Summer 8-15-2016

Use of Proteomics to Probe Dynamic Changes in Cyanobacteria

Amelia Yen Nguyen
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Plant and Microbial Biosciences

Dissertation Examination Committee:
Himadri B. Pakrasi, Chair
Robert E. Blankenship
Joseph M. Jez
Thomas J. Smith
Richard D. Vierstra

Use of Proteomics to Probe Dynamic Changes in Cyanobacteria

by

Amelia Yen Nguyen

A dissertation presented to
the Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2016
St. Louis, Missouri

Table of Contents

List of Figures viii
List of Tables xi
Acknowledgments xiv
Abstract xviii
Chapter 1 Protective Mechanisms in Cyanobacteria 1
1.1 Cyanobacteria 2
1.2 Oxygenic Photosynthesis 2
1.2.1 Photosynthetic Electron Transport Chain 3
1.2.2 Morphologies 3
1.2.3 Habitats 4
1.2.4 Photo-oxidative Damage and Photoprotection Mechanisms 4
1.3 Redox Regulation 5
1.3.1 Protein Redox Modifications 5
1.3.2 Thioredoxin 6
1.4 Phycobilisome Complexes 7
1.4.1 Phycobilisome Synthesis 8
1.4.2 Phycobilisome Degradation 8
1.5 Mass Spectrometry 10
1.6 This Work 11
1.7 References 15
Chapter 2 Proteome-wide Light/Dark Modulation of Thiol Oxidation in Cyanobacteria Revealed by Quantitative Site-Specific Redox Proteomics 19
2.1 Introduction 20
2.1.1 Abstract 20
2.1.2 Redox Biology 20
2.1.3 Thiol Redox Dynamics in Cyanobacteria 21
2.1.4 *In Vivo* Thiol Quantification 22
2.2 Materials and Methods 24
2.2.1 Reagents 24
2.2.2 Cyanobacteria Culture Conditions 24

2.2.3 Sample Preparation 27
2.2.4 Selective Enrichment of Oxidized or Total Cys-peptides 28
2.2.5 On-resin TMT labeling 28
2.2.6 SDS-polyacrylamide Gel Electrophoresis 29
2.2.7 LC-MS/MS Analysis 29
2.2.8 Data Analysis 30
2.2.9 Strains and Culture Preparation for Recombinant Proteins 32
2.2.10 Site-specific Mutagenesis and Protein Expression Vector Construction 33
2.2.11 Expression and Purification of Recombinant G6PDH, Gap2, Sll1621 34
2.2.12 Enzymatic Activity Measurement 35
2.3 Results 36
2.3.1 Quantitative Site-specific Measurements of Reversible Thiol Oxidation 36
2.3.2 Broad Redox Changes on Cys-sites Modulated by Light/Dark 38
2.3.3 Percentage of Reversible Oxidation of Individual Cys-sites 39
2.3.4 Total Protein Abundances 41
2.3.5 Functional Implications of Redox-sensitive Proteins 41
2.3.6 Site-specific Redox Sensitivity Predicts Functional Sites 45
2.3.7 Novel Functional Cys-sites in Redox Regulatory Proteins 47
2.4 Discussion 49
2.5 Acknowledgements 52
2.5.1 Author Contributions 53
2.5.2 Conflict of Interest Statement 53
2.6 References 54
Chapter 3 The Proteolysis Adaptor, NblA, Binds to the N-terminus of β-Phycocyanin: Implications in the Mechanism of Phycobilisome Degradation 76
3.1 Introduction 77
3.1.1 Abstract 77
3.1.2 Phycobilisome Function 78
3.1.3 Phycobilisome Structure 78
3.1.4 Dynamic Remodeling of Phycobilisomes 79
3.1.5 Binding Partners of NblA 82
3.2 Materials and Methods 82
3.2.1 Bacterial Strains and Growth Conditions 82

3.2.2 Overexpression and Purification of $NblA_{2973}$_His_7 Fusion Protein in *E. coli* ... 82
3.2.3 Isolation of Intact PBS ... 84
3.2.4 *In Vitro* Binding Assays ... 84
3.2.5 Chemical Cross-linking ... 84
3.2.6 Sample Digestion, LC-MS/MS, and Data Processing ... 85
3.2.7 Protein Modeling ... 86
3.2.8 Global Protein Docking ... 87
3.2.9 Local Protein Docking ... 88
3.3 Results ... 88
3.3.1 Overexpression and Purification of Fusion Protein ... 88
3.3.2 *In Vitro* Binding Assays and BS^3 Cross-linking ... 89
3.3.3 Mass Spectrometry ... 89
3.3.4 Modeling ... 91
3.4 Discussion ... 91
3.5 Acknowledgements ... 94
3.5.1 Author Contributions ... 95
3.5.2 Conflict of Interest Statement ... 95
3.6 References ... 96
Chapter 4 Conclusions and Future Directions ... 112
4.1 Introduction ... 113
4.2 Redox Regulation ... 113
4.2.1 Reversible Protein Thiol Oxidations ... 114
4.2.2 Quantification of Redox Protein Thiols ... 115
4.2.3 Broad Redox Regulation ... 116
4.3 PBS Structure and Function ... 117
4.3.1 NblA Binding ... 117
4.3.2 PBS Degradation Mechanism ... 118
4.4 Implications ... 119
4.5 Future Directions ... 120
4.5.1 Redox Signaling Network Map ... 120
4.5.2 Protein Proteolysis ... 121
4.5.2 Application to Industry Needs ... 122

4.6 References 123

Appendix A Structural Basis and Evolution of Regulation in Plant Adenosine-5'-Phosphosulfate Kinase 127

A.1 Introduction 128

A.1.1 Abstract 128

A.1.2 Sulfur Metabolism 128

A.2 Materials and Methods 132

A.2.1 Characterization of APSK from *Synechocystis* 6803 132

A.2.2 Crystallography 133

A.2.3 Mutagenesis 133

A.2.4 Enzyme Assays and Redox Titrations 134

A.2.5 Cloning, Expression, Purification, and Analysis of APSK from *Synechocystis* 6803 134

A.3 Results 135

A.3.1 Overall Structure 135

A.3.2 Structure of the Active Site and N-terminal Loop 136

A.3.3 The Cys86-Cys119 Disulfide and Redox-regulation 138

A.3.4 Characterization of APSK from *Synechocystis* 6803 139

A.4 Discussion 140

A.5 Acknowledgements 144

A.5.1 Author Contributions 145

A.5.2 Conflict of Interest Statement 145

A.6 References 146

Appendix B Global Proteomic Analysis Reveals an Exclusive Role of Thylakoid Membranes in Bioenergetics of a Model Cyanobacterium 159

B.1 Introduction 160

B.1.1 Abstract 160

B.1.2 Plasma and Thylakoid Membranes 161

B.2 Materials and Methods 163

B.2.1 Cell Growth and Sample Preparation 163

B.2.2 iTRAQ Labeling and HPLC Fractionation 163

B.2.3 LC-MS/MS Analysis 164

B.2.4 Data Analysis 165

B.2.5 Experimental Design and Statistical Rationale 166

B.2.6 Computational Web-based Tools 166
B.3 Results 166
B.3.1 Identification of PM and TM Proteins 166
B.3.2 Characteristics of Identified Proteins 169
B.3.3 Photosystems and Respiratory Proteins 171
B.3.4 Pigment Biosynthesis and Transport Proteins 172
B.3.5 Proteins Involved in Other Important Cellular Processes 173
B.3.6 Hypothetical and Unknown Proteins 174
B.4 Discussion 175
B.5 Acknowledgements 181
B.5.1 Author Contributions 181
B.5.2 Conflict of Interest Statement 181
Authors declare no competing conflicts of interest. 181
B.6 References 182
Appendix C Non-*tenera* Contamination and the Economic Impact of *SHELL* Genetic Testing in the Malaysian Independent Oil Palm Industry 199
C.1 Introduction 200
C.1.1 Abstract 200
C.1.2 Oil Palm 201
C.2 Material and Methods 203
C.2.1 Sampling of Independent Planting Sites and Nurseries 203
C.2.2 PCR-based Genetic Testing 204
C.2.3 DNA Sequencing 205
C.2.4 Genotype/Phenotype Comparisons 206
C.2.5 Economic Impact Modeling of Comprehensive *SHELL* Genetic Testing 206
C.3 Results 209
C.3.1 Non-*tenera* Contamination Trial Design 209
C.3.2 Non-*tenera* Contamination 209
C.3.2 Identification of Novel Mutant Alleles of *SHELL* 210
C.3.4 Economic Impact Modeling of Comprehensive *SHELL* Genetic Testing 214
C.4 Discussion 215
C.5 Acknowledgements 217
C.5.1 Authors Contribution 218

C.5.2 Conflict of Interest Statement 218
C.6 Reference 219
Appendix D Population Level Coordination of Pigment Response In Individual Cyanobacterial Cells Under Altered Nitrogen Levels 233
D.1 Introduction 234
D.1.1 Abstract 234
D.1.2 Introduction 234
D.2 Material and Methods 238
D.2.1 Cyanobacterial Strains and Culture Conditions 238
D.2.2 Single Cell HCFM Microscopy 239
D.2.3 Spectral Image Analysis 239
D.3 Results 240
D.3.1 Pigment Response to Nitrogen Depravation in *Synechocystis* 6803 Cultures 240
D.3.2 Single Cell Analysis 241
D.4 Discussion 241
D.5 Acknowledgements 241
D.5.1 Author Contributions 242
D.5.1 Conflict of Interest Statement 242
D.6 References 243

List of Figures

Chapter Figures:

Figure 2.1: Overview of the enrichment and site-specific quantification strategy for thiol oxidation. 66

Figure 2.2: Broad Cys redox changes modulated by light/dark..... 68

Figure 2.3: SDS-PAGE gel images of oxidized Cys-peptides..... 69

Figure 2.4: Functional categories of the significant oxidized proteins in the dark and DCMU conditions. 70

Figure 2.5: Broad redox regulations in cyanobacteria. 71

Figure 2.6: Comparison of protein thiol oxdiations levels of individual proteins from different biological processes in light, dark, and DCMU conditions. 72

Figure 2.7: Site-specific redox sensitivity and functional Cys-sites. 73

Figure 2.8: Functional sites of peroxiredoxin (Sll1621) and G6PDH (Zwf). 74

Figure 2.9: Cys-155 in Sll1621 is conserved across many cyanobacterial and other bacterial strains as shown in the 'DNCP' motif. 75

Figure 3.1: Plasmid map of the overexpression of $NblA_{2973}$_His_7 strain. 101

Figure 3.2: Induction, purification, and LC-MS/MS analysis of $NblA_{2973}$_His_7. 102

Figure 3.3: *In vitro* binding and cross-linking of NblA and phycobiliproteins. 103

Figure 3.4: Docking results of an NblA dimer globally docked to α-PC/β-PC monomer for orientation I. 104

Figure 3.5: Docking results of an NblA dimer globally docked to α-PC/β-PC monomer for orientation II..... 106

Figure 3.6: C-terminus of $NblA_{2973}$_His_7 binding to the N-terminus of β-PC. 108

Figure 3.7: C-terminus of $NblA_{2973}$_His_7 binding to the N-terminus of NblA. 109

Figure 3.8: Molecular Docking of NblA Dimer to PC. 110

Appendix Figures:

Figure A.1: Sulfur assimilatory pathways. 153

Figure A.2: Sequence alignment of representative APSK. 154

Figure A.3: Structure of AtAPSK. 155

Figure A.4: Stereoview of the AtAPSK active site. 156

Figure A.5: Functional analysis of redox activity. 157

Figure A.6: Velocity curves of *v* versus APS concentration. 158

Figure B..1: Distribution of ppm mass accuracy for all identified spectra. 192

Figure B..2: Heat map of differentially abundant membrane proteins. 193

Figure B.3: Reverse volcano plot of all quantified protein identifications comparing PM and TM results. 194

Figure B.4: Functional categorization distribution of proteins between PM and TM analyzed preparations. 195

Figure B.5: Predicted topology of identified a TM and b PM proteins in *Synechocystis* 6803.. 196

Figure B.6: Schematic drawing showing distribution of functional roles between PM and TM. 197

Figure B.7: Comparative analysis of membrane proteins. 198

Figure C.1: Non-*tenera* contamination rates within each individual sampled site. 228

Figure C.2: *SHELL* MADS box domain mutations associated with fruit form phenotype. 229

Figure C.3: Fruit form phenotypes 230

Figure C.4: Economic impact modeling of comprehensive *SHELL* genetic testing 231

Figure C.5: Annual economic impact of *SHELL* genetic testing on the oil palm industry, gross national income and the Malaysian government. 232

Figure D.1: Hyperspectral work-flow and pigments measurements. 246

Figure D..2: Changes in the concentration of pigments at different time points under varying nitrogen conditions 247

Figure D.3: Multivariate curve resolution. 248

Figure D.4: Average per cell fluorescence. 249

Figure D.5: Single cell scatter plots for comparing abundances of photosynthetic pigments.... 250

List of Tables

Chapter Tables:

Table 2.1: Oligonucleotides used for site-specific mutagenesis of selected cysteine sites in G6PDH, Gap2 and Sll1621 proteins. 61

Table 2.2: Total identified Cys-sites along with peptide sequences. 63

Table 2.3: Cys-sites with significant redox changes (abundance ratios >1.50, percentage of oxidation >20%). 63

Table 2.4: Total protein abundances from global proteome profiling. 63

Table 2.5: Redox-regulated transcriptional regulators or signaling proteins. 63

Table 2.6: Previously identified protein targets for *Synechocystis* thioredoxin A or glutaredoxin 63

Table 2.7: The levels of relative Cys oxidation on individual Cys-sites for selected proteins with known active or functional Cys-sites. 64

Table 3.1: Strains, plasmids, and oligonucleotides. 100

Appendix Tables:

Table A.1: Oligonucleotides used for generation of strains. 150

Table A.2: Summary of crystallographic statistics. 151

Table A.3: Steady-state kinetic parameters. 152

Table B.1: Identified peptide and proteins details. 185

Table B.2: List and analysis of total proteins detected and quantified. 185

Table B.3: Distribution of TM integral proteins with Log_2 ratio of -2 or less across different pathways/subsystems. 186

Table B.4: Distribution of PM integral proteins with Log_2 ratio of 2 or more across different pathways/subsystems. 189

Table C.1: PCR primers and conditions. 221

Table C.2: Parameters for Economic Impact Model. 224

Table C.3: Non-*tenera* contamination by region. 225

Table C.4: Novel *SHELL* alleles. 226

Table C.5: Concordance of *SHELL* genotype and fruit form phenotype. 227

Table D.1: Average % of original abundance of photosynthetic pigments. 245

Abbreviations

APC	allophycocyanin
ATP	adenosine triphosphate
Chl	chlorophyll
Cyt	cytochrome
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	histidine tag
LC-MS/MS	liquid chromatography tandem mass spectrometry
MS	mass spectrometry
NADP(H)	nicotinamide adenine dinucleotide phosphate
NblA	non-bleaching protein A
Ni-NTA	nickel-nitrilotriacetic acid
OD	optical density
OEC	oxygen evolving complex
ORF	open reading frame
PBS	phycobilisome
PC	phycocyanin
PCR	polymerase chain reaction
PSI and PSII	photosystem I and II, respectively
ROS	reactive oxygen species
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Synechococcus 2973	*Synechococcus* sp. UTEX 2973
Synechocystis 6803	*Synechocystis* sp. PCC 6803
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol

Acknowledgments

I would like to thank my family and friends for all their love and support. I am forever grateful to them for teaching me that the meaning of "family" is not defined by blood. Because of them, I learned how to treat all men and women equally; and most importantly, to love everyone regardless of our differences. They are my motivation to live each day by seeing, hearing, and feeling with kindness.

In addition, I am thankful to my AP Chemistry teacher, Mr. Matthew Cain, for kindling my life long fascination about science. I received an internship in Dr. David Fruman's cancer laboratory at UC Irvine the summer before my senior year in high school because Mr. Cain was willing to vouch for me. I also managed to put together a winning application to earn the Gates Millennium Scholarship, a 10-year full ride scholarship to attend college and graduate school, all thanks to the kind letter of recommendation he wrote for me. I am forever indebted to Mr. Cain for helping me take advantage of these life-changing opportunities.

At UC Berkeley, I learned that scientific research did not necessarily require me to euthanize mice. My first plant biology class taught by Drs. Norman Terry and Anastasios Melis introduced me to the intricacies and quirks of the plants that we see every day; which helped me appreciate the boundless potential plant biochemistry has for alleviating global environmental issues. In my excitement to apply my scientific knowledge, I applied for a research position in Dr. Norman Terry's laboratory. When his manager informed me that a more senior student was given the job, I remained undaunted and personally expressed to Dr. Norman Terry my sincere interest in his research. The next week, I landed myself a project as an undergraduate researcher in his laboratory, where I worked for the rest of college. Thanks to Drs. Terry and Melis, I

learned the hypothesis driven scientific method and cultivated my curiosity to solve mysteries that could advance the world's body of knowledge.

Having experienced the scientific environment in Berkeley, I wanted to expose myself to a new scientific community during graduate school. The vast concentration of plant biotechnological research in St. Louis led me to Washington University. Presently, I am honored to be a National Science Foundation Graduate Research Fellow in Dr. Himadri Pakrasi's laboratory. I would like to thank Drs. Himadri Pakrasi, Joseph Jez, Robert Blankenship, Thomas Smith, and Richard Vierstra for serving on my thesis committee and offering numerous insightful comments that helped shape my dissertation. Under their guidance, I am able to explore my academic and non-academic interests with little apprehension of failure. This freedom to pursue my dreams helped me accomplish my personal goals. Just as importantly, I would like to thank all the past and present members of the Pakrasi, Jez, and Blankenship labs for the countless memorable discussions and experiences that we shared together. Our awesome friendships made any lab setbacks bearable and I will never forget you all.

My journey from working in a cancer laboratory to current interest in photosynthesis fuels my drive to solve challenging questions. Over the years, I realized that to solve world problems, contemporary scientists should not focus on their research alone. An emphasis must be placed on interdisciplinary communication to train and inspire the next generation of leaders. I actively sought out opportunities to learn about the world outside of my research. I found my sense of belonging in graduate school after joining BALSA, a student-led consulting organization. Thanks to the BALSA community, I personally witnessed the countless ways scientists could create broad impact in this world through both academic and non-academic professions.

Last but not least, I would like to thank all the funding agencies for their financial support; which made my research possible. This work was supported by the National Science Foundation Graduate Research Fellowship Program (Grant DGE-1143954), Chemical, Sciences, Geosciences, and Biosciences Division of the Basic Energy Sciences program at the U.S. Department of Energy (Grant DE-FG02-99ER20350), and the Photosynthetic Antenna Research Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Basic Energy Sciences (Grant DE-SC0001035).

Amelia Yen Nguyen

Washington University in St. Louis

August 2016

Dedicated to all my mentors.

ABSTRACT OF THE DISSERTATION

Use of Proteomics to Probe Dynamic Changes in Cyanobacteria

for Arts & Sciences Graduate Students

by

Amelia Yen Nguyen

Doctor of Philosophy in Biology and Biomedical Sciences

Plant and Microbial Biosciences

Washington University in St. Louis, 2016

Professor Himadri B. Pakrasi, Chair

Cyanobacteria are unicellular photosynthetic microorganisms that capture and convert light energy to chemical energy, which is the precursor for feed, fuel, and food. These oxygenic phototrophs appear blue-green in color due to the blue bilin pigments in their phycobilisomes and green chlorophyll pigments in their photosystems. They also have diverse morphologies, and thrive in terrestrial, marine water, fresh water, as well as extreme environments. Cyanobacteria have developed a number of protective mechanisms and adaptive responses that allow the photosynthetic process to operate optimally under diverse and extreme conditions. Prolonged deprivation of essential nutrients, such as nitrogen and sulfur, commonly found in the natural environments cyanobacteria grow in, can disrupt crucial metabolic activities and promote the production of lethal reactive oxygen species. The dynamic remodeling of protein complexes and structures facilitates adaptation to environmental stresses, however, specific protein modifications are poorly understood. Synthetic and systems biology approaches have been used to study how photosynthetic microorganisms optimize their cellular metabolism in response to adverse environmental conditions. To gain insights on how cyanobacteria cope with environmental changes, we created a global proteomics map of redox-sensitive amino acid

residues and examined the degradation of light harvesting apparatus in cyanobacteria. These studies offered significant insights into the broad redox regulation and protein degradation, advancing knowledge of how photosynthetic microbial cells dynamically rely on protective mechanisms to survive changing environmental conditions.

Chapter 1

Protective Mechanisms in Cyanobacteria

1.1 Cyanobacteria

Known as the evolutionary precursors of chloroplasts, cyanobacteria are responsible for oxygenating the Earth 2.4 billion years ago and are major drivers in the global carbon and nitrogen cycles (Buick 2008). Cyanobacteria are oxygenic photosynthetic prokaryotes that generate chemical energy in the form of carbohydrates, which are the precursors for food, feed, fuel, and fiber. In addition, some diazotrophic cyanobacteria can fix atmospheric nitrogen into more accessible forms like nitrate and ammonia (Herrero et al. 2001). The integral protein machineries in oxygenic photosynthesis are the light harvesting phycobilisome (PBS) complex and the reaction centers of Photosystem I (PSI) and Photosystem II (PSII). Due to the brilliantly colored blue bilin pigments in the PBS complexes and green chlorophyll pigments in the photosystems, cyanobacteria were previously referred to as blue-green algae. The PBS complexes are the massive light harvesting antennas in cyanobacteria that harvest light energy and funnel the energy down towards both PSII and PSI to initiate photosynthesis (Bogorad 1975; Liu et al. 2013). PSI and PSII are two thylakoid membrane bound pigment-protein complexes that carry out photosynthesis in a series of light-dependent electron transfer reactions (Pakrasi 1995).

1.2 Oxygenic Photosynthesis

Light is the driving force of oxygenic photosynthesis, a process that oxidizes water molecules into O_2, protons, and electrons and converts CO_2 into desirable chemical products (Buchanan 1991). The splitting of water molecules occurs in the oxygen-evolving complex (OEC) called the manganese cluster (i.e. Mn_4CaO_5) that is located centrally in the PSII. The protons create a proton gradient and the electrons are passed along the photosynthetic electron transport (PET) chain to produce ATP and NADPH, which are energy currencies for the

biosynthesis of energy rich sugars and lipids. Oxygenic photosynthesis occurs in a specialized membrane system call the thylakoid found in photosynthetic organisms like plants algae, and most cyanobacteria.

1.2.1 Photosynthetic Electron Transport Chain

The pigment-protein complexes and redox intermediates in the thylakoid membrane create a PET chain that play a crucial role in the reduction of soluble electron carriers, e.g. NADP(H), and the formation of a trans-thylakoid electrochemical gradient that generates ATP. These processes are known to modify the redox environment in photosynthetic organisms. Alterations of the redox environment occur because photosynthesis operates in the presence of strong oxidants and reductants that lead to the production of potentially destructive chemically reactive species (Shulaev and Oliver 2006; Singh et al. 2010).

1.2.2 Morphologies

In addition to converting carbons to sugars during photosynthesis, cyanobacteria can convert sugars to carbons during respiration too. The photosynthetic and respiratory machineries are both located in the internal thylakoid membrane. The electron transport chains for both processes share several redox active protein components, i.e. the plastoquinone pool, cytochrome b_6f complex, plastocyanin, and cytochrome c_{553} (Vermaas 2001). In addition, as gram-negative bacteria (Hoiczyk and Hansel 2000), all cyanobacteria commonly have an outer membrane, peptidoglycan layer, and plasma membrane, but grow with varied spherical, rod-shaped, or filamentous morphology. Intriguingly, some cyanobacteria can cultivate symbioses with plants, fungi, sponges, and protists (Adams and Duggan 2008).

1.2.3 Habitats

The metabolic flexibility of cyanobacteria allows them to thrive in a variety of environments, from terrestrial to marine conditions. Cyanobacteria are often exposed to fluctuating nutrient and light levels that can range from limiting to in excess. Several protective mechanisms and adaptive responses help cyanobacteria cope with fluctuating environmental conditions. For example, redox regulation and PBS remodeling allow cyanobacteria to exploit environmental conditions and withstand oxidative and nutrient stresses.

1.2.4 Photo-oxidative Damage and Photoprotection Mechanisms

Countless heterotrophic organisms generate numerous chemically reactive species during photosynthesis, respiration, and/or fatty acid oxidation. The presence of strong oxidants and reductants in cyanobacteria inevitably lead to the production of chemically reactive species. Cells undergo oxidative stress in the prolonged presence of (i) reactive oxygen species (ROS), e.g. hydroxyl, peroxide, and superoxide anion radicals, or (ii) mixed reactive nitrogen–oxygen species (RNS), e.g. nitric oxide (NO) and peroxynitrite (ONOO-) (Dalle-Donne et al. 2009). To minimize photo-oxidative damage, photosynthetic organisms have developed a number of protective mechanisms to operate optimally under adverse conditions. For instance, thermal dissipation, antenna size adjustments, state transitions, photo-respiration, water-water cycle, antioxidant systems, and repair & synthesis are employed to combat photo-oxidative damage (Niyogi 1999; Gilmore 1997; Horton et al. 1996; Noctor and Foyer 1998; Demmig-Adams and Adams 1996). How proteins are modified in response to the production of ROS during photosynthesis and respiration remains unclear for some organisms, but recent studies in various biological systems demonstrated that proteins rely on reversible ROS-mediated post-translational modifications to regulate their functions (Rinalducci et al. 2008; Winterbourn and Hampton 2008; Fratelli et al. 2002; Brandes et al. 2011).

1.3 Redox Regulation

Cyanobacteria have developed elaborate mechanisms to prevent excessive absorption of light energy. Although ROS and RNS may form at all light intensities, they become destructive when the amount of light energy available exceeds what could be used for phytochemical conversion of light to chemical energy. Interestingly, these chemically reactive species may play important roles in signaling and maintenance of redox homeostasis in both non-stressed and oxidative stressed conditions. The redox states of proteins in the PET chain depend on the redox environment in cells. Modifications of protein thiols in cysteine (Cys) residues involved in the PET chain regulate gene transcription, stability of PSI and PSII, and post-translational modifications (PTMs) of proteins in cyanobacteria, green algae, and plants (Pfannschmidt et al. 2009; Hihara et al. 2003; Lopez-Maury et al. 2009; Gill et al. 2002). Redox regulation is one mechanism that helps cyanobacteria to withstand oxidative stresses, and is an emerging paradigm in which protein thiols undergo reversible PTMs that affect protein functions.

1.3.1 Protein Redox Modifications

Oxygen evolution in cyanobacteria and plant promotes significant PTMs of the proteins present in the thylakoid membranes. The thylakoid membrane in cyanobacteria is one of the most redox active compartments in biological systems since oxygen is constantly evolved during photosynthesis. However, little is known about the redox proteomics in cyanobacteria. The thiol side chain in Cys often participates in enzymatic reactions and may become inactivated when oxidized, and activated when reduced (Buchanan and Balmer 2005). Additionally, many metabolic pathways in plastids are regulated by the conversion of protein thiols from dithiol (reduced Cys) to disulfide (oxidized Cys) (Navrot et al. 2011). The thiol side chain in Cys, under steady-state or oxidative stresses conditions, can be modified in various forms, including

reversible forms: disulfides, S-nitrosylation (SNO), S-glutathionylation (SSG), sulfenylation (SOH), and irreversible forms: sulfinylation (SO_2H), and sulfonylation (SO_3H) (Bachi et al. 2013; Held and Gibson 2012).

Elucidation of the redox proteome in cyanobacteria under different redox stress conditions will undoubtedly reveal new insights into the functions of proteins that protect photosynthetic organisms under unfavorable conditions. But detection of the dynamic redox states of specific Cys in proteins (as cells are exposed to different environmental stresses) have been limited due to 1) the transient nature of PTMs of thiols and 2) technical challenges associated with site-specific detection of Cys modifications.

Understanding the dynamic redox changes of specific protein thiols on a global scale should provide a comprehensive picture of the key players that protect the photosynthetic apparatus from oxidative stress. The finding that an increasing number of proteins undergo reversible ROS-mediated thiol modifications instead of photo damage in various biological systems supports the rising concept of redox regulation (Fratelli et al. 2002). The extent to which Cys residues are modified depend on the amounts and types of ROS present, the reactivity of the Cys thiols involved, and the kinetic and thermodynamic stability of the respective thiol modifications (Winterbourn and Hampton 2008).

1.3.2 Thioredoxin

Thiols play a central role in coordinating the antioxidant defense network by acting as reversible redox switches. Previous studies suggest that the oxidative PTMs of protein thiols induce redox regulation in cells by switching protein functions on or off. After light activates photosynthesis, electrons flow through the PET chain and reduce critical enzymes like ferredoxin, ferrodoxin-thioedoxin reductase, and thioredoxin (Trx) (Buchanan 1991). Trx has

two Cys in its active site that form a disulfide bond when Trx is inactive. The reduced state of Trx is responsible for the reduction of thiols, which then modulates the activation or inactivation of enzymes linked to PSI and PSII (Buchanan and Balmer 2005), and various other redox-regulated target processes in plants and cyanobacteria (Lindahl and Florencio 2003). In short, the redox states of protein thiols are crucial determinants of multiple metabolic, signaling, and transcriptional processes in cells (Dalle-Donne et al. 2009).

Traditionally, Trx affinity chromatography had been used to identify proteins that are regulated by Trx. The method consists of thioredoxin that is immobilized and mutated at one Cys residue in the active site. Proteins that naturally interact with Trx will form one disulfide bond with Trx, and be trapped to Trx because they cannot form a second disulfide bond to be fully oxidized. The trapped proteins can be identified by first being eluted with DTT, run on a protein gel, and sequenced. The identified proteins are redox sensitive since they interact with Trx, an enzyme that plays a major role in regulating the redox states of proteins. Using this method, less than 100 proteins were confirmed as Trx-regulated before our current study (Lindahl and Florencio 2003; Hosoya-Matsuda et al. 2005; Mata-Cabana et al. 2007). This biochemical approach lost its appeal after scientists refined mass spectrometry (MS) experiments to increase levels of sensitivity and detection of site-specific modifications.

1.4 Phycobilisome Complexes

In cyanobacteria, light harvesting is accomplished by a combination of membrane intrinsic proteins, such as CP43, CP47 (Zak et al. 2001), and IsiA (Bibby et al. 2001), and large extrinsic PBS complexes (Grossman et al. 1993; Bogorad 1975). The primary photosynthetic antenna systems in prokaryotic cyanobacteria, eukaryotic red algae, and cyanelles are PBS complexes that range from 3-7 Mega Daltons in size (Adir et al. 2006). The PBS complexes

function in the process of harvesting sunlight for conversion into cellular fuel. PBS complexes initiate photosynthesis by absorbing light energy and funneling it to the reaction centers in photosystems. PBS may account up to 50% of the total soluble proteins in the cell, thus serving as a large cellular nitrogen reserve (Bogorad 1975). These principal light-harvesting antennas in cyanobacteria are composed of phycobiliproteins, but little is known how redox signaling regulates the functions of such proteins to help cyanobacteria cope with light stress (Campbell et al. 1998).

1.4.1 Phycobilisome Synthesis

PBS complexes are composed of pigment-proteins called phycobiliproteins. The Cys residues in phycobiliproteins are covalently bonded to linear tetrapyrroles. The ordered assembly of PBS begins with the nucleation of the α- and β-phycocyanin (PC) subunits to form a PC monomer that serves as the building block for the PBS complex. Six PC monomers form a PC hexamer. With the help of color-less linker proteins, usually three PC hexamers are linked together to form a PC rod. Ultimately, six peripheral PC rods radiate from the allophycocyanin (APC) core in the PBS in a fan shaped manner (Ajlani and Vernotte 1998). More than a hundred monomers of α- and β-PC make up one PBS complexe, highlighting the enormous amount of phycobiliproteins that must come together to make these massive light apparatuses. Some cyanobacteria also have phycoerythrin (PE), another type of phycobiliproteins that facilitates the harvesting of light energy at lower wavelength compared to the other types of phycobiliproteins.

1.4.2 Phycobilisome Degradation

PBS complexes play a key role in allowing cyanobacteria to cope with light and nutrient stresses. PBS complexes serve as a large nutrient reserve due to the massive number of phycobiliproteins needed to create these complexes. The dynamic remodeling of PBS complexes

provides a mechanism for cyanobacteria to fine-tune their responses to environmental stress. During nitrogen, sulfur, or phosphorous depletion, the degradation of PBS complexes provide building blocks that are essential for the maintenance of crucial metabolic activities and as a form of photoprotection (Adir et al. 2006; Adir et al. 2003). As cellular metabolism slows down during nutrient depletion, the amount of light energy harvested needs to decrease to prevent ROS from being created.

The dynamic remodeling of PBS occurs on a rapid and massive scale in, enabling cyanobacteria to acclimate to changing environmental conditions. This plasticity is best seen during nutrient deprivation, e.g. nitrogen or sulfur deplete condition, which causes cells to bleach (Collier and Grossman 1994). Upon addition of nutrients, cyanobacteria regain their blue-green color due to the re-synthesis of PBS. PBS degradation plays an important function in cell maintenance, growth, and development. The processes of PBS degradation and re-synthesis play a key role for cell survival. If PBS do not degrade during nutrient deprivation, photosynthetic rates will remain unchanged while cellular metabolism slows down, which will cause excessive energy to be absorbed and production of harmful radical species (Adir et al. 2003). Consequently, PBS degradation plays an important function in cell maintenance, growth, and development.

Degradation of PBS is an active, rapid and specific process that occurs on a massive scale. A popular model of PBS degradation begins with the sequential trimming of the peripheral rods, starting at the most distal PC end, and terminates 1 day after nitrogen or sulfur starvation, with complete degradation of the remaining PBS occurring 2 days after continued nutrient depletion in *Synechococcus* sp. PCC 7942 (hereafter referred to as *Synechococcus* 7942) cells are not supplemented with carbon dioxide (Collier and Grossman 1994). An essential factor for the

ATP-dependent degradation of PBS is Non-bleaching A (NblA), a 7-kDa, 59 amino acids polypeptide (Collier and Grossman 1994). NblA is a protein that triggers the active, specific, and massive degradation of PBS in cyanobacteria. This proteolysis adaptor facilitates the interactions between ClpC and a phycobiliprotein targeted for degradation at its N and C-termini, respectively (Karradt et al. 2008; Stanne et al. 2007). ClpC is a chaperon partner of an ATP-dependent protease found in eubacteria and eukaryotes (Stanne et al. 2007; Andersson et al. 2006). Elucidation of how NblA mechanistically facilitates PBS destruction may reveal general principles that govern the specificity of macromolecular complex degradation (Collier and Grossman 1994).

1.5 Mass Spectrometry

Mass spectrometry (MS) enables examination of the PTMs of proteins in cyanobacteria under stress. Due to enormous advancements in MS during the past two decades, the field of proteomics has exploded with powerful techniques that offer accurate information on the composition, dynamics, and structure of proteins (Weisz et al. 2016). The numerous applications MS offers to science triggered the development of redox proteomics, which investigates how oxidative stress modifies protein structure and function (D'Alessandro et al. 2011; Rinalducci et al. 2008). For example, tools have been developed to detect diverse sets of Cys modifications as a result of the global interest in redox biology. Traditional methods used diagonal electrophoresis or antibodies against Cys oxoforms, but they suffer from moderate sensitivity and limited site-specific identification (McDonagh 2009; Seo and Carroll 2009). Methods based on differential chemical labeling have also been developed to identify Cys modifications and their dynamic changes (Bachi et al. 2013; Lin et al. 2010; Forrester et al. 2007; Jaffrey and Snyder 2001; Jaffrey et al. 2001; Izquierdo-Alvarez and Martinez-Ruiz 2011; Nelson et al. 2010; Leichert et al.

2008; Sethuraman et al. 2004; Liu et al. 2004; Liu et al. 2005; Leonard and Carroll 2011). Our collaborators at the Pacific Northwest National Laboratory (PNNL) and others have developed proteomic approaches to enrich protein/peptide thiols for quantitative and site-specific profiling of Cys-containing peptides, Cys-SNO and Cys-SSG (Liu et al. 2004; Liu et al. 2005; Su et al. 2013).

1.6 This Work

An understanding of how cyanobacteria dynamically process information and self-regulate in response to environmental changes would provide significant insights on how biological platforms can be optimized to address global challenges. Cyanobacteria are excellent model organisms for biological research, namely because they grow fast, are transformable, survive in minimal growth nutrients, and do not require arable land. Genetic modifications of cyanobacteria for strategic production of desirable chemicals have the potential to drive innovation in the energy sector to decrease America's dependence on foreign oil, ameliorate environmental pollution, and increase economic security. However, a major challenge in engineering microbial factories for production of biochemicals is the low yield, which is cost-prohibitive for large-scale production. To optimize biological production platforms, we aimed to analyze the protective role protein modifications play in enabling cyanobacteria sense and adapt to their environments.

Several strategies were used to study the mechanisms and functions of protein remodeling in cyanobacteria. Proteomics research was carried out in *Synechocystis* sp. PCC 6803 (hereafter referred to as *Synechocystis* 6803) and *Synechococcus* sp. UTEX 2973 (hereafter referred to as *Synechococcus* 2973). These two cyanobacterial strains were chosen due to the extensive omics and bioinformatics information available for them, ease of genetic

modifications, and fast growing capabilities. To study how cyanobacteria respond to different redox environments, we examined the global redox states of protein thiols in *Synechocystis* 6803 that was exposed to distinct oxidizing and reducing conditions. Tandem mass isobaric tags were used to streamline high-throughput LC-MS/MS runs. Genetic engineering allowed us to investigate the function of specific Cys residues from our large-scale map of redox-sensitive protein thiols. We verified that mutations in redox-active Cys inactivated protein functions. Given that numerous phycobiliproteins have redox-active protein thiols, molecular biology and biochemistry techniques allowed us to investigate how the PBS complexes are degraded. A combined cross-linking and MS approach allowed us to study the disassembly, binding, and structures of the PBS complexes.

To start, the rapid rise in popularity of relying on MS to analyze the structure and function of proteins prompted the creation of redox proteomics, which aims to identify and quantify redox-based changes within the proteome (Butterfield and Dalle-Donne 2012). In Chapter 2, a detailed redox proteomics global map of *Synechocystis* 6803 was created in collaboration with a team of proteomics specialists at the PNNL. The goal was to quantify the magnitude of redox state changes on protein thiols of all proteins in cyanobacteria. A resin-assisted enrichment of oxidized Cys residues coupled with MS detected, on a global scale, redox-sensitive Cys residues *in vivo* that are switched on and off by redox processes. This data facilitates the discovery of the factors responsible for protecting cells from ROS damage and the broad redox regulation of photosynthetic organisms.

Interestingly, the dramatic redox sensitivity in numerous phycobiliproteins highlights how the PBS complexes are redox active centers. This correlates to the role PBS complexes have in harvesting light energy to trigger a series of redox signaling cascade during photosynthesis.

Interests in phycobiliprotein inactivation/activation led to investigations into PBS degradation during environmental stresses. In Chapter 3, I presented novel insights in how a small soluble proteolysis adaptor, NblA, triggers the degradation of the massive PBS complexes. NblA has been shown to be critical in this degradation process (Collier and Grossman 1994). A combined affinity chromatography, cross-linking, and MS approach facilitated investigations of the binding partners of NblA. These findings revealed insights about the protein-protein interactions NblA has and provided low-resolution structural data about the PBS complexes. With help from our collaborators at MIT, we carried out protein docking studies to suggest how NblA could interact with β-PC and hypothesize that the PBS complex dynamically change conformations during nutrient stress.

In conclusion, I have developed an efficient and reliable method for quantitative site-specific profiling of the redox dynamics of protein thiols on a global scale. This map provides insights in how the redox states of Cys residues affect play a role in the redox regulation of broad biological processes like photosynthesis and metabolism. Several novel proteins that presumably play important roles in regulating redox homeostasis in cyanobacteria have been discovered. Furthermore, components of the PBS, PSI, PSII, and other proteins translated from genes in the Core Transcriptional Response (Singh et al. 2010) were significantly oxidized after both dark and DCMU in light exposures. Given the redox sensitivity many phycobiliproteins have, we decided to research the mechanism of NblA-mediated PBS degradation in the fast growing strain *Synechococcus* 2973. A His-tagged version of *nblA* had been overexpressed in *E. coli*, and *in vitro* binding studies suggested that purified NblA specifically binds to PBS. Cross-linking coupled with MS experiments elucidated the binding partners of NblA. NblA may be able bind and disrupt the structural integrity of the PBS complexes via the α-helical structural mimicry

they share with phycobiliproteins (Dines et al. 2008), highly conserved amino acid residues (Karradt et al. 2008), electrostatic interactions, and/or van der Waals interactions. Together, these studies are unraveling the different aspects of light harvesting and the various strategies that cyanobacteria employ to optimize energy capture.

1.7 References

Adams, D.G., Duggan, P.S. (2008). "Cyanobacteria–bryophyte symbioses." *Journal of Experimental Botany*. **59**(5): 1047-1058.

Adir, N., Dines, M., Klartag, M., McGregor, A., Melamed-Frank, M. (2006). "Assembly and disassembly of phycobilisomes." *Complex Intracellular Structures in Prokaryotes*. Springer Berlin Heidelberg, Berlin, Heidelberg: 47-77.

Adir, N., Zer, H., Shochat, S., Ohad, I. (2003). "Photoinhibition - a historical perspective." *Photosynthesis Research*. **76**: 343-370.

Ajlani, G., Vernotte, C. (1998). "Construction and characterization of a phycobiliprotein-less mutant of *Synechocystis* sp. PCC 6803." *Plant Molecular Biology*. **37**(3): 577-580.

Andersson, F.I., Blakytny, R., Kirstein, J., Turgay, K., Bukau, B., Mogk, A., Clarke, A.K. (2006). "Cyanobacterial ClpC/HSP100 protein displays intrinsic chaperone activity." *Journal of Biological Chemistry*. **281**(9): 5468-5475.

Bachi, A., Dalle-Donne, I., Scaloni, A. (2012). "Redox proteomics: chemical principles, methodological approaches and biological/biomedical promises." *Chemical Reviews*. **113**(1): 596-698.

Bibby, T.S., Nield, J., Partensky, F., Barber, J. (2001). "Oxyphotobacteria. Antenna ring around photosystem I." *Nature*. **413**: 590.

Bogorad, L. (1975). "Phycobiliproteins and complementary chromatic adaptation." *Annual Review of Plant Physiology*. **26**(1): 369-401.

Brandes, N., Reichmann, D., Tienson, H., Leichert, L.I., Jakob, U. (2011). "Using quantitative redox proteomics to dissect the yeast redoxome." *Journal of Biological Chemistry*. **286**(48): 41893-41903.

Buchanan, B.B. (1991). "Regulation of CO_2 assimilation in oxygenic photosynthesis: the ferredoxin/thioredoxin system. Perspective on its discovery, present status, and future development." *Archives of Biochemistry and Biophysics*. **288**(1): 1-9.

Buchanan, B.B., Balmer, Y. (2005). "Redox regulation: a broadening horizon." *Annual Review of Plant Biology*. **56**: 187-220.

Buick, R. (2008). "When did oxygenic photosynthesis evolve?" *Philosophical Transactions of The Royal Society B.* **363**(1504): 2731-2743.

Butterfield, D.A., Dalle-Donne, I. (2012). "Redox proteomics." *Antioxidants & Redox Signaling*. **17**(11): 1487-1489.

Campbell, D., Hurry, V., Clarke, A.K., Gustafsson, P., Öquist, G. (1998). "Chlorophyll fluorescence analysis of cyanobacterial photosynthesis and acclimation." *Microbiology and Molecular Biology Reviews*. **62**(3): 667-683.

Collier, J.L., Grossman, A.R. (1994). "A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria." *EMBO Journal*. **13**(5): 1039-1047.

D'Alessandro, A., Rinalducci, S., Zolla, L. (2011). "Redox proteomics and drug development." *Journal of Proteomics*. **74**(12): 2575-2595.

Dalle-Donne, I., Rossi, R., Colombo, G., Giustarini, D., Milzani, A. (2009). "Protein S-glutathionylation: a regulatory device from bacteria to humans." *Trends in Biochemical Sciences*. **34**(2): 85-96.

Demmig-Adams, B., Adams, W.W. III. (1996). "The role of xanthophyll cycle carotenoids in the protection of photosynthesis." *Trends in Plant Science*. **1**(1): 21-26.

Dines, M., Sendersky, E., David, L., Schwarz, R., Adir, N. (2008). "Structural, functional, and mutational analysis of the NblA protein provides insight into possible modes of interaction with the phycobilisome." *Journal of Biological Chemistry*. **283**(44): 30330-30340.

Forrester, M.T., Foster, M.W., Stamler, J.S. (2007). "Assessment and application of the biotin switch technique for examining protein S-nitrosylation under conditions of pharmacologically induced oxidative stress." *Journal of Biological Chemistry*. **282**(19): 13977-13983.

Fratelli, M., Demol, H., Puype, M., Casagrande, S., Eberini, I., Salmona, M., Bonetto, V., Mengozzi, M., Duffieux, F., Miclet, E., Bachi, A., Vandekerckhove, J., Gianazza, E., Ghezzi, P. (2002). "Identification by redox proteomics of glutathionylated proteins in oxidatively stressed human T lymphocytes." *Proceedings of the National Academy of Sciences*. **99**(6): 3505-3510.

Gill, R.T., Katsoulakis, E., Schmitt, W., Taroncher-Oldenburg, G., Misra, J., Stephanopoulos, G. (2002). "Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. PCC 6803." *Journal of Bacteriology*. **184**(13): 3671-3681.

Gilmore, A.M. (1997). "Mechanistic aspects of xanthophyll cycle-dependent photoprotection in higher plant chloroplasts and leaves." *Physiologia Plantarum*. **99**(1): 197-209.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., Collier, J.L. (1993). "The phycobilisome, a light-harvesting complex responsive to environmental conditions." *Microbiological Reviews*. **57**(3): 725-749.

Held, J.M., Gibson, B.W. (2012). "Regulatory control or oxidative damage? Proteomic approaches to interrogate the role of cysteine oxidation status in biological processes." *Molecular & Cellular Proteomics*. **11**(4): 1-14.

Herrero, A., Muro-Pastor, A.M., Flores, E. (2001). "Nitrogen control in cyanobacteria." *Journal of Bacteriology*. **183**(2): 411-425.

Hihara, Y., Sonoike, K., Kanehisa, M., Ikeuchi, M. (2003). "DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803." *Journal of Bacteriology*. **185**(5): 1719-1725.

Hoiczyk, E., Hansel, A. (2000). "Cyanobacterial cell walls: news from an unusual prokaryotic envelope." *Journal of Bacteriology*. **182**(5): 1191-1199.

Horton, P., Ruban, A.V., Walters, R.G. (1996). "Regulation of light harvesting in green plants." *Annual Review of Plant Physiology and Plant Molecular Biology*. **47**: 655-684.

Hosoya-Matsuda, N., Motohashi, K., Yoshimura, H., Nozaki, A., Inoue, K., Ohmori, M., Hisabori, T. (2005). "Anti-oxidative stress system in cyanobacteria. Significance of type II peroxiredoxin and the role of 1-Cys peroxiredoxin in *Synechocystis* sp. strain PCC 6803." *Journal of Biological Chemistry*. **280**(1): 840-846.

Izquierdo-Alvarez, A., Martinez-Ruiz, A. (2011). "Thiol redox proteomics seen with fluorescent eyes: the detection of cysteine oxidative modifications by fluorescence derivatization and 2-DE." *Journal of Proteomics*. **75**(2): 329-338.

Jaffrey, S.R., Erdjument-Bromage, H., Ferris, C.D., Tempst, P., Snyder, S.H. (2001). "Protein S-nitrosylation: a physiological signal for neuronal nitric oxide." *Nature Cell Biology*. **3**(2): 193-197.

Jaffrey, S.R., Snyder, S.H. (2001). "The biotin switch method for the detection of S-nitrosylated proteins. *Science's STKE*. **2001**(86): 1-9.

Karradt, A., Sobanski, J., Mattow, J., Lockau, W., Baier, K. (2008). "NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease." *Journal of Biological Chemistry*. **283**(47): 32394-32403.

Leichert, L.I., Gehrke, F., Gudiseva, H.V., Blackwell, T., Ilbert, M., Walker, A.K., Strahler, J.R., Andrews, P.C., Jakob, U. (2008). "Quantifying changes in the thiol redox proteome upon oxidative stress *in vivo*." *Proceedings of the National Academy of Sciences*. **105**(24): 8197-8202.

Leonard, S.E., Carroll, K.S. (2011). "Chemical 'omics' approaches for understanding protein cysteine oxidation in biology." *Current Opinion in Chemical Biology*. **15**(1): 88-102.

Lin, D., Li, J., Slebos, R.J., Liebler, D.C. (2010). "Cysteinyl peptide capture for shotgun proteomics: global assessment of chemoselective fractionation." *Journal of Proteome Research*. **9**(10): 5461-5472.

Lindahl, M., Florencio, F.J. (2003). "Thioredoxin-linked processes in cyanobacteria are as numerous as in chloroplasts, but targets are different." *Proceedings of the National Academy of Sciences*. **100**(26): 16107-16112.

Liu, H., Zhang, H., Niedzwiedzki, D.M., Prado, M., He, G., Gross, M.L., Blankenship, R.E. (2013). "Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria." *Science*. **342**(6162): 1104-1107.

Liu, T., Qian, W.J., Chen, W.N., Jacobs, J.M., Moore, R.J., Anderson, D.J., Gritsenko, M.A., Monroe, M.E., Thrall, B.D., Camp, D.G. II., Smith, R.D. (2005)." Improved proteome coverage by using high efficiency cysteinyl peptide enrichment: the human mammary epithelial cell proteome." *Proteomics*. **5**(5): 1263-1273.

Liu, T., Qian, W.J., Strittmatter, E.F., Camp, D.G. II., Anderson, G.A., Thrall, B.D., Smith, R.D. (2004). "High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology." *Analytical Chemistry*. **76**(18): 5345-5353.

Lopez-Maury, L., Sanchez-Riego, A.M., Reyes, J.C., Florencio, F.J. (2009). "The glutathione/glutaredoxin system is essential for arsenate reduction in *Synechocystis* sp. strain PCC 6803." *Journal of Bacteriology*. **191**(11): 3534-3543.

Mata-Cabana, A., Florencio, F.J., Lindahl, M. (2007). "Membrane proteins from the cyanobacterium *Synechocystis* sp. PCC 6803 interacting with thioredoxin." *Proteomics*. **7**(21): 3953-3963.

McDonagh, B. (2009). "Diagonal electrophoresis for detection of protein disulphide bridges." *Methods Molecular Biology*. **519**: 305-310.

Navrot, N., Finnie, C., Svensson, B., Hagglund, P. (2011). "Plant redox proteomics." *Journal of Proteomics*. **74**(8): 1450-1462.

Nelson, K.J., Klomsiri, C., Codreanu, S.G., Soito, L., Liebler, D.C., Rogers, L.C., Daniel, L.W., Poole, L.B. (2010). "Use of dimedone-based chemical probes for sulfenic acid detection methods to visualize and identify labeled proteins." *Methods in Enzymology*. **473**: 95-115.

Niyogi, K.K. (1999). "Photoprotection revisted: genetic and molecular approaches." *Annual Review of Plant Physiology and Plant Molecular Biology*. **50**: 333-359.

Noctor, G., Foyer, C.H. (1998). "Ascorbate and glutathione: keeping active oxygen under control." *Annual Review of Plant Physiology and Plant Molecular Biology*. **49**: 249-279.

Pakrasi, H.B. (1995). "Genetic analysis of the form and function of photosystem I and photosystem II." *Annual Review of Genetics*. **29**: 755-776.

Pfannschmidt, T., Brautigam, K., Wagner, R., Dietzel, L., Schroter, Y., Steiner, S., Nykytenko, A. (2009). "Potential regulation of gene expression in photosynthetic cells by redox and energy state: approaches towards better understanding." *Annals of Botany*. **103**(4): 599-607.

Rinalducci, S., Murgiano, L., Zolla, L. (2008). "Redox proteomics: basic principles and future perspectives for the detection of protein oxidation in plants." *Journal of Experimental Botany*. **59**(14): 3781-3801.

Seo, Y.H., Carroll, K.S. (2009). "Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies." *Proceedings of the National Academy of Sciences*. **106**(38): 16163-16168.

Sethuraman, M., McComb, M.E., Heibeck, T., Costello, C.E., Cohen, R.A. (2004). "Isotope-coded affinity tag approach to identify and quantify oxidant-sensitive protein thiols." *Molecular & Cellular Proteomics*. **3**(3): 273-278.

Shulaev, V., Oliver, D.J. (2006). "Metabolic and proteomic markers for oxidative stress. New tools for reactive oxygen species research." *Plant Physiology*. **141**(2): 367-372.

Singh, A., Elvitigala, T., Cameron, J.C., Ghosh, B.K., Bhattacharyya-Pakrasi, M., Pakrasi, H.B. (2010). "Integrative analysis of large scale expression profiles reveals core transcriptional response and coordination between multiple cellular processes in a cyanobacterium." *BioMed Central*. **4**(105): 1-14.

Stanne, T.M., Pojidaeva, E., Andersson, F.I., Clarke, A.K. (2007). "Distinctive types of ATP-dependent Clp proteases in cyanobacteria." *Journal of Biological Chemistry*. **282**(19): 14394-14402.

Su, D., Shukla, A.K., Chen, B., Kim, J.S., Nakayasu, E., Qu, Y., Aryal, U., Weitz, K., Clauss, T.R., Monroe, M.E., Camp, D.G., Bigelow, D.J., Smith, R.D., Kulkarni, R.N., Qian, W.J. (2013). "Quantitative site-specific reactivity profiling of S-nitrosylation in mouse skeletal muscle using cysteinyl peptide enrichment coupled with mass spectrometry." *Free Radical Biology & Medicine*. **57**: 68-78.

Vermaas, W.F.J. (2001). "Photosynthesis and respiration in cyanobacteria." *Encyclopedia of Life Sciences*. 1-7.

Weisz, D.A., Gross, M.L., Pakrasi, H.B. (2016). "The use of advanced mass spectrometry to dissect the life-cycle of photosystem II." *Frontiers in Plant Sciences*. **7**(617): 1-25.

Winterbourn, C.C., Hampton, M.B. (2008). "Thiol chemistry and specificity in redox signaling." *Free Radical Biology and Medicine*. **45**(5): 549-561.

Zak, E., Norling, B., Maitra, R., Huang, F., Andersson, B., Pakrasi, H.B. (2001). "The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes." *Proceedings of the National Academy of Sciences*. **98**(23): 13443-13448.

Chapter 2

Proteome-wide Light/Dark Modulation of Thiol Oxidation in Cyanobacteria Revealed by Quantitative Site-Specific Redox Proteomics

This research was originally published in *Molecular and Cellular Proteomics*. Amelia Y. Nguyen[#]*, Jia Guo[#], Ziyu Dai[#], Dian Su, Matthew J. Gaffrey, Ronald J. Moore, Jon M. Jacobs, Matthew E. Monroe, Richard D. Smith, David W. Koppenaal, Himadri B. Pakrasi, and Wei-Jun Qian[1]. Proteome-wide Light/Dark Modulation of Thiol Oxidation in Cyanobacteria Revealed by Quantitative Site-Specific Redox Proteomics. *Molecular and Cellular Proteomics*. 2014. 13(2):3270-85.

[#]These authors contributed equally to this paper.

*A.Y.N. prepared all cyanobacterial samples, analyzed the data, and wrote the manuscript.

2.1 Introduction

2.1.1 Abstract

Reversible protein thiol oxidation is an essential regulatory mechanism of photosynthesis, metabolism, and gene expression in photosynthetic organisms. Herein, we present a proteome-wide quantitative and site-specific profiling of *in vivo* thiol oxidation modulated by light/dark in the cyanobacterium *Synechocystis* sp. PCC 6803, an oxygenic photosynthetic prokaryote, using a resin-assisted thiol enrichment approach. Our proteomic approach integrates resin-assisted enrichment with isobaric tandem mass tags labeling to enable site-specific and quantitative measurements of reversibly oxidized thiols. The redox dynamics of ~2,100 Cys-sites from 1,060 proteins under light, dark, and 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU, a photosystem II inhibitor) conditions were quantified. In addition to relative quantification, the stoichiometry or percentage of oxidation (reversibly oxidized/total thiols) for ~1,350 Cys-sites was also quantified. The overall results revealed broad changes in thiol oxidation in many key biological processes including photosynthetic electron transport, carbon fixation, and glycolysis. Moreover, the redox sensitivity along with the stoichiometric data enabled prediction of potential functional Cys-sites for proteins of interest. The functional significance of redox-sensitive Cys-sites in NADP-dependent glyceraldehyde-3-phosphate dehydrogenase, peroxiredoxin (AhpC/TSA family protein Sll1621), and glucose 6-phosphate dehydrogenase were further confirmed by site-specific mutagenesis and biochemical studies. Together, our findings provide significant insights into the broad redox regulation of photosynthetic organisms.

2.1.2 Redox Biology

Reversible protein thiol oxidation has been recognized as a fundamental redox regulatory mechanism occurring throughout biology and plays essential roles in photosynthesis, cellular

metabolism, gene expression, and other key biological processes (Buchanan and Balmer 2005; Foyer and Noctor 2009). Protein thiols serve as important redox switches in cells through reversible thiol oxidation by forming diverse post-translational modifications (PTMs), including disulfide (Sato and Inaba 2012), S-nitrosylation (SNO) (Derakhshan et al. 2007; Greco et al. 2006), S-glutathionylation (SSG) (Hess et al. 2005), and S-sulfenylation (SOH) (Paulsen et al. 2012). In the case of photosynthesis, redox regulation is inextricably associated with the photosynthetic electron transport (PET) chain and regulation of gene expression. In light/dark cycles, thiol oxidation (e.g., disulfide formation) through the thioredoxin and glutaredoxin systems modulates the activation status of many enzymes linked to photosystem I (PSI) and photosystem II (PSII), as well as many other redox related processes in plants (Buchanan and Balmer 2005) and cyanobacteria (Lindahl and Florencio 2003). However, it is still largely unknown how broadly the redox process is involved beyond photosynthesis and, importantly, what are the specific cysteine (Cys) sites serving as redox switches in photosynthetic organisms.

2.1.3 Thiol Redox Dynamics in Cyanobacteria

Cyanobacteria are considered useful model organisms for photosynthesis research due to their evolutionary similarities to chloroplasts (Falcon et al. 2010). These oxygenic photosynthetic prokaryotes are also increasingly recognized as microbial biofactories for the production of desired chemicals (Singh et al. 2011), and potentially for biofuels from solar energy and carbon dioxide (Singh et al. 2010; Nogales et al. 2012). Of particular interest is the thiol redox dynamics of proteins localized in the thylakoid lumen as redox regulation directly impacts the PET chain and photosystem stability, which in turn affects light energy utilization efficiency and biofuel production (Rosgaard et al. 2012). In the cyanobacterium *Synechocystis* sp. PCC 6803 (hereafter referred to as *Synechocystis* 6803), modulation of the thiol redox states of the PET chain

(Alfonso et al. 2000) and the redox environment (Cameron and Pakrasi 2010) has been observed in response to a light/dark cycle. The oxidative environment of the dark phase, due to the decreased PSII activity and NADPH production (Alfonso et al. 2000; Pfannschmidt et al. 2009), alters the redox state of PET chain electron carriers by oxidizing the proteins, similar to changes exhibited in the presence of electron transport inhibitors (Alfonso et al. 2000; Pfannschmidt et al. 2009). This in turn suggests that the redox state of the PET chain is a key mechanism of light/dark gene transcript regulation, involving the stability and protein PTMs in response to environmental stresses (Pfannschmidt et al. 2009; Lopez-Maury et al. 2009). In the current study, we applied a recently described quantitative redox proteomic approach (Guo et al. 2014) with the aim of obtaining a proteome-wide view of *in vivo* thiol-oxidation in *Synechocystis* 6803 as modulated by the light/dark cycle. Due to the complexity of the multiple forms of Cys oxidation (*e.g*, disulfide, SNO, SSG, SOH), the current study focuses on the quantification of total reversible thiol oxidation as a starting point of investigating the thiol redox proteome.

2.1.4 *In Vivo* Thiol Quantification

Measurements of *in vivo* thiol redox status at physiological conditions have traditionally been challenging because free thiols are unstable and prone to oxidation during sample preparation with labile reversible oxidized Cys-residues. Although many redox-sensitive proteins have been reported (Lindahl and Kieselbach 2009), in most cases it was unknown which specific Cys-residues were modified *in vivo*. To identify Cys modifications, most initial redox approaches employed differential chemical or isotopic labeling followed by biotin-based affinity enrichment. In these methods, free protein thiols were initially blocked by alkylation and then specific oxidized Cys-residues were selectively reduced and labeled with biotin tags (Jaffrey et al. 2001), fluorophores (Izquierdo-Alvarez and Martinez-Ruiz 2011; Nelson et al. 2010), radioactive

compounds (Leichert and Jakob 2004), or isotope-coded affinity tags (Leichert et al. 2008; Sethuraman et al. 2004) for Western blot, gel electrophoresis, or mass spectrometry (MS) analysis. Alternatively, thiol-reactive chemical probes coupled with click chemistry have recently been applied to identify redox sensitive proteins or Cys-residues (Deng et al. 2013; Weerapana et al. 2010) and *in vivo* thiol redox changes (Sadler et al. 2013). However, most of these approaches do not provide broad site-specific identifications of Cys modifications with accurate quantification. Several groups have recently reported an alternative isobaric iodoTMT or cysTMT switch approach for multiplexed quantification of reversible thiol modifications (Pan et al. 2014; Qu et al. 2014; Murray et al. 2012). However, the enrichment specificity of the anti-TMT-based approach is still a potential factor of concern that may limit the achievable coverage of Cys-peptides since in a recent report only ~21% of enriched peptides were observed to be TMT-labeled peptides using an optimized competitive elution buffer (Qu et al. 2014).

Recently, we and others have developed a resin-assisted enrichment to facilitate more sensitive proteomic identification and quantification of Cys-containing peptides (Liu et al. 2004) and Cys-based reversible modifications (Forrester et al. 2009; Guo et al. 2014). In this study, we extend the resin-assisted enrichment for proteome-wide quantification of *in vivo* reversible thiol oxidation on individual Cys sites under multiple conditions (light, dark, and in the presence of a photosystem II inhibitor DCMU [3-(3,4-dichlorophenyl)-1,1-dimethylurea]) in *Synechocystis* 6803. ~2,100 Cys sites from 1,060 proteins were identified with the vast majority of Cys sites displaying redox changes induced by the light/dark switch. Our results provide a broad quantitative picture of redox-mediated changes of the thiol proteome in cyanobacteria, inferring the extent of thiol-based redox regulation in photosynthetic organisms.

2.2 Materials and Methods

2.2.1 Reagents

Synechocystis 6803 cells were obtained from Washington University (St. Louis, MO). *E. coli* strains were purchased from New England Biolabs (Ipswith, MA). BCA protein assay reagents, silver stain kit, spin columns, tandem mass tags (TMT) reagents, and isopropyl β-D-1-thiogalactopyranoside (IPTG) were purchased from Thermo Fisher Scientific (Rockford, IL). Porcine trypsin was from Promega (Madison, WI). The SeeBlue Plus2 protein standard was from Invitrogen (Carlsbad, CA). Thiopropyl Sepharose 6B affinity resin was from GE Healthcare Bio-Sciences (Pittsburgh, PA). Tris/glycine/SDS (TGS) buffer, Laemmli sample loading buffer, and Tris-HCl precast gel with 4–20% linear gradient were all from Bio-Rad Laboratories (Hercules, CA). Unless otherwise noted, all other chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO).

2.2.2 Cyanobacteria Culture Conditions

Oxygen evolution in cyanobacteria and plant promotes significant post-translational modifications (PTMs) of the proteins present in the thylakoid membranes. The thylakoid membrane in cyanobacteria is one of the most redox active compartments in biological systems since oxygen is constantly evolved during photosynthesis. However, little is known about the redox proteomics in cyanobacteria. The thiol side chain in Cys often participates in enzymatic reactions and may become inactivated when oxidized and activated when reduced (Buchanan and Balmer 2005). Additionally, many metabolic pathways in chloroplastic proteins are regulated by the transitions of protein thiols from dithiol (reduced Cys) to disulfide (oxidized Cys) (Navrot et al. 2011). Elucidation of the redox proteome in cyanobacteria under different

redox stress conditions will undoubtedly reveal new insights into the functions of proteins that protect photosynthetic organisms under unfavorable conditions.

Understanding the dynamic redox changes of specific protein thiols on a global scale should provide a comprehensive picture of the key players that protect the photosynthetic apparatus from oxidative stress. The discovery that an increasing number of proteins undergo reversible ROS-mediated thiol modifications instead of photo damage in various biological systems supports the rising concept of redox regulation (Fratelli et al. 2002). The extent to which cysteine (Cys) are modified depend on the amounts and types of ROS present, the reactivity of the Cys thiols involved, and the kinetic and thermodynamic stability of the respective thiol modifications (Winterbourn and Hampton 2008). Thiols play a central role in coordinating the antioxidant defense network by acting as reversible redox switches. Previous studies suggest that the oxidative PTMs of protein thiols induce redox regulation in cells by switching protein functions on or off. For instance, the reduction of thiols by thioredoxin (Trx) modulates the activation or inactivation of enzymes linked to photosystem (PS) I (PSI) and II (PSII) (Buchanan and Balmer 2005), and various other redox-regulated target processes in plants and cyanobacteria (Lindahl and Florencio 2003). In short, the redox states of protein thiols are crucial determinants of multiple metabolic, signaling, and transcriptional processes in cells (Dalle-Donne et al. 2009).

Countless heterotrophic organisms still generate numerous ROS during respiration and fatty acid oxidation, despite being incapable of evolving oxygen like phototrophic organisms. Heterotrophic organisms like *Escherichia coli* (*E. coli)* and *Saccharomyces cerevisiae* (yeast) have one form of electron transport chain that takes place during respiration. In contrast, photosynthetic organisms contain two types of electron transport chains involved in

photosynthesis and respiration. How proteins change in response to the production of ROS during photosynthesis and respiration remains unclear for some organisms, but recent studies in various biological systems demonstrated that proteins rely on reversible ROS-mediated PTMs to regulate their functions (Rinalducci et al. 2008; Winterbourn and Hampton 2008; Fratelli et al. 2002; Brandes et al. 2011).

The thiol side chain in Cys, under steady-state or oxidative stresses conditions, can be modified in various forms, e.g. reversible forms: disulfides, S-nitrosylation (SNO), S-glutathionylation (SSG), sulfenylation (SOH), irreversible forms: sulfinylation (SO_2H), and sulfonylation (SO_3H) (Bachi et al. 2013; Held and Gibson 2012). But detection of the dynamic redox states of specific Cys in proteins as cells are exposed to different environmental stresses have been limited due to 1) the transient nature of PTMs of thiols and 2) technical challenges associated with site-specific detection of Cys modifications.

Traditionally, Trx affinity chromatography had been used to identify proteins that are regulated by Trx. About 50 novel Trx-regulated proteins were identified by 2007 using this biochemical approach (Hosoya-Matsuda et al. 2005; Lindahl and Florencio 2003; Mata-Cabana et al. 2007). The rapid increase in popularity of deep sequencing and MS opened the doors for the identification of the dynamic PTMs of protein thiols. The numerous applications mass spectrometry offers to science triggered the creation of redox proteomics, which investigates how oxidative stress modifies protein structure and function (D'Alessandro et al. 2011; Rinalducci et al. 2008). Numerous tools have been developed to detect diverse sets of Cys modifications as a result of the global interest in redox biology. Traditional methods used diagonal electrophoresis (McDonagh 2009) or antibodies against Cys oxoforms (Seo and Carroll 2009), but they suffer from moderate sensitivity and limited site-specific identification. Methods

based on differential chemical labeling have also been developed to identify Cys modifications and their dynamic changes (Bachi et al. 2013; Lin et al. 2010; Forrester et al. 2007; Jaffrey and Snyder 2001; Jaffrey et al. 2001; Izquierdo-Alvarez and Martinez-Ruiz 2011; Nelson et al. 2010; Leichert et al. 2008; Sethuraman et al. 2004; Liu et al. 2004; Liu et al. 2005; Leonard and Carroll 2011). Our collaborators at the Pacific Northwest National Laboratory (PNNL) and others have developed proteomic approaches to enrich protein/peptide thiols for quantitative and site-specific profiling of Cys-containing peptides (Liu et al. 2004; Liu et al. 2005), Cys-SNO (Su et al. 2013) and Cys-SSG (Dr. Jia Guo, data not published).

2.2.3 Sample Preparation

Cell pellets were resuspended with 10% (w/v) trichloroacetic acid (TCA) and incubated on ice for 20 min, which leads to partial lysis of the bacterial cells, precipitation of proteins and effectively stops thiol-disulfide exchange reactions. After this incubation, proteins and cell debris were spun down at 4 °C for 30 min at 13,000 *g* and then carefully rinsed the pellet with 500 µl ice cold 10% (w/v) TCA, followed by a second rinse with 200 µl ice-cold 5% (w/v) TCA. Lysis buffer (pH 7.6) containing 100 mM N-ethylmaleimide (NEM), 200 mM Tris-HCI, 10mM EDTA, 0.5% (w/v) SDS and 8 M urea was added to dissolve the above pellet with brief intermittent sonication until the pellet dissolved. The resultant mixture was incubated at 37 °C for 2 h. Bead-beading was performed using 200 µl of 0.1 mm Zirconia/Silica beads to extract more proteins. Another 200 µl of lysis buffer was used to wash beads, and cell lysates were centrifuged at 10,000 × *g* for 10 minutes at 4 °C to pellet cellular debris. Supernatant was transferred to a new 2 ml microcentrifuge tube, and excess NEM was removed by cold acetone precipitation. Proteins were re-dissolved in lysis buffer, and 10 mM DTT was added to the above mixture to reduce the oxidized thiols. After reduction, excessive DTT reagents were removed by buffer exchange with

8 M urea and then Milli-Q water by centrifuge centrifugation at 4,000 × *g* for 30 min at 4 °C. Proteins concentration was determined by the BCA assay and total 500 μg of proteins were used for the following enrichment experiment. ~100 μg of proteins from each sample was subjected to trypsin digestion without enrichment according to the previously described procedure (Qian et al. 2008) for global proteome profiling.

2.2.4 Selective Enrichment of Oxidized or Total Cys-peptides

The samples were transferred to Handee mini spin columns with 35 mg prewet and prewashed Thiopropyl Sepharose 6B resin and 50 mM HEPES buffer as described previously (Guo et al. 2014). To enrich the total Cys-peptides, NEM was not added during sample processing. Enrichment of free Cys-containing proteins was carried out by incubating the samples in a Thermomixer at room temperature with shaking at 850 rpm for 2 h. Nonspecifically bound proteins were removed by washing five times with the following solutions: (1) 8 M urea, (2) 2 M NaCl, (3) 80% (v/v) acetonitrile (ACN) with 0.1% (v/v) trifluoroacetic acid (TFA), and (4) 25 mM HEPES (pH 7.7). To perform the on-resin digestion, 120-μl solution containing 25 mM HEPES buffer (pH 7.7), 1 M urea, 5 mg trypsin, and 1 mM $CaCl_2$ was added to the columns and the samples were incubated at 37 °C with shaking at 850 rpm for 3 h. Nonspecifically bound peptides were washed away using the same procedure as described above for removing nonspecifically bound proteins.

2.2.5 On-resin TMT labeling

On-resin isobaric labeling using six-plex TMT were performed for labeling the captured peptides to achieve relative quantification for oxidized Cys under light, dark or DCMU. For TMT labeling, acetonitrile (40 μl) was added to the manufacturer-provided TMT reagents to dissolve/dilute them. Then, 80 μl of dissolution buffer and above TMT reagent solutions were

added to the spin columns. The labeling reaction was carried out at room temperature with shaking at 850 rpm on a Thermomixer for 1 h. The reaction was stopped by adding 8 µl of 5% (w/w) NH_2OH in 200 mM triethylammonium bicarbonate buffer followed by incubation at room temperature with shaking at 850 rpm for 15 min. The excess TMT reagents were removed by washing five times with the following solutions: (1) 80% (v/v) ACN/0.1% (v/v) TFA and (2) 25 mM ammonium bicarbonate (pH 7.8). Cys-peptides were eluted by incubation with 20 mM dithiothreitol (DTT) in 100 µl of 25 mM ammonium bicarbonate for 30 min followed by rinsing with 100 µl of 80% ACN and 0.1% TFA. Cys-peptide samples were then concentrated in a Thermo Scientific Speed Vac concentrator and adjusted to a final volume of 25 µl with water. The final Cys-peptide sample contained 20 mM DTT, which prevented free thiols from oxidation before LC–MS/MS analysis. Equal amounts of the samples from each labeling channel were then mixed to generate the final sample for LC–MS/MS analyses.

2.2.6 SDS-polyacrylamide Gel Electrophoresis

SDS–PAGE was carried out with a 4–20% (w/v) precast linear gradient Tris–HCl polyacrylamide gel (Bio-Rad) to assess enrichment efficiency and total oxidation levels. Equal volumes (5 µl) of the above Cys-peptide sample and Laemmli sample buffer (Bio-Rad) were mixed and then incubated at 95 °C for 5 min. Meanwhile, 1 ml of See Blue Plus 2 protein standard was directly loaded onto the gel. Gel electrophoresis was run at 170 V for 30 min in Tris/glycine/SDS buffer (Bio-Rad). After electrophoresis, the gel image was developed with silver staining following the manufacturer's standard protocol.

2.2.7 LC-MS/MS Analysis

All TMT-labeled samples were analyzed by LC–MS/MS with two technical replicates. All peptide samples were analyzed by a Waters nano-Aquity UPLC system (Waters Corporation,

Milford, MA) with a homemade 75 μm i.d. × 70 cm reversed-phase capillary column using 3 μm C18 particles (Phenomenex, Torrance, CA, USA). The system was operated at a constant flow of 300 nl/min over 3 h with a gradient starting with 100% of mobile phase A (0.1% (v/v) formic acid in water) to 60% (v/v) of mobile phase B (0.1% (v/v) formic acid in acetonitrile). MS analysis was performed on a Thermo Scientific LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA, USA) coupled with an electrospray ionization interface using home-made 150 μm o.d. × 20 μm i.d. chemically etched electrospray emitters (Kelly et al. 2006). The heated capillary temperature and spray voltage were 350 ºC and 2.2 kV, respectively. Full MS spectra were recorded at resolution of 100 K (m/z 400) over the range of m/z 400–2000 with an automated gain control (AGC) value of 1×10^6. MS/MS was performed in the data-dependent mode with an AGC target value of 3×10^4. The most abundant 10 parent ions were selected for MS/MS using high-energy collision dissociation (HCD) with a normalized collision energy setting of 45. Precursor ion activation was performed with an isolation width of 2 Da, a minimal intensity of 500 counts, and an activation time of 10 ms. A dynamic exclusion time of 45 s was used. Label-free analyses were performed by the same LC-MS platform from three biological replicates under data-dependent acquisition mode using collision-induced dissociation with a normalized collision energy setting of 35.

2.2.8 Data Analysis

LC-MS/MS raw data were converted into dta files using Bioworks Cluster 3.2 (Thermo Fisher Scientific, Cambridge, MA, USA), and an MSGF plus algorithm (v9979 released in March 2014)(Kim et al. 2008) was used to search MS/MS spectra against the *Synechocystis* 6803 FASTA database (3,169 entries, Kazusa Genome Resources 2011-09-09). The key search parameters used were 20 ppm tolerance for precursor ion masses, 0.5 Da tolerance for fragment

ions, partial tryptic search with up to three miscleavages, dynamic oxidation of methionine (15.9949 Da), dynamic NEM modification of Cys (125.0477 Da), and static 6-plex TMT modification of lysine and N-termini of peptides (229.1629 Da). For the label-free global proteome analysis, dynamic oxidation of methionine, NEM modification of Cys, and carbamidomethylation of Cys-residues (57.0215 Da) were applied. Peptides were identified from database searching results applying the following criteria: MSGF E-value <1E-8, Q-value <0.01, and mass measurement error <10 ppm. The decoy database searching methodology (Qian et al. 2005) was used to confirm the final false discovery rate at the unique peptide level to be ~0.3%. Since NEM blocked only the free Cys-residues, all originally oxidized Cys-residues were identified as unmodified Cys. The Cys-sites of oxidation were identified based on each peptide and its corresponding protein sequence using an in-house software tool, Protein Coverage Summarizer (available at panomics.pnnl.gov).

For TMT-based relative quantification, all MS/MS spectra were grouped based on individual Cys-sites and their TMT reporter ion intensities were summed from all spectra corresponding to a given Cys-site. The relative oxidation levels for individual Cys-sites were calculated by dividing the summed intensities for each TMT channel in a given six-plex experiment against the average values across the three conditions. Following this conversion, the data from the two independent 6-plex experiments were merged to provide data with four biological replicates for each condition. The dynamic changes of relative oxidation for individual Cys-sites were then compared by calculating the ratio between dark and light, and the ratio of DCMU and light. At least 50% change in the level of Cys oxidation was considered confident redox changes ($p<0.05$, ANOVA test). To calculate the relative oxidation level of a given

protein, the data from all significant Cys-sites from a specific protein were averaged as the relative oxidation level of specific proteins.

For TMT-based stoichiometric (percent of reversible oxidation) analysis, the ratios of reporter ion intensities between channels for oxidized thiols and channels for total thiols were calculated as a percentage. For the global proteome profiling, label-free spectral counting strategy was applied. Peptides were chosen if they were observed in all three biological replicates, and spectral counts of each peptide were transformed to log2 values and normalized using the central tendency approach. The peptide level was considered significantly different when p-value is less than 0.05 (ANOVA) with at least 50% change between conditions. Biological function categorization of the identified proteins was clustered based on Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analyses using the DAVID functional annotation (Huang da et al. 2009).

2.2.9 Strains and Culture Preparation for Recombinant Proteins

E. coli Top10 cells were used as a host for Gibson assembly (Gibson et al. 2009). The *E. coli* strains BL2 (DE3) and Lemo21 (DE3) (New England Biolabs, Ipswith, MA) were used for protein expression. All bacterial cultures were grown on luria broth (LB) agar plates or in liquid culture media with 100 μg/ml ampicillin. For plasmid DNA purification, a single colony was grown in 10 ml of LB liquid culture medium in 125 ml Pyrex glass Erlenmeyer flask overnight at 37 °C at 250 rpm. Similarly, for *E. coli* protein expression, a single colony was inoculated in 10 ml of LB medium and grown overnight at 37 °C at 250 rpm. Aliquot (2 ml) of overnight culture was transferred into 50 ml LB medium with 100 μg/ml ampicillin to reach the initial optical density (OD_{600}) of 0.15. After the cultures were grown to mid-logarithmic phase with OD_{600} of 0.5-0.6, IPTG was added into the cultures to a final concentration of 0.5 mM. The cultures were

further grown at 30 °C and 250 rpm for 18-24 h. Bacterial cultures (10 ml) were transferred into 15 ml centrifuge tube and the rest cultures into 50 ml centrifuge tubes. The bacteria were pelleted down by centrifugation at 4 °C and 4,500 × *g* for 15 min. Cell pellets were stored at -80 °C freezer until use.

2.2.10 Site-specific Mutagenesis and Protein Expression Vector Construction

The genomic DNA isolation of *Synechocystis* 6803 was mainly followed by DNA extraction protocol from cyanobacteria described by Sinha *et al.*, 2009 at Protocol Online (http://www.protocol-online.org) with modifications. Cyanobacterial cells (20 µl) were resuspended in 1 ml 10 mM Tris buffer with 1mM EDTA, pH 8.0 and pelleted by centrifugation for 2 min at 10,000 *g* at room temperature. The pelleted cells were resuspended in 200 µl of 1% SDS and incubated at 70 °C for 15 min. After heat treatment, 300 mg glass beads (diameter of 0.4-0.6 mm) and 200 µl water saturated phenol (pH 8.0, Invitrogen, Carlsbad, CA) were added and vortexed for 1 min. The cell debris was pelleted by centrifugation at 15,000 × *g* for 5 min at room temperature. The supernatant was further treated with 1 mg/ml ribonuclease A for 30 min at 50 °C and extracted twice with an equal volume of phenol-chloroform (1:1) and centrifuged at 15,000 × *g* for 5 min at room temperature. The genomic DNA in supernatant was precipitated with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volume of ice-cold 95% ethanol at room temperature for 30 min and pelleted by centrifugation at 5,000 × *g* for 10 min at 4 °C. Finally, the genomic DNA was resuspended in 100 µl of the 10 mM Tris buffer (pH 8.0). The site-specific mutagenesis was carried out for the replacement of selected cysteine with serine in cyanobacterial proteins of glucose-6-phosphate dehydrogenase (Zwf, C187S, C265S and C445S), glyceraldehyde-3-phosphate dehydrogenase (Gap2, C154S, C154S/C158S and C292S), and peroxiredoxin (Sll1621, C55S, C155S/C162S and C162S). Cynaobacterial genomic DNA

(50 ng) were used as DNA templates for polymerase chain reaction (PCR) to generate a specific DNA fragment with one unit Phusion high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA), 50 µl reaction mixture (Phusion HF buffer and dNTP) and a pair of corresponding oligonucleotides listed in **Table 2.1**. DNA fragments were separated in 1% agarose gel and purified with QIAquick gel extraction kit (Qiagen, Valencia, CA).

Protein expression vector construction was carried out with a Gibson assembly kit (New England Biolabs, Ipswich, MA). Briefly, the plasmid DNA of protein expression vector pET6xHN-N (Clontech, Mountain View, CA) was linearized by digestion with restriction endonucleases Stu I and Xba I. Proportional amounts of the linearized plasmid DNA of pET6xHN-N and corresponding DNA fragments were mixed and assembled at 50 °C for 1 h. The assembled DNA mixtures were transferred into *E. coli* Top10 cells by chemical transformation for plasmid DNA replication. The plasmid DNA was purified from specific Top10 cells with QIAprep spin miniprep kit (Qiagen, Valencia, CA). DNA sequencing by Beckman Coulter Genomics (Danvers, MA) verified the site-specific mutation in the selected protein expression vector.

2.2.11 Expression and Purification of Recombinant G6PDH, Gap2, Sll1621

The plasmid DNA of designated protein expression vectors was transformed into the *E. coli* BL2 [(DE3), G6PDH and Gap2] or Lemo21 [(DE3), Sll1621] by following the manufacturer's transformation protocol. Since all three proteins were in frame assembled with 6 repeating His-Asn and enterokinase cleavage site at their N-terminal, proteins were purified with His60 Ni superflow gravity columns (Clontech, Mountain View, CA). Briefly, the *E. coli* cell pellets were fully re-suspended in 1 ml ice-cold Ni60 Ni xTractor buffer per 10 ml of the original cultures by pipetting up and down gently. The re-suspended cells were further homogenized with

Cole Parmer 4710 series homogenizer at the output of 3 and 70% cycle duty for 3 × 5 seconds on ice with 1 min interval. The cell lysates were incubated on ice for 15 min and then centrifuged at 10,000 × *g* and 4 °C for 20 min. Finally, the supernatants were passed through 1 ml His60 Ni gravity columns to purify the recombinant proteins with detailed instruction described by the manufacturer's manual (Clontech, Mountain View, CA). Purified proteins (30 μg) were separated by SDS-PAGE gels and visualized by Gel DocTM EZ system (Bio-Rad, Hercules, CA).

2.2.12 Enzymatic Activity Measurement

The G6PDH catalyzes the first step in the pentose phosphate pathway by oxidizing glucose-6-phosphate (6-PG) and reducing $NADP^+$ (nicotinamide adenine dinucleotide phosphate) to NADPH. The rate of NADPH formation is correlated to the G6PDH activity and can be measured with the change in NADPH absorbance at 340 nm. G6PDH (35 μg) and its mutants were mixed with reaction buffer (50 mM phosphate buffer, 10 mM $MgCl_2$, pH 7.5) and 0.67 mM NADP in 1-ml final volume in a 1-ml plastic cuvette and kept at room temperature for 3 min. The reaction kinetics was initiated by addition of 2 mM 6-PG. G6PDH activity was calculated by mainly following the instruction of G6PDH activity calculation described by Sigma-Aldrich (St. Louis, MO). The activities of Gap2 (GAPDH), peroxiredoxin (Sll1621, thioredoxin reductase) and their mutants were determined by KDalertTM GAPDH assay kit (Life Technologies, Grand Island, NY) and thioredoxin reductase assay kit (Sigma-Aldrich, St. Louis, MO), respectively.

For DTT and H_2O_2 treatment, wild-type Gap2 protein were treated by either 10 mM DTT or specific amount of H_2O_2 by directly adding the reagent into the reaction mixture and incubated at room temperature for 2 min prior to initiation of enzyme reactions. For the treatment

of Sll1621, 10 mM DTT was mixed the 100 µl (250 µg) enzyme and kept at room temperature for 15 min. Then, excess DTT was removed by Amicon ultra-0.5 centrifugal filter devices (EMD Millipore, Billerica, MA) with 500 µl of 50 mM phosphate buffer, pH 7.5 for 5 times. The same procedures for DTT treatment and washing were applied prior to additional 15 min treatment with 1 mM H_2O_2. The activities were determined with thioredoxin reductase assay kit.

2.3 Results

2.3.1 Quantitative Site-specific Measurements of Reversible Thiol Oxidation

Resin-assisted thiol-affinity enrichment has recently shown to be effective for identification and quantification of SNO (Forrester et al. 2009; Su et al. 2013) and SSG (Su et al. 2014). Here we adapted this strategy for the quantitative profiling of the reversible thiol oxidation in the *Synechocystis* 6803 proteome in response to light/dark modulation. Briefly, *Synechocystis* 6803 was grown in the presence of oxygen and continuous light to the mid-log phase under continuous light, then exposed to 1) continuous light, 2) darkness, or 3) DCMU in continuous light (hereafter referred to as DCMU) for 2 h before cell harvesting. All endogenous free thiols are initially trapped by acid trapping using TCA precipitation (Leichert and Jakob 2004) followed by blocking with NEM (**Fig. 2.1a**). Following free thiol blocking, the oxidized Cys-residues (i.e., disulfide or other reversible oxidative modifications) are reduced back to free thiols by DTT. The newly formed free thiols from the originally oxidized proteins are then specifically captured by Thiopropyl Sepharose 6B resin through a disulfide exchange reaction (Liu et al. 2004). The enriched proteins are subjected to on-resin trypsin digestion. Resin bound Cys-peptides are further labeled with amine-reactive isobaric TMT reagents (Thompson et al. 2003) to facilitate quantification of the dynamics of thiol oxidation across multiple conditions. The 6-plex TMT reagents allow simultaneous quantification of six samples in a single

experiment. Here two biological replicates of light, dark, and DCMU conditions can be quantified in a single experiment (**Fig. 2.1b**). The TMT labeled Cys-peptides are subjected to LC-MS/MS analyses for site-specific identification and relative quantification of the levels of oxidation on individual Cys sites.

In addition to the relative quantification, our strategy can also be modified to quantify the percentage (or stoichiometry) of thiol oxidation (reversibly oxidized/total thiols) for individual Cys-residues (**Fig. 2.1c**). In this case, a given biological sample is divided into two identical aliquots with one aliquot processed for oxidized thiols only and the other for total thiols without NEM blocking. By using 6-plex TMT labeling, the percentage of reversible oxidation of three replicates can be quantified in a single LC-MS/MS experiment (Leichert et al. 2008). But several caveats associated with this method should be noted. First, the irreversible oxidation was excluded in the measurement of total thiols. Thus, overestimation of the percentage of reversible oxidation might occur if irreversible oxidation is present at high levels (Held et al. 2010). Second, when a given peptide contains more than one Cys-residues, the percentage of oxidation cannot be accurately measured. We note that this is a common limitation shared by several previously reported methods (Leichert et al. 2008; Knoefler et al. 2012) due to the existence of multiple combinatory forms of oxidized Cys-sites for the given sequence.

In our approach, a key step for measuring oxidized thiols is the effective blocking of free thiols**.** To achieve this, we adapted the acid trapping method (Leichert and Jakob 2004; Held et al. 2010) using 10% TCA followed by alkylation with a high concentration of NEM (100 mM) in lysis buffer to block free thiols and prevent artificial thiol oxidation during sample processing. To assess the efficiency of the free thiol blocking, one set of sample pellets from the TCA precipitation were dissolved in lysis buffer containing the alkylating agent, NEM, while another

set was not exposed to NEM. The set of sample pellets in lysis buffer containing NEM was carefully dissolved with sonication so as to not introduce additional oxygen from air and incubated to block free thiols to prevent artificial oxidation. The final level for enriched oxidized Cys-peptides was observed to be very low (**Fig. 2.1d**) compared to the level of total Cys-peptides in the sample pellet not exposed to NEM. The results suggested that there was minimal artificial oxidation as a result of effective blocking with NEM, and evidence that *in vivo* thiol redox state is being preserved. However, for comparison, cell pellet samples were alternatively subjected to bead-beating before NEM incubation. Simply altering the preparation by performing a bead-beating process prior to blocking introduced a sufficient oxidative environment to artificially increase thiol oxidation. The level of oxidized Cys-peptides was nearly the same as the level of total Cys-peptides, abrogating the detection of any differential abundance (**Fig. 2.1e**).

2.3.2 Broad Redox Changes on Cys-sites Modulated by Light/Dark

To identify site-specific redox changes on protein thiols, LC-MS/MS analyses were performed for four biological replicates of light, dark, and DCMU conditions. ~2,100 oxidized Cys-sites from 1,060 proteins were identified and quantified based on TMT reporter ion intensities in MS/MS spectra (**Table 2.2**). Among all identified peptides, ~98% were Cys-containing peptides, in line with the high efficiency of the resin-assisted enrichment (Liu et al. 2004; Su et al. 2013). The overall results (**Figs. 2.2a** and **2.3a**) demonstrate that most protein thiols are less oxidized (or more reduced) under continuous light, but become significantly more oxidized in the dark phase or in the presence of DCMU. DCMU is a specific inhibitor of PSII, which blocks the plastoquinone binding site of PSII and stops the electron flow in the PET chain (Trebst 2007), thus simulating the dark condition where no photosynthetic electron flow occurs. ~80% of the identified Cys-sites were observed with significant redox changes between light,

dark, and DCMU conditions ($p<0.05$, ANOVA test) (**Table 2.2**). After applying a >1.5 in abundance ratios (i.e., >50% change in oxidation levels) and $p<0.05$ cutoff, a total of ~1,100 Cys-sites were observed with substantial redox changes (**Fig. 2.2b**). This observation of increased oxidation under the dark condition is consistent with the understanding that cells under light, with active photosynthesis and electron flow through the PET chain, foster a reductive environment through the regulatory mechanisms of the reduced ferredoxin and thioredoxin systems, and hence are reflected in a more reduced form of thiols from thioredoxin target enzymes (Dai et al. 2004; Buchanan and Balmer 2005). However, the dark conditions, or similarly in the case of DCMU inhibition of the PET chain, facilitates a shift to a more oxidative environment, leading to higher levels of oxidative modifications in target enzymes which modulates enzyme activity across numerous cellular processes.

2.3.3 Percentage of Reversible Oxidation of Individual Cys-sites

We sought to measure the percentage or stoichiometry of reversible oxidation of individual Cys-sites using the strategy outlined in **Fig. 2.1c.** Initial gel-based results showed that a relatively high percentage of oxidation in the DCMU condition via comparison of the overall level of oxidized thiols to total thiols (**Fig. 2.3b**). LC-MS/MS analyses were performed to measure the percentages of oxidation of individual Cys-sites in the DCMU condition only. The percentages of oxidation in light and dark conditions were calculated based on the data from the DCMU condition and the relative abundance changes shown in **Fig. 2.2a**. Using this strategy, stoichiometry data were obtained for 1,351 specific Cys-sites (~65% percentage of all Cys-sites from the dynamic measurements in **Fig. 2.2a**). Due to the limitation for stoichiometric measurements of peptides containing multiple Cys sites, these peptides were excluded for further

analysis. **Fig. 2.2c** shows the distribution of Cys-sites peptides in terms of percentage of oxidation for individual Cys sites (**Table 2.2 and Table 2.3**).

Again, one potential source of bias in calculating this percentage is the unknown levels of irreversible oxidation. To address this issue, we compared the levels of total thiols (free and reversibly oxidized thiols) across light, dark, and DCMU conditions, and no obvious change in the levels of total thiols was observed in two independent experiments (left lanes in **Fig. 2.3c** and Fig. **2.3d**). The results suggested that no significant levels of irreversible oxidation occur in both dark and DCMU conditions because the levels of total thiols (excluding irreversibly oxidized thiols) should decrease if the levels of irreversible oxidation in dark or DCMU are high.

As shown in **Fig. 2.2c** the majority of Cys-residues under the light condition were mostly reduced with only 5-20% in the oxidized form. However, the levels of oxidation were significantly increased in dark and DCMU with the majority of Cys-sites at 20-40% oxidation. The overall levels of oxidation support the models that protein thiols are mostly in a reduced state under physiological conditions due to the highly reducing intracellular environment (Bachi et al. 2013), both darkness and DCMU induced a consistent pattern of increased thiol oxidation in the proteome compared to continuous light (**Fig. 2.2c**). The results suggest that a primary redox-mediated mechanism in light/dark cycle can function by altering the thiol redox states. It is interesting to note that DCMU inhibition consistently resulted in a slightly higher level of thiol oxidation than the dark condition. We note that DCMU effectively inhibits electron flow under continuous light, leading to a state similar to the dark condition; however, the DCMU condition does not inhibit photo-oxidation of PSII and oxygen generation completely. The slightly higher level of thiol oxidation of DCMU condition is presumably due to the photo-oxidation, which is absent in the dark condition.

2.3.4 Total Protein Abundances

To verify that the observed redox changes were not due to the changes in total protein abundances, we performed global proteome profiling across these conditions. Aliquots of all protein samples before enrichment were digested, and subjected to LC-MS/MS analyses. Among the total of 413 proteins with spectra count data, no significant abundance changes were observed for the vast majority of the proteins in **Table 2.4**. Among them, 273 proteins overlapped with those detected with oxidized Cys-residues in **Table 2.2**. Only 7 proteins displayed a consistent increase in abundances (with $p<0.05$) similar to the observed redox changes. The overall results suggest that the short 2 h dark incubation or DCMU inhibition is not long enough to induce significant protein abundance changes as compared to previous large-scale diurnal cycle studies. Such global diurnal proteome effects in cyanobacterium *Cyanothece* sp. ATCC 51142 (Stockel et al. 2011) induced significant changes in the abundance of only ~70 proteins among the total identified ~1,200 proteins in the first 1-3 h of the dark phase during the diurnal cycle. On the other hand, genomic studies for the dynamic transcriptional profiling of light-to-dark transition in *Synechocystis* 6803 has revealed 387 genes of the full-genome (3,169 genes) responded to light exposure within 2 h (Gill et al. 2002), and the expression of 64 genes significantly changed in response to DCMU (Hihara et al. 2003). Interestingly, we have observed a number of transcriptional regulators (**Table 2.5**) are potentially redox-regulated, which may facilitate initial changes in gene expression followed by changes in protein expression at later time points.

2.3.5 Functional Implications of Redox-sensitive Proteins

To further narrow down the redox-sensitive Cys-sites that displayed significant dynamic changes (>1.5-fold) between light and dark conditions, we applied >20% oxidation in dark or DCMU conditions as an additional criterion for selecting those Cys-sites as more likely to induce

a physiological effect (Leichert et al. 2008). This criterion is based on the general notion that the potential physiological impact is larger when the extent of protein being modified is higher. This additional filtering decreased the original list of ~1,100 sites with significant changes to ~600 redox-sensitive sites (from 428 proteins) (**Table 2.3**). Among the 428 identified redox-sensitive proteins, only ~100 proteins in *Synechocystis* 6803 were reported with evidence of redox relevance as interacting with thioredoxin (Mata-Cabana et al. 2007; Perez-Perez et al. 2006); however, direct evidence of oxidation and site information were not available for most of these proteins.

To gain an overall picture of the pathways and processes that are potentially redox-regulated, functional analysis based on Gene Ontology and KEGG pathway information was performed for these redox-sensitive proteins. These proteins were observed to be broadly involved in various biological processes or molecular functions, including amino acid biosynthesis, glycolysis, PET, carbon fixation, and many enzyme classes such as oxidoreductase, ligase, and hydrolase (**Fig. 2.4**). The identified redox-sensitive proteins within the context of several key biological processes including PET, carbon fixation, glycolysis and the Krebs cycle are illustrated in **Fig. 2.5**. The observation of a number of enzymes as redox-sensitive in these processes is consistent with the existing knowledge on functional significance of disulfide proteome in these biological processes (Lindahl et al. 2011), and our quantitative site-specific data provide potential functional sites in the enzymes. **Fig. 2.6** further illustrates the percentages of thiol oxidation of selected proteins in these biological processes, demonstrating the consistent increases in thiol oxidation in response to dark or DCMU and different degrees of relative changes for different proteins.

Our results revealed that several proteins within the PET chain, including the phycobilisome complex and PSII, were redox-sensitive and were in a more reduced state under continuous light compared to the other two conditions. This finding is consistent with the knowledge that protein thiols play a significant role in the redox regulation of the PET chain through the ferredoxin/thioredoxin system (Buchanan and Balmer 2005; Alfonso et al. 2000). Moreover, the redox regulation of the PET chain is known to play a key role in modulating the redox homeostasis and gene expression of cyanobacteria (Alfonso et al. 2000; Rochaix 2013). Among the identified redox-sensitive proteins in photosynthesis (**Figs. 2.5a** and **2.6a**), α-subunits of the ATP synthase complex (AtpA) were identified as thioredoxin targets in *Synechocystis* 6803 (Lindahl and Kieselbach 2009; Lindahl and Florencio 2003). A number of the observed redox-sensitive proteins, e.g. photosystem II manganese-stabilizing polypeptide (PsbO), and F-type H+-transporting ATPase gamma subunit (AtpG), were also reported as thioredoxin targets in *Arabidopsis thaliana* (Lindahl and Kieselbach 2009). The redox changes observed in several phycobiliproteins in this study suggest that the redox changes on Cys-residues play a role in the assembly or stability of the phycobilisome complex.

A number of key enzymes involved in the Calvin cycle, glycolysis, and Krebs cycle have been previously reported as potentially redox-regulated (Lindahl et al. 2011). Our study not only identified the putative redox-regulated enzymes in these processes, but also the potentially functional redox-sensitive Cys-sites. As shown in **Fig. 2.5b and Fig. 2.6b**, most of the enzymes in the Calvin cycle were observed to be redox sensitive, confirming previous knowledge that the entire Calvin cycle is under the control of thioredoxin in both chloroplast and cyanobacteria thylakoid (Lindahl et al. 2011). The thiols in most enzymes were more reduced in the light condition (**Fig. 2.6b**) compared to dark, supporting higher enzyme activities for carbon fixation

in the light condition. Moreover, our results provide furthers insights into the redox regulation on these enzymes. For example, fructose-1,6-bisphosphatase (FBPase) in chloroplasts has been known to undergo light/dark redox transitions for many years (Schurmann 2003). However, the redox regulation of its cyanobacterial counterpart, the bifunctional fructose 1,6-bisphosphatase/sedoheptulose-1,7-biphosphatase (FBPase/SBPase, Slr2094 in *Synechocystis* 6803), has not been well established. For instance, one study reported that the FBPase activity was not regulated by light intensities (Yan and Xu 2008). However, an earlier study reported that the FBPase activity in *Synechococcus* sp. strain 6301 was under a delicate control of the oxidizing and reducing conditions (Udvardy et al. 1982). Similarly, the enzyme phosphoribulokinase (Prk) has been found to be regulated by light/dark and activated through ferredoxin/thioredoxin system in chloroplast of plants (Buchanan 1991); however, it has not been confirmed to undergo light/dark modulation in cyanobacteria, although its activity *in vitro* was observed to be regulated by reversible thiol oxidation based on H_2O_2 and 5,5-dithiobis (2-nitrobenzoic acid) treatments (Kobayashi et al. 2003). Our results clearly show light/dark modulation of the thiol oxidation on these enzymes, providing evidence of potential light/dark modulation of these enzyme activities. Additionally, a number of enzymes were observed as redox-sensitive in glycolysis and the Krebs cycle (**Fig. 2.5c and Fig. 2.6c**), supporting the role of redox-regulation in these metabolic processes. For example, aconitase and isocitrate dehydrogenase were found to be associated with thioredoxin in higher plants, indicating the potential redox regulation of their activities (Balmer et al. 2004). Currently, it is still unclear whether redox-regulation of many potentially redox-sensitive enzymes exists in cyanobacteria under physiological relevant conditions. Further studies need to be carried out in order to fully determine how redox regulation happens in the cells.

2.3.6 Site-specific Redox Sensitivity Predicts Functional Sites

A unique aspect of this study is the proteome-wide site-specific redox sensitivity information for Cys-residues in response to physiological perturbations. One question is whether such site-specific data provides any predictive value for identifying potential functionally active Cys-residues. To examine this predictive potential, we compiled the existing data from identified proteins with known functional Cys-sites (either annotated active Cys-sites or functional information based on site-directed mutagenesis) in *Synechocystis* 6803 as listed in **Table 2.6**. Out of 14 proteins, the functional sites of 12 proteins (8 selected proteins listed in **Table 2.7** were observed to be either the most sensitive or the second most sensitive Cys-residues to light/dark modulation for that particular protein. This provides a level of confidence in linking site-specific redox sensitivity with functional Cys-residues, implying the predictability of the site-specific redox sensitivity data for identifying potential functional sites. Such predictability is in good agreement with results from several recent reactivity profiling studies in other organisms (Deng et al. 2013; Weerapana et al. 2010).

Several examples illustrating the correlation of redox sensitivity with functional sites are shown in **Fig. 2.7**. $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase (Gap2 or GAPDH, GeneBank: X83564) is a classic redox-sensitive enzyme that is specific to the Calvin cycle and non-photosynthetic gluconeogenesis (Koksharova et al. 1998). The functional domain of GAPDH is highly conserved and it is a well-known redox switch in cellular metabolism (Brandes et al. 2009). Our site-specific redox sensitivity data implies that Cys-154 and Cys-158 were the most sensitive to light/dark modulation (**Fig. 2.7a**). The observed similar redox sensitivity for both the active site Cys-154 and adjacent site Cys-158 is consistent with several previously reports on other organisms (Su et al. 2014; Deng et al. 2013). Gap2 in *Synechococcus elongatus* (PDB code: 2D2I) shares an 80% homology with *Synechocystis* 6803 and is conserved

at the region containing the active sites. As shown in **Fig. 2.7b**, the nicotinamide ring of cofactor $NADP^+$ faces towards the catalytic Cys-154 in Gap2 (Kitatani et al. 2006). Previous mutational studies have shown that the catalytic Cys of Gap2 can be oxidized and inactivated by oxidative stress (Su et al. 2013; Deng et al. 2013). To confirm the functional Cys-residues in *Synechocystis* 6803 Gap2, recombinant Gap2 mutants C154S, C154S/C158S and C292S were generated and tested for catalytic activity. Interestingly, all the mutations completely abolished Gap2 enzyme activity (**Fig. 2.7c**), demonstrating that all redox active Cys-sites are crucial for Gap2 enzymatic function. It is surprising to observe that C292S also abolished enzyme activity since it has not been previously reported as a functional site. For this reason, we examined the protein expression levels for these mutants by SDS-PAGE and observed that mutant C292S is essentially not detected (**Fig. 2.7c**), suggesting that Cys-292 is structurally relevant for protein stability, with C292S possibly causing a rapid degradation of the translated protein product. Purified recombinant Gap2 was also treated with various concentrations of H_2O_2 or DTT. The catalytic activity of Gap2 was significantly reduced after exposure to H_2O_2, while increased by DTT (**Fig. 2.7d**), which is in agreement with previous reports (Kitatani et al. 2006; Deng et al. 2013).

Ferredoxin $NADP^+$ reductase (PetH, FNR) is the last enzyme in PET chain for transferring electrons from reduced ferredoxin to NADPH, which serves as the electron donor for the Calvin cycle. PetH is known to be bound to the phycobilisome complexes of *Synechocystis* 6803(van Thor et al. 1999). The activity of PetH is enhanced in light; however, the mechanism of the enzyme activation is not fully elucidated (Talts et al. 2007). Site-specific redox data (**Fig. 2.7e**) shows high redox sensitivity for Cys-231 and Cys-236, and both residues were found within the known FAD binding region. Mutation of Cys-231 and Cys-236 to serine in spinach PetH resulted in decreased enzyme activity due to the altered kinetic properties (Aliverti et al.

1993), supporting a prediction of redox alterations in Cys-231 and Cys-236 as functionally relevant mechanisms for regulating PetH activity in *Synechocystis* 6803. **Fig. 2.7f** shows PetH in *Synechococcus* sp. PCC 7002 (PDB code: 2B5O), which shares a 76% amino acid sequence identity similar with *Synechocystis* 6803 and with a highly conserved FAD binding region. Cys-231 of PetH locates in the nucleotide binding region for FAD.

2.3.7 Novel Functional Cys-sites in Redox Regulatory Proteins

Based upon the predictive observations linking redox sensitivity and functional Cys-residues (**Table 6**), we then attempted to determine whether our data would provide valuable information in the identification of novel functional sites of redox regulatory proteins. One protein of interest is peroxiredoxin, encoded by *sll1621* (GeneBank: Bak49948.1) in the AhpC/TSA family. Peroxiredoxins have been well-recognized as major players in the antioxidant defense system and the redox regulatory network of the plant and cyanobacterial cells (Dietz 2011), and they are conserved markers of circadian rhythms (Edgar et al. 2012). It is known that reduced Sll1621 plays a critical role in protection against photo-oxidative stress, especially under high light conditions (Kobayashi et al. 2004). A *Synechocystis* 6803 *sll1621* deletion mutant showed a severely reduced growth rate compared to wild-type under light (Kobayashi et al. 2004; Hosoya-Matsuda et al. 2005). Since Sll1621 is one of the most abundant and active *Synechocystis* 6803 peroxiredoxin isoforms (Perez-Perez et al. 2009), Sll1621 likely represents most of the peroxiredoxin activity in maintaining redox homeostasis for *Synechocystis* 6803 growth; however, its catalytic activity and functional sites are still unknown. We observed that Cys-155 and Cys-162 are the most redox-sensitive sites among the identified three oxidized Cys **(Fig. 2.8a).** Recombinant protein of *Synechocystis* 6803 Sll1621 and the mutants of C55S, C162S, and C155S/C162S were generated to determine the functions of selected Cys-sites of

Sll1621. The enzyme activity data (**Fig. 2.8b**) confirms that Cys-155 is a potentially novel functional site for the peroxiredoxin activity, whereas both Cys-55 and Cys-162 have no effect on the enzymatic activity. The lack of activity reduction for mutant C162S implies that this residue is possibly exposed to capture redox sensitivity but is not directly related to (or required for) peroxiredoxin activity. Also, the marked effects of DTT or H_2O_2 on its activity were also observed (**Fig. 2.8c**), confirming its roles in redox response. To further examine whether Cys-155 is a conserved residue across difference species, we blasted the sequence motif in UniProt database and confirmed that the sequence motif "DNCP" containing Cys-155 in Sll1621 is indeed conserved among ~50 cyanobacteria strains and up to 70 bacterial strains (**Fig. 2.9**), strongly supporting the functional importance of Cys-155.

Another protein of interest is Glucose-6-Phosphate 1-Dehydrogenase (G6PDH, Zwf, GeneBank: NP_440771.1), which is the first enzyme in the dark-initiated pentose phosphate pathway. Zwf has been reported to be activated by Cys oxidation in dark conditions in plants (Nee et al. 2009); however, the specific Cys-sites responsible for the activation of Zwf are not known in *Synechocystis* 6803. Our data indicates that Cys-265 and Cys-445 are highly sensitive to oxidation and are potential functional sites of Zwf (**Fig. 2.8a**). Additionally, Cys-445 is a conserved Cys-residue among multiple strains of cyanobacteria, such as *Anabaena*, *Nostoc* and *Synechococcus* (Newman et al. 1995). We created Zwf mutants of C187S, C265S and C445S for functional activity assays. Results showed that all three Cys-sites play a role in maintaining Zwf activity with Cys-445 playing the most significant role **(Fig. 2.8d),** which is consistent with the redox sensitivity data. Interestingly, further assays involving the addition of H_2O_2 or DTT did not affect the enzymatic activity of recombinant Zwf (data not shown), showing evidence that Zwf alone may not be sensitive to a general redox status. It has been reported that Zwf is not sensitive

to redox regulation without its allosteric activator (OpcA) present (Hagen and Meeks 2001). This provides a potentially interesting scenario where either Cys-445, 265, and/or 187 are likely functionally involved in Zwf/OpcA interactions facilitating redox regulation of the pentose phosphate pathway and further studies will be necessary to confirm the role of these Cys-residues in redox regulation.

2.4 Discussion

Cyanobacteria are oxygenic photosynthetic prokaryotes that are believed to share a common ancestor with higher plant chloroplasts. As in all photosynthetic organisms, redox regulation via light/dark diurnal cycles is essential for photosynthesis, metabolism, and gene expression, starting from the PET chain through the ferredoxin/thioredoxin system to all downstream target proteins. Reversible oxidation of protein Cys-residues, an important type of PTMs, is a major mechanism of redox regulation where thiols serve as redox switches (Buchanan and Balmer 2005; Lindahl et al. 2011). Despite the significance, the overall knowledge of site-specific redox modifications on Cys, especially under physiological conditions, is still very limited due to the constraint of measurement technologies. A recent study on thiol-based redox modifications under physiological conditions of *C. elegans* only identified <200 Cys-sites as oxidized using OxICAT (Knoefler et al. 2012), illustrating the challenge of identifying redox modification. This work provides a proteome-wide, quantitative and site-specific analysis of light/dark modulated changes of thiol oxidation in cyanobacteria covering ~2,100 Cys-residues, significantly expanding the current repertoire of thiol-based redox modifications under physiological conditions.

Our approach adapted an optimized resin-assisted enrichment approach for achieving site-specific quantification of Cys oxidation across multiple conditions. In addition to the relative

quantification of oxidation levels, our approach was also easily modified to quantify the stoichiometry or fractional occupancy of oxidation of specific Cys-sites, which is particularly important for studying PTMs since functional PTMs are likely to have higher fractional occupancies of modifications in certain biological conditions. We have shown the advantages of enrichment specificity, proteome coverage, and multiplexed quantification for measuring site-specific redox changes (**Fig. 2.1b**) compared to existing methods such as OxiICAT, other biotin-based approaches (Leichert et al. 2008; Forrester et al. 2009), or anti-TMT antibody-based approaches (Pan et al. 2014; Qu et al. 2014; Murray et al. 2012). We, and others, have also shown the potential of applying this resin-assisted approach for studying other types of reversible Cys redox modifications (e.g., S-nitrosylation, S-glutathionylation)(Guo et al. 2014; Su et al. 2014; Forrester et al. 2009).

Significantly, the advantages of this approach enabled quantitative profiling of Cys oxidation in a physiological light-dark environment with in-depth proteome coverage of the Cys redox modifications in cyanobacteria. The observed landscape of dynamic redox changes of ~2,100 Cys-sites from 1,060 proteins (nearly one-third of the genome) is much broader than one would initially expect from light-dark modulation. However, given the fact that tens of thousands of modification sites have been identified in other PTMs such as phosphorylation (Olsen et al. 2010), ubiquitylation (Mertins et al. 2013), and acetylation (Lundby et al. 2012), we assert that this study likely represents the beginning of the expansion on the potential sites of thiol-based redox modifications. Nevertheless, we recognize that there are still several technical limitations of the current approach. First, since this approach measures the level of total reversible thiol oxidation, the exact form of oxidation is not known. However, our approach can be adapted to measure the exact forms of modification such as SSG or SNO by coupling selective reduction

with resin-assisted enrichment as recently described (Guo et al. 2014). Second, our approach does not measure the levels of irreversible oxidation such as S-sulfinic or S-sulfonic acids, which may exist in dark or DCMU conditions. The potential irreversible oxidation could complicate the data interpretation of reversible oxidation; however, the levels of irreversible oxidation under physiological conditions are most likely to be insignificant (based on the results shown in **Fig. 2.3c and Fig. 2.3d**).

It is important to note that the current findings demonstrate a redox rheostat in the cyanobacteria after light-to-dark transition in a global proteome scale, i.e., oxidation and reduction of protein thiols do not function as an all-or-nothing phenomenon. Such dynamic redox changes are actually reflected more by partial activation or inactivation of specific proteins, given that the number of specific proteins reduced/oxidized differs between conditions. This is an interesting observation as even in an environment so regimented as a light/dark cycle, we clearly still detect specific regulatory sites in both the oxidized and reduced forms. The site-specific dynamic data provide important information on redox-sensitive sites for proteins involved in photosynthesis and metabolism (**Fig. 2.5** and **Fig. 2.6**). Many specific enzymes in the Calvin and Krebs cycles, as well as other biosynthetic pathways, were previously identified as thioredoxin-regulated through disulfide formation (Lindahl and Kieselbach 2009). Most of these enzymes are mainly in the activated state with reduced thiols under light condition, and become inactivated by disulfide bond formation under dark condition (Buchanan and Balmer 2005), consistent with our observations.

Similar to other PTM studies, significant challenges remain associated with deciphering which specific proteins and their respective Cys-sites are most likely to be directly (or indirectly) involved in protein function/regulation from such global proteomic data based upon the redox

changes observed in such a relatively straightforward light/dark system. A broad coverage of thiol-based redox modifications is a prerequisite for identifying novel redox regulated proteins or functional sites; however, the large number of available sites is also a challenge for determining functional redox sensitivity. One unique aspect of this study is that the redox sensitivity data for individual Cys-residues provides important predictive information for identifying functional or key regulatory sites for given proteins. The examination of proteins with known functional sites showed good predictability (**Fig. 2.7** and **Table 2.7**). Moreover, the stoichiometry or fractional occupancy data on individual Cys-sites provided additional evidence of functional relevance. Multiple site-mutagenesis and biochemical studies (**Fig. 2.7** and **Fig. 2.8**) provided a range of redox regulatory scenarios showing a proof-of-principle for both confirmation and identification of novel functional sites utilizing the redox data. Therefore, with the redox sensitivity and stoichiometry information on individual sites, this dataset should serve a unique resource for identifying functional sites for redox-regulated proteins as well as for novel redox proteins. Follow-up targeted site-specific mutagenesis and pertinent functional studies will help to elucidate the mechanisms of redox regulation and its impact on protein function, and specific pathways in response to environmental stimuli.

2.5 Acknowledgements

Portions of this work were supported by the DOE Early Career Research Award (to W.J.Q.), DOE grant number DE-FG02-99ER20350 (to H.B.P.), the Environmental Molecular Science Laboratory (EMSL) Research Campaign project, and the DOE Office of Biological and Environmental Research Genome Sciences Program under the Pan-omics project. A.N. has been supported by an NSF Graduate Research Fellowship. Portions of Experimental work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific

user facility at PNNL in Richland, Washington. PNNL is operated by Battelle for the DOE under contract DE-AC05-76RLO-1830.

2.5.1 Author Contributions

W.J.Q. and H.B.P. conceived and supervised the project; J.G., A.Y.N., Z.D., D.S., M.J.G. and R.J.M. performed the experiments and mass spectrometry analyses; J.G., A.Y.N, J.M.J., M.E.M. and W.J.Q. analyzed the data; D.W.K. and R.D.S. contributed to the development of the measurement capabilities used; J.G., A.Y.N., Z.D., H.B.P., and W.J.Q. wrote the manuscript.

2.5.2 Conflict of Interest Statement

The authors declare no competing financial interests.

2.6 References

Alfonso, M., Perewoska, I., Kirilovsky, D. (2000). "Redox control of *psbA* gene expression in the cyanobacterium *Synechocystis* PCC 6803. Involvement of the cytochrome b_6f complex." *Plant Physiology*. **122**(2): 505-516.

Aliverti, A., Piubelli, L., Zanetti, G., Lubberstedt, T., Herrmann, R.G., Curti, B. (1993). "The role of cysteine residues of spinach ferredoxin-$NADP^+$ reductase As assessed by site-directed mutagenesis." *Biochemistry*. **32**(25): 6374-6380.

Bachi, A., Dalle-Donne, I., Scaloni, A. (2013) Redox proteomics: chemical principles, methodological approaches and biological/biomedical promises. *Chemical Reviews*. **113**(1): 596-698.

Balmer, Y., Vensel, W.H., Tanaka, C.K., Hurkman, W.J., Gelhaye, E., Rouhier, N., Jacquot, J.P., Manieri, W., Schurmann, P., Droux, M., Buchanan, B.B. (2004). "Thioredoxin links redox to the regulation of fundamental processes of plant mitochondria." *Proceedings of the National Academy of Sciences*. **101**(8): 2642-2647.

Brandes, N., Reichmann, D., Tienson, H., Leichert, L.I., Jakob, U. (2011). "Using quantitative redox proteomics to dissect the yeast redoxome." *Journal of Biological Chemistry*. **286**(48): 41893-41903.

Brandes, N., Schmitt, S., Jakob, U. (2009). "Thiol-based redox switches in eukaryotic proteins." *Antioxidants & Redox Signaling*. **11**(5): 997-1014.

Buchanan, B.B. (1991). "Regulation of CO_2 assimilation in oxygenic photosynthesis: the ferredoxin/thioredoxin system. Perspective on its discovery, present status, and future development." *Archives of Biochemistry and Biophysics*. **288**(1): 1-9.

Buchanan, B.B., Balmer, Y. (2005). "Redox regulation: a broadening horizon." *Annual Review of Plant Biology*. **56**: 187-220.

Cameron, J.C., Pakrasi, H.B. (2010). "Essential role of glutathione in acclimation to environmental and redox perturbations in the cyanobacterium *Synechocystis* sp. PCC 6803." *Plant Physiology*. **154**(4): 1672-1685.

D'Alessandro, A., Rinalducci, S., Zolla, L. (2011). "Redox proteomics and drug development." *Journal of Proteomics*. **74**(12): 2575-2595.

Dai, S., Johansson, K., Miginiac-Maslow, M., Schurmann, P., Eklund, H. (2004). "Structural basis of redox signaling in photosynthesis: structure and function of ferredoxin:thioredoxin reductase and target enzymes." *Photosynthesis Research*. **79**(3): 233-248.

Dalle-Donne, I., Rossi, R., Colombo, G., Giustarini, D., Milzani, A. (2009). "Protein S-glutathionylation: a regulatory device from bacteria to humans." *Trends in Biochemical Sciences*. **34**(2): 85-96.

Deng, X., Weerapana, E., Ulanovskaya, O., Sun, F., Liang, H., Ji, Q., Ye, Y., Fu, Y., Zhou, L., Li, J., Zhang, H., Wang, C., Alvarez, S., Hicks, L.M., Lan, L., Wu, M., Cravatt, B.F., He, C. (2013). "Proteome-wide quantification and characterization of oxidation-sensitive cysteines in pathogenic bacteria." *Cell Host & Microbe*. **13**(3): 358-370.

Derakhshan, B., Wille, P.C., Gross, S.S. (2007). "Unbiased identification of cysteine S-nitrosylation sites on proteins." *Nature Protocols*. **2**(7): 1685-1691.

Dietz, K.J. (2011). "Peroxiredoxins in plants and cyanobacteria." *Antioxidants & Redox Signaling*. **15**(4): 1129-1159.

Edgar, R.S., Green, E.W., Zhao, Y., van Ooijen, G., Olmedo, M., Qin, X., Xu, Y., Pan, M., Valekunja, U.K., Feeney, K.A., Maywood, E.S., Hastings, M.H., Baliga, N.S., Merrow, M., Millar, A.J., Johnson, C.H., Kyriacou, C.P., O'Neill, J.S., Reddy, A.B. (2012). "Peroxiredoxins are conserved markers of circadian rhythms." *Nature*. **485**(7399): 459-464.

Falcon, L.I., Magallon, S., Castillo, A. (2010). "Dating the cyanobacterial ancestor of the chloroplast." *The ISME Journal*. **4**(6): 777-783.

Forrester, M.T., Foster, M.W., Stamler, J.S. (2007). "Assessment and application of the biotin switch technique for examining protein S-nitrosylation under conditions of pharmacologically induced oxidative stress." *Journal of Biological Chemistry*. **282**(19): 13977-13983.

Forrester, M.T., Thompson, J.W., Foster, M.W., Nogueira, L., Moseley, M.A., Stamler, J.S. (2009). "Proteomic analysis of S-nitrosylation and denitrosylation by resin-assisted capture." *Nature Biotechnology*. **27**(6): 557-559.

Foyer, C.H., Noctor, G. (2009). "Redox regulation in photosynthetic organisms: signaling, acclimation, and practical implications." *Antioxidants & Redox Signaling*. **11**(4): 861-905.

Fratelli, M., Demol, H., Puype, M., Casagrande, S., Eberini, I., Salmona, M., Bonetto, V., Mengozzi, M., Duffieux, F., Miclet, E., Bachi, A., Vandekerckhove, J., Gianazza, E., Ghezzi, P. (2002). "Identification by redox proteomics of glutathionylated proteins in oxidatively stressed human T lymphocytes." *Proceedings of the National Academy of Sciences.* **99**(6): 3505-3510.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. III., Smith, H.O. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nature Methods*. **6**(5): 343-345.

Gill, R.T., Katsoulakis, E., Schmitt, W., Taroncher-Oldenburg, G., Misra, J., Stephanopoulos, G. (2002). "Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. strain PCC 6803." *Journal of Bacteriology*. **184**(13): 3671-3681.

Greco, T.M., Hodara, R., Parastatidis, I., Heijnen, H.F., Dennehy, M.K., Liebler, D.C., Ischiropoulos, H. (2006). "Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells." *Proceedings of the National Academy of Sciences*. **103**(19): 7420-7425.

Guo, J., Gaffrey, M.J., Su, D., Liu, T., Camp, D.G. II., Smith, R.D., Qian, W.J. (2014). "Resin-assisted enrichment of thiols as a general strategy for proteomic profiling of cysteine-based reversible modifications." *Nature Protocols*. **9**(1): 64-75.

Hagen, K.D., Meeks, J.C. (2001). "The unique cyanobacterial protein OpcA is an allosteric effector of glucose-6-phosphate dehydrogenase in *Nostoc punctiforme* sp. ATCC 29133." *Journal of Biological Chemistry*. **276**(15): 11477-11486.

Held, J.M., Danielson, S.R., Behring, J.B., Atsriku, C., Britton, D.J., Puckett, R.L., Schilling, B., Campisi, J., Benz, C.C., Gibson, B.W. (2010). Targeted quantitation of site-specific cysteine oxidation in endogenous proteins using a differential alkylation and multiple reaction monitoring mass spectrometry approach." *Molecular & Cellular Proteomics*. **9**(7): 1400-1410.

Held, J.M., Gibson, B.W. (2012). "Regulatory control or oxidative damage? Proteomic approaches to interrogate the role of cysteine oxidation status in biological processes." *Molecular & Cellular Proteomics*. **11**(4): 1-14.

Hess, D.T., Matsumoto, A., Kim, S.O., Marshall, H.E., Stamler, J.S. (2005). "Protein S-nitrosylation: purview and parameters." *Nature Reviews Molecular Cell Biology*. **6**(2): 150-166.

Hihara, Y., Sonoike, K., Kanehisa, M., Ikeuchi, M. (2003). "DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803." *Journal of Bacteriology*. **185**(5): 1719-1725.

Hosoya-Matsuda, N., Motohashi, K., Yoshimura, H., Nozaki, A., Inoue, K., Ohmori, M., Hisabori, T. (2005). "Anti-oxidative stress system in cyanobacteria. Significance of type II peroxiredoxin and the role of 1-Cys peroxiredoxin in *Synechocystis* sp. strain PCC 6803." *Journal of Biological Chemistry*. **280**(1):840-846.

Huang da, W., Sherman, B.T., Lempicki, R.A. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." *Nature Protocols*. **4**(1): 44-57.

Izquierdo-Alvarez, A., Martinez-Ruiz, A. (2011). "Thiol redox proteomics seen with fluorescent eyes: the detection of cysteine oxidative modifications by fluorescence derivatization and 2-DE." *Journal of Proteomics*. **75**(2): 329-338.

Jaffrey, S.R., Erdjument-Bromage, H., Ferris, C.D., Tempst, P., Snyder, S.H. (2001). "Protein S-nitrosylation: a physiological signal for neuronal nitric oxide." *Nature Cell Biology*. **3**(2): 193-197.

Jaffrey, S.R., Snyder, S.H. (2001). "The biotin switch method for the detection of S-nitrosylated proteins." *Science's STKE*. **2001**(86): 1-9.

Kelly, R.T., Page, J.S., Luo, Q., Moore, R.J., Orton, D.J., Tang, K., Smith, R.D. (2006). "Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry." *Analytical Chemistry*. **78**(22): 7796-7801.

Kim, S., Gupta, N., Pevzner, P.A. (2008). "Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases." *Journal of Proteome Research*. **7**(8): 3354-3363.

Kitatani, T., Nakamura, Y., Wada, K., Kinoshita, T., Tamoi, M., Shigeoka, S., Tada, T. (2006). "Structure of NADP-dependent glyceraldehyde-3-phosphate dehydrogenase from *Synechococcus* PCC 7942 complexed with NADP." *Acta Crystallographica Section F, Structural Biology and Crystallization Communications*. **62**(Pt 4): 315-319.

Knoefler, D., Thamsen, M., Koniczek, M., Niemuth, N.J., Diederich, A.K., Jakob, U. (2012). "Quantitative in vivo redox sensors uncover oxidative stress as an early event in life." *Molecular Cell*. **47**(5): 767-776.

Kobayashi, D., Tamoi, M., Iwaki, T., Shigeoka, S., Wadano, A. (2003). "Molecular characterization and redox regulation of phosphoribulokinase from the cyanobacterium *Synechococcus* sp. PCC 7942." *Plant Cell Physiology*. **44**(3): 269-276.

Kobayashi, M., Ishizuka, T., Katayama, M., Kanehisa, M., Bhattacharyya-Pakrasi, M., Pakrasi, H.B., Ikeuchi, M. (2004). "Response to oxidative stress involves a novel peroxiredoxin gene in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803." *Plant & Cell Physiology*. **45**(3): 290-299.

Koksharova, O., Schubert, M., Shestakov, S., Cerff, R. (1998). "Genetic and biochemical evidence for distinct key functions of two highly divergent GAPDH genes in catabolic and anabolic carbon flow of the cyanobacterium *Synechocystis* sp. PCC 6803." *Plant Molecular Biology*. **36**(1): 183-194.

Leichert, L.I., Gehrke, F., Gudiseva, H.V., Blackwell, T., Ilbert, M., Walker, A.K., Strahler, J.R., Andrews, P.C., Jakob, U. (2008). "Quantifying changes in the thiol redox proteome upon oxidative stress *in vivo*." *Proceedings of the National Academy*. **105**(24): 8197-8202.

Leichert, L.I., Jakob, U. (2004). "Protein thiol modifications visualized *in vivo*." *PLOS Biology*. **2**(11): 1723-1737.

Leonard, S.E., Carroll, K.S. (2011). "Chemical 'omics' approaches for understanding protein cysteine oxidation in biology." *Current Opinion in Chemical Biology*. **15**(1): 88-102.

Lin, D., Li, J., Slebos, R.J., Liebler, D.C. (2010). "Cysteinyl peptide capture for shotgun proteomics: global assessment of chemoselective fractionation." *Journal of Proteome Research*. **9**(10): 5461-5472.

Lindahl, M., Florencio, F.J. (2003). "Thioredoxin-linked processes in cyanobacteria are as numerous as in chloroplasts, but targets are different." *Proceedings of the National Academy of Sciences*. **100**(26): 16107-16112.

Lindahl, M., Kieselbach, T. (2009). "Disulphide proteomes and interactions with thioredoxin on the track towards understanding redox regulation in chloroplasts and cyanobacteria." *Journal of Proteomics*. **72**(3): 416-438.

Lindahl, M., Mata-Cabana, A., Kieselbach, T. (2011). "The disulfide proteome and other reactive cysteine proteomes: analysis and functional significance." *Antioxidants & Redox Signaling*. **14**(12): 2581-2642.

Liu, T., Qian, W.J., Chen, W.N., Jacobs, J.M., Moore, R.J., Anderson, D.J., Gritsenko, M.A., Monroe, M.E., Thrall, B.D., Camp, D.G. II., Smith, R.D. (2005). "Improved proteome coverage by using high efficiency cysteinyl peptide enrichment: the human mammary epithelial cell proteome." *Proteomics*. **5**(5): 1263-1273.

Liu, T., Qian, W.J., Strittmatter, E.F., Camp, D.G. II., Anderson, G.A., Thrall, B.D., Smith, R.D. (2004). "High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology." *Analytical Chemistry*. **76**(18): 5345-5353.

Lopez-Maury, L., Sanchez-Riego, A.M., Reyes, J.C., Florencio, F.J. (2009). "The glutathione/glutaredoxin system is essential for arsenate reduction in *Synechocystis* sp. strain PCC 6803." *Journal of Bacteriology*. **191**(11): 3534-3543.

Lundby, A., Lage, K., Weinert, B.T., Bekker-Jensen, D.B., Secher, A., Skovgaard, T., Kelstrup, C.D., Dmytriyev, A., Choudhary, C., Lundby, C., Olsen, J.V. (2012). "Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns." *Cell Reports*. **2**(2): 419-431.

Mata-Cabana, A., Florencio, F.J., Lindahl, M. (2007). "Membrane proteins from the cyanobacterium *Synechocystis* sp. PCC 6803 interacting with thioredoxin." *Proteomics*. **7**(21): 3953-3963.

McDonagh, B. (2009). "Diagonal electrophoresis for detection of protein disulphide bridges." *Methods in Molecular Biology*. **519**: 305-310.

Menon, B.R., Davison, P.A., Hunter, C.N., Scrutton, N.S., Heyes, D.J. (2010). "Mutagenesis alters the catalytic mechanism of the light-driven enzyme protochlorophyllide oxidoreductase." *Journal of Biological Chemistry*. **285**(3): 2113-2119.

Mertins, P., Qiao, J.W., Patel, J., Udeshi, N.D., Clauser, K.R., Mani, D.R., Burgess, M.W., Gillette, M.A., Jaffe, J.D., Carr, S.A. (2013). "Integrated proteomic analysis of post-translational modifications by serial enrichment." *Nature Methods*. **10**(7): 634-637.

Murray, C.I., Uhrigshardt, H., O'Meally, R.N., Cole, R.N., Van Eyk, J.E. (2012). "Identification and quantification of S-nitrosylation by cysteine reactive tandem mass tag switch assay." *Molecular & Cellular Proteomics*. **11**(2): 1-12.

Navrot, N., Finnie, C., Svensson, B., Hagglund, P. (2011). "Plant redox proteomics." *Journal of Proteomics*. **74**(8): 1450-1462.

Nee, G., Zaffagnini, M., Trost, P., Issakidis-Bourguet, E. (2009). "Redox regulation of chloroplastic glucose-6-phosphate dehydrogenase: a new role for f-type thioredoxin." *FEBS Letters*. **583**(17): 2827-2832.

Nelson, K.J., Klomsiri, C., Codreanu, S.G., Soito, L., Liebler, D.C., Rogers, L.C., Daniel, L.W., Poole, L.B. (2010). "Use of dimedone-based chemical probes for sulfenic acid detection methods to visualize and identify labeled proteins." *Methods in Enzymology*. **473**: 95-115.

Newman, J., Karakaya, H., Scanlan, D.J., Mann, N.H. (1995). "A comparison of gene organization in the *zwf* region of the genomes of the cyanobacteria *Synechococcus* sp. PCC 7942 and *Anabaena* sp. PCC 7120." *FEMS Microbiology Letters*. **133**(1-2): 187-193.

Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O., Thiele, I. (2012). "Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis." *Proceedings of the National Academy of Sciences*. **109**(7): 2678-2683.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., Brunak. S., Mann, M. (2010). "Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis." *Science Signaling*. **3**(104): 1-16.

Pan, K.T., Chen, Y.Y., Pu, T.H., Chao, Y.S., Yang, C.Y., Bomgarden, R.D., Rogers, J.C., Meng, T.C., Khoo, K.H. (2014). "Mass spectrometry-based quantitative proteomics for dissecting multiplexed redox cysteine modifications in nitric oxide-protected cardiomyocyte under hypoxia." *Antioxidants & Redox Signaling*. **20**(9): 1365-1381.

Paulsen, C.E., Truong, T.H., Garcia, F.J., Homann, A., Gupta, V., Leonard, S.E., Carroll, K.S. (2012). "Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity." *Nature Chemical Biology*. **8**(1): 57-64.

Perez-Perez, M.E., Florencio, F.J., Lindahl, M. (2006). "Selecting thioredoxins for disulphide proteomics: target proteomes of three thioredoxins from the cyanobacterium *Synechocystis* sp. PCC 6803." *Proteomics*. **6**(Suppl 1): S186-195.

Perez-Perez, M.E., Mata-Cabana, A., Sanchez-Riego, A.M., Lindahl, M., Florencio, F.J. (2009). "A comprehensive analysis of the peroxiredoxin reduction system in the cyanobacterium *Synechocystis* sp. strain PCC 6803 reveals that all five peroxiredoxins are thioredoxin dependent." *Journal of Bacteriology*. **191**(24): 7477-7489.

Pfannschmidt, T., Brautigam, K., Wagner, R., Dietzel, L., Schroter, Y., Steiner, S., Nykytenko, A. (2009). "Potential regulation of gene expression in photosynthetic cells by redox and energy state: approaches towards better understanding." *Annals of Botany*. **103**(4): 599-607.

Qian, W.J., Kaleta, D.T., Petritis, B.O., Jiang, H., Liu, T., Zhang, X., Mottaz, H.M., Varnum, S.M., Camp, D.G. II., Huang, L., Fang, X., Zhang, W.W., Smith, R.D. (2008). "Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy." *Molecular & Cellular Proteomics*. **7**(10): 1963-1973.

Qian, W.J., Liu, T., Monroe, M.E., Strittmatter, E.F., Jacobs, J.M., Kangas, L.J., Petritis, K., Camp, D.G. II., Smith, R.D. (2005). "Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome." *Journal of Proteome Research*. **4**(1): 53-62.

Qu, Z., Meng, F., Bomgarden, R.D., Viner, R.I., Li, J., Rogers, J.C., Cheng, J., Greenlief, C.M., Cui, J., Lubahn, D.B., Sun, G.Y., Gu, Z. (2014). "Proteomic quantification and site-mapping of S-nitrosylated proteins using isobaric iodoTMT reagents." *Journal of Proteome Research*. **13**(7): 3200-3211.

Rinalducci, S., Murgiano, L., Zolla, L. (2008). "Redox proteomics: basic principles and future perspectives for the detection of protein oxidation in plants." *Journal of Experimental Botany*. **59**(14): 3781-3801.

Rochaix, J.D. (2013). "Redox regulation of thylakoid protein kinases and photosynthetic gene expression." *Antioxidants & Redox Signaling*. **18**(16): 2184-2201.

Rosgaard, L., de Porcellinis, A.J., Jacobsen, J.H., Frigaard, N.U., Sakuragi, Y. (2012). "Bioengineering of carbon fixation, biofuels, and biochemicals in cyanobacteria and plants." *Journal of Biotechnology*. **162**(1): 134-147.

Sadler, N.C., Melnicki, M.R., Serres, M.H., Merkley, E.D., Chrisler, W.B., Hill, E.A., Romine, M.F., Kim, S., Zink, E.M., Datta, S., Smith, R.D., Beliaev, A.S., Konopka, A., Wright, A.T. (2014). "Live cell chemical profiling of temporal redox dynamics in a photoautotrophic cyanobacterium." *ACS Chemical Biology*. **9**(1): 291-300.

Sato, Y., Inaba, K. (2012). "Disulfide bond formation network in the three biological kingdoms, bacteria, fungi and mammals." *FEBS Journal*. **279**(13): 2262-2271.

Schurmann, P. (2003). "Redox signaling in the chloroplast: the ferredoxin/thioredoxin system." *Antioxidants & Redox Signaling*. **5**(1): 69-78.

Seo, Y.H., Carroll, K.S. (2009). "Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies." *Proceedings of the National Academy of Sciences*. **106**(38): 16163-16168.

Sethuraman, M., McComb, M.E., Heibeck, T., Costello, C.E., Cohen, R.A. (2004). "Isotope-coded affinity tag approach to identify and quantify oxidant-sensitive protein thiols." *Molecular & Cellular Proteomics*. **3**(3): 273-278.

Singh, A.K., Elvitigala, T., Cameron, J.C., Ghosh, B.K., Bhattacharyya-Pakrasi, M., Pakrasi, H.B. (2010). "Integrative analysis of large scale expression profiles reveals core transcriptional response and coordination between multiple cellular processes in a cyanobacterium." *BMC Systems Biology*. **4**(105): 1-14.

Singh, R.K., Tiwari, S.P., Rai, A.K., Mohapatra, T.M. (2011). "Cyanobacteria: an emerging source for drug discovery." *Journal of Antibiotics*. **64**(6): 401-412.

Stockel, J., Jacobs, J.M., Elvitigala, T.R., Liberton, M., Welsh, E.A., Polpitiya, A.D., Gritsenko, M.A., Nicora, C.D., Koppenaal, D.W., Smith, R.D., Pakrasi, H.B. (2011). "Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium *Cyanothece* 51142." *PLOS One*. **6**(2): 1-11.

Su, D., Gaffrey, M.J., Guo, J., Hatchell, K.E., Chu, R.K., Clauss, T.R., Aldrich, J.T., Wu, S., Purvine, S., Camp, D.G. II., Smith, R.D., Thrall, B.D., Qian, W.J. (2014). "Proteomic identification and quantification of S-glutathionylation in mouse macrophages using resin-assisted enrichment and isobaric labeling." *Free Radical Biology & Medicine*. **67**: 460-470.

Su, D., Shukla, A.K., Chen, B., Kim, J.S., Nakayasu. E., Qu, Y., Aryal, U., Weitz, K., Clauss, T.R., Monroe, M.E., Camp, D.G. II., Bigelow, D.J., Smith, R.D., Kulkarni, R.N., Qian, W.J. (2013). "Quantitative site-specific reactivity profiling of S-nitrosylation in mouse skeletal muscle using cysteinyl peptide enrichment coupled with mass spectrometry." *Free Radical Biology & Medicine*. **57**: 68-78.

Talts, E., Oja, V., Ramma, H., Rasulov, B., Anijalg, A., Laisk, A. (2007). "Dark inactivation of ferredoxin-NADP reductase and cyclic electron flow under far-red light in sunflower leaves." *Photosynthesis Research*. **94**(1): 109-120.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., Hamon, C. (2003). "Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS." *Analytical Chemistry*. **75**(8): 1895-1904.

Townley, H.E., Sessions, R.B., Clarke, A.R., Dafforn, T.R., Griffiths, W.T. (2001). "Protochlorophyllide oxidoreductase: a homology model examined by site-directed mutagenesis." *Proteins.* **44**(3): 329-335.

Trebst, A. (2007). "Inhibitors in the functional dissection of the photosynthetic electron transport system." *Photosynthesis Research*. **92**(2): 217-224.

Tsukamoto, Y., Fukushima, Y., Hara, S., Hisabori, T. (2013). "Redox control of the activity of phosphoglycerate kinase in *Synechocystis* sp. PCC 6803." *Plant & Cell Physiology*. **54**(4): 484-491.

Udvardy, J., Godeh, M.M., Farkas, G.L. (1982). "Regulatory properties of a fructose 1,6-bisphosphatase from the cyanobacterium *Anacystis nidulans*." *Journal of Bacteriology*. **151**(1): 203-208.

van Thor, J.J., Gruters, O.W., Matthijs, H.C., Hellingwerf, K.J. (1999). "Localization and function of ferredoxin:NADP(+) reductase bound to the phycobilisomes of *Synechocystis*." *EMBO Journal*. **18**(15): 4128-4136.

Weerapana, E., Wang, C., Simon, G.M., Richter, F., Khare, S., Dillon, M.B., Bachovchin, D.A., Mowen. K., Baker, D., Cravatt, B.F. (2010). "Quantitative reactivity profiling predicts functional cysteines in proteomes." *Nature*. **468**(7325): 790-795.

Winterbourn, C.C., Hampton, M.B. (2008). "Thiol chemistry and specificity in redox signaling." *Free Radical Biology & Medicine*. **45**(5): 549-561.

Yan, C., Xu, X. (2008). "Bifunctional enzyme FBPase/SBPase is essential for photoautotrophic growth in cyanobacterium *Synechocystis* sp. PCC 6803." *Progress in Natural Science*. **18**(2): 149-153.

Table 2.1: Oligonucleotides used for site-specific mutagenesis of selected cysteine sites in G6PDH, Gap2 and Sll1621 proteins.

Oligoucleotide name[#]	Oligoucleotide
G6PDH [parent (1/4)[*], C187S(1/2, ¾)[*], C265S (1/5, 6/5)[*] and C445S (1/7, 8/4)[*]]	
1-pET6xHN	TGATGACGATGATAAGGCGGTAACGCTACTCGAAAATC
2-G6PRW1	TTTCTTTGCTGACACTCTGCACTACCCG
3-G6PFW1	AGAGTGTCAGCAAAGAAAATCAGGTGTATCGCATC
4-pET6xHN	GCCAACTCAGCTTCCTCTAAAGTCGGCGCCAGCG
5-G6PRW2	GGTGAGACTAAACAGTTGCATGAGGTG
6-G6PFW2	CAACTGTTTAGTCTCACCGCCATGGACC
7-G6PRW3	CCCAACATGGAATCTAATAAAAGGCGGTGATAG
8-G6PFW3	TTTATTAGATTCCATGTTGGGAGACCAAACCC
Gap2 [parent (9/12)[*], C154S (9/10,11/12)[*], C154S/C158S (9/13,14/15)[*] and C292S (16/17, 18/19)[*]]	
9-pET6xHN	TGATGACGATGATAAGGCGACTAGAGTAGCAATTAACGG
10-GP2RW1	GGTAGTACTACTAGCGTTACTAATTACTTCG
11-GP2FW1	ACGCTAGTAGTACTACCAACTGCCTCG
12-pET6xHN	GCCAACTCAGCTTCCTCTATTTCCAGTTTTTAGCCAC
13-GP2RW2	GGCGAGGCTGTTGGTAGTACTACTAGCGTTACTAATTACTTCG
14-GP2FW2	ACGCTAGTAGTACTACCAACAGCCTCGCCCCATTCGGCA
15-pET6xHN	GCCAACTCAGCTTCCTCTATTTCCAGTTTTTAGCCACAATTTCAGCCAAG
16PET6XHN	TGATGACGATGATAAGGCGACTAGAGTAGCAATTAACGGATTTGGAC
17GPRW3	AGAGCTATCCGTGCCCCGGAAGTC
18GPFW3	TTCCGGGGCACGGATAGCTCTTCCACTGTGGATGGT
19PET6XHN	GCCAACTCAGCTTCCTCTATTTCCAGTTTTTAGCCACAATTTCAGC
Sll1621 [parent (20/23)[*], C55S(20/21,22/23)[*], C155S/C162S(20/24,25/26)[*] and C162S (20/27,28/23)[*]]	
20-pET6xHN	TGATGACGATGATAAGGCGACCCCCGAACGAGTTCC
21-TRRW1	GTTAGAAGAAGAGGTGGGGGTGAAGGCAC
22-TRFW2	CCCCACCTCTTCTTCTAACCACTTGCC
23-pET6xHN	GCCAACTCAGCTTCCTTTAGCCGACAAAAGCTTTAAC
24-TRRW2	ACACTCGAAGGGATCTACGGGAGAGTTGTCACCGA

	ACTCAGGCT
25-TRFW3	AGCCTGAGTTCGGTGACAACTCTCCCGTAGATCCC TTCGAGTGT
26-TRRW3	GCCAACTCAGCTTCCTTTAGCCGACAAAAGCTTTA ACGGGCT
27-TRRW4	TGTCAGCGTCGGAAGACTCGAAGGGATCTACGGGA GAGTTGTCACCGAACTCAGGCTCA
28-TRFW4	TCGGTGACAACTCTCCCGTAGATCCCTTCGAGTCTT CCGACGCTGACACCAT

*: oligionucleotide pair for PCR;

#: the list number of oligonucleotide shown for pairing.

Table 2.2: Total identified Cys-sites along with peptide sequences.

Table 2.3: Cys-sites with significant redox changes (abundance ratios >1.50, percentage of oxidation >20%).

Table 2.4: Total protein abundances from global proteome profiling.

Table 2.5: Redox-regulated transcriptional regulators or signaling proteins.

Table 2.6: Previously identified protein targets for *Synechocystis* thioredoxin A or glutaredoxin

Tables 2.2- 2.6 are available online at:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256482/bin/supp_13_12_3270__index.html

Table 2.7: The levels of relative Cys oxidation on individual Cys-sites for selected proteins with known active or functional Cys-sites.

Gene ID	Accession	Protein name	Cys-sites	% Oxidation (Reversible)	Ratios of oxidation levels[a]	
					Dark/ Light	DCMU /Light
sll0080	P54899	N-acetyl-gamma-glutamyl-phosphate reductase (argC)	151		10.9	22.7
			151, 154[#]		3.3	42.8
			154[#]		3.9	5.0
			86		2.8	3.5
sll0427	P10549	Photosystem II manganese-stabilizing polypeptide (psbO)	48[@]	75%	1.4	2.8
			73	86%	1.5	2.4
sll1342	P80505	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase (gap2)	154[#]		3.4	4.1
			154[#], 158		2.6	2.8
			158		2.6	2.7
			292	31%	1.8	1.7
slr0394	P74421	Phosphoglycerate kinase (pgk)	216	11%	2.2	2.3
			314[@]	19%	3.0	4.5
			94		1.8	1.8

			94, 97		1.3	1.3
			97		1.8	1.8
slr0506	Q59987	Light-dependent NADPH-protochlorophyllide oxidoreductase (por)	37[@]	22%	2.3	2.7
			133	21%	1.3	1.5
			226[@]	21%	3.7	7.5
slr0527	Q55468	Transcription regulator ExsB homolog (queC)	190[@]		1.8	2.7
slr1463	P28371	Elongation factor EF-G (fusA)	105[@]	25%	2.2	2.6
			388	28%	1.3	1.2
			676	30%	2.0	2.2
			68	28%	2.2	2.7
slr1643	Q55318	Ferredoxin-$NADP^+$ oxidoreductase (petH)	213	21%	1.8	2.4
			231[@]		3.4	4.6
			231[@],236		2.7	3.5
			236		3.2	3.1
			371	27%	1.5	1.7

[a] Ratios of Cys oxidation levels (calculated for each condition against light condition).

[#] Active sites of the enzymes.

[@] Functional sites verified by site-directed mutagenesis studies (Tsukamoto et al. 2013; Townley et al. 2001; Menon et al. 2010; Aliverti et al. 1993).

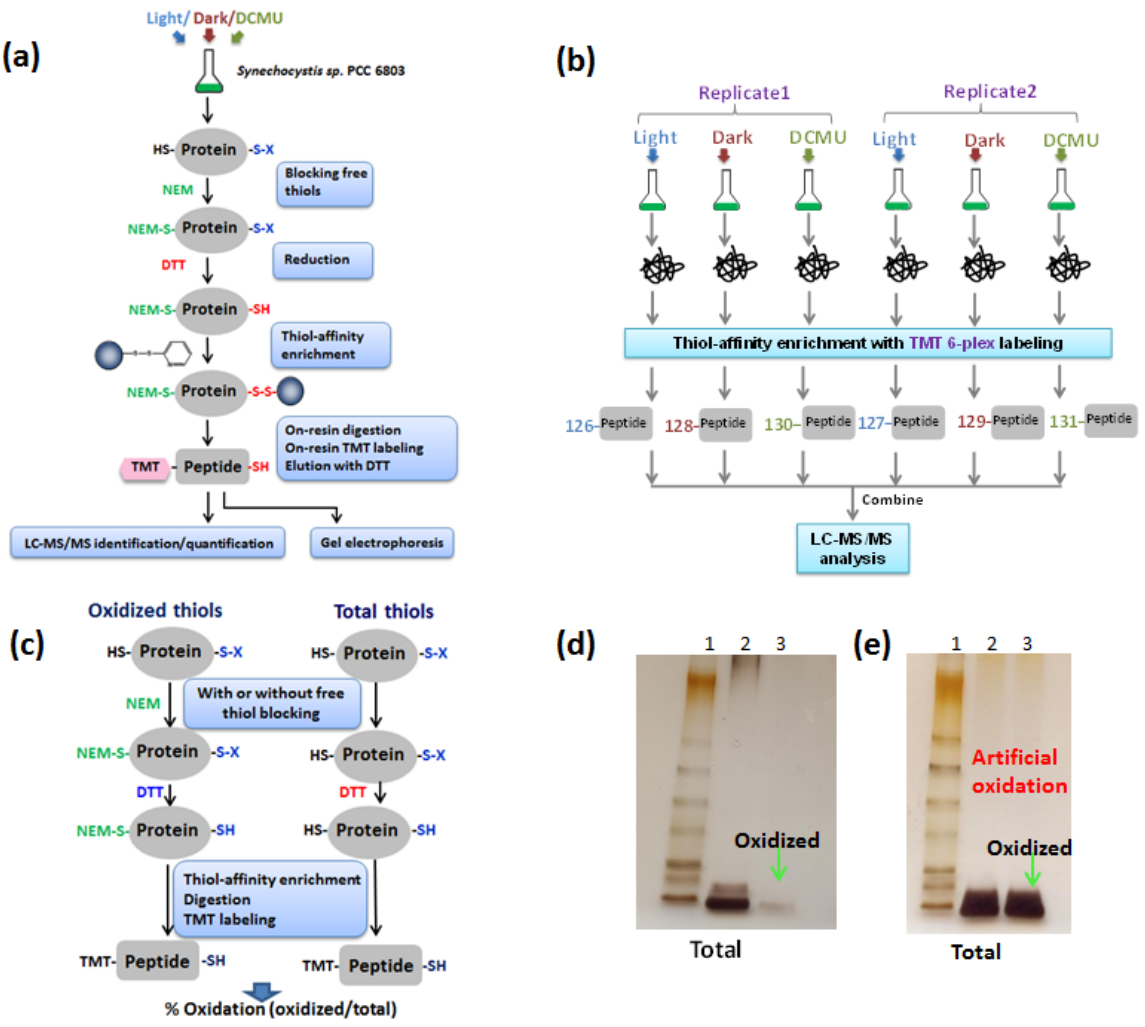


Figure 2.1: Overview of the enrichment and site-specific quantification strategy for thiol oxidation.

a Enrichment and processing workflow by the resin-assisted approach. Note that S-X denotes oxidized thiols. Proteins were extracted from different conditions and free thiols were blocked by NEM alkylation. Oxidized Cys were reduced by DTT and captured by Thiopropyl Sepharose resin. On-resin protein digestion and TMT isobaric labeling of enriched Cys-peptides were carried out, subsequently followed by DTT elution of the enriched peptides for LC-MS/MS

analysis. **b** 6-plex quantitative strategy for profiling Cys redox dynamics. *Synechocystis* were cultured under continuous light, or switched to dark for 2 h, or treated with 10 μM DCMU in light for 2 h. Proteins were extracted and processed to enrich the oxidized Cys-peptides accordingly, and enriched peptides were labeled with TMT reagents with different reporter tags (126-131). The six labeled samples were combined to facilitate MS/MS-based 6-plex quantification. **c** Workflow for quantifying the percentage of reversible Cys oxidation. Equal amounts of protein samples were processed in parallel both with NEM blocking (oxidized thiols) and without NEM blocking (oxidized thiols). Total thiols (free plus reversibly oxidized thiols) are enriched in the sample reduced by DTT without NEM blocking. **d** Gel image of enriched total Cys-containing peptides and oxidized Cys peptides for the samples processed by the incubation with 100 mM NEM before bead-beating. Samples were from the light condition. Lane 1: standard protein ladder; Lane 2: total enriched Cys-peptides; Lane 3: enriched oxidized Cys-peptides. **e** Gel image of enriched total Cys-containing peptides and oxidized Cys peptides for the samples processed with bead-beating prior to incubating with 100 mM NEM. Lane 1: standard protein ladder; Lane 2: total enriched Cys-peptides; Lane 3: enriched oxidized Cys-peptides.

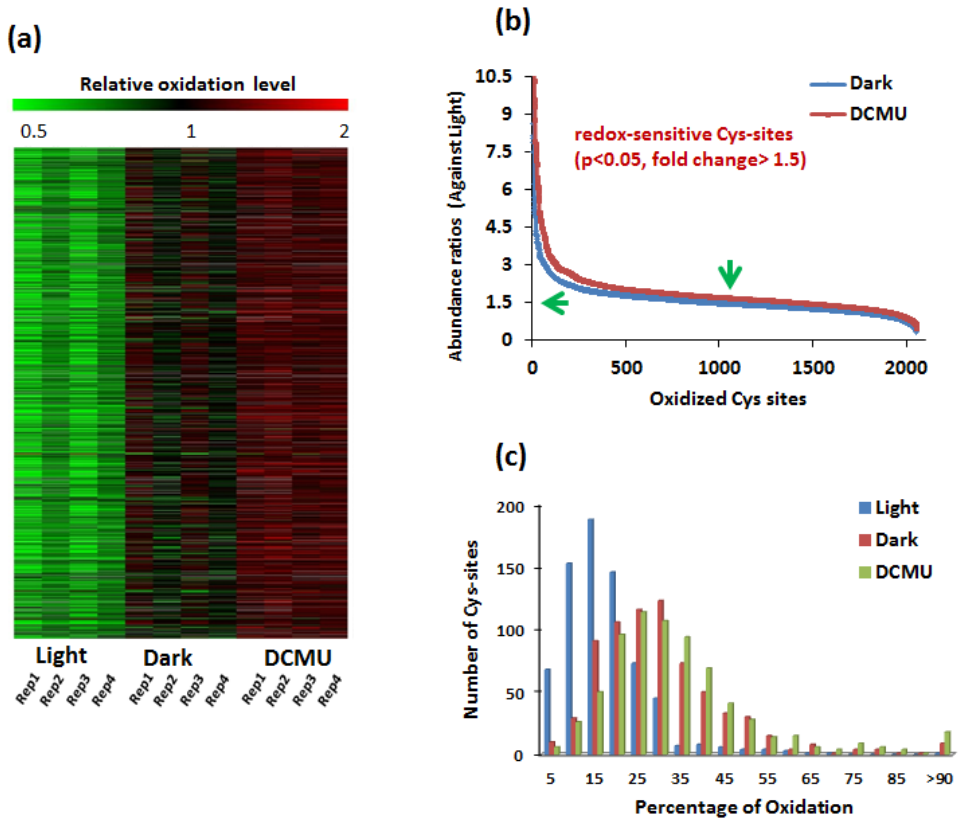


Figure 2.2: Broad Cys redox changes modulated by light/dark.

a The relative oxidation levels of ~2000 identified Cys-sites under light, dark or DCMU conditions. The relative oxidation levels were calculated by dividing the total reporter ion intensity from each TMT channel for a given Cys-site against the average intensity across all conditions for the same Cys-site. **b** The redox sensitivity of individual Cys-sites. The y-axis abundance ratios are expressed as the ratios of oxidation levels under dark or DCMU conditions divided by those under the light condition. Peptides were considered confident redox-sensitive if all four data points were observed ($p<0.05$ and fold change >1.5-fold). **c** Distribution of Cys-sites in terms of the percentage of Cys oxidation for individual Cys-residues. Histograms show a shift to higher levels of oxidation for dark and DCMU compared to light.

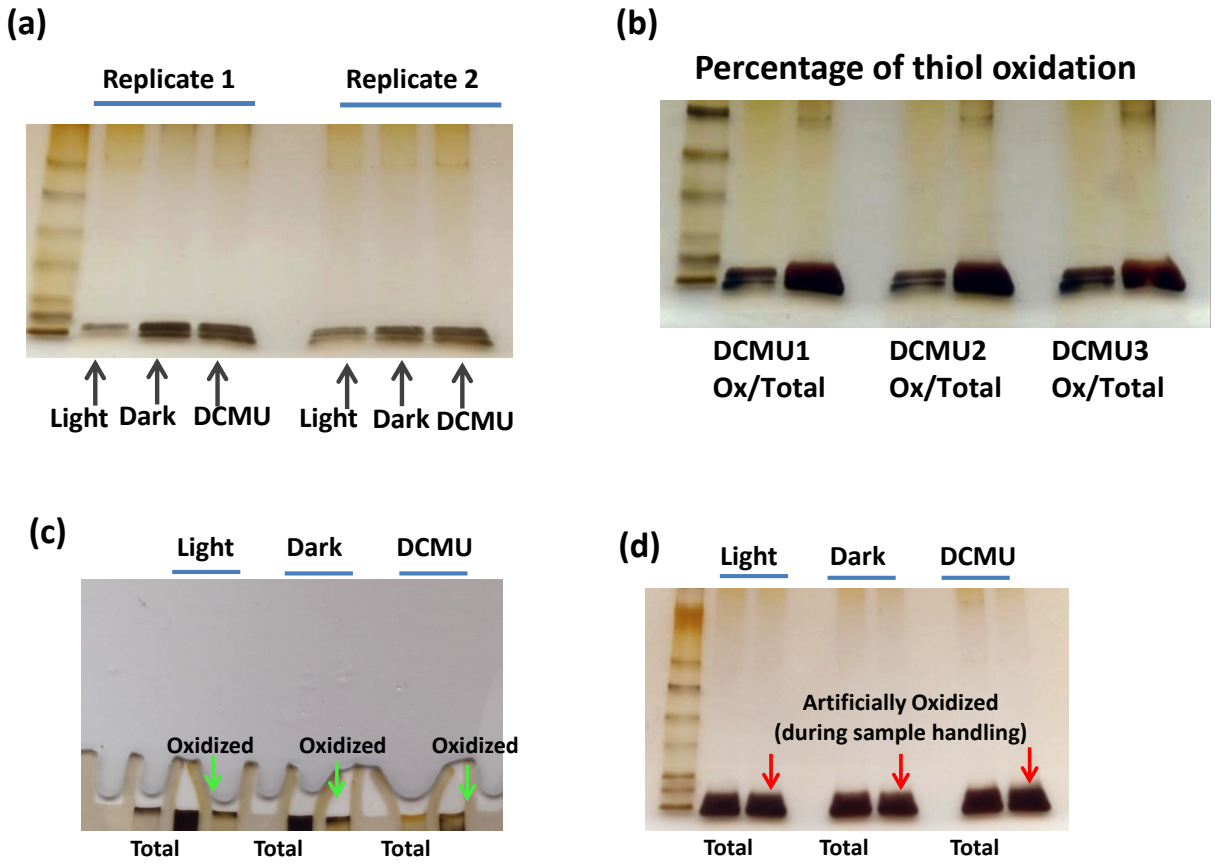


Figure 2.3: SDS-PAGE gel images of oxidized Cys-peptides.

a SDS-PAGE gel image of the enriched oxidized Cys-peptides in light, dark, and DCMU conditions. The intensities of peptide bands indicate an increase in the oxidation levels in dark and DCMU. **b** SDS-PAGE gel image showing the levels of oxidized cysteine-peptides versus total cysteine-peptides from 3 biological replicates under the DCMU condition. **c** Gel image of enriched total Cys-containing peptides and oxidized Cys peptides for the samples processed by incubating with 100 mM NEM before bead-beating. **d** Gel image of enriched total Cys-containing peptides and artificially oxidized Cys peptides during sample handling. The false levels of oxidation were introduced by bead-beating prior to incubating with 100 mM NEM.

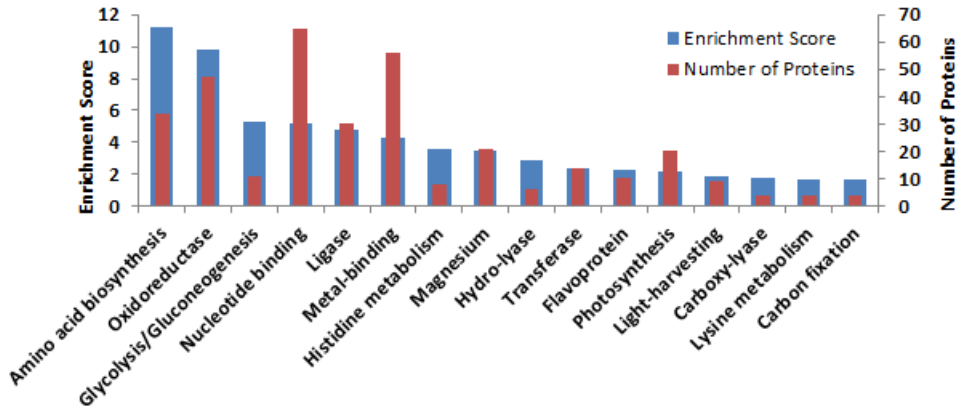


Figure 2.4: Functional categories of the significant oxidized proteins in the dark and DCMU conditions.

Enrichment scores are calculated based on the p-values in a corresponding annotation cluster by DAVID bioinformatics tool, where the p-values associated with annotation terms were calculated by Fisher Exact test. The enrichment scores indicate the ranked biological significance of the functional categories, and the number of proteins indicates the identified redox proteins from each category.

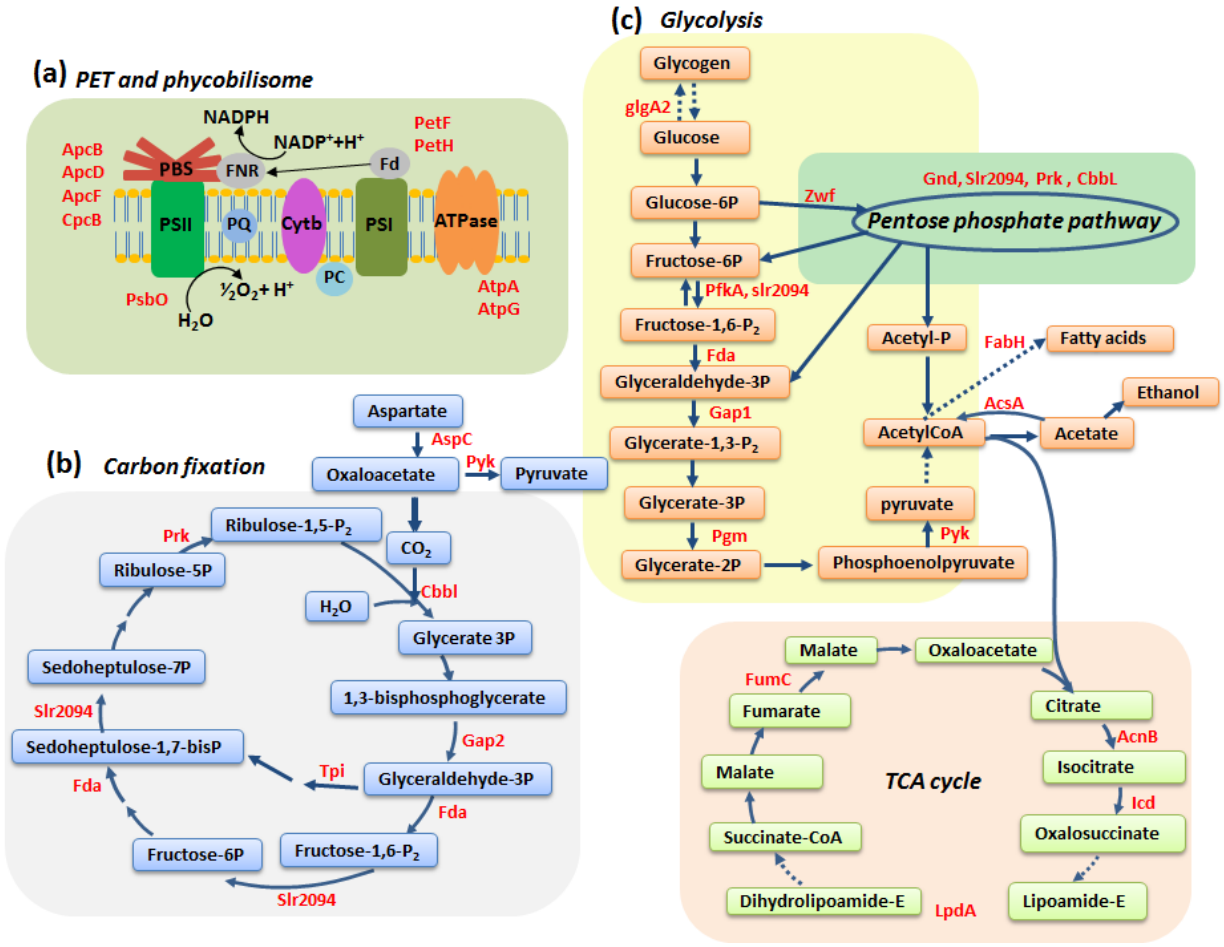


Figure 2.5: Broad redox regulations in cyanobacteria.

Redox-sensitive proteins are involved in a PET, b carbon fixation, and c glycolysis and the Krebs cycle. The identified redox-regulated enzymes are labeled in red.

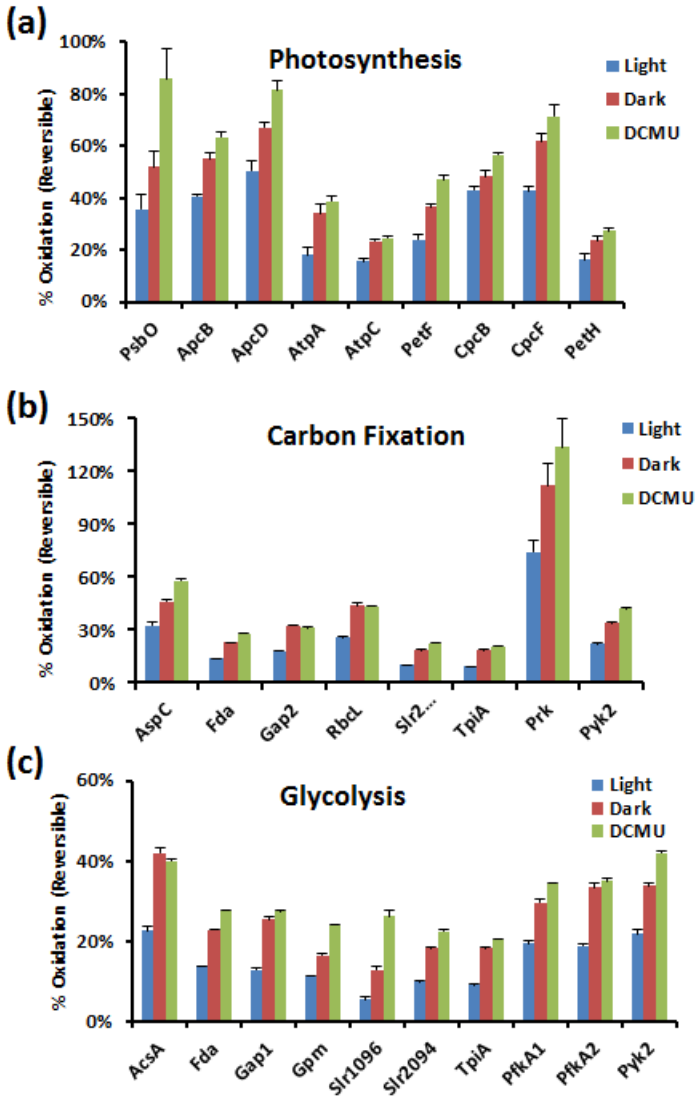


Figure 2.6: Comparison of protein thiol oxdiations levels of individual proteins from different biological processes in light, dark, and DCMU conditions.

Redox-sensitive proteins involved in **a** PET, **b** carbon fixation, and **c** glycolysis were selected. The percent of oxidation for each protein is represented by the Cys-site with the highest stoichiometry of each protein. Error bars are the standard errors from four biological replicates.

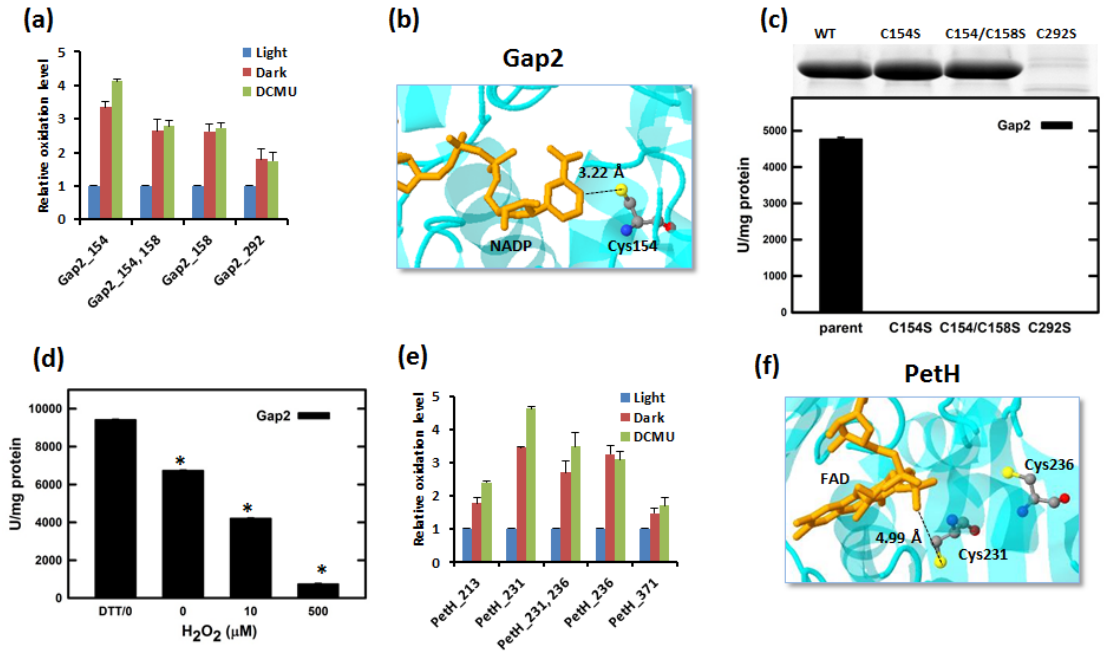


Figure 2.7: Site-specific redox sensitivity and functional Cys-sites.

a Relative oxidation levels of individual Cys-sites of Gap2 as modulated by light, dark, DCMU. **b** Monomer crystal structure of Gap2 (PDB code: 2D2I) in *Synechococcus elongatus* PCC 7942. NADP is shown in stick model; catalytic Cys154 is highlighted in ball-and-stick model. **c** Site-specific mutants of Gap2 abolished its activity. Top: expressed levels of recombinant WT (wild-type), C154S, C154S/C158S, and C292S; Bottom: Gap2 activities of wild-type, C154S, C154S/C158S, and C292S. **d** Redox regulation of Gap2 activity as treated by 10 mM DTT or various concentrations of H_2O_2 for 2 min. The asterisks denote that the differences from the DTT treated are statistically significant ($p < 0.05$). **e** Relative oxidation levels of individual Cys-sites of PetH. **f** Monomer structure of Ferredoxin $NADP^+$ reductase (PetH, FNR) in *Synechococcus* sp. pcc 7002 (PDB code: 2B5O). FAD is shown in stick model, and the FAD binding region Cys231 and Cys236 are highlighted in ball-and-stick model. Quantitative data are represented as the mean ± standard errors.

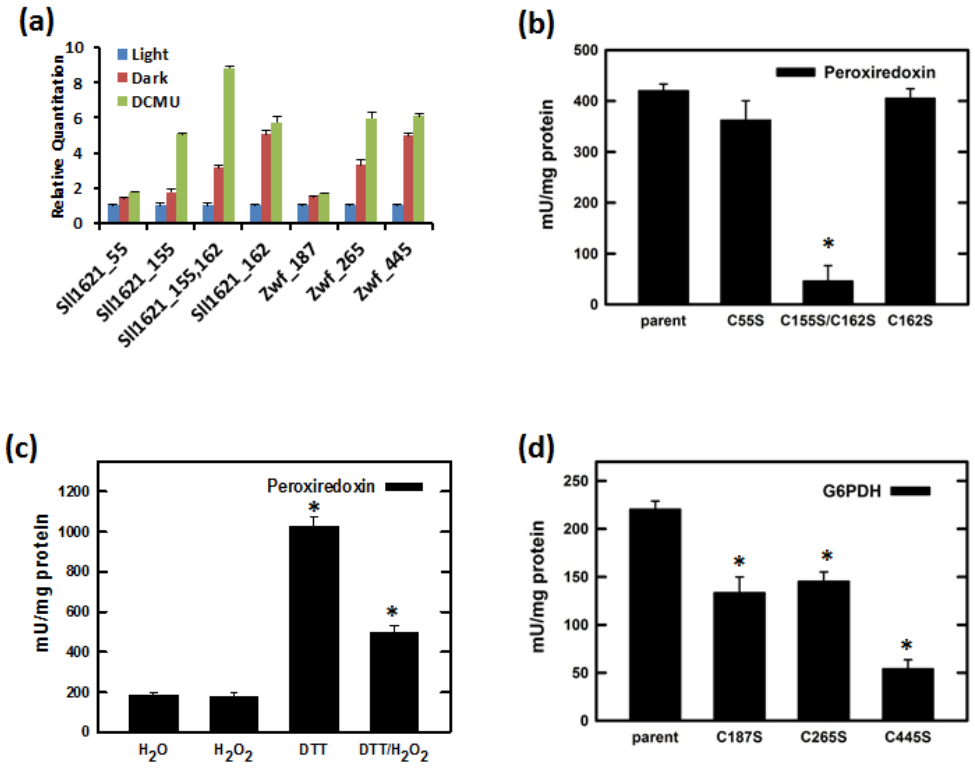


Figure 2.8: Functional sites of peroxiredoxin (Sll1621) and G6PDH (Zwf).

a Relative oxidation levels of individual Cys-sites of Sll1621 and Zwf in light, dark, DCMU. **b** Enzymatic activity of purified peroxiredoxin (Sll1621) wild-type, C55S, C155S/C162S, and C162S mutants. Activity was indicated as mU/ mg protein. mU is defined as nmole/time. **c** Enzymatic activity of purified peroxiredoxin (Sll1621) wild-type treated with water (control), 10 mM DTT for 15 min, and 1 mM H_2O_2 for 15 min after DTT treatment where DTT was removed by buffer exchange. H_2O_2 treatment was performed after the removal of DTT by buffer exchange with phosphate buffer. Same buffer exchange step was applied to DTT treatment prior to activity measurement. **d** Enzymatic activity of purified Zwf (G6PDH) wild-type, C187S, C265S, C445S. All data are represented as the mean ± standard errors. The asterisks denote that the differences from the parent type or untreated are statistically significant ($p < 0.05$).

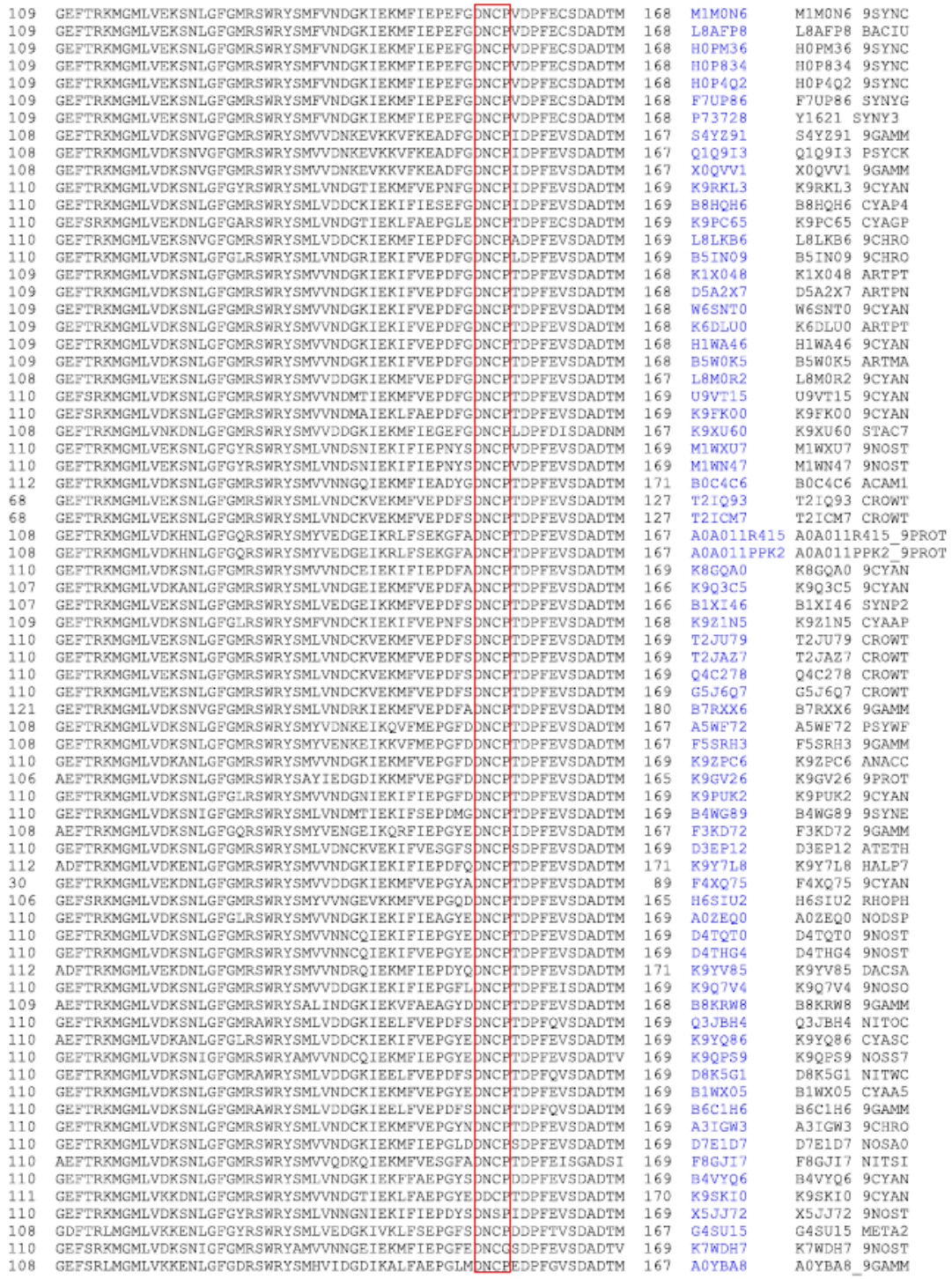


Figure 2.9: Cys-155 in Sll1621 is conserved across many cyanobacterial and other bacterial strains as shown in the 'DNCP' motif.

Chapter 3

The Proteolysis Adaptor, NblA, Binds to the N-terminus of β-Phycocyanin: Implications in the Mechanism of Phycobilisome Degradation

This work will be submitted to *Photosynthesis Research.* Amelia Y. Nguyen*, William P. Bricker, Hao Zhang, Daniel Weisz, Michael Gross, Himadri B. Pakrasi. The Proteolysis Adaptor, NblA, Binds to the N-terminus of β-Phycocyanin: Implications in the Mechanism of Phycobilisome Degradation. *Photosynthesis Research.*

*A.Y.N. designed the project; cloned, overexpressed, and purified C-terminal His tagged NblA; isolated PBS; carried out the binding and cross-linking experiments; analyzed the LC-MS/MS data; and wrote the manuscript.

3.1 Introduction

3.1.1 Abstract

Phycobilisome (PBS) complexes are the massive light-harvesting apparati in cyanobacteria that capture and funnel light energy to the photosystems for conversion into cellular fuel. PBS complexes are dynamically degraded during nutrient deprivation, which causes severe chlorosis, and resynthesized during nutrient repletion. The degradation of PBS complexes occurs rapidly after nutrient step-down, and is specifically triggered by a small protein called non-bleaching protein A (NblA), which serves as a proteolysis adapter that facilitates interactions between a protease and the phycobiliproteins. Little is known about the mode of action of NblA during PBS degradation. In this study, we used a chemical cross-linking coupled LC-MS/MS approach to investigate the interactions between NblA and phycobiliproteins. An isotopically-coded BS^3 cross-linker captured protein interactions formed between immobilized NblA and free phycobiliproteins during pull-down assays. A liquid chromatography and tandem mass spectrometry approach determined the amino acid residues that confer binding between key factors involved in PBS turnover. These results were modeled into existing crystal structures of NblA and PC using protein-docking simulations. Our cross-linking and protein docking results indicate that the K^{52}-T^2 cross-link between NblA and β-phycocyanin (PC) occurs in the central hollow cavity of the PC rods where the C-terminal end of NblA fits in an open groove in β-PC. This finding suggests that the PBS complex undergoes conformational changes during nutrient stress for NblA to gain access to a inaccessible region. In addition, M^1-K^{44} and M^1-K^{52} cross-links between the N-terminus of NblA monomer to the C-terminus of another NblA monomer are consistent with the NblA crystal structure, suggesting that the purified NblA is structurally biologically relevant. This work presents direct evidence that NblA physically interacts with PBS.

3.1.2 Phycobilisome Function

Cyanobacteria are unicellular phototrophs that naturally thrive in diverse terrestrial and marine ecosystems, where they are exposed to scarce or excess levels of nutrients and light levels. The versatility in cyanobacteria's ability to acclimate to environmental changes is due to their ability to rapidly modify their metabolism and dynamically remodel their phycobilisome (PBS) complexes (Grossman et al. 1993) and optimize the utilization of nutrients (Schwarz and Forchhammer 2005). The light-harvesting apparati in cyanobacteria are called PBS complexes (Glazer 1982), which are massive, multi-subunit complexes found on the stromal side of the thylakoid membrane (Bogorad 1975; MacColl 1998). Soluble PBS complexes in cyanobacteria initiate photosynthesis by absorbing and transferring light energy to the reaction centers in both photosystem I (PSI) and photosystem II (PSII) (Liu et al. 2013). The photosystems are found in thylakoid membranes, where photochemical electron transfer reactions occur to convert light energy to chemical energy (Glazer 1982). In nutrient-rich media, cyanobacteria appear blue-green in color owing to blue bilin pigments covalently attached to the PBS complexes and green chlorophyll (Chl) pigments in the photosystems.

3.1.3 Phycobilisome Structure

The phycobiliproteins that make up the PBS assemble into 3-7 megadalton complexes (Adir et al. 2006) that may account up to 50% of the total soluble proteins in the cell; thus serving as a large cellular nitrogen reserve (Bogorad 1975). With the help of colorless linker proteins, the pigment-protein phycobiliproteins associate to form PBS complexes (Elmorjani et al. 1986; MacColl 1998). The major components of the PBS complexes are brilliantly colored red phycoerythrin (PE), blue phycocyanin (PC), and blue-green allophycocyanin (APC) (Grossman et al. 1993; Scheer and Zhao 2008). Structurally, PBS complexes contain three hexamers of six PC rods radiating out from a bi- or tri-cylindrical APC core in a fan-shaped

manner (Grossman et al. 1993; Ajlani and Vernotte 1998; Glazer 1982; Bogorad 1975; Sidler 2004). Given that a monomer of PC is comprised of an α- and β-PC heterodimer (Ajlani and Vernotte 1998), over 100 units of both α- and β-PC are required to form six peripheral PC rods in a PBS complex, highlighting the enormous number of phycobiliproteins that make up one PBS complex.

3.1.4 Dynamic Remodeling of Phycobilisomes

The plasticity of PBS restructuring is best seen during nitrogen, sulfur, or phosphorous depletion (Richaud et al. 2001), where non-diazoptropic cyanobacteria undergo bleaching (Allen and Smith 1969) owing to PC and APC transcript-synthesis repression and protein degradation (Collier and Grossman 1992) to free up macronutrients for maintenance of essential metabolic activities and for photoprotection (Adir et al. 2006). If PBS complexes do not degrade during nutrient deprivation, photosynthetic rates will remain unchanged whereas cellular metabolism slows down, causing excessive energy to be absorbed and harmful radical species to be produced (Adir et al. 2003). Low-level photosynthesis and loss of pigments are crucial for cell survival during nitrogen stress (Sauer et al. 2001). Prolonged deprivation of nitrogen-induced chlorosis in *Synechococcus* sp. PCC 7942 (hereafter referred to as *Synechococcus* 7942) occurs in three phases: 1) rapid decline of phycobiliproteins while chl *a* decreases at a slower rate, 2) progressive chlorophyll *a* and carotenoid decreases, and 3) undetectable levels of pigments and high levels of glycogen (Gorl et al. 1998). Upon exogenous addition of nutrients, cyanobacteria regain their blue-green color because phycobiliproteins are resynthesized (Schwarz and Forchhammer 2005). The processes of PBS degradation and re-synthesis play a key role for cell survival in terms of cell maintenance, growth, and development (Schwarz and Forchhammer 2005).

The remodeling of PBS is an active, reversible, rapid, and specific process that occurs on a massive scale. A popular model of PBS degradation begins with the sequential trimming of the peripheral PC rods, starting at the most distal end, with complete degradation of the remaining PBS occurring two days after continued nutrient depletion in *Synechococcus* 7942 (Collier and Grossman 1994). The degradation of these enormous light-harvesting apparatuses in cyanobacteria is triggered by an ATP-dependent Non-bleaching protein A (NblA) that is 7 kDa and is comprised of 59 amino acids (Collier and Grossman 1994). The *Synechococcus* 7942 *nblA* deletion strain maintains high levels of PBS after nitrogen, sulfur, or phosphorous step down (Collier and Grossman 1994). In most cyanobacteria, levels of *nblA* transcripts and NblA proteins are up-regulated during nitrogen, sulfur, and phosphorous stress (Collier and Grossman 1994; Baier et al. 2001), whereas PBS levels are simultaneously down-regulated (Adir et al. 2006; Baier et al. 2001; Li and Sherman 2002).

NblA is an adapter protein that triggers the degradation of PBS in cyanobacteria (Collier and Grossman 1994; Karradt et al. 2008). A single copy of *nblA* is present in most PBS-containing cyanobacteria and red algae, although there is low similarity between species (i.e., 30% sequence identity) (Baier et al. 2014; Bienert et al. 2005; Dines et al. 2008). Some strains have two copies of *nblA* (e.g., *Nostoc* sp. PCC 7120) (Karradt et al. 2008), which only need the expression of one *nblA* gene for PBS degradation, and *Synechocystis* sp. PCC 6803, for which both genes are required for PBS degradation (Baier et al. 2001). Crystal structures of NblA from *Nostoc* 7120 (Karradt et al. 2008), *Thermosynechococcus vulcanus* (Dines et al. 2008), and *Synechococcus* 7942 (Dines et al. 2008) demonstrate that the NblA monomer has a helix-loop-helix motif that undergoes homo- or hetero-dimerization to form an open four-helix bundle,

which is the basic functional unit of NblA proteins (Bienert et al. 2005; Baier et al. 2014; Dines et al. 2008).

As a proteolysis adaptor, NblA triggers PBS complexes by binding to ClpC and phycobiliproteins on its N-terminus (Karradt et al. 2008) and C-terminus (Bienert et al. 2005), respectively. ClpC functions in protein recognition, unfolding, and chaperoning to the membrane bound ClpP catalytic core for degradation (Olinares et al. 2011; Porankiewicz et al. 1999). The Clp proteasome system serves as an important cellular disposal system. Clp proteins are universally found in eubacteria, including cyanobacteria, and eukaryotes (Andersson et al. 2006; Stanne et al. 2007). Cross-linking studies reveal that ClpC binds to phycobiliproteins, suggesting that the degradation of PBS is regulated by the Clp proteasome system. ClpC is an ATP-dependent, HSP100 chaperon partner that is essential and constitutively expressed in cyanobacteria and plants (Andersson et al. 2006). In *Nostoc* sp. PCC 7120, NblA binding activates ClpC by initiating its oligiomerization into a hetero-hexamer with three NblA dimers (Karradt et al. 2008).

The sequential mechanism of how NblA triggers and relies on the Clp proteasome system to disassemble PBS complexes is poorly understood. Random mutagenesis studies of NblA demonstrate that variable residues residues in the middle part of NblA play a crucial role in mediating PBS degradation *in vivo* (Dines et al. 2008). Combined with the α-helical structural similarities between NblA and phycobiliprotiens, structural mimicry may be the method for NblA binding to the PBS complexes (Dines et al. 2008). Site-directed mutagenesis and pull-down assays, however, highlight several conserved amino acid residues in NblA that specifically bind to ClpC and α-PC (Karradt et al. 2008; Stanne et al. 2007; Dines et al. 2008; Bienert et al. 2005). Western Blots show that NblA also binds to APC subunits (Bienert et al. 2005). Several

reports indicate that NblA does not bind to β-PC (Bienert et al. 2005; Luque et al. 2003). Recent results, however, show that immobilized *Synechocystis* sp. PCC 6803 NblA1/NblA2 can pull down β-PC from *Synechocystis* sp. PCC 6803 crude cell lysate (Baier et al. 2014).

3.1.5 Binding Partners of NblA

In this report, we describe a combined *in vitro* binding assay, cross-linking, and mass spectrometry (MS) approach to investigate the binding patterns of NblA in the cyanobacterium *Synechococcus elongatus* UTEX 2973 (hereafter referred to as *Synechococcus* 2973) that doubles in ~ 2 hours and has only one copy of *nblA* that is sufficient to mediate PBS degradation (Yu et al. 2015). *In vitro* binding assays combined with chemical cross-linking and MS prove that the C-terminal K^{52} in NblA binds to the N-terminal T^{2} in β-PC. Protein-docking models provided insights on how NblA may interact with the PC structure.

3.2 Materials and Methods

3.2.1 Bacterial Strains and Growth Conditions

All *E. coli* strains were grown under standard conditions (Sambrook and Russell 2001), and supplemented with 100 μg/mL Ampicillin (Amp). *E. coli* cultures induced with 1mM IPTG were kept at 16 ^{o}C for 96 h. Details about strains, plasmids, and oligonucleotides are listed in **Table 3.1**.

3.2.2 Overexpression and Purification of $NblA_{2973}$_His_7 Fusion Protein in *E. coli*

Overexpression of *$nblA_{2973}$_His* from *Synechococcus* 2973 was driven by the T7 promoter in pET21a expression vector. We placed a TEV cleavage site between *nblA* and the His tag. The resultant plasmid, pSL2432, was transformed into the *E. coli* expression host BL21DE3 to make pSL2436 (**Fig. 3.1**). The *E. coli* cells transformed with pSL2436 were grown in LB at 37

^{o}C to O.D.$_{600}$ 0.5. WT NblA$_{2973}$ with ENLYFQGSSHHHHHHH appended on the C-terminus end was induced with 1 mM IPTG (Sigma) in SL2436 *E. coli* cells at 16 ^{o}C for 96 hours to discourage the formation of inclusion bodies (**Fig. 3.2a**). Urea (2 M) was used to increase the solubility of NblA$_{2973}$_His$_7$ fusion proteins (Dines et al. 2007). Cell pellets were washed with Triton X-100 to remove cell membrane and wall debris. Before elution, proteins were refolded on the column with an extensive wash step in the absence of urea. Purification of NblA$_{2973}$_His$_7$ was facilitated by using 2 M urea to solubilize the protein from inclusion bodies in the pellet fraction and by washing extensively (Dines et al. 2007). Cell pellets were dissolved in lysis buffer (50 mM Tris pH 8, 20 mM Imidazole, 500 mM NaCl, 10% glycerol, 1% Tween 20) in the presence of a protease inhibitor cocktail, DNase, and lysozyme and sonicated to lyse cells. The cell lysate was centrifuged for 1 h at 20,200 *x g* and 4 ^{o}C. Inclusion bodies were recovered by saving the pellet fraction. The pellet was washed by re-suspending with wash buffer A (50 mM Tris pH 8, 20 mM imidazole, 500 mM NaCl, 10% glycerol, 0.5% Triton X-100, and 2 M urea). Five 20 min centrifugations at 20,200 *x g* and 4 ^{o}C were carried out to remove *E. coli* membrane and cell-wall materials. The pellet was re-suspended in wash buffer B (50 mM Tris pH 8, 20 mM imidazole, 500 mM NaCl, 10% glycerol, and 2 M urea), and then submitted to two rounds of centrifugation for 20 min at 20,198.7 *x g* and 4 ^{o}C to recover the pellet. Wash buffer B was combined with wash buffer C (50 mM Tris pH 8, 20 mM imidazole, 500 mM NaCl, and 10% glycerol) to create a linear gradient running from 2 M to 0 M urea (with NaCl constant in the buffer) at 1 mL/min, was used to refold the protein on the column before elution. A step gradient of 100, 200, 300, 400, and 500 mM imidazole was used to elute the bound protein with an elution buffer D (50 mM Tris pH 8, 100-500 mM Imidazole, 500 mM NaCl, 10% glycerol).

3.2.3 Isolation of Intact PBS

The NblA-mediated PBS degradation mechanism was studied in cyanobacterium *Synechococcus* 2973. Preparation of PBS from crude *Synechococcus* 2973 and *Synechocystis* 6803 was carried out as described by Ajlani *et al* 1995 (Ajlani et al. 1995). A sucrose gradient containing 0.8 M potassium phosphate pH 7.0 buffer was used in combination with ultracentrifugation (SW41Ti rotor at 209,490 *x g* and 20 ^{o}C) to isolate intact PBS from *Synechococcus* 2973. Sucrose gradients consisted of 3 mL 1.0 M sucrose, 2.5 mL 0.75 M sucrose, 2.5 mL 0.5 M sucrose, and 2.0 mL 0.25 M sucrose. The PBS from *Synechocystis* 6803 was used as a control to check the specificity of $NblA_{2973}$ binding to PBS from *Synechococcus* 2973.

3.2.4 *In Vitro* Binding Assays

Purified $NblA_{2973}$_His_7 fusion protein was mixed with PBS via affinity chromatography. $NblA_{2973}$_His_7 was first incubated with Ni-NTA resin (Qiagen) overnight at 9 ^{o}C with gentle agitation. Unbound proteins were removed by five-column volumes of wash buffer C (10 mM HEPES pH 7.8, 50 mM NaCl, and 5 mM imidazole). NblA-loaded resin was incubated with isolated PBS overnight at 9 ^{o}C with gentle agitation, and then incubated at room temperature for 1 h with gentle agitation. Bound proteins were eluted with buffer B (10 mM HEPES pH 7.8, 50 mM NaCl, and 200 mM imidazole). The molar ratios of NblA:PBS were optimized in many trials.

3.2.5 Chemical Cross-linking

Isotopically labeled 1 mM *N*-hydroxysuccinimide ester-based chemical cross-linker BS^3-H_{12}/D_{12} (Creative Molecules, Inc., http://www.creativemolecules.com/) was incubated overnight at 9 ^{o}C with the elution fractions from *in vitro* pull-down assays to capture any transient protein interactions of NblA and PBS. BS^3 reactive groups on both ends underwent covalent binding to

the primary amines of lysine and at the N-terminus (Leitner et al. 2014) and linked any reactive site 11.4 Å apart. Unbound proteins were washed with HEPES buffer. A final concentration of 1.0 mM BS^3 yielded the best results. The BS^3 cross-linking reaction was quenched by incubating with 2 M excess Tris-HCl to BS^3 for 20 min. Samples were analyzed with gel electrophoresis by using SDS-PAGE, Coomassie Brilliant Blue R250, and UV-enhanced illumination of bilin pigments (**Fig. 3.3**).

3.2.6 Sample Digestion, LC-MS/MS, and Data Processing

For MS analysis, cross-linking products were precipitated with acetone by using a 2D-cleanup Kit (GE Healthcare) and digested in-solution with LysC (Wako Chemicals, Richmond, VA) and Trypsin (Promega, Madison, WI) by following a previously published method (Leitner et al. 2014). In brief, protein pellets were re-suspended in 20 µL 8 M urea (Sigma-Aldrich, St. Louis, MO), then incubated with 2.5 mM tris(2-carboxyethyl)phosphine (Sigma-Aldrich, St. Louis, MO) at 37 ^{o}C for 30 min and 5 mM iodoacetamide (Sigma-Aldrich, St. Louis, MO) for 30 min at room temperature (21 ^{o}C). LysC (Wako Chemicals, Richmond, VA) stock solution (2 µL of 0.5 µg/µL) was added to each cross-linking sample at 37 ^{o}C for 2 h. The protein solution was diluted into 100 mM Tris (Sigma-Aldrich, St. Louis, MO) and incubated with trypsin overnight at 37 ^{o}C. Formic acid (0.1%, Sigma-Aldrich, St. Louis, MO) was added to the samples to quench the digestion.

LC-MS/MS comprised of an Ultimate 3000 Nano LC system (Thermo Scientific Dionex, Sunnyvale, CA) attached on-line to a Q Exactive Plus mass spectrometer (Thermo Fisher, Waltham, MA) was used to analyze the cross-linked sample digests. A 5 µL aliquot of each sample was loaded onto a guard column (Acclaim PepMap100, 100 µm × 2 cm, C18, 5 µm, 100 Å; Thermo Scientific Dionex) in Solvent A (water with 0.1% formic acid), and the peptides in

the sample were then separated on a C_{18} reversed-phase column (Magic, 0.075 mm ×150 mm, 5 μm, 120 Å, Michrom Bioresources, Inc., Auburn, CA) custom-packed at a flow rate of 4.5 μL/min. A linear 90-min gradient from 5-95% solvent B (80% acetonitrile, 20% water, 0.1% formic acid), followed by a 10-min hold at 95% solvent B, was used for peptide elution. After elution, peptides were admitted into the mass spectrometer via a PicoView Nanospray Source (PV550, New Objective, Inc., Woburn, MA) with a spray voltage of 1.8 kV. The mass spectrometer was operated in the positive-ion mode with standard data-dependent acquisition settings as described previously (Liu et al. 2016) with the following modifications: for each precursor-ion scan, the top 15 ions with minimal intensity of 4 x 10^4 counts were selected for fragmentation with an isolation width of 3.0 *m/z* and normalized collision energy of 30% of maximum. Data processing and analysis used Protein Prospector (see references cited in (Liu et al. 2016).

3.2.7 Protein Modeling

To gain insight into the NblA – PC binding modes, a two-step protein-protein docking methodology was used to elucidate likely structural motifs. First, the NblA protein dimer was globally docked from two potential starting configurations onto the PC monomeric sub-unit, consisting of an α-PC and a β-PC protein monomer, with an additional β-PC protein necessary to define the NblA binding pocket. Second, the most likely structural binding mode from each starting configuration was retained and used as an initial structure for a local protein docking routine. The best scoring structures from this local search were retained as the most likely structural motifs for NblA – PC binding.

3.2.8 Global Protein Docking

Initial models for the NblA protein dimer and the monomeric sub-unit of PC were from the X-ray crystal structures for NblA from *Synechococcus* 7942 (PDB ID: 3CS5) (Dines et al. 2008) and PC from *Synechococcus* 7942 (PDB ID: 4H0M) (Marx and Adir 2013) were used. For each of the two starting configurations was generated 100,000 globally docked models of the NblA – PC binding motif, by using the protein-protein docking protocol (Gray et al. 2003; Wang et al. 2007) implemented in the Rosetta 3.4 software package (Leaver-Fay et al. 2011).

To prepare for the docking simulation, initial structures were first relaxed with all heavy atoms constrained using the Rosetta's relax protocol. This initial step minimized steric clashes due to irregularities in the X-ray crystal structures. Next, the docking partners were "pre-packed" by using Rosetta's docking prepack protocol, which moved the docking partners out of contact, optimized the side chains, and then brought the docking partners back to their initial positions. Subsequently, global protein-protein docking was performed by generating 100,000 decoys of the NblA – PC binding motif. For global docking, the NblA dimer was moved out of contact with PC, randomly oriented, and brought back into contact with PC. During this step, PC was held in its initial orientation. Next, the rigid-body orientation of NblA was energy minimized, and positions of NblA and PC side-chains were optimized before scoring the docked conformation (Leaver-Fay et al. 2013).

For each starting conformation, the top 10 scoring structures were retained, and the distance between residues K^{52} on NblA and T^{2} on β-PC was calculated to accommodate a cross-link between these residues. The ProDy package (Bakan et al. 2011) was used to extract atomic level structure from residues K^{52} on NblA and T^{2} on β-PC, and estimate the cross-linking distance from the centroids of the two residues. The cross-linking distance was ranked from lowest to highest, and the binding motif with the lowest cross-linking distance was retained for

subsequent local protein docking (**Fig. 3.4** and **Fig. 3.5**) from global docking results for the two starting conformations).

3.2.9 Local Protein Docking

By using the best globally docked NblA – PC structure from both starting conformations, a second local protein docking search was performed. This refining step was similar to global docking except that the NblA dimer was not randomly oriented prior to rotamer sampling and side-chain optimization. Instead, the initial structure was randomly perturbed by using a Gaussian function for the translational and rotational components, where starting structures were within a 3 Å translation and an 8° rotation of the initial structure. Local protein docking was performed for 25,000 decoys of the NblA – PC binding motif at each starting conformation. For each starting conformation, the top 10 structures were sorted according to their total docking scores, and the estimated cross-linking distance from K^{52} on NblA and T^{2} on β-PC was calculated from the centroids of the two residues. The binding motifs with the lowest cross-linking distance as candidate structures were retained.

3.3 Results

3.3.1 Overexpression and Purification of Fusion Protein

Overexpression and purification of fusion protein – To study the binding partners of NblA, we expressed and purified $NblA_{2973}$_His_7 fusion protein in *E. coli* cells (**Fig. 3.2a**). SDS-PAGE gels confirmed that $NblA_{2973}$_His_7 fusion proteins were purified well at the 7 kDa range (**Fig. 3.2b**). Bottom-up proteomics (LC-MS/MS) analyses identified 100% of the $NblA_{2973}$_His_7 sequence (**Fig. 3.2c**).

3.3.2 *In Vitro* Binding Assays and BS^3 Cross-linking

To determine the specific amino acid residues involved in the binding interactions between NblA and phycobiliproteins, we used isotopically labeled BS^3 cross-linkers to capture any transient protein-protein interactions. Affinity columns demonstrate that purified NblA from *Nostoc* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 bind to α- and β- PC, and APC proteins from crude cell extracts (Baier et al. 2014; Bienert et al. 2005). BS^3 are 11.4 Å spacers that have reactive groups on both ends that form covalent bonds with the primary amines in lysine residues and the N-termini of proteins (Leitner et al. 2014), and possibly the secondary amines in serine, and tyrosine residues (Rappsilber 2011; Sinz 2006). Using a 1:1 mixture of BS^3 labeled with 12 hydrogens and 12 deuteriums encodes the cross-linked peptides and facilitates the identification of true cross-links. A true protein-protein interaction will be labeled 50% of the time with both light (H_{12}) and heavy (D_{12}) forms of BS^3. Consequently, a doublet that has a 12 Dalton shift will show up in the mass spectra, indicating a real cross-link. Careful analysis of the MS^2 chromatograms is required to identify and sequence correctly the peptides that are participating in the binding reaction. SDS-PAGE and $ZnSO_4$ enhanced UV illumination of bilin pigments visualized that the binding of NblA to α- and β-PC was captured by 1 mM BS^3 cross-linker *in vitro*. Extensive optimizations of the ratios between proteins and cross-links were necessary to produce true cross-links. A 7:1 molar ratio of NblA:PBS in the presence of 1 mM BS^3 was the best condition for capturing the binding between NblA and phycobiliproteins (**Fig. 3.3**).

3.3.3 Mass Spectrometry

Studies show that the N- and C- termini of NblA bind to the ClpC protease and phycobiliproteins, respectively (Baier et al. 2014; Baier et al. 2004; Bienert et al. 2005; Karradt et al. 2008) although older work suggests that NblA preferentially binds to only the α-PC (Luque et al. 2003; Bienert et al. 2005). We used a combined pull-down assay, cross-linking, and MS

work-flow to provide direct evidence that NblA binds to PBS, specifically NblA can actually bind to the β-PC.

We analyzed the mass spectral doublet and MS^2 fragmentation spectra to demonstrate that K^{52} from NblA is cross-linked to T^2 from β-PC (**Fig. 3.6a and Fig. 3.6b**). Manual verification of the MS^2 chromatograms allowed confident determination of this cross-link (**Fig. 3.6c**). This cross-link suggests that the C-terminus of NblA is in close proximity to the N-terminus of β-PC (i.e., MAHENIFK^{52}GMIR (NblA) to T^2FDAFTK (β-PC)) (**Fig**. **3.6d**). The "MIR" region in the C-terminus end of NblA plays a critical role in mediating PBS degradation. Sequence alignment of NblA demonstrates that the "MIR" region is conserved (Bienert et al. 2005). R56C mutagenesis in NblA inhibits PBS degradation during nutrient step-down (Dines et al. 2008). The M54 and R56 residues from $NblA_{2973}$ are analogous to the L^{51} and K^{53} residues in $NblA_{7120}$, and mutagenesis of L^{51} and K^{53} prohibits binding of α-PC to $NblA_{7120}$ (Bienert et al. 2005; Karradt et al. 2008). Intriguingly, the T^2 from β-PC residue that is cross-linked to K^{52} from NblA is located inside the PC trimer, indicating that NblA is able gain access to the interior region of the PC rod via a hollow cavity inside the PC rod (Marx and Adir 2013), a region that is not easily accessible to solvents.

We also found NblA-NblA cross-links by interrogating the extracted mass chromatogram (**Fig. 3.7a**), where analysis of the MS^2 fragmentations (**Fig. 3.7b**) indicates that QK^{44}MAHENIFK from NblA is cross-linked to M^{1}LPPLPDFSLSVEQQFDLQK from NblA (**Fig. 3.7c and 3.7d**). A cross-link between M^{1}LPPLPDFSLSVEQQFDLQK from NblA to MAHENIFK^{52}GMIR from NblA also formed (not shown). These cross-links highlight how the N-terminus of NblA is in close proximity to the C-terminus of NblA. These cross-links are

consistent with the crystal structure of NblA, where the M^1 – K^{44} and M^1 – K^{52} residues are within an acceptable distance for cross-links to form.

3.3.4 Modeling

Structural analysis of the cross-links revealed that the N-terminus of β-PC is located in the interior region of the PC trimer, which is potentially accessible from above or inside the rod structure (**Fig. 3.8a**). To initiate PC degradation, NblA needs to infiltrate the PC rod for the C-terminus of NblA to bind to the N-terminus of β-PC (**Fig. 3.8a**). A monomer of β-PC has a groove on the left side located near the T^2 position identified in cross-linking. This vacancy is of similar shape to that of the α-helices near the C-terminus of an NblA dimer, and docking calculations (**Fig. 3.8b**) reveal this geometrical arrangement to be conducive to NblA – PC binding, as well as bringing the K^{52} residue on NblA in close proximity to T^2 on β-PC. In addition, the inside of the PC rod contains a groove near each of the T^2 residues on β-PC, indicating an alternative, or initial mode of NblA – PC binding (**Fig. 3.8c**). Detailed information on the modeling approach is contained in the Experimental Procedures.

3.4 Discussion

The combined chemical cross-linking and MS approach is a powerful structural proteomics tool used to investigate protein-protein interactions and establish low-resolution structures and their conformational changes (Bricker et al. 2015; Sinz 2014). But despite the powerful strengths of this method, there are many challenges to analyzing cross-links correctly. For example, the tryptic digest of cross-linked proteins is an extremely complex mixture eveb for LC-MS/MS analysis. In addition, cross-links tend to be in low-abundance compared to the un-linked peptides. Lastly, unambiguous identification can be difficult, especially with a large candidate protein database. To obviate these problems, we used a mixture of isotopically labeled

BS^3 and Protein Prospector, taking advantage of the characteristic doublet in the mass chromatograms in a high-throughput manner.

In this study, we were particularly interested in discovering where NblA binds to the PBS complex to understand how NblA regulates PBS degradation. We induced a C-terminal His-tagged NblA protein from *Synechococcus* 2973 by 1 mM IPTG at low temperatures for several days (**Fig. 3.2a**). We then purified $NblA_{2973}$_His_7 from *E. coli* with the assistance of urea to increase the solubility and yield **(Fig**. **3.2b**). Optimizations of the molar ratios of $NblA_{2973}$_His_7, isolated PBS complexes from *Synechococcus* 2973, and BS^3 cross-linker demonstrated that the key factor to cross-link formation is the cross-linker concentration. In the absence of the isotopically labeled BS^3 H_{12}/D_{12} cross-linker, we observed a faint band with an approximate MW of 34 kDa. The presence of 1 mM BS^3 encouraged the formation of a denser band, suggesting that BS^3 captured the transient protein-protein interactions in the cross-linking reactions. We assign a band at 34 kDa to a PC hetero-dimer. In addition, the large band that migrated corresponding to 21 kDa is a combination of α- and β- PC and α- and β- APC (**Fig**. **3.3**). There may be other cross-linked products that simply were not resolved by SDS-PAGE and $ZnSO_4$ enhanced UV illumination of bilins.

MS analysis of the cross-link samples provided direct evidence that NblA interacts with the PBS complexes, and we were able to assign the specific amino acid residues participating in the binding sites. To start, the $M^1 – K^{44}$ (**Fig**. **3.7d**) and $M^1 – K^{52}$ cross-links between NblA-NblA confirm that NblA was properly refolded on the column after solubilization by urea. In addition, the $K^{52} –T^2$ **(Fig**. **3.6d**) cross-link between NblA-β-PC is intriguingly positioned in the central cavity of the PC rod. Protein docking studies suggest that NblA fits in the open groove near the N-terminus of β-PC, via one or a combination of structural mimicry (Dines et al. 2008), binding

to conserved residues (Karradt et al. 2008), electrostatic-binding association, and/or van der Waals' forces. The open groove is also in close proximity to the N-termini of both α and β-PC. Interestingly, studies in *Nostoc* sp. PCC 7120 show that NblA binds to amino acid residues 16-39 of α-PC (Bienert et al. 2005). NblA binding inside the hollow cavity of the PC rods has implications about the structural rearrangement and degradation of PBS complexes.

Given that NblA only functions as a proteolysis adaptor, the key factors that actually degrade phycobiliproteins belong to the Clp proteasome system. ClpC is an essential protein chaperone (Karradt et al. 2008) that is constitutively expressed in microbes and plants (Olinares et al. 2011). After activation by NblA, ClpC chaperones oligomerize into NblA/ClpC hetero-hexamers that are several hundred kDa in size (Kirstein et al. 2006; Karradt et al. 2008). Cross-linking studies demonstrate that ClpC binds to PBS complexes (Stanne et al. 2007), and confocal imaging shows that NblA co-localizes to photosynthetic membranes where PBS complexes are present (Sendersky et al. 2014). These findings imply that NblA facilitates the interactions between ClpC and phycobiliproteins (Karradt et al. 2008) in PBS complexes that are still attached to the thylakoid membrane. ClpC has the ability to unfold and shuttle proteins to the ClpP catalytic protease core, which is usually located in the membrane (Olinares et al. 2011). It appears unlikely that these large NblA/ClpC complexes could fit inside the inner hollow cavity of the PC rod if PBS complexes are rigid structures. Perhaps during nutrient stress, PBS complexes undergo a conformational change. For instance, the extension or disassociation of the linker proteins that hold the massive PBS complexes together may possibly allow NblA the opportunity to enter the hollow cavity of the PC rod during nutrient stress. NblA binding may also disrupt the structural integrity of the entire PBS complex to make the structure more

accessible for degradation. Although compelling, experimental verification is needed for several aspects of this model.

In summary, we successfully discovered that NblA binds to the interior region of the central cavity of the PC rods (**Fig. 3.8**). This finding presents new directions of research to elucidate the sequential mechanism of how NblA mediates PBS degradation. To start, how many copies of NblA binds simultaneously to the PC rods remains a mystery. Densitometric analyses of the LiDS gels of His_6-NblA1-phycobiliprotein complexes from *Tolypothrix* PCC 7601 indicate that the molar ratio of NblA1/PC(α+β) is 0.65, after corrected for the difference in molecular mass between these subunits (Luque et al. 2003). This suggests that the number of NblA needed to trigger PBS degradation is less than the amount of PC available. Binding between NblA and APC proteins were not detected by gels in *Tolypothrix* sp. PCC 7601 (Luque et al. 2003). But recent imaging studies have shown that NblA can actually interact with the APC proteins (Sendersky et al. 2015). Further analysis of NblA binding to the APC core would provide greater insights on how NblA triggers degradation of the PBS complexes. Lastly, the mechanistic details of how ClpC/NblA hetero-oligomer denatures and shuttles phycobiliproteins to the proteolytic catalytic center is unclear. Continued explorations of how the Clp proteasome system regulates the levels of phycobiliproteins would offer more insights into the roles of Clp proteins in cyanobacteria, other eubacteria and plants.

3.5 Acknowledgements

We would like to thank Drs. Aparna Nagarajan, Michelle Liberton, Haijun Liu, all the past and current members of the Pakrasi laboratory for collegial discussions and advices. Funding for this study was provided by the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science,

Office of Basic Energy Sciences under award number DE-SC 0001035. AYN has been supported by an NSF Graduate Research Fellowship grant number DGE-1143954. Funding for WPB has been provided by ARO MURI W911NF1210420. Mass spectrometry was supported in part by the NIH, NIGMS Grant No. P41GM103422 to MLG.

3.5.1 Author Contributions

A.Y.N. and H.B.P. conceived the project; A.Y.N., and H.Z. performed the experiments; A.Y.N., H.Z., and D.W. analyzed the mass spectrometry data; W.P.B. designed and implemented the docking models, and analyzed the docking modeling data; M.G. and H.B.P contributed to the development of the measurement capabilities used; A.Y.N., W.P.B., D.W., M.G., and H.B.P. wrote the manuscript.

3.5.2 Conflict of Interest Statement

The authors declare no competing financial interests.

3.6 References

Adir, N., Dines, M., Klartag, M., McGregor, A., Melamed-Frank, M. (2006). "Assembly and disassembly of phycobilisomes." *Complex Intracellular Structures in Prokaryotes*. Springer Berlin Heidelberg, Berlin, Heidelberg: 47-77.

Adir, N., Zer, H., Shochat, S., Ohad, I. (2003). "Photoinhibition - a historical perspective." *Photosynthesis Research*. **76**: 343-370.

Ajlani, G., Vernotte, C. (1998). "Construction and characterization of a phycobiliprotein-less mutant of *Synechocystis* sp. PCC 6803." *Plant Molecular Biology*. **37**(3): 577-580.

Ajlani, G., Vernotte, C., DiMagno, L., Haselkorn, R. (1995). "Phycobilisome core mutants of *Synechocystis* sp. PCC 6803." *Biochimica et Biophysica Acta - Bioenergetics*. **1231**(2): 189-196.

Allen, M.M., Smith, A.J. (1969). "Nitrogen chlorosis in blue-green algae." *Archives Microbiology*. **69**(2): 114-120.

Andersson, F.I., Blakytny, R., Kirstein, J., Turgay, K., Bukau, B., Mogk, A., Clarke, A.K. (2006). "Cyanobacterial ClpC/HSP100 protein displays intrinsic chaperone activity." *Journal of Biological Chemistry*. **281**(9): 5468-5475.

Baier, A., Winkler, W., Korte, T., Lockau, W., Karradt, A. (2014). "Degradation of phycobilisomes in *Synechocystis* sp. PCC 6803: evidence for essential formation of an NblA1/NblA2 heterodimer and its codegradation by a Clp protease complex." *Journal of Biological Chemistry*. **289**(17): 11755-11766.

Baier, K., Lehmann, H., Stephan, D.P., Lockau, W. (2004). "NblA is essential for phycobilisome degradation in *Anabaena s*p. strain PCC 7120 but not for development of functional heterocysts." *Microbiology*. **150**: 2739-2749.

Baier, K., Nicklisch, S., Grundner, C., Reinecke, J., Lockau, W. (2001). "Expression of two *nblA*-homologous genes is required for phycobilisome degradation in nitrogen-starved *Synechocystis* sp. PCC 6803." *FEMS Microbiology Letters*. **195**:35-39

Bakan, A., Meireles, L.M., Bahar, I. (2011). "ProDy: protein dynamics inferred from theory and experiments." *Bioinformatics*. **27**(11): 1575-1577.

Bienert, R., Baier, K., Volkmer, R., Lockau, W., Heinemann, U. (2005). "Crystal structure of NblA from *Anabaena* sp. PCC 7120, a small protein playing a key role in phycobilisome degradation." *Journal of Biological Chemistry*. **281**(8): 5216-5223.

Bogorad, L. (1975). "Phycobiliproteins and Complementary Chromatic Adaptation." *Annual Review of Plant Physiology*. **26**(1): 369-401.

Bricker, T.M., Mummadisetti, M.P., Frankel, L.K. (2015). "Recent advances in the use of mass spectrometry to examine structure/function relationships in photosystem II." *Journal of Photochemistry and Photobiology.* **15**(Part B): 227-246.

Collier, J.L., Grossman, A.R. (1992). "Chlorosis induced by nutrient deprivation in *Synechococcus* sp. strain PCC 7942: not all bleaching is the same." *Journal of Bacteriology*. **174**(14): 4718-4726.

Collier, J.L., Grossman, A.R. (1994). "A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria." *EMBO Journal*. **13**(15): 1039-1047.

Dines, M., Sendersky, E., David, L., Schwarz, R., Adir, N. (2008). "Structural, functional, and mutational analysis of the NblA protein provides insight into possible modes of interaction with the phycobilisome." *Journal of Biological Chemistry*. **283**(44): 30330-30340.

Dines, M., Sendersky, E., Schwarz, R., Adir, N. (2007). "Crystallization of sparingly soluble stress-related proteins from cyanobacteria by controlled urea solublization." *Journal of Structural Biology*. **158**(1): 116-121.

Elmorjani, K., Thomas, J.C., Sebban, P. (1986). "Phycobilisomes of wild type and pigment mutants of the cyanobacterium *Synechocystis* PCC 6803." *Archives Microbiology*. **146**(2): 186-191.

Glazer, A.N. (1982). "Phycobilisomes: structure and dynamics." *Annual Review of Microbiology*. **36**: 173-198.

Gorl, M., Sauer, J., Baier, T., Forchhammer, K. (1998). "Nitrogen-starvation-induced chlorosis in *Synechococcus* PCC 7942: adaptation to long-term survival." *Microbiology*. **144**(Pt 9): 2449-2458.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., Baker, D. (2003). "Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations." *Journal of Molecular Biology*. **331**(1): 281-299.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., Collier, J.L. (1993). "The phycobilisome, a light-harvesting complex responsive to environmental conditions." *Microbiological Reviews*. **57**(3): 725-749.

Karradt, A., Sobanski, J., Mattow, J., Lockau, W., Baier, K. (2008). "NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease." *Journal of Biological Chemistry*. **283**(47): 32394-32403.

Kirstein, J., Schlothauer, T., Dougan, D.A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B., Turgay, K. (2006). "Adaptor protein controlled oligomerization activates the AAA+ protein ClpC." *EMBO Journal*. **25**(7): 1481-1491.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., Davis, I.W., Cooper, S., Treuille, A., Mandell, D.J., Richter, F., Ban, Y.E., Fleishman, S.J., Corn, J.E., Kim, D.E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J.J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J.J., Kuhlman, B., Baker, D., Bradley, P. (2011). "ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules." *Methods in Enzymology*. **487**: 545-574.

Leaver-Fay, A., Tyka, M., Jacak, R., Song, Y., Kellogg, E.H., Thompson, J., Davis, I.W., Pache, R.A., Lyskov, S., Gray, J.J., Kortemme, T., Richardson, J.S., Havranek, J.J., Snoeyink, J., Baker, D., Kuhlman, B. (2013) "Scientific benchmarks for guiding macromolecular energy function improvement." *Methods in Enzymology*. **523**: 109-143.

Leitner A, Walzthoeni T, Aebersold R (2014). "Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline." *Nature Protocols*. **9**(1): 120-137.

Li, H., Sherman, L.A. (2002). "Characterization of *Synechocystis* sp. strain PCC 6803 and deltanbl mutants under nitrogen-deficient conditions." *Archives Microbiology*. **178**(4): 256-266.

Liu, H., Zhang, H., Niedzwiedzki, D.M., Prado, M., He, G., Gross, M.L., Blankenship, R.E. (2013). "Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria." *Science.* **342**(6162): 1104-1107.

Liu, H., Zhang, H., Orf, G.S., Lu, Y., Jiang, J., King, J.D., Wolf, N.R., Gross, M.L., Blankenship, R.E. (2016). "Dramatic domain rearrangements of the cyanobacterial orange carotenoid protein upon photoactivation." *Biochemistry*. **55**(7): 1003-1009.

Luque, I., Ochoa de Alda, J.A., Richaud, C., Zabulon, G., Thomas, J.C., Houmard, J. (2003). "The NblAI protein from the filamentous cyanobacterium *Tolypothrix* PCC 7601: regulation of its expression and interactions with phycobilisome components." *Molecular Microbiology*. **50**(3): 1043-1054.

MacColl, R. (1998). "Cyanobacterial phycobilisomes." *Journal of Structural Biology*. **124**(2-3): 311-334.

Marx, A., Adir, N. (2013). "Allophycocyanin and phycocyanin crystal structures reveal facets of phycobilisome assembly." *Biochimica et Biophysica Acta*. **1827**(3): 311-318.

Olinares, P.D., Kim, J., van Wijk, K.J. (2011). "The Clp protease system; a central component of the chloroplast protease network." *Biochim et Biophys Acta.* **1807**(8): 999-1011.

Porankiewicz, J., Wang, J., Clarke, A.K. (1999). "New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond." *Molecular Microbiology*. **32**(3): 449-458.

Rappsilber, J. (2011). "The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes." *Journal of Structural Biology*. **173**(3): 530-540.

Richaud, C., Zabulon, G., Joder, A., Thomas, J.C. (2001). "Nitrogen or sulfur starvation differentially affects phycobilisome degradation and expression of the *nblA* gene in *Synechocystis* strain PCC 6803." *Journal of Bacteriology*. **183**(10): 2989-2994.

Sambrook, J., Russell, D.W. (2001). "Molecular Cloning: A Laboratory Manual." *Cold Spring Harbor Laboratory Press*. **1**(4): 1-1881.

Sauer, J., Schreiber, U., Schmid, R., Volker, U., Forchhammer, K. (2001). "Nitrogen starvation-induced chlorosis in *Synechococcus* PCC 7942. Low-level photosynthesis as a mechanism of long-term survival." *Plant Physiology*. **126**(1): 233-243.

Scheer, H., Zhao, K.H. (2008). "Biliprotein maturation: the chromophore attachment." *Molecular Microbiology*. **68**(2): 263-276.

Schwarz, R., Forchhammer, K. (2005). "Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular responses." *Microbiology*. **151**(Pt 8): 2503-2514.

Sendersky, E., Kozer, N., Levi, M., Garini, Y., Shav-Tal, Y., Schwarz, R. (2014). "The proteolysis adaptor, NblA, initiates protein pigment degradation by interacting with the cyanobacterial light-harvesting complexes." *Plant Journal*. **79**(1): 118-126.

Sendersky, E., Kozer, N., Levi, M., Moizik, M., Garini, Y., Shav-Tal, Y., Schwarz, R. (2015). "The proteolysis adaptor, NblA, is essential for degradation of the core pigment of the cyanobacterial light-harvesting complex." *Plant Journal*. **83**(5): 845-852.

Sidler, W. (2004). "Phycobilisome and phycobiliprotein structures." *The Molecular Biology of Cyanobacteria*. **1**: 139-216.

Sinz, A. (2006). "Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions." *Mass Spectrometry Reviews*. **25**(4): 663-682.

Sinz, A. (2014). "The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks." *Expert Review of Proteomics*. **11**(6): 733-743.

Stanne, T.M., Pojidaeva, E., Andersson, F.I., Clarke, A.K. (2007). "Distinctive types of ATP-dependent Clp proteases in cyanobacteria." *Journal of Biological Chemistry.* **282**(19): 14394-14402.

Wang, C., Bradley, P., Baker, D. (2007). "Protein-protein docking with backbone flexibility." *Journal of Molecular Biology*. **373**(2): 503-519.

Weisz, D.A., Gross, M.L., Pakrasi, H.B. (2016). "The use of advanced mass spectrometry to dissect the life-cycle of photosystem II." *Frontiers in Plant Sciences*. **7**: 1-25.

Yu, J., Liberton, M., Cliften, P.F., Head, R.D., Jacobs, J.M., Smith, R.D., Koppenaal, D.W., Brand, J.J., Pakrasi, H.B. (2015). "*Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO_2." *Scientific Reports*. **5(**8132): 1-10.

Table 3.1: Strains, plasmids, and oligonucleotides.

Strains, plasmids, or oligonucleotides	**Description**	**Source of reference**
Strains		
Synechococcus 2973	Wild-type fast growing cyanobacterium	(Yu et al. 2015)
Synechococcus 2973 Δ*nblA*	Fast growing cyanobacterium with *nblA* deleted	(Yu et al. 2015)
DH5αMCR *E. coli*	Background strain	
Transformed pET21a/ $NblA_{2973}$_His_7	*E. coli* transformed with expression vector carrying $nblA_{2973}$_$His7$	This reference
Plasmids		
pET21a	Expression vector containing multiple cloning sites	
pET21a/ $NblA_{2973}$_His_7	Vector for expression of $nblA_{2973}$_His_7	This reference
Oligonucleotides[a]		
# nbl1: NblA.TEV.His.F	AGATATACATatgctccccctctccccg	
# nbl2: pET21a.R	gggggagcatATGTATATCTCCTTCTTAAAGTTAAAC	
# nbl3: pET21a.F	GAGAACCTGTACTTCCAGGGCAGTAGTCACCACCACCACC ACCAC	
# nbl4: NblA.TEV.His.R	ACTACTGCCCTGGAAGTACAGGTTCTCacttccttggcgaatcatgC	

[a]Oligonucleotides read in 5’ to 3’ direction.

Figure 3.1: Plasmid map of the overexpression of $NblA_{2973}$_His_7 strain.

ORF of *nblA* was appended with a C-terminal TEV and His tag in the pET21a expression vector. The T7 promoter was used to overexpress *nblA*.

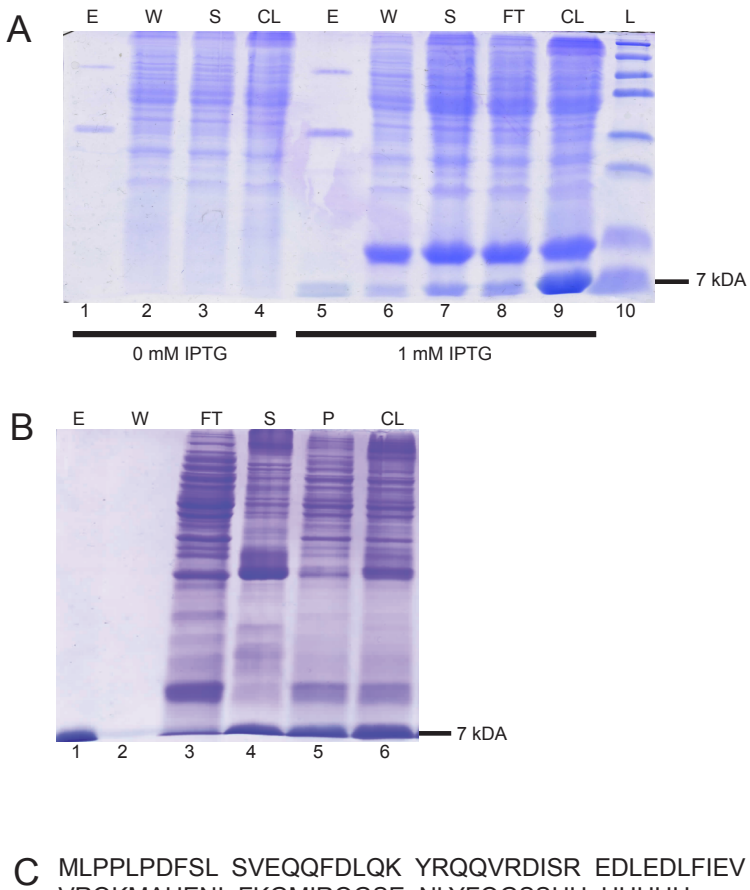


Figure 3.2: Induction, purification, and LC-MS/MS analysis of $NblA_{2973}$_His_7.

a The soluble $NblA_{2973}$_His_7 7-kDa fusion protein was induced with 1 mM IPTG (lanes 5-9) for 96 hours at 16°C, as shown by SDS-PAGE and Coomassie Blue R250 staining. **b** 2 M urea solubilized the $NblA_{2973}$_His_7 that was trapped in the pellet fraction after sonication (lane 5). $NblA_{2973}$_His_7 was washed and refolded extensively on the column before elution. The molecular mass and purity of $NblA_{2973}$_His_7 (lane 1) was verified by SDS-PAGE and Coomassie Blue R250 staining. **c** In-gel digestion of purified $NblA_{2973}$_His_7 was carried out for bottom-up proteomics. 100% coverage of the protein was detected. Abbreviations: L – Ladder, CL – Cell Lysate, FT – Flow Through, S – Supernatant, W – Wash, and E – Elution.

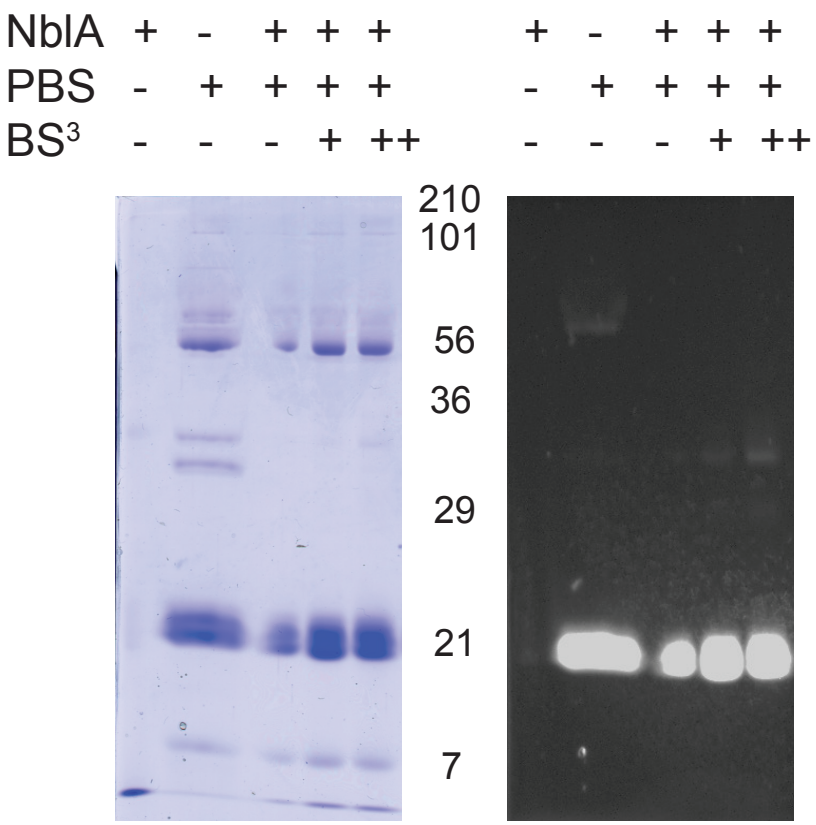


Figure 3.3: *In vitro* binding and cross-linking of NblA and phycobiliproteins.

$NblA_{2973}$_His_7 bound NTA-Ni resin was incubated with isolated PBS from *Synechococcus* 2973 in the following concentrations of BS^3: 0 mM (-), 0.5 mM (+), and 1.0 mM (++). Higher concentration of BS^3 yielded higher molecular weight bands, suggesting that cross-links between $NblA_{2973}$_His_7 and PBS_{2973} had formed.

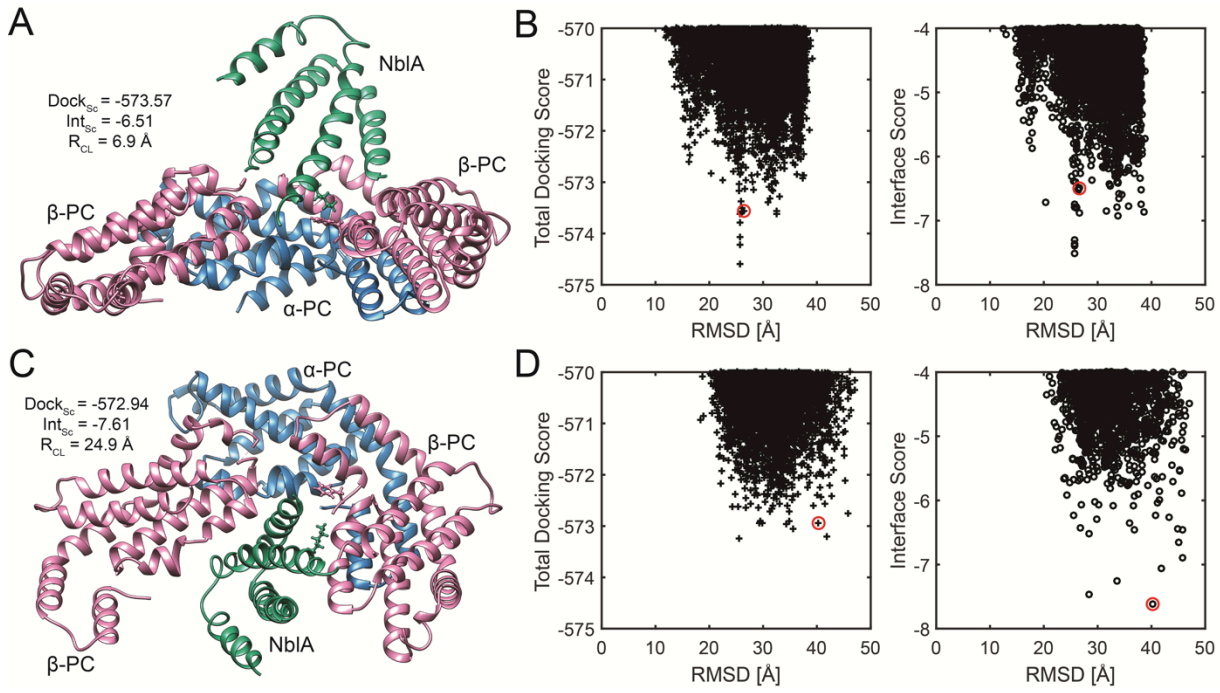


Figure 3.4: Docking results of an NblA dimer globally docked to α-PC/β-PC monomer for orientation I.

a Best globally docked pose sorted by lowest cross-linking distance [$\mathbf{R}_{CL}$] between residue K^{52} of NblA (green) and residue T^{2} of β-PC (pink) after extracting the top ten scoring structures from 100,000 docked poses from initial orientation I. Total docking score [$Dock_{Sc}$] and interface score [Int_{Sc}] are also reported for the pose in addition to cross-linking distance [$\mathbf{R}_{CL}$]. **b** Total docking score for the 100,000 globally docked poses vs. RMSD [Å] from initial orientation I, and Interface score for the 100,000 globally docked poses vs. RMSD [Å] from initial orientation I. The location of the best globally docked pose from orientation I is circled in red on both plots. **c** Best globally docked pose sorted by lowest cross-linking distance [$\mathbf{R}_{CL}$] between residue K^{52} of NblA (green) and residue T^{2} of β-PC (pink) after extracting the top ten scoring structures from 100,000 docked poses from initial orientation II. Total docking score [$Dock_{Sc}$] and interface score [Int_{Sc}] are also reported for the pose in addition to cross-linking distance [$\mathbf{R}_{CL}$]. **d** Total

docking score for the 100,000 globally docked poses vs. RMSD [Å] from initial orientation II, and Interface score for the 100,000 globally docked poses vs. RMSD [Å] from initial orientation II. The location of the best globally docked pose from orientation II is circled in red on both plots.

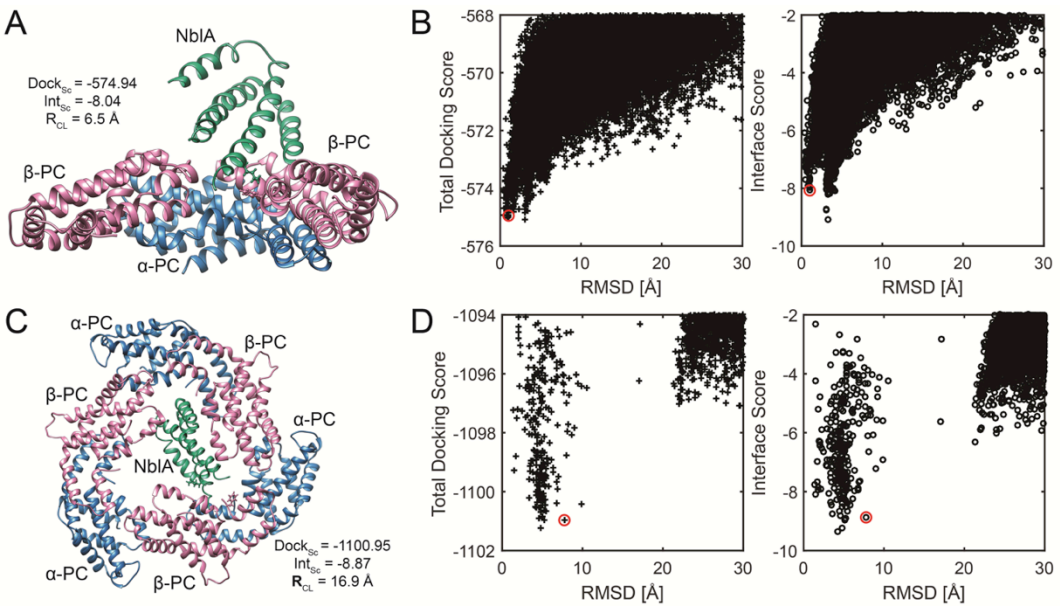
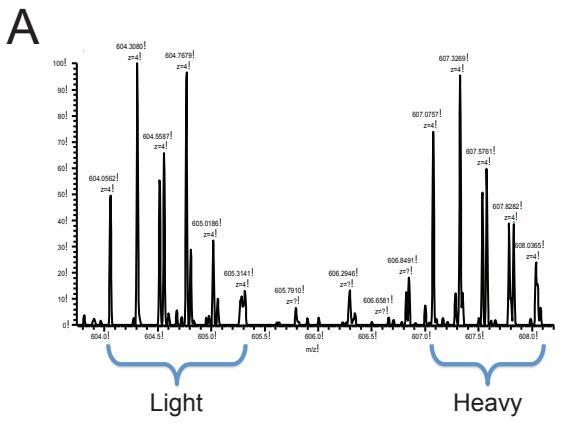
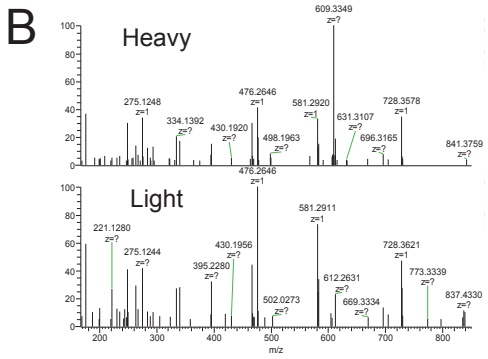


Figure 3.5: Docking results of an NblA dimer globally docked to α-PC/β-PC monomer for orientation II.

a Best locally docked pose sorted by lowest cross-linking distance [$\mathbf{R}_{CL}$] between residue K^{52} of NblA (green) and residue T^{2} of β-PC (pink) after extracting the top ten scoring structures from 25,000 docked poses from initial orientation I. Total docking score [$Dock_{Sc}$] and interface score [Int_{Sc}] are also reported for the pose in addition to cross-linking distance [$\mathbf{R}_{CL}$]. **b** Total docking score for the 25,000 locally docked poses vs. RMSD [Å] from initial orientation I, and Interface score for the 25,000 locally docked poses vs. RMSD [Å] from initial orientation I. The location of the best locally docked pose from orientation I is circled in red on both plots. **c** Best locally docked pose sorted by lowest cross-linking distance [$\mathbf{R}_{CL}$] between residue K^{52} of NblA (green) and residue T^{2} of β-PC (pink) after extracting the top ten scoring structures from 25,000 docked poses from initial orientation II. Total docking score [$Dock_{Sc}$] and interface score [Int_{Sc}] are also reported for the pose in addition to cross-linking distance [$\mathbf{R}_{CL}$]. **d** Total docking score for the 25,000 locally docked poses vs. RMSD [Å] from initial orientation II, and Interface score for the

25,000 locally docked poses vs. RMSD [Å] from initial orientation II. The location of the best locally docked pose from orientation II is circled in red on both plots.





C

Fragment Ion	m/z, predicted	m/z, light	m/z, heavy
$b3^{+1}$; H	340.1443	340.1442	340.1436
$b4^{+1}$; E	469.1869	469.1849	469.1855
$b5^{+1}$; N	583.2299	583.2303	583.2303
$b6^{+1}$; I	696.3139	696.3096	696.3165
$y1^{+1}$; R	175.1195	175.1188	175.1188
$y2^{+1}$; I	288.2036	288.2022	288.2021
$y4^{+1}$; G	476.2655	476.2646	476.2646
$y6^{+4}$; F	430.23013	430.1956	430.1920
Fragment Ion	**m/z, predicted**	**m/z, light**	**m/z, heavy**
$y2^{+1}$; T	248.161	248.1602	248.1596
$y3^{+1}$; F	395.2294	395.228	395.2288
$y4^{+1}$; A	466.2666	466.2674	466.2639
$y5^{+1}$; D	581.2935	581.2911	581.292
$y6^{+1}$; F	728.3619	728.3621	728.3578

D

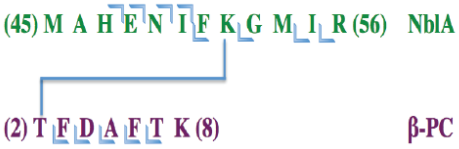
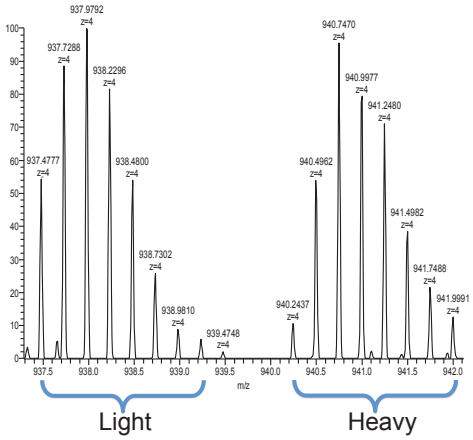


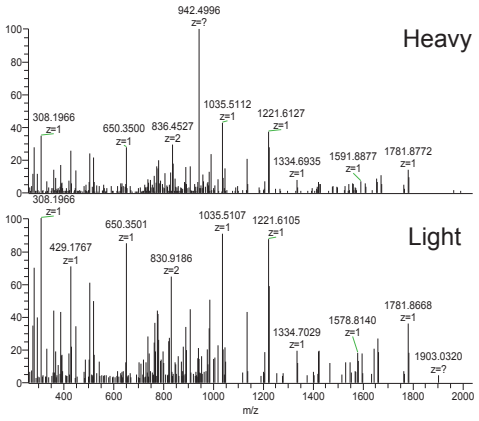
Figure 3.6: C-terminus of $NblA_{2973}$_His_7 binding to the N-terminus of β-PC.

a 12 Dalton mass shift doublet in MS^1 chromatogram suggests a real cross-link has been detected. **b** MS^2 fragmentations of the isotopic peaks from MS^1. **c** In-depth analysis of all the fragments from the MS^2 confirms the identity of the two proteins participating in the binding reaction. **d** Illustration of the T^2 β-PC and M^{52} NblA residues are participating in the cross-linking reaction.

A



B



C

Fragment Ion	m/z, predicted	m/z, light	m/z, heavy
$y2^{+1}$; F	294.1818	294.1806	294.1809
$y3^{+1}$; I	407.2658	407.266	407.2625
$y4^{+1}$; N	521.3088	521.2865	521.2869
$y5^{+1}$; E	650.3513	650.3501	650.3500
$y6^{+1}$; H	787.4103	787.4088	787.4139
$y7^{+1}$; A	858.4474	858.4438	858.4429
$y6^{+3}$; H	263.14162	263.1022	263.1012

Fragment Ion	m/z, predicted	m/z, light	m/z, heavy
$b2^{+3}$; L	543.28285	543.2672	543.2617
$b7^{+5}$; D	430.22648	430.178	430.1775
$y2^{+1}$; Q	275.1719	275.1713	275.1706
$y3^{+1}$; L	388.256	388.254	388.2546
$y4^{+1}$; D	503.2829	503.2813	503.2817
$y5^{+1}$; F	650.3513	650.3501	650.3500
$y6^{+1}$; Q	778.4099	778.4099	778.4098
$y7^{+1}$; Q	906.4685	906.4690	906.4692
$y8^{+1}$; E	1035.5111	1035.5107	1035.5112
$y9^{+1}$; V	1134.5795	1134.5802	1134.579
$y10^{+1}$; S	1221.6115	1221.6105	1221.6127
$y11^{+1}$; L	1334.6956	1334.7029	1334.6935
$y12^{+1}$; S	1421.7276	1421.7292	1421.7228
$y13^{+1}$; F	1568.796	1568.812	1568.7925
$y15^{+1}$; P	1780.8757	1780.8772	1780.8823

D

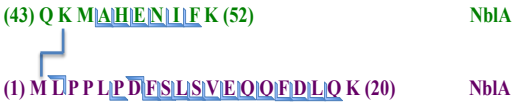


Figure 3.7: C-terminus of NblA$_{2973}$_His$_7$ binding to the N-terminus of NblA.

a 12 Dalton mass shift doublet in MS^1 chromatogram suggests a real cross-link has been detected. **b** MS^2 fragmentations of the isotopic peaks from MS^1. **c** In-depth analysis of all the fragments from the MS^2 confirms the identity of the two proteins participating in the binding reaction. **d** Illustration of the M^1 NblA and K^{44} NblA residues are participating in the cross-linking reaction.

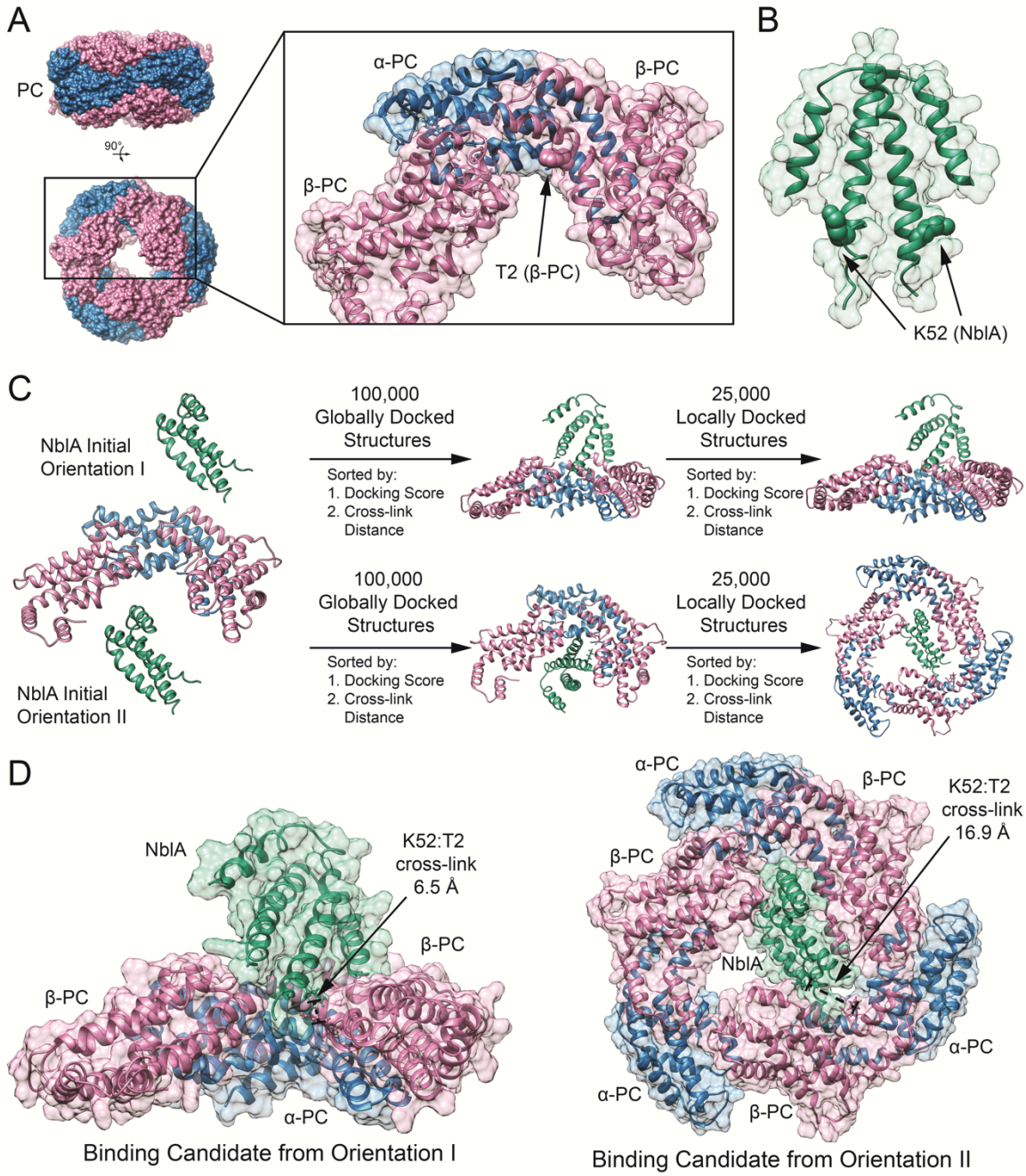


Figure 3.8: Molecular Docking of NblA Dimer to PC.

a X-ray structure of Phycocyanin (PC) from *Synechococcus* 7942 (PDB ID: 4H0M), showing top and side view of hexameric structure. Monomeric sub-unit highlighted including one α-PC (blue) and one β-PC (pink), with an additional β-PC (pink) to define the NblA binding pocket. The N-terminus of β-PC is located in the interior hollow cavity of the PC rod. **b** X-ray structure of NblA

dimer from *Synechococcus* 7942 (PDB ID: 3CS5), where residue K^{52} is highlighted due to cross-link with β-PC. **c** Protein-protein docking protocol using RosettaDock. Two initial orientations are chosen, simulating an NblA dimer (green) docking to PC from above (orientation I) and from inside the rod (orientation II). First, 100,000 globally docked candidates structures are generated, sorted by docking score and subsequently sorted by closest cross-link between T^2 (β-PC) and K^{52} (NblA). Global docking details are presented in Fig. S2. Second, the best globally docked candidate is the initial structure for generating 25,000 locally docked candidate structures, sorted again by docking score and subsequently sorted by closest cross-link between T^2 (β-PC) and K^{52} (NblA). For orientation II, the trimeric PC structure (three sub-units each of α-PC and β-PC) is used to accurately define the NblA binding pocket inside the PC rod. Local docking details are presented in Fig. S3. **d** Binding candidates for NblA dimers docked to PC from the two initial orientations. Residues T^2 from β-PC and K^{52} from NblA are highlighted due to cross-link.

Chapter 4

Conclusions and Future Directions

4.1 Introduction

The goal of our studies was to investigate the protective mechanisms used by cyanobacteria to overcome environmental stresses. Protein post-translational modifications (PTMs) and catabolism play a key role in maintaining cell homeostasis, but these processes are poorly understood. For this dissertation, we focused on the following two research areas: 1) dynamic site-specific redox changes on protein thiols that inactivate/activate protein functions and 2) disassembly of massive light harvesting apparatuses. By developing and optimizing powerful proteomics work-flows, we have created a global map of the redox states of thiols under different oxidizing conditions and discovered the binding partners of NblA, a proteolysis adaptor. A comprehensive list of all redox-sensitive Cys residues in a model cyanobacterium was determined, which provided insights into the broad and massive signaling network that promotes cell acclimation to environmental perturbations. Given that numerous phycobiliproteins have redox-active protein thiols, we explored the mechanism of how the PBS complexes become degraded. Our biochemical and mass spectrometry (MS) studies provided direct evidence that NblA interacts specifically to the N-terminal end of β-PC, a region inside the central cavity of the PC rod. This finding offers insights in the structural rearrangements the PBS complexes may have during nutrient deprivation.

4.2 Redox Regulation

Light is the driving force that powers photosynthesis and is known to modify the redox environment in the cell, which impacts the photosynthetic electron transport (PET) chain and photosystems stability. Alterations of the redox environment occur because photosynthesis operates in the presence of strong oxidants and reductants that lead to the production of damaging chemically reactive species. Upon light activation, electrons flow through the PET

chain, and are transferred to enzymes like ferredoxin, then ferrodoxin-thioredoxin reductase, which then reduce a crucial enzyme call thioredoxin (Trx), which has two Cys in its active site (Buchanan 1991). The reduced state of thioredoxin is responsible for redox-dependent regulation of thiol enzymes.

Redox signaling is the science of cellular communication and information processing involving the electron-transfer processes. Redox-based PTMs are conserved throughout evolution and influence many aspects of cellular physiology (Stamler et al. 2001). Thiol redox state plays essential roles in the regulation of metabolic, signaling, and transcriptional process in cells (Buchanan and Balmer 2005). Crucial biological processes like photosynthesis, cellular metabolism, and gene expression are regulated by the redox states of proteins.

4.2.1 Reversible Protein Thiol Oxidations

During active photosynthesis, cyanobacteria and chloroplasts experience highly oxidizing conditions that undoubtedly promote significant PTMs in proteins. Reversible reduction and oxidation of proteins enable thiol residues to serve as dynamic redox switches. The emerging principle of redox regulations revolves around the concept that proteins thiols undergo reversible PTMs that affect protein functions. Which lead to the suggestion that thiols act like sub-cellular redox switches. The term “reversible redox switches” have been coined to describe how the dynamic redox state changes of protein thiols may determine the inactivation/activation of enzymatic reactions. The types of Cys modifications are diverse. For example, the thiol side chain in cysteines can become oxidized to acids, form disulfide bonds, or interact with RNS and GSH to become mixed disulphides (Dalle-Donne et al. 2009). There are other different types of redox modifications on proteins thiols. Unraveling the dynamics of such modifications should provide a comprehensive understanding of how proteins protect cells from redox stresses.

4.2.2 Quantification of Redox Protein Thiols

We developed and optimized a method to detect on a global-scale oxidized Cys peptides. We accomplished this by enriching oxidized Cys peptides on a column and using quantitative isobaric labeling to quantify the extent of oxidation on individual thiols and redox dynamics under different oxidizing conditions. Cells were grown in optimal conditions to the exponential growth phase, and then exposed to one of three conditions: 1) continuous light, 2) continuous dark, and 3) continuous light and a chemical inhibitor DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea). Continuous light allows pigment proteins to initiate photosynthesis, which creates a reducing environment as enzymes along the PET chain gains electrons. Continuous darkness promoted an oxidizing environment since water cannot be split to yield electrons that could pass through the PET chain. The presence of DCMU in continuous light artificially creates an enhanced oxidizing environment since the chemical completely inhibits photosynthesis.

A resin-assisted thiol-affinity enrichment followed by LC-MS/MS analyses of four biological replicates collectively led to the identification and quantification of 1,060 out of 3,672 predicted proteins in *Synechocystis* 6803 (i.e. 28.9%), among them more than half of the proteins were redox-sensitive. A quantitative and site-specific analysis of ~2,100 Cys sites across diverse protein functional processes provides a detail map of redox sensitive protein thiols that are switched on and off by redox processes. This data facilitates the discovery of factors responsible for protecting cells from ROS damage and provides a broad quantitative picture of thiol-based redox regulation in photosynthetic organisms. There are consistent increases in the levels of Cys thiol oxidation in the dark compared to the light condition. DCMU inhibition leads to further increase in the level of oxidation. **Fig. 2.2a** illustrates that peptides are more oxidized under dark and DCMU conditions compared to the light condition. About ~1,000 out of the ~2,100 peptides

detected were classified as redox sensitive residues because of their 1.5 fold change in oxidation levels in both dark against light and in DCMU against light.

The changes of Cys oxidation levels demonstrate that protein thiols are less reduced after a switch to darkness or in the presence of DCMU, which is probably caused by the interruption of electron flow through the PET chain. Our large-scale proteomic map is a valuable tool that helps scientists focus on select dynamic redox changes in protein thiols to investigate how cells cope with environmental stresses.

4.2.3 Broad Redox Regulation

We used the DAVID online tool to perform bioinformatics analysis on our large dataset to identify the functional information about our 523 redox-sensitive proteins. Of the 523 significantly oxidized, 316 proteins have broad functional annotations that confirm they participate in different processes that are regulated by Trx, such as, amino acid biosynthesis, vitamin biosynthesis, starch degradation, glycolysis, DNA replication, Calvin cycle, nitrogen metabolism, sulfur metabolism, fatty acid biosynthesis, pentose phosphate cycle, and oxidative stress. ~200 proteins identified as redox-sensitive have unknown functions. Redox changes of protein thiols from enzymes in the PET chain control many known processes in an oxygenic photosynthetic organism. Of particular interest is the increased oxidation of Cys residues from phycobiliproteins in darkness and DCMU compare to the light conditions. Out of 13 types of phycobiliproteins, we have identified 10 redox-sensitive Cys residues of interest from 7 different phycobiliproteins; 3 are novel redox-sensitive proteins; 4 have been found by earlier studies; 2 additional phycobiliproteins are regulated by thioredoxin, but are not redox-sensitive. All of these redox-sensitive residues have a 1.5+ fold change for dark/light and DCMU/light. Consequently, phycobiliproteins were chosen for in-depth characterizations.

4.3 PBS Structure and Function

The light harvesting PBS complexes in cyanobacteria form massive multi-subunit structures on the stromal side of the thylakoid membrane. The major components of PBS complexes are brilliantly colored blue phycocyanin (PC), blue-green allophycocyanin (APC), and in some cyanobacteria, red phycoerythrin (PE). Multiple copies of these different phycobiliproteins and color-less linker proteins assemble into a 3-7 Mega Dalton holo-complex (Adir et al. 2006).

PBS complexes can account up to half of the total soluble proteins in cells and serve as a large internal nutrient reserve for cells (Grossman et al. 1993). During nutrient stresses, proteolysis of PBS promotes cell survival. Within hours after nitrogen, sulfur, or phosphorous starvation, cyanobacteria bleach as their PBS complexes are degraded (Allen and Smith 1969). This process frees up essential proteins for the maintenance of essential metabolic activities and uncouples PBS from the photosystems to prevent excessive photo-oxidative damage to the important photosynthetic protein machineries in the thylakoid membranes (Adir et al. 2003).

4.3.1 NblA Binding

Genetic screens for mutants that are unable to degrade their PBS complexes during nutrient starvation helped discovered NblA, a integral player in the molecular pathway of PBS degradation (Collier and Grossman 1994). As a soluble 7-kDa proteolysis adaptor, NblA, triggers the degradation of PBS by facilitating the interactions between a protease chaperone, ClpC, and a phycobliprotein (Karradt et al. 2008; Baier et al. 2004). But how NblA mediates PBS degradation is poorly understood. To investigate the binding partners of NblA, C-terminal His-tagged NblA was overexpressed, purified, bound to Ni-NTA resin, and incubated with isolated PBS. Phycobiliproteins that interacted with NblA were pull-downed and chemically cross-linked with an isotopically labeled BS^3 cross-linker. LC-MS/MS analysis was performed in

collaboration with proteomics experts from the MS facility at Washington University in St. Louis. A cross-link between the C-terminal end of NblA and N-terminal end of β-PC provided direct evidence NblA binds to the PBS complex. Specifically, NblA binds to the inner core cavity of the PC rod, a region that is structurally not readily accessible. Protein docking experiments with modeling experts in the Department of Biological Engineering at MIT demonstrated that NblA fits best in a groove between the α- and β-PC subunits that is potentially accessible from above or inside the PC rod structure. This finding insinuates that the PBS complex is dynamically changing to allow NblA access to the inner cavity of the PC rod. NblA may be able bind and disrupt the structural integrity of the PBS complexes via the alpha helical structural mimicry they share with phycobiliproteins (Dines et al. 2008), highly conserved amino acid residues (Karradt et al. 2008), electrostatic interactions, and/or van der Waals interactions.

4.3.2 PBS Degradation Mechanism

The proteolysis of the light harvesting antennas in cyanobacteria is initiated immediately after nutrient step-down as the expressions of the *nblA* gene (Collier and Grossman 1994) and NblA protein (Baier et al. 2001) are up-regulated. A possible NblA-mediated PBS degradation model could begin with the hetero-oligomerization of NblA with ClpC, an ATP-dependent chaperon (Stanne et al. 2007; Karradt et al. 2008). Upon forming large hexamers, the NblA-ClpC structure migrates and interacts with the PBS complexes in a way where the C-terminus of NblA binds to the N-terminus of β-PC, a region that is potentially accessible from above or inside interior hollow cavity of the PC rod (**Fig. 3.8**). Our protein docking modeling suggests that the K^{52} residue in NblA best fits in an internal groove within the PC structure that is near the T^{2} residue in β-PC. Upon NblA binding to β-PC, ClpC would be able to denature the phycobiliprotein in preparation for degradation by the Clp protease. The PBS complex must be

dynamically undergoing conformational changes for the NblA-ClpC hetero-oligomer to enter the internal hollow cavity of the PC rod.

4.4 Implications

Cyanobacteria are microscopic prokaryotes that have the potential to produce massive amounts of chemical energy by harvesting light energy to initiate photosynthesis. Combined with their remarkably fast growth and minimal nutrient requirements, cyanobacteria are ideal model organisms for biofuels research. Our investigations have provided insights into a significant signaling network in cyanobacteria. MS analysis of the *in vivo* redox reactions demonstrated thiol-based redox regulation and signaling control a broad range of cellular processes. In addition, our global map contains data on the specific protein thiols that are modified during redox stress. Knowledge of key redox active Cys residues in proteins that sensitive to redox signaling offers scientist the opportunity to strategically focus on key factors that affect their functional area of interest.

One area of great bioengineering interest is the optimization of photosynthesis by fine-tuning the light harvesting apparatuses in photosynthetic organisms during environmental stresses. Given how numerous phycobilproteins have redox-sensitive protein thiols, we explored the regulation of PBS complexes inactivation and turnover. NblA-mediated PBS degradation intriguing revolves around a very small proteolysis adaptor that could trigger the degradation of massive PBS complexes. An understanding of how NblA tags phycobiliproteins for degradation could potentially open the doors for precise manipulation of protein proteolysis. This could transform the world of synthetic and systems biology by degrading toxic protein intermediates build-up or inactivation of select proteins on cue.

4.5 Future Directions

Dynamic protein modifications and degradation facilitate cell survival during environmental perturbations. Recent advances in MS made it possible to analyze transient *in vitro* and *in vivo* protein modifications. The redox state of protein thiols have proven to play a crucial role in the activation/inactivation of protein functions. We have developed an optimized redox proteomics work-flow that captured on a large-scale the redox states of all free Cys residues in *Synechocystis* 6803. This global map of redox-sensitive protein thiols provides a comprehensive signaling network of an internal communication pathway cyanobacteria depend on to acclimate to light stress. However, characterizations of the redox-sensitive protein thiols are necessary to verify their role in protein function. In addition, our cross-linking and MS data had shown that NblA binds to β-PC in a region that is located in the hollow cavity of the PC rod. How NblA enters the inside of the PC rod is still a mystery.

4.5.1 Redox Signaling Network Map

In response to environmental stimuli, cells rely on redox signaling as an *in vivo* form of communication. Characterization of redox-sensitive protein thiols detected by our redox proteomics experiments would be a great future aim. Given the broad regulation of protein thiols, Cys residues from specific biological pathways could be strategically examined to create functional redox signaling network map. These network maps would be invaluable tools to deciphering the roles of the Cys proteome. For instance, our redox proteomics studies have elucidated numerous redox-sensitive protein thiols in various phycobiliproteins. Mutagenesis of these Cys residues combined with 77 K fluorescence emission spectroscopy could examine the energy transfer between PBS to PSII and PSI. Given that fluorescence occurs during the transfer from energy from one pigment to another pigment, increases in fluorescence may suggest increases in energy transfer.

In addition, Trx proteins are master regulators of redox signaling due to their ability to modify the redox states of protein thiols. By investigating the most redox-sensitive protein thiols from the global redox map we created in Chapter 2, we could discover new Trx-like proteins. These Cys residues in cyanobacteria could be mutated to serine, given their similar amino acid structures. Then the following analyses could help determine the roles of these Cys residues of these point mutants: growth rates, PSII/PSI ratios, kinetic spectroscopic measurements of PSII and PSI activities, oxygen evolution and consumption, and 77 K fluorescence emission spectroscopy. Discovery of novel Trx-like proteins would help paint a more complete picture of how the redox states of protein thiols are modified.

4.5.2 Protein Proteolysis

Our intriguing discovery that NblA interacts with the N-terminal end of β-PC, a region that is structurally located inside the hollow cavity of the PC rod, presents new research directions. A future investigation could explore the conformational changes within PC and APC. Any structural rearrangements that these linker proteins may have during nutrient stress are poorly understood. Several different color-less linker proteins play an integral role in assembling the massive PBS complexes together. Possible extensions of these linker proteins may promote openings in the PC subunits that allow NblA access to the interior regions of the PC rod. Another direction could examine where NblA interacts with APC. Whether NblA interacts with only the proteins located on the surface or if it could bind with the proteins inside the core is unknown. In addition, knowledge of how many copies of NblA that are necessary to initiate PBS degradation could provide immense insights on how a small NblA dimer could trigger such drastic remodeling of the massive light harvesting antennas in cyanobacteria. Lastly, the Clp proteasome system regulates protein turnover in countless organisms from eubacteria to plants, but little is

known how Clp proteins degrade PBS complexes. Continued research in NblA-mediated PBS degradation could possibly reveal information on how Clp proteins dispose unwanted proteins in cyanobacteria.

4.5.2 Application to Industry Needs

How photosynthetic microorganisms optimize their cellular metabolism in response to adverse environmental conditions is an area of great interest in the field of bioenergy. A major challenge to engineering microbial factories for production of biochemicals is the low yield, which is cost-prohibitive for large-scale production. The ability to control the level of outputs in an organism by fine-tuning the redox states of protein thiols and protein turnover would promote innovation in mixotrophic microbial platforms for energy production.

Future high performing microbial chassis could have three key features. First, dynamic modulations between rapid cell growth to synthesis of a highly valued product, depending on the levels of an exogenous inducer. Second, programmed protective mechanisms through modifications of protein thiols and protein degradation to minimize levels of lethal intermediates. Third, a molecular sensor that prevents low-performing cells access to nutrients, thus leading to cell death, to account for productivity variations between cells (Xiao et al. 2016). Only cells that produce high titers of a preferred molecule can consume nutrients and live. These fine-tuning capabilities will decrease production time and enable efficient usage of nutrients to drastically reduce production costs. Overall, cross-functional teams could create a high throughput pipeline that designs, synthesizes, and tests cells. The objective is to identify strains with fast growth, low nutrient needs, and high production of bioenergy, which are all crucial features in the creation of scalable microbial factories.

4.6 References

Adir, N., Dines, M., Klartag, M., McGregor, A., Melamed-Frank, M. (2006). "Assembly and disassembly of phycobilisomes." *Complex Intracellular Structures in Prokaryotes*. Springer Berlin Heidelberg, Berlin, Heidelberg: 47-77.

Adir, N., Zer, H., Shochat, S., Ohad, I. (2003). "Photoinhibition - a historical perspective." *Photosynthesis Research*. **76**: 343-370.

Ajlani, G., Vernotte, C. (1998). "Construction and characterization of a phycobiliprotein-less mutant of *Synechocystis* sp. PCC 6803." *Plant Molecular Biology*. **37**(3): 577-580.

Ajlani, G., Vernotte, C., DiMagno, L., Haselkorn, R. (1995). "Phycobilisome core mutants of *Synechocystis* sp. PCC 6803." *Biochimica et Biophysica Acta - Bioenergetics*. **1231**(2): 189-196.

Allen, M.M., Smith, A.J. (1969). "Nitrogen chlorosis in blue-green algae." *Archives Microbiology*. **69**(2): 114-120.

Andersson, F.I., Blakytny, R., Kirstein, J., Turgay, K., Bukau, B., Mogk, A., Clarke, A.K. (2006). "Cyanobacterial ClpC/HSP100 protein displays intrinsic chaperone activity." *Journal of Biological Chemistry*. **281**(9): 5468-5475.

Baier, A., Winkler, W., Korte, T., Lockau, W., Karradt, A. (2014). "Degradation of phycobilisomes in *Synechocystis* sp. PCC 6803: evidence for essential formation of an NblA1/NblA2 heterodimer and its codegradation by a Clp protease complex." *Journal of Biological Chemistry*. **289**(17): 11755-11766.

Baier, K., Lehmann, H., Stephan, D.P., Lockau, W. (2004). "NblA is essential for phycobilisome degradation in *Anabaena s*p. strain PCC 7120 but not for development of functional heterocysts." *Microbiology*. **150**: 2739-2749.

Baier, K., Nicklisch, S., Grundner, C., Reinecke, J., Lockau, W. (2001). "Expression of two *nblA*-homologous genes is required for phycobilisome degradation in nitrogen-starved *Synechocystis* sp. PCC 6803." *FEMS Microbiology Letters*. **195**:35-39

Bakan, A., Meireles, L.M., Bahar, I. (2011). "ProDy: protein dynamics inferred from theory and experiments." *Bioinformatics*. **27**(11): 1575-1577.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., (2000). "The Protein Data Bank." *Nucleic Acids Research*. **28**(1): 235-242.

Bienert, R., Baier, K., Volkmer, R., Lockau, W., Heinemann, U. (2005). "Crystal structure of NblA from *Anabaena* sp. PCC 7120, a small protein playing a key role in phycobilisome degradation." *Journal of Biological Chemistry*. **281**(8): 5216-5223.

Bogorad, L. (1975). "Phycobiliproteins and Complementary Chromatic Adaptation." *Annual Review of Plant Physiology*. **26**(1): 369-401.

Bricker, T.M., Mummadisetti, M.P., Frankel, L.K. (2015). "Recent advances in the use of mass spectrometry to examine structure/function relationships in photosystem II." *Journal of Photochemistry and Photobiology.* **15**(Part B): 227-246.

Buchanan, B.B. (1991). "Regulation of CO_2 assimilation in oxygenic photosynthesis: the ferredoxin/thioredoxin system. Perspective on its discovery, present status, and future development." *Archives of Biochemistry and Biophysics*. **288**(1): 1-9.

Buchanan, B.B., Balmer, Y. (2005). "Redox regulation: a broadening horizon." *Annual Review of Plant Biology.* **56**: 187-220.

Collier, J.L., Grossman, A.R. (1992). "Chlorosis induced by nutrient deprivation in *Synechococcus* sp. strain PCC 7942: not all bleaching is the same." *Journal of Bacteriology*. **174**(14): 4718-4726.

Collier, J.L., Grossman, A.R. (1994). "A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria." *EMBO Journal*. **13**(15): 1039-1047.

Dalle-Donne, I., Rossi, R., Colombo, G., Giustarini, D., Milzani, A. (2009). "Protein S-glutathionylation: a regulatory device from bacteria to humans." *Trends in Biochemical Sciences*. **34**(2): 85-96.

Dines, M., Sendersky, E., David, L., Schwarz, R., Adir, N. (2008). "Structural, functional, and mutational analysis of the NblA protein provides insight into possible modes of interaction with the phycobilisome." *Journal of Biological Chemistry*. **283**(44): 30330-30340.

Dines, M., Sendersky, E., Schwarz, R., Adir, N. (2007). "Crystallization of sparingly soluble stress-related proteins from cyanobacteria by controlled urea solublization." *Journal of Structural Biology*. **158**(1): 116-121.

Elmorjani, K., Thomas, J.C., Sebban, P. (1986). "Phycobilisomes of wild type and pigment mutants of the cyanobacterium *Synechocystis* PCC 6803." *Archives Microbiology*. **146**(2): 186-191.

Glazer, A.N. (1982). "Phycobilisomes: structure and dynamics." *Annual Review of Microbiology*. **36**: 173-198.

Gorl, M., Sauer, J., Baier, T., Forchhammer, K. (1998). "Nitrogen-starvation-induced chlorosis in *Synechococcus* PCC 7942: adaptation to long-term survival." *Microbiology*. **144**(Pt 9): 2449-2458.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., Baker, D. (2003). "Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations." *Journal of Molecular Biology*. **331**(1): 281-299.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., Collier, J.L. (1993). "The phycobilisome, a light-harvesting complex responsive to environmental conditions." *Microbiological Reviews*. **57**(3): 725-749.

Karradt, A., Sobanski, J., Mattow, J., Lockau, W., Baier, K. (2008). "NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease." *Journal of Biological Chemistry*. **283**(47): 32394-32403.

Kirstein, J., Schlothauer, T., Dougan, D.A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B., Turgay, K. (2006). "Adaptor protein controlled oligomerization activates the AAA+ protein ClpC." *EMBO Journal*. **25**(7): 1481-1491.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., Davis, I.W., Cooper, S., Treuille, A., Mandell, D.J., Richter, F., Ban, Y.E., Fleishman, S.J., Corn, J.E., Kim, D.E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J.J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J.J., Kuhlman, B., Baker, D., Bradley, P. (2011). "ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules." *Methods in Enzymology*. **487**: 545-574.

Leaver-Fay, A., Tyka, M., Jacak, R., Song, Y., Kellogg, E.H., Thompson, J., Davis, I.W., Pache, R.A., Lyskov, S., Gray, J.J., Kortemme, T., Richardson, J.S., Havranek, J.J., Snoeyink, J., Baker, D., Kuhlman, B. (2013) "Scientific benchmarks for guiding macromolecular energy function improvement." *Methods in Enzymology*. **523**: 109-143.

Leitner A, Walzthoeni T, Aebersold R (2014). "Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline." *Nature Protocols*. **9**(1): 120-137.

Li, H., Sherman, L.A. (2002). "Characterization of *Synechocystis* sp. strain PCC 6803 and deltanbl mutants under nitrogen-deficient conditions." *Archives Microbiology*. **178**(4): 256-266.

Liu, H., Zhang, H., Niedzwiedzki, D.M., Prado, M., He, G., Gross, M.L., Blankenship, R.E. (2013). "Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria." *Science.* **342**(6162): 1104-1107.

Liu, H., Zhang, H., Orf, G.S., Lu, Y., Jiang, J., King, J.D., Wolf, N.R., Gross, M.L., Blankenship, R.E. (2016). "Dramatic domain rearrangements of the cyanobacterial orange carotenoid protein upon photoactivation." *Biochemistry*. **55**(7): 1003-1009.

Luque, I., Ochoa de Alda, J.A., Richaud, C., Zabulon, G., Thomas, J.C., Houmard, J. (2003). "The NblAI protein from the filamentous cyanobacterium *Tolypothrix* PCC 7601: regulation of its expression and interactions with phycobilisome components." *Molecular Microbiology*. **50**(3): 1043-1054.

MacColl, R. (1998). "Cyanobacterial phycobilisomes." *Journal of Structural Biology*. **124**(2-3): 311-334.

Marx, A., Adir, N. (2013). "Allophycocyanin and phycocyanin crystal structures reveal facets of phycobilisome assembly." *Biochimica et Biophysica Acta*. **1827**(3): 311-318.

Olinares, P.D., Kim, J., van Wijk, K.J. (2011). "The Clp protease system; a central component of the chloroplast protease network." *Biochim et Biophys Acta.* **1807**(8): 999-1011.

Porankiewicz, J., Wang, J., Clarke, A.K. (1999). "New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond." *Molecular Microbiology*. **32**(3): 449-458.

Rappsilber, J. (2011). "The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes." *Journal of Structural Biology*. **173**(3): 530-540.

Richaud, C., Zabulon, G., Joder, A., Thomas, J.C. (2001). "Nitrogen or sulfur starvation differentially affects phycobilisome degradation and expression of the *nblA* gene in *Synechocystis* strain PCC 6803." *Journal of Bacteriology*. **183**(10): 2989-2994.

Sambrook, J., Russell, D.W. (2001). "Molecular Cloning: A Laboratory Manual." *Cold Spring Harbor Laboratory Press*. **1**(4): 1-1881.

Sauer, J., Schreiber, U., Schmid, R., Volker, U., Forchhammer, K. (2001). "Nitrogen starvation-induced chlorosis in *Synechococcus* PCC 7942. Low-level photosynthesis as a mechanism of long-term survival." *Plant Physiology*. **126**(1): 233-243.

Scheer, H., Zhao, K.H. (2008). "Biliprotein maturation: the chromophore attachment." *Molecular Microbiology*. **68**(2): 263-276.

Schwarz, R., Forchhammer, K. (2005). "Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular responses." *Microbiology*. **151**(Pt 8): 2503-2514.

Sendersky, E., Kozer, N., Levi, M., Garini, Y., Shav-Tal, Y., Schwarz, R. (2014). "The proteolysis adaptor, NblA, initiates protein pigment degradation by interacting with the cyanobacterial light-harvesting complexes." *Plant Journal*. **79**(1): 118-126.

Sendersky, E., Kozer, N., Levi, M., Moizik, M., Garini, Y., Shav-Tal, Y., Schwarz, R. (2015). "The proteolysis adaptor, NblA, is essential for degradation of the core pigment of the cyanobacterial light-harvesting complex." *Plant Journal*. **83**(5): 845-852.

Sidler, W. (2004). “Phycobilisome and phycobiliprotein structures.” *The Molecular Biology of Cyanobacteria*. **1**: 139-216.

Sinz, A. (2006). “Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions.” *Mass Spectrometry Reviews*. **25**(4): 663-682.

Sinz, A. (2014). “The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks.” *Expert Review of Proteomics*. **11**(6): 733-743.

Stamler, J.S., Lamas, S., Fang, F.C. (2001). “Nitrosylation: the prototypic redox-based signaling mechanism.” *Cell*. **106**(6): 675-683.

Stanne, T.M., Pojidaeva, E., Andersson, F.I., Clarke, A.K. (2007). “Distinctive types of ATP-dependent Clp proteases in cyanobacteria.” *Journal of Biological Chemistry.* **282**(19): 14394-14402.

Wang, C., Bradley, P., Baker, D. (2007). “Protein-protein docking with backbone flexibility.” *Journal of Molecular Biology*. **373**(2): 503-519.

Xiao, Y., Bowen, C.H., Liu, D., Zhang, F. (2016). “Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis.” *Nature Chemical Biology*. **12**(5): 339-344.

Yu, J., Liberton, M., Cliften, P.F., Head, R.D., Jacobs, J.M., Smith, R.D., Koppenaal, D.W., Brand, J.J., Pakrasi, H.B. (2015). “*Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO_2.” *Scientific Reports*. **5(**8132): 1-10.

Appendix A

Structural Basis and Evolution of Regulation in Plant Adenosine-5'-Phosphosulfate Kinase

This work was originally published in the *Proceedings of National Academy of Sciences*. Geoffrey E. Ravilious, Amelia Y. Nguyen*, Julie A. Francois, Joseph M. Jez. Structural Basis and Evolution of Redox Regulation in Plant Adenosine-5'-Phosphosulfate Kinase. *Proceedings of National Academy of Sciences*. 2012. 109(1):309-14.

*A.Y.N. cloned, overexpressed, and purified SynAPSK, performed the kinetics experiments, and analyzed the data.

A.1 Introduction

A.1.1 Abstract

Adenosine-5'-phosphosulfate (APS) kinase (APSK) catalyzes the phosphorylation of APS to 3'-phosphoadenosine-5'-phosphosulfate (PAPS). In *Arabidopsis thaliana*, APSK is essential for reproductive viability and competes with APS reductase to partition sulfate between the primary and secondary branches of the sulfur assimilatory pathway; however, the biochemical regulation of APSK is poorly understood. The 1.8 Å resolution crystal structure of AtAPSK in complex with AMP-PNP, Mg^{2+}, and APS provides the first view of the Michaelis complex for this enzyme and reveals the presence of an intersubunit disulfide bond between Cys86 and Cys119. Functional analysis of AtAPSK demonstrates that reduction of Cys86-Cys119 resulted in a 17-fold higher k_{cat}/K_m and a 15-fold increase in K_i for substrate inhibition by APS compared to the oxidized enzyme. The C86A/C119A mutant was kinetically similar to the reduced wild-type enzyme. Gel- and activity-based titrations indicate that the midpoint potential of the disulfide in AtAPSK is comparable to that observed in APS reductase. Both cysteines are invariant among the APSK from plants, but not other organisms, which suggests redox-control as a novel regulatory feature of the plant APSK. Based on structural, functional, and sequence analyses, we propose that the redox-sensitive APSK evolved after bifurcation of the sulfur assimilatory pathway in the green plant lineage and that changes in redox environment resulting from oxidative stresses may affect partitioning of APS into the primary and secondary thiol metabolic routes by having opposing effects on APSK and APS reductase in plants.

A.1.2 Sulfur Metabolism

Sulfur is an essential element for all living organisms and is required for the biosynthesis of a diverse array of metabolites and macromolecules (Leustek et al. 2000; Yi et al. 2010a; Yi et al. 2010b; Takahashi et al. 2011). Plants and prokaryotes are the primary assimilatory organisms

that convert inorganic sulfate (SO_4^{2-}), the predominant form of environmental sulfur, into physiologically useful forms of sulfur (Patron et al. 2008; Marzluf 1997). The metabolic organization of sulfur assimilation varies among plants and microbes (**Fig. A.1**) (Leustek et al. 2000; Yi et al. 2010a; Yi et al. 2010b; Patron et al. 2008; Takahashi et al. 2011; Marzluf 1997; Sekowska et al. 2000; Bick et al. 2000). In yeast, fungi, and enterobacteria, including *Escherichia coli,* sulfate is incorporated into adenosine-5'-phosphate (APS), then converted to 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as a biologically "activated" compound that is reduced to sulfide (Marzluf 1997; Sekowska et al. 2000). In other sulfate-assimilating bacteria, such as *Pseudomonas aeruginosa*, APS can be used for reduction of sulfide (Bick et al. 2000). Plants possess bifurcated thiol metabolism pathways, which may reflect the metabolic needs of sessile organisms adapted to a range of environmental stresses and nutrient fluctuations (**Fig. A.1**) (Patron et al. 2008). These pathways branch after formation of APS. The primary sulfur metabolic route in plants uses APS as the activated high-energy compound for sulfur reduction and the production of cysteine, which is crucial for the synthesis of proteins, methionine, iron-sulfur clusters, vitamin cofactors, and compounds that protect against oxidative stresses, including glutathione and phytochelatin peptides (Noctor and Foyer 1998; Meyer 2008; Cobbett and Goldsbrough 2002; Galant et al. 2011). Alternatively, APS can be converted into PAPS to provide a sulfate-donor for the modification of multiple natural products, brassinosteroid and jasmonate hormones, phytosulfokines, and other sulfonated molecules (Klein and Papenbrock 2004; Halkier and Gershenzon 2006; Amano et al. 2007). Partitioning of sulfate at the branch-point between primary (reductive) and secondary metabolic pathways in plants is controlled by APS reductase (APSR) and APS kinase (APSK), respectively (Rausch and Wachter 2005; Mugford et al. 2011).

APSR catalyzes the glutathione-dependent reduction of APS to sulfite (SO_3^{2-}) and AMP. Extensive studies of APSR in plants demonstrate a critical role for this enzyme in regulating flux through the primary sulfur assimilatory pathway (Bick et al. 2001; Martin et al. 2005; Loudet et al. 2007; Scheerer et al. 2010). Modulation of activity by redox changes in the APSR from *Arabidopsis thaliana* (AtAPSR) occurs through a regulatory disulfide bond, which upon reduction attenuates activity (Bick et al. 2001). Additional evidence suggests that APSR is important for supplying reduced sulfur under stress conditions, such as nutrient deprivation and chilling (Phartiyal et al. 2008). Conversely, application of cysteine and glutathione to Arabidopsis results in decreased APSR transcript levels and activity (Vauclare et al. 2002). In contrast to APSR, the role of APSK in plant thiol metabolism has only begun to be examined.

Recent studies in Arabidopsis revealed that APSK is essential for plant reproduction and viability (Mugford et al. 2011; Mugford et al. 2009; Kopriva et al. 2009; Mugford et al. 2010). Use of T-DNA knockout lines targeting each of the four AtAPSK isoforms established a connection between this branch of sulfur metabolism and glucosinolate biosynthesis (Mugford et al. 2009; Yatusevich et al. 2010). Subsequent analysis of triple AtAPSK knockout lines demonstrated that AtAPSK1 alone could maintain wild-type levels of growth and development in Arabidopsis (Mugford et al. 2010). Moreover, measurements of flux through the primary sulfur assimilation pathway was increased when APSK activity was reduced, which implies that alterations in either APSR or APSK activity aids in partitioning of sulfur between the two branches of thiol metabolism (Mugford et al. 2011).

In plants, the biochemical regulation of APSK is not well understood and no study to date has linked stress conditions to the control of its activity. APSK catalyzes the transfer of the γ-phosphate from ATP to the 3'-hydroxyl group of the APS adenine ring to yield PAPS and ADP.

The reaction sequence suggested by kinetic studies of the APSK from *Penicillium chrysogenum, E. coli*, and *A. thaliana* follows an ordered mechanism with substrate inhibition by APS (Satishchandran and Markham 1989; Lillig et al. 2001; MacRae and Segel 1999). Although no structural information for a plant APSK is available, x-ray crystal structures of the *P. chrysogenum* APSK and bifunctional ATP sulfurylase-APSK from human, *Aquifex aeolicus* and *Thiobacillus denitrificans* have been determined and reveal a canonical a/b purine nucleotide binding fold (MacRae et al. 2000; Lansdon et al. 2002; Yu et al. 2007; Gay et al. 2009; Sekulic et al. 2007a; Sekulic et al. 2007b). For the bifunctional human enzyme, the substrate inhibition by APS on APSK activity is linked to a ~20 amino acid N-terminal loop of low sequence homology (Sekulic et al. 2007b); **Fig. A.2**); however, the general role of this region across different species remains unclear. In addition, earlier studies of AtAPSK suggest that oxidation reduces specific activity (Lillig et al. 2001).

To better understand the molecular function of the plant APSK, we determined the three-dimensional structure of AtAPSK isoform 1, lacking its chloroplast localization leader, in complex with b,g-imidoadenosine-5′-triphosphate (AMP-PNP), Mg^{2+}, and APS. Although the overall structure is similar to that of the APSK from fungi and humans, the AtAPSK•AMP-PNP•Mg^{2+}•APS complex new insight on the reaction sequence of this enzyme and reveals an intersubunit disulfide bond formed between Cys86 in the N-terminal loop of one subunit and Cys119 in the adjacent monomer. Kinetic analysis and redox titrations of wild-type and the C86A/C119A mutant of AtAPSK indicate that the disulfide bond forms at a physiologically relevant redox potential and affects catalytic efficiency and substrate inhibition by APS. Because sequence comparisons suggest a novel regulatory element of the plant APSK, we also functionally compare AtAPSK and the APSK from the cyanobacterium *Synechocystis* sp. PCC

6803, which is evolutionarily related to plant chloroplasts (Raven and Allen 2003), to demonstrate that a homolog lacking the cysteines is not redox-sensitive. Based on structural, functional, and sequence analyses, we propose that evolution of a thiol-based redox switch in the plant APSK evolved with bifurcation of the sulfur assimilatory pathway later in the plant lineage and that changes in redox environment resulting from oxidative stresses may coordinate flux between the primary and secondary thiol metabolic routes by modulating APSR and APSK activity in plants.

A.2 Materials and Methods

A.2.1 Characterization of APSK from *Synechocystis* 6803

Generation of the pET-28a-AtAPSKΔ77 bacterial expression vector, which encodes AtAPSK isoform 1 lacking the plastid localization sequence and with an N-terminal hexahistidine tag, was previously described (Phartiyal et al. 2006). Transformed *E. coli* BL21(DE3) were grown at 37 °C in Terrific broth containing 50 μg mL^{-1} kanamycin until A_{600nm} ~ 0.8. After induction with 1 mM isopropyl 1-thio-β-D-galactopyranoside, the cultures were grown at 20 ^{o}C overnight. Cells were pelleted by centrifugation and resuspended in 50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol (βME), 10% (v/v) glycerol and 1% Tween-20. After sonication and centrifugation, the supernatant was passed over a Ni^{2+}-nitriloacetic acid (Qiagen) column. The column was washed with buffer minus Tween-20 and the bound protein eluted using 250 mM imidazole in wash buffer. The eluent was dialyzed overnight at 4 °C against 25 mM HEPES (pH 7.5), 200 mM KCl, 5% glycerol, and 5 mM βMe, and then loaded onto a Superdex-200 26/60 HiLoad FPLC size-exclusion column equilibrated in the same buffer. Fractions corresponding to the major protein peak were pooled, and judged to be >95% pure by SDS-PAGE. Protein was concentrated (Amicon) to 10 mg mL^{-1} with protein concentration

determined using a molar extinction coefficient (ε_{280nm} = 22,430 M^{-1} cm^{-1}) calculated in ProtParam (http://web.expasy.org/protparam). Protein was dialyzed against 25 mM HEPES (pH 7.5), 200 mM KCl, 10% glycerol, and 5 mM βMe, flash frozen in liquid nitrogen, and stored at -80 °C. For crystallization and redox titrations, βMe was removed by buffer exchange.

A.2.2 Crystallography

Crystals of the AtAPSK•AMP-PNP•Mg^{2+}•APS complex grew at 4°C in hanging drops from a 1:1 ratio of protein pre-incubated with 5 mM APS, 5 mM AMP-PNP, and 10 mM $MgCl_2$ and crystallization buffer (100 mM HEPES (pH 7.5), 200 mM $MgCl_2$, and 15-17.5% PEG 2,000). For data collection, crystals were transferred to cryoprotectant (mother liquor containing ligands and 20% glycerol) and then frozen in liquid nitrogen. Diffraction data (100 K) were collected at beam 19-ID of the Advanced Photon Source - Argonne National Laboratory. Diffraction intensities were integrated, merged, and scaled using the HKL3000 software suite (Otwinowski and Minor 1997). The structure of AtAPSK was determined by molecular replacement with PHASER (McCoy et al. 2007) using the structure of APSK from *P. chrysogenum* (PDB: 1M7G; (Lansdon et al. 2002)) as the search model. After iterative rounds of model building in COOT (Emsley et al. 2010) and refinement in PHENIX (Adams et al. 2010), the R-factors converged to those reported in Table S1. Coordinates and structure factors for the AtAPSK•AMP-PNP•Mg^{2+}•APS complex have been deposited in the RCSB Protein Data Bank.

A.2.3 Mutagenesis

The AtAPSK C86A/C119A mutant was generated using the QuikChange PCR method (Stratagene) with pET-28a-AtAPSKΔ77 as the template. Two pairs of oligonucleotides (**Table A.1**) were used to sequentially generate the double-mutant. Mutations were confirmed by

sequencing (Washington University DNA Sequencing Facility). Protein expression and purification were performed as described for wild-type protein.

A.2.4 Enzyme Assays and Redox Titrations

Enzymatic activity of purified wild-type and mutant AtAPSK was determined spectrophotometrically using a coupled assay (MacRae and Segel 1999). Steady-state kinetic parameters were determined by initial velocity experiments with data fitted to a general substrate inhibition model, $v = (V_{max}[S])/(K_M+[S]*((1+[S])/K_i)$. The effect of redox potential on AtAPSK activity was examined with assays containing defined ratios of reduced (GSH) and oxidized (GSSG) glutathione (Sigma-Aldrich). Protein (5 μM) was equilibrated in degassed 100 mM HEPES (pH 7.5 or 7.0), 200 mM KCl, 5% glycerol and 20 mM GSH/GSSG for 60 min at 25 ^{o}C. Aliquots were removed and transferred to the coupled assay (final protein concentration of 10-30 nM) and initial velocities measured. Values for E_M were determined by fitting titration data to the Nernst equation, $E_h = E_m + (RT/nF)(\ln([GSSG]/[GSH]^2)$ with RT/F = 25.7 mV (44, 47) and *n* = 2 using Kaleidagraph (Synergy Software). The effect of redox potental on AtAPSK was also evaluated by non-reducing SDS-PAGE. Protein was equilibrated with varying ratios of GSSG/GSH, DTT_{red}, or DTT_{ox}, and then separated into reduced and oxidized forms by SDS-PAGE. The fraction of reduced versus oxidized AtAPSK was calculated by quantifying the fractions of oxidized cross-linked (44 kDa) and reduced monomeric (22 kDa) species by densiometry (Quantity One software; BioRad).

A.2.5 Cloning, Expression, Purification, and Analysis of APSK from *Synechocystis* 6803

The cDNA encoding APSK from *Synechocystis* sp PCC. 6803 was PCR-amplified using PfuUltraHF polymerase and gene-specific oligonucleotides (**Table A.1**). The PCR product was

digested and cloned into the NheI and EcoRI sites of pET28a. Protein expression, purification, and activity assays were performed as described for AtAPSK.

A.3 Results

A.3.1 Overall Structure

AtAPSK isoform 1 (residues 78-276; (Phartiyal et al. 2006)) lacking the 77 amino acid N-terminal chloroplast localization sequence was used for protein crystallography. The structure of AtAPSK in complex with AMP-PNP, Mg^{2+}, and APS was solved by molecular replacement using the *P. chrysogenum* APSK (Lansdon et al. 2002) as a search model with three monomers in the asymmetric unit (**Table A.2** and **Fig. A.3**). Chain A forms a crystallographic dimer and chains B and C yield a second non-crystallographic dimer.

The overall fold of each monomer in the dimer consists of a variable N-terminal region (residues 80-98), which includes a1 and entwines the adjacent monomer, a canonical a/b-purine nucleotide binding domain (b1-a2-b2-a3-b3-a4-b4-a7), and a small domain, which includes a5 and a6, that caps the nucleotide binding sites (**Fig. A.3a**). Although AtAPSK shares a common fold with the bifunctional PAPS synthetase from human (1.2 Å r.m.s.d. for 187 C_a-atoms; 51% sequence identity; 35) and the *P. chrysogenum* APSK (1.3 Å r.m.s.d. for 176 C_a-atoms; 55% sequence identity; 32), the positioning of the N-terminal region differs in these structures (**Fig. A.3b**). This region is characterized by an unstructured loop followed by a ~7 amino acid a-helix (a1) in various APSK structures (Lansdon et al. 2002; Yu et al. 2007; Gay et al. 2009; Sekulic et al. 2007a; Sekulic et al. 2007b) and displays sequence divergence between the APSK homologs from bacteria, cyanobacteria, green algae, fungi, plants, and human (**Fig. A.2).**

In the initial $2F_o$-F_c maps of the AtAPSK structure, clear electron density for AMP-PNP, APS, and Mg^{2+} was observed (**Fig. A.3c**). Additional contiguous density was observed between

Cys86 in monomer B and Cys119 of a2 in monomer C (**Fig. A.3d**), Cys86 in monomer C and Cys119 in monomer B, and between Cys86 of monomer A and Cys119 in the crystallographic symmetry-mate of monomer A. The distance between the sulfur atoms of each cysteine falls within the range of expected bond length (~2.05 Å) for a disulfide linkage. The Cys86(A)-Cys119(A) and Cys86(B)-Cys119(C) disulfides were fully oxidized and the Cys86(C)-Cys119(B) disulfide partially oxidized. The Cys forming the disulfide are invariant across the APSK from plants and mosses, but are generally missing in the homologs from other organisms (**Fig. A.2**).

A.3.2 Structure of the Active Site and N-terminal Loop

The structure of AtAPSK with a non-hydrolyzable ATP analog (AMP-PNP), Mg^{2+}, and APS bound in the active site provides the first view of the Michaelis complex in the APSK reaction sequence (**Fig. A.4**). APS is locked in the active site through multiple binding contacts (**Fig. A.4**). The adenine ring is stacked between Phe150 and Phe232. Arg141 and Asn158 provide bridging interactions between the sulfate and phosphate groups of the substrate. Arg155 and the backbone nitrogen of Ile181 also interact with the sulfate and phosphate groups, respectively. Asp138 forms a bidentate interaction with the ribose hydroxyl groups to position the 3'-OH proximal to the g-phosphate group of AMP-PNP for the ensuing phosphoryl transfer reaction.

Binding of AMP-PNP is mediated almost exclusively through the phosphate moieties of the ligand to residues in the active site and an extensive water network centered on the bound Mg^{2+} ion (**Fig. A.4**). The adenine ring is positioned between Pro257 and Arg215. Extending into the active site, phosphate groups of AMP-PNP form multiple contacts with main-chain atoms of residues 110-116 in the P-loop (39). The a-phosphate group is bound by interactions with the

side-chain hydroxyl group and backbone nitrogen of Thr116 and the backbone nitrogen of Gly113. Lys114 contacts both the b- and g-phosphates with Ser110 and Lys228 forming additional interactions to the g-phosphate. The b- and g-phosphates of AMP-PNP, three water molecules, the hydroxyl-groups of Ser115 and Ser179, and the carboxylate of Asp136 octahedrally coordinate the Mg^{2+} ion. Asp136 and Asp138 form part of the DGDN-loop that bridges the APS and ATP•Mg^{2+} binding sites and are critical for catalytic activity in APSK (MacRae et al. 2000; Sekulic et al. 2007a; Singh and Schwartz 2003).

Although not immediately in contact with ligands in the AtAPSK active site, part of the N-terminal loop (residues 80-84) of the adjacent monomer abuts the DGDN-loop and the side-chain of His84 hydrogen bonds to a water molecule in the network that coordinates the Mg^{2+} ion (**Fig. A.4**). The structure of the AtAPSK N-terminal region is nearly identical to that of the APSK domain from human PAPS synthetase in complex with ADP and PAPS, but differs from the *P. chrysogenum* APSK apoenzyme and ADP•APS complex (**Fig. A.3b**). Changes in the N-terminal region of PAPS synthetase were suggested to promote conformational changes that stabilize interactions between the DGDN-loop and substrates during the catalytic cycle (Yu et al. 2007).

Interestingly, the average B-factors for the N-terminal loop residues (80-90) in the monomers containing the fully oxidized Cys86(A)-Cys119(A) and Cys86(B)-Cys119(C) disulfides were 39 $Å^2$ and 47 $Å^2$, respectively, which were lower than the average B-factor of 56 $Å^2$ in the monomer with the partially oxidized Cys86(C)-Cys119(B) disulfide bond. Crystallographic analysis of the Cys86-Cys119 disulfide bond in AtAPSK suggests that redox-changes may alter the positioning and/or mobility of the N-terminal loop near the active site and potentially affect enzymatic activity.

A.3.3 The Cys86-Cys119 Disulfide and Redox-regulation

Formation of the Cys86-Cys119 disulfide linkage in the N-terminal region of AtAPSK may have a functional role in modulating enzymatic activity. Non-reducing SDS-PAGE analysis of AtAPSK incubated with either dithiothreitol (DTT_{red}) or trans-4,5-dihydroxy-1,2-dithiane (DTT_{ox}) showed the enzyme as reduced monomeric (22 kDa) and oxidized dimeric (44 kDa) forms, respectively (**Fig. A.5a**). The same analysis using the C86A/C119A mutant confirmed that mutation of the cysteines prevented cross-linking of the two monomers (**Fig. A.5a**). Both the wild-type and C86A/C119A mutant AtAPSK migrated as dimers upon size-exclusion chromatography in either the presence or absence of reducing agents.

Kinetic analysis of AtAPSK incubated in the presence of either DTT_{red} or DTT_{ox} demonstrated that reduction of the disulfide bond resulted in enhanced catalytic efficiency (k_{cat}/K_m) and a decreased effect of substrate inhibition by APS (**Table A.3** & **Fig. A.6**). Reduced AtAPSK was 17-fold more efficient and had a K_i value for APS 15-fold higher compared to the oxidized protein (**Table A.3**). Moreover, the K_i values of the reduced protein increased dramatically with decreasing ATP concentrations (**Fig. A.6d**); however, substrate inhibition by APS was comparable at all ATP concentrations for the oxidized enzyme (**Fig. A.6e**). Steady-state kinetic analysis of the C86A/C119A mutant showed similar kinetic parameters for both forms of the enzyme that were comparable to those of the reduced AtAPSK (**Table A.3**). The C86A/C119A mutant treated with DTT_{red} showed little difference in kinetic parameters with respect to $AtAPSK_{red}$. Similarly, treatment of the mutant with DTT_{ox} caused a modest 1.5-fold decrease in activity compared to DTT_{red}, but had little effect on either K_m or K_i. These results suggest that redox environment may have a role in regulating the activity of AtAPSK.

To determine the redox midpoint potential (E_m) for the Cys86-Cys119 disulfide, titrations using a gel-based system and activity assays were performed (**Figs. 4b & 4c**). AtAPSK was

incubated in solutions containing varied ratios of oxidized:reduced glutathione (GSSG:GSH; total concentration = 20 mM). Aliquots from incubations (E_h = -340 to -240 mV) were analyzed by non-reducing SDS-PAGE and activity assays. Triplicate titrations of AtAPSK at pH 7.5 yielded average E_m values of -295 ± 12 mV and -286 ± 18 mV in the gel-and activity-based assays, respectively. At pH 7, the titrations yielded average E_m values of -249 ± 9 mV and -260 ± 17 mV in the gel-and activity-based assays, respectively. Compared to the E_m values of other redox-regulated plant enzymes, which range from -390 to -237 mV (18, 41-45), these values suggest that changes in redox environment may modulate APSK activity *in vivo,* which could provide a strategy for modulating flux through the primary and secondary sulfur metabolism pathways in plants.

A.3.4 Characterization of APSK from *Synechocystis* 6803

Comparison of APSK sequences indicates that the residues corresponding to Cys86 and Cys119 are invariant in the enzymes from monocot and dicot plants and the mosses *Selaginella moellendorffii* and *Physcomitrella patens* (**Fig. A.2b**). Cys119 is conserved in the APSK from green alga *Chlorella variabilis* and *Chlamydomonas reinhardtii*, but is not found in the enzyme from cyanobacteria *Synechocystis* and *Cyanothece* (**Fig. A.2a**). These sequence alignments suggest that redox-regulation of APSK appeared later in the lineage of photosynthetic organisms. To confirm that an evolutionarily earlier homolog is not redox-sensitive, the APSK from *Synechocystis* sp. PCC 6803 was cloned, expressed, purified, and assayed. Steady-state kinetic analysis of SynAPSK in the presence of either DTT_{red} or DTT_{ox} indicates that its activity does not change with redox environment (**Table A.3**).

A.4 Discussion

The sulfur assimilation pathway in plants supports sulfur reduction and the synthesis of cysteine in the primary metabolic branch and the generation of sulfonated molecules for specialized metabolism in the secondary branch. Partitioning of sulfate at the branchpoint between primary and secondary metabolic pathways in plants is balanced by the activities of APSR and APSK (**Fig. A.1**) (Rausch and Wachter 2005; Mugford et al. 2011). Recent work reveals an essential role of APSK for the reproductive viability of Arabidopsis (Mugford et al. 2011; Mugford et al. 2009; Mugford et al. 2010; Kopriva et al. 2009), but the biochemical regulation of this enzyme in plants is not well understood. Structural and functional studies of AtAPSK provide new insights on the molecular basis for PAPS formation and redox-control of this important metabolic branchpoint enzyme.

The reaction sequence of APSK is proposed to follow an ordered mechanism with ATP•Mg^{2+} binding first, followed by APS, catalysis, release of PAPS, and release of ADP•Mg^{2+} (Renosto et al. 1984). Structures of the *P. chrysogenum* APSK apoenzyme (Lansdon et al. 2002), the *P. chrysogenum* APSK•ADP•Mg^{2+}•APS dead-end complex (Yu et al. 2007), the ADP•Mg^{2+}•PAPS product complex of the APSK domain from human PAPS synthetase (Sekulic et al. 2007a), and AtAPSK•AMP-PNP•Mg^{2+}•APS Michaelis complex (**Fig. A.3** and **Fig. A.4**) clearly define the active site changes accompanying ATP-dependent phosphorylation of APS to PAPS. Shifts in the side-chains of residues in the P-loop (Ser110-Thr116 in AtAPSK) position the phosphate backbone of either ADP or AMP-PNP into the active site with additional conformational changes in residues of the APS site allowing for binding of this ligand. These alterations are accompanied by ordering of the 'lid'-domain, which in AtAPSK contains a5 and a6, over the active site to constrain the flexibility of the structure.

These crystal structures suggest critical roles for the Mg^{2+} ion in catalysis and organization of the active site architecture. The interactions centered on the Mg^{2+} ion intricately link the ATP binding site, the active site water network, and residues in the catalytically essential DGDN-loop, which includes Asp136 and Asp138 of AtAPSK. Sekulic et al. (Sekulic et al. 2007a) noted that Mg^{2+} binding induces the DGDN-loop to switch from an inactive to active conformation. Contacts from Asp136 to the Mg^{2+} and active site water molecules position Asp138 in proximity to the ribose hydroxyl groups of APS. In the AtAPSK structure, the carboxylate oxygens of Asp138 are 2.67 Å and 3.15 Å from the 3'-OH and 2'-OH groups of the APS ribose, respectively (**Fig. A.4**). This suggests that the negatively-charged oxygen of Asp138 serves as a general base in the catalytic mechanism to abstract a proton from the 3'-hydroxyl group of APS and promotes its nucleophilic attack on the g-phosphate group of ATP.

Phosphorylation of APS to yield PAPS requires an orchestrated series of structural changes to organize the APSK active site for catalysis. Comparisons of various APSK in complex with different ligands (**Fig. A.3b**) suggest that the conformational flexibility of the N-terminal loop region likely affects these structural changes. Within the active site, this loop sterically contacts the DGDN-loop (**Fig. A.4**) and is proposed to position catalytically essential residues to promote ligand binding and phosphoryl transfer (Sekulic et al. 2007a; Sekulic et al. 2007b). Deletion of the N-terminal region from the APSK domain of human PAPS synthetase indicates that this region, which is not in direct contact with ligands in the active site, is responsible for substrate inhibition by APS (Sekulic et al. 2007b). The AtAPSK structure reveals a new regulatory control feature in the N-terminal loop, which is unique to plants.

Formation of a disulfide bond between Cys86 and Cys119 in the N-terminal loop region cross-links subunits of the AtAPSK dimer (**Figs. A.3d** and **Fig. A.4**), decreases catalytic

efficiency, and enhances substrate inhibition by APS (**Table A.2**). Comparisons of the kinetic properties of the reduced and oxidized forms of AtAPSK indicate that reduction of the disulfide yields a more active enzyme. Mutagenesis of the two cysteines in AtAPSK also mimics reduction of the disulfide, as demonstrated by activity assays (**Table A.3**) and gel-based analysis (**Fig. A.5a**). Moreover, redox-titrations of the disulfide bond using both gel-based and activity-based assays indicate that the midpoint potential is within a physiologically range (**Figs. A.5b** and **A.5c**) (Bick et al. 2001; Bick et al. 1998; Jez et al. 2004; Hothorn et al. 2006; Hicks et al. 2007; Gromes et al. 2008; Hutchison and Ort 1995). The redox midpoint potential of the Cys86-Cys119 disulfide in AtAPSK is comparable to that determined for AtAPSR (E_m = -255 at pH 7). Interestingly, both AtAPSK and AtAPSR are closer to the redox midpoint for GSH/GSSG than to the reported values associated with plant thioredoxins (Dangoor et al. 2009; Hirasawa et al. 1999; Collin et al. 2003). Structurally, changes in the oxidation state of Cys86 and Cys119 in AtAPSK likely alter the dynamic movements of the N-terminal loop to affect both catalytic efficiency and substrate inhibition. Physiologically, the Cys86-Cys119 disulfide bond may act as a regulatory switch to coordinate flux between the primary and secondary branches of sulfur assimilation in Arabidopsis and other plants.

Several plant thiol metabolism enzymes (i.e., APSK, APSR and glutamate-cysteine ligase (GCL); **Fig. A.1**) possess cysteines that form disulfide bonds to regulate enzymatic activity in response to changes in redox environment (Bick et al. 2001; Hothorn et al. 2006; Hicks et al. 2007; Gromes et al. 2008). This suggests that oxidative stresses may simultaneously affect multiple proteins across sulfur assimilation and metabolism in plants. Because APSR and APSK share similar redox midpoint potentials and partition the flow of APS into different branches of sulfur metabolism (Rausch and Wachter 2005; Mugford et al. 2011), coordinated biochemical

regulation of their activities may be important to meet metabolic demands for sulfur reduction and/or production of PAPS for specialized sulfonation reactions. Redox-regulation of these branchpoint proteins would provide a mechanism for controlling sulfur allocation between the two metabolic routes. For example, under oxidative stress conditions that increase demand for sulfur reduction to support glutathione synthesis for maintaining redox state, formation of the disulfides in APSR and APSK would increase and decrease their activities, respectively, and result in direction of APS into cysteine and glutathione synthesis and away from synthesis of PAPS. Moreover, previous studies demonstrate that formation of a key disulfide bond in the plant GCL under mulitple oxidative stress conditions activates this enzyme to increase glutathione synthesis (Jez et al. 2004; Hothorn et al. 2006; Hicks et al. 2007; Gromes et al. 2008). Thus, both APSR and GCL in the primary thiol metabolic pathway would be active in response to stress and APSK activity attenuated in the secondary pathway. Further experimental studies in plants that analyze metabolite distribution and sulfur pathway flux in response to biotic and abiotic stresses that alter oxidation state is required to explore this potential biochemical regulatory mechanism.

Sequence comparisons of the APSK from plants, mosses, fungi, human, green algae, and cyanobacteria show that active site residues are highly conserved (Mugford et al. 2010; Mugford et al. 2009; Kopriva et al. 2009; Yatusevich et al. 2010; Satishchandran and Markham 1989; Lillig et al. 2001; MacRae et al. 2000; Lansdon et al. 2002; Yu et al. 2007; Gay et al. 2009; Sekulic et al. 2007a; Sekulic et al. 2007b) and that the N-terminal regions are divergent (**Fig. A.2**), which may reflect the need for specialization of APSK regulation in organisms with differing metabolic demands for sulfur assimilation. Residues corresponding to Cys86 and Cys119 of AtAPSK are nearly invariant across the homologs from plants and mosses with some

unicellular alga and yeast retaining Cys119; however, in the APSK from organisms lacking plastids neither cysteine is found (**Fig. A.2b**). In addition, of the four APSK isoforms in Arabidopsis (Phartiyal et al. 2008), the disulfide cysteines are found in the three plastid-localized isoforms, but only Cys119 is conserved in the cytosolic isoform. Given that the chloroplast is a highly redox active organelle (Noctor and Foyer 1998; Meyer 2008; Galant et al. 2011), this difference between AtAPSK isoforms suggests the evolution of a specialized regulatory control in the enzymes of this organelle. Through the lineage from cyanobacteria to green algae to mosses and 'higher' plants, there is also a shift in the organization of the sulfur assimilation pathway from a linear one in cyanobacteria to a branched pathway in the later evolving mosses and plants (**Fig. A.1**) (Patron et al. 2008; Raven and Allen 2003). Thus, the redox-sensitive disulfide observed in the plant and moss APSK may be a later adaptation to the development of chloroplasts and branched sulfur metabolism pathways in these multicellular photosynthetic organisms and reflects a greater need for regulatory control in response to more specialized metabolic demands for coordination of flux between the primary and secondary sulfur pathways by reciprocally modulating APSR and APSK activity in plants.

A.5 Acknowledgements

We thank the Pakrasi lab for providing *Synechocystis* genomic DNA. This work was supported by the National Science Foundation (MCB-0904215). Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Advanced Photon Source, a national user facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research (DE-AC02-06CH11357).

A.5.1 Author Contributions

G.E.R., J.M.J. designed research; G.E.R., A.Y.N., J.A.F. performed research; G.E.R., A.Y.N., J.M.J. analyzed data; and G.E.R., J.M.J. wrote the paper.

A.5.2 Conflict of Interest Statement

The authors declare no conflict of interest.

A.6 References

Adams, P., Afonine, P., Bunkoczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L., Kapral, G., Grosse-Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T., Zwart, P. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallographica Section D Biological Crystallography*. **66**(2): 213-221.

Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M., Matsubayashi, Y. (2007). "Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in Arabidopsis." *Proceedings of the National Academy of Sciences*. **104**(46): 18333-18338.

Bick, J.A., Åslund, F., Chen, Y., Leustek, T. (1998). "Glutaredoxin function for the carboxyl-terminal domain of the plant-type 5′-adenylylsulfate reductase*." Proceedings of the National Academy of Sciences*. **95**(14): 8404-8409.

Bick, J.A., Dennis, J., Zylstra, G., Nowack, J., Leustek, T. (2000). "Identification of a new class of 5'-adenylylsulfate (APS) reductases from sulfate-assimilating bacteria." *Journal of Bacteriology*. **182**(1): 135-142.

Bick, J.A., Setterdahl, A., Knaff, D., Chen, Y., Pitcher, L., Zilinskas, B., Leustek, T. (2001). "Regulation of the plant-type 5'-adenylyl sulfate reductase by oxidative stress." *Biochemistry*. **40**(30): 9040-9048.

Cobbett, C., Goldsbrough, P. (2002). "Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis." *Annual Review of Plant Biology*. **53**:159-182

Collin, V., Issakidis-Bourguet, E., Marchand, C., Hirasawa, M., Lancelin, J., Knaff, D., Miginiac-Maslow, M. (2003). "The Arabidopsis plastidial thioredoxins: new functions and new insights into specificity." *Journal of Biological Chemistry*. **278**(26): 23747-23752.

Dangoor, I., Peled-Zehavi, H., Levitan, A., Pasand, O., Danon, A. (2009). "A small family of chloroplast atypical thioredoxins." *Plant Physiology*. **149**(3): 1240-1250.

Emsley, P., Lohkamp, B., Scott, W., Cowtan, K. (2010). "Features and development of Coot." *Acta Crystallographica Section D Biological Crystallography*. **66**(4): 486-501.

Galant, A., Preuss, M., Cameron, J., Jez, J. (2011). "Plant glutathione biosynthesis: diversity in biochemical regulation and reaction products." *Frontiers in Plant Science*. **2**(45): 1-7.

Gay, S., Segel, I., Fisher, A. (2009). "Structure of the two-domain hexameric APS kinase from *Thiobacillus denitrificans*: structural basis for the absence of ATP sulfurylase activity." *Acta Crystallographica Section D Biological Crystallography*. **65**(10): 1021-1031.

Gromes, R., Hothorn, M., Lenherr, E., Rybin, V., Scheffzek, K., Rausch, T. (2008). "The redox switch of gamma-glutamylcysteine ligase via a reversible monomer-dimer transition is a mechanism unique to plants." *Plant Journal*. **54**(6): 1063-1075.

Halkier, B.A., Gershenzon, J. (2006). "Biology and biochemistry of glucosinolates." *Annual Review of Plant Biology*. **57**: 303-333.

Hicks, L.M., Cahoon, R.E., Bonner, E.R., Rivard, R.S., Sheffield, J., Jez, J.M. (2007). "Thiol-based regulation of redox-active glutamate-cysteine ligase from *Arabidopsis thaliana*." *Plant Cell*. **19**(8): 2653-2661.

Hirasawa, M., Schurmann, P., Jacquot, J., Manieri, W., Jacquot, P., Keryer, E., Hartman, F., Knaff, D. (1999). "Oxidation-reduction properties of chloroplast thioredoxins, ferredoxin:thioredoxin reductase, and thioredoxin f-regulated enzymes." *Biochemistry*. **38**(16): 5200-5205.

Hothorn, M., Wachter, A., Gromes, R., Stuwe, T., Rausch, T., Scheffzek, K. (2006). "Structural basis for the redox control of plant glutamate cysteine ligase." *Journal of Biological Chemistry*. **281**(37): 27557-27565.

Hutchison, R., Ort, D. (1995). "Measurement of equilibrium midpoint potentials of thiol/disulfide regulatory groups on thioredoxin-activated chloroplast enzymes." *Methods in Enzymology*. **252**: 220-228.

Jez, J., Cahoon, R., Chen, S. (2004). "Arabidopsis thaliana glutamate-cysteine ligase: functional properties, kinetic mechanism, and regulation of activity." *Journal of Biological Chemistry*. **279**(32): 33463-33470.

Klein, M., Papenbrock, J. (2004). "The multi-protein family of Arabidopsis sulphotransferases and their relatives in other plant species." *Journal of Experimental Botany*. **55**(404): 1809-1820.

Kopriva, S., Mugford, S., Matthewman, C., Koprivova, A. (2009). "Plant sulfate assimilation genes: redundancy versus specialization." *Plant Cell Reports*. **28**(12): 1769-1780.

Lansdon, E., Segel, I., Fisher, A. (2002). "Ligand-induced structural changes in adenosine 5'-phosphosulfate kinase from *Penicillium chrysogenum*." *Biochemistry*. **41**(46): 13672-13680.

Leustek, T., Martin, M.N., Bick, J.A., Davies, J.P. (2000). "Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies." *Annual Review Plant Physiology and Plant Molecular Biology*. **51**: 141-165.

Lillig, C.H., Schiffmann, S., Berndt, C., Berken, A., Tischka, R., Schwenn, J.D. (2001). "Molecular and catalytic properties of *Arabidopsis thaliana* adenylyl sulfate (APS)-kinase." *Archives of Biochemistry and Biophysics*. **392**(2): 303-310.

Loudet, O., Saliba-Colombani, V., Camilleri, C., Calenge, F., Gaudon, V., Koprivova, A., North, K., Kopriva, S., Daniel-Vedele, F. (2007). "Natural variation for sulfate content in *Arabidopsis thaliana* is highly controlled by APR2." *Nature Genetics*. **39**(7): 896-900.

MacRae, I.J., Segel, I. (1999). "Adenosine 5'-phosphosulfate (APS) kinase: diagnosing the mechanism of substrate inhibition." *Archives of Biochemistry and Biophysics*. **361**(2): 277-282.

MacRae, I.J., Segel, I.H., Fisher, A.J. (2000). "Crystal structure of adenosine 5'-phosphosulfate kinase from *Penicillium chrysogenum*." *Biochemistry*. **39**(7): 1613-1621.

Martin, M., Tarczynski, M., Shen, B., Leustek, T. (2005). "The role of 5'-adenylylsulfate reductase in controlling sulfate reduction in plants." *Photosynthesis Research*. **86**(3): 309-323.

Marzluf, G.A. (1997). "Molecular genetics of sulfur assimilation in filamentous fungi and yeast." *Annual Review of Microbiology*. **51**: 73-96.

McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., Read, R. (2007). "Phaser crystallographic software." *Journal of Applied Crystallography*. **40**(4): 658-674.

Meyer, A.J. (2008). "The integration of glutathione homeostasis and redox signaling." *Journal of Plant Phyiology*. **165**(13): 1390-1403.

Mugford, S.G., Lee, B., Koprivova, A., Matthewman, C., Kopriva, S. (2011). "Control of sulfur partitioning between primary and secondary metabolism." *Plant Journal*. **65**(1): 96-105.

Mugford, S.G., Matthewman, C.A., Hill, L., Kopriva, S. (2010). "Adenosine-5′-phosphosulfate kinase is essential for Arabidopsis viability." *FEBS Letters*. **584**(1): 119-123.

Mugford, S.G., Yoshimoto, N., Reichelt, M., Wirtz, M., Hill, L., Mugford, S.T., Nakazato, Y., Noji, M., Takahashi, H., Kramell, R., Gigolashvili, T., Flügge, U.I., Wasternack, C.,

Gershenzon, J., Hell, R., Saito, K., Kopriva, S. (2009). “Disruption of adenosine-5′-phosphosulfate kinase in Arabidopsis reduces levels of sulfated secondary metabolites.” *Plant Cell*. **21**(3): 910-927.

Noctor, G., Foyer, C.H. (1998). “ASCORBATE AND GLUTATHIONE: Keeping active oxygen under control.” *Annual Review of Plant Physiology and Plant Molecular Biology*. **49**: 249-279.

Otwinowski, Z., Minor, W. (1997). “Processing of X-ray diffraction data collected in oscillation mode.” *Methods in Enzymology*. **276**: 307-326.

Patron, N.J., Durnford, D., Kopriva, S. (2008). “Sulfate assimilation in eukaryotes: fusions, relocations and lateral transfers.” *BMC Evolutionary Biology*. **8**(39): 1-14.

Phartiyal, P., Kim, W., Cahoon, R., Jez, J., Krishnan, H. (2006). “Soybean ATP sulfurylase, a homodimeric enzyme involved in sulfur assimilation, is abundantly expressed in roots and induced by cold treatment.” *Archives of Biochemistry and Biophysics*. **450**(1): 20-29.

Phartiyal, P., Kim, W., Cahoon, R., Jez, J., Krishnan, H. (2008). “The role of 5'-adenylylsulfate reductase in the sulfur assimilation pathway of soybean: molecular cloning, kinetic characterization, and gene expression.” *Phytochemistry*. **69**(2): 356-364.

Rausch, T., Wachter, A. (2005). “Sulfur metabolism: a versatile platform for launching defence operations.” *Trends in Plant Sciences*. **10**(10): 503-509.

Raven, J.A., Allen, J.F. (2003). “Genomics and chloroplast evolution: what did cyanobacteria do for plants?” *Genome Biology*. **4**(3): 209-209.

Renosto, F., Seubert, P., Segel, I. (1984). “Adenosine 5'-phosphosulfate kinase from *Penicillium chrysogenum*. Purification and kinetic characterization.” *Journal of Biological Chemistry*. **259**(4): 2113-2123.

Satishchandran, C., Markham, G.D. (1989). “Adenosine-5'-phosphosulfate kinase from *Escherichia coli* K12. Purification, characterization, and identification of a phosphorylated enzyme intermediate.” *Journal of Biological Chemistry*. **264**(25): 15012-15021.

Scheerer, U., Haensch, R., Mendel, R.R., Kopriva, S., Rennenberg, H., Herschbach, C. (2010). “Sulphur flux through the sulphate assimilation pathway is differently controlled by adenosine 5′-phosphosulphate reductase under stress and in transgenic poplar plants overexpressing γ-ECS, SO, or APR.” *Journal of Experimental Botany*. **61**(2): 609-622.

Sekowska, A., Kung, H., Danchin, A. (2000). “Sulfur metabolism in *Escherichia coli* and related bacteria: facts and fiction.” *Journal of Molecular Microbiology Biotechnology*. **2**(2): 145-177.

Sekulic, N., Dietrich, K., Paarmann, I., Ort, S., Konrad, M., Lavie, A. (2007a). Elucidation of the active conformation of the APS kinase domain of human PAPS synthetase 1. *Journal of Molecular Biology*. **367**(2): 488-500.

Sekulic, N., Konrad, M., Lavie, A. (2007b). “Structural mechanism for substrate inhibition of the adenosine 5'-phosphosulfate kinase domain of human 3'-phosphoadenosine 5'-phosphosulfate synthetase 1 and its ramifications for enzyme regulation.” *Journal of Biological Chemistry*. **282**(30): 22112-22121.

Singh, B., Schwartz, N. (2003). “Identification and functional characterization of the novel BM-motif in the murine phosphoadenosine phosphosulfate (PAPS) synthetase.” *Journal of Biological Chemistry*. **278**(1): 71-75.

Takahashi, H., Kopriva, S., Giordano, M., Saito, K., Hell, R. (2011). "Sulfur assimilation in photosynthetic organisms: molecular functions and regulations of transporters and assimilatory enzymes." *Annual Review of Plant Biology*. **62**: 57-184.

Vauclare, P., Kopriva, S., Fell, D., Suter, M., Sticher, L., von Ballmoos, P., Krahenbuhl, U., den Camp, R., Brunold, C. (2002). "Flux control of sulphate assimilation in *Arabidopsis thaliana*: adenosine 5'-phosphosulphate reductase is more susceptible than ATP sulphurylase to negative control by thiols." *Plant Journal*. **31**(6): 729-740.

Yatusevich, R., Mugford, S., Matthewman, C., Gigolashvili, T., Frerigmann, H., Delaney, S., Koprivova, A., Flugge, U., Kopriva, S. (2010). "Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*." *Plant Journal*. **62**(1): 1-11.

Yi, H., Galant, A., Ravilious, G.E., Preuss, M.L., Jez, J.M. (2010a). Sensing sulfur conditions: simple to complex protein regulatory mechanisms in plant thiol metabolism." *Molecular Plant*. **3**(2): 269-279.

Yi, H., Ravilious, G.E., Galant, A., Krishnan, H.B., Jez, J.M. (2010b). "From sulfur to homoglutathione: thiol metabolism in soybean." *Amino Acids*. **39**(4): 963-978.

Yu, Z., Lansdon, E., Segel, I., Fisher, A. (2007). "Crystal structure of the bifunctional ATP sulfurylase-APS kinase from the chemolithotrophic thermophile *Aquifex aeolicus*." *Journal of Molecular Biology*. **365**(3): 732-743.

C86A-F	5’-dCTCGACAAATATAAAGTGGCATGA*AGC*TTCTGTTG-3’
C86A-R	5’-dCTGTCTATCAACTTTCTCAACAGAA*GCT*TCATGCC-3’
C119A-F	5’-dGGGAAGAGTACTTTGGCT*GCT*GCTTTG -3’
C119A-R	5’-dCAACATCTGATTCAAAGC*AGC*AGCCAAAG -3’
SynAPSK-F	5'-dTTTATGGCTAGC**ATG**CAACAACGTGGCGTAAC-3’
SynAPSK-R	5'-dATAGAATTCG**TTA**GCCCTCGATATATTTTAGATCTACTAGCTTCTG-3’

Table A.1: Oligonucleotides used for generation of strains.

The AtAPSK C86A/C119A and overexpression of APSK from *Synechocystis* sp. PCC 6803 strains were made. Mutated codons are italized, restriction sites are underlined, and start/stop codons are shown in bold.

Crystal	AtAPSK•AMP-PNP•Mg^{2+}•APS
Space group	C2
Cell dimensions	*a* = 121.1 Å, *b* = 95.31 Å, *c* = 73.33 Å; b = 114.1°
Data Collection	
Wavelength (Å)	0.979
Resolution range (Å) (highest shell resolution)	30.5 - 1.80 (1.83 - 1.80)
Reflections (total/unique)	255,905 / 69,476
Completeness (highest shell)	98.0% (98.4%)
<I/s> (highest shell)	34.1 (1.7)
R_{sym}[a] (highest shell)	3.2% (44.1%)
Model and Refinement	
R_{cryst}[b] / R_{free}[c]	17.3 / 20.1
No. of protein atoms	4,525
No. of water molecules	479
No. of ligand atoms	177
R.m.s. deviation, bond lengths (Å)	0.006
R.m.s. deviation, bond angles (°)	1.08
Avg. B-factor ($Å^2$) - protein, waters, ligands	39.5, 50.0, 46.0
Stereochemistry: most favored, allowed, generously allowed	97.6, 2.1, 0.2%

Table A.2: Summary of crystallographic statistics.

aRsym = S|Ih - <Ih>|/SIh, where <Ih> is the average intensity over symmetry. bRcryst = S|Fo - <Fc>|/SFo, where summation is over the data used for refinement. cRfree is defined the same as Rcryst, but was calculated using 5% of data excluded from refinement.

	k_{cat} (s^{-1})	K_m (μM)	K_i (μM)	k_{cat} / K_m (M^{-1} s^{-1} x 10^8)
$AtAPSK_{red}$	272 ± 39	0.48 ± 0.41	37.5 ± 6.9	5.67
$AtAPSK_{ox}$	14.1 ± 2.3	0.43 ± 0.26	2.51 ± 0.84	0.328
$C86A/C119A_{red}$	239 ± 44	1.90 ± 0.53	34.8 ± 2.7	1.25
$C86A/C119A_{ox}$	203 ± 31	0.87 ± 0.49	41.6 ± 7.1	2.33
$SynAPSK_{red}$	18.6 ± 4.5	0.33 ± 0.17	27.6 ± 3.5	0.56
$SynAPSK_{ox}$	15.9 ± 2.2	0.57 ± 0.26	21.6 ± 3.7	0.28

Average values ± S.E. (n = 3) are shown.

Table A.3: Steady-state kinetic parameters.

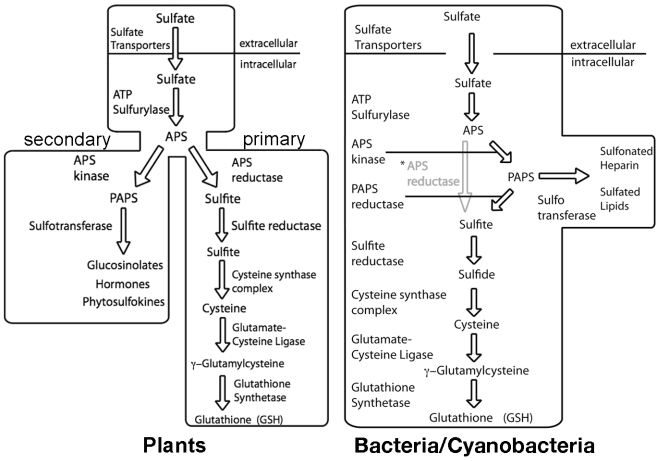


Figure A.1: Sulfur assimilatory pathways.

In plants (left) and bacteria/cyanobacteria (right). Various bacteria and cyanobacteria use either APS kinase and PAPS reductase (black arrows) or APS reductase (grey arrow and text marked with an asterisk) for formation of sulfite. Abbreviations are: APS, adenosine 5'-phosphosulfate; PAPS, 3'-phosphoadenosine-5'-phosphosulfate.

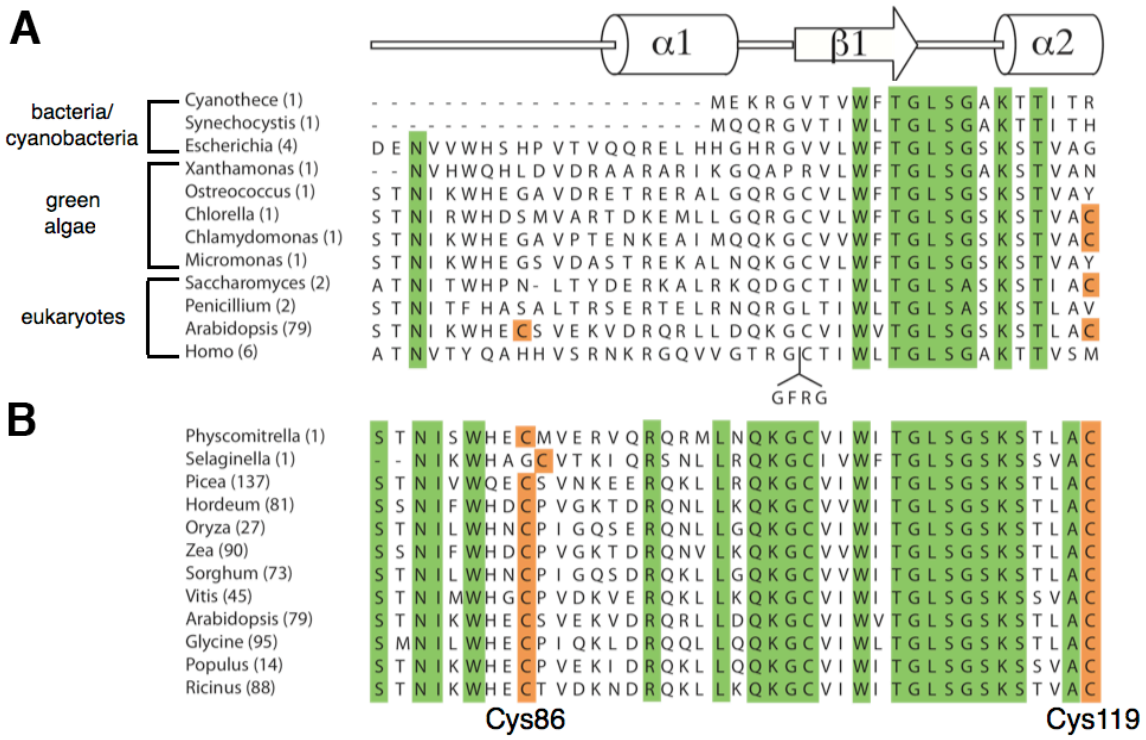


Figure A.2: Sequence alignment of representative APSK.

From bacteria, cyanobacteria, green algae, and eukaryotes **a** and plant species **b**. The secondary structure of the AtAPSK is shown above the alignment. Invariant residues are highlighted in green. Conserved cysteines corresponding to those in AtAPSK that form the disulfide bond are highlighted in orange.

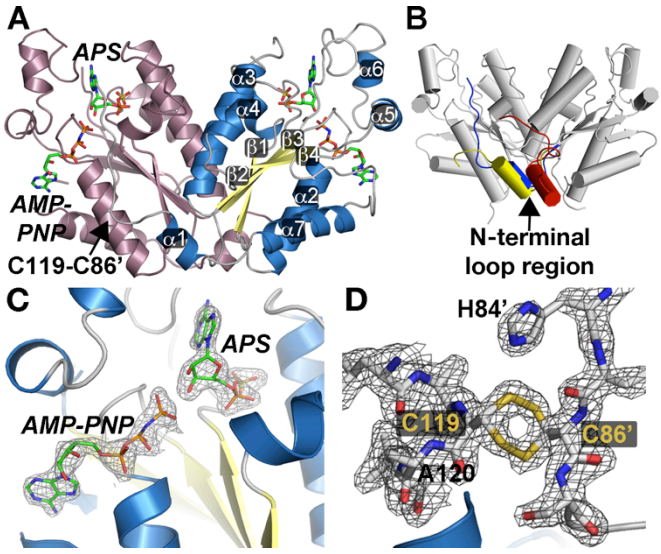


Figure A.3: Structure of AtAPSK.

a Ribbon diagram of the dimer. Each monomer is colored in rose and blue, respectively. The positions of AMP-PNP and APS are shown as stick molecules. The position of the C86-C119 disulfide bond is indicated in the left monomer. Secondary structure features are labeled in the right monomer. **b** Structural comparison of APSK structures. The overall AtAPSK structure is shown in grey. The position of the N-terminal region of the AtAPSK•AMP-PNP••Mg^{2+}•APS complex is in blue. The N-terminal of the APSK domain from the human PAPS synthetase•ADP•PAPS complex is nearly identical to that of AtAPSK. The N-terminal of *P. chrysogenum* APSK apoenzyme is in yellow and the position in the *P. chrysogenum* APSK•ADP•APS complex is in red. **c** The $2F_o$-F_c omit map (1.5 s) for AMP-PNP and APS in the AtAPSK active site. **d** The $2F_o$-F_c omit map (1.5 s) for Cys86-Cys119 disulfide bond. Note that the cysteine residues are found in alternate conformations.

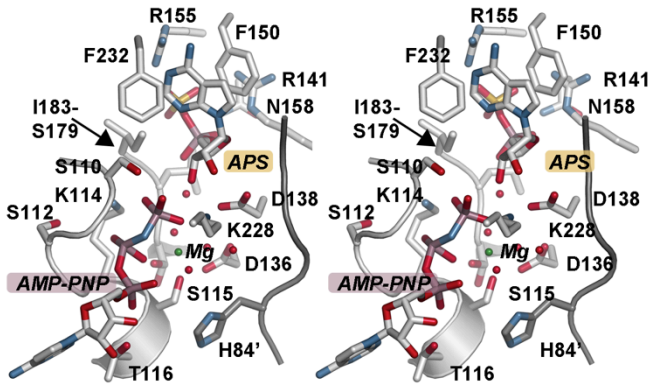


Figure A.4: Stereoview of the AtAPSK active site.

The N-terminal loop of the adjacent monomer is shown as the dark gray loop, which includes His84.

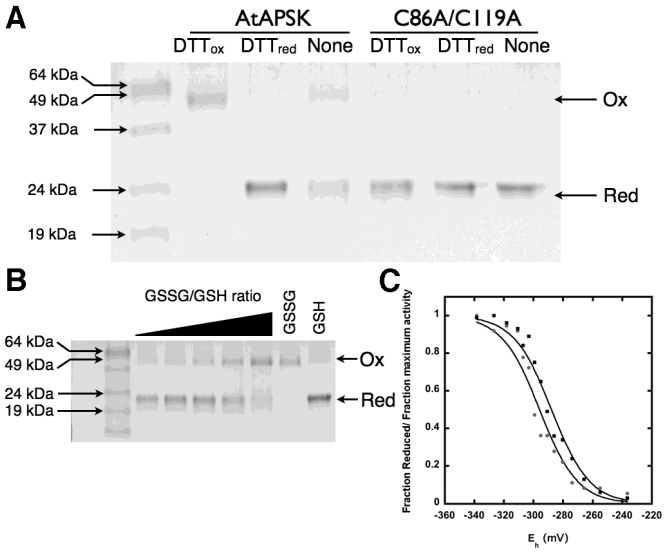


Figure A.5: Functional analysis of redox activity.

a SDS-PAGE of wild-type and C86A/C119A AtAPSK incubated in the absence or presence of either 5 mM reduced DTT (DTT_{red}) or trans-4,5-dihydroxy-1,2-dithiane (DTT_{ox}). Protein (0.3-1 μg) was stained with Coomassie blue. Arrows on the left side of the gel indicate positions of the molecular weight markers. Arrows on the right indicate positions of the oxidized dimeric and reduced monomeric forms. **b** Gel-based redox-titration of AtAPSK. Protein was incubated in increasing ratios of oxidized:reduced glutathione (GSSG:GSH; 20 mM total; E_h = -240 to -340 mV). **c** Comparison of gel-based (squares) and activity-based (circles) redox titrations. Solid lines represent fits to the Nernst equation.

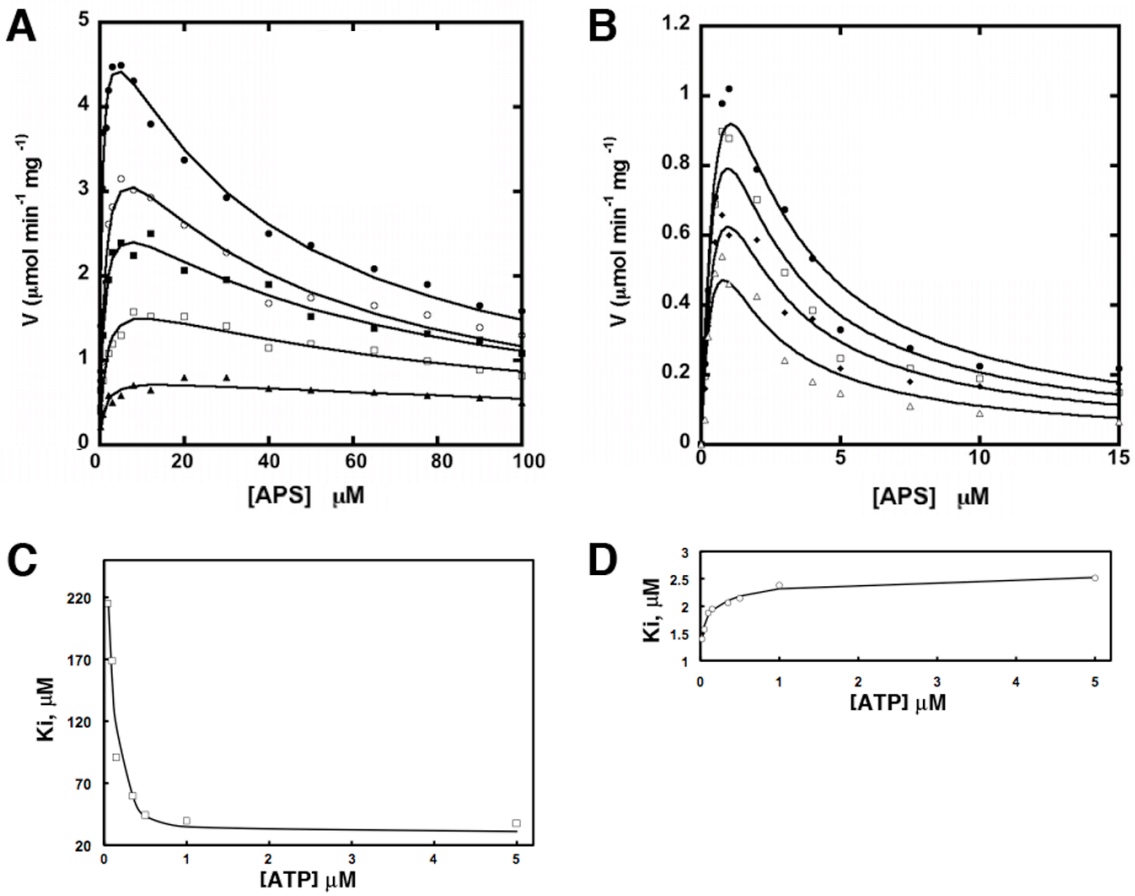


Figure A.6: Velocity curves of *v* versus APS concentration.

For **a** $AtAPSK_{red}$, **b** $AtAPSK_{ox}$, and **c-d** the effect of [ATP] on K_i of $AtAPSK_{red}$ and $AtAPSK_{ox}$. **a** Initial velocities were measured at 5 mM (●), 1 mM (○), 0.35 mM (■), 0.15 mM (□) and 0.05 mM (▲) ATP. The curves represent best fits to the general substrate inhibition model described in Materials and Methods. **b** Initial velocities were measured at 5 mM (●), 0.5 mM (□), 0.1 mM (◆), 0.025 mM (△) ATP. **c-d** Effect of ATP concentration on the inhibition constants (K_i) of $AtAPSK_{red}$ (**c**) and $AtAPSK_{ox}$ (**d**).

Appendix B

Global Proteomic Analysis Reveals an Exclusive Role of Thylakoid Membranes in Bioenergetics of a Model Cyanobacterium

This work was originally published in *Molecular and Cellular Proteomics*. Michelle Liberton[#], Rajib Saha[#], Jon M. Jacobs[#], Amelia Y. Nguyen*, Marina A. Gritsenko, Richard D. Smith, David W. Koppenaal, and Himadri B. Pakrasi. Global Proteomic Analysis Reveals an Exclusive Role of Thylakoid Membranes in Bioenergetics of a Model Cyanobacterium. *Molecular and Cellular Proteomics*. 2016. 15(6):2012-32.

[#]These authors contributed equally to this work.

*A.Y.N isolated the membrane samples, performed the Western Blots, analyzed the data, and wrote the paper.

B.1 Introduction

B.1.1 Abstract

Cyanobacteria are photosynthetic microbes with highly differentiated membrane systems. These organisms contain an outer membrane, plasma membrane, and an internal system of thylakoid membranes where the photosynthetic and respiratory machinery are found. This existence of compartmentalization and differentiation of membrane systems poses a number of challenges for cyanobacterial cells in terms of organization and distribution of proteins to the correct membrane system. Proteomics studies have long sought to identify the components of the different membrane systems in cyanobacteria, and to date about 450 different proteins have been attributed to either the plasma membrane or thylakoid membrane. Given the complexity of these membranes, many more proteins remain to be identified, and a comprehensive catalog of plasma membrane and thylakoid membrane proteins is needed. Here we describe the identification of 635 differentially localized proteins in *Synechocystis* sp. PCC 6803 by quantitative iTRAQ isobaric labeling; of these, 459 proteins were localized to the plasma membrane and 176 were localized to the thylakoid membrane. Surprisingly, we found over 2.5 times the number of unique proteins identified in the plasma membrane compared to the thylakoid membrane. This suggests that the protein composition of the thylakoid membrane is more homogeneous than the plasma membrane, consistent with the role of the plasma membrane in diverse cellular processes including protein trafficking and nutrient import, compared to a more specialized role for the thylakoid membrane in cellular energetics. Thus, our data clearly define the two membrane systems with distinct functions. Overall, the protein compositions of the *Synechocystis* 6803 plasma membrane and thylakoid membrane are quite similar to that of the plasma membrane of *Escherichia coli* and thylakoid membrane of Arabidopsis chloroplasts, respectively.

Synechocystis 6803 can therefore be described as a gram-negative bacterium with an additional internal membrane system that fulfills the energetic requirements of the cell.

B.1.2 Plasma and Thylakoid Membranes

Photosynthetic microbes such as the cyanobacterium *Synechocystis* sp. PCC 6803 convert light to cellular energy, an ability that makes these organisms of particular interest in renewable energy studies. Cyanobacteria typically have a Gram-negative-type cell envelope consisting of a plasma membrane (PM), peptidoglycan layer, and outer membrane. These microbes also have an internal thylakoid membrane (TM) system where the protein complexes of the photosynthetic and respiratory electron transfer chains function. The presence of these differentiated membrane systems makes cyanobacteria more complex than other gram-negative bacteria. There is considerable interest in understanding the roles of the membrane systems and their relation with each other. Our studies using electron tomography revealed that the TM in the cyanobacterium *Cyanothece* sp. ATCC 51142 forms a complicated network of membranes that enclose a single lumenal space (Liberton et al. 2011). Several studies have probed the question of whether the PM and TM are contiguous, or if these two systems are physically independent (Liberton et al. 2006; van de Meene et al. 2006; Schneider et al. 2007). Recent reports have proposed the existence of sites of "hemifusion" between PM and TM, which can be analyzed as a subfraction of the PM and used to further clarify the targeting pathways between the membrane systems (Pisareva et al. 2011). Similarly, the existence of a membrane subfraction that associates with both PM and TM has been proposed (Nickelsen et al. 2011; Rengstl et al. 2011). Thus, identifying the protein composition of the different membrane systems is of considerable interest in understanding the form and function of cyanobacterial membranes.

Several previous studies have begun to catalog the protein complement of the cyanobacterial membrane systems. One study of the PM proteome used two-dimensional gel electrophoresis coupled with mass spectrometry (MS) to identify 57 proteins, of which 17 are integral membrane proteins and 40 are peripheral proteins (Huang et al. 2002). Another study identified 51 integral PM proteins by peptide mass fingerprinting (Pisareva et al. 2007). Isolated TM samples were used to identify 76 proteins from 1- and 2-D gels by MALDI-TOF MS (Srivastava et al. 2005). A study of both isolated PM and TM samples probed by nano-LC separation and MS/MS identified 379 different proteins (Pisareva et al. 2011), of which 237 were uniquely localized to either PM or TM. However, all together to date only about 450 different proteins have been identified as localized to the PM or TM.

In order to comprehensively detect and identify proteins localized to the PM and TM, we applied a sensitive LC-MS/MS based analysis pipeline for the identification and quantification of this protein complement. This resulted in the identification of 635 proteins observed with significantly different localizations across PM and TM from purified membrane samples isolated from *Synechocystis* sp. PCC 6803 (hereafter, *Synechocystis* 6803). This is a large increase in the number of differentially abundant proteins compared to previous studies and offers considerable insight into the composition of PM and TM. Our study found a larger number of proteins uniquely localized in PM (459) compared to TM (176). The overall protein composition of PM was characterized by proteins involved in transport, secretion, and trafficking, whereas the TM protein composition described a specialized membrane system dedicated to the energetics of electron transport, highlighting the very different roles these membrane systems have in cyanobacterial cellular metabolism. Comparison of the *Synechocystis* 6803 membrane systems with the *E. coli* PM and Arabidopsis TM showed how an oxygenic phototrophic bacterium

modified the Gram-negative PM for specific purposes while creating a specialized internal membrane compartment for photosynthetic electron transfer.

B.2 Materials and Methods

B.2.1 Cell Growth and Sample Preparation

Synechocystis 6803 cells were grown in BG11 at 30°C under 30 μmole photons·m^{-2}·s^{-1} white light. Membrane isolation and two-phase partitioning were performed as described (Keren et al. 2005). Two-phase systems were prepared from stock solutions of 20% (w/w) Dextran T-500 and 40% (w/w) polyethylene glycol 3350.

Protein samples were incubated in 8 M urea, 100 mM ammonium bicarbonate, pH 8.0 solution containing 5 mM dithiothreitol at 56°C for 45 min with constant shaking at 800 rpm in Thermomixer R (Eppendorf, NY). Alkylation was performed with 20 mM iodoacetamide at 37°C in the dark with constant shaking (800 rpm in Thermomixer), followed by an 8 fold dilution with 25 mM ammonium bicarbonate, pH 8.0 containing 1 mM $CaCl_2$. Tryptic digestion with Sequencing Grade Modified Trypsin (Promega, WI) was performed at 1:50 enzyme-to-substrate ratio for 4 h at 37°C. The digested samples were then acidified with 10% trifluoroacetic acid to ~pH 3 and 5% acetonitrile was added to the digested samples prior to desalting. SPE C-18 columns (SUPELCO Discovery) were used for cleanup of the resultant peptide mixture, and samples were concentrated in a SpeedVac SC250 Express (ThermoSavant) followed by BCA assay to determine final peptide concentration.

B.2.2 iTRAQ Labeling and HPLC Fractionation

Isobaric labeling of peptides using 2 separate four-plex iTRAQ™ reagents was performed according to the manufacturer's instructions (AB Sciex, Foster City, CA) and as previously described (Welkie et al. 2014). Labeled peptide samples were separated at a flow rate

of 0.5 ml/min on a reverse phase Waters XBridge C18 column (250 mm × 4.6 mm column containing 5 μm particles, and a 4.6 mm × 20 mm guard column) using an Agilent 1200 HPLC System equipped with a quaternary pump, degasser, diode array detector, Peltier-cooled auto-sampler and fraction collector (both set at 4ºC). Approximately 120 μg of labeled tryptic peptides was suspended in buffer A (10 mM triethylammonium bicarbonate, pH 7.5) and loaded onto the column After the sample loading, the C18 column was washed for 35 min with solvent A, before applying the LC gradient. The LC gradient started with a linear increase of solvent A to 10% B (10 mM triethylammonium bicarbonate, pH 7.5, 90% acetonitrile) for 10 min, then linearly increased at 15 min to 20% B, 30 min to 30% B, 15 min to 35% B, 10 min to 45% B and another 10 min to 100% solvent B. Using an automated fraction collector, 96 fractions were collected for each sample, lyophilized and reconstituted into 12 fractions prior to LC-MS/MS analysis.

B.2.3 LC-MS/MS Analysis

All iTRAQ™-labeled fractions were analyzed by LC–MS/MS. Each sample was loaded onto a homemade 65 cm × 75 mm i.d. reversed-phase capillary column using 3 mm C18 particles (Phenomenex, Torrance, CA, USA). The HPLC system consisted of a custom conuration of 100 ml Isco Model 100DM syringe pumps (Isco, Lincoln, NE, USA), two-position Valco valves (Valco Instruments Co., Houston, TX, USA), and a PAL autosampler (Leap Technologies, Carrboro, NC, USA) that allowed fully automated sample analysis across four HPLC columns (Livesay et al. 2008). The system was operated at a constant pressure of 10,000 psi over 3 h with an exponential gradient starting with 100% of mobile phase A (0.1% (v/v) formic acid in water) to 60% (v/v) of mobile phase B (0.1% (v/v) formic acid in acetonitrile). MS analysis was performed on a Thermo Scientific LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific,

San Jose, CA, USA) coupled with an electrospray ionization interface using homemade 150-mm o.d. × 20-mm i.d. chemically etched electrospray emitters (Kelly et al. 2006). Full MS spectra were recorded at resolution of 100 K (m/z 400) over the range of m/z 400–2000 with an automated gain control (AGC) value of 1×10^6. MS/MS was performed in the data-dependent mode with an AGC target value of 3×10^4. The most abundant 10 parent ions were selected for MS/MS using high-energy collision dissociation with a normalized collision energy setting of 45. Precursor ion activation was performed with an isolation width of 2 Da, a minimal intensity of 500 counts, and an activation time of 10 ms.

B.2.4 Data Analysis

LC–MS/MS raw data were converted into dta files using Bioworks Cluster 3.2 (Thermo Fisher Scientific, Cambridge, MA, USA), and MSGF+ algorithm (Kim et al. 2008) was used to search MS/MS spectra against *Synechocystis* 6803 (NCBI 2011-02-28, 3672 entries). The key search parameters used were 20 ppm tolerance for precursor ion masses, +0.5 Da and -0.5 Da window on fragment ion mass tolerances (Kim et al. 2010), no limit on missed cleavages, partial tryptic search, no exclusion of contaminants, dynamic oxidation of methionine (15.9949 Da), static IAA alkylation on cysteine (57.0215 Da), and static iTRAQ modification of lysine and N-termini (+144.1021 Da). No additional mass shifts were performed on the data. The decoy database searching methodology (Elias and Gygi 2007; Qian et al. 2005) was used to control the false discovery rate at the unique peptide level to ~0.1% (Kim et al. 2008). Only proteins containing multiple peptide identifications with reporter ion intensities were quantified. The ppm distribution of identified spectra is shown is **Figure B.1** and protein and peptide data can be found in **Table B.1**. The mass spectrometry proteomics data have been deposited to the

ProteomeXchange Consortium (Vizcaino et al. 2014) via the PRIDE partner repository with the dataset identifier PXD003079 and 10.6019/PXD003079.

B.2.5 Experimental Design and Statistical Rationale

The experimental design incorporated three biological replicates of PM and TM samples, across two independent four-plex iTRAQ experiments to cover all six samples, where final quantitative comparisons included the averaged technical replicate values for TM-2 and TM-3. For quantification purposes, peptide reporter ion intensities were captured across all channels and compared by calculating the summed peptide intensity values for TM and PM samples. Summed protein values were then scaled within each experiment and then central tendency normalized and statistically compared across biological replicates using ANOVA with membrane type as a fixed effect and using the program DAnTE (Polpitiya et al. 2008) for final comparisons.

B.2.6 Computational Web-based Tools

We used SignalP 4.1 server (Petersen et al. 2011) (http://www.cbs.dtu.dk/services/SignalP/) for identifying putative signal peptides as well as their cleavage sites. In addition, we utilized THHMM server 2.0 (Krogh et al. 2001) (http://www.cbs.dtu.dk/services/TMHMM/) and LipoP 1.0 server (Juncker et al. 2003) (http://www.cbs.dtu.dk/services/LipoP/) to identify TMHs and lipoproteins, respectively, in identified proteins.

B.3 Results

B.3.1 Identification of PM and TM Proteins

In order to analyze isolated plasma membrane (PM) and thylakoid membrane (TM) samples, we previously devised a 2D protocol for membrane isolation with a short preparation

time (Keren et al. 2005). This procedure used multiple rounds of a polymer two-phase isolation step, which is a fast procedure that maintained the photochemical activity of the isolated membranes. As markers for the purity of the isolated membranes, antibodies raised against the NrtA and CP47 proteins were used that were found exclusively in PM and TM, respectively, and immunoblotting results indicated that membrane fractions isolated using this procedure are highly purified (Keren et al. 2005).

PM and TM samples prepared using this procedure were analyzed by a sensitive LC-MS/MS-based pipeline for the identification and quantification of proteins. Quantitative iTRAQ 4-plex isobaric labeling coupled with pre-MS high pH reversed phase peptide separations were used to quantitatively capture a triplicate comparison of PM and TM samples (see Experimental Procedures for details). This resulted in the identification of 1496 proteins with appropriate quantitative values for statistical comparison (**Table B.2**). Of these 1496 proteins, 635 proteins were observed with significantly different localizations across PM and TM (p-value <0.05) (**Fig. B.2**, **Table B.2**, **Fig. B.3**), and 861 proteins were found to be present in both PM and TM (**Table B.2**).

Of the 861 proteins found in both PM and TM, 257 of these are predicted to be hypothetical and 88 are unknown proteins (**Table B.2**). Of the remaining 516 proteins, many of these are known soluble proteins of high cellular abundance, such as phycobilisome light-harvesting antenna subunits, and are present in both isolated PM and TM membrane samples. Interestingly, a number of PSI and PSII proteins (e.g., the PSII D1 protein and PSI proteins PsaB and PsaK) are present in both PM and TM. Another major group (i.e., 91) of such proteins belong to translational processes such as aminoacyl tRNA synthetase and tRNA modification, protein modification and degradation, and ribosomal protein synthesis and modification. The

remaining proteins are distributed across several pathways/subsystems including biosynthesis of amino acids and cofactors, energy metabolism, regulatory and transport processes.

This identification of 635 proteins with significantly different localizations across PM and TM provides insight into the unique properties and characteristics of these two membrane systems. In addition, this study represents a more comprehensive investigation, in terms of both depth of coverage and quantitation, compared to the most recent study of isolated PM and TM samples (Pisareva et al. 2011). This previous study, which also employed LC-MS/MS analysis but utilized spectral count approaches, identified a total of 379 proteins, 237 of which were designated as differentially localized to PM or TM (Pisareva et al. 2011). Note that this comparison considers only the number of proteins identified in either PM or TM, but not in both, in these studies. As iTRAQ labeling and quantitative comparison were used in the current study, we focused on the differential abundance of proteins to determine membrane localization, recognizing that low abundance signals can be generated by iTRAQ in both TM and PM preparations, and it is likely that the sensitivity of the analyses coupled with known co-fragmentation and isolation issues with iTRAQ labeling (Wuhr et al. 2012) contribute to this overlap.

Of these 635 differentially localized proteins identified in our study, 459 and 176 proteins were localized to PM and TM, respectively (**Table B.2** and **Fig. B.2**). This was a surprising result, with over 2.5 times the number of proteins identified in PM compared to TM, suggesting that the protein composition of TM is more limited and specialized than PM. Proteins involved in protein trafficking, nutrient transport, cofactor biosynthesis, cell envelope function, secretion, and small molecule transport were found predominately in the PM (**Fig. B.4**). In comparison,

TM was characterized by proteins involved in metabolism, ATP generation, electron transport, and photosynthesis (**Fig. B.4**).

Another possible rationale for this finding could be that more abundant proteins in TM, such as PSI and PSII components, bias against peptide/protein identification compared to PM. However, considering the sensitivity of the analyses and the large number of overall proteins identified, it is unlikely that there are a significant number of proteins specific to TM that have been missed. An additional consideration is the overall abundance of PM versus TM in the cell. We chose to compare similar peptide/protein amounts between PM and TM for optimizing coverage and comparison. This could introduce a bias for greater PM protein coverage; however, when viewing the volcano plot (Supplemental Fig. S2), the right side (PM specific) is more consistently populated at higher p-values and fold changes compared to the left side (TM specific), which interestingly appears more variable in its protein distribution, regardless of which cutoff value is used. Furthermore, annotation of the findings is consistent with the diversity of roles of PM in cellular processes. Though there may be limited biases in the protein distributions, overall our results describe independent membrane systems that have unique, highly specific cellular roles.

B.3.2 Characteristics of Identified Proteins

Important characteristics of membrane proteins describe their types (e.g., integral membrane proteins or soluble proteins) and provide other relevant information such as the number of transmembrane helices (TMH) and the protein orientation across the membrane (for transmembrane proteins) or presence/absence of a signal peptide (for soluble proteins). From the 176 TM proteins identified in the current study, the number of TMH was predicted using TMHMM, and lipoprotein N-terminal signals were predicted using SignalP/LipoP. Based on this

analysis, Figure 3a shows the predicted topologies of identified TM proteins. Eighty-seven (of 176) TM proteins (49%) were identified as integral membrane proteins. About 45% of these integral membrane proteins have only one TMH, while about 20% have more than five. Seven (of 87) integral membrane proteins have N-terminal signal peptides as identified by SignalP 4.1; these include three photosystem II proteins (Sll1418 (PsbP), Sll1194 (PsbU), and Sll0427 (PsbO)), the photosystem I reaction center subunit III precursor/plastocyanin docking protein PsaF (Sll0819), the cytochrome c550 protein (Sll0258), a H^+/Ca^{+2} exchanger protein (Slr1336), and a hypothetical protein (Slr1273). Five of the identified non-integral TM proteins contain an N-terminus with a consensus pattern for lipoproteins. There are five proteins with a Sec signal among the rest of the proteins (i.e., soluble proteins). Finally, 79 of the 84 soluble proteins do not have any N-terminal signal peptide and hence are considered as peripheral proteins on the cytoplasmic side of the thylakoid membrane.

Based on similar analysis as that for TM proteins, **Fig. B.5b** shows the predicted topologies of identified PM proteins. Of 459 PM proteins, 150 (33%) were identified as integral membrane proteins. About 37% of these 150 integral membrane proteins have only one TMH while about 28% have more than five. Eight (of 150) integral membrane proteins have N-terminal signal peptides, including three proteins (Slr1744, Slr0089, and Slr1897) from cell envelope, fatty acid, and transport/binding processes, respectively, and five other hypothetical/unknown proteins. Nine of the identified non-integral PM proteins contain an N-terminus with a consensus pattern for lipoproteins, whereas 22 others were detected as soluble proteins with a Sec signal on the periplasmic side. Since the remaining 278 of the soluble proteins do not have any N-terminal signal peptide, they are considered as peripheral proteins on the cytoplasmic side of the plasma membrane.

Synechocystis 6803 is predicted to have more than 800 membrane proteins (Wang et al. 2009; Gao et al. 2015), or ~22% of the predicted 3672 open reading frames. In this study, we have identified a total of 237 integral membrane proteins differentially localized to TM or PM (87 in TM and 150 in PM). The remaining set of integral membrane proteins (~500) are proteins that are potentially found in OM, in both PM and TM, or were not identified in our analysis. Our strategy was successful in identifying a substantial number of integral membrane proteins with more than five TMHs (17 in TM and 42 in PM) (**Fig. B.5**), demonstrating the ability of this approach to identify highly hydrophobic proteins.

B.3.3 Photosystems and Respiratory Proteins

The current work identified a majority of the subunits of important photosynthetic complexes localized to the TM, including photosystem I (PSI), photosystem II (PSII), cytochrome b_6f, and ATP synthase. Similarly, major respiratory complexes (e.g., NADH dehydrogenase and cytochrome b_6f) are also mostly located in the TM. However, of the two subunits of cytochrome oxidase, one is located in the TM (Slr1136) and the other in the PM (Sll0813).

Table B.3 shows the most abundant integral TM proteins and includes many important proteins from photosynthetic and respiratory metabolism. Included in this set are four PSI proteins (Slr1834 (PsaA), Sll0819 (PsaF), Slr1655 (PsaL), and the PSI assembly related protein Sll0226), all which have one or more TM TMH. Note that PsaF is the only PSI protein having N-terminal signal peptides (see above). There are five other PSI-related proteins including Sll0563 (PsaC), Slr0737 (PsaD), and Ssr2831 (PsaE) identified as peripheral TM proteins on the cytoplasmic side (**Fig. B.5**).

Out of eight PSII proteins (Psb subunits with one or more TMH), four proteins, Sll0427

(PsbO), Sll1418 (PsbP), Sll1194 (PsbU), and Sll0258 (PsbV), have a single TMH each with N-terminal signal peptides. Of these, PsbO, PsbU, and PsbV are included with the most abundant proteins in **Table B.3**. Three other PSII-related proteins were detected as soluble proteins: Slr2034 (Ycf48) on the lumenal side and two Psb28 subunits (Slr1739 and Sll1398) on the cytoplasmic side. This localization of Psb28 is consistent with previous reports (Dobakova et al. 2009). Consistent with a recent report (Selao et al. 2016), the D1 protein processing protease, CtpA (Slr0008) was found as an integral TM protein (**Table B.2**).

Of seven subunits of ATP synthase located in TM, Sll1323 (AtpG(β)) and Sll1322 (AtpI(α)) (**Table B.3**) have one and five TM TMH, respectively, with the remaining being soluble proteins on the cytoplasmic side. There are 9 NADH dehydrogenase subunits identified in TM, five (NdhA, NdhB, NdhE, NdhF1 and NdhF4) as integral and four others as soluble proteins. However, one of the two Type 2 dehydrogenase subunits (Sll1484) is found as an integral PM protein, while the other (Slr1743) is a soluble TM protein on the cytoplasmic side. Among the remaining proteins, two cytochrome b_6f complex proteins (PetA and PetC) were identified in the TM. Overall, almost the entire electron transport machinery functions in the TM and these comprise a large portion of the proteins listed in **Table B.3**.

B.3.4 Pigment Biosynthesis and Transport Proteins

Our analysis revealed the location of 11 proteins from pigment biosynthesis (3 from carotenoid and 8 from cobalamin, heme, phycobilin and porphyrin metabolism). All but ferrochelatase (HemH/ScpA) were identified as PM soluble proteins on the cytoplasmic side.

A total of 62 transport (and binding) proteins were identified in the PM. Among these proteins, 34 proteins were PM integral proteins, including components of ABC transporters, permease proteins, biopolymer transporter system, P-type ATPase, and metal ion/cation transport

system proteins. Of the 35 most abundant integral PM proteins identified (**Table B.4**), half were classified as hypothetical/unknown, and the majority of the categorized proteins were transport/binding proteins.

The periplasmic sugar-binding protein of an ABC transporter, Slr1897, is the only protein with a single transmembrane span and N-terminal signal peptides. Seven PM proteins including Slr0040 (CmpA), Sll1450 (NrtA), and Slr0447 (UrtA) were detected as lipoproteins, while two others (NatB and Slr2043) as soluble periplasmic proteins. The remaining 19 proteins (mostly ATP binding proteins of ABC/urea transporter) are soluble proteins on the PM cytoplasmic side. Compared to the large number of transport proteins located in PM, only 7 such proteins are located in the TM. Five of these that involve metal (i.e., Ca^{+2} and Na^{+}) and urea transport systems were found to be integral TM proteins, while (similar to PM) the remaining two were ATP binding proteins located on the cytoplasmic side of the membrane.

B.3.5 Proteins Involved in Other Important Cellular Processes

Of the 50 proteins from cell envelope and cellular processes that were identified in the current study, 10 proteins, including 5 putative porins and PilQ, are soluble PM proteins on the periplasmic side, while solute-binding protein Slr1962 and putative endoglucanase Slr0897 were detected as a PM lipoprotein. Note the remaining proteins including pilus biogenesis protein Slr0063 were detected as soluble PM proteins on the cytoplasmic side. Fourteen other proteins including MurC, PilC, PilA2, Ctr1, and TaxD1 are integral PM proteins. Among these proteins, N-acetylmuramoyl-L-alanine amidase (Slr1744) has a single transmembrane span with N-terminal signal peptides. In addition, 5 proteins (e.g., PilA1) were identified as integral TM proteins. Only one protein from the amino acid metabolism, cytochrome *b* subunit of nitric oxide reductase (NorB), and two proteins (DesD and Sll0418) from fatty acid metabolism were

identified as integral TM proteins, while just one protein, gamma-tocopherol methyltransferase (Slr0089) from fatty acid metabolism, was identified as a PM protein.

Compared to proteins from metabolic/transport processes, only a handful of proteins involved in translation and regulation were identified in PM/TM by our current work. These include a protease (Slr0535) as an integral PM protein, and the tRNA synthetase AsnS, and the cis-trans isomerase Sll0408 as integral TM proteins. From regulatory mechanism, IcfG (Carbon metabolism regulatory protein), Slr1225 and Slr0599 (Serine/Threonine kinase), and two component system (Hik6/10/12/21/31) proteins were detected as integral proteins in the PM.

B.3.6 Hypothetical and Unknown Proteins

Hypothetical proteins are predicted from nucleic acid sequences without experimental evidence. Oftentimes, these proteins are associated with low identity to known/annotated proteins (Lubec et al. 2005). Unknown proteins are also predicted by bioinformatics tools, but unlike hypothetical proteins, these are experimentally proven to exist without any biochemical characterization (Lubec et al. 2005). Based on the annotations provided on the Cyanobase database (Nakao et al. 2010), the *Synechocystis* genome contains about 33% hypothetical and 18% unknown ORFs. Our current study reveals about 25% and 13% of PM proteins as hypothetical and unknown, respectively. The corresponding percentages are 32% (hypothetical) and 10% (unknown) for TM proteins.

Of 57 hypothetical TM proteins, 27 are integral, 11 have one predicted TMH, 7 have two, and the remaining 9 have three or more helices. One of these proteins (Slr1273) has an N-terminal signal peptide. In addition, 2 of the identified non-integral hypothetical TM proteins contain an N-terminus with a consensus pattern for lipoproteins, whereas 3 others are soluble proteins on the lumen side. Finally, the remaining 25 proteins are peripheral proteins on the

cytoplasmic side of TM. Among 17 unknown TM proteins, 10 were identified as integral, 7 have one predicted TMH, and the remaining 3 have four or more helices. Sll0022 was identified as a non-integral lipoprotein, while the remaining 6 are soluble proteins on the cytoplasmic side.

Of 117 hypothetical PM proteins, 51 are integral, 19 have one predicted TMH, 18 others (three sets of 6 proteins) have two, three, and four helices, respectively, and 14 have five or more helices. Two such proteins (Slr0200 and Slr0765) were found to have N-terminal signal peptides. Four other proteins were identified as soluble proteins on the periplasmic side, whereas the remaining 62 are peripheral proteins on the cytoplasmic side of PM. Among these 62 unknown PM proteins, 25 were identified as integral, 12 have one predicted TMH, 3 have two, 3 others have three, and 7 have four or more helices. Of these 25, Slr01257 and Ssr0693 have N-terminal signal peptides. The remaining proteins include 2 non-integral lipoproteins, 3 soluble PM proteins on the periplasmic side, and 32 peripheral proteins on the cytoplasmic side.

B.4 Discussion

Of the 635 differentially abundant proteins identified in our study, 459 proteins were localized to PM and 176 were localized to TM. **Fig. B.2** shows the clear quantitative differentiation between these two localized protein groups; however, interestingly, we also observed more quantitative variation within the biological replicates of TM or PM protein groups than anticipated (more so for the TM preparation, as also seen in **Fig. B.2**), likely hinting at a more dynamic localization of proteins between biological replicates. This is difficult to determine, however, as the scale and ratio type data of iTRAQ quantification is somewhat limited for this type of comparison. Regardless, the fact that over 2.5 times the number of proteins appear localized to PM compared to TM suggests that the protein composition of TM is more limited and specialized compared to PM. Functional annotation of these differential

proteins also confirms, that as a whole, our data describe two membrane systems that have very different roles: PM is involved in transport, secretion, trafficking, and general "gatekeeping" functions, whereas TM is devoted to the energetics of electron transport and cellular metabolism (**Fig. B.4**). Based on our analysis, it is evident that a higher percentage of PM proteins are soluble proteins (i.e., 65% PM soluble proteins vs. 48% TM soluble proteins), while the opposite is the case with integral membrane proteins (i.e., 32% PM integral proteins vs. 49% TM integral proteins). Therefore, these topological/functional variations are well correlated with their stated roles. An overview of the distribution of cellular processes between PM and TM is depicted in **Fig. B.7**.

The association between PM and TM in cyanobacteria has long been a topic of interest, and even though we technically identified the majority of proteins across both membranes, we assert that these data provide further evidence that PM and TM are separate membrane systems, since given the differences in protein composition described here, it is unlikely that PM and TM are contiguous. Other studies have investigated a membrane subfraction existing between PM and TM, but our purification procedure does not result in the isolation of this subfraction. In fact, we did not identify the PratA protein (Slr2048), a marker for this subfraction, in our analysis. The PSII biogenesis and repair cycle has recently been explored by a combination of aqueous two-phase partitioning, epitope tagging, and radioactive pulse chase, and it was determined that processing of the D1 protein occurs in the TM (Selao et al. 2016), a finding that is consistent with our localization of the CtpA processing protease in the TM (**Table B.3**). However, this study found the vesicle-inducing protein 1 (Vipp1, Sll0617) predominately in the PM (Selao et al. 2016), with a weaker signal found in the TM, while our analysis identified Vipp1 as a TM protein (**Table B.3**). Interestingly, we identified several photosystem-related proteins in the set

of proteins shared between PM and TM, including PsbA3 (Sll1867), PsbA1 (Slr1181), PsaB (Slr1835), and PsbB (Slr0906).

One of the goals of this study was to generate a comprehensive list of uniquely localized proteins for both PM and TM (**Table B.2**). We therefore examined the scope of the current work in comparison to the most comprehensive previous study, that of (Pisareva et al. 2011). Of the 176 proteins we identified localized to TM, 149 of these were uniquely identified in our analysis, while 27 proteins were also identified by (Pisareva et al. 2011). However, an additional 73 proteins were uniquely localized to the TM by (Pisareva et al. 2011), but not in the current study. Similarly, of the 459 proteins we identified as localized to PM, 71 of these were also identified by Pisareva et al in the PM, with 86 additional proteins uniquely localized to the PM by Pisareva et al. **Table B.2** summarizes the proteins uniquely identified by Pisareva et al in both PM and TM. It is interesting to note that the vast majority (84%) of these 73 proteins uniquely localized to the TM by Pisareva et al were identified/quantified with one spectra (61 out of 73), from which it is difficult to accurately determine or quantify appropriate localization. This most likely explains the limited overlap between the two studies, and in this regard, our current, more in-depth study, which included pre-MS fractionation along with quantitative iTRAQ labeling performed in biological triplicate, provides the necessary quantitative information to inform upon the previously limited protein identifications. This more comprehensive survey of TM components included key photosynthetic and electron transport components, i.e., 7 different NADH dehydrogenase subunits (Subunits 2 and 5 as integral and the rest as soluble proteins), 7 photosystem I proteins (3 integral and 4 soluble), 9 photosystem II proteins (5 integral and 4 soluble), and 5 ATP synthase subunits (2 integral and 3 soluble).

A similar comparison in terms of PM proteins across these two data sets revealed overlap of 71 proteins, while 388 proteins were identified as PM specific in the current study and 86 identified as PM specific by (Pisareva et al. 2011) (**Fig. B.7b**). Among these unique proteins, 160 (out of 388) and 49 (out of 86) are either hypothetical or have unknown functions. Of the remaining 228 PM proteins detected in the current analysis, 53 are integral membrane proteins (such as two-component hybrid system proteins, ATP binding proteins on ABC transporters, and Na^+/H^+ antiporter), 7 are substrate-binding lipoproteins, 11 are soluble proteins on the periplasmic end (e.g., putative porins), and the remaining ones are peripheral proteins on the cytoplasmic side.

Cross-comparison between the current dataset and that in (Pisareva et al. 2011) (see Supplemental Table S2) also revealed multiple instances where proteins previously deemed localized to TM or PM were in fact found quantitatively enriched in the other membrane. Seven TM proteins from the current study were identified in the PM by Pisareva et al., and 16 PM proteins from the current study were found in the TM by Pisareva et al. Excluding hypothetical/unknown proteins, 3 TM proteins identified in the current study included two (Sll0897 and Sll1260) peripheral proteins on the cytoplasmic end, plus lipoprotein Sll0915. Out of 10 PM proteins, 4 are integral membrane proteins (e.g., two-component regulatory proteins and binding proteins), 2 are lipoproteins (GgtB and Slr0804) with the remaining being soluble proteins. The remainder of unique proteins found by (Pisareva et al. 2011) (57 TM and 79 PM proteins) were either not detected or excluded through our analysis pipeline requirement of multiple peptide identifications per protein, (27 TM and 38 PM proteins), or quantitatively rejected due to p-value criterion (30 TM and 96 PM proteins). Overall, these proteins are distributed across different functional categories; however, a significant portion of these has

photosynthesis/respiratory function or are hypothetical/unknown proteins. A detailed comparison of TM/PM proteins between these two data sets is included in Supplemental Table S2.

Synechocystis 6803 is a gram-negative photosynthetic bacterium, and therefore, the general PM features are predicted to resemble those of a gram-negative bacterium such as *E. coli*. In addition, since a cyanobacterial ancestor was the progenitor of plant chloroplasts, the *Synechocystis* 6803 TM membrane properties might also be correlated to that of a plant species such as the model plant *Arabidopsis thaliana*. Based on this premise, we carried out a comparative study by using literature data available on topologies of *E. coli* PM (von Heijne 2006) and Arabidopsis TM (Ferro et al. 2010) (**Fig. B.7**). As shown in **Fig. B.7a**, both *E. coli* and *Synechocystis* 6803 have an almost similar amount of PM proteins (17% vs. 13% of total proteins in *E. coli* (Tatusova et al. 2015) and *Synechocystis* 6803 (UniProt: a hub for protein information 2015), respectively), but two categories (namely, transport proteins and proteins involved in cellular processes such as biogenesis, cell envelope development, DNA replication/repair, and regulation) differ considerably. In contrast to 13% of PM proteins involved in transport in *Synechocystis* 6803, *E. coli* has 40% proteins involved in various transport activities (e.g., export/import, active/passive transport). Interestingly, a significant portion of *Synechocystis* 6803 transport proteins is involved in generic ABC transporters, whereas the majority of *E. coli* transporter proteins are substrate specific (e.g., amino acid transporters). This difference might be primarily due to their physiological differences: while *Synechocystis* 6803, as a photosynthetic bacterium, is able to synthesize all essential amino acids, *E. coli* is equipped to import these when available, without expending energy on their production (Shimizu 2014). Compared to a mere 6% of *E. coli* PM proteins, a staggering 31% *Synechocystis* 6803 PM proteins are involved in various cellular processes such as cell envelope development,

DNA modification, restriction, modification, recombination, and repair, regulatory functions, transcription, and translation. Another interesting difference is the localization of electron transfer chain: while in *E. coli* the entire electron transfer chain is in the PM, in *Synechocystis* 6803 it mostly functions in the TM. It is interesting to note that the evolutionary line of bacteria involves the transition from anoxygenic to oxygenic photosynthesis, and in purple anoxygenic photosynthetic bacteria, reaction centers are located in a structure called the intracytoplasmic membrane, which is a specialized invagination of the plasma membrane (Sener et al. 2007). This is in contrast to the autonomous nature of the PM and TM in cyanobacteria.

Fig. B.7b shows the comparison between Arabidopsis and *Synechocystis* 6803 TM proteins. While proteins from the majority of categories, such as photosystems, binding, and NDH, have similar contributions across these two photosynthetic species, the main differences lie in the corresponding contributions of metabolic and unknown proteins. About 44% of *Synechocystis* 6803 TM proteins have unknown functions, which might be due to differences in the levels of annotation of these two species. In contrast, compared to less than 1% of *Synechocystis* 6803 TM metabolic proteins, about 14% of Arabidopsis TM proteins function as metabolic proteins mainly in vitamin, pigment, and lipid metabolism. Given that a cyanobacterial ancestor was the progenitor of chloroplasts via an endosymbiotic event, these differences point to the changes that have occurred during the evolution of modern-day chloroplasts.

To summarize, the *Synechocystis* 6803 PM functions similarly to a gram-negative bacterial PM with additional activities across different cellular processes, while the TM behaves in a slightly different way compared a plant TM, with lesser activities in metabolic processes. Overall, the *Synechocystis* 6803 membrane systems can be considered as a gram-negative bacterial PM membrane system with bioenergetic electron transfer functioning in the TM.

B.5 Acknowledgements

The studies at Washington University were supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award # DE-FG02-99ER20350 to HBP. The studies at PNNL were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy Office of Biological and Environmental Research national scientific user facility located at Pacific Northwest National Laboratory in Richland, Washington. Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract No. DE-AC05-76RLO 1830. AYN has been partially supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1143954).

B.5.1 Author Contributions

R.D.S., D.W.K., and H.B.P. designed research; M.L., R.S., J.M.J., A.Y.N., and M.A.G. performed research; J.M.J., M.A.G., and R.D.S. contributed new reagents or analytic tools; M.L., R.S., J.M.J., A.Y.N., M.A.G., D.W.K., and H.B.P. analyzed data; M.L., R.S., J.M.J., A.Y.N., M.A.G., R.D.S., D.W.K., and H.B.P. wrote the paper.

B.5.2 Conflict of Interest Statement

Authors declare no competing conflicts of interest.

B.6 References

Dobakova, M., Sobotka, R., Tichy, M., Komenda, J. (2009). "Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium *Synechocystis* sp. PCC 6803." *Plant Physiology*. **149**(2): 1076-1086.

Elias, J.E., Gygi, S.P. (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." *Nature Methods*. **4**(3): 207-214.

Ferro, M., Brugiere, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Le Gall, S., Kieffer-Jaquinod, S., Bruley, C., Garin, J., Joyard, J., Masselon, C., Rolland, N. (2010). "AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins." *Molecular & Cellular Proteomics*. **9**(6): 1063-1084.

Gao, L., Wang, J., Ge, H., Fang, L., Zhang, Y., Huang, X., Wang, Y. (2015). "Toward the complete proteome of *Synechocystis* sp. PCC 6803." *Photosynthesis Research*. **126**(2-3): 203-219.

Huang, F., Parmryd, I., Nilsson, F., Persson, A.L., Pakrasi, H.B., Andersson, B., Norling, B. (2002). "Proteomics of *Synechocystis* sp. strain PCC 6803: identification of plasma membrane proteins." *Molecular & Cellular Proteomics*. **1**(12): 956-966.

Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., Krogh, A. (2003). "Prediction of lipoprotein signal peptides in Gram-negative bacteria." *Protein Science*. **12**(8): 1652-1662.

Kelly, R.T., Page, J.S., Luo, Q., Moore, R.J., Orton, D.J., Tang, K., Smith, R.D. (2006). "Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry." *Analytical Chemistry*. **78**(22): 7796-7801.

Keren, N., Liberton, M., Pakrasi, H.B. (2005). "Photochemical competence of assembled photosystem II core complex in cyanobacterial plasma membrane." *Journal of Biological Chemistry*. **280**(8): 6548-6553.

Kim, S., Gupta, N., Pevzner, P.A. (2008). "Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases." *Journal of Proteome Research*. **7**(8): 3354-3363.

Kim, S., Mischerikow, N., Bandeira, N., Navarro, J.D., Wich, L., Mohammed, S., Heck, A.J., Pevzner, P.A. (2010). "The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search." *Molecular & Cellular Proteomics*. **9**(12): 2840-2852.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L. (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." *Journal of Molecular Biology*. **305**(3): 567-580.

Liberton, M, Austin, J.R. II., Berg, R.H., Pakrasi, H.B. (2011). "Unique thylakoid membrane architecture of a unicellular N_2-fixing cyanobacterium revealed by electron tomography." *Plant Physiology*. **155**(4): 1656-1666.

Liberton, M., Howard, B.R., Heuser, J., Roth, R., Pakrasi, H.B. (2006). "Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803." *Protoplasma*. **227**(2-4): 129-138.

Livesay, E.A., Tang, K., Taylor, B.K., Buschbach, M.A., Hopkins, D.F., LaMarche, B.L., Zhao, R., Shen, Y., Orton, D.J., Moore, R.J., Kelly, R.T., Udseth, H.R., Smith, R.D. (2008). "Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses." *Analytical Chemistry*. **80**(1): 294-302.

Lubec, G., Afjehi-Sadat, L., Yang, J.W., John, J.P. (2005). "Searching for hypothetical proteins: theory and practice based upon original data and literature." *Progress in Neurobiology*. **77**(1-2): 90-127.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T., Nakamura, Y. (2010). "CyanoBase: the cyanobacteria genome database update 2010." *Nucleic Acids Research*. **38**(Database issue): D379-381.

Nickelsen, J., Rengstl, B., Stengel, A., Schottkowski, M., Soll, J., Ankele, E. (2011). "Biogenesis of the cyanobacterial thylakoid membrane system--an update." *FEMS Microbiology Letters*. **315**(1): 1-5.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H. (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." *Nature Methods*. **8**(10): 785-786.

Pisareva, T., Kwon, J., Oh, J., Kim, S., Ge, C., Wieslander, A., Choi, J.S., Norling, B. (2011). "Model for membrane organization and protein sorting in the cyanobacterium *Synechocystis* sp. PCC 6803 inferred from proteomics and multivariate sequence analyses." *Journal of Proteome Research*. **10**(8): 3617-3631.

Pisareva, T., Shumskaya, M., Maddalo, G., Ilag, L. (2007). "Identification of novel integral plasma membrane proteins." *FEBS Journal.* **274**(3): 791-804.

Polpitiya, A.D., Qian, W.J., Jaitly, N., Petyuk, V.A., Adkins, J.N, Camp DG II., Anderson, G.A., Smith, R.D. (2008). "DAnTE: a statistical tool for quantitative analysis of -omics data." *Bioinformatics*. **24**(13): 1556-1558.

Qian, W.J., Liu, T., Monroe, M.E., Strittmatter, E.F., Jacobs, J.M., Kangas, L.J., Petritis, K., Camp, DG II., Smith, R.D. (2005). "Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome." *Journal of Proteome Research*. **4**(1): 53-62.

Rengstl, B., Oster, U., Stengel, A., Nickelsen, J. (2011). "An intermediate membrane subfraction in cyanobacteria is involved in an assembly network for Photosystem II biogenesis." *Journal of Biological Chemistry*. **286**(24): 21944-21951.

Schneider, D., Fuhrmann, E., Scholz, I., Hess, W.R., Graumann, P.L. (2007). "Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes." *BMC Cell Biology*. **8**(39): 1-10.

Selao, T.T., Zhang, L., Knoppova, J., Komenda, J., Norling, B. (2016). "Photosystem II assembly steps take place in the thylakoid membrane of the cyanobacterium *Synechocystis* sp. PCC 6803." *Plant Cell Physiology*. **57**(1): 95-104.

Sener, M.K., Olsen, J.D., Hunter, C.N., Schulten, K. (2007). "Atomic-level structural and functional model of a bacterial photosynthetic membrane vesicle." *Proceedings of National Academy of Sciences*. **104**(40): 15723-15728.

Shimizu, K. (2014). "Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stress." *Metabolites*. **4**(1): 1-15.

Srivastava, R., Pisareva, T., Norling, B. (2005). "Proteomic studies of the thylakoid membrane of *Synechocystis* sp. PCC 6803." *Proteomics*. **5**(18): 4905-4916.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., Tolstoy, I. (2015). "RefSeq microbial genomes database: new representation and annotation strategy." *Nucleic Acids Research*. **43**(7): 3872.

UniProt Consortium. (2015). "UniProt: a hub for protein information." *Nucleic Acids Research*. **43**(Database issue): D204-212.

van de Meene, A.M., Hohmann-Marriott, M.F., Vermaas, W.F., Roberson, R.W. (2006). "The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC 6803." *Archives of Microbiology*. **184**(5): 259-270.

Vizcaino, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N., Binz, P.A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R.J., Kraus, H.J., Albar, J.P., Martinez-Bartolome, S., Apweiler, R., Omenn, G.S., Martens, L., Jones, A.R., Hermjakob, H. (2014). "ProteomeXchange provides globally coordinated proteomics data submission and dissemination." *Nature Biotechnology*. **32**(3): 223-226.

von Heijne, G. (2006). "Membrane-protein topology." *Nature Reviews Molecular Cell Biology*. **7**(12): 909-918.

Wang, Y., Xu, W., Chitnis, P.R. (2009). "Identification and bioinformatic analysis of the membrane proteins of *Synechocystis* sp. PCC 6803." *Proteome Science*. **7**(11): 1-12.

Welkie, D., Zhang, X., Markillie, M.L., Taylor, R., Orr, G., Jacobs, J., Bhide, K., Thimmapuram, J., Gritsenko, M., Mitchell, H., Smith, R.D., Sherman, L.A. (2014). "Transcriptomic and proteomic dynamics in the metabolism of a diazotrophic cyanobacterium, *Cyanothece* sp. PCC 7822 during a diurnal light-dark cycle." *BMC Genomics*. **15**(1185): 1-16.

Wuhr, M., Haas, W., McAlister, G.C., Peshkin, L., Rad, R., Kirschner, M.W., Gygi, S.P. (2012). "Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster." *Analytical Chemistry*. **84**(21): 9214-9221.

Table B.1: Identified peptide and proteins details.

Table B.2: List and analysis of total proteins detected and quantified.

Tables B.1-B.2 can be found at:

http://www.mcponline.org/content/early/2016/04/07/mcp.M115.057240/suppl/DC1.

Table B.3: Distribution of TM integral proteins with Log_2 ratio of -2 or less across different pathways/subsystems.

Proteins	Function	Subsystem/Pathway	Log_2 ratio (PM/TM)	No of TMH[a]
Sll0026	NADH dehydrogenase subunit 5 (involved in constitutive, low affinity CO2 uptake)	Photosynthesis and respiration	-3.65	16
Sll0223	NADH dehydrogenase subunit 2	Photosynthesis and respiration	-4.25	14
Sll0226	photosystem I assembly related protein	Photosynthesis and respiration	-2.01	2
Sll0258	photosystem II PsbV protein	Photosynthesis and respiration	-3.50	1
Sll0262	acyl-lipid desaturase (delta 6)	Fatty acid, phospholipid and sterol metabolism	-2.16	5
Sll0408	peptidyl-prolyl cis-trans isomerase	Translation	-2.93	1
Sll0427	photosystem II PsbO protein	Photosynthesis and respiration	-4.85	1
Sll0450	cytochrome b subunit of nitric oxide reductase	Amino acid biosynthesis	-3.38	13
Sll0522	NADH dehydrogenase subunit 4L	Photosynthesis and respiration	-2.85	3
Sll0556	Na+/H+ antiporter	Transport and binding proteins	-2.36	12
Sll0819	photosystem I PsaF protein	Photosynthesis and respiration	-2.40	2
Sll0851	photosystem II CP43 protein	Photosynthesis and	-2.59	7

		respiration		
Sll1147	glutathione S-transferase	Biosynthesis of cofactors, prosthetic groups, and carriers	-3.05	3
Sll1194	photosystem II PsbU protein	Photosynthesis and respiration	-3.50	1
Sll1316	cytochrome b6-f complex iron-sulfur subunit (Rieske iron sulfur protein)	Photosynthesis and respiration	-4.50	1
Sll1322	ATP synthase A chain of CF(0)	Photosynthesis and respiration	-2.02	5
Sll1323	ATP synthase subunit b' of CF(0)	Photosynthesis and respiration	-2.84	1
Sll1471	phycobilisome rod-core linker polypeptide	Photosynthesis and respiration	-2.78	1
Sll1513	c-type cytochrome synthesis protein	Other categories	-3.81	8
Sll1694	pilin polypeptide PilA1	Cellular processes	-2.28	1
Sll1702	hypothetical protein YCF51	Hypothetical	-3.34	2
Sll1784	periplasmic protein, function unknown	Unknown	-2.78	1
Slr0228	cell division protein FtsH	Cellular processes	-2.72	2
Slr1336	H+/Ca2+ exchanger	Transport and binding proteins	-3.96	11
Slr1434	pyridine nucleotide transhydrogenase beta subunit	Biosynthesis of cofactors, prosthetic groups, and carriers	-2.44	9
Slr1645	photosystem II Psb27 protein	Photosynthesis and respiration	-2.15	1
Slr1655	photosystem I subunit XI	Photosynthesis and	-2.61	2

		respiration		
Slr1834	photosystem I PsaA protein	Photosynthesis and respiration	-2.65	9
Slr2087	c-type cytochrome biogenesis protein Ccs1	Other categories	-2.27	3
Ssr3451	photosystem II PsbE protein	Photosynthesis and respiration	-4.15	1
Sll0862	hypothetical protein	Hypothetical	-2.84	11
Sll5034	hypothetical protein	Hypothetical	-2.96	1
Slr0637	hypothetical protein	Hypothetical	-3.68	1
Slr0813	hypothetical protein	Hypothetical	-2.54	3
Slr0962	unknown protein	Unknown	-4.26	3
Slr1261	hypothetical protein	Hypothetical	-2.08	2
Slr1273	hypothetical protein	Hypothetical	-2.22	1
Slr1470	hypothetical protein	Hypothetical	-2.70	1
Slr1471	hypothetical protein	Hypothetical	-3.39	3
Slr1624	hypothetical protein	Hypothetical	-2.71	1
Ssl0410	unknown protein	Unknown	-2.84	1
Ssl1328	hypothetical protein	Hypothetical	-3.92	1
Ssl5113	unknown protein	Unknown	-2.47	1
Ssr2406	unknown protein	Unknown	-5.86	2

[a]Number of TMH was predicted using TMHMM-v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0)

Table B.4: Distribution of PM integral proteins with Log_2 ratio of 2 or more across different pathways/subsystems.

Proteins	Function	Subsystem/Pathway	Log_2 ratio (PM/TM)	No of TMH[a]
Sll0108	ammonium/methylammonium permease	Transport and binding proteins	3.01	12
Sll0169	cell division protein Ftn2 homolog	Cellular processes	2.04	1
Sll0574	probable permease protein of lipopolysaccharide ABC transporter	Transport and binding proteins	3.08	6
Sll0855	putative channel transporter	Transport and binding proteins	2.01	10
Sll0993	potassium channel	Transport and binding proteins	2.67	3
Sll1276	ATP-binding protein of ABC transporter	Transport and binding proteins	2.09	3
Sll1695	pilin polypeptide PilA2	Cellular processes	3.19	1
Slr0114	putative PP2C-type protein phosphatase	Unknown	2.71	2
Slr0369	RND multidrug efflux transporter	Transport and binding proteins	2.05	11
Slr0593	cAMP binding membrane protein	Unknown	2.11	6
Slr0615	ATP-binding protein of ABC transporter	Transport and binding proteins	2.53	6
Slr0678	biopolymer transport ExbD like protein	Transport and binding proteins	2.56	1

Slr0798	zinc-transporting P-type ATPase (zinc efflux pump) involved in zinc tolerance	Transport and binding proteins	2.27	5
Slr1149	ATP-binding protein of ABC transporter	Transport and binding proteins	2.66	5
Slr1216	Mg2+ transport protein	Transport and binding proteins	2.12	3
Slr1423	UDP-N-acetylmuramate-alanine ligase	Cell envelope	2.91	1
Slr1515	putative membrane protein required for bicarbonate uptake	Hypothetical	2.54	12
Slr1575	probable potassium efflux system	Transport and binding proteins	2.28	2
Sll0267	unknown protein	Unknown	2.14	3
Sll0283	hypothetical protein	Hypothetical	2.31	4
Sll0384	unknown protein	Unknown	2.01	4
Sll0505	hypothetical protein	Hypothetical	2.15	2
Sll0602	hypothetical protein	Hypothetical	2.25	1
Sll0727	hypothetical protein	Hypothetical	2.21	11
Sll1608	hypothetical protein	Hypothetical	2.09	7
Sll2003	hypothetical protein	Hypothetical	2.02	11
Slr0060	unknown protein	Unknown	2.09	1

Slr0594	hypothetical protein	Hypothetical	2.65	4
Slr0625	hypothetical protein	Hypothetical	2.59	11
Slr0818	hypothetical protein	Hypothetical	2.02	1
Slr0960	unknown protein	Unknown	2.31	3
Slr1257	unknown protein	Unknown	2.05	2
Slr1875	hypothetical protein	Hypothetical	2.05	4
Slr1927	hypothetical protein	Hypothetical	2.16	2
Slr2011	hypothetical protein	Hypothetical	2.24	3

[a]Number of TMH was predicted using TMHMM-v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0)

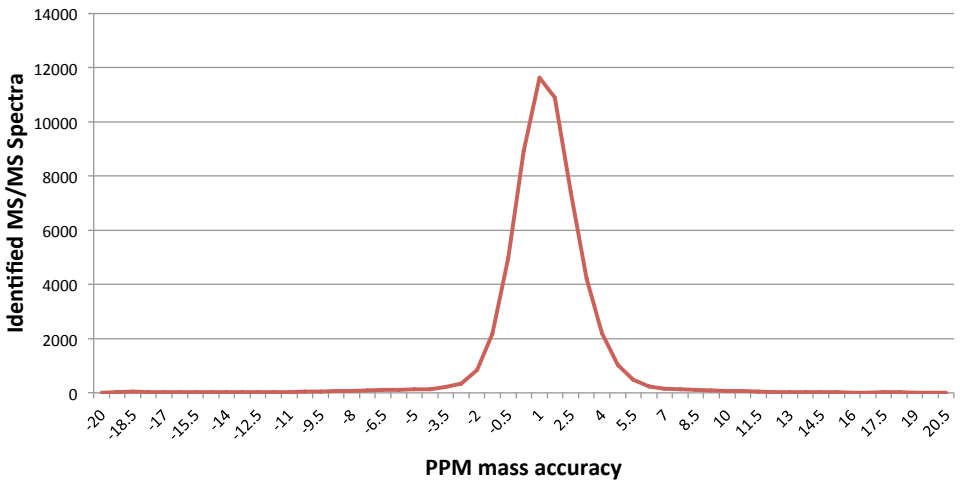


Figure B.1: Distribution of ppm mass accuracy for all identified spectra.

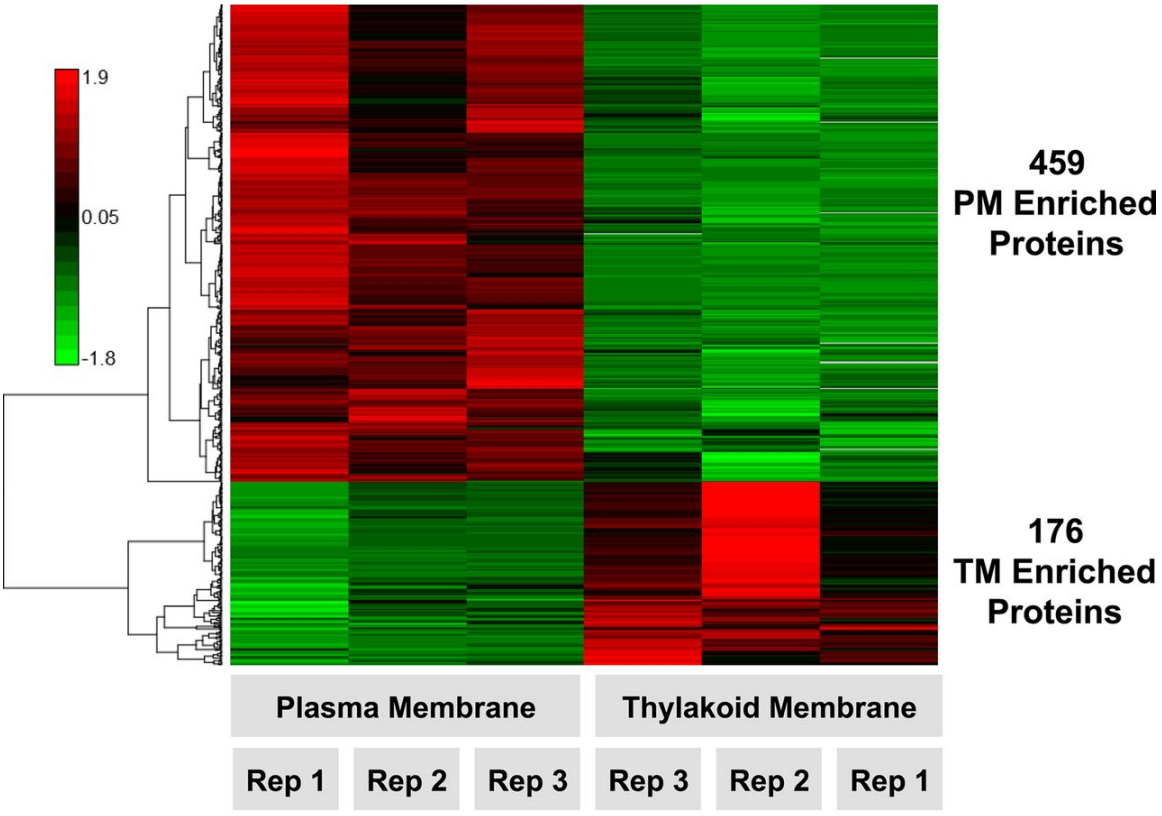


Figure B.2: Heat map of differentially abundant membrane proteins.

LC-MS/MS quantification of >1400 proteins resulted in the identification of 459 PM and 176 TM proteins observed with significantly different membrane localization after direct comparison, p-value <0.05, between plasma and thylakoid membrane sample preparations. Proteins were hierarchically clustered using a Pearson correlation distance metric. Three biological replicates of PM and TM samples were analyzed (as shown), which included the additional averaged technical replicate values for TM rep 2 and TM rep 3. The red-green color scale depicts normalized $\log_2$ ratio values of PM to TM abundance of any individual membrane protein (Red: high, Green: low).

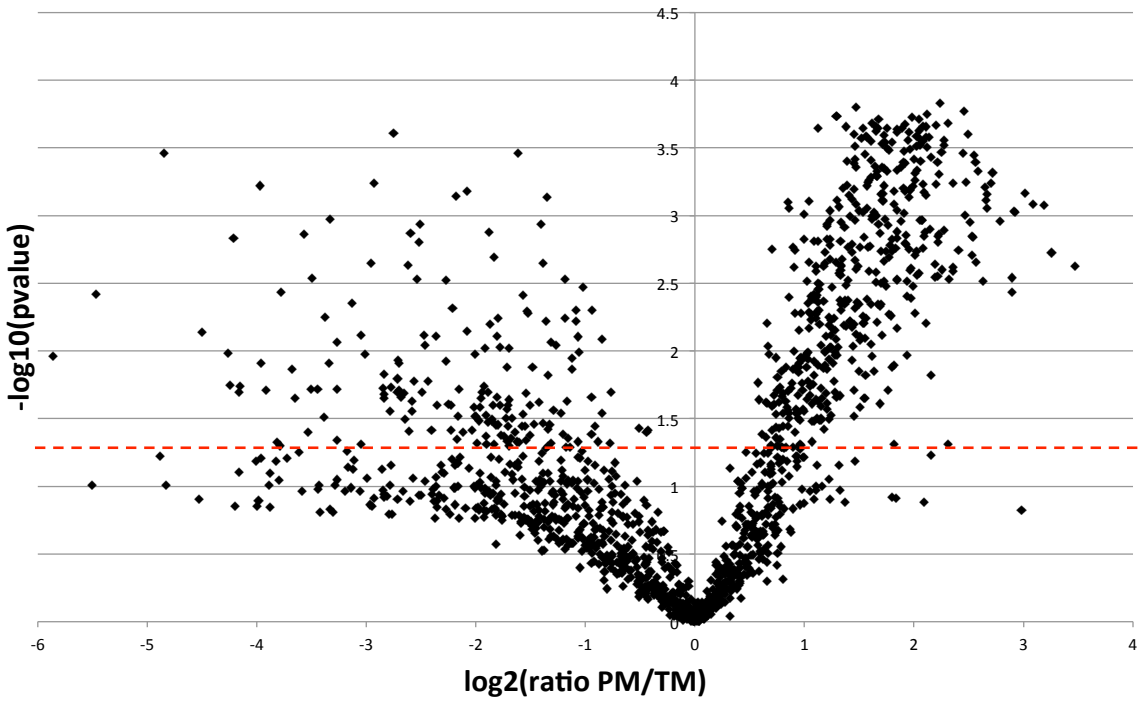


Figure B.3: Reverse volcano plot of all quantified protein identifications comparing PM and TM results.

Shown in red dotted line is the >0.05 p-value cut-off, resulting in the 635 proteins deemed differentially abundant between PM and TM preparations.

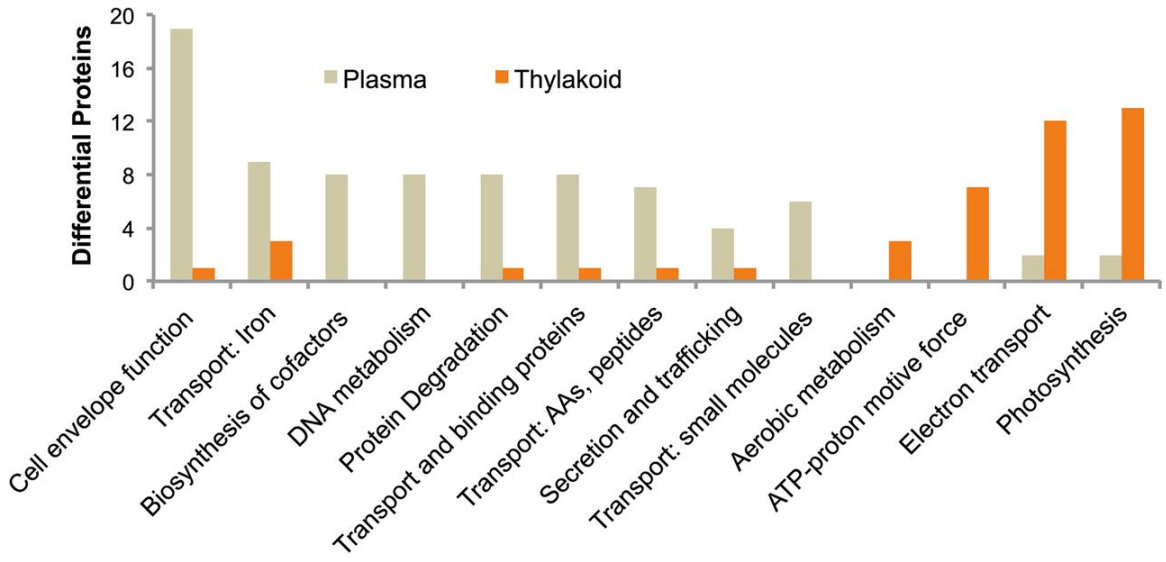


Figure B.4: Functional categorization distribution of proteins between PM and TM analyzed preparations.

Shown is the most abundant and relevant functional categorization populated by proteins enriched in either PM or TM preparations, demonstrating clear segregation of cellular processes. Categorization is based upon CyanoBase annotation.

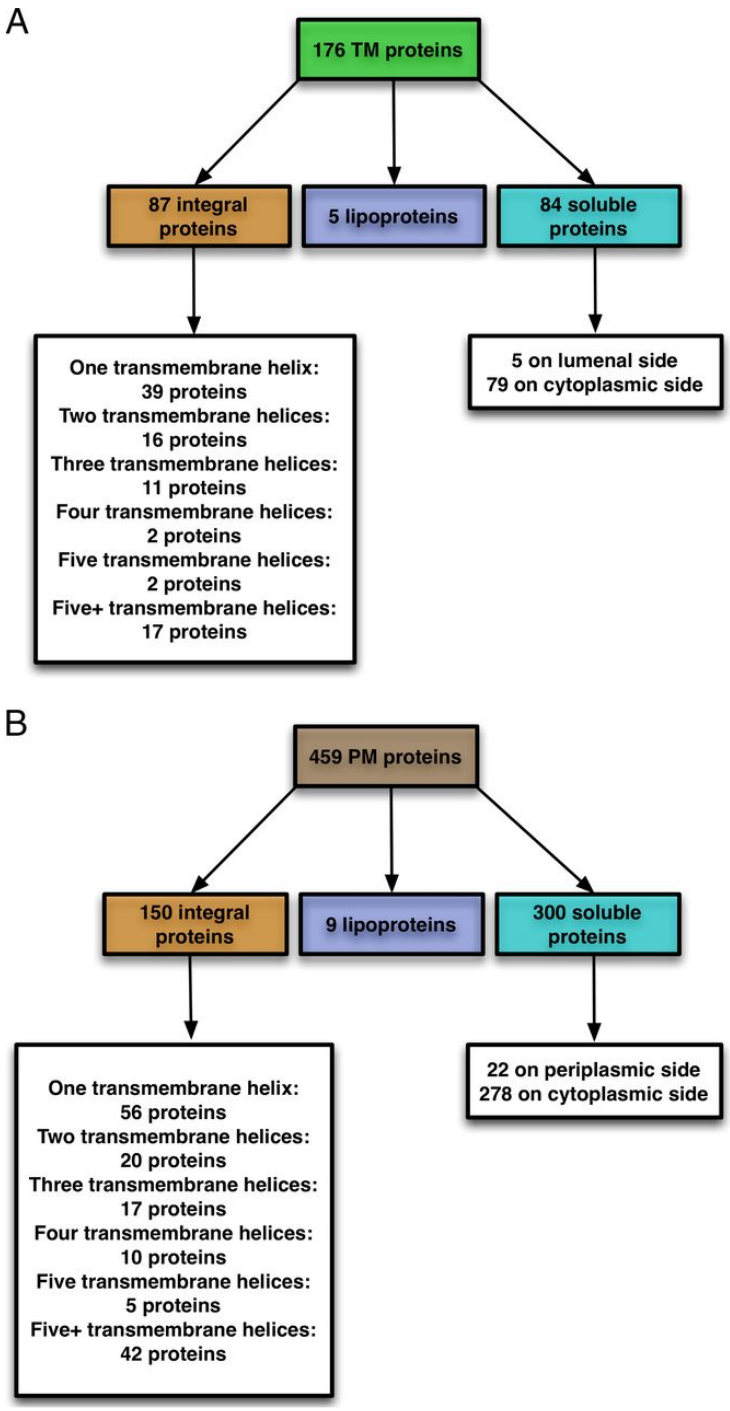


Figure B.5: Predicted topology of identified a TM and b PM proteins in *Synechocystis* 6803.

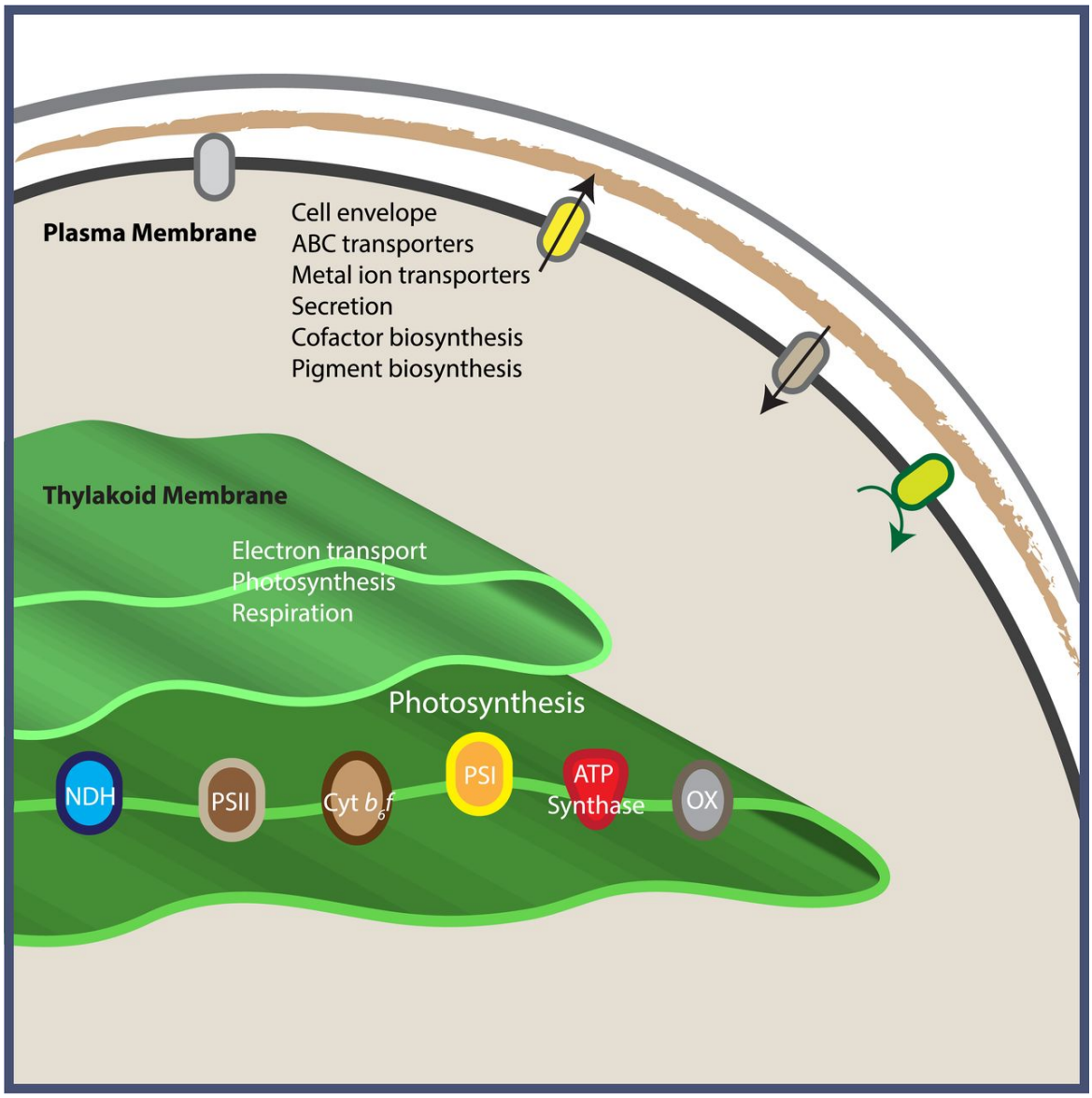


Figure B.6: Schematic drawing showing distribution of functional roles between PM and TM.

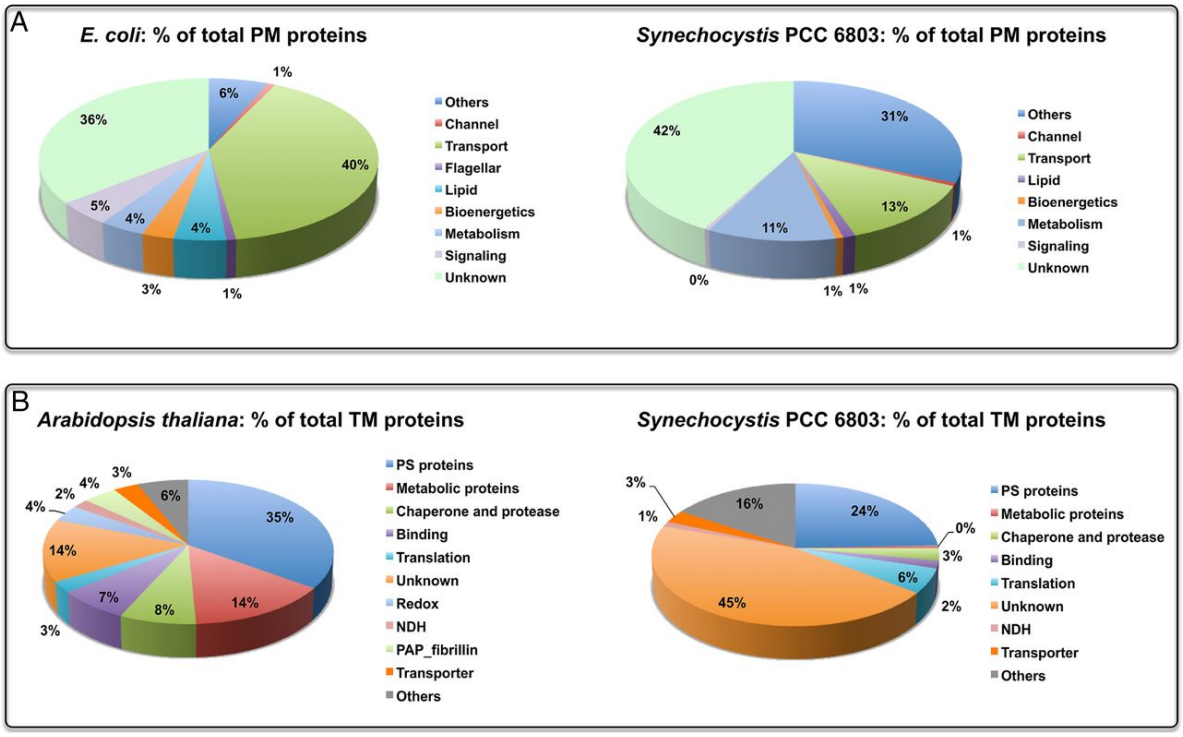


Figure B.7: Comparative analysis of membrane proteins.

a *E. coli* PM vs. *Synechocystis* PM, and **b** Arabidopsis TM vs. *Synechocystis* TM.

Appendix C

Non-*tenera* Contamination and the Economic Impact of *SHELL* Genetic Testing in the Malaysian Independent Oil Palm Industry

This work was originally published in *Frontiers in Plant Science*. Leslie Cheng-Li Ooi, Eng-Ti Leslie Low, Meilina Ong Abdullah, Rajanaidu Nookiah, Ngoot Chin Ting, Jayanthi Nagappan, Mohamad Arif Abdul Manaf, Kuang Lim Chan, Mohd Amin Halim, Norazah Azizi, Wahid Omar, Abdul Jalil Murad, Nathan Lakey, Jared M. Ordway, Anthony Favello, Muhammad A. Budiman, Andrew Van Brunt, Melissa Beil, Michael T. Leininger, Nan Jiang, Steven W. Smith, Clyde R. Brown, Alex Chun Seng Kuek, Shabani Bahrain, Allison Hoynes-O'Connor, Amelia Y. Nguyen*, Hemangi G. Chaudhari, Shivam A. Shah, Yuen May Choo, Ravigadevi Sambanthamurthi, and, Rajinder Singh. Non-*tenera* Contamination and the Economic Impact of *SHELL* Genetic Testing in the Malaysian Independent Oil Palm Industry. *Frontiers in Plant Science*. 2016. 7(771):1-13.

*A.Y.N. created the economic model, analyzed the data, and wrote the paper.

C.1 Introduction

C.1.1 Abstract

Oil palm (*Elaeis guineensis*) is the most productive oil bearing crop worldwide. It has three fruit forms, namely *dura* (thick-shelled), *pisifera* (shell-less) and *tenera* (thin-shelled), which are controlled by the *SHELL* gene. The fruit forms exhibit monogenic co-dominant inheritance, where *tenera* is a hybrid obtained by crossing maternal *dura* and paternal *pisifera* palms. Commercial palm oil production is based on planting thin-shelled *tenera* palms, which typically yield 30% more oil than *dura* palms, while *pisifera* palms are female-sterile and have little to no palm oil yield. It is clear that *tenera* hybrids produce more oil than either parent due to single gene heterosis. The unintentional planting of *dura* or *pisifera* palms reduces overall yield and impacts land utilization that would otherwise be devoted to more productive *tenera* palms. Here we identify three additional novel mutant alleles of the *SHELL* gene, which encode a type II MADS-box transcription factor, and determine oil yield via control of shell fruit form phenotype in a manner similar to two previously identified mutant *SHELL* alleles. Assays encompassing all five mutations account for all *dura* and *pisifera* palms analyzed. By assaying for these variants in 10,224 palms or seedlings, we report the first large scale accurate genotype-based determination of the fruit forms in independent oil palm planting sites and in the nurseries that supply them throughout Malaysia. The measured non-*tenera* contamination rate (10.9% overall on a weighted average basis) underscores the importance of *SHELL* genetic testing of seedlings prior to planting in production fields. By eliminating non-*tenera* contamination, comprehensive *SHELL* genetic testing can improve sustainability by increasing yield on existing planted lands. In addition, economic modelling demonstrates that *SHELL* gene testing will confer substantial annual economic gains to the oil palm industry, to Malaysian gross national income and to Malaysian government tax receipts.

C.1.2 Oil Palm

The *Elaeis* genus comprises two interfertile species, *E. guineensis* of West African origin and *E. oleifera* of Central and South American origin (Zeven 1965; Cochard et al. 2005). Due to higher oil yield, *E. guineensis* is the predominant species in commercial planting sites, and oil palm cultivation accounts for 45% of edible oil production worldwide. Oil palm has three naturally occurring fruit forms (*dura*, *pisifera* and *tenera*) which vary in oil yield depending on the presence and thickness of the distinct lignified shell surrounding the fruit kernel. *Dura* fruits have a thick shell (2-8 mm) and produce approximately 5.3 tons of oil per hectare per year. *Pisifera* fruits have no shell, are typically female sterile, and bunches prematurely rot prior to oil production. Crosses of *dura* and *pisifera* palms produce hybrids having *tenera* fruit with a relatively thin shell (0.5-3 mm) surrounded by a distinct fiber ring and can produce from 7.4 tons (Hartley 1988) to as high as 13.6 tons of oil per hectare per year, as seen in elite individuals (Sharma and Tan 1999). Therefore, *tenera* hybrids produced by selection of maternal (*dura*) and paternal (*pisifera*) pools are the basis for commercial palm oil production throughout southeast Asia (Rajanaidu et al. 2000).

The shell fruit form phenotype displays co-dominant monogenic inheritance (Beirnaert and Vanderweyen 1941). Our previous identification of the *SHELL* gene and two mutations responsible for *tenera* and *pisifera* fruit forms explained the single gene heterosis exhibited in *tenera* palms (Singh, Low et al. 2013). The discovery was recently independently verified and converted to a PCR based assay (Ritter et al. 2015). SHELL is a type II MADS-box transcription factor homologous to *Arabidopsis* SEEDSTICK (STK) and rice OsMADS13, which are members of transcription factor networks controlling differentiation of the ovule, seed and lignified endocarp in *Arabidopsis* (Favaro et al. 2003; Pinyopich et al. 2003; Dinneny and Yanofsky 2005) and ovule differentiation and female fertility in rice (Dreni et al. 2007). MADS-

box proteins function through heterodimerization with other MADS-box family members, and as predicted by homology, wild-type SHELL heterodimerizes with the rice SEPALLATA MADS-box protein, OsMADS24, in yeast two-hybrid assays (Singh, Low, et al., 2013). The sh^{MPOB} mutant allele, identified among descendants of the Nigerian *tenera* accession T128, is a T-to-C polymorphism that results in a leucine-to-proline amino acid change within the highly conserved MADS-box domain. The sh^{AVROS} mutant allele, identified within a multigenerational pedigree spanning five-decades segregating for the Congo-derived AVROS *pisifera* allele, is an A-to-T polymorphism resulting in a lysine-to-asparagine substitution, which is two amino acids carboxy-terminal to the sh^{MPOB} amino acid change (Singh et al. 2013). Both mutations occur within the α-helical structure, characteristic of all MADS-box domains, which is involved in both heterodimerization and DNA binding. Palms producing thick-shelled *dura* fruits are homozygous for the wild-type ($Sh^{DeliDura}$) nucleotide at each of the two variant nucleotide positions. Heterozygosity for either the sh^{MPOB} or sh^{AVROS} alleles ($Sh^{DeliDura}/sh^{MPOB}$ or $Sh^{DeliDura}/sh^{AVROS}$) results in palms producing thin-shelled *tenera* fruits. Palms homozygous for either mutation or heteroallelic for both mutations (a sh^{AVROS} mutation on one chromosome and a sh^{MPOB} mutation on the other) produce shell-less *pisifera* fruits. Cases of heteroallelic *pisifera* palms indicated that sh^{AVROS} and sh^{MPOB} alleles do not complement each other, thus confirming the identity of the *SHELL* gene (Singh et al. 2013).

The intention of the oil palm industry is to plant only high-yielding *tenera* hybrids in production fields. However non-*tenera* contamination can arise due to several reasons, including unintentional use of pollen from a non-*pisifera* palm, self-pollination of *dura* parental palms, open pollination of *dura* parental palms by surrounding *dura* palms, and imprecise selection of seeds or seedlings (Corley 2005). The phenotypic identification of non-*tenera* contaminant palms

is based on the observation of a cross-section of mature fruit. Since fruiting first occurs three to four years after field planting, the replacement of contaminants is not economical due to the large size of palms at this stage. With the discovery of *SHELL* and the mutations responsible for fruit form, DNA-based fruit form prediction is possible in nurseries prior to field planting.

We report the application of *SHELL* genetic testing in the assessment of non-*tenera* contamination rates in planting and nursery sites throughout Malaysia. Contamination rates are substantially higher than the <1% theoretically achievable through careful control of pollination (Corley 2005). DNA sequencing identified three novel mutant alleles of *SHELL,* each resulting in the *tenera* fruit form when heterozygous. Like the sh^{AVROS} and sh^{MPOB} mutations, each mutant allele is a missense mutation resulting in substitution of a conserved amino acid within the SHELL MADS-box domain, indicating that all five *SHELL* mutations share a common or closely related molecular mechanism impacting oil palm fruit form.

We also constructed a 48-parameter, 4-stage economic model to determine the losses incurred by the inadvertent cultivation of non-*tenera* palms in Malaysia. The findings suggest that comprehensive DNA-based screening and removal of non-*tenera* contaminants at the nursery will increase yield and result in substantial economic gains while optimizing use of existing planted area – a positive step towards sustainability.

C.2 Material and Methods

C.2.1 Sampling of Independent Planting Sites and Nurseries

Non-*tenera* contamination rates were assessed in representative independent planting sites surrounding the Malaysian Palm Oil Board's (MPOB's) 6 research stations, located in Peninsular Malaysia (Teluk Intan, Kluang, Keratong, Hulu Paka), Sarawak (Sessang) and Sabah (Lahad Datu). Non-*tenera* contamination rates were also assessed in three independent nursery

sites in each of seven sampling sub-regions encompassing the six geographical locations selected for the independent planters. The seven sampling sub-regions include Perak, Selangor, Negeri Sembilan, Melaka, Johor and Sarawak. In total, approximately 200 palms from each of 36 independent planting sites (n = 6,272) and from each of 21 nursery sites (n = 3,952) were randomly selected and genotyped by allele-specific PCR.

C.2.2 PCR-based Genetic Testing

In a total of 10,224 samples, genotypes at the sh^{AVROS} and sh^{MPOB} variant nucleotide positions were determined by SureSawit™ *SHELL* Kit (Orion Biosains, http://orionbiosains.com/products/suresawit shell/protocol/sybr-green-dye-detection). Approximately 20 ng of genomic DNA were used as template for each of four independent allele-specific PCR reactions: (i) exon 1 specific primers in which one primer overlapped the sh^{MPOB} variant nucleotide with the wild-type base and amplified only the wild-type allele at the variant nucleotide position, (ii) exon 1 specific primers in which one primer overlapped the sh^{MPOB} variant nucleotide with the sh^{MPOB} base and amplified only the sh^{MPOB} allele at the variant nucleotide position, (iii) exon 1 specific primers in which one primer overlapped the sh^{AVROS} variant nucleotide with the wild-type base and amplified only the wild type allele at the variant nucleotide position, and iv) exon 1 specific primers in which one primer overlapped the sh^{AVROS} variant nucleotide with the mutant base and amplified only the sh^{AVROS} allele at the variant nucleotide position. PCR wells were scored for amplification at end point by SYBR Green emission. Genotypes were scored based on the amplification pattern across the four reactions. Samples that are wild-type at both variant positions ($Sh^{DeliDura}/Sh^{DeliDura}$) amplify in reactions (i) and (iii) only. $Sh^{DeliDura}/sh^{MPOB}$ samples amplify in reactions (i), (ii) and (iii) only. $Sh^{DeliDura}/sh^{AVROS}$ samples amplify in reactions (i), (iii) and (iv) only. sh^{AVROS}/sh^{AVROS} samples

amplify in reactions (i) and (iv) only. sh^{MPOB}/sh^{MPOB} samples amplify in reactions (ii) and (iii) only. Finally, heteroallelic sh^{AVROS}/sh^{MPOB} samples amplify in all four reactions. Primer sequences and PCR conditions are provided in **Table C.1**.

C.2.3 DNA Sequencing

A total of 1,132 palm or seedling samples were genotyped by allele-specific PCR as wild-type at both the sh^{AVROS} and sh^{MPOB} variant nucleotide positions. Although these palms or seedlings would be predicted to be *dura* phenotype based on the genotypes at the sh^{AVROS} and sh^{MPOB} positions alone, they could, in fact, be *tenera* phenotype due to *SHELL* mutations independent of the sh^{AVROS} and sh^{MPOB} mutations. Therefore, exon 1 of *SHELL* was PCR amplified using flanking intronic primers (Singh et al. 2013) and Sanger sequenced. Primer sequences were confirmed to be unique in the reference *pisifera* genome and free of polymorphic nucleotides (Singh, Ong-Abdullah, et al., 2013). *SHELL* exon 1 was amplified from 20 ng of purified genomic DNA under standard PCR amplification conditions (Singh, Low, et al., 2013). Amplicons were treated with exonuclease I and shrimp alkaline phosphatase to remove unincorporated primers and deoxynucleotides. An aliquot of each amplicon was Sanger sequenced (ABI 3730) using an amplicon specific primer. All sequencing data was aligned to the reference *pisifera* genome sequence, and data were analyzed to determine the genotype at each exon 1 base. *SHELL* exons 2-7 were similarly Sanger sequenced in a subset of samples. Exons 3-6 were amplified using an intron 2 forward primer and an intron 6 reverse primer. Exons 2 and 7 were separately amplified using intronic primers flanking each exon. All primer sequences and PCR conditions are provided in **Table C.1**.

C.2.4 Genotype/Phenotype Comparisons

To determine concordance of fruit form phenotype with *SHELL* genotype, 512 previously unanalyzed samplings were phenotyped by visual inspection of shell thickness in cross-sectioned ripe fruits and genotyped by Sanger sequencing of *SHELL* exon 1, as described above. Phenotyping and genotyping were performed by different individuals. Individuals involved in phenotyping were blinded to genotyping calls, and individuals involved in genotyping were blinded to phenotyping calls. Genotypes and phenotypes were compared only after each set of calls was finalized. To address initial genotype/phenotype discordances, 15 discordant palms, as well as 15 palms that were concordant in the initial analysis, were sampled a second time for independent blinded phenotyping and genotyping as described above.

C.2.5 Economic Impact Modeling of Comprehensive *SHELL* Genetic Testing

To assess the economic impact of *SHELL* genetic testing on the Malaysian oil palm industry, Malaysian gross national income (GNI) and tax revenues of the Malaysian government, a 48-parameter, four-stage economic model that involves breeders, nurseries, planting sites and mills was constructed (**Table C.2**). In the model, two scenarios were compared – a baseline scenario representing the current state of the independent planters and a *SHELL* gene testing scenario, which forecasts changes in the production of crude palm oil (CPO), palm kernel oil (PKO), and palm kernel cake (PKC) in the deployment of comprehensive DNA testing and culling at the nursery stage. Contamination rates utilized for economic modeling were based on weighted averages of observed contamination in the surveyed independent planting sites, as these values represent the closest approximation of contamination currently in production fields. Observed contamination rates within each planting region were weighted according to the hectare size of planting area for each region, resulting in the *pisifera* (2.8%) and *dura* (8.1%) contamination rates utilized in the baseline scenario. Note that non-weighted contamination rates

are reported in **Table C.3**. Assuming that comprehensive *SHELL* gene testing would enable the molecular based identification and subsequent culling of all contaminant *pisifera* and *dura* palms at the nursery stage, we modelled 100% *tenera* rates in the selected sites in the *SHELL* gene testing scenario.

Monte Carlo simulation was used to mimic variation around 4 key fruit composition parameters including fruit to bunch, mesocarp to fruit, shell to fruit and kernel to fruit for *tenera* and *dura* palms, where 1,000 values were selected at random from normal distributions of each parameter for each fruit form for each year in the 60 -year simulation (**Table C.2**). For *pisifera* palms in the baseline scenario, we assumed these trees produced no fruit and were a complete loss.

We held planted area managed by independent planters at reported 2015 levels throughout the simulation (0.809M HA, representing 15% of the 5.39M HA of total planted area in Malaysia), and assumed an average planting density of 143 palms per hectare. Assumptions surrounding the age structure and productivity of palms in year 0 of the simulation are summarized in **Table C.2**, as are mill extraction efficiencies and other parameters.

The mass of CPO, PKO, and PKC for the baseline and *SHELL* testing scenarios were computed from the following mass balance equations:

(i) FFB (fresh fruit bunch) = EFB (empty fruit bunch) + PF (palm fruit);

(ii) PF (palm fruit) = Msc (mesocarp) + PK (kernel) + PKS (palm kernel shell);

(iii) Msc (mesocarp) = CPO (crude palm oil) + MF (mesocarp fruit fibre) + $POME_{MSC}$ (fraction of palm oil mill effluent from mesocarp);

(iv) PK (kernel) = PKO (palm kernel oil) + PKC (palm kernel cake) + WW_{PK} (fraction of palm kernel mill waste water from kernel).

Substituting terms of equation (i) with equations (ii), (iii) and (iv) yields equation (v):

(v) $FFB = EFB + CPO + MF + POME_{MSC} + PKO + PKC + WW_{PK} + PKS.$

To determine the selling price (P) of CPO and PKO at a future date (*i.e.*, future month A), a regression of monthly closing prices denominated in Malaysian Ringgit from January 1995 through April 2015 for CPO (http://www.indexmundi.com/commodities/?commodity=palm oil&months=360&currency=myr) and from January 1996 through April 2015 for PKO (http://www.indexmundi.com/commodities/?commodity=palm-kernel oil&months=360&currency=myr), was computed and the following exponential trend equations and effective inflation rates were determined and used in the model (historic prices for PKO denominated in Malaysian Ringgit were not available in 1995):

P_{CPO}(month A) = 1107.9 x e(0.0039 x B); an effective inflation rate of 4.79%

P_{PKO}(month A) = 1602.5 x e(0.0037 x B); an effective inflation rate of 4.54%

where B equals the number of months from January 1995 to Month A.

A price inflation rate of 4.79% (based on the effective inflation rate of CPO) and a bond rate of 3.55% were used for time value of money corrections, an economic multiplier of 1.511 (as determined for oil palm primary products in Malaysia) (Bekhet 2011) was used to convert industry economic gains into increases in gross national income, and a weighted average tax rate of 25% (the standard Malaysian corporate income tax rate since 2009) applied to increases in gross national income was used to compute increases in Malaysian government tax receipts. The costs of sampling and culling labor at the nursery, shipping sampled material to a centralized laboratory and the costs of *SHELL* allele testing were not included in the model.

C.3 Results

C.3.1 Non-*tenera* Contamination Trial Design

Adult palms and seedlings were selected among 6 planting sites and 3 nearby independent nurseries within multiple geographical regions throughout peninsular Malaysia, Sabah and Sarawak (**Table C.3**). The number of sampled palms per region ranged from 539 to 1,133, for a total of 10,224 independent palms. To ensure that the palms sampled from each site were representative, plants were randomly selected from several operational areas within each site. *SHELL* genotypes for two previously described mutations (sh^{AVROS} and sh^{MPOB}) were determined by allele-specific PCR assays designed to amplify either the *dura* or *pisifera* allele (Materials and Methods). In addition to sh^{AVROS} and sh^{MPOB} mutant alleles, Table 1 includes three novel *SHELL* mutations described below.

C.3.2 Non-*tenera* Contamination

Palms were classified as genetically non-*tenera* if the genotype was either homozygous wild type at both variant nucleotide positions and therefore *dura* fruit form, or homozygous or heteroallelic for sh^{AVROS} and/or sh^{MPOB} mutant alleles (sh^{AVROS}/sh^{AVROS}, sh^{MPOB}/sh^{MPOB} or sh^{AVROS}/sh^{MPOB}) and therefore *pisifera* fruit form. Palms that were homozygous or heteroallelic for any of three novel *SHELL* gene mutations, described below, were also classified as non *tenera.* Palms identified as genetically *dura* were confirmed by DNA sequencing to have no other *SHELL* gene mutations that could have caused the palm to be a *tenera* phenotype palm (Materials and Methods). The unadjusted non-*tenera* contamination across all palms sampled was 10.7% (95% confidence interval (CI) 10.1% - 11.9%) (**Table C.3**). The average contamination rates from each region were then weighted by each region's respective total oil palm planted area resulting in a national weighted average contamination rate in independent planting sites of 10.9% (2.8% *pisifera* and 8.1% *dura*).

Both *dura* and *pisifera* contamination were observed in each region with *dura* contamination exceeding that of *pisifera* in each case (**Table C.3**). This finding indicates that contamination is not completely explained by self or open pollination of *dura* palms by *dura* pollen (which would yield only *dura* seeds) or by unintentional pollination of *dura* palms by *tenera* pollen (which would yield *dura* or *tenera* seeds), as *pisifera* seeds could only be produced from *tenera* x *pisifera* or *tenera* x *tenera* crosses. Regional unweighted non-*tenera* contamination ranged from 2.5% (Region 4) to 23.1% (Region 7). There was significant variation in non-*tenera* contamination between different sites within a given region, with the exception of the two lowest overall contamination sites (Regions 3 and 4) (**Fig. C.1**). Furthermore, individual sites with very high contamination rates were identified, for example, site 7a with 63.5% (95% CI 56.5% - 70.1%) non-*tenera* palms and site 11b with 60.1% (95% CI 52.9% - 67.0%) non-*tenera* palms. Only 6 of 57 sites had no detectible non-*tenera* palms (**Fig. C.1**). Palms of all age groups sampled (from less than 5 years to greater than 10 years after planting) showed non-*tenera* contamination, indicating that current controlled pollination practices have not overcome the problem.

C.3.2 Identification of Novel Mutant Alleles of *SHELL*

We previously identified two independent but closely related mutations within the *SHELL* MADS-box domain (sh^{AVROS} and sh^{MPOB}) that each results in the *tenera* fruit form phenotype when heterozygous and the *pisifera* fruit form phenotype when homozygous or heteroallelic with each other (Singh et al. 2013). However, recognizing that additional mutant alleles of *SHELL* could exist in commercial material, exon 1 of the *SHELL* gene (encoding the entire MADS-box domain) was sequenced in each of the DNA samples with a wild-type genotype at both the sh^{AVROS} and sh^{MPOB} nucleotide variant positions (Materials and Methods).

Among 1,132 palms genotyped as homozygous wild-type at the sh^{AVROS} and sh^{MPOB} nucleotide positions, 235 (20.8%) were heterozygous for one of three novel non-synonymous nucleotide variants within the MADS-box coding region (**Table C.4**, **Fig. C.2a**), and no palms were found to be homozygous for these variants. Two variants (sh^{MPOB2} (lysine-to-glutamine) and sh^{MPOB4} (lysine-to-asparagine)) result in the substitution of the same conserved lysine six residues amino-terminal to the sh^{MPOB} leucine-to-proline position. The sh^{MPOB3} variant results in an alanine-to-aspartate substitution 10 residues carboxy-terminal to the sh^{AVROS} lysine-to-asparagine position (**Fig. C.2**). The three novel mutant *SHELL* alleles varied in frequency. The sh^{MPOB2} variant was detected in 1.4% of the 10,224 palms analyzed (**Table C.4**), and was present in 7 of the 13 geographic regions at rates ranging from 0.1% (Region 1) to 6.4% (Region 10). The sh^{MPOB3} variant was detected in 7 of 13 regions, at a rate ranging from 0.6% (Region 3) to 2.1% (Region 1), representing 0.8% of all palms analyzed. Finally, the sh^{MPOB4} variant was detected in only two palms (in Region 1 and Region 3), representing 0.02% of all palms analyzed. Given the conservation of these residues within the MADS-box domain motif (**Fig. C.2b**) and the proximity to the mutations previously shown to determine fruit form phenotype (Singh, Low, et al., 2013), we hypothesized that these variants represent novel mutant alleles of *SHELL* that also control oil palm shell fruit form. Note that the non-*tenera* contamination rates described in **Table C.3** and **Fig. C.1** were calculated based on all five *SHELL* gene variants. All seven exons of *SHELL* were sequenced in a panel of 112 palms that were homozygous wild-type at all five variant nucleotide positions. While three synonymous single nucleotide polymorphisms (SNP) were identified within exon 7, these would not be predicted to impact fruit form as they code for the wild-type amino acid. One non-synonymous SNP resulting in a

conservative glutamate-to-aspartate substitution 12 amino acids from the SHELL carboxy terminus was identified.

To demonstrate that the three novel mutant *SHELL* alleles conferred the shell fruit form trait, adult oil palms from populations in which the alleles were detected were sampled and scored for fruit form. In total, 512 sampled palms with mature fruit bunches were visually phenotyped and DNA sequenced (Materials and Methods). In an initial genotype-by-phenotype comparison, 97.1% of genotypes matched the predicted phenotype, assuming that all 5 mutant alleles result in *tenera* phenotype when heterozygous and *pisifera* phenotype when homozygous. A 2.9% discordance rate is within the accuracy norms of visual phenotyping in plantation settings (Singh et al. 2013). The 15 discordant palms included 9 $Sh^{DeliDura}/sh^{AVROS}$ palms and 1 $Sh^{DeliDura}/sh^{MPOB2}$ palm phenotyped as *dura*, 2 sh^{AVROS}/sh^{AVROS} palms and 1 $Sh^{DeliDura}/ Sh^{DeliDura}$ palm phenotyped as *tenera* and 2 $Sh^{DeliDura}/sh^{AVROS}$ palms phenotyped as *pisifera*. There were no discordances in which a genotypically *pisifera* palm was phenotyped as *dura*, or *vice versa,* as these phenotypes are the most easily discriminated by visual inspection of shell thickness. However, in addition to a low rate of visual phenotyping inaccuracy, discordances could arise due to inaccuracies of genotyping as well as due to potential sample collection and processing errors that may arise, especially under conditions where hundreds to thousands of samples are collected in field settings. To address this, the 15 discordant palms, as well as 15 palms that were concordant in the initial analysis, were sampled a second time for independent blinded phenotyping and genotyping (Materials and Methods). In this independent round of sampling, no phenotype or genotype changes occurred among the 15 originally concordant palms, and therefore these remained genotype-by-phenotype concordant. However, among the 15 originally discordant palms, all 15 were either scored as a different phenotype than in the first sampling (6

palms) or genotyping of the repeated sampling yielded a different genotype than the first sampling (9 palms), and these changes resolved each of the 15 discordances. These results suggest that the very low rate of discordances was due to a combination of phenotyping and sample processing inaccuracies. First, all genotyping in this genotype-by-phenotype comparison was performed by Sanger DNA sequencing, which has a very low error rate. Second, if genotyping inaccuracies were the cause of discordances, then genotype calls should have changed at a similar rate among the initially concordant and discordant palm sets. However, 60% of initial discordant palms changed genotype upon second sampling compared to no palms changing genotype within the initially concordant set ($p<0.0003$, chi squared test). To further support this conclusion, we performed three independent PCR amplification and Sanger sequencing assays on each of the 30 original sampling DNA preps, as well as each of the 30 resampled DNA preps. In every case (180 independent assays), genotype calls were consistent between replicated assays of the same sample and agreed with the original genotype call of the same prep, thus demonstrating the accuracy of genotype calls.

After resolution of rare discordances, the five *SHELL* MADS box domain mutations accounted for 100% of observed *tenera* and *pisifera* phenotypes (**Table C.5**, **Fig. C.3**). Among the 422 phenotypically *tenera* palms, 385 were heterozygous $sh^{AVROS}/Sh^{DeliDura}$, 2 were heterozygous $sh^{MPOB}/Sh^{DeliDura}$, 2 were heterozygous $sh^{MPOB2}/Sh^{DeliDura}$, 31 were heterozygous $sh^{MPOB3}/Sh^{DeliDura}$ and 2 were heterozygous $sh^{MPOB4}/Sh^{DeliDura}$. All 56 phenotypically *dura* palms were wild-type at each of the five variant positions. Finally, among 34 phenotypically *pisifera* palms, 33 were homozygous sh^{AVROS}/sh^{AVROS} and one was heteroallelic with the sh^{MPOB2} allele on one chromosome and the sh^{AVROS} allele on the other. This heteroallelic *pisifera* palm suggests that sh^{MPOB2} does not complement the sh^{AVROS} allele, thus confirming that the sh^{MPOB2} mutation

(and, likely, the sh^{MPOB4} mutation, as it involves substitution of the same amino acid) functionally determines oil palm shell fruit form. Furthermore, the 100% concordance between heterozygosity for sh^{MPOB3} and *tenera* phenotype strongly supports the conclusion that this allele is functionally equivalent to the other four mutant alleles. The finding that all *tenera* (n = 442) and all *pisifera* (n = 34) phenotype palms could be explained by heterozygous (*tenera*) or homozygous/heteroallelic (*pisifera*) mutations involving the five *SHELL* MADS box domain mutations indicates that the single conservative missense amino acid substitution detected in exon 7 likely does not impact shell fruit form phenotype. While additional mutant alleles impacting fruit form may be present in other oil palm populations, the data confirm that the three novel *SHELL* alleles affect fruit form and suggest that these five mutations are responsible for at least the vast majority of *tenera* and *pisifera* phenotypes within this wide sampling of planting sites and nurseries.

C.3.4 Economic Impact Modeling of Comprehensive *SHELL* Genetic Testing

Nursery stage DNA testing of the estimated 3.9 million palms planted at independent planting sites each year would enable nursery operators to cull contaminant palms before they are field planted. To assess the economic impact of *SHELL* genetic testing in the independent sector, a 48-parameter, four-stage economic model that involves breeders, nurseries, planting sites and mills was constructed (**Table C.2**, Materials and Methods). The monthly closing prices for crude palm oil (CPO) and palm kernel oil (PKO) and the fit of the regression curves for CPO and PKO are plotted in **Figs. C.4a** and **Fig. C.4b**, respectively, where the regression equations were able to explain 48.43% and 41.68% of the historical pricing data, and regression predictions were used for future CPO and PKO pricing in the model (Materials and Methods). The economic model determined two curves, which compute the annual gains at steady state excluding

sampling, testing and culling costs i) to the oil palm industry, ii) to Malaysian gross national income (GNI), and iii) to the Malaysian government in the form of increased tax revenues. The gains are determined as a function of the weighted average national independent planting site *pisifera* (**Fig. C.4c**) and *dura* (**Fig. C.4d**) contamination prevented in the *SHELL* gene screening scenario (Materials and Methods). The *SHELL* gene-screening scenario would increase GNI, industry income, and government tax income as low yielding contaminant palms are replaced by high yielding *tenera* palms. At steady state, *SHELL* gene DNA testing in the independent sector alone would add Malaysian ringgit (RM) 1.05 billion to Malaysian GNI annually by comprehensive screening (or RM 272 in gains per screened palm - including all *tenera* and non-*tenera* palms tested), RM 0.693 BN of increased production annually to oil palm industry members (or RM 180 per screened palm) and RM 0.26 BN of increased tax receipts annually (or RM 68 in new taxes per screened palm) (**Fig. C.5**). Using the average RM to USD exchange rate over the past 20 years of 3.5, *SHELL* gene DNA testing would add $300M USD to Malaysian GNI annually (or $77.71 USD per screened palm), $198M USD of increased production annually to oil palm industry members (or $51.42 USD per screened palm), and $74.28M USD of increased tax receipts annually (or $19.43 USD per screened palm) (**Fig. C.5**).

C.4 Discussion

Since the publication of the *E. guineensis* and *E. oleifera* reference genome sequences in 2013 (Singh et al. 2013), several critical milestones toward oil palm industry sustainability have been achieved. These advances include the identifications of the *SHELL* gene (Singh et al. 2013), the *VIRESCENS* gene and mutations responsible for oil palm fruit colour (Singh et al. 2014), as well as the *MANTLED* gene and the epigenetic abnormality responsible for the somaclonal mantling phenotype (Ong-Abdullah et al. 2015). These discoveries have introduced strategies for

genetic or epigenetic testing impacting sustainability by ensuring that land devoted to oil palm production is optimally utilized. The present study will enhance the sensitivity of DNA-based screening for non-*tenera* contamination through the addition of three novel mutant alleles of *SHELL* to the two previously identified alleles (Singh et al. 2013). These five mutations, clustering within a highly conserved 19 amino acid span of the MADS box domain, were sufficient to account for all *tenera* and *pisifera* phenotype palms within the diverse geography surveyed.

These findings, for the first time, also enabled a direct assessment of non-*tenera* contamination in independent planting sites and nurseries across Malaysia. Within these populations, contamination rates were substantially higher than the <1% contamination rate theoretically achievable through careful control of pollination (Corley 2005). Although contamination rates varied significantly both between surveyed regions (ranging from 2.5% to 30.0%) and between specific sites within those regions, no region was found to be free of contamination. Overall, the measured non-*tenera* contamination rate across all regions/sites on a weighted average basis was 10.9% (8.1% *dura* and 2.8% *pisifera*).

Given directly measured contamination rates among independent planting sites and nurseries, it is possible to model the real-world economic impact of the implementation of comprehensive *SHELL* genetic testing at the nursery stage in this sector, which would effectively eliminate the inadvertent planting of non-*tenera* materials. Economic modeling predicts that the comprehensive genetic testing to prevent cultivation of non-*tenera* palms in the independent sector alone would add RM 1.05 BN ($300M USD) to Malaysian gross national income annually, which represents RM 272 ($77.71 USD) in gains per screened palm. Screening would contribute RM 0.693 BN ($198M USD) to oil palm industry members annually representing RM

180 ($51.42 USD) per screened palm. Finally, *SHELL* gene testing would increase the Malaysian tax revenues annually by RM 0.26 BN ($74.28M USD) or RM 68 ($19.43 USD) for each palm screened. Therefore, *SHELL* genetic testing would have a major positive economic impact while improving oil palm sustainability by optimizing the utilization of existing planted area.

Although over 10,000 palms and seedlings derived from multiple sites within each of 13 geographic regions were analyzed in the present study, it is noted that this survey of 36 independent planting sites and 21 nurseries represents only a small sampling of the over 200,000 independent planting sites throughout Malaysia. The findings reveal that non-*tenera* contamination is likely a concern throughout the entire palm industry, and they demonstrate that comprehensive genetic testing will have major positive environmental and economic impacts. However, substantially larger surveys of contamination, utilizing now available *SHELL* genetic testing, will be a critical area of future oil palm research.

C.5 Acknowledgements

The authors would like to thank MPOB for permission to publish this manuscript. We would also like to thank the staff of MPOB's Breeding and Tissue Culture Unit at the various research stations for their assistance in the sampling process. We also extend our appreciation to the Integration Research & Extension Division as well as the Licensing & Enforcement Division, especially Ms. Nur Hanani Mansor, Mr. Moktarudin Hj. Ahmad and Mr. Azhar Ahmad for their assistance in conducting this study. We thank Rob Martienssen for critical advice and guidance and for helpful comments on preparation of the manuscript.

C.5.1 Authors Contribution

E.L., M.O.A., R.N., N.L., J.O., R.Sa. and R.Si. conceptualized the research programme. L.O., E.L., M.O.A., R.N., M.M., W.O., A.M., N.L., J.O., Y.C., R.Sa. and .R.Si. designed the experiments and coordinated the project. L.O., N.T., M.A.B., A.V.B., M.B., M.L., Y.C. and R.Si. conducted laboratory experiments. L.O., E.L., M.O.A., N.L., J.O., M.A.B., A.V.B., N.J., M.B., M.L., S.W.S., R.Sa. and R.Si. assisted in data analysis. E.L., J.N., K.C., M.H., N.A. and S.W.S. performed bioinformatics analysis. L.O., T.F., A.K. and S.B. coordinated collection of samples. R.N., N.L., A.H.O., A.N., H.C. and S.A.S. participated in the economic analysis. L.O., E.L., M.O.A., R.N., N.L., J.O., R.Sa. and R.Si. prepared and revised the manuscript. All authors approved the final manuscript.

C.5.2 Conflict of Interest Statement

A.K. and S.B. are employees of Orion Biosains, a biotechnology company located in Puchong, Selangor, Malaysia that offers *SHELL* genetic testing to the oil palm industry under a royalty bearing license from the Malaysian Palm Oil Board. Orion Biosains is an affiliate of Orion Genomics, LLC based in St. Louis, MO. N.L., J.O. and T.F. are directors of Orion Biosains. N.L., J.O., T.F., M.A.B., A.V.B., M.B., M.L., N.J., S.W.S. and C.B. are employees and stock holders of Orion Genomics, LLC. L.O., E.L., M.O.A., R.N., N.T., J.N., M.M., K.C., M.H., N.A., W.O., A.M., Y.C., R.Sa. and R.Si. are employees of the Malaysian Palm Oil Board. L.O., E.L., M.O.A., R.N., S.W.S., N.L., J.O., A.V.B., M.A.B., R.Sa. and R.Si. are co-inventors on pending patents covering *SHELL* genetic testing. All other authors declare no competing interests relevant to this work.

C.6 Reference

Beirnaert, A., and Vanderweyen, R. (1941). "Contribution a l'etude genetique et biometrique des varieties d'Elaeis guineensis Jacquin." *Publications Institut National Pour L'etude Agronomique Congo Belge Serie Science.* **27**: 1-101.

Bekhet, H.A. (2011). "Output, income and employment multipliers in Malaysian Economy: Input-output approach." *International Business Research.* **4**(1): 208-223.

Cochard, B., Amblard, P., and Durand-Gasselin, T. (2005). "Oil palm genetic improvement and sustainable development." *Oleagineux, Corps Gras, Lipides.* **12**: 141-147.

Corley, R. H. V. (2005). "Illegitimacy in oil palm breeding - A review." *Journal of Oil Palm Research.* **17**: 64-69.

Dinneny, J. R., and Yanofsky, M. F. (2005). "Drawing lines and borders: how the dehiscent fruit of Arabidopsis is patterned." *Bioessays.* **27**(1): 42-49.

Dreni, L., Jacchia, S., Fornara, F., Fornari, M., Ouwerkerk, P. B., An, G., *et al.* (2007). "The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice." *Plant Journal.* **52**(4): 690-699.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., *et al*. (2003). "MADS-box protein complexes control carpel and ovule development in Arabidopsis." *Plant Cell.* **15**(11): 2603-2611.

Hartley, C. 1988. "*The Oil Palm*: Longman."

Ong-Abdullah, M., Ordway, J. M., Jiang, N., Ooi, S. E., Kok, S. Y., Sarpan, N., *et al*. (2015). "Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm." *Nature.* **525**(7570): 533-537.

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., *et al.* (2003). "Assessing the redundancy of MADS-box genes during carpel and ovule development." *Nature.* **424**(6944): 85-88.

Rajanaidu, N., Kushairi, A., Rafii, M., Mohd Din, A., Maizura, I., and Jalani, B. S. (2000). "Advances in oil palm research." Bangi, Selangor: Malaysian Palm Oil Board.

Ritter, E., Lopez de Armentia, E., Erika, P., Herrero, J., Niggrum, Y. P., Santika, B., *et al.* (2015). "Development of a molecular marker system to distinguish shell thinkness in oil palm genotypes." *Euphytica.* **207**: 367-376.

Sharma, M., and Tan, Y. P. (1999). "Oil palm breeding programmes and the performance of DxP planting materials at United Plantations Berhad." *Proceeding Seminar on Sourcing of Oil Palm Planting Materials for Local and Overseas Joint Venture*. Malaysia: Palm Oil Research Institute.

Singh, R., Low, E. T., Ooi, L. C., Ong-Abdullah, M., Nookiah, R., Ting, N. C., *et al.* (2014). "The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB." *Nature Communications.* **5**(4106): 1-8.

Singh, R., Low, E. T., L. Ooi, L. C., Ong-Abdullah, M., Ting, N. C., Nagappan, J., *et al.* (2013). "The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK." *Nature.* **500**(7462): 340-344.

Singh, R., Ong-Abdullah, M., Low, E. T., Manaf, M. A., Rosli, R., Nookiah, R., *et al.* (2013). "Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds." *Nature.* **500**(7462): 335-339.

Zeven, A. C. (1965). "The origin of the oil palm." *Journal of the Nigerian Institute of Oil Palm Research.* **4**: 218-225.

Allele-specific genotyping primers			
Assay ID (as described in text)	Amplified Allele	Allele-specific primer[a]	Second primer
i	Wildtype allele at *sh*^{MPOB} variant nucleotide	GCAAACGCCGAAATGGACTGA*T	GTTGCCCTAGTCAAGGTAAAACTC*T
ii	Mutant allele at *sh*^{MPOB} variant nucleotide	GCAAACGCCGAAATGGACTTT*C	GTTGCCCTAGTCAAGGTAAAACTC*T
iii	Wildtype allele at *sh*^{AVROS} variant nucleotide	CACAAAGGACAGACAACTCATAAGG*T	GAGAATCGAGATTTGAGCCG*T
iv	Mutant allele at *sh*^{AVROS} variant nucleotide	CAAAGGACAGACAACTCATAAGT*A	GAGAATCGAGATTTGAGCCG*T
Amplification primers for Sanger DNA Sequencing			
Assay ID (as described in text)	Forward primer	Reverse primer	Sequencing Primer
Exon 1	tcagaatttaaagaaactaaacttcacgta	ggatcagggataaaagggaag	gatctatggaaattaataagtcaatatgtcagtatgtg
Exon 2	gaaaattatatggttagaaaatgttactaaaggt	aggactccgacatagtcttaatagttagtagta	gaaaattatatggttagaaaatgttactaaaggt
Exons 3-6	tgaatcatcaaatttattacattttagcct	tgagaaaatgcattttaaacataagtcac	tgaatcatcaaatttattacattttagcct
Exon 7	tgctgatttctctatttatcgatgc	gatctactcaatagatgatggttggtg	tgctgatttctctatttatcgatgc
PCR conditions (allele specific assays and amplification for Sanger DNA sequencing)			
Step	Condition		
1	94^{o}C, 2 minutes		
2			

	94°C, 30 seconds		
3	58°C, 1 minute	Steps 2-4 repeated 34 cycles	
4	72°C, 1 minute		
5	72°C, 5 minutes		
* indicates phosphorothionate bond between 3'N and 3'N-1 oligonucleotide positions.			
[a] Mismatched bases outside the variant nucleotide position were intentionally included to increase specificity of allele specific primers.			

Table C.1: PCR primers and conditions.

Definition		Value
Contamination Rates:		
Initial Fraction of Tenera		89.10%
Initial Fraction of Dura		8.10%
Initial Fraction of Pisifera		2.80%
Final Fraction of Tenera		100.00%
Final Fraction of Dura		0.00%
Final Fraction of Pisifera		0.00%
Plantation:		
Number of trees per hectare		143
Total Planted Area (in HA)		5,392,235
Fraction Cultivated by Ind. Smallholders		15.00%
Total Planted Area of Ind. Smallholders (in HA)		808,835
Age Structure and Productivity:		
Initial Fraction of immature palms (age <4)		13.00%
Initial Fraction of young palms (age 5-8)		17.00%
Initial Fraction of prime palms (age 9-18)		35.00%
Initial Fraction of aging palms (age 19-24)		20.00%
Initial Fraction of old palms (age 25-30)		15.00%
Productivity - young palms (MT FFB/HA/YR)(D&T)		15
Productivity - prime palms (MT FFB/HA/YR)(D&T)		23
Productivity - aging palms (MT FFB/HA/YR)(D&T)		17
Productivity - old palms (MT FFB/HA/YR)(D&T)		15
Milling:		
CPO mill extraction efficiency	ECPO	86.00%
PKO mill extraction efficiency	EPKO	97.00%
Valuation:		

Initial CPO Price (in RM)	PCPO	2,869
Initial PKO Price (in RM)	PPKO	3,781
Initial PKC Price (in RM)	PPKC	370
CPO Price Inflation Rate	RI	4.79%
Bond Rate for Malaysia	RB	3.55%
MY Corporate Income Tax Rate	Tax	25.00%
MY Econ Multiplier (Oil Palm Primary Products)	M	1.511
Fruit Composition (Mass Ratios):		
Tenera: fruit to bunch (Avg.)	PF/FFB(T)(Avg.)	0.649
Tenera: fruit to bunch (Std. Dev.)	PF/FFB(T)(Std. Dev.)	0.041
Dura: fruit to bunch (Avg.)	PF/FFB(D)(Avg.)	0.612
Dura: fruit to bunch (Std. Dev.)	PF/FFB(D)(Std. Dev.)	0.038
Tenera: mesocarp to fruit (Avg.)	Msc/PF(T)(Avg.)	0.865
Tenera: mesocarp to fruit (Std. Dev.)	Msc/PF(T)(Std. Dev.)	0.027
Dura: mesocarp to fruit (Avg.)	Msc/PF(D)(Avg.)	0.612
Dura: mesocarp to fruit (Std. Dev.)	Msc/PF(D)(Std. Dev.)	0.026
Tenera: shell to fruit (Avg.)	PKS/PF(T)(Avg.)	0.07
Tenera: shell to fruit (Std. Dev.)	PKS/PF(T)(Std. Dev.)	0.008
Dura: shell to fruit (Avg.)	KS/PF(D)(Avg.)	0.306
Dura: shell to fruit (Std. Dev.)	PKS/PF(D)(Std. Dev.)	0.026
Tenera: kernel to fruit (Avg.)	PK/PF(T)(Avg.)	0.065
Tenera: kernel to fruit (Std. Dev.)	PK/PF(T)(Std. Dev.)	0.005
Dura: kernel to fruit (Avg.)	PK/PF(D)(Avg.)	0.082
Dura: kernel to fruit (Std. Dev.)	PK/PF(D)(Std. Dev.)	0.01
Tenera: CPO content of mesocarp (Avg.)	CPO/Msc(T)(Avg.)	0.55
Dura: CPO content of mesocarp (Avg.)	CPO/Msc(D)(Avg.)	0.55
Tenera: PKO content of kernel (Avg.)	CPO/Msc(T)(Avg.)	0.47

Dura: PKO content of kernel (Avg.)	CPO/Msc(D)(Avg.)	0.47

Abbreviations: MT, million tonnes; FFB, fresh fruit bunch; HA, hectare; T, tenera; D, dura; CPO, crude palm oil; PKO, palm kernel oil; MY, Malaysia; PF, palm fruit; Msc, mesocarp; PKS, palm kernel shell; PK, palm kernel; PKC, palm kernel cake; RM, Malaysian Ringgit

Table C.2: Parameters for Economic Impact Model.

Region[a]	No. sites surveyed	Palms genotyped	*tenera*[b] (%)	*dura*[c] (%)	*pisifera*[d] (%)	Unadjusted non-*tenera* contamination[e]
1	6	1,132	88.9	7.4	3.7	11.1
2	6	1,133	89.1	9.7	1.1	10.9
3	6	882	96.4	2.8	0.8	3.6
4	6	1,092	97.5	2.3	0.2	2.5
5	3	539	92.8	7.1	0.2	7.2
6	3	571	79.7	20.1	0.2	20.3
7	3	576	76.9	22.7	0.3	23.1
8	3	571	84.2	15.6	0.2	15.8
9	3	550	95.8	3.8	0.4	4.2
10	6	921	91.6	8.1	0.2	8.4
11	6	1,112	81.0	11.8	7.2	19.0
12	3	572	88.5	11.4	0.2	11.5
13	3	573	94.1	5.6	0.3	5.9
Total	57	10,224	89.3	9.2	1.5	10.7

[a] Regions within peninsular Malaysia (1-9) and Sabah & Sarawak (10-13) were surveyed.

[b] *Tenera* genotypes include $Sh^{DeliDura}/sh^{AVROS}$, $Sh^{DeliDura}/sh^{MPOB}$, $Sh^{DeliDura}/sh^{MPOB2}$, $Sh^{DeliDura}/sh^{MPOB3}$ or $Sh^{DeliDura}/sh^{MPOB4}$.

[c] *Dura* genotypes are wild-type at all five variant nucleotide positions.

[d] *Pisifera* genotypes include sh^{AVROS}/sh^{AVROS}, sh^{MPOB}/sh^{MPOB} or sh^{AVROS}/sh^{MPOB2}.

[e] Note that national weighted average contamination rate adjusted for the regional planting area is 10.9%.

Table C.3: Non-*tenera* contamination by region.

	Potential *dura* contaminants[a]	*sh*^{MPOB2}	*sh*^{MPOB3}	*sh*^{MPOB4}
Region 1	110	1 (0.1%)[b]	24 (2.1%)	1 (0.1%)
Region 2	126	-	16 (1.4%)	-
Region 3	32	-	5 (0.6%)	1 (0.1%)
Region 4	36	-	11 (1.0%)	-
Region 5	37	1 (0.2%)	-	-
Region 6	114	18 (3.2%)	-	-
Region 7	129	12 (2.1%)	2 (0.3%)	-
Region 8	89	5 (0.9%)	-	-
Region 9	20	-	-	-
Region 10	145	59 (6.4%)	11 (1.2%)	-
Region 11	199	52 (4.7%)	16 (1.4%)	-
Region 12	65	-	-	-
Region 13	30	-	-	-
Total	1,132	148 (1.4%)	85 (0.8%)	2 (0.02%)

[a] Number of palms genotyped as wild-type at both the *sh*^{AVROS} and *sh*^{MPOB} variant nucleotide positions.

[b] Integer value represents number of palms per site heterozygous for the indicated nucleotide variant, and parentheses denote the percentage of the indicated heterozygous palms among all palms analyzed per site.

Table C.4: Novel *SHELL* alleles.

Predicted Phenotypes[a]	Genotypes	Observed Phenotypes[b]		
		tenera	*pisifera*	*dura*
tenera	$sh^{AVROS}/Sh^{DeliDura}$	385	-	-
	$sh^{MPOB}/Sh^{DeliDura}$	2	-	-
	$sh^{MPOB2}/Sh^{DeliDura}$	2	-	-
	$sh^{MPOB3}/Sh^{DeliDura}$	31	-	-
	$sh^{MPOB4}/Sh^{DeliDura}$	2	-	-
pisifera	sh^{AVROS}/sh^{AVROS}	-	33	-
	sh^{AVROS}/sh^{MPOB2}	-	1	-
dura	$Sh^{DeliDura}/Sh^{DeliDura}$	-	-	56

[a] Fruit form phenotypes predicted by the indicated genotype.

[b] Fruit form phenotypes determined by visual inspection.

Table C.5: Concordance of *SHELL* genotype and fruit form phenotype.

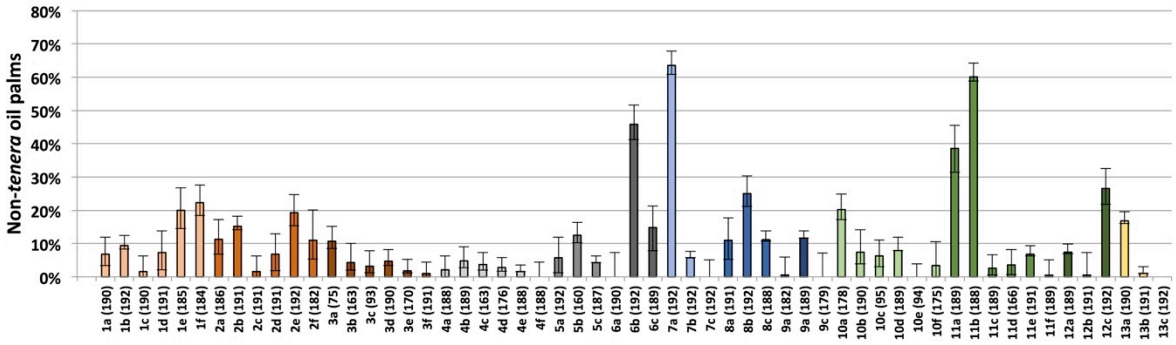


Figure C.1: Non-*tenera* contamination rates within each individual sampled site.

Palms and nursery seedlings genotyped as either wild-type (*dura*), homozygous mutant for any of the five identified *SHELL* gene MADS box mutation (*pisifera*) or heteroallelic for any two of the five mutants (*pisifera*) were classified as non-*tenera*. On the x-axis, numbers 1-13 indicate sampled regions according to Table C.3. Letters represent each independent site sampled within the indicated region. Numbers in parentheses indicate the number of palms or seedlings genotyped per region/site. Error bars indicate 95% confidence intervals.

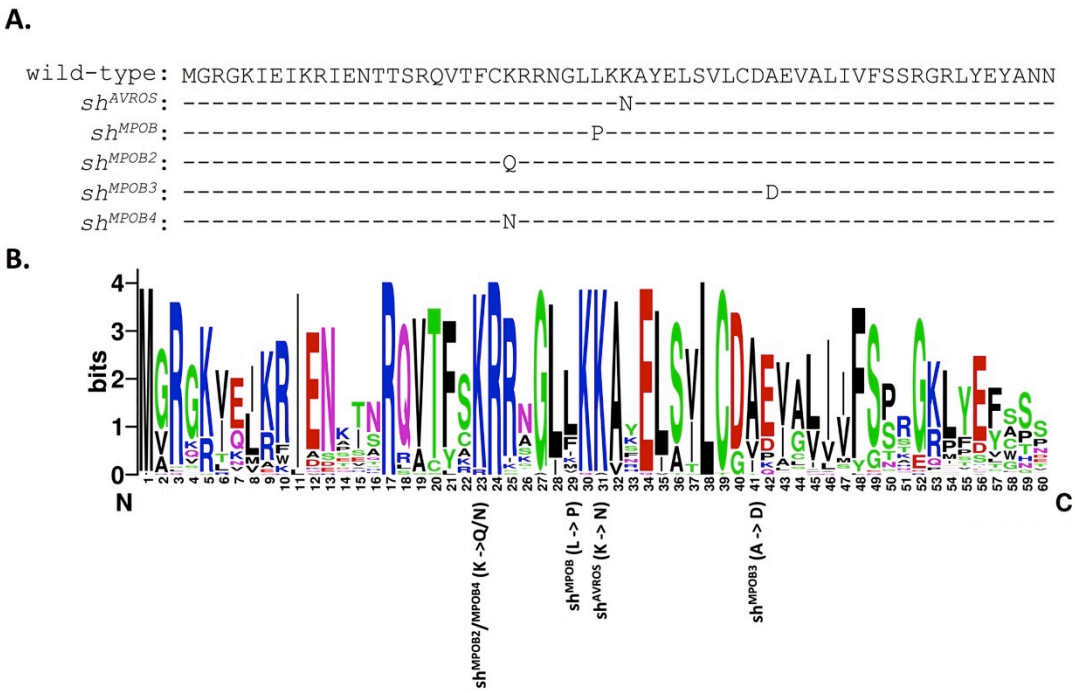


Figure C.2: *SHELL* MADS box domain mutations associated with fruit form phenotype.

a The amino acid sequence of the wild-type (*dura*) SHELL MADS box domain is shown at the top. Amino acid substitutions encoded by the sh^{AVROS}, sh^{MPOB}, sh^{MPOB2}, sh^{MPOB3} and sh^{MPOB4} mutations are each indicated below the wild-type amino acid sequence. **b** Logo plot of amino acid conservation among 51 independent oil palm MADS box gene predictions within the oil palm reference genome sequence. The positions of substituted amino acids encoded by the five *SHELL* mutations are indicated below the plot. The sh^{MPOB2} (K-to-Q) and sh^{MPOB4} (K-to-N) mutations alter the same amino acid (position 23 in the Logo plot). The sh^{MPOB}, sh^{AVROS}, and sh^{MPOB3} mutations alter amino acid positions 29, 31 and 41 in the Logo plot, respectively. Each mutation results in alteration of the most prevalent amino acid to an amino acid not utilized in any oil palm MADS box gene prediction at that position.

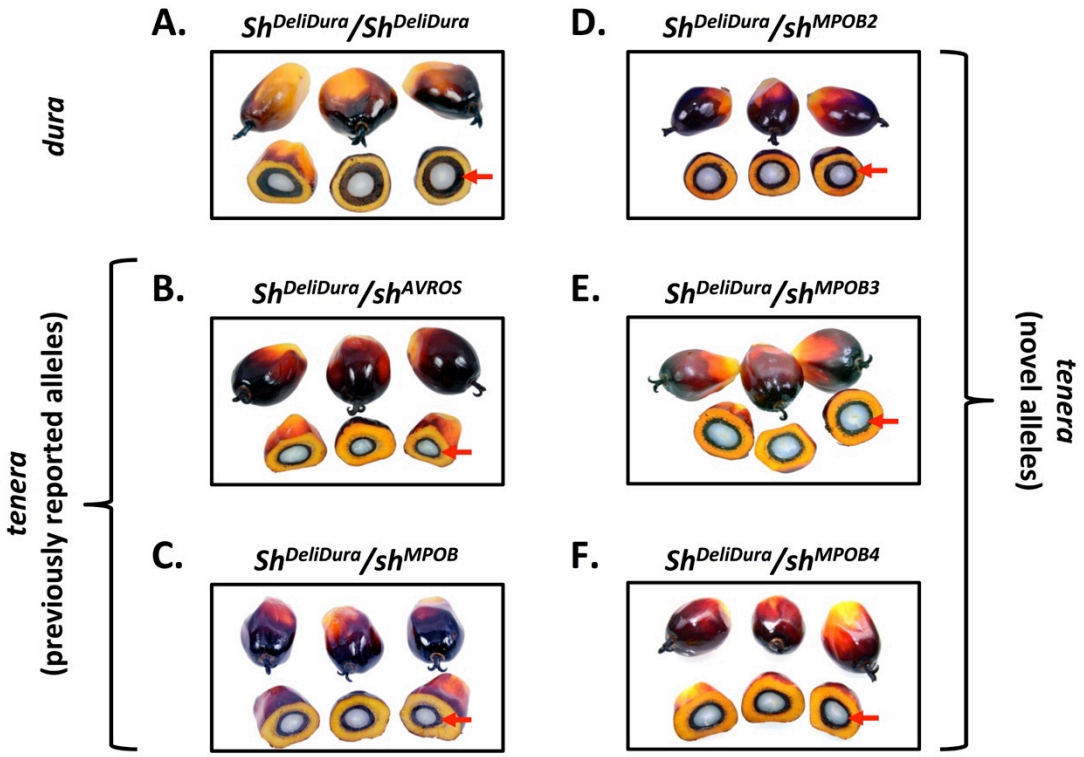


Figure C.3: Fruit form phenotypes.

Photographs of three whole fruits (top) and three cross-sectioned fruits (bottom) from a wild-type *dura* palm **a** and from a $Sh^{DeliDura}/sh^{AVROS}$ **b**, $Sh^{DeliDura}/sh^{MPOB}$ **c**, $Sh^{DeliDura}/sh^{MPOB2}$ **d**, $Sh^{DeliDura}/sh^{MPOB3}$ **e** or $Sh^{DeliDura}/sh^{MPOB4}$ **f** *tenera* palm are shown. Arrows indicate the thick shell of *dura* fruit **a** and the thin shell of *tenera* fruits from palms with different heterozygous mutant alleles **b-e**.

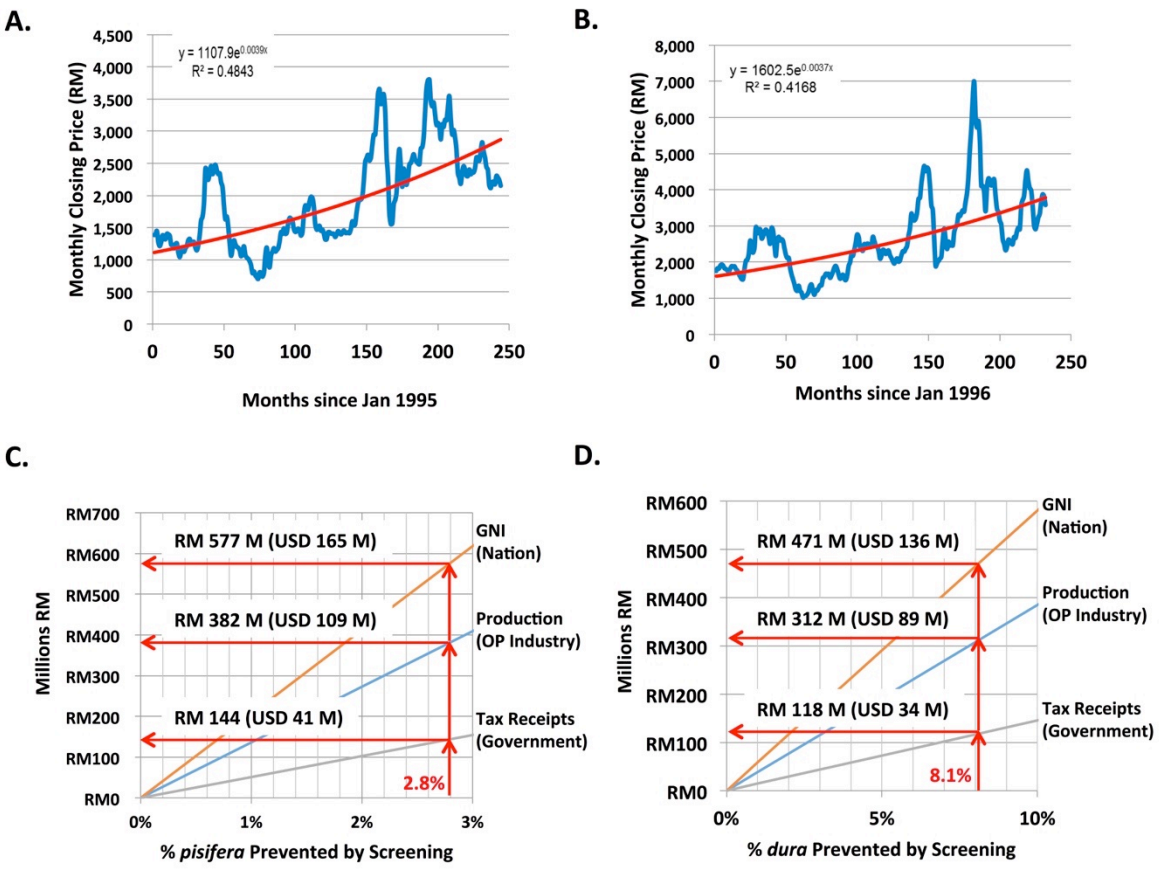


Figure C.4: Economic impact modeling of comprehensive *SHELL* genetic testing.

a-b To determine selling price of CPO **a** and PKO **b** at a future date, monthly closing prices (in Malaysian Ringgit, RM) are plotted from January 1995 through April 2015 **a** and from January 1996 through April 2015 (historic prices for PKO denominated in Malaysian Ringgit were not available in 1995) **b**. A regression of monthly closing prices was computed as described in Materials and Methods. **c-d** Annual economic impact of preventing *pisifera* **c** and *dura* **d** contamination in independent planting sites at steady state. RM to U.S. dollar (USD) conversions are based on an average exchange rate over the past 20 years of 3.5RM/USD.

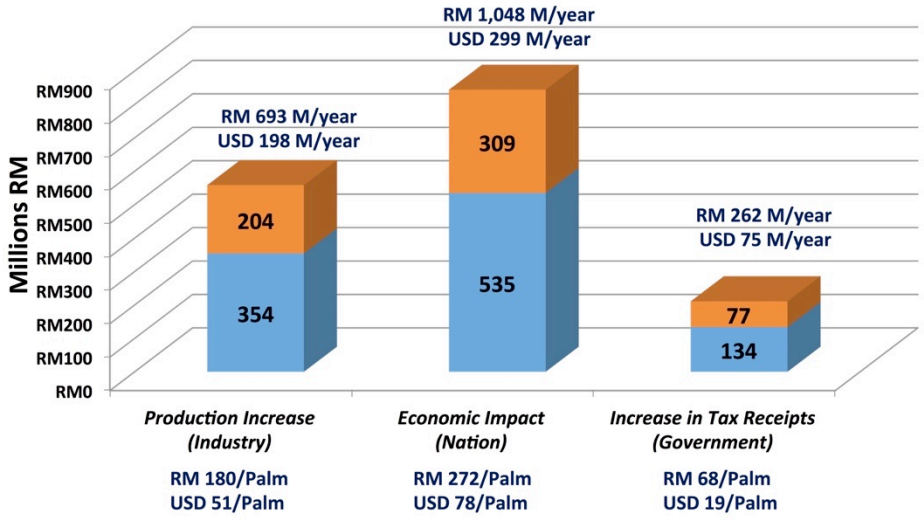


Figure C.5: Annual economic impact of *SHELL* genetic testing on the oil palm industry, gross national income and the Malaysian government.

Calculated annual increases (Malaysian Ringgit, RM) are plotted as proportions obtained by genetic screening based prevention of *dura* (blue) and *pisifera* (orange) contamination among independent planting sites. Gains per individual palm screened prior to commercial planting including *tenera* and non-*tenera* palms are indicated below each column. RM to U.S. dollar (USD) conversions are based on an average exchange rate over the past 20 years of 3.5RM/USD.

Appendix D

Population Level Coordination of Pigment Response In Individual Cyanobacterial Cells Under Altered Nitrogen Levels

This work will be submitted to *Cell Reports*. Jaclyn Murton, Aparna Nagarajan, Amelia Y. Nguyen*, Michelle Liberton, Harmony Hancock, Himadri B. Pakrasi[2] Jerilyn A. Timlin. Population Level Coordination of Pigment Response In Individual Cyanobacterial Cells Under Altered Nitrogen Levels. *Cell Reports.*

*A.Y.N. sent samples and wrote the paper.

D.1 Introduction

D.1.1 Abstract

In cyanobacteria, phycobilisome (PBS) pigment-protein complexes harvest light and transfer the energy to reaction centers. Cyanobacteria respond to changes in nutrient availability by altering PBS complexes. We have characterized the pigment response to nitrogen depletion and repletion at the subcellular level in individual live *Synechocystis* sp. PCC 6803 cells using hyperspectral confocal fluorescence microscopy (HCFM) and multivariate image analysis. Our results show that PBS degradation and re-synthesis are well coordinated, with homogeneous cell populations undergoing pigment modifications. Chlorophyll fluorescence originating from the photosystems decreased during nitrogen starvation, but the photosystem I/II ratio was unchanged, and no alteration in subcellular chlorophyll localization was found. We observed differential rod and core pigment responses to nitrogen deprivation, suggesting that PBSs undergo a stepwise degradation process. Compared to previous studies that assayed cellular response to nitrogen starvation at the population level, this analysis provides insights into nutrient stress response at the single cell level.

D.1.2 Introduction

Cyanobacteria are unicellular photosynthetic microbes that played a central role in oxygenating the Earth's atmosphere 2.4 billion years ago (Buick 2008). Cyanobacteria are found ubiquitously throughout the Earth's biosphere where they are crucial contributors to the global carbon and nitrogen cycles (Zehr et al. 2001). These organisms have diverse morphologies, including spherical, rod-shaped, and filamentous forms, and have the metabolic flexibility to thrive in environments ranging from terrestrial, freshwater, and marine habitats to more extreme conditions like deserts and hot springs. The ability of cyanobacteria to use light energy to fix

carbon dioxide and produce oxygen makes these organisms of critical importance in the biosphere, and of particular interest in biotechnological studies.

Cyanobacterial cells contain pigment-proteins that function to harvest and transfer light energy to the reaction centers in the thylakoid membranes to power photochemistry. The blue bilin and green chlorophyll (Chl) pigments impart the characteristic blue-green color to cyanobacteria. The bilin-containing phycobilisomes (PBS) are the main light-harvesting antenna in cyanobacteria. Bilins are covalently attached to proteins to form phycobiliproteins, of which there are three major types: phycocyanin (PC), allophycocyanin (APC), and phycoerythrin (PE). These phycobiliproteins are arranged in PBS pigment-protein complexes, giant membrane extrinsic structures of 3-5 MDa in size (Adir et al. 2006). In a well studied cyanobacterial model organism, *Synechocystis sp.* PCC 6803 (hereafter *Synechocystis* 6803), PBS are hemidiscoidal complexes consisting of a tricylindrical APC central core and six PC peripheral rods radiating from the core (MacColl 1998). The pigmented phycobiliproteins are interspaced with non-pigmented linker proteins in the PBS structure. PBS may account for up to 60% of the total soluble protein in the cell, thus serving as a large cellular nitrogen reserve (Bogorad 1975).

Cyanobacteria are subjected to varying concentrations of nutrients in their diverse habitats. Acclimation to fluctuating environmental conditions can be facilitated by plasticity in pigment abundance and localization. The dynamic nature of PBS is best observed during nutrient deprivation (nitrogen or sulfur depletion), when cells bleach due to the degradation of PBS (Collier and Grossman 1994). The degradation of phycobilisomes supplies macronutrients for cellular use during nutrient deprivation conditions, and may prevent photosystems from undergoing photoinhibition and production of harmful radical species (Adir et al. 2006). Consequently, PBS degradation plays an important function in cell survival.

A number of proteins have been identified that are involved in PBS degradation. An essential factor for the ATP-dependent degradation of PBS is non-bleaching A (NblA) (Collier and Grossman 1994). NblA is a protein that triggers the degradation of PBS in cyanobacteria by serving as an adapter protein to facilitate the interaction of a protease with phycobiliproteins (Baier et al. 2004; Karradt et al. 2008). NblA is a ~7kDa, ~60 amino acid polypeptide with low homology (about 30% sequence identity) (Dines et al. 2008). Most cyanobacteria and red algae have one copy of the *nblA* gene; *Synechocystis* 6803 is an exception that has two copies (Baier et al. 2001). Upon deletion of *nblA,* cells maintain high levels of PBS after nutrient step down (Collier and Grossman 1994). Additional proteins including NblR, NblS, NblB, RpaB, and NtcA have been identified that function with NblA to regulate PBS degradation during nutrient deprivation (Grossman et al. 1993; van Waasbergen et al. 2002; Dolganov and Grossman). Physiological experiments conducted during nutrient starvation in previous studies showed that essentially all PBS are degraded in wild-type cells within 48 hours (Collier and Grossman 1994; Li and Sherman 2002). Chl levels decreased dramatically following PBS degradation, with <1% of the original Chl remaining after 2 weeks of nitrogen deprivation (Gorl et al. 1998). Upon addition of nutrients, cyanobacteria regain their blue-green color due to the re-synthesis of PBS. The process of re-synthesis is also rapid, and cells can regenerate pigmentation with the re-addition of nitrate back to the culture even with prolonged nitrogen starvation of up to a month (Gorl et al. 1998).

Degradation of PBS is an active, rapid and specific process that occurs on a massive scale. Earlier studies have monitored the rate of degradation by measuring the decrease of the pigment peaks in whole cell absorption spectra in bulk cultures during nutrient depletion. A model of PBS degradation developed using this approach describes the sequential trimming of

the peripheral phycocyanin rods, starting at the most distal end, with complete degradation of the remaining PBS occurring within 2 days of continued nutrient depletion in *Synechococcus* sp. PCC 7942 (hereafter referred to as *Synechococcus* 7942) (Collier and Grossman 1994).

Advances in microcopy in recent years have made it possible to study pigment localization at the single cell level, however challenges with spectral resolution between pigments remains (Tóth et al. 2013). Previous techniques were unable to separate overlapping pigments (e.g. PSI/PSII or APC/PC), localize subcellular compartments, and image single cell population. Hyperspectral confocal fluorescence microscopy (HCFM) is a technique that allows the spectral resolution of pigments with similar fluorescence emission, and when combined with multivariate curve resolution (MCR) algorithms, pure fluorescent component spectra can be extracted from the emission of live cells. The resulting data sets can show the distribution of pigments within cells. Hyperspectral confocal fluorescence microscopy has been previously applied to isolate spectra from highly overlapping pigments in photosynthetic organisms**,** specifically APC, PC, PSI and PSII in *Synechocystis* 6803 (Vermaas 2008), and to analyze and compare the distribution of pigments in a group of PBS mutants with increasingly truncated antenna complexes (Collins et al.).

In this study, we explored the plasticity of pigment response in live *Synechocystis* 6803 by applying a quantitative imaging approached to characterize, at the subcellular level, the pigment response to nitrogen depletion and repletion conditions using HCFM. In a new application of HCFM, we coupled HCFM with single cell analysis to quantify changes in the different pigment components in individual single cells in order to gain insights into the degradation and re-synthesis of PBS on a population level. We carried out single cell quantification of pigment dynamics by analyzing PBS degradation in response to nitrogen

depletion to explore pigment changes in wild type *Synechocystis* 6803 during nitrogen starvation. Our results show that PBS degradation and re-synthesis are well coordinated, with homogeneous cell populations undergoing pigment modifications. Chl fluorescence originating from the photosystems decreased during nitrogen starvation, and the PBS to Chl ratio changed dramatically (~4x) under nitrogen depletion conditions. However, the photosystem I/II ratio was unchanged, and no alteration in subcellular chlorophyll localization was found. We observed differential rod and core pigment responses to nitrogen deprivation, suggesting that PBSs undergo a stepwise degradation process. These data provide insights into how individual pigment proteins react to changes in extracellular nitrogen at the single cell level.

D.2 Material and Methods

D.2.1 Cyanobacterial Strains and Culture Conditions

Synechocystis sp. PCC 6803 was grown photoautotrophically in BG11 medium with 1.76 M $NaNO_3$ as a nitrogen source. The nitrogen source was replaced with 1.76 M NaCl to generate nitrogen depleted media. Cultures were incubated at 30^0C while continuously shaking at 150 rpm (VWR Orbital Shaker) under constant, cool-white light (30 μmol photons•m^{-2}•s^{-1})**.** To perform the experiment, three replicate cultures were started by transferring 20 mL of *Synechocystis* culture (10^7 cells/mL) into 25 mL BG11 media containing nitrogen and incubated for three days under standard culture conditions. After three days each culture was subdivided into two equal volumes, harvested by centrifugation (500 X g for 10 min) and washed twice with either nitrogen replete BG11 media (BG11 (+N)) or nitrogen depleted media (BG11 (-N)). After washing, the cells were resuspended in 50 mL BG11 (+N) or BG11 (-N) and cultured in a 500 mL baffled Erlenmeyer flask for 24 hours. Both cultures were harvested by centrifugation after 24 hours, resuspended in media (+ or – N, depending on experimental condition to be tested) and

cultured for another 24 hours. This process was repeated for all the required experimental time points. Bulk culture absorbance spectra (400-800 nm) were obtained from all cultures/timepoints using a plate reader (BioTek Eon). Samples were prepared hyperspectral confocal fluorescence microscopy (HCFM) images by withdrawing 25 μL of the first concentrated cell pellet (prior to any washing steps). This aliquot was resuspended in 100 uL of media (+ or – N, depending on experimental condition to be tested) and 8uL of the resulting culture was placed on an agar coated slide. Cells were allowed to settle for 60 sec and a coverslip was applied (#1.5), excess culture wicked from the edges, and sealed with nail polish. Imaging was performed immediately.

D.2.2 Single Cell HCFM Microscopy

Hyperspectral confocal fluorescence images were acquired using a custom hyperspectral confocal fluorescence microscope described previously (Sinclair et al 2006). In brief, 3 uW of 488 nm laser excitation was focused onto the sample through a 60x oil-immersion objective (Nikon Plan Apochromat; NA=1.4) to a diffraction-limited spot. Fluorescence emission was collected through the same objective and dispersed by a prism spectrometer onto the focal plane of an electron-multiplied CCD array (iXon DU897U, Andor Technologies). The per pixel dwell time was 240 ms. The image was formed by raster scanning the beam over the sample with step size of 0.12 um. This generates images with diffraction limited lateral spatial resolution (240 nm). A total of 256 images, each containing 44,100 spectra were collected.

D.2.3 Spectral Image Analysis

Hyperspectral images were preprocessed and subsequently analyzed using multivariate image analysis methods to extract the underlying spectral components and calculate their relative contributions to each image pixel as described in Jones et. al. (Jones 2012). Representative images from each time point and sample were combined into one image data set and MCR was

executed with non-negativity constraints on all image pixels above the background. This resulted in a 5 (or 6) component spectral model that explained >99.4% of the spectral variance. An autofluorescent component was included in the analysis to account for dead or dying cells. Concentration maps indicating the abundance and location of each component were generated using a classical least squares analysis. The resulting concentration maps were segmented using a modified watershed transformation algorithm to identify individual cells. Automated cell segmentation was verified and edited manually. Single cell statistics were calculated for individual cells. Dead and dying cells were identified by their extremely high autofluorescent abundance and/or high PBS fluorescence and excluded from calculation of single cell statistics.

D.3 Results

D.3.1 Pigment Response to Nitrogen Depravation in *Synechocystis* 6803 Cultures

Synechocystis 6803 cells were starved for nitrogen and samples were collected at 0 h, 24 hr, and 48 hr time points, and at 24 hr after nitrogen readdition (**Fig. D.1a**). These samples were used for absorbance measurements and HCFM. The absorbance peak originating from PBS (~625 nm) gradually declined over 24 hr of nitrogen depletion and recovered to above original levels upon nitrogen repletion (**Fig. D.1b**). Likewise, the chlorophyll absorbance peak (680 nm) declined with nitrogen depletion and recovered above original levels when nitrogen was added back to the media (**Fig. D.1b**).

The PC and Chl concentrations per OD_{730} were measured at the different time points during nitrogen depletion and readdition (**Fig. D.2**). The PC content in the control culture gradually increased by a factor of 1.7 over 48 hr, while the PC content in the nitrogen depleted culture dramatically decreased by a factor of 6.5 within 24 hr (**Fig. D.2a**). When nitrogen was added back to the depleted culture, the PC concentration increased 12 fold to normal levels. The

Chl response to changing nitrogen levels exhibited a similar pattern to PC, although diminished. The average chlorophyll concentration per OD_{730} decreased by a factor of 1.8 in nitrogen depleted media and exhibited a 2.4 fold increase in upon nitrogen repletion (**Fig. D.2b**). Interestingly, the average chlorophyll concentration per OD_{730} in the nitrogen depleted culture did not appear to recover to normal levels 24 hours after nitrogen was added back to the media. To elucidate further, we investigated the ratio of PC to chlorophyll (**Fig. D.2c**), which generally remained constant over 48 hours in the cultures with nitrogen in the media. The ratio decreased fivefold in the nitrogen depleted media, following the same trend observed in the adaptive response of PC to variable nitrogen levels (**Fig. D.2a**). Upon repletion of nitrogen, the ratio of PC to chlorophyll increased seven fold (**Fig. D.2c**).

D.3.2 Single Cell Analysis

This section is incomplete because our collaborator at Sandia National laboratory needs to discuss about cell finder and single cell statistics. The goal was to investigate changes in pigments in intact *Synechocystis* 6803 cells, we used HCFM combined with MCR analysis.

D.4 Discussion

Still in progress.

D.5 Acknowledgements

The authors acknowledge Dr. Michael B. Sinclair for the maintenance and use of the hyperspectral confocal fluorescence microscope used for these studies and Howland Jones and Stephen Anthony for the previous development of the multivariate curve resolution and cell segmentation software used in this study. This work was supported as part of the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award #DE-SC0001035.

D.5.1 Author Contributions

Conceptualization, H.B.P. and J.A.T.; Methodology, H.B.P and J.A.T.; Investigation, J.M., A.N., A.Y.N., M.L., and H.H.; Formal Analysis and Validation, J.M. and J.A.T., Writing original draft, J.M., A.N., A.Y.N., M.L., H.B.P. and J.A.T.; Writing review and editing, J.M., A.N., A.Y.N., M.L., H.H. H.B.P and J.A.T.; Visualization, J.M. and J.A.T.; Supervision, J.A.T. and H.B.P; Funding acquisition, J.A.T. and H.B.P.

D.5.1 Conflict of Interest Statement

The authors declare no conflict of interest.

D.6 References

Adir, N., Dines, M., Klartag, M., McGregor, A., Melamed-Frank, M. (2006). "Assembly and disassembly of phycobilisomes." *Complex Intracellular Structures in Prokaryotes*. Springer Berlin Heidelberg, Berlin, Heidelberg: 47-77.

Baier, K., Lehmann, H., Stephan, D.P., Lockau, W. (2004). "NblA is essential for phycobilisome degradation in *Anabaena s*p. strain PCC 7120 but not for development of functional heterocysts." *Microbiology*. **150**: 2739-2749.

Baier, K., Nicklisch, S., Grundner, C., Reinecke, J., Lockau, W. (2001). "Expression of two *nblA*-homologous genes is required for phycobilisome degradation in nitrogen-starved *Synechocystis* sp. PCC 6803." *FEMS Microbiology Letters*. **195**:35-39.

Bogorad, L. (1975). "Phycobiliproteins and complementary chromatic adaptation." *Annual Review of Plant Physiology*. **26**(1): 369-401.

Buick, R. (2008). "When did oxygenic photosynthesis evolve?" *Philosophical transactions of the royal society B*. **363**(1504): 2731-2743.

Collier, J.L., Grossman, A.R. (1994). "A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria." *EMBO Journal*. **13**(15): 1039-1047.

Collins, A.M., Liberton, M., Jones, H.D.T., Garcia, O.F., Pakrasi, H.B., Timlin, J.A. (2012). "Photosynthetic pigment localization and thylakoid membrane morphology are altered in *Synechocystis* 6803 phycobilisome mutants." *Plant Physiology*. **158**(4): 1600-1609.

Dines, M., Sendersky, E., David, L., Schwarz, R., Adir, N. (2008). "Structural, functional, and mutational analysis of the NblA protein provides insight into possible modes of interaction with the phycobilisome." *Journal of Biological Chemistry*. **283**(44): 30330-30340.

Dolganov, N., Grossman, A.R. (1999). "A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes." *Journal of Bacteriology.* **181**(2): 610-617.

Gorl, M., Sauer, J., Baier, T., Forchhammer, K. (1998). "Nitrogen-starvation-induced chlorosis in *Synechococcus* PCC 7942: adaptation to long-term survival." *Microbiology*. **144**(Pt 9): 2449-2458.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., Collier, J.L. (1993). "The phycobilisome, a light-harvesting complex responsive to environmental conditions." *Microbiological Review*. **57**(3): 725-749.

Karradt, A., Sobanski, J., Mattow, J., Lockau, W., Baier, K. (2008). "NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease." *Journal of Biological Chemistry*. **283**(47): 32394-32403.

Li, H., Sherman, L.A. (2002). "Characterization of *Synechocystis* sp. strain PCC 6803 and deltanbl mutants under nitrogen-deficient conditions." *Archives of Microbiology*. **178**(4): 256-266.

MacColl, R. (1998). "Cyanobacterial phycobilisomes." *Journal of Structural Biology*. **124**(2-3): 311-334.

Tóth, T., Chukhutsina, V.V., Krumova, S.B., Gombos, Z., van Amerongen, H. (2013). "Fluorescence Lifetime Imaging Microscopy of *Synechocystis* WT Cells — Variation in Photosynthetic Performance of Individual Cells in Various Strains of *Synechocystis* sp. PCC 6803." Photosynthesis Research for Food, Fuel and the Future: 15th International

Conference on Photosynthesis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 139-142.

van Waasbergen, L.G., Dolganov, N., Grossman, A.R. (2002). "*nblS*, a gene involved in controlling photosynthesis-related gene expression during high light and nutrient stress in *Synechococcus elongatus* PCC 7942." *Journal of Bacteriology*. **184**(9): 2481-2490.

Vermaas, W.F.J., Timlin, J.A., Jones, H.D.T., Sinclair, M.B., Neiman, L.T., Hamad, S.W., Melgaard, D.K., Haaland, D.M. (2008). "*In vivo* hyperspectral confocal fluorescence imaging to determine pigment localization and distribution in cyanobacterial cells." *Proceedings of the National Academy of Sciences*. **105**(10): 4050-4055.

Zehr, J., Waterbury, J., Turner, P., Montoya, J., Omoregie, E., Steward, G., Hansen, A., Karl, D. (2001). "Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean." *Nature*. **412**(6847): 635-638.

Pigment	% of original abundance	
	following N depletion	following N repletion
PC	3	88
APC	9	87
Chl-PSI	43	80
Chl-PSII	38	95

Table D.1: Average % of original abundance of photosynthetic pigments.

Following nitrogen depletion (24 hrs) and repletion (24 hrs) calculated from the single cell hyperspectral confocal fluorescence images.

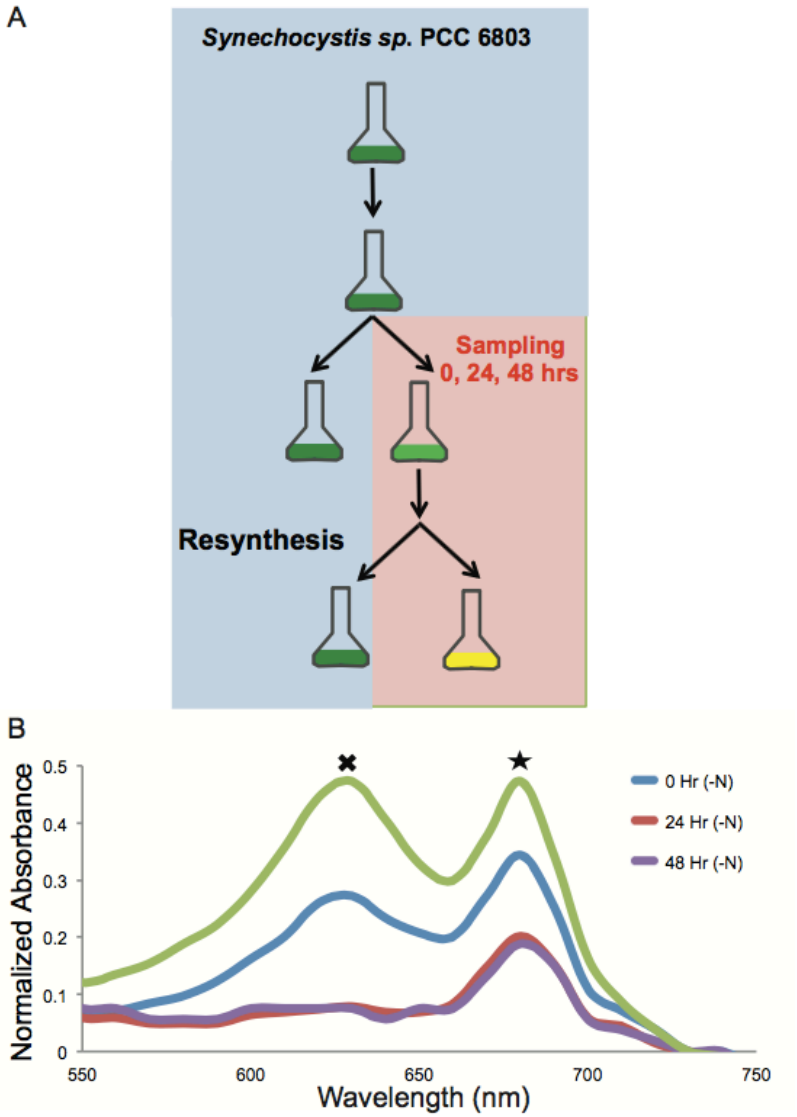


Figure D.1: Hyperspectral work-flow and pigments measurements.

a Schematic of experimental design. **b** Bulk absorbance measurements of *Synechocystis* cultures under varying nitrogen conditions. Absorbance spectra normalized to A730. ✖ indicates the phycobilin peak, while ★ indicates the chlorophyll peak.

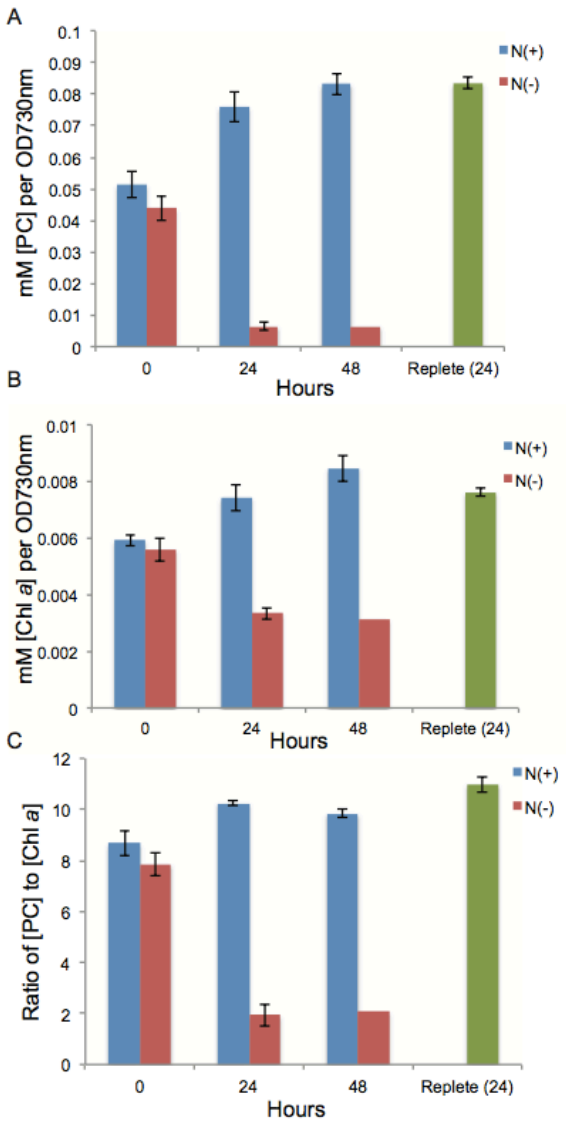


Figure D.2: Changes in the concentration of pigments at different time points under varying nitrogen conditions.

a Phycobilin content per cell. **b** Chlorophyll content per cell. **c** Ratio of phycobilin to chlorophyll content per cell.

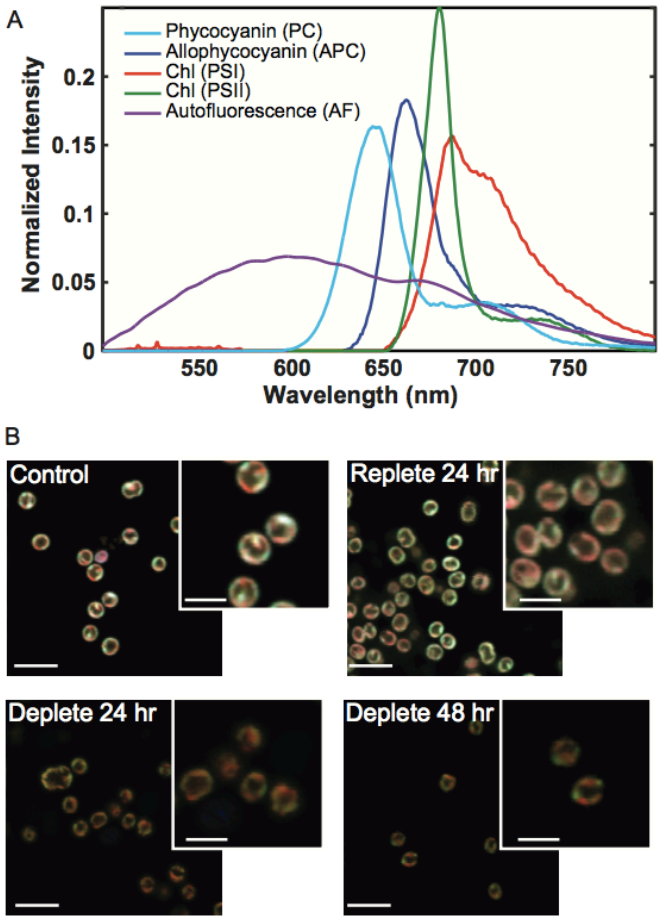


Figure D.3: Multivariate curve resolution.

Results from hyperspectral confocal fluorescence images of single *Synechocystis* 6803 cells under a time course of varying nitrogen conditions. **a** Spectral model. Fluorescence emission spectra corresponding to four photosynthetic pigments and a broad autofluorescence emission spectrum were mathematically isolated. A 6th spectrum, a flat offset, has been omitted for simplicity. **b** RGB images corresponding to the relative abundance of the four photosynthetic pigments. Red: Chlorophyll (PSI), Green: Chlorophyll (PSII), Blue: phycocyanin + allophycocyanin. Scale bars = 5 μm in large images, 2.5 μm in the zoomed inset images. Colorscales are identical for all images to facilitate comparison between images.

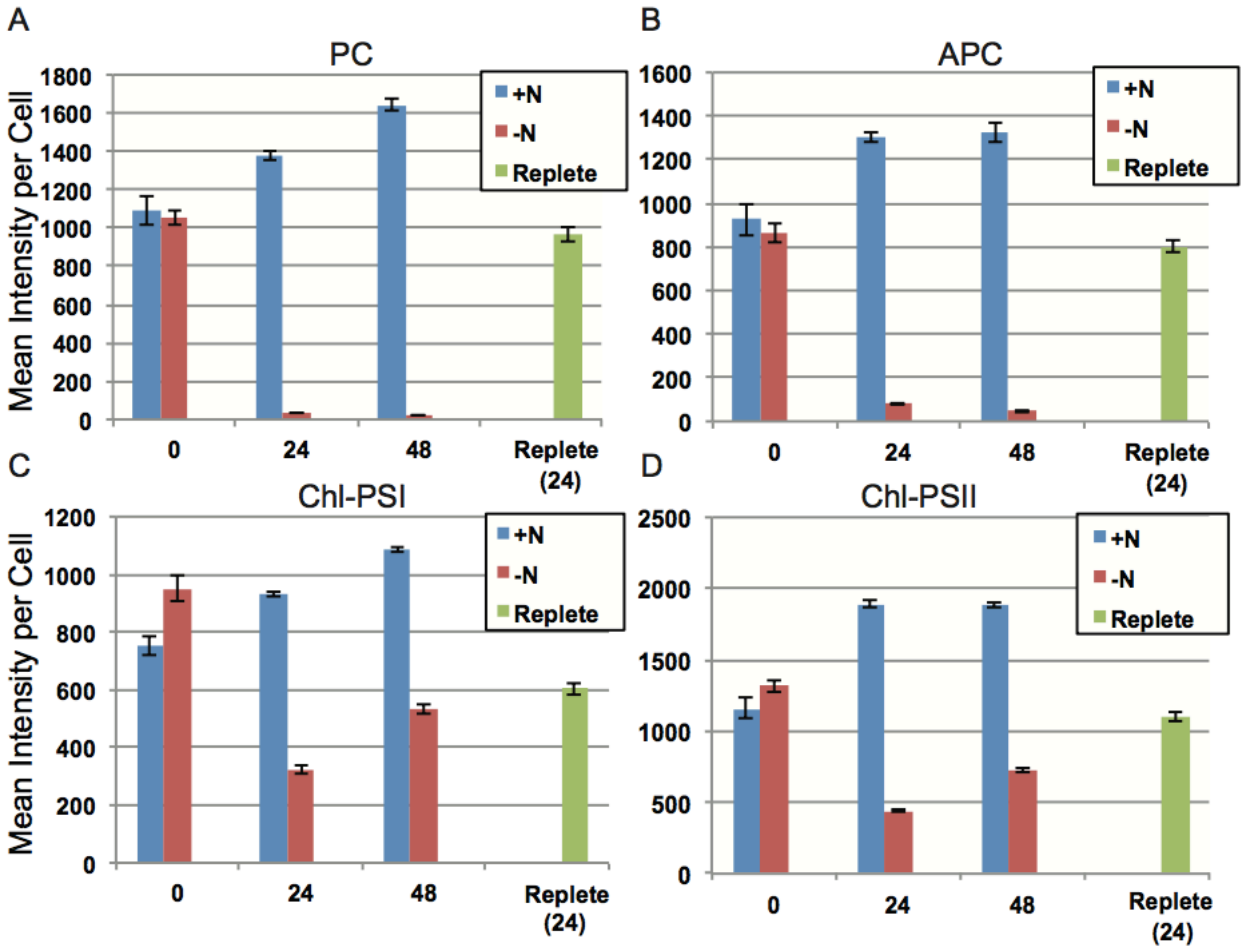


Figure D.4: Average per cell fluorescence.

Intensities of photosynthetic pigments from hyperspectral confocal fluorescence images of *Synechocystis* 6803 under a time course of varying nitrogen conditions. **a** Phycocyanin, **b** Allophycocyanin, **c** Chlorophyll (PS1), and **d** Chlorophyll (PSII). Error bars represent the standard error of the single cell measurements.

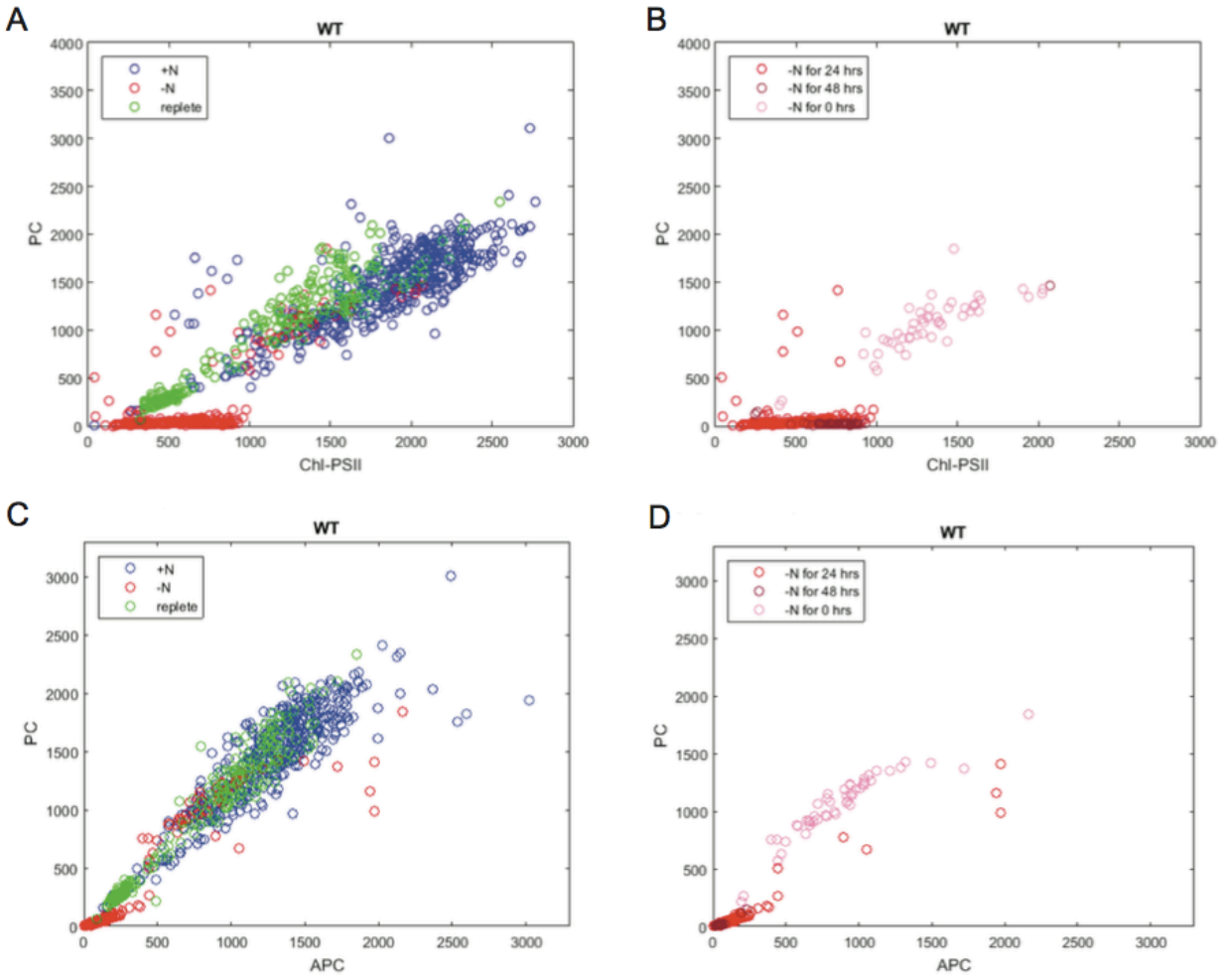


Figure D.5: Single cell scatter plots for comparing abundances of photosynthetic pigments. **a** & **b** Chl-PSII vs. phycocyanin content and **c** & **d** phycocyanin vs. allophycocyanin content in individual cells. **a** & **c** All time points are represented in control (blue circles), nitrogen deplete (red circles) and replete conditions (green circles). **b** & **d** Same data as **a** & **c**, however, only showing deplete conditions and color coding for 0, 24, and 48 hrs in nitrogen deplete environment.

AMELIA YEN NGUYEN

aynguyen@wustl.edu | 714.475.9150

EXECUTIVE SUMMARY

- Protein biochemist with over 7 years of life sciences research and 3 years of consulting experiences
- Productive independent researcher with excellent communication and management skills
- Employed methods in genetics, biochemistry, and mass spectromety to study different proteins and peptides
- Served teammates by advising consulting projects with a focus on technology assessment and market research

EDUCATION

Washington University in St. Louis, Division of Biology and Biomedical Sciences, St. Louis, MO

Ph.D. in Plant and Microbial Biosciences (August 2016)

- National Science Foundation Graduate Research Fellowship, >$132,000, 5-year award
- Presidential Management Fellow STEM Program Finalist, 2-year training at a government agency

University of California, Berkeley, College of Natural Resources, Berkeley, CA

B.S. in Genetics and Plant Biology (May 2011)

- Gates Millennium Scholarship, 2007-present, >$400,000, 10-year as needed award
- Semester study and research at Tohoku University in Sendai, Japan

RESEARCH EXPERIENCE

Washington University in St. Louis, St. Louis, MO, USA

Ph.D. Candidate, laboratory of Dr. Himadri Pakrasi, 50 hours/week (2011-2016)

- Elucidating the degradation mechanism of light harvesting protein complexes
- Developed a global proteomics map to examine changes in redox-sensitive residues
- Led a multi-institution collaboration, resulting in a 1st authored paper in *Molecular & Cellular Proteomics*
- Created molecular tools for a 2nd authored paper in *Proceedings of the National Academy of Sciences*
- Initiated a 17 member-lab improvement program, resulting in a comprehensive lab manual

University of California, **Berkeley,** Berkeley, CA

Undergraduate Research Assistant, laboratory of Dr. Norman Terry, 20 hours/week (2009-2011)

- Managed and led reclaimation of boron contaminated studies
- Analyzed and wrote a literary review of native boron tolerant plants for reports to Rio Tinto Borax

BUSINESS & MANAGEMENT EXPERIENCE

The Biotechnology and Life Science Advising Group, St. Louis, MO

Director of Finance, ~5 hours/week (2015-2016)

- Managing the finances of an 80+ people, >$60,000 annual revenue non-profit company
- Created a working group and library to teach ~20 members about finances and accounting

Project Advisor, Manager, and Consultant, 15 hours/week (2013-2016)

- Leading >10 cross-functional team-based consulting projects
- Assisting firms in overcoming obstacles and exploring opportunities, specifically in
- technology assessment, market research, competitor analysis, and patentability review
- Delivering data-driven recomendations in written reports and oral presentations to clients
- Submitted a joint paper in *Frontiers in Plant Sciences* with client that is currently in press

American Society for Cell Biology – Keck Graduate Institute, Claremont, CA

2-week managing science in the biotech industry course attendee (July 2015)

- 1 of 40 (out of 591 applicants) awarded a full scholarship to attend the course
- Received formal training in bioscience management
- Managed team-based consulting project and presented deliverables to professors and trustees

Promoting Science Policy, Education, and Research, St. Louis, MO

Career Development Leader, *Webmaster*, and *Founding Officer,* 7 hours/week (2012-2014)

- Advocated the interplay between public policy and scientific research and enriched students' career development by organizing policy panel discussions and speaker series
- Presented an environmental initiative at the Clinton Global Initiative University

PUBLICATIONS

- **Nguyen AY,** Bricker, W, Zhang H, Gross M, Pakrasi HB**.** 2016. Proteolysis Adaptor, NblA, Binds to β-PC: Implications into PBS Degradation. *Photosynthesis Research.* In Preparation.
- Murton J, Nagarajan A, **Nguyen AY**, Liberton M, Hancock H, Pakrasi HB, Timlin J. 2016. Effect of Nitrogen Starvation on Pigment Localization and Stoichiometry in the Cyanobacterium *Synechocystis* sp. PCC 6803. *Cell Reports.* In Preparation.
- Malaysian Palm Oil Board, *et al*. 2016. Non-*tenera* contamination and the economic impact of *SHELL* genetic testing in the Malaysian independent oil palm industry. *Frontiers in Plant Science*. 7(771):1-13.
- Liberton M*, Saha R*, Jacobs JM*, **Nguyen AY**, Gritsenko MA, Smith RD, Koppenaal DW, Pakrasi HB. 2016. Global Proteomic Analysis of Isolated Plasma Membrane and Thylakoid Membrane from the Cyanobacterium *Synechocystis* sp. PCC 6803. *Molecular & Cellular Proteomics*. 15(6):2021-32. (*Co-First Authors).
- Guo J*, **Nguyen AY***, Dai Z*, Su D, Gaffrey MJ, Moore RJ, Jacobs JM, Monroe ME, Smith RD, Koppenaal DW, Pakrasi HB, Qian WJ. 2014. Proteome-wide Light/Dark Modulation of Thiol Oxidation in Cyanobacteria Revealed by Quantitative Site-Specific Redox Proteomics. *Molecular & Cellular Proteomics*. 13(12): 3270-85. (*Co-First Authors).
- Ravilious GE, **Nguyen A**, Francois JA, Jez JM. 2011. Structural basis and evolution of redox regulation in plant adenosine-5'-phosphosulfate kinase. *Proceedings of the National Academy of Sciences.* 109(1):309-14.

TEACHING EXPERIENCE

Washington University in St. Louis, St. Louis, MO

- Guest Lecturer, Bioenergy (Biology 4830/EECE 5830), Spring 2016
- Guest Lecturer, Lab Investigations with Model Organisms (Biology 579), Summer 2013
- Teaching Assistant, Principles of Biology I (Biology 2960), Spring 2013

University of California, **Berkeley,** Berkeley, CA

- SAT Math Instuctor, People's Test Preparation Service, 2010-2011

SKILLS & INTERESTS

- Languages- English (fluent), Vietnamese (conversational), and Japanese (beginner)
- Software- Microsoft Office, Illustrator, and Photoshop (all at an advance level)
- Sports- Running, Tennis, and Badminton

REFERENCES

Dr. Himadri Pakrasi

Department of Biology

Washington University in St. Louis

314-935-6853

pakrasi@wustl.edu

Dr. Joseph Jez

Department of Biology

Washington University in St. Louis

314-935-3376

jjez@wustl.edu

Dr. Robert Blankenship

Department of Biology

Washington University in St. Louis

314-935-7971

blankenship@wustl.edu