my mental processes. His valuable advice helped me examine a problem in different perspectives and broaden my way of thinking. I would not have made it so far without his help in many aspects of my research.

I would like to thank Dr. Veerendra Kumar for teaching and sharing his expertise in X-ray crystallography. I have benefitted substantially from his tutoring. I would also like to thank my lovely lab members who have provided me with a great amount of assistance in times of need. It is indeed a pleasure to work with such wonderful colleagues.

I would like to give my earnest gratitude to my close friends. They are a great source of happiness in my life. I am grateful that they have always been with me through the ups and downs during the past years, and have always had faith in me.

In addition, I would like to thank A*STAR Graduate Academy for providing me the opportunity to partake this PhD and to fund my AGS scholarship. It has always been a dream for me to join the A*STAR family since I was young. I am immensely grateful that I have joined the family. I would like to thank IMCB, NISB and SBS from NTU for allowing me to use their state-of-the-art facility for my research work. I would like to thank Australian Synchrotron and National Synchrotron Radiation Research Center at Taiwan, and Swiss Light Source in Switzerland for the synchrotron facility for X-ray data collection.

Last but not least, I am truly thankful to my beloved family members, who have been taking care of my emotional needs and giving me enormous amount of encouragement throughout this journey. They are my greatest mental support and my strongest pillars in life.

ACKNOWLEDGEMENT-----	3
TABLE OF CONTENTS-----	6
ABSTRACT-----	9
LIST OF FIGURES-----	10
LIST OF TABLES-----	12
LIST OF ABBREVIATIONS-----	13

Chapter 1: Structural and functional studies Actin Interacting Protein 5, a novel actin assembly regulator in *Saccharomyces cerevisiae*

1. INTRODUCTION-----	16
1.1 Polarized cell growth in budding yeast-----	16
1.2 Polarization establishment requires a complex of polarity proteins -----	17
1.2.1 Actin and actin structure-----	18
1.2.2 Dynamics of actin filaments-----	21
1.2.3 Actin and polarisome complex-----	24
1.2.3.1 Nucleation factors-----	25
1.2.3.2 Nucleation promoting factors-----	29
1.2.3.3 Elongation factors-----	30
1.3 Aip5 as a novel class of actin NF and an NPF of Bni1-----	31
1.3.1 Aip5, a previously unknown polarisome protein in budding yeast-----	31
1.3.2 Aip5 localizes to the bud tip in presence of Spa2-----	32
1.3.3 Localization of Aip5 partially depends on NPF Bud6 and NF Bni1-----	33
1.3.4 Aip5 as a potent NF and an NPF of Bni1-----	34
1.3.5 Aip5 functions as an NF and NPF of Bni1 through C-terminal domain-----	36
1.4 Aims and objectives-----	37
2. MATERIALS AND METHODS-----	38
2.1 Molecular cloning-----	38
2.2 Aip5-C mutagenesis-----	38
2.3 Protein expression and purification-----	40
2.4 Crystallization-----	42
2.5 Data collection and structure determination-----	43
2.6 Oligomeric state determination by analytical size-exclusion chromatography-----	43
2.7 Pyrene fluorescence assay of actin assembly-----	44
2.8 Microscale Thermophoresis (MST) -----	45

2.9 Protein sequence prediction and evolutionary analysis-----	46
2.10 Deposited data-----	46
3. RESULTS-----	47
3.1 Small scale expression of Aip5-C-----	47
3.2 Purification of Aip5-C in large scale-----	48
3.3 Crystallization of Aip5-C-----	50
3.4 Data collection and refinement of Aip5-C-----	53
3.5 Crystal structure of Aip5-C-----	54
3.5.1 Presence of a thioredoxin fold structure in Aip5-C-----	55
3.5.2 Strong polar network within the dimer interface of Aip5-C-----	56
3.6 Sequence alignment and structural comparison of Aip5-C with cGrx2-----	57
3.7 Aip5 nucleation activity depends on a loop region that binds to G-actin-----	59
3.7.1 Binding affinity between G-actin and Aip5-C loop via MST assay-----	63
3.8 Structure of the mutant Aip5-C-3m (L1150A, N1162A, S1165C)-----	64
3.8.1 Structure-function relation through comparison of Aip5-C and Aip5-C-3m-----	67
4. DISCUSSION-----	70
4.1 Effect of protein fusion tag on crystallization-----	70
4.2 Aip5 has a thioredoxin fold with a different function-----	72
4.3 Structural conformation and loop orientation in Aip5-C are crucial for its activity-----	73
5. FUTURE WORK-----	74
REFERENCES-----	75

Chapter 2: Structural and functional studies of CbbX, a red-type Rubisco activase in *Cyanidioschyzon merolae*

1. INTRODUCTION-----	81
1.1 Ever-increasing demand for sustainable resources-----	81
1.2 Photosynthesis-----	81
1.2.1 Light dependent reactions-----	82
1.2.2 Light independent reactions-----	84
1.3 Rubisco-----	86
1.3.1 Structure of rubisco-----	87
1.4 Rubisco activase-----	90
1.4.1 CbbX-----	91
1.4.2 CbbX in <i>Cyanidioschyzon merolae</i> -----	92
1.5 Aims and objectives-----	95

2. MATERIALS AND METHODS-----	96
2.1 Molecular cloning-----	96
2.2 Mutagenesis on CbbX complex-----	96
2.3 Small scale protein expression and purification-----	98
2.4 Large scale protein expression and purification-----	100
2.5 Crystallization-----	101
2.6 Data collection-----	103
2.7 Protein sequence prediction and alignment-----	104
3. RESULTS-----	105
3.1 CbbX(pt) -----	105
3.1.1 Small scale expression of full length CbbX(pt) -----	105
3.1.2 Large scale purification of CbbX(pt) -----	106
3.1.3 Crystallization of full length CbbX(pt) -----	107
3.1.4 Data collection and refinement of CbbX(pt) -----	109
3.1.5 Overall crystal structure of CbbX(pt) from <i>C. merolae</i> -----	110
3.1.6 Structural comparison between CbbX(pt) and 3SYL-----	111
3.1.7 Structural comparison of CbbX(pt) with 3CF3-----	113
3.2 CbbX(nuc)-----	116
3.2.1 Large scale purification of full length CbbX(nuc) -----	116
3.2.2 Crystallization of full length CbbX(nuc) -----	118
3.3 CbbX complex-----	118
3.3.1 Large scale purification of CbbX complex-----	118
3.3.2 Crystallization of CbbX complex-----	121
3.3.3 Protein crystallization by mutagenesis of CbbX surface residues-----	122
3.3.4 Small scale expression of CbbX complex upon surface mutations-----	123
3.3.5 Large scale purification of Δ CbbX complex-----	124
3.3.6 Crystallization of Δ CbbX complex-----	126
4. DISCUSSION-----	127
4.1 Protein expression and purification-----	127
4.2 Crystallization of CbbX(pt), CbbX(nuc), and CbbX complex-----	129
4.3 Crystallographic data for CbbX(pt) from <i>C. merolae</i> -----	130
4.4 Structural analysis of CbbX(pt) from <i>C. merolae</i> -----	131
5. FUTURE WORK-----	133
REFERENCES-----	135

ABSTRACT

Polarisome, a fuzzy multiprotein complex, regulates polarized cell growth in both budding yeast and filamentous fungi through actin polymerization. In budding yeast, multiple polarisome proteins have been identified to interact with each other in nucleating actin filament through formin Bni1. However, molecular mechanism of dynamic regulation of the polarisome proteins in generating actin filaments largely remains unknown. In this study, we reported a previously uncharacterized gene YFR016C that encodes a novel fourth type of actin nucleation factor, named Aip5, which also functions as a potent nucleation-promoting factor of Bni1. Through X-ray crystallography, we have successfully solved the structure of Aip5 C-terminal domain (Aip5-C), the functional component for actin nucleating activity of Aip5. We have also solved the Aip5-C mutant structure, of which the activity was completely abolished. Structural analysis of both the wild type and mutant Aip5-C revealed that Aip5 C-terminal domain directly interacts with G-actin via a loop region for nucleating actin filaments, which was corroborated by various biochemical experiments. Future work involves co-crystallization of Aip5-C with actin or soaking of Aip5-C peptide with actin to further complement our study through structural means.

As the most abundant protein on earth, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is the key enzyme for carbon fixation during photosynthesis. Due to its sluggish enzymatic activity, Rubisco activase is required to speed up Rubisco's activity. Protein CbbX has been recently identified as the Rubisco activase in red algae. In the unicellular red algae *Cyanidioschyzon merolae*, there are two isoforms (plastid and nuclear) of the CbbX protein that bind to form a functional Rubisco activase – namely CbbX(pt) and CbbX(nuc). Using X-ray crystallography, the structure of CbbX(pt) was solved at 3.9 Å. Ribulose 1,5-bisphosphate (RuBP) and Adenosine triphosphate (ATP) binding sites were identified by superimposing with homologous structures, among which, both regions appeared to be conserved. CbbX(nuc) and CbbX complex crystals were successfully produced. However, more efforts are required to improve the present crystal diffraction, as well as to determine the complex structure to unveil the underlying catalytic and regulatory mechanism.

LIST OF FIGURES

Figure 1	Molecular Pathways for Cell Polarity Development in Budding Yeast Epithelial Cells.
Figure 2	Structure of uncomplexed actin in the ADP state.
Figure 3	Mechanisms that govern actin turnover.
Figure 4	Polarized cell growth requires the help from polarisome complex.
Figure 5	Factors involved in actin nucleation and polymerization.
Figure 6	Aip5 is a novel polarisome protein that nucleates actin polymerization-dependent and independent of Bni1.
Figure 7	Aip5 localization changes upon Latrunculin A treatment.
Figure 8	Aip5 functions as an NF and an NPF of Bni1.
Figure 9	Aip5 functions as an NF and NPF of Bni1 through C-terminal domain.
Figure 10	Soluble constructs of Aip5-C.
Figure 11	SDS-PAGE from PPP showing soluble expression for different constructs.
Figure 12	Large scale purification of Aip5-C.
Figure 13	Large scale purification of SeMet Aip5-C.
Figure 14	Crystals of Aip5-C.
Figure 15	Crystals of SeMet Aip5-C.
Figure 16	Determination of oligomeric state of Aip5-C in solution using superdex75 10/300 column.
Figure 17	Crystal structure of Aip5-C homodimer.
Figure 18	Sequence alignment of Aip5-C with structural homologue cGRx2.
Figure 19	Superposition of Aip5-C with homologue cGrx2.
Figure 20	Electron density observed for the loop region in Aip5-C and superposition of protomer A and B of Aip5-C.
Figure 21	Sequence alignment of Aip5-C among pathogenic filamentous fungi, illustrated by Jalview using the Clustal Omega method.
Figure 22	Actin nucleating activity of Aip5-C variants.
Figure 23	Hydrophobic interaction by Leu1150 from both protomers in Aip5-C.
Figure 24	Thermophoresis binding curves of 2 μ M G-actin titrated with peptides of loop and loop-L1150A, respectively.
Figure 25	Crystal structure of Aip5-C triple mutant.
Figure 26	Determination of oligomeric state of Aip5-C-3m in solution using superdex75 10/300 column.
Figure 27	Servers used to analyze total buried surface area and solvent accessible area in native and mutant Aip5-C.
Figure 28	Superposition of protomers in Aip5-C and Aip5-C-3m.
Figure 29	Superposition of Aip5-C dimer with Aip5-C-3m dimer.
Figure 30	Orientation of the loop sites in Aip5-C and Aip5-C-3m.
Figure 31	Crystals of Aip5-C with N-terminal His6 tag.

Figure 32	Structure of a chloroplast.
Figure 33	Photosynthesis starts with light dependent reactions.
Figure 34	Overview of Calvin-Benson cycle.
Figure 35	Structures of form I & form II Rubisco.
Figure 36	Structures of form III & form IV Rubisco.
Figure 37	Sequence alignment of CbbX(pt) and CbbX(nuc) with another red-type rubisco activase from <i>R. sphaeroides</i> .
Figure 38	Small-scale expression tests of CbbX(pt) using A) pH 8.0 buffer and B) pH 10.5 buffer.
Figure 39	Large scale purification of CbbX(pt).
Figure 40	Crystals of full length CbbX(pt).
Figure 41	Crystal structure of CbbX(pt) from <i>C. merolae</i> shown in ribbon representation.
Figure 42	Superimposition of crystal structures of CbbX(pt) with its homologue 3SYL shown in ribbon representation.
Figure 43	Conservation of the residues at RuBP binding site in crystal structures of both CbbX(pt) and 3SYL.
Figure 44	Sequence alignment of CbbX(pt) from <i>C. merolae</i> and its homolog 3SYL from <i>R. sphaeroides</i> using MAFFT multiple alignment and ESPrit programs.
Figure 45	Superimposition of crystal structures of CbbX(pt) with another homologue 3CF3 shown in ribbon representation.
Figure 46	Conservation of the residues at ATP binding site in crystal structures of both CbbX(pt) and 3CF3.
Figure 47	Sequence alignment of CbbX(pt) from <i>C. merolae</i> and homolog 3CF3 from <i>M. musculus</i> using MAFFT multiple alignment and ESPrit programs.
Figure 48	Large scale purification of CbbX(nuc).
Figure 49	Crystals of full length CbbX(pt).
Figure 50	Purification of CbbX complex.
Figure 51	Crystals of CbbX complex.
Figure 52	Small scale expression tests of CbbX complex upon surface mutations.
Figure 53	Purification of Δ CbbX complex with surface mutations.
Figure 54	Crystals of Δ CbbX complex.
Figure 55	Anion exchange purification of CbbX(pt) using 5 ml HiTrap Q HP column.
Figure 56	Various optimization results of CbbX complex.
Figure 57	Electron density for CbbX(pt) from <i>C. merolae</i> .

LIST OF TABLES

Table 1	Summary of crystallographic data and refinement statistics for Aip5-C, SeMet Aip5-C, and Aip5-C-3m.
Table 2	Aip5-C amino acid sequences and various mutation sites.
Table 3	Summary of crystallographic data and refinement statistics for CbbX(pt) from <i>C. merolae</i>
Table 4	Surface mutations of CbbX.

LIST OF ABBREVIATIONS

2PG	2-phospho-glycolate
3PGA	3-phospho-D-glycerate
ABPs	Actin-binding proteins
Ah	Actin-binding helices
Aip5	Actin-interacting protein 5
ANS	Actin-nucleating sequences
APC	Adenomatous polyposis coli
ARP	Actin-related protein
AS	Australian Synchrotron
ATP	Adenosine triphosphate
CO₂	Carbon dioxide
DAD	Diaphanous autoregulatory Domain
DNase I	Deoxyribonuclease I
Ena/VASP	Enabled/vasodilator-stimulated phosphoprotein
ETC	Electron transport chain
Fd	Ferredoxin
FH	Formin homology domain
FHA	Forkhead-associated domain
FNR	Ferredoxin NADP reductase
G3P	Glyceraldehyde 3-phosphate
GAPs	GTPase-activating proteins
GEFs	Guanylyl nucleotide exchange factors
IDRs	Intrinsically disordered regions
IPTG	Isopropyl-b-D-thiogalactopyranoside
LLR	Leucine-rich repeat
MBL	Monomer-binding linker
Mkk1	Mitogen-activated protein kinase kinase 1
Mpk1	Mitogen-activated protein kinase 1
MWCO	Molecular weight cut off
NF	Nucleation factor
NPF	Nucleation-promoting factor

NSRRC	National Synchrotron Radiation Research Center
O₂	Oxygen
OD	Optical density
PC	Plastocyanin
PDB	Protein Data Bank
PPP	Protein Production Platform
PQ	Plastoquinone
PSI	Photosystems I
PSII	Photosystems II
PTMs	Post-translational modifications
Rca	Rubisco activase
RMSD	Root-mean-square deviation
Rubisco	D-ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
SDS-PAGE	Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis
SH3	Src homology 3
SLS	Swiss Light Source
TMR	Tetramethylrhodamine-5-maleimide
WASP	Wiskott-Aldrich syndrome family protein
WH2	WASP-Homology 2, or Wiskott-Aldrich homology 2

Chapter 1: Structural and functional studies of actin interacting protein 5 (Aip5), a novel actin assembly regulator in *Saccharomyces cerevisiae*

1. INTRODUCTION

1.1 Polarized cell growth in budding yeast

Cell polarity refers to the intrinsic asymmetrical organization of different cellular components (Orlando and Guo 2009). Polarized cell growth and directional cell division are essential for eukaryotic development given their contribution to the morphology, structure and function of a cell. The precise establishment and maintenance of cell polarity are central to various functions in diverse cell types, for instance, nutrient transport, neuronal signaling, and motility (Chiou, Balasubramanian et al. 2017). These functions might vary greatly at the level of tissue or organ. However, polarized cell growth tends to follow common underlying principles at the cellular level.

Polarized cell growth occurs in budding yeast during different stages of its life cycle. It can be initiated when there are sufficient nutrients, or a lack of specific element such as nitrogen. Another circumstance where polarized cell growth can be stimulated is during the mating response (Madden and Snyder 1998). During the late G1 phase in a cell cycle, the cell becomes committed and no longer requires extracellular signals to proceed to the S phase. It is also during this time point that budding yeast selects its budding site in two discrete spatial patterns – axial and bipolar budding, in which MAT α and MAT α cells adopt axial pattern and create bud site adjacent to the preceding division site, while MAT α /MAT α cells adopt bipolar pattern and construct bud site either near the preceding division site or opposite of that (Chant and Pringle 1995).

1.2 Polarization establishment requires a complex of polarity proteins

Septins and actins are two important cytoskeleton proteins that act as spatial cues in generation of the bud sites (Ayscough, Stryker et al. 1997). In axial budding, assembled septins select an axial bud site by colocalizing two proteins Bud3 and Bud4. Additional protein Axl1 facilitates this selection process (Drubin and Nelson 1996). In bipolar budding, actin, instead of septin, is responsible for determining the budding pattern. Actin recruits Sla1, Sla2, and Rvs167 that trigger the subsequent assembly of polarized patch at the budding site (Drubin and Nelson 1996, Madden and Snyder 1998).

To initiate polarity establishment in budding yeast, Bud5, a guanylyl nucleotide exchange factor (GEF) activates Rsr1 GTPase (also known as Bud1), which in turn binds Cdc24, a GTPase-activating protein (GAP). Activated Cdc24 then recruits another GTPase Cdc42 to the bud site marked by septins or actins (**Figure 1**). The cascade of events leads to further recruitment of the actin cytoskeleton and the assembly of a huge protein complex associated to actin, which is known as the polarisome, leading to plasma membrane growth and cell wall assembly (Drubin and Nelson 1996).

Figure 1. Molecular Pathways for Cell Polarity Development in Budding Yeast Epithelial Cells. The actin and septin cytoskeleton proteins are shown in both green and blue to emphasize on their dual roles as spatial cues and as proteins required to propagate the cues. (Figure adapted and modified from Drubin and Nelson 1996)

1.2.1 Actin and actin structure

Being the key player in polarization establishment and the most abundant intracellular protein in eukaryotes, actin is a cytoskeletal component that is highly conserved across species. It helps to maintain cell morphology, regulate intracellular trafficking, and contribute to motility and contractility of a cell during cell division. Actin has two distinguishing features: 1) intrinsic ATPase activity, this means that it is capable of assembling and disassembling itself in the cell; 2) protein scaffold, able to bind to an array of actin associated proteins to propagate its function downstream (Ayscough, Stryker et al. 1997).

The first actin structure was solved in 1990 with G-actin bound to deoxyribonuclease I (DNase I) (Kabsch, Mannherz et al. 1990). Subsequently more actin structures have been reported by complexing with a variety of actin-binding proteins (ABPs). Due to

spontaneous polymerization of actin, actins have to be either modified chemically or mutated at specific sites to prevent self-polymerization (Dominguez and Holmes 2011), as polymerized actin poses great difficulty in crystallization.

Actin belongs to a structural superfamily that consists of sugar kinases, hexokinases, and stress-70 proteins, with two main domains classified into the larger or smaller domain (**Figure 2**) (Hurley 1996). The two main domains are separated by a nucleotide binding cleft located in the centre. Both domains have a signature topology that is the five β strands comprising a β -meander and a right handed β - α - β (Kabsch, Mannherz et al. 1990). Conventionally, actin was divided into four subdomain nomenclature (Dominguez and Holmes 2011). The smaller domain is made up of subdomain 1 and 2, while the larger one is made up of subdomain 3 and 4 (**Figure 2**). Subdomain 1 and 3 form the barbed end, while subdomain 2 and 4 form the pointed end. Besides its intrinsic nucleotide binding activity, actin is able to complex with many ABPs at the barbed end for a variety of cellular activities (Ayscough, Stryker et al. 1997).

Figure 2. Structure of uncomplexed actin in the ADP state. Four subdomains of actin are represented in different colors. The DNase I binding loop, which is folded as an α helix in this structure, is located in subdomain 2. Tetramethylrhodamine-5-maleimide (TMR) is covalently attached to Cys374, and binds in a hydrophobic pocket near the COOH-terminus in actin subdomain 1. (Figure adapted and modified from Otterbein, Graceffa et al. 2001)

Given its ability to polymerize under physiological condition, actin is not able to be crystallized on its own (Dominguez and Holmes 2011). Structures of actin in complex with some binding partners such as DNase I, profilin or gelsolin have been solved, however, these binding proteins prohibit the polymerization of actin (Porta and Borgstahl 2012). To elucidate the structure of uncomplexed actin, a fluorescent probe, tetramethylrhodamine-5-maleimide (TMR) was covalently bound to actin to block actin polymerization (Otterbein, Graceffa et al. 2001). This allowed actin to be crystallized without any protein partners in the ADP state.

Actin crystals were produced in solution soaked with ATP, in which ATP was hydrolyzed to ADP. In the structure, ADP was bound to the nucleotide binding cleft

in the center of all four subdomains (Otterbein, Graceffa et al. 2001). The overall structure of actin bound to ADP was similar to previously published ATP-bound state, with the exception that there was a slight orientation difference of about 10° and 5° rotation of subdomains 2 and 4 in the ADP-bound structure, respectively (Otterbein, Graceffa et al. 2001). TMR was covalently linked to Cys374 and bound to a small hydrophobic pocket in subdomain 1 that prevented actin polymerization (**Figure 2**). Majority of the ABPs interacts with actin around the cleft between subdomain 1 and 3.

1.2.2 Dynamics of actin filaments

Actin transits between two principal forms, globular, monomeric (G-actin), and filamentous polymeric (F-actin) forms (Visa and Percipalle 2010). G-actin is a 42 kDa actin monomer that polymerizes to be F-actin, also known as the microfilament, which provides structural support for the cell and involves various cellular processes. To initiate actin polymerization, it requires actin nucleation, which is the formation of a small actin aggregate comprising three actin monomers (Lodish, Berk et al. 2000). Actin monomer has an asymmetric structure that consists of close-fitting binding sites allowing another two actin monomers to bind to it in a head-to-tail fashion (Dominguez and Holmes 2011). Each actin monomer has a rotation of 166° along the axis, which contributes to the helical appearance of actin filaments. The distinctive polarity of actin filaments results in a barbed (+) end and a pointed (-) end (Cooper GM. 2000).

Transition from G-actin to F-actin is largely modulated by ATP hydrolysis, ionic strength under physiological level and accessory proteins that bind to actin to effect

the downstream cascade of activities (Wegner and Isenberg 1983). While actin monomers can be polymerized/depolymerized from either the pointed or barbed end, the rate of association/dissociation is different for each end. Its dissociation rate is independent of concentration of free actin monomers, but its association rate is proportional to concentration of free actin monomers (Yarmola, Dranishnikov et al. 2008). This critical concentration helps to maintain the equilibrium between monomers and filaments. It can be depicted by the equation where $K_{\text{off}} = C_{\text{critical}} \times K_{\text{on}}$ (C_{critical} refers to the critical concentration of free actin monomers required to maintain the equilibrium; K_{off} refers to the rate of dissociation, K_{on} refers to the rate of association) (Cooper GM. 2000). In the ATP state, actin monomers bind to the barbed (+) end and rapidly polymerize. The barbed end tends to elongate 5-10 times faster than the pointed end. As a result, different critical concentrations of actin monomers are required for polymerization at barbed and pointed ends to support the equilibrium. Once hydrolysis occurs, ATP is converted into ADP, ADP-actin quickly dissociates from the pointed (-) end. This steady state of actin polymerization/depolymerization is known as actin filament tread milling (Otterbein, Graceffa et al. 2001). In the process, free actin monomers transit between critical concentrations needed for polymerization at both ends, where the loss of monomers from the pointed end balances out the addition of monomers to the barbed end (**Figure 3**).

Besides tread milling, actin turnover is governed by two more mechanisms, which are dynamic instability and annealing/fragmentation (Littlefield and Fowler 2002). Under dynamic instability (**Figure 3**), actin monomers polymerize from both barbed and pointed ends due to higher affinity of actin filaments for ATP-capped actin monomers. Simultaneously, actin filaments can also depolymerize from both barbed

and pointed ends due to lower affinity of actin filaments for ADP-capped actin monomers. Dissociation of actin monomers in the ADP state from the barbed end is the rate-limiting step for actin filament turnover under dynamic instability. Actin filaments may undergo another turnover mechanism that is annealing and fragmentation (**Figure 3**). Polymerization/depolymerization occurs when actin fragments join or break off from the actin filament. Dissociation of actin monomer then occurs from the actin fragments generated from barbed and pointed ends. The exact mechanism of which end of the filament the actin fragments associate to or dissociate from is unknown, so does the role of ATP hydrolysis under this process. One of the factors that might affect the filament turnover rate is the size of actin fragments, which are generally less than ten monomers.

Figure 3. Mechanisms that govern actin turnover. A) Treadmilling occurs through a steady state association of ATP-actin at barbed ends and dissociation of ADP-actin at pointed ends. B) Dynamic instability involves ATP-capped filaments that polymerize due to their high affinity for actin monomers and ADP-capped filaments that depolymerize due to their low affinity for actin monomers. C) Another mechanism that controls actin turnover is annealing and fragmentation at filament ends. The average size of the fragments can be the limiting factor for this process. (Figure adapted and modified from Littlefield and Fowler 2002)

1.2.3 Actin and polarisome complex

During polarization establishment, actin plays dual roles as both the spatial cue for the initiation of a bud site and the subsequent cytoskeletal component in the cell (Ayscough, Stryker et al. 1997). Actin cables assemble at the bud tip and escort secretory vesicles and organelles along the cables (**Figure 4**), emphasizing the polarity established by temporal and spatial cues during bud site selection. Assembly of the actin cables is a complex and dynamic process that cannot be accomplished without the help of a vast number of actin-binding proteins (ABPs) in the cell (Pollard and Borisy 2003). These ABPs possess different functions, ranging from sequestration of actin monomer, capping of the barbed (+) or pointed (-) end, filament fragmentation and cross-linking and many more. Numerous structures of G-actin bound to ABPs have been solved by x-ray crystallography. Some structures of F-actin bound to ABPs have also been reconstructed at low resolutions by Cryo-electron microscopy (Dominguez and Holmes 2011).

Under in vitro condition, altering the ionic strength in solution can change actin polymerization (Wen and Rubenstein 2009). In cells, ionic strength is maintained constantly. Thus, a more complicated mechanism is involved in actin polymerization. A multiprotein complex known as the polarisome, gathers at the bud tip during polarization and assemble actin cables at the site of cell growth in both yeast and filamentous fungi. In the budding yeast, polarisome consists of different classes of proteins with specific role, such as nucleation factors (NFs, e.g. Bni1), nucleation promoting factors (NPFs, e.g. Pea2 and Bud6), and scaffold proteins (e.g. Spa2/Sph1) (Graziano, DuPage et al. 2011, Tu, Graziano et al. 2012). These proteins tend to have high level of intrinsically disordered regions (IDRs), which are important for protein

coacervation and phase separation (Miao, Tipakornsawapak et al. 2018). Proteins with a high percentage of IDRs often contain sites for post-translational modifications (PTMs) such as phosphorylation. PTMs play a crucial role in regulating protein function by varying the protein conformation and affecting its interaction with other binding partners (Groban, Narayanan et al. 2006).

Figure 4. Polarized cell growth requires the assistance of polarisome complex. Different protein members of the polarisome complex gather at the bud tip to direct polarized cell growth through a cascade of events. The polarisome promotes transport of damaged proteins from daughter to mother cells as well, which involves actin cables. (Figure adapted and modified from Liu, Larsson et al. 2010)

1.2.3.1 Nucleation factors

There are three main classes of nucleation factors identified to date - Arp2/3 complex, the formins, and tandem-monomer-binding nucleators that include Spire, cordon-bleu (Cobl), and leiomodin (Lmod) (Chesarone and Goode 2009).

Arp2/3 complex, the first class of nucleation factors discovered, is a protein complex consisting of seven subunits (Welch, DePace et al. 1997) (**Figure 5A**). Of these subunits, structure of two actin-related proteins ARP2 and ARP3 closely resemble that of G-actin. In addition to Arp2/3 complex, the Wiskott-Aldrich syndrome family

proteins (WASP) are required to nucleate actins (Yarar, To et al. 1999). The WASP proteins have WCA domains, where W is the WH2 domain responsible for binding G-actin, CA domain comprises the Arp2/3 binding central/acidic sequence responsible for Arp2/3 binding. The Arp2/3 complex initiates actin polymerization by binding to a pre-existing actin filament and to the CA domain of WASP, generating a nucleation core from which the actin extends (Bieling, Hansen et al. 2018). This results in new filaments growing on pre-existing filaments to form branched actin filaments. Each new actin filament extends from the existing filament at a highly characteristic 70° angle, leading to a Y-branched network of functional actin cytoskeleton (Goley, Rammohan et al. 2010).

Formins, the second class of nucleation factors and with majority as effector of Rho-GTPase, have multiple highly conserved domains such as formin homology (FH) domains, WH2 (WASP-Homology 2, or Wiskott-Aldrich homology 2) domain, Forkhead-associated (FHA) domain, Diaphanous autoregulatory (DAD) domain, and PDZ domain (Liu, Linardopoulou et al. 2010) (**Figure 5A**). Formins are generally recognized by the presence of three FH domains FH1, FH2, and FH3. All members of the formins contain FH1 and FH2 domains, and some contains FH3 domain (Petersen, Nielsen et al. 1998). FH1 is a proline-rich domain that regulates its interaction with a diverse pool of proteins, such as ABP profilin, and proteins that contain SH3 (Src homology 3) or WW domain (Frazier and Field 1997). The FH2 domain is responsible for its homodimeric self-assembly, by which two formins would bind to each other. Moreover, it is this domain that accounts for its actin nucleating activity (Le Clainche and Carlier 2008). Formins polymerize actin through a different mechanism as compared to Arp2/3 complex. Instead of generating branched actin

filaments, formin dimer nucleates actin by binding and assembling unbranched actin to form actin cables (Chesarone and Goode 2009). Monomeric FH2 domain that caps the barbed end of actin filament does not have a nucleating function, while dimeric one stabilizes actin dimers by binding to the barbed end of actin and initiates its assembly (Dominguez and Holmes 2011). Furthermore, formins act as elongation factors. The proline-rich FH1 domain in formins binds to a pool of monomeric actins that are sequestered by profilin, facilitating its delivery to the barbed end of the actin filament to allow elongation. Together, FH1 and FH2 domains in formins cooperate to nucleate and elongate the unbranched actin filament (Michelot, Guérin et al. 2005, Chesarone and Goode 2009). Formin in association with the barbed end of the growing actin filament also inhibits various capping proteins in the cell from abolishing polymerization abruptly (Chesarone and Goode 2009).

The third class of nucleation factor is tandem-monomer-binding nucleator (**Figure 5A**). This is a relatively new class of nucleation factors compared to Arp2/3 and formins. Some of the members include Spire, cordon-bleu (Cobl), leiomodulin (Lmod), JMY and adenomatous polyposis coli (APC) (Lei, Omotade et al. 2016). They contain tandem G-actin binding motifs, which are typically WH2 domains that recruit actin monomers to form an actin nucleation seed/nucleus in preparation for filament elongation. Some members contain other actin-binding motifs such as monomer-binding linker (MBL) in Spire and JMY, actin-binding helices (Ah), the leucine-rich repeat (LRR) in leiomodulin, and the actin-nucleating sequences (ANS1, 2) in APC (Firat-Karalar and Welch 2011). Though these members nucleate actin by forming nucleation seeds, each nucleator generates nucleation seeds with different structural arrangements. A recent study has revealed that actin monomers are arranged into

either a side-by-side or a straight-longitudinal conformation that is subsequently added to the growing filament by Spire (Ducka, Joel et al. 2010). It has been suggested that similar to Spire, JMY tethers actin monomers along the long-pitch helix of actin filament, while Cobl, Lmod and APC stabilize short-pitch trimeric actin nuclei and add on to the actin filament (Namgoong, Boczkowska et al. 2011). However, structural information on the detailed mechanism of how each member nucleates actin is limited and, further investigation to unravel their mechanisms is needed.

Figure 5. Factors involved in actin nucleation and polymerization. A) Three classes of actin nucleation factors. Arp2/3 complex recruits actin monomers via WH2 domain in WASP and generate branched actin filaments. Formins nucleate actin by stabilizing actin dimers and binding to the barbed ends to allow addition of actin subunits. Spire, Cobl and Lmod contain tandem WH2 domains that associate with actin monomers. Each member nucleates actin by forming nucleation seeds with different structural arrangements. B) Two classes of elongation factors. Formins act as elongation factor by protecting the barbed end of actin filament from capping proteins through their dimeric FH2 domains. Ena/VASP is a relatively new class of actin elongation factors that bind to the barbed ends of actin through its tetrameric formation and accelerate actin elongation (Figure adapted and modified from Chesarone and Goode 2009).

1.2.3.2 Nucleation promoting factors

As mentioned earlier, nucleating activity of nucleation factors (NFs) only is insufficient to activate actin filament assembly. Nucleation promoting factors (NPFs) are essential in boosting nucleation factors' activity (Chesarone and Goode 2009). For example, Arp2/3 complex requires WASP, which is a NPF, to be activated prior to actin nucleation (Padrick, Doolittle et al. 2011). In budding yeast, formins Bni1 and Bnr1 are NFs involved in establishing polarized cell growth, however, they require the help from a NPF Bud6 to enhance their activity. Studies have shown that Bud6 promotes Bni1-mediated actin filament assembly during the nucleation stage by gathering and delivering actin monomers to the Diaphanous Autoregulatory Domain (DAD) of Bni1 (Graziano, DuPage et al. 2011). Deletion of Bud6 in budding yeast led to a stark loss of actin cables. Contrary to profilin, which promotes elongation of actin filament, Bud6 only affects actin assembly during the nucleation stage (Graziano, DuPage et al. 2011). Bud6 and profilin work synergistically to assemble filaments.

Pea2, another NPF found in the polarisome complex, is responsible for the localization of Bni1 to the bud tip where polarization of the cell originates, thereby regulating assembly of actin cables (Wang, Huang et al. 2016). During mating, Pea2 is needed for bipolar budding. Pea2 is composed of a coiled-coil domain, its nucleating promoting activity is inter-dependent on the activity of polarisome scaffold protein Spa2 (Sheu, Santos et al. 1998). Spa2 acts as a scaffold for Mitogen-activated protein kinase kinase 1 (Mkk1) and Mitogen-activated protein kinase 1 (Mpk1) that are involved in the signaling pathway responsible for cell wall integrity (Levin 2005). Current studies showed that mutation in Pea2 resulted in failure of Spa2 localization

to the bud site, though the expression of Spa2 was unaffected. While mutation in Spa2 led to disturbing Pea2 expression (Sheu, Santos et al. 1998). This indicated that Spa2 and Pea2 function as a complex in formation of bipolar budding and provide proper guidance in polarized cell growth during mating.

1.2.3.3 Elongation factors

Nucleation factors generate actin nucleation seeds with the help from NPFs. Once the nucleation seeds are formed, actin filaments extend progressively from the barbed ends until there is an exhaustion of free actin monomers, or capping proteins come into play and to terminate the elongation process (Chesarone and Goode 2009). There is an intrinsically high association rate for capping proteins to the barbed ends of actin filaments (Shekhar and Carlier 2016), thus it is important to keep capping proteins under stringent control to prevent them from hindering filament elongation. Cells express various proteins that bind to actin filaments and move along the filaments, inhibiting capping of the barbed ends by capping proteins. These proteins are known as the elongation factors (Harris and Higgs 2004, Chesarone and Goode 2009). Two classes of proteins under elongation factors have been discovered to date, which are formins and Ena/VASP (**Figure 5B**).

Formins are both a nucleation factor and an elongation factor (Chesarone and Goode 2009). Dimeric formin associates with the barbed ends of an actin dimer through its FH2 domain, while FH1 domain recruits profilin-actin complexes, and deliver the actin to the growing end of the filaments that are capped by formins FH2 domains (Breitsprecher and Goode 2013). The cooperation between FH1 and FH2 domains of

formins accelerates the elongation process by one to five folds (Chesarone and Goode 2009).

Enabled/vasodilator-stimulated phosphoprotein (Ena/VASP) is a class of highly conserved actin regulatory proteins. They consist of several domains such as EVH1/2 and a central proline-rich region (PRR) (Han, Chung et al. 2002). The EVH1 domain is responsible for localization and specific ligand binding; the C-terminal EVH2 caps barbed ends of actin filament; the central proline-rich region consists of binding motifs that recognize profilin-actin complexes and G-actins (Renfranz and Beckerle 2002). In comparison with formins, Ena/VASP associates with profilin-actin complexes to deliver actin to the barbed end of the growing filament. They also protect the barbed ends from inhibition by capping proteins. In contrast to formins with a dimeric form to elongate actin filament, Ena/VASP has a tetrameric form (Chesarone and Goode 2009). Furthermore, Ena/VASP requires significantly higher concentration to exert the same effect as formins, which could be potentially due to the difference between their binding affinities to actin barbed ends (Chesarone and Goode 2009, Breitsprecher, Kieseewetter et al. 2011).

1.3 Aip5 as a novel class of actin NF and an NPF of Bni1

1.3.1 Aip5, a previously unknown polarisome protein in budding yeast

Actin-interacting protein 5 (Aip5) is a novel polarisome protein that has not been well characterized up to date (Shih, Reck-Peterson et al. 2005). It is a functionally unknown protein encoded from the gene YFR016C in budding yeast. In collaboration with Prof. Miao Yansong's lab, we proposed Aip5 as the fourth class of actin

nucleation factor that nucleate actin filaments and a new NPF to augment Bni1's nucleation activity.

Aip5 was first mentioned by Shih, Reck-Peterson et al. in a study carried out in 2005. Immunoprecipitation coupled with tandem mass spectrometry analysis were performed to investigate on the pool of proteins that associate with Spa2 in regulating cell polarity (Shih, Reck-Peterson et al. 2005). Among the proteins identified, the protein encoded by YFR016C was one of the new binding partners of Spa2. Recent studies conducted by Miao, Wong et al. on in vitro reconstitution of actin cable assembly by Bni unveiled this uncharacterized protein again, which is encoded by YFR016C, and named as Aip5 (Miao, Wong et al. 2013). Aip5 appears to be evolutionarily conserved only among fungal species, including various pathogenic filamentous fungi for human and plants.

1.3.2 Aip5 localizes to the bud tip in presence of Spa2

To examine the localization of Aip5 in the polarisome, Aip5 was tagged with green fluorescent protein at its C-terminal. It was observed that the localization of Aip5 depends fully on the presence of Spa2. Using PHYRE2 and ANCHOR prediction softwares, Spa2 is predicted to have an intrinsically disordered N-terminus and an ordered C-terminus (1306-1466aa), which is similar to the predicted secondary structure of Aip5. Based on our collaborator's results, Aip5's localization signal was completely abolished in presence of a truncated Spa2-C mutant (1-1305aa). Spa2 Δ C mutant was still able to localize to the bud tip and bud neck and behaved likewise comparing to its wild type counterpart (**Figure 6A and 6B**). The results indicated that

Spa2 most probably localizes to the bud tip via its N-terminus, while it recruits Aip5 via its C-terminus.

Figure 6. Aip5 is a novel polarisome protein that nucleates actin polymerization-dependent and independent of Bni1. A) Representative maximum Z-projection images of Aip5-GFP protein localization at the presumptive bud site in wild-type and polarisome protein mutants. S, small bud tip; M, medium bud tip; BN, bud neck. Scale bar = 3 μ m. B) Domain schematic of the polarisome scaffold protein Spa2. Three classes of Aip5 localization that are quantitatively determined by the ratio of fluorescence signal intensity (R) of Aip5-GFP at presumptive bud site and cytoplasm. (Aip5-GFP, n=83; *spa2Δ* Aip5-GFP, n=60; *bni1Δ* Aip5-GFP, n=51; *bud6Δ* Aip5-GFP, n=69). **Figure adapted from Prof. Miao Yansong's lab, courtesy of Ms Xie Ying.**

1.3.3 Localization of Aip5 partially depends on NPF Bud6 and NF Bni1

More studies revealed that localization of Aip5 in the polarisome depends partially on Bni1 and Bud6. Through tagging of fluorescent protein, Aip5 was shown to colocalize with other polarisome members Spa2, Bni1, and Bud6. Mutations of both Bni1 and Bud6 led to diminished Aip5-GFP localization (**Figure 6A and 6B**). Localization of NF Bni1 depends highly on F-actin in the cell. When cells are treated with Latrunculin A, a potent inhibitor of actin polymerization, F-actin structures will be rapidly disrupted in the cell. Incubation of Bni1-GFP with Latrunculin A for 30 min showed a drastic reduction in cytosolic Bni1, instead, Bni1 was found to be concentrated at the bud cortex with abnormal localization. Among these cells, roughly about 30% of the cells showed Bni1 localized at the bud neck as a ring (Ozaki-

Kuroda, Yamamoto et al. 2001). Similarly, Aip5 in our study demonstrated comparable level of abnormal localization at the bud tip when incubated with Latrunculin A. Aip5 patches produced by Latrunculin A treatment remained partially in Spa2 mutant (**Figure 7A and 7B**). However, they vanished completely in Spa2 and Bni1 double mutant. This implied a potential interaction between Aip5 and Bni1.

Figure 7. Aip5 localization changes upon Latrunculin A treatment. A) Representative maximum Z-projection fluorescence images of Aip5-GFP localization at the presumptive bud site in wild-type and several polarisome protein mutants with 10 μ M Lat A treatment for 30 min before imaging. B) Quantification of relative ratio (R) of Aip5-GFP signal intensity at medium bud tip in the conditions of (A). Aip5-GFP, n=112 cells; *spa2Δ* Aip5-GFP, n=82 cells; *spa2Δ bni1Δ* Aip5-GFP, n=51 cells; *bni1Δ* Aip5-GFP, n=56 cells. **Figure adapted from Prof. Miao Yansong's lab, courtesy of Ms Xie Ying.**

1.3.4 Aip5 as a potent NF and an NPF of Bni1

Given that Bni1 and Bud6 are known components of the polarisome involved in actin nucleation, we were interested to find out whether Aip5 has similar function. Aip5 was expressed and purified using a yeast-based galactose-inducible gene expression system (Ying, Xiao et al. , Miao, Han et al. 2016). Results from pyrene actin polymerization assay demonstrated actin nucleation activity for Aip5, which promoted actin filament assembly in a concentration-dependent manner ranging from submicromolar to micromolar concentrations (**Figure 8A**). In another pyrene actin

polymerization assay Aip5 enhanced the actin filament assembly effect of Bni1 (FH1-COOH domain) (**Figure 8B**). Both Aip5 and Bni1-FH1COOH could reduce actin filament lengths while increasing actin filament number. When Aip5 was combined with Bni1-FH1COOH, the number of short actin filament escalated significantly (**Figure 8C and 8D**). To test the effect of Aip5 wild type and mutant on the rate of polarized cell growth, Latrunculin A was added to perturb actin filament assembly. Aip5 mutant was shown to be more sensitive to the perturbation as compared to its wild type. In the case of Aip5 and Bni1 double mutants, Latrunculin A addition led to stronger cell growth inhibition. Together, the evidence demonstrated that Aip5 functions as an NF and an NPF of Bni1.

Figure 8. Aip5 functions as an NF and an NPF of Bni1. A) Pyrene actin polymerization reaction with increased concentration of full length Aip5. B) Pyrene actin polymerization reaction of Aip5 in the presence of Bni1FH1COOH. C) Representative fluorescence images of Acti-stain 488 phalloidin labeled-actin filaments that are polymerized by Aip5 in the absence and presence of Bni1FH1COOH for 30 min. Scale bar represents 5 μ m. D) Quantification of the number and length of actin filament in (C). A total of 300 filaments from 15 images at an area size of 80 μ m² were analyzed in each sample. **Figure adapted from Prof. Miao Yansong's lab, courtesy of Ms Xie Ying.**

1.3.5 Aip5 functions as both an NF and an NPF of Bni1 through its C-terminal domain

Similar to Spa2, Aip5 was predicted to have an unstructured N-terminus and a well-structured C-terminal domain from residues 1132 to 1234. Based on the prediction, several constructs were generated from the C-terminal region. Among these, the construct with residues 1110-1233 (named as Aip5-C) displayed high solubility in bacterial protein expression system. We then tested its activity through pyrene actin polymerization assays, and found that it displayed robust nucleating activity (**Figure 9A**). At the same time, its activity was also inhibited by a capping reagent cytochalasin D (CD) that caps the barbed ends of actin filaments (**Figure 9B**). This indicated that Aip5 binds to the barbed end of the actin filament via its C-terminal domain. To determine the binding affinity between Aip5-C and G-actin, Alexa-488 labelled 30 nM Aip5-C was titrated with non-polymerized actin (NP-actin) and the K_d was determined to be approximately $1.49 \pm 0.94 \mu\text{M}$ from anisotropy binding measurements (**Figure 9C**). We then decided to examine whether Aip5-C was as potent as formin, which is a strong NF that displaces capping protein from barbed ends of actin filaments. CapZ was chosen as the capping protein for the experiment due to its high affinity for barbed ends (Schafer, Jennings et al. 1996). Our results showed that Aip5-C could not restore actin elongation that was blocked by CapZ like formin (**Figure 9D**). However, when co-incubated with Bni1-FH1COOH, Aip5-C enhanced the activity of Bni1 displacing CapZ from the filament barbed ends (**Figure 9D**). All above results clearly showed that Aip5 acts as both an NF and an NPF of Bni1, through the C-terminal region.

Figure 9. Aip5 functions as an NF and NPF of Bni1 through C-terminal domain. A) Aip5-C nucleates spontaneous actin polymerization using pyrene actin polymerization assay. B) Actin polymerization by Aip5-C with or without 100 nM cytochalasin D in pyrene actin assay. C) Fluorescence anisotropy binding measurements of Alexa-488 labelled 30 nM Aip5-C that is titrated by non-polymerized actin (NP-actin). Data are represented by circles from three biological replicates, and Hill slope equation fit is shown as solid line. D) Barbed end capping assay using 0.5 μ M seeds, 10 nM CapZ, and additional 0.5 μ M monomeric pyrene actin, in the presence of Aip5-C with or without Bni1FH1COOH. **Figure adapted from Prof. Miao Yansong's lab, courtesy of Ms Xie Ying.**

1.4 Aims and objectives

Being a single-celled eukaryote, the genome of budding yeast *Saccharomyces cerevisiae* is the first fully sequenced eukaryotic genome (Goffeau, Barrell et al. 1996, Duina, Miller et al. 2014). Budding yeast has been proven to be a useful model for studying molecular mechanisms that regulate polarized cell growth because of its simple genetics and asymmetric cell division (Higuchi-Sanabria, Pernice et al. 2014).

Given the novelty and significance of Aip5 in actin nucleation and a potentially new NF and NPF in actin polymerization in budding yeast, we are interested to

biochemically and structurally characterize it, in collaboration with Prof. Miao Yansong's lab. Particularly for its functional C-terminal domain, it would be of our great interest to elucidate its structure and unravel its functional mechanism. This project involves:

- Purify and crystallize Aip5-C
- Conduct structural comparisons between Aip5-C and its homologs
- Generate Aip5-C mutants and test their effects on actin nucleation
- Purify and crystallize Aip5-C mutant
- Conduct structural analysis and comparison between Aip5-C and Aip5-C mutant
- Unveil potential binding site on Aip5-C and perform binding test between Aip5-C and Aip5-C mutant with G-actin

2. MATERIALS & METHODS

2.1 Molecular cloning

The construct for Aip5-C was designed consisting of residues 1110-1233. Cloning of Aip5 C-terminal was conducted by Protein Production Platform (PPP, NTU). pNIC28-Bsa4 vector that includes a cleavable N-terminal His6 tag and kanamycin resistance gene was used for the cloning. Aip5-C mutants were generated by site-directed mutagenic PCR with Q5 High-Fidelity DNA polymerase following the manufacturer's protocol, and sequences of the mutants were confirmed by DNA sequencing from 1st Base.

2.2 Aip5-C mutagenesis

Site directed mutagenesis was utilized to introduce single amino acid mutation into native Aip5-C. Reaction mixture in 25 μ l contained 5 μ l 5 $\times$ Q5 reaction buffer, 0.5 μ l

10mM dNTP, 1.25 μ l 10 μ M each of forward and reverse primers, 50ng Aip5-C DNA template, 5 μ l 5 $\times$ Q5 High GC Enhancer (optional), and 0.5 U (0.25 μ l) Q5 High-Fidelity DNA Polymerase. Listed below are the forward and reverse primers used for each mutagenesis of Aip5-C.

Aip5-C Ser1165Cys

F 5' –ATCGCTTATGCACAATCCTGACGGCCAACCGAA– 3'

R 5' – GGATTGTGCATAAGCGATTTGTTCTGGGGATC –3'

Aip5-C Leu1150Ala

F 5' – GTCAGCAGCCGGAGGCGGTTTC – 3'

R 5' – CCGGCTGCTGACGTGTATATATATACAGG –3'

Aip5-C Asn1162Ala

F 5' – AGAACAGCGCGCTTAAGCACAATCCTG – 3'

R 5' – TAAGCGCGCTGTTCTGGGGATCATATG –3'

Aip5-C L1150A/N1162A/S1165C triple mutant

F 5' – CCCCAGAACAGCGCGCTTATGCAC – 3'

R 5' – ATCATATGGAAACCGCCTC –3'

Aip5-C Met1157Ala

F 5' – TTCCATGCGATCCCCAGAACAAATCGCTTAA – 3'

R 5' – GGATCGCATGGAAACCGCCTCCG –3'

Aip5-C Δ 1143-1233

F 5' – TCCAATCCCCTGTATATATATACACGTCCTAGC – 3'

R 5' – TATACAGGGGATTGGAAGTACAGGTTCTCG –3'

Aip5-C T1148_M1157del (loop deletion)

F 5' – GTATATATATACATCCCCAGAACAAATCGCTTAAG – 3'

R 5' – TGTTCCTGGGGATGTATATATATACAGGTTCTTTTC –3'

The reaction mixtures were subjected to the following thermocycling conditions: 1 × 98°C for 30 seconds, 16 × (98°C for 10 seconds, a gradient annealing temperatures ranging from 58°C to 68°C for 30 seconds, 72°C for 4 min and 10 seconds). Final extension was carried out at 72°C for 4 min and the reaction mixtures were kept at 4°C. For enzymatic digestion, 1 µl of DpnI and 2 µl of 10× cutsmart buffer (NEW England BioLabs) were added to the 20 µl reaction mixture and incubated at 37°C for 1 hour. The digested mixture was then transformed into *E. coli* DH5α and plated onto 2× YT agar with 50 µg/ml of kanamycin. Plasmid DNA was extracted and verified by sequencing to ensure the right mutation was introduced.

2.3 Protein expression and purification

Expression plasmids containing Aip5-C or Aip5-C mutants were transformed into *E. coli* strain BL21(DE3) through heat shock. Bacterial cells carrying the recombinant plasmids were grown in 2xYT medium with the addition of 50 µg/ml of Kanamycin at 37 °C to an optical density (OD) 600 of 0.6. Protein expression was induced by 0.5 mM Isopropyl-b-D- thiogalactopyranoside (IPTG) (Gold Biotechnology) at 16°C overnight.

The same *E. coli* strain was used for transformation of Selenomethionine-labeled Aip5-C (SeMet Aip5-C). Cells were grown at 37 °C in M9 minimal medium supplemented with glucose, standard trace elements, and the corresponding antibiotic. The temperature was cooled down to 16°C when the optical density (OD) 600 of the cells reached 1.0. Then a mixture of amino acids with the exception of L-methionine was added, followed by the addition of 0.025 mg/ml of L-selenomethionine. The cells

were allowed to grow for 20 min prior to induction with 0.5 mM IPTG at 16°C overnight.

The cells were harvested by centrifugation at 4°C, 5,000 g for 15 min. The pellet was resuspended and lysed by EmulsiFlex-C3 (www.avestin.com). The lysate was clarified by centrifugation at 48,300 g for 40 min. The supernatant was purified by nickel affinity chromatography. Fractions of the target protein were pooled following gradient elution with increasing imidazole concentration. The protein was dialyzed against the buffer 20 mM Tris-HCl pH 8.0, 150 mM NaCl, with the addition of His6-tagged TEV protease to cleave the His6-tag. A second nickel affinity chromatography was performed to remove the cleaved His6-tag and TEV protease. The protein was further purified by size-exclusion chromatography using a HiLoad 16/600 Superdex75 column (GE Healthcare) with elution buffer 20 mM Tris-HCl pH 8.0, 150 mM NaCl. The target protein was concentrated to a final concentration of 18.5 mg/ml for crystallization screening.

Actin purification was carried out by Ms Xie Ying from Miao Yansong's lab. To prepare monomeric Ca^{2+} -ATP-actin, 2.5 g rabbit muscle acetone powder (Pel-Freez Inc, USA) was dissolved in G-buffer (5 mM Tris-HCl pH 8.0, 0.2 mM ATP, 0.1 mM CaCl_2 , 0.5 mM DTT) for 3- 4 times in 4°C for 30 min, the supernatant was filtered through cheese cloth and further cleared by centrifugation at 27,170 g at 4°C for 30 min in the JA25.50 rotor. The collected supernatant actin was polymerized by adding 50 mM KCl and 2 mM MgCl_2 for 1 hour, stirred gently and followed by addition of 0.8 M KCl for 30 min at 4 °C. The polymerized F-actin was pelleted down by centrifugation at 142,160 g using Ti45 rotor (Beckman Coulter, Inc. USA) for 3 hour,

and then depolymerized by homogenized and sonication (3 sec on, 30 sec off, 3-4 cycles in total). A further 48-hour dialysis against G-buffer at 4°C was performed with 3 times changing buffer in between. The top 2/3 supernatant actin fraction was collected after the centrifugation at 197,810 g for 2 h and 30 min using Ti55 rotor (Beckman Coulter, Inc., USA). The supernatant was further loaded for gel filtration chromatography as Sephacryl S-300 HR in G-buffer with 0.1 % sodium azide to obtain monomeric Ca²⁺-ATP-actin.

2.4 Crystallization

Crystal screening using commercially available kits for native Aip5-C, SeMet Aip5-C, and Aip5-C-L46A/N54A/S58C mutant was carried out through sitting drop vapor-diffusion method using 96 well intelli-plate® (Art Robbins Instruments). Gryphon LCP (Art Robbins Instruments) and Nanoliter Pipetting System mosquito® (TTP LabTech) were used to perform the screening. Three drops with protein to reservoir ratios (0.2 µl : 0.1 µl, 0.15 µl : 0.15 µl, 0.1 µl : 0.2 µl) were set up per condition in each well of the intelli-plate®. The plates were incubated in RockImager (Formulatrix) at 20 °C for over 2 weeks. Native Aip5-C crystal was crystallized in 0.1 M Tris-HCl pH 7.0, 20% polyethylene glycol 1000 with a protein to reservoir ratio at 2.0 µl to 1.0 µl. SeMet Aip5-C crystal was crystallized in 0.2 M potassium sodium tartrate and 20% polyethylene glycol 3350 with a protein to reservoir ratio at 2.0 µl to 1.0 µl. Aip5-C-L46A/N54A/S58C was crystallized in sodium phosphate citrate pH 4.2, 40% polyethylene glycol 300 with a protein to reservoir ratio at 2.0 µl to 1.0 µl. Crystals were picked directly from the drop and cryo-protected in reservoir solution supplemented with 15-25% v/v glycerol with the exception of Aip5-C-

L46A/N54A/S58C, which did not require additional glycerol. Crystals were flash frozen in liquid nitrogen prior to data collection.

2.5 Data collection and structure determination

Diffraction data for native Aip5-C and multiple wavelength anomalous dispersion (MAD) data for SeMet Aip5-C were collected at BL13B1 beamline of the National Synchrotron Radiation Research Centre, Taiwan. The dataset for Aip5-C-L46A/N54A/S58C mutant was collected at in-house X-ray facility using Rigaku FR-X. The position of selenium site was located by SHELDX. The initial experimental phase was calculated using PHENIX AUTOSOL (Adams et al., 2010), followed by model building in PHENIX AUTOBUILD (Adams et al., 2010). The diffraction data of native and mutant Aip5 were collected, subsequently indexed, integrated and scaled using HKL3000 (Otwinowski et al.). Structure of mutant Aip5-C was determined by molecular replacement with PHASER (Adams et al., 2010) using native Aip5-C structure as a search model. All structures were built using PHENIX AUTOBUILD (Adams et al., 2010) and manually improved in COOT (Emsley et al., 2010). Subsequent refinement was performed with a combination of COOT and PHENIX.REFINE. Water molecules were included in the last round of refinement. The crystallographic statistics of the structures are summarized in **Table 1**. All structural figures were prepared with PyMol (Schrödinger Team, 2010).

2.6 Oligomeric state determination by analytical size-exclusion chromatography

Size-exclusion chromatography was performed for Aip5-C Aip5-C mutant in superdex 75 10/300 GL (GE healthcare) equilibrated in buffer 20 mM Tris-HCl pH 8.0, 150 mM NaCl. 100 µl of 2 mg/ml of each purified Aip5-C/Aip5-C variants were

injected into the column, and chromatography was carried out at 4 °C in an AKTA Purifier system. Elution volumes of each Aip5-C/Aip5-C variants were recorded. 100 µl of a mixture containing a series of standard proteins dissolved in buffer containing 20 mM Tris pH 8.0, 150 mM NaCl was injected into the column. Elution volumes of the standard proteins were used to calculate the molecular size and oligomeric state of Aip5-C and different Aip5-C variants.

2.7 Pyrene fluorescence assay of actin assembly

Pyrene fluorescence assays were done by Ms Xie Ying from Miao Yansong's lab. Pyrene-labeled actin was purchased from Cytoskeleton Inc. A 10 µM G-actin (5% pyrene actin) was firstly converted to Mg^{2+} -ATP-actin, and then mixed rapidly with various proteins in the G buffer. The spontaneous actin polymerization was initiated by 10× KME buffer mix (10 mM $MgCl_2$, 10 mM EGTA, and 500 mM KCl), with a total reaction volume as 120 µl. The pyrene-actin fluorescence signal was monitored in a plate reader Cytation 5 (BioTek, USA) at excitation and emission wavelengths of 365 and 407 nm, respectively. Experiments were performed with at least two biological replicates. The graphs shown in the figures are representative results plotted by GraphPad Prism 6.

For testing the barbed end capping activity, 10 µM of actin was polymerized in the F-buffer (5 mM Tris-HCl pH 8.0, 0.2 mM ATP, 0.1 mM $CaCl_2$, 0.5 mM DTT, 50 mM KCl, 1 mM $MgCl_2$ and 1 mM EGTA) at room temperature for 2 h. Actin seeds were created by shearing the filaments through a gauge 27^{1/2} needle for 20 times. 0.5 µM actin seeds were mixed with 0.5 µM G-actin (5% pyrene actin), the desired proteins, cytochalasin D (MP Biomedicals, USA), or CapZ (Hypermol, Germany) at room

temperature for 5 min before adding 10× KME to initiate actin assembly. To determine the actin critical concentration for polymerization, a various concentration of Mg^{2+} -ATP-actin (from 0.05 μM to 3 μM , 30% pyrene actin) were polymerized in the F-buffer with desired proteins at room temperature for 16 h, the pyrene fluorescence signal was recorded by the plate reader and plotted by GraphPad Prism 6. To test the protein activities in depolymerizing actin filament, 5 μM of actin (30% pyrene actin) was polymerized in the F-buffer at room temperature for 2 h before diluted to 0.1 μM . The decrease of pyrene fluorescence was monitored after the supplement of Cofilin or Aip5. The actin polymerization graphs were all plotted in the GraphPad Prism 6.

2.8 Microscale thermophoresis (MST)

Binding affinity between actin and Aip5-C loop peptides (loop: TSLAGGGFHM; loop-L1150A: TSAAGGGFHM) was measured by MST method. Aip5 peptides were mixed with 2 μM G-actin at a 1:1 ratio in a total volume of 20 μl , where various stock concentrations of Aip5 peptide were prepared with 2 mM as the highest concentration. The binding reactions were carried out in G-buffer with additional 0.05% Tween 20 at room temperature. Samples were loaded into MonolithTM NT.115 standard capillaries (Nanotemper Technologies), which were incubated at 25°C in the MonolithTM NT.115 apparatus (Nanotemper Technologies) before measurement. The data were collected at 25°C using the red LED at 5% (GREEN filter; excitation 515-525, emission 560-585 nm) and IR-Laser power at 80%. Data analyses and the binding curves were fit with Kd mode in MO analysis software (Nanotemper Technologies).

2.9 Protein sequence prediction and evolutionary analysis

To identify the Aip5 homologs and perform evolutionary analysis, the Aip5-C sequence was submitted as a query sequence using NCBI bBlastP with default parameters. The species identities scored greater than 50% were chosen. The sequence alignment was performed in the online server Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and the phylogenetic tree was generated by the interactive tree of life (iTOL) (<http://itol.embl.de/>). ANCHOR (<http://anchor.enzim.hu/>) and PHYRE2 protein fold recognition server (<http://www.sbg.bio.ic.ac.uk/~phyre/>) were used, respectively, in identifying intrinsically disordered region and conserved folded-domains. Structural homologs of Aip5-C were identified from Dali server (<http://ekhidna2.biocenter.helsinki.fi/dali/>). The sequence alignment of Aip5-C and its structural homologs was carried out through T-COFFEE (<http://tcoffee.crg.cat/>) and subsequently generated by ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/ESPript/>).

2.10 Deposited data

The coordinates and structure factors of both Aip5-C and Aip5-C-3m have been deposited in Protein Data Bank (PDB) with accession code 6ABR and 6ABS respectively.

3. RESULTS

3.1 Small scale expression of Aip5-C

Given the significance of C-terminal region of Aip5 in its biochemical activity, its structural information would provide atomic insights into its function in actin polymerization. Structure prediction using online server (Xtal Predict) predicts the C-terminal region of Aip5 adopts secondary structures consisting of helices and beta sheets. Based on the structure prediction, several truncations of the Aip5 C-terminus were designed and sent to NTU Protein Production Platform (PPP) for expression tests. Among these, construct with residues number 1110-1233 displayed high soluble expression.

Figure 10. Soluble constructs of Aip5-C. Different constructs of Aip5 C-terminus tested by NTU PPP. Green indicates soluble expression, while yellow indicates poor soluble expression.

Figure 11. SDS-PAGE from PPP showing soluble expression for different constructs. Construct H8 (Aip5 residues 1110-1233), highlighted in red rectangular box, exhibited high soluble expression.

3.2 Purification of Aip5-C in large scale

With the preliminary results from PPP, Aip5 residues 1110-1233, thereafter referred to as Aip5-C was chosen to be purified further in large scale. Large scale purification of Aip5-C was carried with two steps: 1) nickel affinity chromatography; 2) size exclusion chromatography (HiLoad 16/60 Superdex 75 pg column) (**Figure 11**).

The pI of Aip5-C is predicted to be 6.05 and its molecular weight is approximately 16.9 kDa (N-terminal His6 tag included). Thus the pH of the buffers used for purification was set to be 8.0. Both lysis and eluting buffers used during nickel affinity purification were phosphate based, consisting of 50 mM NaPi pH 8.0, 300 mM NaCl, while eluting buffer contained an additional 500 mM of imidazole. The protein was eluted with 13.2% imidazole, equivalent to approximately 66 mM of imidazole. Results were confirmed by Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), showing proteins bands of the correct molecular weight corresponding to the peak (**Figure 12A**). Target fractions were pooled and subjected to overnight cleavage via TEV protease. The Aip5-C with His-tag removed was passed through a second round of nickel affinity column to get rid of the uncleaved protein and subjected to size exclusion column.

The elution volume for Aip5-C from the 16/60 Superdex 200 pg column was 74.6 ml. The yield of the protein was extremely high and could be reflected from the UV absorption, which exceeded 2000 mAU (**Figure 12B**). Distinctive peak of Aip5-C elution could not be spotted due to its high yield, which caused the UV absorption to be saturated. Target protein was analyzed by SDS-PAGE, indicating high purity. We

also checked the sample for possible DNA contamination using NanoDrop2000 (Thermo Scientific). The final 260/280 ratio of the sample was 0.52, indicating no DNA contamination. Consequently, Aip5-C was concentrated to 18.5 mg/ml, and subjected to crystal screening.

Figure 12. Large scale purification of Aip5-C. A) Nickel affinity purification. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted with 13.2% of imidazole (equivalent to 66 mM). SDS-PAGE showing target protein bands corresponding to the peak. B) Size exclusion purification using 16/60 Superdex 75 pg column. Target protein was eluted out at 74.6 ml and target fractions were checked by SDS-PAGE. Final protein yield and protein purity were extremely high.

SeMet Aip5-C was purified using the same procedure as the above description (**Figure 12**). Briefly, protein lysate was clarified and purified through nickel affinity column, followed by overnight cleavage using TEV protease to remove the His6-tag and subsequent size exclusion chromatography (**Figure 13A and 13B**). Elution profile in the size exclusion column looked similar to the native Aip5-C. Target

fractions were pooled and concentrated to 18.5 mg/ml in preparation for crystal screening.

Figure 13. Large scale purification of SeMet Aip5-C. A) Nickel affinity purification. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted with 15.6% of imidazole (equivalent to 78 mM). SDS-PAGE showing target protein bands corresponding to the peak. B) Size exclusion purification using 16/60 Superdex 75 pg column. Target protein was eluted out at 77.1 ml and target fractions were checked by SDS-PAGE. Final protein yield and protein purity were relatively high.

3.3 Crystallization of Aip5-C

Protein concentration for Aip-C was set to be 18.5 mg/ml prior to crystal screening. Six screening kits were tried including Index, Pegion, Wizard, Cryo1&2, Memfrac/natrix and SaltRX, among which, Index, Pegion, and Wizard gave crystal hints. One good condition from Wizard consisting of 0.1M Tris-HCl pH 7.0 and 20% polyethylene glycol 1000 was chosen for further optimization, eventually growing nice and huge crystals (**Figure 14**). Crystals were first tested using in-house X-ray

Rigaku FR-X, which diffracted to approximately 2 Å, and subsequently sent to National Synchrotron Radiation Research Centre, Taiwan for data collection.

Figure 14. Crystals of Aip5-C. Nice and huge crystals were developed at a concentration of 18.5mg/ml in Wizard condition consisting of 0.1M Tris-HCl pH 7.0 and 20% polyethylene glycol 1000. Crystals were cryoprotected in mother liquor condition supplemented with 25% glycerol. These crystals diffracted to 2 Å.

Due to the novelty of Aip5-C, no similar structural homolog of Aip5-C could be used to solve its structure. As a result, two strategies were employed at the same time in attempt of solving the structure. The first strategy uses Single-wavelength anomalous dispersion of Sulphur atoms, known as native Sulphur-SAD. The second strategy uses Selenomethionine incorporation multiwavelength anomalous dispersion, SeMet-MAD. For S-SAD, mutagenesis was carried out to mutate serine residues to cysteine residues at different sites within the protein. For SeMet labeling, bacteria were cultured in M9 minimum media supplemented with selenomethionine to incorporate SeMet into the protein. SeMet labeled Aip5-C was purified with the aforementioned procedure. Protein concentration for SeMet Aip5-C was set to be 18 mg/ml. Four screening kits were tried including Index, Pegion, Cryo1&2, and Crystal Screen.

Among which, Pegion gave rise to nice crystals in almost every condition (**Figure 15**). Crystals were picked directly from the screening plate containing the condition 0.2 M potassium sodium tartrate and 20% polyethylene glycol 3350, and cryoprotected in the exact same condition with addition of 25% glycerol. These crystals diffracted to 2.5Å and were sent to National Synchrotron Radiation Research Centre, Taiwan for data collection.

Figure 15. Crystals of SeMet Aip5-C. Crystals of various shapes were developed at a concentration of 18mg/ml in almost every condition from Pegion screening kit. Crystals shown here were from Pegion condition consisting of 0.2 M potassium sodium tartrate and 20% polyethylene glycol 3350. Crystals were cryoprotected in mother liquor condition supplemented with 25% glycerol. These crystals diffracted to 2.5 Å.

3.4 Data collection and refinement of Aip5-C

Table 1: Summary of crystallographic data and refinement statistics for Aip5-C, SeMet Aip5-C, and Aip5-C-3m.

		Selenomethionine derivative		
	Native	MAD _{Infection}	MAD _{Remote}	Aip5-C-3m
PDB ID	6ABR			6ABS
Data Collection				
Wavelength (Å)	1.0000	0.9788	0.9636	1.5418
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions				
a, b, c (Å)	32.57, 82.77, 34.28	32.51, 82.47, 34.23	32.51, 82.46, 34.20	38.94, 54.60, 128.65
α, β, γ (°)	90, 107.367, 90	90, 107.466, 90	90, 107.464, 90	90, 90, 90
Resolution (Å)	50.0–2.0 (2.10–2.0)	50.0–2.5 (2.60–2.50)	50.0–2.5 (2.60–2.50)	50.0–2.2 (2.30–2.20)
No. of unique reflections	11332	10716	5830	14417
R _{sym} (%)	5.2 (14.2)	5.3 (10.7)	5.1 (10.3)	11.1 (63.9)
I/σ ₁	19.4 (7.8)	16.4 (7.6)	14.6 (7.6)	14 (2.3)
Completeness (%)	96.8 (89.1)	97.0 (98.3)	97.0 (98.6)	98.3 (99.0)
Redundancy	3.8 (3.4)	2.4 (2.4)	2.4 (2.4)	4.9 (3.8)
Refinement				
Resolution (Å)	32.7–2.0			37.3–2.2
R _{work} /R _{free} (%)	22.6/26.2			19.7/24.7
No. of reflections	11303			14179
No. of atoms	1548			1632
Protein	1480			1606
Ligand/ion	-			26
Water	68			90
<i>B</i> -factors	28.2			34.44
Protein	27.9			34.0
Ligand/ion	-			37.1
Water	32.8			42.6
rmsd				
Bond lengths (Å)	0.007			0.008
Bond angles (°)	1.17			1.12
Ramachandran				
Favored (%)	95.45			97.85
Allowed (%)	4.55			2.15
Outlier (%)	0			0

MAD diffraction data were collected from one crystal.
Values in parentheses are for the highest-resolution shell.

Diffraction data of a single Aip5-C crystal collected from the 13B1 beamline in Taiwan showed diffraction up to 2.0 Å. The space group was determined to be P2₁, with cell dimensions of a=32.573 Å, b=82.775 Å, c=34.278 Å; $\alpha=90^\circ$, $\beta=107.367^\circ$, $\gamma=90^\circ$ (**Table 1**). Two sets of data were collected from a single SeMet Aip5-C crystal at inflection and remote wavelengths. The position of selenium site was located by SHELDX. The initial experimental phase was calculated based on SeMet-MAD data using PHENIX AUTOSOL (Adams et al., 2010), followed by model building in PHENIX AUTOBUILD (Adams et al., 2010). The native structure of Aip5-C was successfully solved with the help from SeMet labeled Aip5-C. Subsequent model building and refinement through Coot (Emsley, Lohkamp et al. 2010) improved R_{work} and R_{free} values of Aip5-C to 0.226 and 0.262, respectively, at a resolution of 2.0 Å.

3.5 Crystal structure of Aip5-C

The structure of Aip5-C (PDB ID: 6ABR) was determined by selenomethionine-multiple wavelength anomalous diffraction (SeMet-MAD) phasing method. The protein was crystalized in the monoclinic space group P2₁, with two molecules per asymmetric unit, and gel filtration chromatography result of Aip5-C further confirmed the dimer state of Aip5-C (**Figure 16**). The elution volume of Aip5-C in superdex75 10/300 was 12 ml, which was compared with the elution volumes of a series of standard proteins consisting of blue dextran, conalbumin, ovalbumin, and ribonuclease A, to calculate its molecular weight in solution. Theoretically, molecular weight of a protomer of Aip5-C is 14480 Da, based on the elution volume of Aip5-C, it was estimated to be approximately 24 kDa, indicating dimer formation of Aip5-C in solution.

Figure 16. Determination of oligomeric state of Aip5-C in solution using superdex75 10/300 column. The elution volumes of a series of standard proteins consisting of blue dextran, conalbumin, ovalbumin, and ribonuclease A with known molecular weights and oligomeric states were used to extrapolate the calculated molecular weight of Aip5-C in solution. The molecular weights for the standard proteins are as follow: blue dextran – 2000 kDa, conalbumin – 76 kDa, ovalbumin – 45kDa, and ribonuclease A 13.7 kDa. Aip5-C was determined to be a dimer in solution, which is consistent with its crystallographic oligomeric state.

3.5.1 Presence of a thioredoxin fold structure in Aip5-C

Each molecule of Aip5-C displays a compact structure exhibiting a thioredoxin fold.

Both protomers A and B of Aip5-C consist of four beta strands in the core region flanked by four alpha helices (**Figure 17A**). This structural arrangement can be subdivided into an N-terminal $\beta 1\alpha 1\beta 2$ motif and a C-terminal $\beta 3\beta 4\alpha 3\alpha 4$ motif, representing a typical thioredoxin fold, which is common in thiol-disulfide oxidoreductase (Martin 1995). The two motifs of this thioredoxin fold are connected by $\alpha 2$ with some loop residues.

In the N-terminal region, the $\beta 1$ and $\beta 2$ run almost parallel, which are connected by $\alpha 1$. In contrast, the $\beta 3$ and $\beta 4$ are anti-parallel and connected by a short loop. Alpha helices are arranged in a characteristic manner as well. $\alpha 1$ and $\alpha 3$ from N- and C-

terminals of Aip5-C orientate in positions almost parallel to each other and on one side of the beta strands (**Figure 17A**). $\alpha 2$ that connects the C-terminal and N-terminal regions is located on the other side of the beta strands and is perpendicular to $\alpha 1$ in both protomer A and protomer B of Aip5-C. Together, the alpha helices sandwich the beta strands, thus aiding towards maintaining a compact thioredoxin fold.

Figure 17. Crystal structure of Aip5-C homodimer. A) The two protomers A and B are colored brightorange and violetpurple, respectively. B) Zoom in of the hydrogen bond network within the dimer interface of Aip5-C. $\beta 2$ from both protomers of Aip5-C dimer forming hydrogen bonds with water, in particularly water 3, 5, 10, 18, 20, 101 are involved.

3.5.2 Strong polar network within the dimer interface of Aip5-C

Dimerization of Aip5-C involves $\beta 2$ of each protomer in the crystal structure. The $\beta 2$ strands in protomers A and B of Aip5-C run anti-parallel to each other. Interactions within the dimer interface are predominantly by hydrogen bonds. Waters 3, 5, 10, 18, 20, 101 form the rich hydrogen-bonding network that mediates the dimerization between the two $\beta 2$ within a dimer (**Figure 17B**). Of which, Thr1176 (protomer A) interacts with Asp1183 (protomer B) via water10 and 101. Carbonyl oxygen of the C-alpha of Tyr1177 from protomer A interacts indirectly with the carbonyl oxygen of

the C-alpha of Tyr1177 from protomer B via water 5 and 18 through hydrogen bonds (**Figure 17B**). The network interaction of these hydrogen bonds stabilizes the dimer formation in Aip5-C.

3.6 Sequence alignment and structural comparison of Aip5-C with cGrx2

Aip5-C adopts a classical thioredoxin fold, which typically comprises approximately 80 residues (Martin 1995). Different proteins may contain extra residues in addition to the fold. Thioredoxin fold is commonly found in oxidoreductases. Oxidoreductases consist of a CxxC motif, an important switch for the activity of many redox proteins. The two cysteines are separated by two other residues and are responsible for catalysis and reduction of disulfide bonds. Generally an alpha helix is present downstream of this CxxC motif in the thioredoxin fold (Martin 1995).

A structural homologue cGrx2 was chosen from the closest matches identified by the DALI server. cGrx2 (PDB ID: 4tr0), a glutaredoxin derived from *Clostridium oremlabdii*, was co-crystallized with its substrate glutathione disulphide (GDS). It was chosen to compare with our Aip5-C due to the presence of substrate at its active binding site, so as to possibly shed some light on Aip5-C's binding mechanism.

Sequence alignment of Aip5-C with cGrx2, which possesses thioredoxin biochemical activity showed that the consensus CxxC motif is absent in Aip5, instead, IPRT motif is present (**Figure 18**). Isoleucine and threonine are known not to have any redox activity.

Figure 18. Sequence alignment of Aip5-C with structural homologue cGrx2. CxxC motif that is important for redox activity is absent in Aip5-C. Aip5-C also possesses a much longer loop region (SLAGGGFH) between $\beta 1$ and $\alpha 1$ compared to cGrx2, highlighted in red rectangular box.

Superposition of a protomer of Aip5-C with a protomer of cGrx2 indicated an RMSD volume of 3 Å. The CxxC motif in cGrx2 that is important for redox activity lies between the last residue of the loop connecting $\beta 1$ and $\alpha 1$, and the third residue on $\alpha 1$ (**Figure 19**). Interestingly, besides the lack of CxxC motif, Aip5-C has a much longer loop region (SLAGGGFH) between $\beta 1$ and $\alpha 1$ compared to cGrx2 (**Figure 18 and 19**). In the structure of cGrx2, GDS was observed to nestle above the CxxC motif, where the two cysteines of the motif are in oxidized form (**Figure 19**). Superposition of Aip5-C with cGRX2 revealed that the loop region in Aip5-C (residues 1148 to 1156), which is predominantly hydrophobic, occupies exactly the same space as GDS, perhaps impeding GDS binding. (**Figure 19**).

Figure 19. Superposition of Aip5-C with homologue cGrx2. Superposition of Aip5-C with cGrx2 showed that Aip5-C consists of an IxxT motif instead of CxxC motif present in cGrx2, and revealed a longer loop region between $\beta 1$ and $\alpha 1$ in Aip5-C, which impedes substrate binding in cGrx2.

3.7 Aip5 nucleation activity depends on a loop region that binds to G-actin

The loop (SLAGGGFH) between $\beta 1$ and $\alpha 1$ in Aip5-C is a short extension of amino acids that are highly hydrophobic, indicating potential ability to bind and interact with endogenous peptides in cell. Despite not being able to carry out redox activity, it is fair to ask whether Aip5-C exerts its actin assembly property through a different mechanism involving this loop.

Superposition of protomer A and B in Aip5-C showed an RMSD value of 1.48, which indicates vast similarity between the two protomers, except for a lack of clear electron density in the loop region between $\beta 1$ and $\alpha 1$ in protomer B (**Figure 20B**). The only

selenomethionine used to solve the structure is located near the loop, thus the stabilization of the loop region in Aip5-C is serendipitous for structure determination, which can be seen from its clear electron density in Protomer A (**Figure 20A**). Fo-Fc electron density map of the loop region contour level was set with σ_i at 1.0.

Figure 20. Electron density observed for the loop region in Aip5-C and superposition of protomer A and B of Aip5-C. A) Clear electron density map is observed wrapping around the loop region in protomer A of Aip5-C, reinforcing the stability of this loop. Fo-Fc electron density map was calculated with σ_i at 1.0. M1157 is located at the end of the loop, which was important for structure determination during SeMet-MAD phasing. The loop also consists of L1150, which is a crucial residue for Aip5's activity. B) Comparison between Aip5-C protomer A and B. Both protomers are highly similar except for the lack of electron density at the loop region in protomer B of Aip5-C.

Mutations	Sites of mutations
Aip5-C-LD	DSTTQTTEQSKKNNDKPQDVITTSEIRKLNEKEPVYIY TS LAGGGF HM IPRTNRLSTILTANR IPFTYRDLGTDDEARKVWKTFSGRSLPGVVVRGHNDLIGNWEEIEEANEDYKLRELIYDTI
Aip5-C-L1150A	DSTTQTTEQSKKNNDKPQDVITTSEIRKLNEKEPVYIY TS LAGGGF HM IPRTNRLSTILTANR IPFTYRDLGTDDEARKVWKTFSGRSLPGVVVRGHNDLIGNWEEIEEANEDYKLRELIYDTI
Aip5-C-N1162A	DSTTQTTEQSKKNNDKPQDVITTSEIRKLNEKEPVYIY TS LAGGGF HM IPRT N R L STILTANR PFTYRDLGTDDEARKVWKTFSGRSLPGVVVRGHNDLIGNWEEIEEANEDYKLRELIYDTI
Aip5-C-S1165C	DSTTQTTEQSKKNNDKPQDVITTSEIRKLNEKEPVYIY TS LAGGGF HM IPRTNRL S TILTANR PFTYRDLGTDDEARKVWKTFSGRSLPGVVVRGHNDLIGNWEEIEEANEDYKLRELIYDTI
Aip5-C-3m	DSTTQTTEQSKKNNDKPQDVITTSEIRKLNEKEPVYIY TS LAGGGF HM IPRT NRLS TILTANR IPFTYRDLGTDDEARKVWKTFSGRSLPGVVVRGHNDLIGNWEEIEEANEDYKLRELIYDTI

Table 2. Aip5-C amino acid sequences and various mutation sites. The deletion of loop and sites of residue mutations on Aip5-C were highlighted in red in the sequences.

The absence of the stable loop (SLAGGGFH) in Aip5-C from the homolog cGrx2 motivated us to test whether this loop contributes to its G-actin recognition and actin nucleation. We generated a protein mutant of Aip5-C, Aip5-C-LD, by removing the loop (**Table 2**). Results showed that the loop deletion completely diminished the nucleation activity of Aip5-C (**Figure 22A**). In the loop region, L1150, is highly conserved among multiple pathogenic filamentous fungi homologs (**Figure 21**). Furthermore, a closer look at L1150 revealed that it established hydrophobic contact with L1150 in the second protomer of Aip5, enhancing the stability of a dimer formation in Aip5-C (**Figure 23**). Thus we generated Aip5-C-L1150A mutant protein. To investigate whether the amino acids adjacent to the loop region also contribute to Aip5 activity, we extended our comparison with two more single mutants (Aip5-C-N1162A and Aip5-C-S1165C) and one additional triple mutant (Aip5-C-3m, L1150A, N1162A, S1165C) (**Table 2**). The activities of all Aip5-C variants were compared by calculating the initial polymerization rates using the slope of the secant lines of the first 60 s to 180 s of normalized data. Among all three single mutants, Aip5-C-L1150A showed the most drastic reduction in polymerization rate (**Figure 22B**), emphasizing the significance of L1150 in Aip5-C's activity through the extended loop. Noticeably, the additional S1165C and N1162A mutations to L1150A (Aip5-C-3m) resulted in a complete abolishment of the nucleation activity of Aip5-C (**Figure 22B and 22C**).

Figure 21. Sequence alignment of Aip5-C among pathogenic filamentous fungi, illustrated by Jalview using the Clustal Omega method. L1150 in the loop of Aip5-C is highly conserved among pathogenic filamentous fungi. Mutations in the Aip5-C-3m are indicated with red star. **Figure adapted from Prof. Miao Yansong's lab, courtesy of Ms Xie Ying.**

Figure 22. Actin nucleating activity of Aip5-C variants. A) Pyrene actin polymerization by an increasing concentration of Aip5-C-LD. B) Fold increase of rate of actin assembly with the indicated Aip5-C variants after being normalized to actin-alone reaction. The plot mean values from at least two experiments are listed. C) Pyrene actin polymerization reaction with 5 μ M Aip5-C or Aip5-C-3m. **Figure adapted from Prof. Miao Yansong's lab, courtesy of Ms Xie Ying.**

Figure 23. Hydrophobic interaction by Leu1150 from both protomers in Aip5-C. The hydrophobic establishment enhances dimer formation in native Aip5-C. However, such establishment disappeared in Aip5-C-3m mutant structure due to a major conformational change upon mutations.

3.7.1 Binding affinity between G-actin and Aip5-C loop via MST assay

To further investigate the role of the loop region in the activity of Aip5-C, we synthesized a peptide consisting of the loop residues TSLAGGGFHM, and a mutant peptide TS Δ AGGGFHM by mutating L to A. We tested the direct binding between G-actin and Aip5-C loop using a label-free microscale thermophoresis (MST) assay. G-actin was titrated with increasing concentrations of the loop peptide (TSLAGGGFHM). Results showed direct binding between Aip5-C loop peptide and G-actin at a $K_d = 98.91 \pm 76.26 \mu\text{M}$, which is around two-fold of the L1150A mutant peptide ($K_d = 199.27 \pm 49.96 \mu\text{M}$) (**Figure 24**), indicating the importance of L1150 for G-actin binding.

Figure 24. Thermophoresis binding curves of 2 μ M G-actin titrated with peptides of loop and loop-L1150A, respectively. Three biological replicates were shown with error bar of $\pm$ S.D. Direct binding was detected between Aip5 loop peptide and G-actin at a $K_d = 98.91 \pm 76.26 \mu\text{M}$, which is around two-fold of the L1150A mutant peptide ($K_d = 199.27 \pm 49.96 \mu\text{M}$).

3.8 Structure of the mutant Aip5-C with triple mutation L1150A, N1162A, S1165C (Aip5-C-3m)

Given that the Aip5-C-3m completely lost its activity, it would be interest to investigate its structure and function correlation. Through extensive efforts, Aip5-C-3m (PDB ID: 6ABS) was crystallized and the structure was solved with molecular replacement method using the native Aip5-C structure as the search model.

The mutant Aip5 is still able to maintain a thioredoxin fold, similar to that of the native protein (**Figure 25A**). The mutant Aip5-C still forms a crystallographic homodimer, gel filtration profile also showed a dimeric state of Aip5-C-3m in solution (**Figure 26**). However, the dimer interface was completely changed. Two sets of secondary structural motifs are involved in the dimer formation in the mutant Aip5-C-3m including $\beta 2$ (protomer A)- $\alpha 1$ (protomer B), and $\alpha 2$ (protomer A)- $\beta 2$

(protomer B), respectively (**Figure 25B**). In addition to polar interaction predominantly contributing to dimer formation, the residues S1165C and Tyr1177 from protomer A of mutant Aip5-C directly interact with Thr1176 and Lys1141 from protomer B, via hydrogen-bonds (**Figure 25B**). An exception of a small hydrophobic patch is also present along the dimer interface, which is made up of protomer A Phe1175, Tyr1177, and protomer B Pro1143 (**Figure 25C**). These contacts hold the dimer together and the distance between the homodimer becomes much closer in the triple mutant as compared to the native structure.

Figure 25. Crystal structure of Aip5-C triple mutant. A) Structure of Aip5-C-3m retains the thioredoxin fold and exists as a homodimer similarly to its native counterpart. B) New points of contact observed in Aip5-C-3m that hold the homodimeric protein. Interactions are largely polar. C) Hydrophobic interactions spotted between the dimer interface as well, involving protomer A Phe1175, Tyr1177, and protomer B Pro1143.

Two web servers were used to validate the total buried surface area and the contact residues involved in the dimer interface of the native and mutant Aip5-C – PDBePISA (<http://www.ebi.ac.uk/pdbe/pisa/>) and InterProSurf Protein-Protein

Interaction Server (<http://curie.utmb.edu/>). The total buried surface area for the mutant Aip5-C protomer A and B are 7.8% and 8% compared to 6.5% and 7.8% in native Aip5-C, respectively, as shown in the summary table from PDBePISA server (**Figure 27A**). Native Aip5-C dimer formation involves a strong water network and no residue between the dimer interface is in direct contact. However, mutant Aip5-C dimer becomes much closer due to the conformational change. InterProSurf server indicated a list of residues present at protein interface, among which, residues N1162A, S1165C, Phe1175 and Tyr1177 from protomer A, Lys1141, Pro1145 and Thr1176 from protomer B are the main residues that contribute to the dimer formation in the mutant Aip5-C (**Figure 27B**). This result corresponds with our structural analysis.

Figure 26. Determination of oligomeric state of Aip5-C-3m in solution using superdex75 10/300 column. The elution volumes of a series of standard proteins consisting of blue dextran, conalbumin, ovalbumin, and ribonuclease A with known molecular weights and oligomeric states were used to extrapolate the calculated molecular weight of Aip5-C-3m in solution. The molecular weights for the standard proteins are as follow: blue dextran – 2000 kDa, conalbumin – 76 kDa, ovalbumin – 45kDa, and ribonuclease A 13.7 kDa. Aip5-C-3m was determined to be a dimer in solution, which is in consistence with its crystallographic oligomeric state.

Figure 27. Servers used to analyze total buried surface area and solvent accessible area in native and mutant Aip5-C. A) Results from PDBePISA server, indicating an increase of total buried surface area for mutant Aip5-C. B) Results from InterProSurf showing residues found at protein interface.

3.8.1 Structure-function relation through comparison of Aip5-C and Aip5-C-3m

Upon mutations, Aip5-C native and mutant protomer structures remained similar with differences in the N-terminal, where the N-terminal loop in native Aip5-C folded into an alpha helix in Aip5-C-3m (**Figure 28A and 28B**). Superposition of individual protomers in Aip5-C and Aip5-C-3m revealed missing of the loops that are important for its activity in Aip5-C-3m, which indicated increased loop flexibility (**Figure 28A and 28B**). Superposition of protomer A of native and mutant Aip5-C showed an RMSD of 2.75 Å, while superposition of protomer B showed an RMSD of 1.88 Å.

Figure 28. Superposition of protomers in Aip5-C and Aip5-C-3m. A) Superposition of protomer A from Aip5-C and Aip5-C-3m. B) Superposition of protomer B from Aip5-C and Aip5-C-3m. Structures of protomers are vastly similar except for the missing of loop sites in both protomer A and B in Aip5-C-3m due to increased flexibility. Also, N-terminal region in Aip5-C-3m is folded into a helix instead of a loop present in Aip5-C.

However, superposition of the homodimer of both native and mutant Aip5-C displayed significant conformational change between the structures. By aligning with protomer A, the RMSD between two protomers are 3.5 Å (**Figure 29**). N1162A, originally located in $\alpha 1$ in native Aip5-C, now in $\alpha 2$ of protomer A in the mutant Aip5-C, forms new hydrophobic interaction with T1176 in $\beta 2$ of protomer B (**Figure 25B**). S1165C in $\alpha 2$ of protomer A further stabilizes the dimer formation by interacting with T1176 in $\beta 2$ of protomer B through hydrogen bonds, pulling the two protomers in the mutant closer to each other (**Figure 25B**). The loop between $\beta 1$ and $\alpha 1$ in the native Aip5-C consists of L1150A, which was significant for its activity, changed its orientation in the mutant structure (**Figure 29**). In the native Aip5-C, the loops in both protomers almost run parallel to each other, with the heads of $\alpha 1$ s pointing towards each other (**Figure 30**). These two loop regions are very close to

each other, where L1150 from both protomers established hydrophobic contact (**Figure 23**). However, this establishment disappeared in the mutant Aip5-C. The loop sites in both protomers are far apart, with the heads of α 1s pointing in opposite direction of each other (**Figure 30**). The conformational change in the mutant structure led to the change of orientation of the loop regions in both protomers (**Figure 29**). Note that this loop was shown to be crucial for actin nucleation activity of Aip5. The relative positioning of the two loops within the homodimer could be therefore, significantly important for its involvement in an optimum activity of Aip5.

Figure 29. Superposition of Aip5-C dimer with Aip5-C-3m dimer. Significant conformational change observed upon mutations, which led to the change of orientation of the loop regions in both protomers of Aip5-C-3m. Loop sites that consist of L1150A mutation are highlighted in black squared box.

Figure 30. Orientation of the loop sites in Aip5-C and Aip5-C-3m. The relative positioning of the two loops within the homodimer could be important for its involvement in an optimum activity of Aip5.

4. DISCUSSION

4.1 Effect of protein fusion tag on crystallization

All constructs designed for the C-terminal of Aip5 included a cleavable N-terminal His6 tag for ease of purification. Initial rounds Aip5-C purification involved nickel affinity column followed by size exclusion chromatography. His6 tag was kept intact during crystallization trials. Two concentrations were tested for crystallization – 10mg/ml and 20 mg/ml. Screening was conducted using Index, Crystal Screen, Pegion, PegRX and salt RX. Crystal hits were observed from Index condition consisting of 0.2 M potassium sodium tartrate and 20% polyethylene glycol 3350. Further optimization yielded nice and huge crystal that diffracted to 2.3Å. The crystal took more than three weeks to appear and grew to its fullest size after 4 weeks. However, there was only one crystal spotted from the entire optimization plate comprising 24 wells and 72 protein drops at three different protein to reservoir ratios.

Subsequently, attempts to repeat the crystal were failed due to unknown reasons. Upon considering possible factors that might affect crystallization, we decided to cleave off the N-terminal His6 tag. His6 tags are not usually expected to affect a protein's activity and crystallization due to their small size (Booth, Schlachter et al. 2018). However, more evidences are showing otherwise. Presence of His6 tag increases *B*-factor and possibly makes crystallization more difficult (Carson, Johnson et al. 2007). Once N-terminal His6 tag was cleaved from Aip5-C, multiple conditions from different screening kits gave rise to crystal hits. Crystals were readily reproducible and possessed high diffracting quality (**Figure 14**). Our results showed that protein fusion tag that is as small as a His tag could likely affect protein crystallization.

A

B

Figure 31. Crystals of Aip5-C with N-terminal His6 tag. A) Crystal hits were observed in Index condition 0.2 M k Na tartrate and 20% polyethylene glycol 3350. B) Further optimization yielded nice and huge crystal that diffracted to 2.3 Å. However, we were unable to repeat the crystal condition. Subsequent His6 tag cleavage from the Aip5-C produced repeatable crystals with good diffracting quality.

4.2 Aip5 has a thioredoxin fold with a different function

Structure of Aip5-C revealed a thioredoxin fold. It is a distinctive structural feature present in different classes of proteins, for instance thioredoxin, glutaredoxin, glutathione S-transferase, DsbA and glutathione peroxidase (Martin 1995, Atkinson and Babbitt 2009). All have a common property of interacting with substrates that contain cysteine, and involve redox activities. Two classes of the proteins thioredoxin and glutaredoxin utilize a highly conserved motif CxxC to reduce disulfide bond in their substrates (Berndt, Lillig et al. 2008). DsbA contains the CxxC motif as well, but acts as an oxidant instead of a reductant. While the other two classes of proteins glutathione S-transferase and glutathione peroxidase do not contain CxxC motif at their active sites, but they interact with glutathione, which is a cysteine-containing substrate (Atkinson and Babbitt 2009). Although all of the above mentioned classes of proteins are from distinct families, they have a common feature that involves the interaction with substrates comprising a thiol or disulphide group.

Here we reported a novel polarisome protein Aip5 with a thioredoxin fold structure that acts as an actin assembly regulator. Sequence alignment with structural homologue cGrx2, a typical bacterium glutaredoxin, showed that Aip5 lacks CxxC motif and voids redox activity (**Figure 18 and 19**). It also revealed the presence of a hydrophobic and stable loop (SLAGGGFH) between $\beta 1$ and $\alpha 1$ of the thioredoxin fold in Aip5 (**Figure 20**). Results from our studies indicated that Aip5 acts to nucleate actin via its C-terminal by binding to actin through the loop region (**Figure 9**). Deletion of the loop region led to complete abolishment of Aip5-C's activity (**Figure 22**). In addition, this result was further supported by MST assay, which detected

direct binding between the loop of Aip5-C and G-actin at a $K_d = 98.91 \pm 76.26 \mu\text{M}$ (**Figure 24**). Taken together, our results implied Aip5 is a novel NF and an NPF of Bni1 with its nucleating property, and possibly a new class of proteins that possess the thioredoxin fold with a different function.

4.3 Structural conformation and loop orientation in Aip5-C are crucial for its activity

Pyrene fluorescence assay showed complete abolishment of activity for mutations in Aip5-C-3m (**Figure 22**). Aip5-C with individual mutations, Aip5-C-L1150A, Aip5-C-N1162A, and Aip5-C-S1165C indicated reduced nucleating activity in Aip5, with Aip5-C-L1150A having the most potent effect (**Figure 22**). Aip5-C-3m still adopts a homodimeric thioredoxin fold structure (**Figure 25**). Superposition of protomers of Aip5-C and Aip5-C-3m showed similar structures with subtle differences in N-terminus (**Figure 28**). However the overall structures of Aip5-C and Aip5-C-3m exhibited major change in conformation when compared as homodimers that appears to be the functional unit (**Figure 29 and 30**). This conformational change led to the formation of new sets of contact points between the dimer in Aip5-C-3m and brought the two protomers closer together, resulting in a more compacted homodimer (**Figure 25 and 27**). More importantly, the conformational change destabilized the loop sites in Aip5-C, which can be observed from the lack of electron density at the loop regions in both protomers in Aip5-C-3m (**Figure 28**). It also altered the loops orientation in Aip5-C-3m to be pointing in opposite directions instead of pointing towards each protomer in Aip5-C. All of which contributed to the loss of activity in Aip5-C triple mutant.

5. FUTURE WORK

With the discovery of a novel polarisome protein and evidence establishing the functions of Aip5, we have successfully solved the structure of its functional domain in the C-terminus. We demonstrated that Aip5 functions through a loop region in the thioredoxin fold, and backed up with biochemical assays to show direct binding of Aip5 to actin through the loop. It would be of a great complement to obtain the co-crystallized structure of Aip5 to actin. Crystallizing actin poses great challenge due to its spontaneous polymerization in solution. Few methods of keeping actin in the non-polymerized state are available. For example, tagging a fluorescent probe tetramethylrhodamine-5-maleimide (TMR) covalently to Cys374 in actin to prevent polymerization, or introducing mutations on either barbed end or pointed end in actin to prevent polymerization, or adding macrolide Latrunculin A at equal molar ratio to actin to prevent polymerization (Yarmola, Somasundaram et al. 2000, Pelikan Conchaudron, Didry et al. 2006, Chen, Ni et al. 2013). Factors have to be considered to decide which method is suitable for co-crystallization. We can subsequently make attempts to co-crystallize Aip5 with actin or to soak actin with Aip5 peptides to get the complex structure. Nevertheless, our work unveiled a novel actin nucleator Aip5, alongside the rest of the protein members in the polarisome that direct polarized cell growth, and provided insights into its molecular mechanism through structural and functional studies.

REFERENCES

- (1994). "The CCP4 suite: programs for protein crystallography." *Acta Crystallogr D Biol Crystallogr* **50**(Pt 5): 760-763.
- Atkinson, H. J. and P. C. Babbitt (2009). "An Atlas of the Thioredoxin Fold Class Reveals the Complexity of Function-Enabling Adaptations." *PLoS Computational Biology* **5**(10): e1000541.
- Atkinson, H. J. and P. C. Babbitt (2009). "Glutathione Transferases Are Structural and Functional Outliers in the Thioredoxin Fold." *Biochemistry* **48**(46): 11108-11116.
- Ayscough, K. R., J. Stryker, N. Pokala, M. Sanders, P. Crews and D. G. Drubin (1997). "High Rates of Actin Filament Turnover in Budding Yeast and Roles for Actin in Establishment and Maintenance of Cell Polarity Revealed Using the Actin Inhibitor Latrunculin-A." *The Journal of Cell Biology* **137**(2): 399-416.
- Berndt, C., C. H. Lillig and A. Holmgren (2008). "Thioredoxins and glutaredoxins as facilitators of protein folding." *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783**(4): 641-650.
- Bieling, P., S. D. Hansen, O. Akin, T. D. Li, C. C. Hayden, D. A. Fletcher and R. D. Mullins (2018). "WH2 and proline - rich domains of WASP - family proteins collaborate to accelerate actin filament elongation." *The EMBO Journal* **37**(1): 102-121.
- Booth, W. T., C. R. Schlachter, S. Pote, N. Ussin, N. J. Mank, V. Klapper, L. R. Offermann, C. Tang, B. K. Hurlburt and M. Chruszcz (2018). "Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability." *ACS Omega* **3**(1): 760-768.
- Breitsprecher, D., A. K. Kiesewetter, J. Linkner, M. Vinzenz, T. E. B. Stradal, J. V. Small, U. Curth, R. B. Dickinson and J. Faix (2011). "Molecular mechanism of Ena/VASP-mediated actin-filament elongation." *The EMBO Journal* **30**(3): 456-467.
- Chant, J. and J. R. Pringle (1995). "Patterns of bud-site selection in the yeast *Saccharomyces cerevisiae*." *J Cell Biol* **129**(3): 751-765.
- Chen, X., F. Ni, X. Tian, E. Kondrashkina, Q. Wang and J. Ma (2013). "Structural Basis of Actin Filament Nucleation by Tandem W Domains." *Cell Reports* **3**(6): 1910-1920.
- Chesarone, M. A. and B. L. Goode (2009). "Actin Nucleation and Elongation Factors: Mechanisms and Interplay." *Curr Opin Cell Biol* **21**(1): 28-37.
- Chesarone, M. A. and B. L. Goode (2009). "Actin Nucleation and Elongation Factors: Mechanisms and Interplay." *Current opinion in cell biology* **21**(1): 28-37.
- Chiou, J. G., M. K. Balasubramanian and D. J. Lew (2017). "Cell Polarity in Yeast." *Annu Rev Cell Dev Biol* **33**: 77-101.
- Dominguez, R. and K. C. Holmes (2011). "Actin Structure and Function." *Annual Review of Biophysics* **40**(1): 169-186.
- Dominguez, R. and K. C. Holmes (2011). "Actin Structure and Function." *Annu Rev Biophys* **40**: 169-186.
- Drubin, D. G. and W. J. Nelson (1996). "Origins of cell polarity." *Cell* **84**(3): 335-344.

Ducka, A. M., P. Joel, G. M. Popowicz, K. M. Trybus, M. Schleicher, A. A. Noegel, R. Huber, T. A. Holak and T. Sitar (2010). "Structures of actin-bound Wiskott-Aldrich syndrome protein homology 2 (WH2) domains of Spire and the implication for filament nucleation." Proceedings of the National Academy of Sciences **107**(26): 11757-11762.

Duina, A. A., M. E. Miller and J. B. Keeney (2014). "Budding yeast for budding geneticists: a primer on the *Saccharomyces cerevisiae* model system." Genetics **197**(1): 33-48.

Ellis, R. J. (2010). "Biochemistry: Tackling unintelligent design." Nature **463**(7278): 164-165.

Firat-Karalar, E. N. and M. D. Welch (2011). "New mechanisms and functions of actin nucleation." Current opinion in cell biology **23**(1): 4-13.

Frazier, J. A. and C. M. Field (1997). "Actin cytoskeleton: Are FH proteins local organizers?" Current Biology **7**(7): R414-R417.

Freeman, S. (2014). Biological science.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin and S. G. Oliver (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-547.

Goley, E. D., A. Rammohan, E. A. Znameroski, E. N. Firat-Karalar, D. Sept and M. D. Welch (2010). "An actin-filament-binding interface on the Arp2/3 complex is critical for nucleation and branch stability." Proceedings of the National Academy of Sciences **107**(18): 8159-8164.

Graziano, B. R., A. G. DuPage, A. Michelot, D. Breitsprecher, J. B. Moseley, I. Sagot, L. Blanchoin and B. L. Goode (2011). "Mechanism and cellular function of Bud6 as an actin nucleation-promoting factor." Mol Biol Cell **22**(21): 4016-4028.

Graziano, B. R., A. G. DuPage, A. Michelot, D. Breitsprecher, J. B. Moseley, I. Sagot, L. Blanchoin and B. L. Goode (2011). "Mechanism and cellular function of Bud6 as an actin nucleation-promoting factor." Molecular Biology of the Cell **22**(21): 4016-4028.

Groban, E. S., A. Narayanan and M. P. Jacobson (2006). "Conformational Changes in Protein Loops and Helices Induced by Post-Translational Phosphorylation." PLOS Computational Biology **2**(4): e32.

Han, Y. H., C. Y. Chung, D. Wessels, S. Stephens, M. A. Titus, D. R. Soll and R. A. Firtel (2002). "Requirement of a vasodilator-stimulated phosphoprotein family member for cell adhesion, the formation of filopodia, and chemotaxis in *dictyostelium*." J Biol Chem **277**(51): 49877-49887.

Harris, E. S. and H. N. Higgs (2004). "Actin Cytoskeleton: Formins Lead the Way." Current Biology **14**(13): R520-R522.

Higuchi-Sanabria, R., W. M. Pernice, J. D. Vevea, D. M. Alessi Wolken, I. R. Boldogh and L. A. Pon (2014). "Role of asymmetric cell division in lifespan control in *Saccharomyces cerevisiae*." FEMS Yeast Res **14**(8): 1133-1146.

Hurley, J. H. (1996). "The sugar kinase/heat shock protein 70/actin superfamily: implications of conserved structure for mechanism." Annu Rev Biophys Biomol Struct **25**: 137-162.

Kabsch, W., H. G. Mannherz, D. Suck, E. F. Pai and K. C. Holmes (1990). "Atomic structure of the actin: DNase I complex." Nature **347**: 37.

Kabsch, W., H. G. Mannherz, D. Suck, E. F. Pai and K. C. Holmes (1990). "Atomic structure of the actin:DNase I complex." Nature **347**(6288): 37-44.

Langer, G., S. X. Cohen, V. S. Lamzin and A. Perrakis (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." *Nat Protoc* **3**(7): 1171-1179.

Le Clainche, C. and M. F. Carlier (2008). "Regulation of actin assembly associated with protrusion and adhesion in cell migration." *Physiol Rev* **88**(2): 489-513.

Lei, W., O. F. Omotade, K. R. Myers and J. Q. Zheng (2016). "Actin cytoskeleton in dendritic spine development and plasticity." *Current Opinion in Neurobiology* **39**: 86-92.

Levin, D. E. (2005). "Cell Wall Integrity Signaling in *Saccharomyces cerevisiae*." *Microbiology and Molecular Biology Reviews* **69**(2): 262-291.

Littlefield, R. and V. M. Fowler (2002). "A minor actin catastrophe." *Nature Cell Biology* **4**: E209.

Liu, B., L. Larsson, A. Caballero, X. Hao, D. Öling, J. Grantham and T. Nyström (2010). "The Polarisome Is Required for Segregation and Retrograde Transport of Protein Aggregates." *Cell* **140**(2): 257-267.

Liu, R., E. V. Linardopoulou, G. E. Osborn and S. M. Parkhurst (2010). "Formins in Development: Orchestrating Body Plan Origami." *Biochimica et biophysica acta* **1803**(2): 207-225.

Long, F., A. A. Vagin, P. Young and G. N. Murshudov (2008). "BALBES: a molecular-replacement pipeline." *Acta Crystallogr D Biol Crystallogr* **64**(Pt 1): 125-132.

Madden, K. and M. Snyder (1998). "Cell polarity and morphogenesis in budding yeast." *Annu Rev Microbiol* **52**: 687-744.

Martin, J. L. (1995). "Thioredoxin —a fold for all reasons." *Structure* **3**(3): 245-250.

Matsuzaki, M., O. Misumi, I. T. Shin, S. Maruyama, M. Takahara, S. Y. Miyagishima, T. Mori, K. Nishida, F. Yagisawa, K. Nishida, Y. Yoshida, Y. Nishimura, S. Nakao, T. Kobayashi, Y. Momoyama, T. Higashiyama, A. Minoda, M. Sano, H. Nomoto, K. Oishi, H. Hayashi, F. Ohta, S. Nishizaka, S. Haga, S. Miura, T. Morishita, Y. Kabeya, K. Terasawa, Y. Suzuki, Y. Ishii, S. Asakawa, H. Takano, N. Ohta, H. Kuroiwa, K. Tanaka, N. Shimizu, S. Sugano, N. Sato, H. Nozaki, N. Ogasawara, Y. Kohara and T. Kuroiwa (2004). "Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D." *Nature* **428**(6983): 653-657.

Meijer, W. G., A. C. Arnberg, H. G. Enequist, P. Terpstra, M. E. Lidstrom and L. Dijkhuizen (1991). "Identification and organization of carbon dioxide fixation genes in *Xanthobacter flavus* H4-14." *Mol Gen Genet* **225**(2): 320-330.

Miao, Y., X. Han, L. Zheng, Y. Xie, Y. Mu, J. R. Yates, 3rd and D. G. Drubin (2016). "Fimbrin phosphorylation by metaphase Cdk1 regulates actin cable dynamics in budding yeast." *Nat Commun* **7**: 11265.

Miao, Y., T. Tipakornsawapak, L. Zheng, Y. Mu and E. Lewellyn (2018). "Phospho-regulation of intrinsically disordered proteins for actin assembly and endocytosis." *Febs j.*

Miao, Y., C. C. Wong, V. Mennella, A. Michelot, D. A. Agard, L. J. Holt, J. R. Yates, 3rd and D. G. Drubin (2013). "Cell-cycle regulation of formin-mediated actin cable assembly." *Proc Natl Acad Sci U S A* **110**(47): E4446-4455.

Michelot, A., C. Guérin, S. Huang, M. Ingouff, S. Richard, N. Rodiuc, C. J. Staiger and L. Blanchoin (2005). "The Formin Homology 1 Domain Modulates the Actin Nucleation and Bundling Activity of *Arabidopsis* FORMIN1." *The Plant Cell* **17**(8): 2296-2313.

Murchie, E. H., M. Pinto and P. Horton (2009). "Agriculture and the new challenges for photosynthesis research." *New Phytol* **181**(3): 532-552.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 355-367.

Namgoong, S., M. Boczkowska, M. J. Glista, J. D. Winkelman, G. Rebowski, D. R. Kovar and R. Dominguez (2011). "Mechanism of actin filament nucleation by Vibrio VopL and implications for tandem-W domain nucleation." *Nature structural & molecular biology* **18**(9): 1060-1067.

Orlando, K. and W. Guo (2009). "Membrane organization and dynamics in cell polarity." *Cold Spring Harb Perspect Biol* **1**(5): a001321.

Otterbein, L. R., P. Graceffa and R. Dominguez (2001). "The Crystal Structure of Uncomplexed Actin in the ADP State." *Science* **293**(5530): 708-711.

Otwinowski, Z. and W. Minor (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, Academic Press. **Volume 276**: 307-326.

Ozaki-Kuroda, K., Y. Yamamoto, H. Nohara, M. Kinoshita, T. Fujiwara, K. Irie and Y. Takai (2001). "Dynamic Localization and Function of Bni1p at the Sites of Directed Growth in *Saccharomyces cerevisiae*." *Molecular and Cellular Biology* **21**(3): 827-839.

Padrick, S. B., L. K. Doolittle, C. A. Brautigam, D. S. King and M. K. Rosen (2011). "Arp2/3 complex is bound and activated by two WASP proteins." *Proceedings of the National Academy of Sciences* **108**(33): E472-E479.

Pelikan Conchaudron, A., D. Didry, K. H. Le, E. Larquet, N. Boisset, D. Pantaloni and M. F. Carlier (2006). "Analysis of tetramethylrhodamine-labeled actin polymerization and interaction with actin regulatory proteins." *J Biol Chem* **281**(33): 24036-24047.

Petersen, J., O. Nielsen, R. Egel and I. M. Hagan (1998). "FH3, A Domain Found in Formins, Targets the Fission Yeast Formin Fus1 to the Projection Tip During Conjugation." *The Journal of Cell Biology* **141**(5): 1217-1228.

Pollard, T. D. and G. G. Borisy (2003). "Cellular motility driven by assembly and disassembly of actin filaments." *Cell* **112**(4): 453-465.

Porta, J. C. and G. E. O. Borgstahl (2012). "Structural basis for profilin-mediated actin nucleotide exchange." *J Mol Biol* **418**(0): 103-116.

Robert, X. and P. Gouet (2014). "Deciphering key features in protein structures with the new ENDscript server." *Nucleic Acids Res* **42**(Web Server issue): W320-324.

Schafer, D. A., P. B. Jennings and J. A. Cooper (1996). "Dynamics of capping protein and actin assembly in vitro: uncapping barbed ends by polyphosphoinositides." *J Cell Biol* **135**(1): 169-179.

Shekhar, S. and M.-F. Carlier (2016). "Single-filament kinetic studies provide novel insights into regulation of actin-based motility." *Molecular Biology of the Cell* **27**(1): 1-6.

Sheu, Y.-J., B. Santos, N. Fortin, C. Costigan and M. Snyder (1998). "Spa2p Interacts with Cell Polarity Proteins and Signaling Components Involved in Yeast Cell Morphogenesis." *Molecular and Cellular Biology* **18**(7): 4053-4069.

Shih, J. L., S. L. Reck-Peterson, R. Newitt, M. S. Mooseker, R. Aebersold and I. Herskowitz (2005). "Cell Polarity Protein Spa2P Associates With Proteins

Involved In Actin Function In *Saccharomyces Cerevisiae*." Molecular Biology of the Cell **16**(10): 4595-4608.

Tu, D., B. R. Graziano, E. Park, W. Zheng, Y. Li, B. L. Goode and M. J. Eck (2012). "Structure of the formin-interaction domain of the actin nucleation-promoting factor Bud6." Proc Natl Acad Sci U S A **109**(50): E3424-3433.

Vagin, A. and A. Teplyakov (2010). "Molecular replacement with MOLREP." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 22-25.

Visa, N. and P. Percipalle (2010). "Nuclear Functions of Actin." Cold Spring Harbor Perspectives in Biology **2**(4): a000620.

Wang, H., Z. X. Huang, J. Y. Au Yong, H. Zou, G. Zeng, J. Gao, Y. Wang, A. H. Wong and Y. Wang (2016). "CDK phosphorylates the polarisome scaffold Spa2 to maintain its localization at the site of cell growth." Mol Microbiol **101**(2): 250-264.

Wegner, A. and G. Isenberg (1983). "12-fold difference between the critical monomer concentrations of the two ends of actin filaments in physiological salt conditions." Proc Natl Acad Sci U S A **80**(16): 4922-4925.

Welch, M. D., A. H. DePace, S. Verma, A. Iwamatsu and T. J. Mitchison (1997). "The Human Arp2/3 Complex Is Composed of Evolutionarily Conserved Subunits and Is Localized to Cellular Regions of Dynamic Actin Filament Assembly." The Journal of Cell Biology **138**(2): 375-384.

Wen, K.-K. and P. A. Rubenstein (2009). "Differential Regulation of Actin Polymerization and Structure by Yeast Formin Isoforms." The Journal of Biological Chemistry **284**(25): 16776-16783.

Yarar, D., W. To, A. Abo and M. D. Welch (1999). "The Wiskott–Aldrich syndrome protein directs actin-based motility by stimulating actin nucleation with the Arp2/3 complex." Current Biology **9**(10): 555-S551.

Yarmola, E. G., D. A. Dranishnikov and M. R. Bubb (2008). "Effect of Profilin on Actin Critical Concentration: A Theoretical Analysis." Biophysical Journal **95**(12): 5544-5573.

Yarmola, E. G., T. Somasundaram, T. A. Boring, I. Spector and M. R. Bubb (2000). "Actin-latrunculin A structure and function. Differential modulation of actin-binding protein function by latrunculin A." J Biol Chem **275**(36): 28120-28127.

Ying, X., H. Xiao and M. Yansong "An Effective Recombinant Protein Expression and Purification System in *Saccharomyces cerevisiae*." Current Protocols in Molecular Biology **0**(0): e62.

**Chapter 2: Structural and functional studies of
CbbX, a red-type Rubisco activase in
*Cyanidioschyzon merolae***

1. INTRODUCTION

1.1 Ever-increasing demand for sustainable resources

With fast increasing global population, sustainable resources are needed more than ever. It is estimated that one in seven people are severely malnourished due to the lack of access to food (Foley, Ramankutty et al. 2011). This places an enormous demand on agricultures. However, with limited amount of cultivable lands, increasing crop production has become a critical issue. Among which, improving photosynthesis of the crops could be a potential solution (Zhu, Long et al. 2010).

For more than 50 years, the focus has been on genetic manipulation, usage of pesticides, fertilizers and water to increase crop yield. However, improved agricultural practices have reached a theoretical limit currently (Huner, Dahal et al. 2014). Further rise in crop productivity requires other efforts. As an alternative strategy, enhancing photosynthetic efficiency would increase the conversion of solar energy by plants into biomass, thereby leading to higher crop yield (Murchie, Pinto et al. 2009).

1.2 Photosynthesis

Sunlight is the primary source of energy for nearly all lives on earth. Organisms such as plants, photosynthetic bacteria, and algae utilize a process known as **photosynthesis** to convert this light energy into chemical energy, which propels the synthesis of organic compounds from carbon dioxide (CO₂) and water (H₂O). This is unarguably one of the most important biological processes present to date (Bryant and Frigaard 2006). The end products of this process are oxygen and carbohydrates.

Photosynthesis generally occurs in chloroplasts (**Figure 32**). A chloroplast contains many membrane-bound flattened disks known as the thylakoid membranes, which are stacked together to form grana (Croce and van Amerongen 2014). The inner aqueous phase of a single thylakoid membrane is the lumen, while the inner aqueous phase of a chloroplast is the stroma. Photosynthesis can be divided into two stages – light dependent and light independent reactions. While most of the components required for light-dependent reactions are found in the thylakoid membranes, enzymes needed for carbon assimilation are located within the stroma (Kaiser, Schroppel-Meier et al. 1986).

Nature Reviews | Molecular Cell Biology

Figure 32: Structure of a chloroplast. Found in the cells of algae and green plants, chloroplasts are organelles bound by double layer membrane that conduct photosynthesis. They contain stacks of thylakoid membranes, which are stacked to form grana. Grana are interconnected via another membrane known as the stroma lamellae. The fluid within the chloroplasts surrounding the thylakoid membranes is the stroma, while the space inside thylakoid membranes is the lumen. Major components of photosynthesis including PSI, PSII, cytochrome-b6f complex, and ATP synthase are embedded in the thylakoid membranes. (Figure adapted from Nelson and Ben-Shem 2004)

1.2.1 Light dependent reactions

Light dependent reactions take place in the thylakoid membranes during photosynthesis. There are four major players involved – photosystems I (PSI), photosystems II (PSII), cytochrome b6f, and ATP synthase. PSI and PSII are large protein assemblies comprising hundreds of pigments (Croce and van Amerongen 2014).

As depicted in **Figure 33**, the whole process starts off when a pigment molecule in the reaction center of PSII absorbs solar energy in the form of photon. The photon then excites an electron within the pigment. The excited electron is transferred from PSII to plastoquinone (PQ) pool, which accepts a proton from the stroma. The reaction center chlorophyll of PSII is now oxidized and able to receive electrons from water, splitting water into O_2 and protons in the lumen. Meanwhile, the reduced PQ moves through the membrane from PSII to cytochrome-b6f complex, where PQ gets oxidized and releases its proton in the lumen. This results in transferring a proton from stroma to lumen. The electron is further transferred to PSI via an electron transport chain (ETC) with the help of cytochrome-b6f complex and plastocyanin (PC). The reduced chlorophyll at the reaction center of PSII gets excited and transfers the electron to ferredoxin (Fd), which resides in the stroma. The excited Fd then transfers the electron to and reduces $NADP^+$ via the ferredoxin NADP reductase (FNR) (Michelet, Zaffagnini et al. 2013).

The series of electron transfer creates an electrochemical proton gradient across the thylakoid membrane (between the stroma and the lumen), which drives ATP synthesis

by another important enzyme, the ATP synthase (Goss and Hanke 2014). The electron is eventually transferred to NADP⁺ to form NADPH in the stroma.

Figure 33: Photosynthesis starts with light dependent reactions. Light reactions occur in the thylakoid membrane. When a light photon excites a pigment molecule in PSII, the reaction center chlorophyll molecule loses an electron. The electron is transferred from PSII to PSI with the help from PQ, cytochrome-b6f complex, and PC. Further transfer of electron to Fd reduces NADP⁺ via FNR. The series of electron transfer generates a proton gradient across the thylakoid membrane that energizes ATP synthase to produce ATP. (Figure adapted from Michelet, Zaffagnini et al. 2013)

1.2.2 Light independent reactions

Light independent reactions, on the other hand, take place in the stroma that is located outside of the thylakoid membranes, involving a series of redox reactions to assimilate atmospheric CO₂ into organic compounds. The light independent reactions are mainly driven by the Calvin cycle (**Figure 34**), and fed by the products from light dependent reactions (ATP and NADPH) (Raines 2003). There are three stages in the Calvin cycle: carbon fixation, reduction, and regeneration (Berg, Tymoczko et al. 2002).

The first stage involves fixation of CO₂ to form 3-phospho-D-glycerate (3PGA). CO₂ from the atmosphere enters the plant via stomata, where it reaches the mesophyll cells. It then diffuses into the stroma of chloroplasts. An important enzyme known as the D-ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) covalently adds one molecule of CO₂ to one molecule of ribulose 1,5-bisphosphate (RuBP), producing two molecules of 3PGA. At the end of stage one, three molecules of CO₂ will be fixed into six molecules of 3PGA.

The second stage involves the reduction of 3PGA to form glyceraldehyde 3-phosphate (G3P). This reaction requires products from the light dependent reactions (ATP and NADPH). Six molecules of ATP and NADPH each are utilized to reduce six molecules of 3PGA into six molecules of G3P.

The last stage involves the regeneration of RuBP in order to fix more CO₂. Out of the six molecules of G3P, five of them are needed to regenerate three molecules of RuBP, in the expense of three molecules of ATP. The remaining G3P is shuttled out of the cycle for glucose and other organic compounds production (Michelet, Zaffagnini et al. 2013).

The reaction occurs in a cycle.

Figure 34: Overview of Calvin-Benson cycle. The Benson cycle is a series of redox reactions to assimilate atmospheric CO_2 into organic compounds. Rubisco is the crucial enzyme that initiates the carbon assimilation. (Figure adapted and modified from Freeman 2014)

1.3 Rubisco

Rubisco, the most abundant protein on earth, is a vital enzyme responsible for fixing atmospheric CO_2 into organic carbons via Calvin cycle during photosynthesis. It catalyzes the covalent addition of CO_2 to the five-carbon sugar RuBP and leads to a net gain of carbon for biomass production (Stotz, Mueller-Cajar et al. 2011). However, each molecule of Rubisco is only capable of fixating three to ten carbon dioxide molecules per second, making it a major rate-limiting player in the Calvin cycle (Ellis 2010). This contributes to the large amount of Rubisco found in plants in compensation for this drawback.

Besides carboxylation of RuBP, Rubisco is also responsible for the oxygenation of RuBP. When Rubisco adds oxygen (O_2) onto RuBP, it gives off one molecule of PGA and one molecule of 2-phospho-glycolate (2PG). A series of reactions then use ATP

as energy and convert 2PG into GPA, while release CO₂ at the same time (Tcherkez 2015). This process of utilizing O₂ and releasing CO₂ is known as the photorespiration, which greatly mimics the aerobic respiration. It is a wasteful process, as it requires both carbon and energy, thus diminishing photosynthetic efficiency.

Given the sluggish enzymatic activity and slow turnover rate of Rubisco, improving the efficiency of Rubisco has become an important goal to increase plant productivity (Lin, Occhialini et al. 2014).

1.3.1 Structure of rubisco

There are four different forms (form I to IV) of Rubisco found in nature, categorized according to their molecular structure, sequence identity, sequence conservation of important residues for carbon fixation, and their ability to fix carbon (Park, Hwang et al. 2009). Each form has similar yet unique features and functions. Generally they consist of dimers of large subunits, small subunits, or a mixture of both (Tabita, Satagopan et al. 2008). Large subunit has a molecular weight of 50,000 Da, while small subunit weighs approximately 15,000 Da. One of the common features fundamental to all four forms is the large subunit dimer (Tabita, Satagopan et al. 2008).

As the most abundant type of Rubisco, form I is the predominant form present in terrestrial and marine plants, eukaryotic algae, cyanobacteria, and most phototrophic and chemolithoautotrophic proteobacteria (Tabita, Hanson et al. 2007). Form I Rubisco comprises eight large and eight small catalytic subunits (**Figure 35**). Large

subunits form the catalytic core with four repeating dimers, while eight small subunits cover the catalytic core with four on top and four at the bottom. Together they form a hexadecameric complex. Form I Rubisco can be further divided into four subgroups form IA-D (Park, Hwang et al. 2009). Types IA and B are the green-type Rubisco: IA – purple bacterial green-type Rubisco, IB – cyanobacterial/plant green-type Rubisco. Types IC and D are the red-type Rubisco: IC – purple bacterial red-type Rubisco, ID – eukaryotic non-green algal red-type Rubisco. It has been reported that there could be a new subgroup E under the red-type Rubisco based on the multiple sequence alignments and phylogenic studies done by Park, Hwang et al. 2009.

Form II Rubisco generally consists of large subunit dimers (**Figure 35**). Different source of organisms have varying numbers of oligomers, which range from two to eight oligomers of the dimer (Tabita, Satagopan et al. 2008). Form II Rubisco is found in proteobacteria and dinoflagellates that fix carbon (Tabita, Hanson et al. 2007). However, they generally have a poor CO₂/O₂ specificity.

Figure 35: Structures of form I & form II Rubisco. Form I Rubisco is the predominant type of Rubisco. It is also the only form of Rubisco that contains small subunit. It consists of four dimers of large subunits forming the catalytic core, which is covered by four small subunits on top and four at the bottom of the core. Form II Rubisco forms dimers of large subunits ranging from two to eight oligomers depending on the source of origin. (Figure adapted and modified from Tabita, Satagopan et al. 2008).

Form III Rubisco, found in some thermophilic archaea such as *Methanocaldococcus jannaschii* and *Archaeoglobus fulgidus* that are obligate anaerobes and survive at thermophilic temperatures, catalyzes carbon fixation independent of the Calvin cycle (**Figure 36**). These organisms have shown to fix carbon via a modified acetyl-coenzyme A pathway instead of the typical Calvin cycle pathway (Finn and Tabita 2003). Form III Rubisco is made of dimers of the large subunit, arranged in either a single dimer of two large subunits, or five dimers with a pentagonal symmetry (Kitano, Maeda et al. 2001).

Form IV Rubisco is in the form of a large subunit dimer, also termed as Rubisco-like Protein (RLP) (**Figure 36**). Unlike the typical forms of Rubisco, RLP is unable to fix carbon due to substitutions of many key residues at the catalytic site (Tabita, Satagopan et al. 2008). However, they might possess other catalytic functions. RLPs are present in organisms that participate in sulphur metabolism, for instance, *Chlorobaculum tepidum*, *Bacillus subtilis*, *Geobacillus kaustophilus*, and *Microcystis aeruginosa* (Singh and Tabita 2010).

Figure 36: Structures of form III & form IV Rubisco. Form III Rubisco comprises five dimers of large subunits arranged in pentagonal symmetry. It carries out carbon fixation in a Calvin cycle-independent manner. Form IV Rubisco is known as the RLP, which usually stays as a single large subunit dimer, and is incapable of carbon fixation (Figure adapted and modified from Tabita, Satagopan et al. 2008).

1.4 Rubisco activase

Rubisco activase (Rca) was first discovered from an Rca mutant found in *Arabidopsis thaliana*, via a high CO₂/low CO₂ screening that detects mutants with deficits in photosynthesis pathway (Somerville, Portis et al. 1982). The mutation specifically affects the activation of Rubisco and not other enzymes of photosynthesis.

Prior to fixing carbon, Rubisco has to bind a non-substrate activator CO₂. The active site of Rubisco is found in the large subunit and is carbamylated by CO₂, which reacts with the ε-amino group of lysine 201 at the catalytic site, to form carbamate (Saito, Ashida et al. 2009). However, Rubisco in this form is yet to be active, as it requires the addition of a cofactor Mg²⁺ ion. The binding of Rubisco to this divalent metal ion acts to stabilize the negatively charged carbamate, and thereafter, activates the Rubisco (Jensen 2000). Once activated, Rubisco binds its substrate RuBP, and initiates the fixation process.

The rate of reaction is slow and could only last a couple of min before it deteriorates quickly. This is due to the tight binding of RuBP to Rubisco, which has spontaneously lost the activator CO₂. Binding of RuBP prematurely causes a huge conformational change in Rubisco and renders it inactive (Jensen 2000). To compensate this shortfall, Rca helps to speed up the process by liberating the bound RuBP from Rubisco, thus allowing the activator CO₂ and Mg²⁺ to bind to Rubisco once more. Regulation of Rubisco by Rca ensures photosynthetic efficiency reaching its optimal.

All plants and green algae contain Rca, which is an ATPase associated with diverse cellular activities (AAA^+) protein that facilitates the dissociation of inhibitor RuBP from Rubisco in an ATP-dependent manner (Stotz, Mueller-Cajar et al. 2011). However, no Rca or Rca homologues were found in organisms with red-type Rubisco. In contrast, a gene named *cbbX* was discovered to be repeatedly associated with gene encoding the Rubisco subunits (Mueller-Cajar, Stotz et al. 2011).

1.4.1 CbbX

The *cbbX* gene lies downstream of Rubisco small subunit (*cbbS*) in proteobacteria, and encodes a protein of approximately 35kDa (Meijer, Arnberg et al. 1991). It has been reported that CbbX is a transcriptional regulator of Rubisco operon and a molecular chaperon of Rubisco subunit assembly (Fujita, Tanaka et al. 2008). CbbX protein in the form IC is present in photosynthetic organisms with red-type of Rubisco in proteobacteria, while form ID is present in eukaryotic red algae (Fujita, Tanaka et al. 2008).

Similar to Rca found in the green algae and higher plant, CbbX from the red-type algae is known to be an AAA^+ protein that plays a critical role in carbon assimilation (Mueller-Cajar, Stotz et al. 2011). It releases RuBP from Rubisco and activates Rubisco. However, the mechanism by which CbbX activates Rubisco is rather different from Rca (Robinson and Portis 1989). Contrary to Rca, CbbX shows no ATPase activity in absence of RuBP, and its activity increases when RuBP is added. The activity increases even further with the addition of inhibited Rubisco (RuBP-Rubisco complex) (Mueller-Cajar, Stotz et al. 2011). Meanwhile, activated Rubisco

also increases the catalytic activity of CbbX similar to that of the inhibited Rubisco (Mueller-Cajar, Stotz et al. 2011).

Several studies have shown that defects of CbbX protein found in *Ralstonia eutropha* and *Rhodobacter shaeroides* severely impaired their photosynthetic activity (Gibson and Tabita 1997, Bowien and Kusian 2002). This re-emphasizes the significance of CbbX in Rubisco activation and carbon fixation.

1.4.2 CbbX in *Cyanidioschyzon merolae*

Cyanidioschyzon merolae is a photoautotrophic, unicellular red algae found in sulphate-rich hot springs. It is capable of thriving in high temperature and highly acidic environments (Kobayashi, Harada et al. 2014). The cell compartment organization of *C. merolae* is very simple: only one nucleus, one plastid with plastid nucleoid located in the center, one mitochondrion, and a simple set of membrane structures (a single of each of Golgi body, endoplasmic reticulum, microbody, and a few lysosome like structures) (Matsuzaki, Misumi et al. 2004). It lacks a vacuole and a cell wall. The genome of *C. merolae* was also the first full algal genome (all three organelles) to be sequenced in 2004 (Matsuzaki, Misumi et al. 2004). Due to its simple architecture and easy synchronization of the division cycles, *C. merolae* has emerged as an excellent model organism to study organelle division, biochemistry, and structural biology (Kanesaki, Imamura et al. 2012).

In the plastid genome, there lies an operon comprising *rbcL* (the large chain gene of Rubisco), *rbcS* (the small chain gene of Rubisco), and *cbbX* organized in a tandem array (Ohta, Matsuzaki et al. 2003). The *cbbX* gene contains a nucleotide-binding

motif that regulates the transcription of Rubisco operon in red algae (Fujita, Ehira et al. 2008) (**Figure 37**). It is believed that the plastid encoded cbbX gene came from a β -proteobacterium via horizontal gene transfer. In the mean time, a cbbX homologue was found in the nuclear genome and classified into another cluster together with the nucleomorph-encoded cbbX gene of *Guillardia theta* (Fujita, Ehira et al. 2008). The encoded nuclear CbbX protein contains a putative peptide of 94 amino acids that targets to the plastid.

Figure 37: Sequence alignment of CbbX(pt) and CbbX(nuc) with another red-type rubisco activase from *R. sphaeroides*. Secondary structure elements of CbbX from *R. sphaeroides* (PDB ID code 3SYL) are shown above the sequences. Walker A motif is underlined with blue rectangular box. Walker B motif is underlined with pink rectangular box. Pore loop is underlined with orange rectangular box. Arginine finger is indicated with a red star. The figure was prepared using Esript.

The proteins that the two cbbX genes encoded by the nuclear and plastid genomes of *C. merolae* are nonfunctional in isolation, but together form a thermostable

heterooligomeric Rca that can use both α -proteobacterial and red algal-inhibited rubisco complexes as a substrate (Loganathan, Tsai et al. 2016). The sequence identity of plastid and nuclear CbbX proteins is approximately 50.9%, excluding the putative peptide present in nuclear CbbX. Both proteins have an Walker ATP-binding motifs (A-GXXGTGKT, B-VLFIDE) (**Figure 37**), a pore loop region and arginine finger (Fujita, Ehira et al. 2008). A red-type rubisco activase from *Rhodobacter sphaeroides* was used as an reference, since it possesses high sequence similarity to CbbX(nuc) and CbbX(pt). A sequence alignment of the three proteins indicated that residues that are functionally important in the bacterial activase from *R. sphaeroides* are in general highly conserved in CbbX(pt) and CbbX(nuc) too, as multiple functional motifs are found in all three proteins such as the Walker A and B motifs (**Figure 37**).

The data from our collaborator Prof. Mueller-Cajar's group has suggested that CbbX from *C. merolae* is the most thermostable Rubisco activase discovered thus far (Loganathan, Tsai et al. 2016). The ATPase activity of CbbX increases steadily with an increase in temperature of up to 55°C. It has also been shown that only the hetero-oligomer state of both nuclear and plastid CbbX possesses the maximum ATPase activity, while individual isoforms of CbbX from *C. merolae* failed to show any ATPase activity (Loganathan, Tsai et al. 2016). Contrary to the structure of rubisco activase from *Rhodobacter sphaeroides* solved by Mueller-Cajar et al. (Mueller-Cajar, Stotz et al. 2011), which requires only a single protein to perform its function, the results emphasized the fact that CbbX from *C. merolae* requires both the nuclear and plastid isoforms to become an active Rubisco activase.

As the most important rate limiting factor in photosynthesis, Rubisco is at the heart to improve photosynthesis efficiency, this makes investigation on Rubisco activase very appealing. Understanding its catalytic mechanism via structural analysis would certainly pave a way to improve photosynthesis in photoautotrophic organisms. By obtaining the crystal structure of CbbX from *C. merolae*, we would be able to know the arrangement of its subunits in the hetero oligomer state. This offers insights into how they cooperate as a whole to carry out its catalytic function as a Rubisco activase. Since *C. merolae* thrives under high temperature, it also provides us a chance to look into the molecular determinants that contribute to its thermostability, which could potentially be applied to other organisms via genetic engineering to further improve photosynthetic efficiency under heat stress conditions.

1.5 Aims and objectives

This project is in collaboration with Prof. Mueller-Cajar's group. Given the scientific importance of Rubisco activase, we are interested to elucidate the structure of CbbX from *C. merolae* (both isoforms). In this chapter, the plastid isoform of CbbX will be named as CbbX(pt), while the nuclear isoform will be named as CbbX(nuc). Aims and objectives of this project include:

- Purify and crystallize both nuclear and plastid isoforms of CbbX from *C. merolae* to elucidate their crystal structures through x-ray crystallography
- Conduct structural comparison between CbbX from *C. merolae* and its homologs from other organisms
- Purify and crystallize CbbX complex from *C. merolae* to elucidate their crystal structures through x-ray crystallography
- Investigate on the binding of CbbX (both isoforms) to its co-factor and substrates, for instance, ATP, and RuBP

2. MATERIALS & METHODS

2.1 Molecular cloning

Cloning of CbbX(pt) was done by Protein Production Platform (PPP, NTU). Full length CbbX(pt) was cloned into pNIC-CH2 vector, which consists of a C-terminal His₆ tag. The vector antibiotic resistance is kanamycin. The recombinant plasmid was transformed into *Escherichia coli* BL21 DE3 cells that contains chloramphenicol resistance.

Cloning of CbbX(nuc) (nuclear isoform of CbbX from *C. merolae*) was done by our collaborator Mueller-Cajar's group. Full length CbbX(nuc) was cloned into pHUE vector, which contains an N-terminal His6-ubiquitin tag. To assist the folding and expression of CbbX(nuc), another vector PBAD carrying GroEL/ES was co-transformed with CbbX(nuc) in pHUE vector into *E. coli* C41 DE3 cells. The antibiotic resistance for pHUE vector is ampicillin, and the antibiotic resistance for pBAD vector is chloramphenicol. Unlike BL21 DE3 cells, C41 DE3 cells have no inherit antibiotic resistance.

2.2 Mutagenesis on CbbX complex

Site directed mutagenesis was utilized to remove His6 tag from CbbX(pt) in preparation of CbbX(pt) and CbbX(nuc) co-purification. It was also employed to introduce single amino acid mutations onto CbbX(pt) and CbbX(nuc) in hope of improving diffracting quality of CbbX complex crystals. Reaction mixture in 25 μ l contained 5 μ l 5 $\times$ Q5 reaction buffer, 0.5 μ l 10mM dNTP, 1.25 μ l 10 μ M each of forward and reverse primers, 50 ng of either CbbX(pt) or CbbX(nuc) DNA template, 5 μ l 5x Q5 High GC Enhancer (optional), and 0.5 U (0.25 μ l) Q5 High-Fidelity DNA

Polymerase. Listed below are the forward and reverse primers used for each mutagenesis of CbbX(pt) and CbbX(nuc).

CbbX(pt) His6 tag removal (CbbX(pt) Δ His6)

F 5' – AGTCGTTTGTTC AAGCGTGAGGATCCGAATTCGAG – 3'

R 5' – CTCGAATTCGGATCCTCACGCTTGAAACAAACGACT –3'

CbbX(pt) Lys270Ala, Lys271Ala

F 5' – CAGGCGCGGCGCTAACCAAATTTGATTTGG – 3'

R 5' – TTAGCGCCGCGCCTGGTGTAGCAAAGATA –3'

CbbX(pt) Lys96Ala, Lys97Ala

F 5' – CTATATTGCGGCGGGTCATTTAATTACGGTG – 3'

R 5' – ATGACCCGCCGCAATATAGCCTAAACGATA –3'

CbbX(nuc) Glu162Ala, Glu163Ala, Gln164Ala

F 5' – TCATGGCAGCGGCACGTGAGGATCTCGTGGTCATTT – 3'

R 5' – TCACGTGCCGCTGCCATGACGTTTCAGCAAAATCTCG –3'

CbbX(nuc) Gln16Ala

F 5' – TTCTTGCAAAAGAGGGCGTGCTCGATAT – 3'

R 5' – ACGCCCTCTTTTGCAAGAACTCGTC –3'

CbbX(nuc) Gln16Ala, Lys17Ala

F 5' – TTCTTGCAAGCAGAGGGCGTGCTCGATAT – 3'

R 5' – ACGCCCTCTGCTGCAAGAACTCGTC –3'

CbbX(nuc) Gln16Ala, Lys17Ala, Glu18Ala

F 5' – TTCTTGCAAGCAGCAGGGCGTGCTCGATAT – 3'

R 5' – ACGCCTGCTGCTGCAAGAACTCGTC –3'

The reaction mixtures were subjected to the following thermocycling conditions: 1 × 98°C for 30 seconds, 16 × (98°C for 10 seconds, a gradient annealing temperatures ranging from 58°C to 68°C for 30 seconds, 72°C for 4 min and 10 seconds). Final extension was at 72°C for 4 min and the reaction mixtures were kept at 4°C. For enzymatic digestion, 1 µl of DpnI and 2 µl of 10x cutsmart buffer (NEW England BioLabs) were added to the 20 µl reaction mixture and incubated at 37°C for 1 hour. The digested mixture was then transformed into E. coli DH5α and plated onto 2x YT agar with 50 µg/ml of kanamycin. Plasmid DNA was extracted and verified by sequencing to ensure the right mutation was introduced.

2.3 Small scale protein expression and purification

Transformants of full length CbbX(pt) were grown overnight in 5 mL of LB broth at 37°C with 50 µg/mL of kanamycin and 34 µg/mL of chloramphenicol. The culture was then transferred into 50 mL of LB broth containing the same antibiotics. Protein expression was induced with 0.5 mM of IPTG when the optical density (OD600) of the culture reached 0.6. The culture was shaken at 16°C overnight. The cells were then centrifuged at 4°C, 5,000 g for 10 min.

To prepare the cells for lysis, the pellet was resuspended in lysis buffer containing 50 mM Tris-HCl pH 10.5 and 300 mM NaCl. The cells were lysed for 5 min at amplitude of 40% with 3 seconds interval (130 watts, 30KHz, SONICS Vibra-Cell™). The lysate was centrifuged and the supernatant was incubated with nickel sepharose beads (GE Healthcare) for more than 1 h. The incubated supernatant was transferred to the Poly-prep chromatography column (BIO-RAD) to be purified. Flow through was first collected. The column was then washed with washing buffer

containing 50 mM Tris-HCl pH 10.5, 300 mM NaCl, and 20 mM imidazole to remove non-specifically bound contaminants. The target protein was finally eluted out with eluting buffer comprising 20 mM Tris-HCl pH 10.5, 300 mM NaCl, and 500 mM imidazole. SDS-PAGE was performed to check the expression of the full length CbbX(pt).

Small scale expression and purification for CbbX(nuc) were done by our collaborator Mueller-Cajar's group using similar methods as described above. The antibiotics added to the bacterial culture media were 100 µg/ml of ampicillin and 34 µg/ml of chloramphenicol, respectively. The lysis buffer for CbbX(nuc) contains 50 mM Tris-HCl pH 8.0, 50 mM NaCl, and 10 mM imidazole, and the buffer to wash out contaminants consists of 50 mM Tris-HCl pH 8.0, 50 mM NaCl, and 25 mM imidazole. The concentration of imidazole was increased to 200 mM in the eluting buffer.

Small scale expression and purification for CbbX complex and Δ CbbX complex were carried out by co-lysing CbbX(pt)/ Δ CbbX(pt) with CbbX(nuc). Lysis buffer contained 50 mM Na₂HPO₄, pH 8.0, 500 mM NaCl, 10% glycerol, 10.5 mM MgCl₂, and 500 units of DNaseI. Purification protocol followed the same method as described above.

2.4 Large scale protein expression and purification

Transformants of CbbX(pt) were grown in 50 mL of LB medium overnight with 50 µg/mL of Kanamycin and 34 µg/mL of Chloramphenicol. The bacterial culture was then inoculated into six flasks containing 800 mL LB medium each, supplemented

with antibiotics. The media was shaken at 37°C at 200 rpm until OD600 reached 0.6. The temperature was then reduced to 16°C and protein expression was induced with 0.5 mM of IPTG. The cells were cultured for 20 h and harvested by centrifugation at 4°C, 500 X g for 30 min. The pellet was resuspended in lysis buffer containing 50 mM Tris-HCl, pH 10.5, 500 mM NaCl, 10% glycerol, 10.5mM MgCl₂, and 500 units of DNaseI, then lysed using French press (GlenMills). After centrifugation, the supernatant was filtered using 0.45 µm pore-size filters (Sterivex, Millipore), and subjected to two rounds of purification steps, that were carried out with AKTA Purifier (GE Healthcare) at 4°C.

The filtered supernatant was first run through a nickel affinity column (GE Healthcare), which was pre-equilibrated with the buffer 50 mM Tris-HCl pH 10.5, 500 mM NaCl, 10% glycerol, and 40 mM imidazole. The target protein was eluted out by a linear gradient increase of imidazole (40 to 500 mM). Fractions were chosen to run on a 10% SDS-PAGE to check for protein purity. Fractions containing the target protein were pooled together and further purified by size-exclusion chromatography with HiLoad 26/60 Superdex 200 pg column (GE healthcare). The column was pre-equilibrated with the buffer containing 50 mM Tris-HCl, pH 10.5, 500 mM NaCl, 10% glycerol. The target protein was concentrated to a final concentration of 5 mg/mL, snap froze, and stored at -80°C freezer before crystallization screens.

Similarly, CbbX(nuc) was prepared with the same procedure. Upon overnight induction, the cells were harvested and the pellet was suspended in lysis buffer containing 50 mM Na₂HPO₄, pH 8.0, 500 mM NaCl, 10% glycerol, 10.5 mM MgCl₂,

and 500 units of DNaseI. The proteins were first purified using nickel affinity column with a different set of His A and His B buffers (His A: 50 mM Na₂HPO₄ pH 8.0, 500 mM NaCl, 10% glycerol, 20 mM imidazole pH 8.0; His B: 50 mM Na₂HPO₄ pH 8.0, 500 mM NaCl, 10% glycerol, 500 mM imidazole, pH 8.0). The eluent was checked against SDS-PAGE to locate the target protein. Fractions containing target protein were collected, and the sample was further purified by size-exclusion chromatography using HiLoad 16/60 Superdex 200 pg column. The buffer for this step consists of 20 mM Tris-HCl, pH 8.0 and 150 mM NaCl. The final concentration of target protein was kept at 6.8 mg/ml.

For CbbX complex/ Δ CbbX complex, CbbX(pt)/ Δ CbbX(pt) and CbbX(nuc) were cultured separately and co-lysed in buffer containing 50 mM Na₂HPO₄, pH 8.0, 500 mM NaCl, 10% glycerol, 10.5 mM MgCl₂, and 500 units of DNaseI. Subsequent purification and buffers used followed CbbX(nuc) purification protocol, with the addition of overnight cleavage of His6 – ubiquitin tag from CbbX(nuc) using enzyme USP2 prior to size exclusion chromatography.

2.5 Crystallization

Crystal screening for CbbX(pt), CbbX(nuc) and CbbX complex were carried out through sitting drop vapor- diffusion method using 96 well intelli-plate[®] (Art Robbins Instruments). Different crystallization screening kits were used, which include Wizard Screen (Emerald BioStructures), PegRx Screen, and Crystal Screen. All screening kits are purchased from Hampton Research unless otherwise stated. The automated dispensing robot Phoenix (Art Robbins Instruments) was used to perform the screening. Three drops with different protein to reservoir ratios (0.2 μ l : 0.1 μ l, 0.15 μ l

: 0.15 μ l, 0.1 μ l : 0.2 μ l) were set up per condition in each well of the intelli-plate®. The plates were then incubated in RockImager (Formulatrix) at 20 °C for over 2 weeks, during which, images of crystal growth were automatically recorded.

Once crystal hits were observed, the conditions were further optimized via hanging drop vapor diffusion method using VDX 24-well plates (Hampton Research). CbbX(pt) crystals grew in 0.1 M MES pH 6, 0.05 M CaCl₂, and 45% polyethylene glycol 200 condition, and 0.1 M trisodium citrate pH 5.5, 0.1 M lithium sulfate monohydrate, and 20% w/v polyethylene glycol 1,000 condition. CbbX(nuc) crystals grew in 0.1 M imidazole pH 7.0, 20% polyethylene glycol MME2K, 0.2 M ammonium citrate condition, 0.1 M trisodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2 M ammonium acetate condition, and 0.1 M tris pH 7.8, 16% polyethylene glycol 3350, and 2% dioxane condition. CbbX complex crystals grew in 1M ammonium sulphate, 0.1M Tris 8.5 condition, 0.1M tri-sodium citrate 5.6, 0.1M lithium sulphate monohydrate, 12% polyethylene glycol 6000 condition, 0.1M tri-sodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2M ammonium acetate condition, and 0.2M lithium sulphate monohydrate, 0.1M Tris pH 8.5, 40% polyethylene glycol 400 condition. Δ CbbX complex crystals grew in 0.1M lithium sulfate monohydrate, 0.1M tri-sodium citrate pH 5.6, 4% polyethylene glycol 400 condition, and 0.1M lithium sulfate monohydrate, 0.1M tri-sodium citrate pH 5.5, 20% polyethylene glycol 1,000 condition.

Crystals were picked and cryoprotected in reservoir condition containing 10-20% glycerol (v/v) or an increased percentage of polyethylene glycol depending on the

conditions of the crystal. The crystals were flash-cooled in liquid nitrogen prior to data collection.

2.6 Data collection

Crystals of CbbX(pt) were sent to National Synchrotron Radiation Research Center (NSRRC) in Taiwan, and screened on 13B1 beamline. Crystals of one particular condition, 0.05M CaCl₂, 0.1M MES, pH6, polyethylene glycol 200 45%, diffracted up to 3.9 Å. Data was collected and processed by HKL2000 (Otwinowski and Minor 1997) to P4 space group. It was subsequently reindexed to P4₃ using Reindex from CCP4 program suite (Collaborative Computational Project, Number 4, 1994). The structure of CbbX(pt) was later solved by Balbes subroutine of CCP4 on-line (Long, Vagin et al. 2008). A structural homologue 3SYL (PDB ID) was identified by Balbes as a suitable model for molecular replacement. Molecular replacement was performed by Molrep (Vagin and Teplyakov 2010). Model building was done using ARP/wARP (Langer, Cohen et al. 2008) and refinement using refmac (Murshudov, Skubak et al. 2011). Subsequent rounds of model building and structure refinement were carried out through COOT (Emsley and Cowtan, 2004) and PHENIX (Afonine et al., 2012) until the R_{work} and R_{free} values could no longer be improved further.

Crystals of CbbX(nuc) were flown to Australian Synchrotron (AS) and screened on AX1 beamline. Data were processed by XDS (Kabsch 2010) at the AS. Crystals of CbbX(nuc) generally diffracted between 6 to 9 Å and no data were collected due the unsatisfactory diffraction of the crystals.

Crystals of CbbX complex were sent to beamline PX1 of the Swiss Light Source (SLS) (Villigen, Switzerland). Similar to CbbX(nuc) crystals, no data were collected due to the lack of diffraction of the crystals.

2.7 Protein sequence prediction and alignment

Structural homologs of CbbX were identified from Balbes subroutine of CCP4 on-line (Long, Vagin et al. 2008) and Dali server (<http://ekhidna2.biocenter.helsinki.fi/dali/>). Sequence alignment of CbbX and its structural homologs was carried out through T-COFFEE (<http://tcoffee.crg.cat/>) and subsequently generated by ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/ESPript/>).

3. RESULTS

3.1 CbbX(pt)

3.1.1 Small scale expression of full length CbbX(pt)

The pI of CbbX(pt) is predicted to be 9.3 and it has a molecular weight of approximately 34.5 kDa. Based on the pI of the target protein, two different buffer pH were tested during the small scale expression and purification – pH 8.0 and pH 10.5.

The results were then checked against SDS-PAGE.

Figure 38: Small-scale expression tests of CbbX(pt) using A) pH 8.0 buffer and B) pH 10.5 buffer. A) Target protein showed both poor expression and solubility. Protein band was nearly invisible under the elution lane on the gel. B) Target protein with the correct molecular weight could be clearly seen under the elution lane, highlighted in red box. Both expression and solubility were improved with the change to a higher pH buffer.

Both lysis and eluting buffers were Tris-HCl based, consisting of 50 mM Tris-HCl pH 8.0 or pH 10.5 and 300 mM NaCl, while eluting buffer contains an additional 500 mM of imidazole. From the SDS-PAGE results (**Figure 38**), the target protein lysed and purified using pH 10.5 buffer had much better solubility. A protein band could be clearly observed at around 35 kDa level of the protein ladder, which corresponds to the molecular weight of CbbX(pt) (**Figure 38B**). CbbX(pt) was therefore proceeded to large scale purification using Tris-HCl pH 10.5 buffer.

3.1.2 Large scale purification of CbbX(pt)

CbbX(pt) was purified using a two-step procedure that includes usage of nickel affinity column, followed by HiLoad 26/60 Superdex 200 pg column.

The yield of the protein was relatively high and could be reflected from the UV absorption, which exceeded 1000 mAU (**Figure 39A**). The protein was eluted at ~27.5% imidazole, equivalent to approximately 137.5 mM of imidazole. Results were confirmed by SDS-PAGE, showing proteins bands of the correct molecular weight corresponding to the peak (**Figure 39A**). Fractions containing the target protein were pooled together, concentrated and loaded onto HiLoad 26/60 Superdex 200 pg column for further purification.

Figure 39: Large scale purification of CbbX(pt). A) Nickel affinity purification. Y-axis: absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted with 27.5% of imidazole (equivalent to 137.5 mM). SDS-PAGE showing target protein bands corresponding to the peak. Protein yield was relatively high. B) Size exclusion purification of CbbX(pt) using 26/60 Superdex 200 pg column. Y-axis: absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted out at 134 ml. SDS-PAGE showing target protein bands corresponding to the peak.

The elution volume for CbbX(pt) from the 26/60 Superdex 200 pg column was 134 ml. The elution profile has a characteristic long tail dragging towards the right side of the profile (**Figure 39B**). Target protein was confirmed by SDS-PAGE, and tested for possible DNA contamination using NanoDrop2000 (Thermo Scientific). The final 260/280 ratio was 0.76, which is within the acceptable range. However, the tail dragging observed from the elution profile could indicate inhomogeneity of the target protein. Therefore, to check for homogeneity of the target protein, fractions in the peak were divided into three parts – left, middle, and right. Each part was concentrated to 5 mg/ml, and subjected to crystal screens in separate attempts. Out of the three parts tested, only the left part of the peak could be crystalized.

3.1.3 Crystallization of full length CbbX(pt)

The protein concentration was set at 5 mg/ml prior to crystal screening. Three different screening kits were tried, which include PegRX Screen, Wizard Screen, and Crystal Screen. All three screenings gave rise to crystal hints. Among which, PegRX Screen had the most hits.

Two different shapes of crystals were observed in the screenings – rod and cubic shape. The crystals in PegRX Screen generally appeared to be bigger. Two particular conditions bearing different shapes of crystals were chosen to be optimized first, which were 1) 0.1 M MES pH 6, 0.05 M CaCl₂, polyethylene glycol 200 45% condition (**Figure 40A**) and 2) 0.1 M trisodium citrate pH 5.5, 0.1 M lithium sulfate monohydrate, 20% w/v polyethylene glycol 1,000 condition (**Figure 40B**). However, crystals of trisodium citrate condition did not diffract in a preliminary test using our in-house Bruker single crystal x-ray diffractometer, while the rod-shape like crystal in MES and CaCl₂ condition diffracted up to 8 Å. This condition was therefore further optimized. The crystals were sent to the National Synchrotron Radiation Research Center in Taiwan for diffraction test, among which, a particular crystal diffracted up to 3.9 Å (**Figure 40C**).

3.1.4 Data collection and refinement of CbbX(pt)

Table 3: Summary of crystallographic data and refinement statistics for CbbX(pt) from *C. merolae*

Data collection	
Wavelength (Å)	1.0000
Space group	P4 ₃
Unit cell dimensions	
a,b,c (Å)	a=109.66, b=109.66, c=87.44
α, β, γ (°)	$\alpha=90$ $\beta=90$ $\gamma=90$
Resolution (Å)	57.79-3.90
No. of unique reflections	14807
Rsym (%)	11.6(37.5)
I/ σ I	12.1(2.1)
Completeness (%)	96.7(98.4)
Redundancy	4.4(2.2)
Refinement	
Resolution (Å)	57.79-3.9
No. of unique reflections	14701
R _{work} /R _{free}	26.2/32.6
No. of atoms	3657
Protein	3636
Ligand/ion	-
Water	21
B-factors	
Proteins	86.5
Rmsd from idealty	
Bond length (Å)	0.013
Bond angle (°)	1.650
Ramachandran	
Favored (%)	79.81
Allowed (%)	18.08
Outlier (%)	2.12

Values in parentheses are for the highest-resolution shell.

Diffraction data of a single CbbX(pt) crystal collected from the 13B1 beamline showed diffraction up to 3.9 Å. The space group was determined to be P4₃, with cell dimensions of a=109.66, b=109.66, c=87.44; $\alpha=\beta=\gamma=90^\circ$ (**Table 3**). There are two

CbbX(pt) protomers in one asymmetrical unit. Structural homologue 3SYL (Mueller-Cajar, Stotz et al. 2011), which is the crystal structure of an AAA⁺ protein CbbX from *R. sphaeroides*, was taken from the PDB for molecular replacement to solve the structure of CbbX(pt). Subsequent model building through Coot (Emsley, Lohkamp et al. 2010) improved R_{work} and R_{free} values of CbbX(pt) to 0.262 and 0.326 respectively.

3.1.5 Overall crystal structure of CbbX(pt) from *C. merolae*

The structure of CbbX(pt) from *C. merolae* contains two protomers per asymmetric unit. It has the typical characteristics of the architecture of AAA⁺ family proteins (Mueller-Cajar, Stotz et al. 2011).

Figure 41: Crystal structure of CbbX(pt) from *C. merolae* shown in ribbon representation. The space group is determined to be $P4_3$, with two CbbX(pt) protomers in an asymmetrical unit. Chain A of CbbX(pt) is depicted in red, while chain B is depicted in green.

A protomer of CbbX(pt) contains an α/β sub-domain at its N-terminal, and a smaller α -helical sub-domain at its C-terminal (**Figure 41**). The two subdomains are connected by a short linker in between. The N-terminal domain is wedge-shaped and has sheets of beta-strands in parallel. It is the ATP-binding domain, where the typical

Walker A and B motifs are found. Walker A and B mutants have shown to be inactive in ATP hydrolysis and Rubisco activation (Mueller-Cajar, Stotz et al. 2011). The shorter C-terminal α -helical sub-domain consists of several alpha helices, which are important for distinguishing AAA⁺ family of proteins from other nucleotide-binding proteins (Hanson and Whiteheart 2005). The dimer solved was most probably a crystallographic dimer, as CbbX from *C. merolae* is speculated to be a functional hexamer in its biological assembly.

3.1.6 Structural comparison between CbbX(pt) and 3SYL

Superimposition of a protomer of CbbX(pt) from *C. merolae* with a protomer of its homologs 3SYL from *R. sphaeroides* showed very similar secondary structures, which could be observed from the overlapping beta sheets and alpha helices. Superimposition of CbbX(pt) to 3SYL had a RMSD of 1.7 Å, suggesting a high level of convergence for the superimposed models (**Figure 41**).

Figure 42: Superimposition of crystal structures of CbbX(pt) with its homologue 3SYL shown in ribbon representation. Red: crystal structure of CbbX(pt) from *C. merolae*. Pale cyan: crystal structure of 3SYL from *R. sphaeroides*. Sulphate ions from the 3SYL structure are depicted in yellow. Superimposition of the two structures showed extremely high level of conservation, especially at the RuBP binding region, which is found in the C-terminal α -helical sub-domain of both structures.

CbbX requires RuBP for activation prior to activating Rubisco (Loganathan, Tsai et al. 2016). In the structure solved by Mueller-Cajar, Stotz et al. (Mueller-Cajar, Stotz et al. 2011), two sulphate ions that derive from their crystallization condition can be spotted in the C-terminal α -helical sub-domain and one more sulphate ion can be seen in the N-terminal Walker A motif in the α/β sub-domain (**Figure 42**). The distance of the sulphate ions in the α -helical sub-domain is equivalent to the distance of phosphate-phosphate in the RuBP. This probably suggested an allosteric binding site for RuBP, which is further confirmed by the successful fitting of RuBP in its fully extended conformation into the pocket (Stotz, Mueller-Cajar et al. 2011).

Figure 43: Conservation of the residues at RuBP binding site in crystal structures of both CbbX(pt) and 3SYL. Red: crystal structure of CbbX(pt) from *C. merolae*. Pale cyan: crystal structure of 3SYL from *R. sphaeroides*. Sulphate ions from the 3SYL structure are depicted in yellow. There is a high level of conservation at the RuBP binding site, which can be seen from the overlapping residues around the sulphate ions. Surrounding residues are labeled accordingly.

Zooming into the allosteric binding site for RuBP found in the C-terminal α -helical sub-domain, it can be clearly seen that the residues surrounding the binding site in CbbX(pt) from *C. merolae* are highly conserved, for instance, Arg 238, 256, 260, Tyr 234, and Asn 252. The high level of structural conservation indicates that the

allosteric binding site for RuBP in CbbX(pt) from *C. merolae* could be at the same position, as compared to CbbX from *R. sphaeroides* (**Figure 43**).

Sequence alignment of CbbX(pt) and 3SYL was done using MAFFT multiple sequence alignment and ESPrit programs (Robert and Gouet 2014) to further confirm the structural conservation. Vast similarity of the secondary structures between two proteins can be observed from the overlapping alpha helices and beta sheets. Residues such as Arg 238, 256, 260, Tyr 234, and Asn 252 highlighted with blue stars (**Figure 44**) are highly conserved and consistent with the crystal structure of CbbX(pt).

Figure 44: Sequence alignment of CbbX(pt) from *C. merolae* and its homolog 3SYL from *R. sphaeroides* using MAFFT multiple alignment and ESPrit programs. Red and yellow shaded boxes indicate the conserved and conserved-changed residues, respectively. Black coils and arrows represent secondary structures. Coils indicate alpha helices, while arrows indicate beta strands. From the sequence alignment, most of the secondary structures of CbbX(pt) and 3SYL are well conserved. Residues that are conserved at the RuBP binding site are indicated with the blue stars – Arg 238, 256, 260, Tyr 234, and Asn 252.

3.1.7 Structural comparison of CbbX(pt) with 3CF3

CbbX from *C. merolae* is an ATPase and has an ATP-binding site at the N-terminal α/β sub-domain, therefore, another structure 3CF3 (DeLaBarre and Brunger 2003) was used to superimpose with CbbX(pt) to check for conservation at the ATP binding site (**Figure 45**).

3CF3 is the crystal structure of an ATPase protein p97/VCP (valosin-containing protein). It is another classical example of AAA⁺ family of proteins, which is highly involved in altering the morphology of nuclear and mitotic Golgi membranes.

Figure 45: Superimposition of crystal structures of CbbX(pt) with another homologue 3CF3 shown in ribbon representation. Red: crystal structure of CbbX(pt) from *C. merolae*. Pale green: crystal structure of 3CF3 from *S. cerevisiae*. ADP from the 3CF3 structure is depicted in yellow. Superimposition of the two structures showed an RMSD of 4 Å, with few residues conserved at the ATP binding site.

The crystal structure of 3CF3 includes ADP molecule at the ATP binding site since it was co-crystallized with ADP, while CbbX(pt) is in the apo form. A protomer of 3CF3 with the ADP molecule was superimposed with the protomer of apo CbbX(pt) (**Figure 45**). The superimposition of two protein protomers showed certain extent of structural change, in which the RMSD of both structures is 4 Å, aligned over 208 residues.

Comparing the ATP binding sites of CbbX(pt) and 3CF3 (**Figure 46**), there are few residues from CbbX(pt) that are conserved, for example, Lys 79 and Thr 80. These residues might probably interact in an electrostatic or hydrophobic manner with the di-phosphate of the ADP, inferring from the position of the ADP molecule. Apart from Lys 79 and Thr 80, Leu 218 and Gln 222 might also play a part in ATP binding. Gln 222 could interact with ADP thru charge interactions, while Leu 218 might be interacting with the ADP hydrophobically.

Figure 46: Conservation of the residues at ATP binding site in crystal structures of both CbbX(pt) and 3CF3. Red: crystal structure of CbbX(pt) from *C. merolae*. Pale cyan: crystal structure of 3CF3 from *M. musculus*. ADP from the 3SYL structure is depicted in yellow. The structure of the ATP binding site in both proteins have certain degree of similarity, with some conserved residues around the ADP binding site of both structures including Lys 79 and Thr 80, which are labeled accordingly.

Observed from the sequence alignment (Robert and Gouet 2014) of CbbX(pt) and 3CF3, there is certain degree of similarity in the secondary structures of both proteins (**Figure 47**). Residues involved in ATP binding such as Lys 79 and Thr 80 from CbbX(pt) are conserved, which is consistent with the superimposition of CbbX(pt) and 3CF3 crystal structures. Other residues, for instance, Leu 218 and Gln 222 are not

conserved, but they are most likely involved in ATP binding as well, through hydrophobic and charge-charge interactions respectively.

Figure 47: Sequence alignment of CbbX(pt) from *C. merolae* and homolog 3CF3 from *S. cerevisiae* using MAFFT multiple alignment and ESPrit programs. Red and yellow shaded boxes indicate the conserved and conserved-changed residues, respectively. Black coils and arrows represent secondary structures. Coils indicate alpha helices, while arrows indicate beta strands. From the sequence alignment, the secondary structures of CbbX(pt) and 3SYL have shown some conservation. Residues that are conserved at the ATP binding site are indicated with blue stars.

3.2.1 Large scale purification of full length CbbX(nuc)

Figure 48: Large scale purification of CbbX(nuc). A) Nickel affinity purification. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted with 43.3% of imidazole (equivalent to 216.5mM). SDS-PAGE showing target protein bands and contaminants, only fractions corresponding to the 3rd peak were collected for next round of purification. B) Size exclusion purification of CbbX(nuc) using 16/60 Superdex 200 pg column. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted out at 68 ml. SDS-PAGE showing target protein bands corresponding to the 2nd peak on the gel filtration profile.

The elution volume for CbbX(pt) was 68 ml (**Figure 48B**), which confirmed to be the target protein after running SDS-PAGE. Despite the presence of a prominent peak that was eluted out before the target protein peak, it appeared to be nothing on the SDS-PAGE (refer to lane 3 on **Figure 48B**). This means that the contaminant was most likely not a protein, and it might be a nucleic acid aggregate. The shape of the peak for CbbX(nuc) resembled that of CbbX(pt), which had a dragging tail towards the right side of the profile. To ensure homogeneity of the target protein, only fractions within the first one third of the peak area were collected, concentrated, and subjected to crystal screening.

3.2.2 Crystallization of full length CbbX(nuc)

Among the screenings carried out for crystallization of CbbX(nuc). Many conditions from PegRx gave crystal hits. Similar to CbbX(pt), most of the crystals appeared to be rod-like shape.

Figure 49: Crystals of full length CbbX(pt). A) Crystals growing in 0.1 M imidazole pH 7.0, 20% polyethylene glycol MME2K, 0.2 M ammonium citrate. B) Crystals growing in 0.1 M trisodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2 M ammonium acetate. C) Crystals growing in 0.1 M tris pH 7.8, 16% polyethylene glycol 3350, and 2% dioxane. All crystals observed generally appeared to be rod shape. D) Crystals optimized from 0.1 M imidazole pH 7, 20% polyethylene glycol MME2K, and 0.2 M ammonium citrate, which diffracted between 5 to 7 Å.

Bigger crystals were randomly picked for in-house diffraction testing using Bruker single crystal x-ray diffractometer, which showed no diffraction. Three conditions were subjected to further optimization, which included 1) 0.1 M imidazole pH 7.0, 20% polyethylene glycol MME2K, 0.2 M ammonium citrate (**Figure 49A**), 2) 0.1 M

trisodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2 M ammonium acetate, and 3) 0.1 M tris 7.8, 16% polyethylene glycol 3350, 2% v/v dioxane. Initial crystals could be observed within two days. But crystals grew bigger in a two-week period.

Crystals of CbbX(nuc) grew in different conditions were sent to the Australian Synchrotron for diffraction test. Among which, crystals which grew in the condition that had trisodium citrate and tris as buffer diffracted weakly between 13 to 19 Å (**Figure 49B & 49C**). Crystals that grew in the condition consisting of imidazole as buffer had relatively better diffraction, which was between 5 to 7 Å (**Figure 49D**). However, the diffraction was still considered suboptimal, thus no data was collected for CbbX(nuc) to be processed, further crystallization optimization will be required.

3.3 CbbX complex

3.3.1 Large scale purification of CbbX complex

CbbX(pt) Δ His6 and CbbX(nuc) were cultured separately and co-lysed in buffer containing 50 mM Na₂HPO₄, pH 8.0, 500 mM NaCl, 10% glycerol, 10.5 mM MgCl₂, and 500units of DNaseI. Subsequently it was purified with nickel affinity column followed by overnight cleavage of His6-ubiquitin tag from CbbX(nuc) using USP2 enzyme. Cleaved CbbX complex was subjected to another round of nickel affinity column to remove the His6-ubiquitin tag, followed by size exclusion chromatography on the next day.

Figure 50: Purification of CbbX complex. A) Nickel affinity purification. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted with 16.4% and 24.6% of imidazole (equivalent to 82 mM and 123 mM, respectively). SDS-PAGE showing target protein bands from two peaks in nickel affinity profile. Only fractions from second peak were collected for further purification. B) Size exclusion purification of CbbX complex using 16/60 Superdex 200 pg column. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted out at 48.9 ml and 64.5 ml in the gel filtration column. SDS-PAGE showing target protein bands from two peaks in the gel filtration profile. Only proteins from second peak of the profile were collected and subjected to crystal screening.

Prior to purification of CbbX complex, mutagenesis of CbbX(pt) was done to remove its His6 tag. This was to prevent non-specific binding of CbbX(pt) to the nickel affinity column during co-purification of CbbX(pt) and CbbX(nuc), which could potentially cause false positive result of complex formation. Nickel affinity profile of CbbX complex yielded two peaks eluted with 82 mM and 123 mM of imidazole (**Figure 50A**). Both peaks corresponded to target complex according to SDS-PAGE (**Figure 50A**). Both peaks were collected and concentrated separately. However, the peak that eluted with 82 mM of imidazole precipitated heavily during concentration, thus excluded from subsequent purification steps. While the peak eluted with 123 mM

of imidazole was successfully collected and cleaved by USP2 enzyme overnight to remove the His6-ubiquitin tag from CbbX(nuc) in the complex. The tagless complex was then passed through 16/60 Superdex 200 pg column. Two elution volumes were again observed, at 48.9 ml and 64.5 ml (**Figure 50B**). The first peak consisted of both cleaved and uncleaved CbbX(nuc) in the complex, while the second peak comprised only cleaved CbbX complex as observed on SDS-PAGE (**Figure 50B**). Therefore, only second peak of target complex was pooled and subjected to crystal screenings.

3.3.2 Crystallization of CbbX complex

Out of all the screening kits tested, Memfac/natrix, PegRx, and Cryo1&2 gave crystal hits. Two conditions from Memfac/natrix include 1M ammonium sulphate, 0.1M Tris 8.5, and 0.1M tri-sodium citrate pH 5.6, 0.1M lithium sulphate, 12% polyethylene glycol 6000 (**Figure 51A and 51B**). One condition from PegRx consists of 0.1M tri-sodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2M ammonium acetate (**Figure 51C**). The condition from Cryo1&2 consists of 0.2M lithium sulphate monohydrate, 0.1M Tris pH 8.5, 40% polyethylene glycol 400 (**Figure 51D**). All conditions observed had shower-like crystals.

Figure 51: Crystals of CbbX complex. A) Crystals from Memfac/natrix condition 1M ammonium sulphate, 0.1M Tris 8.5. B) Crystals from Memfac/natrix condition 0.1M tri-sodium citrate 5.6, 0.1M LiSO₄, 12% polyethylene glycol 6000. C) Crystals from PegRx condition 0.1M tri-sodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2M ammonium acetate. D) Crystals from Cryo1&2 condition 0.2M lithium sulphate monohydrate, 0.1M Tris pH 8.5, 40% polyethylene glycol 400. All conditions produced small shower-like crystals.

All conditions were optimized by varying pH values and percentage of precipitants. However, shape and size of the crystals remained similar upon optimization. Crystals from PegRx condition were picked and sent to beamline PX1 of the Swiss Light Source (Villigen, Switzerland). However these crystals did not show diffraction and no data were collected as a result.

3.3.3 Protein crystallization by mutagenesis of CbbX surface residues

Due to persistent poor diffraction of both CbbX(nuc) and CbbX complex crystals, mutagenesis of CbbX surface residues was proposed in hope of improving crystal

diffraction. SERp server (<http://services.mbi.ucla.edu/SER/intro.php>) was used to suggest mutations of the surface residues on CbbX to augment its crystallizability. The method employs generation of crystal contacts via surface entropy reduction. Different clusters of mutations on both CbbX(pt) and CbbX(nuc) were suggested based on the results from the server (**Table 4**). Clusters were chosen according to the SERp score, which indicates high entropy residue.

Table 4. Surface mutations of CbbX. Different clusters of surface mutations suggested by SERp server on both CbbX(pt) and CbbX(nuc).

Proposed mutations	CbbX(pt)	CbbX(nuc)
Cluster 1	K270A, K271A	E162A, E163A, Q164A
Cluster 2	K96A, K97A	Q16A, K17A, E18A
Cluster 3	E217A, Q218A, Q219A	K6A, E7A

3.3.4 Small scale expression of CbbX complex upon surface mutations

Upon mutating Cluster 1 of both CbbX(pt) and CbbX(nuc) (**Table 4**), small scale expression tests were carried out. Different combinations of wild type CbbX and mutated CbbX were considered, for instance, Δ CbbX(pt) mixed with Δ CbbX(nuc), wild type CbbX(pt) mixed with Δ CbbX(nuc), and Δ CbbX(pt) mixed with wild type CbbX(nuc) (**Figure 52**).

In the event of Δ CbbX(pt) mixed with Δ CbbX(nuc), both proteins became insoluble and appeared in the pellet fraction (**Figure 52A**). When wild type CbbX(pt) mixed with Δ CbbX(nuc), majority of the complex was insoluble as well (**Figure 52B**).

However, when Δ CbbX(pt) copurified with wild type CbbX(nuc), fractions of soluble CbbX(nuc) and Δ CbbX(pt) could be pulled down together (**Figure 52B**).

Figure 52: Small scale expression tests of CbbX complex upon surface mutations. A) SDS-PAGE showing Δ CbbX(pt) copurified with Δ CbbX(nuc). Both proteins became insoluble. Insoluble bodies of CbbX(pt) and CbbX(nuc) are highlighted in black and red boxes, respectively. B) SDS-PAGE showing CbbX(pt) copurified with Δ CbbX(nuc) and Δ CbbX(pt) copurified with CbbX(nuc). Only Δ CbbX(pt) copurified with CbbX(nuc) showed soluble complex under elute fraction.

3.3.5 Large scale purification of Δ CbbX complex

From results of small scale expression on CbbX complex with surface mutations, wild type CbbX(nuc) and Δ CbbX(pt) were chosen to proceed to large scale purification. Purification protocol followed wild type CbbX complex, which involved nickel affinity column and overnight cleavage of His6-ubiquitin tag from CbbX(nuc) and a final step of size exclusion chromatography.

Figure 53: Purification of Δ CbbX complex with surface mutations. A) Nickel affinity purification. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted with 15.4% of imidazole (equivalent to 77mM). SDS-PAGE showing target protein bands from three peaks in nickel affinity profile, only fractions from second peak were collected for further purification. B) Size exclusion purification of CbbX complex using 16/60 Superdex 200 pg column. Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. Target protein was eluted out at 45.7 ml and 65.8ml in the gel filtration column. SDS-PAGE showing target protein bands from two peaks in the gel filtration profile. Only proteins from second peak of the profile were collected and subjected to crystal screening.

During nickel affinity purification, Δ CbbX complex was eluted at three different peaks (**Figure 53A**). However, only the second peak that was eluted with 77mM of imidazole corresponded to Δ CbbX complex at the right molar ratio. CbbX(nuc) and CbbX(pt) always bind to each other with 1:1 molar ratio. The most right peak in the profile appeared to contain mostly CbbX(nuc) only (**Figure 53A**). Thus only fractions corresponding to the second peak in the nickel affinity profile were collected for overnight cleavage using USP2 enzyme. Once cleaved, Δ CbbX was subjected to size exclusion chromatography using 16/60 Superdex 200 pg column.

Δ CbbX was eluted at two different volumes during size exclusion chromatography – 45.7ml and 65.8ml (**Figure 53B**). The first peak at 45.7ml in the profile corresponded to aggregated Δ CbbX complex, while the second peak at 65.8ml corresponded to soluble fractions of Δ CbbX complex with the right binding ratio based on SDS-PAGE (**Figure 53B**). Elution volumes of Δ CbbX complex were similar compared to wild type CbbX complex. Only fractions corresponding to the second peak were pooled and concentrated to 13 mg/ml prior to crystal screenings.

3.3.6 Crystallization of Δ CbbX complex

Crystallization of Δ CbbX complex was carried out using four screening kits including Index, Memfac/natrix, PegRX, and Crystal Screen. Among the screenings, Memfac/natrix, PegRX conditions gave crystal hits. Urchin-like crystals were observed in Memfac/natrix condition consisting of 0.1M lithium sulfate monohydrate, 0.1M tri-sodium citrate pH 5.6, 4% polyethylene glycol 400 (**Figure 54A**), and PegRX condition consisting of 0.1M lithium sulfate monohydrate, 0.1M tri-sodium citrate pH 5.5, 20% polyethylene glycol 1,000 (**Figure 54B**). Although the two conditions were from different screening kits, components in the conditions were highly similar. Both of which contained 0.1M of lithium sulfate monohydrate and 0.1M of tri-sodium citrate. Both conditions also contained polyethylene glycol as the precipitant.

Crystals appeared to be crowded and in urchin-like shape. Further optimization is required to obtain bigger and single crystals that could be picked and tested.

Figure 54: Crystals of Δ CbbX complex. A) Crystals from Memfac/natrix condition consisting of 0.1M lithium sulfate monohydrate, 0.1M tri-sodium citrate pH 5.6, 4% polyethylene glycol 400. B) Crystals from PegRX condition consisting of 0.1M lithium sulfate monohydrate, 0.1M tri-sodium citrate pH 5.5, 20% polyethylene glycol 1,000. Both conditions gave rise to urchin-like crystals.

4. DISCUSSION

4.1 Protein expression and purification

Prior to small scale expression test, several parameters of the protein have to be determined, for instance, the isoelectric point (pI), and the molecular weight (MW) of the protein. These parameters will affect the type and the composition of the buffers to be used. It is recommended to use a buffer pH differing from the pI of the protein by 1 to 2, while incompatible buffer or buffer pH has a great impact on the stability and solubility of the protein. This could be seen from the small-scale test results of CbbX(pt). In which case, CbbX(pt) in higher-than-pI buffer was much more stable, and had obviously much better solubility compared to CbbX(pt) in the lower-than-pI buffer (**Figure 38**).

For large-scale purification of CbbX(pt) and CbbX(nuc), a two-step process was chosen instead of the usual three-step process. Anion exchange was attempted during

the initial round of purification. This step generally serves to remove DNA contaminants. However, target protein failed to adhere to the HiTrap Q HP column unexpectedly.

Figure 55: Anion exchange purification of CbbX(pt) using 5 ml HiTrap Q HP column. A) Y-axis: Absorption at 280 nm in mAU; X-axis: elution volume in ml. A major portion of the protein was found in the flow-through while loading sample onto the column. Another huge portion of the protein was flowed out of the column before the salt gradient elution started (from fraction A1 to A6). The last portion of the target protein was gradually eluted out with approximately 38.7% of salt, equivalent to 387 mM. B) SDS-PAGE showing target protein bands corresponding to the flow-through, and the two major peaks observed from the elution profile.

Majority of the protein had flowed out of the column even before the gradient elution started (**Figure 55A**). There was also a considerable amount of target protein found in the flow-through (see lane 2 in **Figure 55B**) while loading protein sample into the column. Therefore, the 5 ml HiTrap Q HP column had failed to capture majority of the protein. There could be several reasons contributing to it, for instance, buffer incompatibility, protein inhomogeneity or aggregation. Due to the unsatisfactory outcome, anion exchange was omitted from the purification process. In order to get rid of the possible DNA contamination, 500 units of DNase I was added into the lysis buffer prior to purification. The addition of DNase I had successfully removed majority of the DNA contaminants, reducing the measurement of 260/280 ratio to between 0.7 to 0.8 subsequently.

4.2 Crystallization of CbbX(pt), CbbX(nuc), and CbbX complex

Having crystallized both CbbX(pt) and CbbX(nuc), there are some common characteristics that could be observed. Majority of the crystals appeared to be rod-like shape. Both proteins seemed to prefer conditions with certain salts, precipitants and additives for example, tri-sodium citrate, ammonium citrate, ammonium acetate, and imidazole. While they also preferred to grow in conditions containing polyethylene glycol 400, polyethylene glycol 3350, and polyethylene glycol MME2K. Crystals of both CbbX(pt) and CbbX(nuc) could be observed in 24 to 48 h. But it took approximately one to two weeks for the crystals to reach maximum size.

Both wild type and Δ CbbX complex crystals appeared to be small in size. Similarly to CbbX(pt) and CbbX(nuc) crystals, they preferred to grow in conditions containing tri-sodium citrate, lithium and ammonium salts, and as well as having polyethylene glycol as their precipitant. However, optimization to produce bigger diffracting crystals posed a challenge. Traditional methods of varying the pH of crystallizing components and changing the percentage of precipitant were carried out. These generally increased the size of the crystals, but results were suboptimal and diffraction remained to be a problem. For instance, a condition from Cryo1&2 consisting of 0.2M lithium sulphate, 0.1M Tris pH 8.5, and 40% polyethylene glycol 400 gave rise to CbbX complex crystals. Upon optimizing, crystals grew bigger slightly and became less crowded, which were still diffracting badly (**Figure 56A**). Various other methods have been tried including streak seeding, additive screening, and dehydration through gradient addition of glycerol (**Figure 56B**). All gave rise to similar crystals with no diffraction.

Figure 56: Various optimization results of CbbX complex. A) Crystals from Cryo1&2 condition consisting of 0.2M lithium sulphate, 0.1M Tris pH 8.5, 40% polyethylene glycol 400. Optimization of the condition by varying pH of Tris and changing the percentage of polyethylene glycol 400 yielded bigger and less crowded crystals. Diffraction of these crystals remained unsatisfactory. B) Crystals from PegRx condition 0.1M tri-sodium citrate pH 5.5, 24% polyethylene glycol 400, 0.2M ammonium acetate. Streak seeding was done on this condition. However, only microcrystals along the streak line were observed after days of seeding.

4.3 Crystallographic data for CbbX(pt) from *C. merolae*

Shown in **Table 3**, the resolution for indexing and merging of the diffraction images was cut off at 3.9 Å in $P4_3$ spacegroup. The resolution limit chosen was reasonably good, since the completeness was 96.7% overall and 98.4% at the highest resolution with $I/\sigma I$ value of 4.4 overall and 2.2 at the highest resolution. Despite the suboptimal resolution, the structure of CbbX(pt) was later solved with R_{work}/R_{free} of acceptable values using Balbes from CCP4 online.

In the structure, electron density ($2F_o - F_c$) shows unambiguous occupancy of the secondary structures and important motifs such as Walker A & B (**Figure 57**). The projection and orientation of many individual key residues in the ATP and the RuBP binding site are also clearly depicted, allowing a fairly adequate analysis of the CbbX(pt) structure.

Figure 57: Electron density for CbbX(pt) from *C. merolae*. A) Overall structure of CbbX(pt) wrapped around by electron density map ($2F_o - F_c$). The secondary structures in CbbX(pt) occupied the electron density map unambiguously. B) Zoomed-in depiction of CbbX(pt) in the electron density map. RuBP binding site was shown in this case. Major residues involved in RuBP binding such as Arg 238, 256, 260, Tyr 234, and Asn 252 are clearly shown, which occupied the electron density map precisely.

4.4 Structural analysis of CbbX(pt) from *C. merolae*

The crystal structure of CbbX(pt) from *C. merolae* is another typical example of the AAA+ family of proteins, having all the essential characteristics of AAA+ proteins. The solved structure showed two monomers in one asymmetrical unit. A single monomer consists of 2 domains – the amino-terminal α/β sub-domain and the carboxy-terminal α -helical subdomain. The ATP binding site lies in the α/β sub-domain, while the RuBP binding site lies in the α -helical subdomain. RuBP binds to an allosteric site in the C-terminal α -helical subdomain while activating CbbX. CbbX then carries out its ATPase activity at the N-terminal region and activates Rubisco.

All structural features observed in CbbX(pt) from *C. merolae* have shown the major characteristics of an AAA+ type of protein, which are in consistent with its homologues 3SYL and 3CF3.

When superimposing with 3SYL from *R. sphaeroides*, the level of structural similarity and sequence conservation have shown to be extremely high. Especially in the RuBP binding site, several residues near the region are highly conserved, for instance, Arg 256, Tyr 234, and Asn 252.

When superimposing with 3CF3 from *S. cerevisiae*, a certain level of structural similarity could be observed, though not as high as 3SYL from *R. sphaeroides*. Few residues at the ATP binding region are conserved, such as Lys 79 and Thr 80. Other residues near the region that are most likely involved in ATP binding include Gln 222 and Leu 218. The corresponding residues from its homologue 3CF3 are Asn 660 and Ile 656 respectively. Despite being not conserved, both sets of proteins (Gln 222 vs Asn 660, Leu 218 vs Ile 656) have many similarities in nature. Gln 222 and Asn 660 have the same charge and differ only by one carbon length. Thus it is reasonable to predict that Gln 222 in CbbX(pt) could interact with the N3 in the ring of adenine via charge-charge interaction. As for Leu 218 and Ile 656, both residues are hydrophobic. Thus it indicates that Leu 218 in CbbX(pt) most probably interacts with ATP through hydrophobic interactions. Therefore, despite the fact that these residues are not conserved in CbbX(pt) from *C. merolae*, they could possibly play a role in ATP binding.

In the structure of 3CF3, several alpha helices around the ATP binding site have a slightly different orientation compared to the structure of CbbX(pt) (**Figure 45**). One possible reason could be due to the presence of ADP molecule. Once ADP was trapped in the binding site in 3CFC, it triggered a slight conformational change/domain movement of 3CF3, causing the alpha helices around the binding region to move towards the pocket and make interactions with the ADP molecule. This could be the reason why the conformation of CbbX(pt), where alpha helices around the binding pocket are in slightly different orientation, in the absence of ADP molecule.

5. FUTURE WORK

Although the structure of CbbX(pt) has been solved, the resolution of the crystals could be further improved. More crystals of the same condition, as well as other conditions, should be produced and tested to obtain a better resolution data set for processing.

Crystals of CbbX(nuc) and CbbX complex have been successfully produced, but the diffractions of these crystals were relatively poor, and absent in the case of CbbX complex. Numerous methods have been carried out to improve the size and diffraction of the crystals, but efforts were to no avail.

Focus has been shifted to introducing surface mutations to CbbX(nuc) and CbbX complex. First clusters of mutations for both CbbX(pt) and CbbX(nuc) were completed. Proteins were purified and subjected to crystal screenings. However, more optimization needs to be done to produce single diffracting crystals. At the same time,

second cluster of mutations could be carried out simultaneously to test for their expression and subsequent purification to prepare for new rounds of crystal screenings.

In addition, co-crystallization of substrates and co-factors as such ATP and RuBP with both CbbX(pt) and CbbX(nuc) could be studied to unveil their binding mechanisms.

All in all, rubisco activase plays a central role in regulating the activity of Rubisco, here we reported the crystal structure of plastid isoform of CbbX from *C. merolae*, which was solved with a diffraction of 3.9 Å. Key features of AAA⁺ proteins were conserved in the structure, and high similarity between CbbX(pt) and its homologue 3SYL was observed with an RMSD of 1.7 Å. Superimposition with another homologue 3CF3 showed conservation at the ATP binding site as well. Though nonfunctional per se, it certainly provides us a glimpse of the structure of CbbX protein from *C. merolae*.

REFERENCES

- Bowien, B. and B. Kusian (2002). "Genetics and control of CO₂ assimilation in the chemoautotroph *Ralstonia eutropha*." *Arch Microbiol* **178**(2): 85-93.
- Bryant, D. A. and N. U. Frigaard (2006). "Prokaryotic photosynthesis and phototrophy illuminated." *Trends Microbiol* **14**(11): 488-496.
- Croce, R. and H. van Amerongen (2014). "Natural strategies for photosynthetic light harvesting." *Nat Chem Biol* **10**(7): 492-501.
- DeLaBarre, B. and A. T. Brunger (2003). "Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains." *Nat Struct Biol* **10**(10): 856-863.
- Ellis, R. J. (2010). "Biochemistry: Tackling unintelligent design." *Nature* **463**(7278): 164-165.
- Finn, M. W. and F. R. Tabita (2003). "Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea." *J Bacteriol* **185**(10): 3049-3059.
- Foley, J. A., N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockstrom, J. Sheehan, S. Siebert, D. Tilman and D. P. Zaks (2011). "Solutions for a cultivated planet." *Nature* **478**(7369): 337-342.
- Freeman, S. (2014). *Biological science*.
- Fujita, K., S. Ehira, K. Tanaka, K. Asai and N. Ohta (2008). "Molecular phylogeny and evolution of the plastid and nuclear encoded *cbbX* genes in the unicellular red alga *Cyanidioschyzon merolae*." *Genes Genet Syst* **83**(2): 127-133.
- Fujita, K., K. Tanaka, Y. Sadaie and N. Ohta (2008). "Functional analysis of the plastid and nuclear encoded *CbbX* proteins of *Cyanidioschyzon merolae*." *Genes Genet Syst* **83**(2): 135-142.
- Gibson, J. L. and F. R. Tabita (1997). "Analysis of the *cbbXYZ* operon in *Rhodobacter sphaeroides*." *J Bacteriol* **179**(3): 663-669.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin and S. G. Oliver (1996). "Life with 6000 genes." *Science* **274**(5287): 546, 563-547.
- Goss, T. and G. Hanke (2014). "The end of the line: can ferredoxin and ferredoxin NADP(H) oxidoreductase determine the fate of photosynthetic electrons?" *Curr Protein Pept Sci* **15**(4): 385-393.
- Han, Y. H., C. Y. Chung, D. Wessels, S. Stephens, M. A. Titus, D. R. Soll and R. A. Firtel (2002). "Requirement of a vasodilator-stimulated phosphoprotein family member for cell adhesion, the formation of filopodia, and chemotaxis in *dictyostelium*." *J Biol Chem* **277**(51): 49877-49887.
- Hanson, P. I. and S. W. Whiteheart (2005). "AAA+ proteins: have engine, will work." *Nat Rev Mol Cell Biol* **6**(7): 519-529.
- Huner, N. P., K. Dahal, L. V. Kurepin, L. Savitch, J. Singh, A. G. Ivanov, K. Kane and F. Sarhan (2014). "Potential for increased photosynthetic performance and crop productivity in response to climate change: role of CBFs and gibberellic acid." *Front Chem* **2**: 18.

Jensen, R. G. (2000). "Activation of Rubisco regulates photosynthesis at high temperature and CO₂." *Proc Natl Acad Sci U S A* **97**(24): 12937-12938.

Kaiser, W. M., G. Schroppel-Meier and E. Wirth (1986). "Enzyme activities in an artificial stroma medium : An experimental model for studying effects of dehydration on photosynthesis." *Planta* **167**(2): 292-299.

Kanesaki, Y., S. Imamura, A. Minoda and K. Tanaka (2012). "External light conditions and internal cell cycle phases coordinate accumulation of chloroplast and mitochondrial transcripts in the red alga *Cyanidioschyzon merolae*." *DNA Res* **19**(3): 289-303.

Kitano, K., N. Maeda, T. Fukui, H. Atomi, T. Imanaka and K. Miki (2001). "Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry." *Structure* **9**(6): 473-481.

Kobayashi, Y., N. Harada, Y. Nishimura, T. Saito, M. Nakamura, T. Fujiwara, T. Kuroiwa and O. Misumi (2014). "Algae sense exact temperatures: small heat shock proteins are expressed at the survival threshold temperature in *Cyanidioschyzon merolae* and *Chlamydomonas reinhardtii*." *Genome Biol Evol* **6**(10): 2731-2740.

Lin, M. T., A. Occhialini, P. J. Andralojc, M. A. Parry and M. R. Hanson (2014). "A faster Rubisco with potential to increase photosynthesis in crops." *Nature* **513**(7519): 547-550.

Loganathan, N., Y.-C. C. Tsai and O. Mueller-Cajar (2016). "Characterization of the heterooligomeric red-type rubisco activase from red algae." *Proceedings of the National Academy of Sciences* **113**(49): 14019-14024.

Matsuzaki, M., O. Misumi, I. T. Shin, S. Maruyama, M. Takahara, S. Y. Miyagishima, T. Mori, K. Nishida, F. Yagisawa, K. Nishida, Y. Yoshida, Y. Nishimura, S. Nakao, T. Kobayashi, Y. Momoyama, T. Higashiyama, A. Minoda, M. Sano, H. Nomoto, K. Oishi, H. Hayashi, F. Ohta, S. Nishizaka, S. Haga, S. Miura, T. Morishita, Y. Kabeya, K. Terasawa, Y. Suzuki, Y. Ishii, S. Asakawa, H. Takano, N. Ohta, H. Kuroiwa, K. Tanaka, N. Shimizu, S. Sugano, N. Sato, H. Nozaki, N. Ogasawara, Y. Kohara and T. Kuroiwa (2004). "Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D." *Nature* **428**(6983): 653-657.

Meijer, W. G., A. C. Arnberg, H. G. Enequist, P. Terpstra, M. E. Lidstrom and L. Dijkhuizen (1991). "Identification and organization of carbon dioxide fixation genes in *Xanthobacter flavus* H4-14." *Mol Gen Genet* **225**(2): 320-330.

Michelet, L., M. Zaffagnini, S. Morisse, F. Sparla, M. E. Perez-Perez, F. Francia, A. Danon, C. H. Marchand, S. Fermani, P. Trost and S. D. Lemaire (2013). "Redox regulation of the Calvin-Benson cycle: something old, something new." *Front Plant Sci* **4**: 470.

Mueller-Cajar, O., M. Stotz, P. Wendler, F. U. Hartl, A. Bracher and M. Hayer-Hartl (2011). "Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase." *Nature* **479**(7372): 194-199.

Murchie, E. H., M. Pinto and P. Horton (2009). "Agriculture and the new challenges for photosynthesis research." *New Phytol* **181**(3): 532-552.

Nelson, N. and A. Ben-Shem (2004). "The complex architecture of oxygenic photosynthesis." *Nat Rev Mol Cell Biol* **5**(12): 971-982.

Ohta, N., M. Matsuzaki, O. Misumi, S. Y. Miyagishima, H. Nozaki, K. Tanaka, I. T. Shin, Y. Kohara and T. Kuroiwa (2003). "Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*." *DNA Res* **10**(2): 67-77.

Park, S. W., E. H. Hwang, H. S. Jang, J. H. Lee, B. S. Kang, J. I. Oh and Y. M. Kim (2009). "Presence of duplicate genes encoding a phylogenetically new subgroup of form I ribulose 1,5-bisphosphate carboxylase/oxygenase in *Mycobacterium* sp. strain JC1 DSM 3803." Res Microbiol **160**(2): 159-165.

Raines, C. A. (2003). "The Calvin cycle revisited." Photosynth Res **75**(1): 1-10.

Robinson, S. P. and A. R. Portis, Jr. (1989). "Adenosine triphosphate hydrolysis by purified rubisco activase." Arch Biochem Biophys **268**(1): 93-99.

Saito, Y., H. Ashida, T. Sakiyama, N. T. de Marsac, A. Danchin, A. Sekowska and A. Yokota (2009). "Structural and functional similarities between a ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO)-like protein from *Bacillus subtilis* and photosynthetic RuBisCO." J Biol Chem **284**(19): 13256-13264.

Singh, J. and F. R. Tabita (2010). "Roles of RubisCO and the RubisCO-like protein in 5-methylthioadenosine metabolism in the Nonsulfur purple bacterium *Rhodospirillum rubrum*." J Bacteriol **192**(5): 1324-1331.

Somerville, C. R., A. R. Portis and W. L. Ogren (1982). "A Mutant of *Arabidopsis thaliana* Which Lacks Activation of RuBP Carboxylase In Vivo." Plant Physiol **70**(2): 381-387.

Stotz, M., O. Mueller-Cajar, S. Ciniawsky, P. Wendler, F. U. Hartl, A. Bracher and M. Hayer-Hartl (2011). "Structure of green-type Rubisco activase from tobacco." Nat Struct Mol Biol **18**(12): 1366-1370.

Tabita, F. R., T. E. Hanson, H. Li, S. Satagopan, J. Singh and S. Chan (2007). "Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs." Microbiol Mol Biol Rev **71**(4): 576-599.

Tabita, F. R., S. Satagopan, T. E. Hanson, N. E. Kreel and S. S. Scott (2008). "Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships." J Exp Bot **59**(7): 1515-1524.

Tcherkez, G. (2015). "The mechanism of Rubisco-catalysed oxygenation." Plant Cell Environ.

Zhu, X. G., S. P. Long and D. R. Ort (2010). "Improving photosynthetic efficiency for greater yield." Annu Rev Plant Biol **61**: 235-261.