RESPONSE OF *SYNECHOCYSTIS* SP. PCC 6803 TO PHOTOPERIOD AND PHOSPHATE ALTERATIONS USING FUNCTIONAL PROTEOMICS APPROACHES

Chee Sian GAN (B. ENG)

Thesis submitted for the degree of Doctor of Philosophy (PhD)

to The University of Sheffield, Sheffield, UK

On completion of research in the Biological and Environmental Systems Group within the Department of Chemical and Process Engineering

December, 2006

This thesis is dedicated to my late father, Mr. Yong Lea GAN.

"I hope I make you proud, dad!"

Declaration

This is to declare that this thesis is an account of the author's work carried out and completed in The University of Sheffield, UK, except where acknowledgements are made. This work has not been submitted for any other degrees.

C S Gan

Chee Sian GAN (Candidate)

Acknowledgments

The road toward the completion is always the most difficult part of this entire journey. From a clueless engineering graduate until today for what I am, I am gladful along this gruelling journey, there is always someone watching over my shoulder.

1. My late-farther, Mr. Yong Lea GAN. He has been the main driving force in my research. I will always remember him. A special gratitude to my mom and family members for all these years of endless love and support.
2. My highest appreciation to my supervisor Prof. Phillip C. Wright for his excellent guidance, support and inspiration. I thank him for his confidence and trust throughout this entire journey.
3. I thank Dr. Adam M. Burja for his exceptional support and constructive advice on this research work. His unselfishness and fruitful comments has brought me through the most difficult times of my research. I also acknowledge Dr. Helia Radianingtyas for her helps during the early phase of my research.
4. A special acknowledgement to Dr. (Shirly) Poh Kuan CHONG. She is always there whenever I was sad, frustrated, discouraged or in troubled. I thank her for the support all these years. I also extended my gratitude to her, together with Mr Trong Khoa PHAM for the collaboration work in Chapter 4, and his endless helps whenever I needed him.
5. I also thank Prof. Kenneth F. Reardon (from Department of Chemical and Biological Engineering, Colorado State University) for his excellent technical advice, and collaboration work in Chapter 3 (Shotgun Prefractionation Study), Chapter 5 and 6 (Light-Dark Study), and Chapter 7 (Phosphate Starvation Study).
6. I express my gratitude to Dr. Alireza Fazeli and Ms. Sarah Elliot (from Academic Unit of Reproductive and Developmental Medicine, University of Sheffield) for their technical helps and the access of facilities to real-time PCR.

7. I would like to thank Prof. Geoff McMullan and Dr. Nigel G. Ternan (from School of Biomedical Sciences, University of Ulster) for technical advices, and carrying out the phosphatase enzyme assay in Chapter 7.

8. Many thanks to Dr. Bram Snijders, Dr. Mark Dickman and Mr. Mark Scaife for all their technical advices and guidance, specifically to Dr. Martin Barrios-Llerena for his teaching during the early phase of my research.

9. I would like to acknowledge Prof. Birgitta Norling (from Department of Biochemistry and Biophysics, Stockholm University, Sweden) and Dr. Catherine Biggs (from Department of Chemical and Process Engineering, University of Sheffield) for being my external and internal examiner respectively. Many of their critical views and suggestions have been incorporated into this thesis.

10. Finally my gratitude to The University of Sheffield for funding my research work over these three years period.

List of Publications

Parts of the work in this thesis have been published as below:

(J – Journal Paper, O – Oral Presentation, P – Poster Presentation)

(J1) C. S. Gan, K. F. Reardon and P. C. Wright, "Comparison of Protein and Peptide Prefractionation Methods for the Shotgun Proteomic Analysis of *Synechocystis* sp. PCC 6803", *Proteomics*, 2005, 5(9), 2468-78.

(J2) P. K. Chong, C. S. Gan, T. K. Pham and P. C. Wright, "Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) Reproducibility: Implication of Multiple Injections", *Journal of Proteome Research*, 2006, 5(5), 1232-40.

(J3) M. E. Barrios-Llerena, P. K. Chong, C. S. Gan, A. P. L. Snijders, K. F. Reardon and P. C. Wright, "Shotgun Proteomics of Cyanobacteria - Applications of Experimental and Data-Mining Techniques", *Briefings in Functional Genomics and Proteomics*, 2006, 5(2), 121-32.

(J4) T. K. Pham, P. K. Chong, C. S. Gan, P. C. Wright, "Proteomic Analysis of *Saccharomyces cerevisiae* under High Gravity Fermentation Condition", *Journal of Proteome Research*, 2006, 5(12), 3411-19.

(J5) C. S. Gan, P. K. Chong, T. K. Pham and P. C. Wright, "Technical, Experimental and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ)", *Journal of Proteome Research*, 2007, 6(2), 821-27.

(J6) C. S. Gan, A. M. Burja, A. Fazeli, K. F. Reardon and P. C. Wright, "Central Carbon Metabolism Reconstruction of *Synechocystis* sp. PCC 6803 During Light Acclimation by Quantitative Proteomic Approach" (In Preparation)

(J7) C. S. Gan, A. M. Burja, A. Fazeli, K. F. Reardon and P. C. Wright, "Circadian Rhythm Study of *Synechocystis* sp. PCC 6803 Under a 12-hour Dark and Light Cycle" (In Preparation)

(J8) C. S. Gan, N. G. Ternan, G. McMullan, K. F. Reardon and P. C. Wright, "Global Proteomic Response of *Synechocystis* sp. PCC 6803 During Phosphate Starvation" (In Preparation)

(O1) C. S. Gan, A. M. Burja, K. F. Reardon and P. C. Wright, "Transcriptome and Proteome Dynamics of *Synechocystis sp.* PCC 6803 During Light-Dark Cycling", American Institute of Chemical Engineers (AICHE), Annual Meeting 2004, Austin, TX, Nov 7-12th, 2004.

(O2) C. S. Gan, A. M. Burja, K. F. Reardon and P. C. Wright, "System Dynamics of *Synechocystis sp.* PCC 6803 During Light-Dark Cycling", 229th American Chemical Society (ACS) National Meeting, San Diego, CA, Mar 13-17th, 2005.

(O3) C. S. Gan, A. M. Burja, K. F. Reardon and P. C. Wright, "Transcriptome and Proteome Dynamics of *Synechocystis sp.* PCC 6803 During Light-Dark Cycling", 7th World Congress of Chemical Engineering (IChemE), Glasgow, Scotland, Jul 10-14th, 2005.

(O4) C. S. Gan, A. M. Burja, K. F. Reardon and P. C. Wright, "Towards a Systems Biology Understanding of the Dynamics of *Synechocystis sp.* PCC 6803 During Light-Dark Cycling", American Institute of Chemical Engineers (AICHE), Annual Meeting 2005, Cincinnati, OH, Nov, 2005.

(O5) C. S. Gan, K. F. Reardon and P. C. Wright, "Proteomics-Based Systems Biology Study of the Phosphorus Starvation Response in the Cyanobacterium *Synechocystis sp.* strain PCC 6803", American Institute of Chemical Engineers (AICHE), Annual Meeting 2005, Cincinnati, OH, Nov, 2005.

(O6) C. S. Gan, P. K. Chong, K. F. Reardon and P. C. Wright, "Analysis of Proteomes Using Shotgun Approaches", Proteomics Method Forum (PMF), University of Manchester, 17-18th Nov, 2005.

(O7) P. K. Chong, C. S. Gan, T. K. Pham and P. C. Wright, "Application and Reliability of Isobaric Mass Tagging Technology in Proteomics", 232nd ACS National Meeting, San Francisco, CA, Sep 10-14th, 2006.

(O8) M. E. Barrios-Llerena, P. K. Chong, C. S. Gan, A. P. L. Snijders, K. F. Reardon and P. C. Wright, "Enhancing Shotgun Proteomics Data Analysis with Bioinformatics Tools: Application to Cyanobacterial Proteomics", American Institute of Chemical Engineers (AICHE), Annual Meeting 2006, San Francisco, CA, Nov, 2006.

(O9) A. L. Knorr, C. S. Gan, K. F. Reardon, P. C. Wright and R. Srivastava, "Comparative Metabolic Modeling with Integrated Proteomic Analysis of *Synechocystis sp.* PCC 6803

during Light Versus Dark Cycling", American Institute of Chemical Engineers (AICHE), Annual Meeting 2006, San Francisco, CA, Nov, 2006.

(O10) C. S. Gan, P. K. Chong and P. C. Wright, "Assessing iTRAQ Reliability", Proteomics Method Forum (PMF), University of Sheffield, Nov 30th – Dec 1st, 2006.

(P1) C. S. Gan, A. M. Burja, and P. C. Wright, "Characterization of the Transient Behavior of *Synechocystis sp.* PCC6803 as an Aid to Metabolic Pathways Elucidation", Society for General Microbiology (SGM), 154th Meeting, University of Bath, UK, Mar 29th - Apr 2nd, 2004.

(P2) C. S. Gan, K. F. Reardon and P. C. Wright, "Comparison of Protein Prefractionation Methods for MDLC Analysis of the *Synechocystis* sp. PCC 6803 Proteome", 6th Siena Meeting, From Genome to Proteome: Biomarker Discovery and Proteome Imaging, Siena, Italy, Aug 30th - Sep 2nd , 2004.

(P3) C. S. Gan, A. M. Burja, K. F. Reardon and P. C. Wright, "Towards Cyanobacterial Systems Biology: Transcriptome and Proteome Dynamics of *Synechocystis sp.* PCC 6803 During Light-Dark Cycling", Biochemical Society Systems Biology Focused Meeting, University of Sheffield, UK, Jan 12-14th, 2005.

(P4) A. P. L. Snijders, C. S. Gan, P. K. Chong, A. Sterling, M. Baumert, P. J. Jackson, K. F. Reardon and P. C. Wright, "Use of Automated Nano Electrospray Tandem Mass Spectrometry for Analysis of Complex Proteomes", Annual Meeting 2005, American Society for Mass Spectrometry, San Antonio, Texas, May, 2005.

(P5) C. S. Gan, A. M. Burja, K. F. Reardon and P. C. Wright, "Transcriptome and Proteome Dynamics of *Synechocystis sp.* PCC 6803 During Light-Dark Cycling", 3rd Joint BSPR/EBI Meeting, Cambridge, UK, Jul 12-14th, 2006.

(P6) M. E. Barrios-Llerena, P. K. Chong, C. S. Gan, A. P. L. Snijders, K. F. Reardon and P. C.Wright, "Shotgun Proteomics of Cyanobacteria: Applications of Experimental and Data Mining Techniques", 3rd Joint BSPR/EBI Meeting, Cambridge, UK, Jul 12-14th, 2006.

TABLE OF CONTENT

Note: Due to excessive amount of supporting information, Appendix A to D are available as soft-copy (CD format) attached at the back of this thesis; while Appendix E to N are available as hard-copy.

Abbreviations

Symbol	Full Description
A	absorbance
ACN	acetonitrile
amu	atomic mass unit
bp	base pairs
CHAPS	3-{(3-cholamidopropyl)dimethylamonio}-1-propanesulfonate
CID	collision-induced dissociation
CV	coefficient of variation
1-D	one-dimensional
2-D	two-dimensional
Da	dalton (molecular mass)
2-DE	two-dimensional gel electrophoresis
DIGE	fluorescence difference gel electrophoresis
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
HEPES	*N*-(2-hydroxyethyl)piperazine-2'-(2-ethanesulfonic acid)
HPLC	high-performance liquid chromatography
HTML	hypertext mark-up language
id	inside diameter
IEF	isoelectric focusing
IPG	immobilized pH gradient
kbp	kilobase pairs
kDa	kilodalton (molecular mass)
LC	liquid chromatography
MALDI-MS	matrix-assisted laser desorption/ionization - mass spectrometry
Mbp	megabase pairs
MES	2-(*N*-morpholino)ethanesulfonic acid
MOPS	3-(*N*-morpholino)propanesulfonic acid
Mw	molecular weight
MS	mass spectrometry
MS/MS	tandem mass spectrometry
m/z	mass-to-charge ratio
NAD	nicotinamide adenine dinucleotide
NAD(P)	nicotinamide adenine dinucleotide phosphate
od	outside diameter
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCC	Pasteur Culture Collection
PCR	polymerase chain reaction
p*I*	isoelectric point
Pi	inorganic phosphate
ppm	parts per million
PTM	post-translational modification
RP	reversed phase
rpm	revolutions per minute
RT-PCR	reverse transcriptase-PCR
SD	standard deviation
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
SELDI	surface-enhanced laser desorption/ionization
SPE	solid-phase extraction

Symbol	Full Description
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	trifluoroacetic acid
TIC	total ion current
TOF	time of flight
Tris	tris(hydroxymethyl)aminomethane
U	unit
URL	uniform resource locator
UV	ultraviolet
Vh	volt x hours
WWW	World Wide Web

Abstract

The consideration of possible future clean-fuel (H_2), biodegradable plastic and potential anti-cancer drug has made the fresh-water unicellular cyanobacterium, *Synechocystis* sp. PCC 6803 one of the most attractive cyanobacteria strains to commercial and industrial markets. To fully exploit the potential of these natural products, it is essential to understand the cellular response of this organism toward environmental stimulations. In this study, *Synechocystis* sp. was subjected to qualitative 'shotgun'-based proteomic analysis, and a total of 776 unique proteins, covered 24% of the entire proteome, were successfully identified. The use of *Synechocystis* sp. as a model of organism also revealed the potential of gel-based prefractionation over more classical chromatography approaches. The reliability of the iTRAQ-mediated quantitative proteomic approach was also assessed using *Synechocystis* sp., *Sulfolobus solfataricus* (an Archaea) and *Saccharomyces cerevisiae* (an Eukaryotes). Based on multiple replicate analyses, biological variation was found to have the greatest impact on results, and thus biological replicates were recommended for inclusion in all future iTRAQ experiments. This investigation also revealed at least ±50% variation is necessary for significant protein expression to have occurred. The functional proteomic approaches in this thesis covered two main fundamental ideas: firstly to investigate the basic cellular response of this organism toward photoperiod (i.e. light-dark cycle); secondly, to examine the cellular stress response as a consequence of macro-nutrient deficiency, specifically phosphate starvation. The cellular response as a result of circadian influence was also studied via combining results from both transcriptomic and proteomic approaches; and possible interactions (e.g. post-translational modification) between mRNA and protein were also discussed.

Chapter 1

Introduction to the Thesis

1.1 Introduction

In this thesis, *Synechocystis* sp. PCC 6803 is used as a model organism. This cyanobacterium is known for its capacity in biological hydrogen gas (H_2) production [1-3], cyanophycin [4-6] and polyhydroxyalkanoate (PHA) [7-9] synthesis, as well as the ability to produce naturally occurred carotenoid [10-12] and tocopherol [13-15] compounds.

Biological hydrogen gas (H_2) production in cyanobacteria has been reported as a real alternative energy source to substitute for the near-depleted and costly fossil fuels, because this approach is pollution-free, highly efficient, clean, renewable and environmental friendly [16, 17]. It has been suggested that the cost of H_2 produced from a photo-bioreactor (US\$25/m^3) is significantly lower than the conventional photovoltaic water-splitting process (US\$170/m^3) [18]. However to achieve a complete change-over as the main energy provider, it is essential to focus research and development on increasing the rate of H_2 synthesis, as well as the total yield of H_2 produced [19].

Besides the energy demand, the increasing pressure on the availability of fossil fuel has also triggers doubt in petroleum-based polymers which in turn has caused the market to focus attention on the bio-polymer option, for example cyanophycin and poly-hydroxy butyrate (PHB) which are found mainly in *Bacillus* spp. and cyanobacteria [20]. Cyanophycin is a biopolymer and can be used as a substitute of polyaspartic acid as a biodegradable polymer [21]. Similarly polyhydroxyalkanoate (PHA) is another naturally produced polyester from cyanobacteria [22]. The 2005 world market for biodegradable polymers was over 110 million pounds (lbs), and the projection amount in 2010 is over 200 million lbs [23].

On the other hand, carotenoids and tocopherols are produced in cyanobacteria as an antioxidant compound, as well as a photo-protective agent during light acclimation [24, 25]. Besides their primary role during photosynthesis, they are also known to act as food colorants [26, 27], nutraceuticals, cosmetic ingredients, pharmaceuticals [28, 29], and anti-carcinogens in cancer treatment [30, 31]. The 2004 world market for carotenoids was estimated at over US\$ 880

million, and the projection figure in 2009 is over US$ 1 billion [32]. Facing the huge potential market in coming years, little is known about the regulation of the carotenoid and tocopherol pathways [33, 34]. Understanding the metabolic flux in these pathways would prove valuable to the metabolic engineering of microbial carotenoid (and tocopherol) production [35].

1.2 Objectives and Aims

The aim of this thesis is to gain a thorough and global understanding of the *Synechocystis* sp. PCC 6803 at the 'Systems Microbiology' level, by employing qualitative and quantitative proteomics, together with transcriptomic analysis to examine the cellular response toward various environmental stimulations. It is hope to gain a deeper understanding at both transcription and translation levels of these stimuli triggered through the induction of stress responses, specifically during light and dark acclimation (photoperiod), and phosphate starvation.

1.3 The Overview

(Chapter 2) This chapter introduces the background of *Synechocystis* sp. PCC 6803, as well as its potential biotechnology applications in alternative fuel, bio-degradable polymer and anti-oxidant compounds. The importance of Proteomics application in 'Systems Microbiology' level is also discussed.

(Chapter 3) One of the primary objectives is to seek (a solution) and develop a qualitative proteomic approach that will enhance (and increase) the identification percentages of proteins annotated in the genome. In the meantime it is also aimed to assess the feasibility of *Synechocystis* sp. PCC 6803 as a test subject in proteomic investigation. This is achieved through the employment of various prefractionation workflows, with a combination of gel-based and gel-free approaches at both the protein and peptide levels [36]. In this study, two tiers of separation were carried out in protein and peptide levels. A total of 6 protein-peptide prefractionation workflows were compared, with 1D polyacrylamide gel electrophoresis

(PAGE), weak-anion exchange (WAX) and isoelectric focusing (IEF) separation in protein level, followed by strong cation exchange (SCX) and IEF separation in peptide level

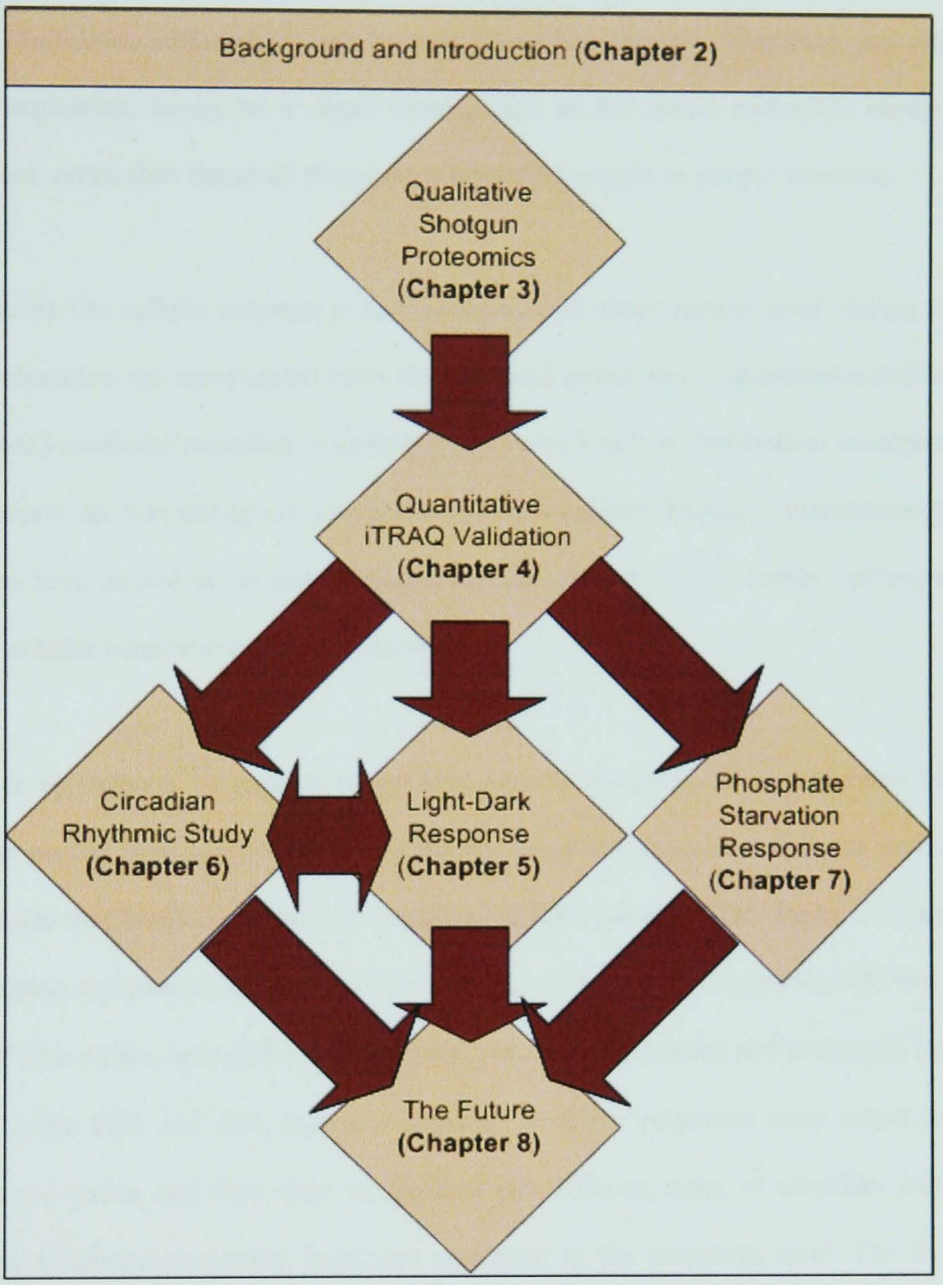


Figure 1.1. The overview picture of the designated investigations carried out in this thesis. The relationship between each chapters also illustrated in the diagram.

(Chapter 4) To gain an in-depth understanding of protein expression, iTRAQ-mediated quantitative proteomic is employed. This technique has the capacity to process up to 4 different samples simultaneously, using unique amine-reactive isobaric tagging system [37]. In order to assess the reproducibility of this quantitative approach, experiments were designed such that the

technical, experimental and biological variations from the samples could be estimated through replicate analyses [38] and multiple injections [39]. This is accomplished by comparing the sample variations from three different domains of life: eukaryotes (*Saccharomyces cerevisiae*), archaea (*Sulfolobus solfataricus*) and bacteria (*Synechocystis* sp.). Therefore, any changes in protein expression would be a direct consequence of the actual biological change in the experiment, rather than the small discrepancy caused by sample to sample variation.

(Chapter 5) The cellular response at both proteome and transcriptome levels during light (and dark) acclimation was investigated using the combined techniques of quantitative real-time PCR and iTRAQ-mediated quantitative proteomic. A central carbon metabolism reconstruction of *Synechocystis* sp. was attempted. This is the first time a global 'shotgun' quantitative proteomic study has been carried out in this organism, and hopefully it would provide further perception into the cellular behaviour during the photoperiod.

(Chapter 6) Under a 12 hour dark and light cycles, *Synechocystis* sp. exhibited the classic circadian profile. This study aims to establish the circadian response at both the transcriptional and translational levels as a result of alterations in light intensity. This study is designed such that the gene and protein profiles could be captured across a 24 hour period, with two different types of time-points, intended to highlight the transient (short-term) and prolonged (long-term) effects under light and dark regimes. Different circadian responses were noted among the investigated genes, and they were categorised into different types of circadian profiles. The circadian influence on protein level was compared to the transcript level. The discrepancy between these levels is also discussed.

(Chapter7) This chapter aims to investigate the cellular stress response when cells were subjected to phosphate starvation. This is achieved via the iTRAQ-mediated approach to study the global proteomic response of phosphate-depleted *Synechocystis* sp. cells. In total, three different phosphate-reduced levels were examined, and the proteome response during phosphate

starvation was noted for major metabolic pathways, as well as the phosphate stress-induced proteins.

(Chapter 8) This chapter summarises the main points and key findings from above. Future recommendations and suggestions are also included.

Chapter 2

Synechocystis sp. PCC 6803:

A Fuel, Biopolymer and Drug Factory

2.1 Cyanobacteria

Cyanobacteria are believed to be one of the earliest forms of life to have evolved on Earth, and thought to be responsible for the formation of the biosphere today [40, 41]. The early fossil record suggests that cyanobacteria can be traced back up to 3.5 billion years ago [40]. The ability to produce oxygen and carry out photosynthesis, thought to have evolved in these organisms and then proliferated into higher plants and algae, has made cyanobacteria one of the most fascinating microbial subjects in this, the post-genomic era [42]. They remain one of the largest sub-groups of Gram-negative prokaryotes, with 150 genera and well over 1,000 species having been described [43]. They are found almost in every conceivable habitat, from salt water, to fresh water, to bare rock, to soil and extreme environments. These organisms exist unicellular (refer to Figure 2.1), colonial or filamentous forms, and some of them have the ability to fix nitrogen gas (N_2) from the atmosphere [44].

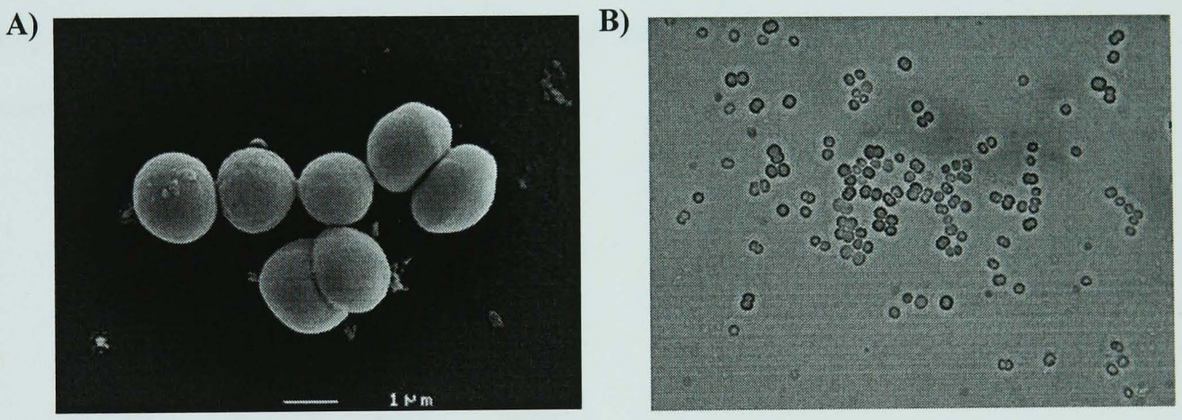


Figure 2.1. The unicellular cyanobacterium *Synechocystis* sp. PCC 6803 under (A) electron microscope, and (B) light microscope.

The capability of these organisms to produce secondary metabolites, such as cyanotoxins like Nodularin and Microcystin, has attracted some commercial and industrial interest [45]. At least one secondary metabolite, cyanovirin, has shown potential anti-HIV activity [46]. However there remain some noxious species that produce neurotoxins, hepatotoxins, cytotoxins, and endotoxins, which makes them harmful to humans and animals [47, 48]. In contrast, some species are available as food, such as *Arthrospira platensis* (*Spirulina*), where it has been

recommended as part of the diet in human food sources to reduce health risk [49, 50]. A more detailed review of marine natural products from cyanobacteria is covered elsewhere [45].

2.2 *Synechocystis* sp. PCC 6803

Synechocystis sp. PCC 6803 is a unicellular, non-nitrogen fixing cyanobacteriium (refer to Figure 2.1) which possess the physiological characteristics of the gram-negative bacteria, with the exception that it has a stronger and more rigid peptidoglycan outer layer [43]. Furthermore, *Synechocystis* sp. PCC 6803 is the first of its kind to have its entire genome fully sequenced (by the Kazusa DNA Research Institute, Chiba, Japan) [51]. This genome consists of approximately 3.6 million base pairs, a GC content of 47.7% (Figure 2.2) and an estimated 3,168 protein encoding genes. When the sequence was first released in 1996, approximately 45% of the sequences did not have any similarity with genes previously registered through GenBank and only 4.6% of the genome was found to be identical to reported genes [52]. However this did not stop researchers from embracing this organism as an ideal type strain for cyanobacterial research.

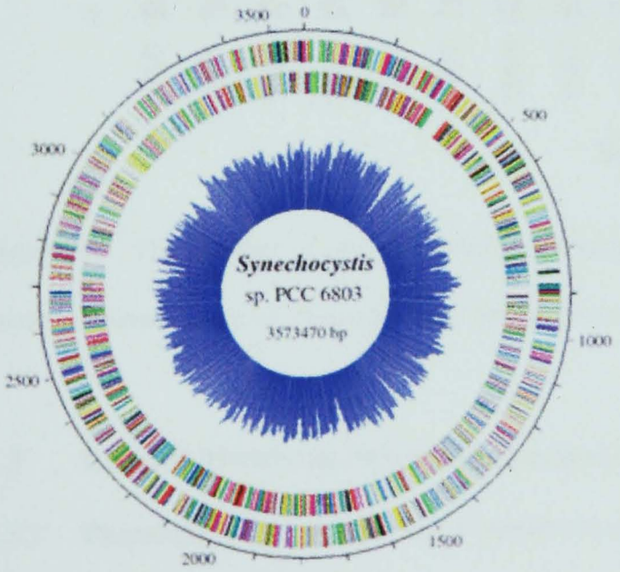


Figure 2.2. Image of the circular genome sequence clock of *Synechocystis* sp. PCC 6803 (reproduced from CyanoBase, http://www.kazusa.or.jp/cyano/).

Since the genome sequence was made available in 1998 [53], the total number of research articles related to both fundamental and applied research into *Synechocystis* sp. (using keyword search of '*Synechocystis*' in the title only) has exceeded more than a thousand publications,

based on a search of PubMed Central in the NCBI database (http://www.pubmedcentral.nih.gov/). Before 1996, the number of articles with *Synechocystis* sp. in the title was constant at below 50, with an average of 33 articles per year. Since then, a gradual increase was noted until a proliferation of publications began to appear in 2000 as indicated in Figure 2.3, where the article number rose to more than 80 per year. The highest record was in 2005 with nearly 100 articles per year. However, it must be reminded that the search was only based on the title of the article and restricted to the NCBI database, therefore the actual number of articles related to *Synechocystis* sp. is likely to be much higher than this.

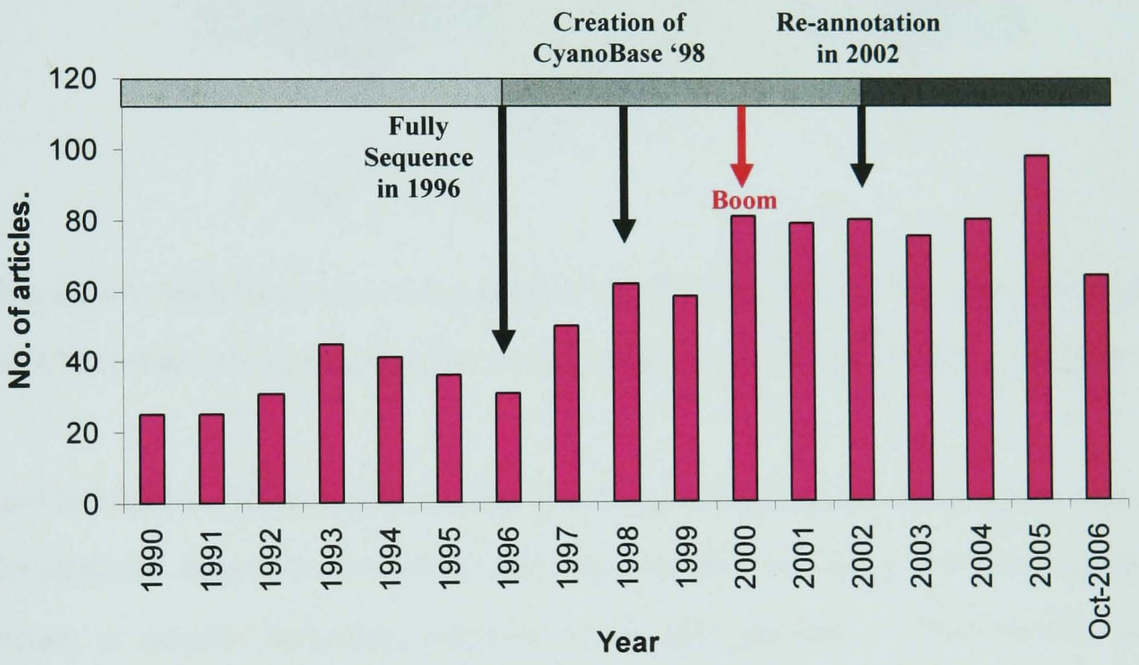


Figure 2.3. The number of research articles produced per year, retrieved from NCBI database, using keyword search of '*Synechocystis*' in the title only.

2.3 Central Metabolic States of *Synechocystis* sp. PCC 6803

2.3.1 Photoautotrophic – The 'light' metabolism

The most obvious feature of photoautotrophic metabolism in *Synechocystis* sp. PCC 6803 condition is the activation of photosynthetic processes, in particular the Photosystem I [54, 55] and Photosystem II [56, 57] reaction centres. Photosynthesis consists of two main reaction types, namely the 'light reactions' and the 'dark reactions' [58]. Energy from the light in the

form of photons is converted into ATP and NADPH through the light reactions; whereas the dark reactions, known as the Calvin cycle, convert CO_2 into hexose sugars and other organic compounds. The main fundamental difference between these two reactions is that the dark reactions do not depend on the presence of the light, whereas the light reactions do [58].

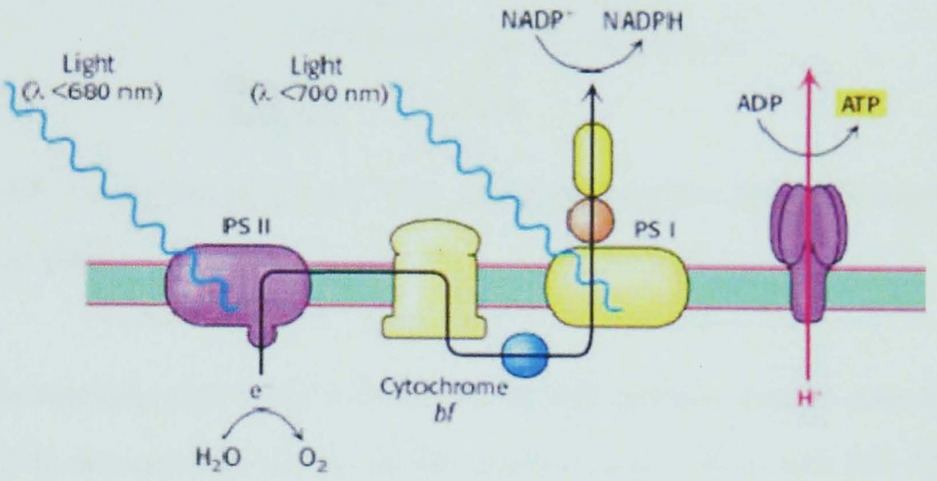


Figure 2.4 A simplified version of Photosystem I (PSI), Photosystem II (PSII), Cytochrome b_6f and ATP synthase response when exposed to light (adapted and modified from Berg *et al.*[59]).

Upon illumination, the chlorophyll and light harvesting antenna in cyanobacteria, known as the phycobilisome, absorb the light energy [58]. The cells then use energy from the absorbed photons to generate high-energy electrons in the light reactions of photosynthesis, as summarised in Figure 2.4. The electrons are derived by splitting two water (H_2O) molecules to produce one O_2 molecule and four electrons (e^-). These electrons are then used to regenerate NADP+ to NADPH. Similarly, the electrons also channel through an electron-transport chain, forming/creating a proton-motive force across the membrane. This proton-motive force then manoeuvres ATP synthesis to form ATP through the ATP synthase [58]. The photosynthesis process depends on the interaction between two Photosystem reaction centres. The high-energy electrons generated from the photosynthetic light reaction will first flow through Photosystem II, then through the membrane bound complex of cytochrome b_6f, before flowing through Photosystem I [58] (Figure 2.4).

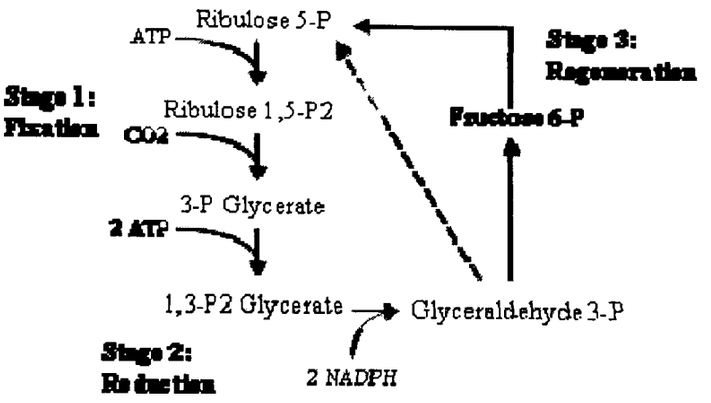


Figure 2.5. A simplified version of Calvin cycle in photosynthetic organisms (modified from Berg *et al.* [59]).

After the formation of ATP and NADPH from the light reactions, the cells enter the second phase of the photosynthesis process, the dark reactions, or the Calvin cycle [58]. The Calvin cycle can be divided into three stages. These are 'fixation', 'reduction' and 'regeneration' stages (refer to Figure 2.5). During the fixation stage, CO_2 is fixed from atmosphere by ribulose 1,5-bisphosphate to produce 3-phosphoglycerate. Then 3-phosphoglycerate is reduced to produce hexose sugar via the glycolysis/gluconeogenesis intermediates of glyceraldehyde 3-phosphate and fructose 6-phosphate during the reduction stage. At the final regeneration stage, ribulose 1,5-bisphosphate is regenerated from fructose 6-phosphate and glyceraldehyde 3-phosphate via a complex series of reactions, ultimately to maintain the continuation of the Calvin cycle (Figure 2.5). This cycle depends strongly on the substrate availability of ribulose-1,5-bisphosphate carboxylase (RuBisCO), the enzyme involved in the reduction stage of the Calvin cycle [58]. Furthermore this enzyme is highly regulated by light [60]. Some of the reactions in the Calvin cycle, where hexose sugar forms, are similar to those in the gluconeogenesis pathway (refer to Figure 2.6); with the exception that glyceraldehyde 3-phosphate dehydrogenase in chloroplasts is NADPH-specific rather than NADH [58].

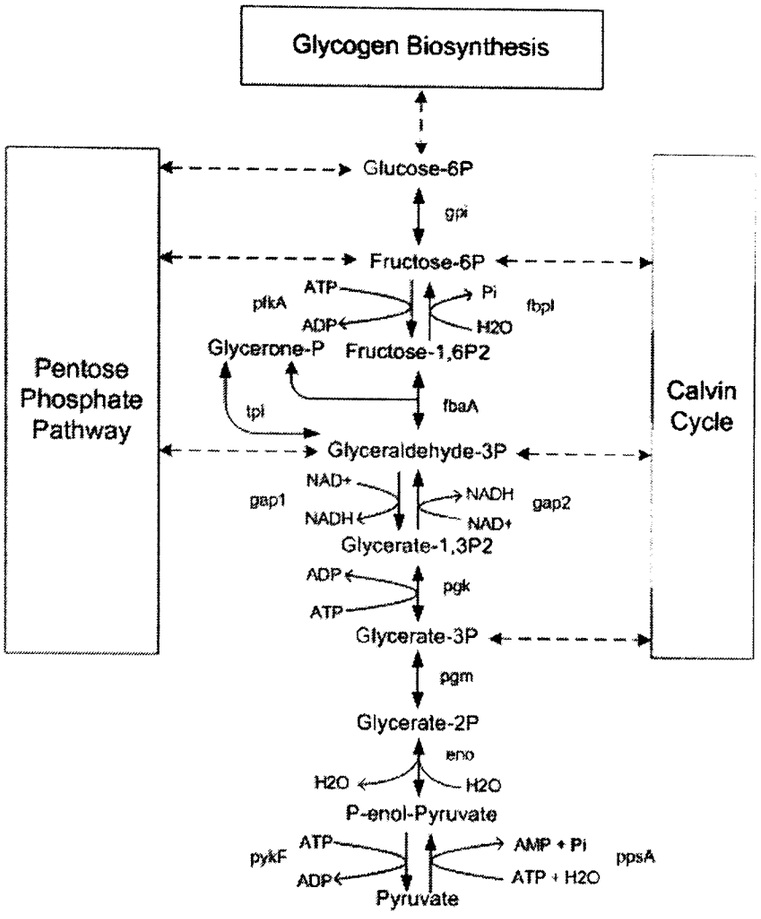


Figure 2.6 The *Synechocystis* sp. PCC 6803 glycolysis and gluconeogenesis pathways, as reconstructed from the KEGG (http://www.genome.jp/kegg/) genome annotation.

In a metabolic flux study carried out under mixotrophic (in light with an additional carbon source) conditions, it was reported that the Calvin cycle was significantly activated, suggesting that light is the preferred energy source for *Synechocystis* sp. PCC 6803 despite the presence of excess carbon sources in the growth environment [61]. However, most of the enzymes involved in glycolysis/gluconeogenesis metabolism did not show any significant alteration in the enzymatic activity when compared to the photoautotrophic or heterotrophic conditions [62]. Similarly at the transcript level, most genes involved in central energy metabolism did not show any significant changes relative to other nutritional modes [63].

2.3.2 Heterotrophic – The 'dark' metabolism

Glycogen is the main energy reserve for cyanobacterial cells, especially during dark conditions. It can be replaced by glucose because the degradation of glycogen begins with the conversion to

glucose-6-phosphate. It has been widely reported that *Synechocystis* sp. PCC 6803 is able to grow under photoautotrophic (with light but without an external carbon source), heterotrophic (without light but with an external carbon source) and mixotrophic (with light and external carbon source) conditions [64]. However, it is generally understood that *Synechocystis* sp. will not grow in sustained darkness with an externally supplied carbon source, unless a daily pulse of white light is provided. This condition is normally termed as light-activated heterotrophic growth (LAHG) [64].

The conversion of glycogen to glucose 6-phosphate and then to pyruvate occurs via the glycolytic pathway in photosynthetic organisms (Figure 2.6). The main function of glycolysis is to generate ATP, as well as glyceraldehyde 3-phosphate, the basic building blocks from which the organism can make a wide variety of compounds, at the expense of glucose. The rate of glycolysis is governed by several key enzymes, specifically phosphofructokinase (Pfk) and pyruvate kinase (Pyk) [58]. The NADH formed in the glycolysis process transfers its electrons to O_2 through the electron transport chain, thus regenerating NAD^+. In a similar manner to photosynthetic processes, the electrons from NADH also form a proton-motive force, and indirectly result in the formation of ATP through the ATP synthase. It is worth pointing out that the source of these high-energy electrons is different from oxidative phosphorylation (from NADH in the dark) and photosynthesis (from H_2O in the light) process [58].

Based on the enzyme assays of many key glycolytic and pentose phosphate pathway (PPP) enzymes, Knowles *et al.* reported that the activities of many of the enzymes was unaffected by growth conditions, except for glucokinase and pyruvate kinase where their activities were significantly higher in *Synechocystis* sp. in heterotrophic conditions [62]. In a separate study, Yang *et al.* found that the metabolic flux through the oxidative pentose phosphate pathway and major glycolytic pathways was considerably effected during heterotrophic conditions for *Synechocystis* sp. [63]. For example, glucose-6-phosphate dehydrogenase and 6-phosphoglunate dehydrogenase were highly active for glucose oxidation. In spite of this, the mRNA abundance levels of gene encoding for many key enzymes in central carbon metabolism were not

significantly altered by shifts in growth conditions [63]. Nonetheless, it has been well-documented that the enzymes within central energy metabolism tend to differentially regulate when the energy source (i.e. the growth condition type) available to the cyanobacteria cells is fundamentally altered [65-68]. On the other hand, a change to the heterotrophic condition always results in a deactivation of the photosynthetic pathways. This is observable at both the transcriptional and protein levels [63, 66, 67].

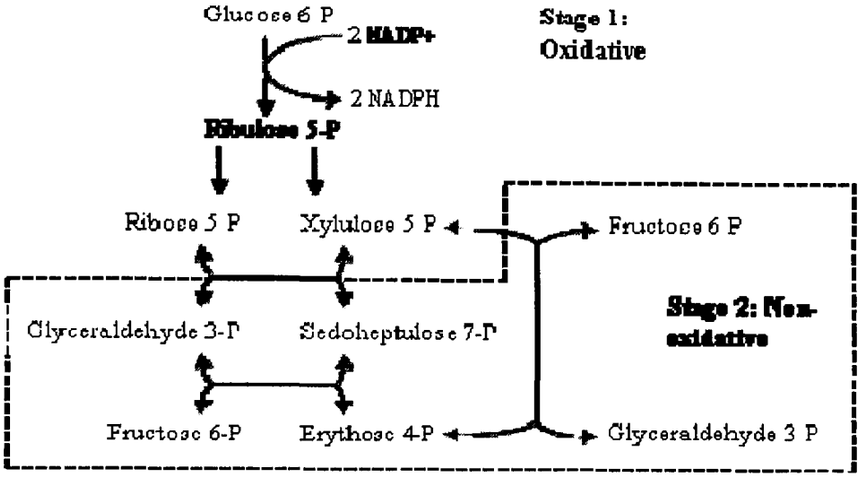


Figure 2.7. A simplified version of oxidative and non-oxidative pentose phosphate pathway (modified from Berg *et al.* [59]).

It is worth mentioning that many common enzymes and intermediates are shared between the Calvin cycle and the pentose phosphate pathway. In fact, the pentose phosphate pathway is the complete opposite of the Calvin cycle, which is used to break down glucose molecule into CO_2 thus to generate NADPH. When the growth condition changes, such as switching from light to dark, this pathway becomes inevitably important, since NADPH is required for the biosynthesis of many bio-molecules such as fatty acid and amino acids [58].

The pentose phosphate pathway comprises two stages, known as the 'oxidative' and the 'non-oxidative' stages (Figure 2.7) [58]. The generation of NADPH in the oxidative stage involves oxidation of glucose 6-phosphate to ribulose 5-phosphate, or ribose 5-phosphate. The second non-oxidative stage, on the other hand, is responsible for the conversion of 5-carbon sugars

(Figure 2.7). The 5-carbon sugars can be used in the synthesis of RNA, DNA or nucleotide co-enzymes [58]. The primary control factor in the oxidative stage is the concentration of NADP+ in the cytoplasm. In contrast, the non-oxidative stage is predominantly governed by the substrate availability [58]. Therefore, cyanobacteria have a very robust central metabolism between the Calvin cycle and pentose phosphate pathway when subjected to different nutritional modes [69].

2.4 *Synechocystis* sp. as a Fuel Factory?

One of the key features of *Synechocystis* sp., like any other cyanobacteria, is its capability to photosynthesise, i.e. to photo-convert CO_2 and H_2O into carbohydrates and O_2. The primary photosynthesis reaction is a water-splitting process with the generation of O_2 and electrons in the form of reduced NADP (NADPH) [70]. The electrons from NADPH can then be used for assimilation of CO_2, nitrates, sulfates and other nutrients. In the case of high 'reducing pressure' in cells, such as under anaerobic conditions, or limitation on electrons acceptors, the reduced NADP can be utilized via the hydrogenase reaction, involving the bi-directional hydrogenase [71]. Today, with the increasing demands on the identification of alternative energy sources to substitute for fossil fuels, hydrogen gas is amongst the top-rated possible source of such energy [17]. This may make cyanobacteria attractive as source organisms for the new touted hydrogen economy. The following paragraphs place cyanobacterium hydrogenase work into context and provide some basis for why *Synechocystis* sp. PCC6803 is an attractive organism for this process.

There are a few broad enzymes in cyanobacteria that are involved in hydrogen gas metabolism. These can be categorised into three groups: (1) the nitrogenases; (2) the uptake hydrogenases, and (3) the bidirectional or reversible hydrogenases [72]. In general, there are two systems involved in cell metabolism and operates with H_2, specifically a) nitrogenase and uptake hydrogenase for NO_3 assimilation; and b) bidirectional hydrogenase for regulation of redox state. The nitrogenase system (encoded by *nif* genes) is responsible for the nitrogen gas assimilation process in all nitrogen-fixing cyanobacteria, by catalysing the reduction of N_2 to

ammonia (NH_3), and simultaneously producing H_2 as a by-product. The uptake hydrogenase system (encoded by the *hup* gene), on the other hand, catalyses the consumption of H_2 produced from the nitrogenase system. While the bidirectional or reversible hydrogenase system (encoded by *hox* gene) has the capability to drive the H_2 reaction either forward (producing) or backward (consuming). Specifically, *Synechocystis* sp. PCC 6803 possesses the (NiFe)-dependent bidirectional hydrogenase system, which is encoded by 5 *hox* genes and two unknown open reading frames [73]. This is interesting, as hydrogen production in this organism is governed by a single set of genes, the *hox* operon that contains *hox*E, *hox*F, *hox*H, *hox*U and *hox*Y genes [73, 74], rather than in competition with either the nitrogenase or uptake hydrogenase systems as occurs in all nitrogen-fixing cyanobacteria.

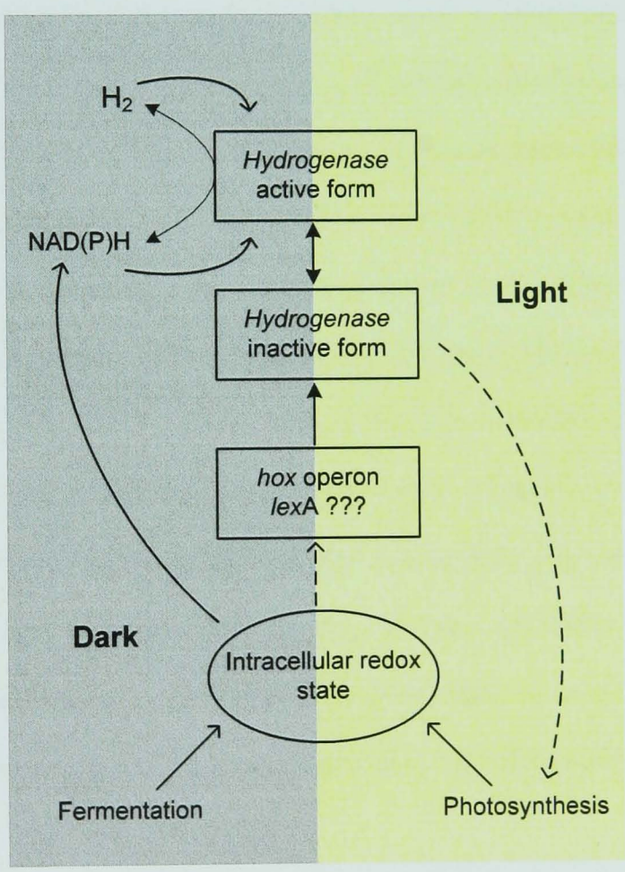


Figure 2.8. The proposed scheme of regulation of bidirectional hydrogenase (*hox* operon) expression in *Synechocystis* sp. PCC 6803. Closed arrow (➔) indicates reaction; open arrow (→) indicates regulation; open arrow with dotted line (----→) indicates putative regulation. This figure is adapted from Antal *et al.* with modifications [3].

The *hox* operon in *Synechocystis* sp. has been suggested as being part of the respiration system [75] and may also be involved in the photosynthesis process, as an electron valve [76]. Furthermore, an association with the fermentation process has also being proposed [77]. Yet, in fact, the actual functional mechanism for this operon has yet to be determined. A recent study of the *Synechocystis* sp. PCC 6803 bidirectional hydrogenase operon has proposed a tentative regulation mechanism as illustrated in Figure 2.8 [3]. Moreover, *hox* genes are known to be transcribed and expressed in both aerobic and anaerobic conditions [72]. Furthermore the *hox* operon also depends on O_2 and NAD(P)H availability, since the reaction is driven forward by NAD(P)H and repressed by O_2 [71].

The identification of the *lex*A promoter as a redox control and transcription regulator located in the promoter region of *hox*E in *Synechocystis* sp. [78, 79] has sparked an argument regarding to the actual functional role of this gene. It was thought previously that the *lex*A repressor in *Synechocystis* sp. is similar to the SOS-regulon in *Escherichia coli* (responsible for DNA damage repair system [80]) from the sequence annotation [51]. Then it was proven that this gene does not behave like the SOS-regulon, instead being essential for survival under inorganic carbon starvation conditions [81]. More recently, there was another suggestion that the LexA protein in *Synechocystis* sp. serves as a regulator of redox-responsive *crh*R (RNA helicase) gene expression [82]. Although there are many suggestions regarding the possible *in situ* functional role of *lex*A in *Synechocystis* sp., all researchers now agree that *lex*A is not involved in DNA damage repair system, and it is involved in the redox state of reaction, either in controlling hydrogen metabolism (*hox*) or in the separation of the double helix (*crh*R).

Despite all of these uncertainties, Cournac *et al.* used a *ndh*B deletion *Synechocystis* sp. mutant (M55), defective in the type I NADPH-dehydrogenase complex, and noticed a longer and more sustainable production of hydrogen gas, at a rate of 6 μmol of H_2 (mg of chlorophyll)$^{-1}$ h^{-1} when compared to the wild-type, with negligible hydrogen uptake and minimal amount of O_2 produced [71]. This study is the first indication that electron flux can be redirected to the *hox*

system using the M55 mutant. The idea of enhancing hydrogen production using metabolic engineering approaches has been deeply discussed elsewhere [83]. However an optimisation in hydrogen production does not often complement the natural operating system in the cell, which might cause some offset in the system, especially noticeable in growth rate retardation. Alternative methods such as pre-incubation of cells with pure H_2 [3] or pure methane (CH_4) [2] were also reported as effective in enhancing the hydrogen production rate by 2-fold and 4-fold respectively compared to the wild-type.

Further improvements in hydrogen production are also noticed when *Synechocystis* sp. PCC 6803 is grown under nitrate-limited conditions [3, 84]. A rate of 235.5 nmol H_2 $(\mu g\ Chl)^{-1}\ hr^{-1}$ is reported based on the *in vivo* methyl viologen measurements of *Synechocystis* sp. under nitrate-limiting conditions [1], compared to only 6 nmol of H_2 $(\mu g\ of\ Chl)^{-1}\ h^{-1}$ in nitrate-rich conditions [71]. It is not yet understood if *Synechocystis* sp. exerts similar response to those of nitrogen-fixing cyanobacteria, since *Synechocystis* sp. does not fix N_2 from the atmosphere. Furthermore, the incremental increase in hydrogen production under nitrate starvation conditions is not a direct consequence from the increase in transcript level [84], which further raises doubts regarding the role of nitrate, or the involvement of nitrogen metabolism during hydrogen production. Similar to nitrate deprivation conditions, sulphur-limitation also appears to enhance hydrogen production in *Synechocystis* sp. Antal and Lindblad reported that the hydrogen evolution rate in sulphur-deprived *Synechocystis* sp. cells is four time higher (measured at 90 nmol H_2 $(mg\ protein)^{-1}\ hr^{-1}$) than those supplement with sulphur in normal conditions [2]. *Chlamydomonas reinhardtii*, an alga, also shows an enhancement in H_2 production when grown under sulphur starvation conditions [85].

One common problem amongst all studies carried out so far is the difference in techniques used to obtain the rate of hydrogen production and hydrogenase activity (either *in vivo* or *in vitro*). This is especially noticeable in their units. It seems the methods vary according to researcher preference, with measurements normalised against either against total protein concentration (per mg of protein) or chlorophyll content (per mg of chlorophyll). This variation has made the

comparison between studies especially difficult, when a comparison of the hydrogenase production rate is needed, for example, between the sulphur-deprived (90 nmol H_2 (mg protein)$^{-1}$ hr^{-1}) [2] and the nitrate-limited (235.5 nmol H_2 (μg Chl)$^{-1}$ hr^{-1}) [1] condition. Besides this problem, it is also generally thought that *in vitro* hydrogen evolution rates as measured through the reduced methyl viologen method is widely accepted, however there exists several other techniques such as Clark-type electrode [86] and gas exchange mass spectrometry [71]. Without taking into account the bias in each instruments or methods, it is difficult to directly compare values at this time. The progress in this field has been sluggish over the decades, with only minor advancements.

2.5 *Synechocystis* sp. as Biopolymer Factory?

2.5.1 Cyanophycin

Cyanophycin (or multi-L-arginyl-poly-L-aspartic acid) is a water-insoluble polymer and a non-ribosomally produced amino acid polymer in cyanobacteria [87]. The purified form of cyanophycin can be converted into a polymer with a reduced arginine content [88], and becomes a substitute of polyaspartic acid as a biodegradable substitute for synthetic polyacrylate in various industrial processes [21]. With the increasing pressure on the availability of petroleum-based polymers, bio-based polymers have become of increasing interest to the biotechnology sector [20]. This biopolymer is synthesized from arginine and aspartate in an ATP-dependent reaction catalyzed by a single enzyme, cyanophycin synthetase [89] (refer to Figure 2.9). *Synechocystis* sp. PCC 6803 contains one cyanophycin synthetase, *cph*A and one intracellular cyanophycinase, *cph*B in the genome [51, 90]. It is generally thought that cyanophycin serves as a temporary nitrogen storage compound during the transition from exponential to stationary growth phases [91]. Berg *et al.* carried out a detailed study to understand the actual mechanism of cyanophycin production, and the specificity of the Cph enzymes in *Synechocystis* sp. [92]. They also discovered that the arginine residue can be replaced by the lysine residue, but at a much lower rate of incorporation. Since then many studies have been focused on the molecular characterisation of the (*Synechocystis* sp.) cyanophycin synthetase (*cph*) [87, 89, 90, 93, 94].

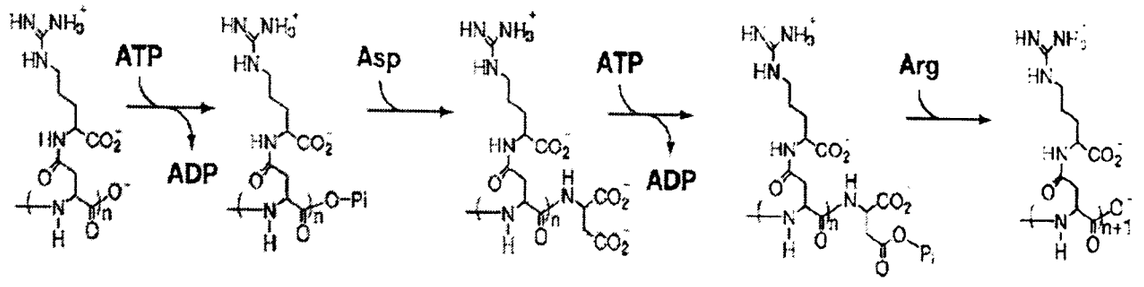


Figure 2.9. The proposed mechanism for cyanophycin production catalyzed by the cyanophycin synthase from *Anabaena* sp. 29413. This figure is adapted from Stubbe *et al.* [95].

The production of cyanophycin is reportedly a cellular response towards nutrient-deprivation (except nitrogen limitation) [96]. Evidence exists that when cells encounter chloramphenicol stress [97] or light limitation environments [98], the organism reacts by increasing the cyanophycin content. Interestingly, in *Synechocystis* sp., cyanophycin is synthesized mainly from the nitrogen source available in the growth medium, with a slower synthesis rate also observed from the degradation of cellular proteins [5]. However Kolodny *et al.* argued that the nitrogen incorporation is depended on the source of nitrogen (ammonium or nitrate), and if the cells are subjected to nitrogen starvation beforehand [99].

Zuther *et al.* investigated the accumulation of cyanophycin in a salt-sensitive mutant 549 of *Synechocystis* sp. PCC 6803 [100]. In mutant 549, a putative glycoprotease (*gcp*) deletion mutant was created and subjected to salt stress. In salt stress conditions, the amount of cyanophycin accumulated in the *gcp* mutant was approximately 40 times higher (21 mg per mg of protein) than in mutant 549 alone (0.5 mg per mg of protein). Besides salt-stress environments, cyanophycin accumulation in *Synechocystis* sp. is also noted in phosphate starvation conditions [101]. This study showed that the control mechanism of cyanophycin production and accumulation is governed by the stoichiometric ratio of arginine to the total nitrogen content in the cell, and suggested a complex interaction between cyanophycin synthesis and arginine catabolism in *Synechocystis* sp. This is in agreement with the suggestion that the rate of synthesis is depended upon the activation of a key enzyme in arginine biosynthesis, the

N-acetyl-L-glutamate kinase [102]. The studies so far have been focused on efforts to understand the governing factors in cyanophycin synthesis. However it still remained unclear if the presence or absence of the cyanophycin synthetase gene in the cell would cause any significant changes in the cellular level. This is answered by Li *et al.*, when they discovered that cells degrade the phycobilisomes as a nitrogen reserve in *cph*A and *cph*B deletion mutants in *Synechocystis* sp. [103]. From the observation above, they concluded that cyanophycin is not essential in the non-nitrogen fixing cyanobacterium, *Synechocystis* sp. compared to nitrogen-fixing cyanobacteria, since *Synechocystis* sp. could utilise phycobilisomes in the absence of cyanophycin.

The first large scale production (500 litre culture volume) of cyanophycin using *cph*A gene from *Synechocystis* sp. PCC 6803 was conducted in the recombinant *E. coli* strain DH1 [4]. A maximum cyanophycin content of up to 24% of cellular dry matter (CDM) was obtained. The study has been a success, and immediately provided the benchmark for the large scale fermentation process in cyanophycin production. Further improvements in cyanophycin production came in a study conducted by Elbahloul *et al.*, when protamylasse (a residual compound rich in carbon and nitrogen obtained from the production of starch) was incorporated into the media, and resulted into a maximum cyanophycin content of up to 28% of CDM [6]. Besides using *E. coli* as a host strain, Voss *et al.* also considered *Pseudomonas putida*, *Ralstonia eutropha*, *Bacillus megaterium* and *Corynebacterium. glutamicum* as candidate strains [104]. By using the *cph* gene from various cyanobacteria such as *Anabaena* sp. strain PCC 7120, *Synechococcus* sp strain MA19, and *Synechocystis* strain PCC 6308, the highest amount of enzymatic activity is obtained in *E. coli* as the host strain. However, despite this finding, the recommended host strains are *P. putida* and *R. eutropha*, because these gram-negative bacteria are cheaper in overall production cost, albeit they produce reasonable amount of cyanophycin (up to 20% of CDM) compared to *E. coli*. An example of cyanophycin accumulation in the *Synechocystis* sp. PCC 6803 is illustrated in Figure 2.10.

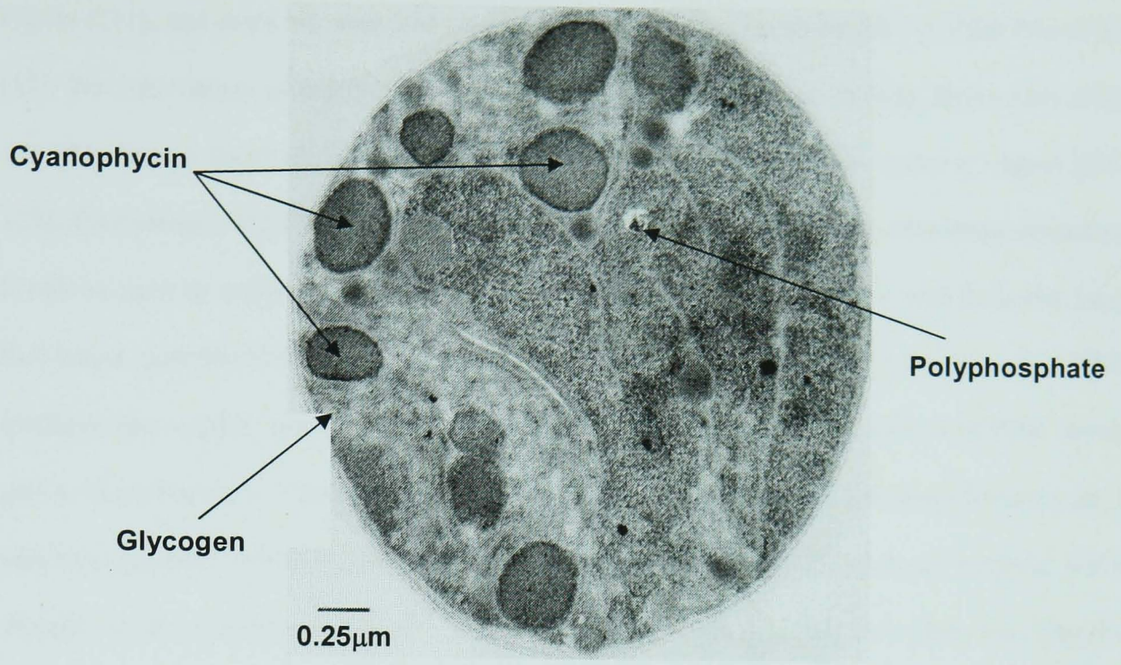


Figure 2.10. Electron microscopy image of the cellular structures of *Synechocystis* sp. during cyanophycin accumulation. This figure is adapted from Zuther *et al.* [100] with modifications.

Since then, further developments have been carried out using *P. putida* and *R. eutropha*. By using a polyhydroxyalkanoate (PHA)-deletion mutant with *cph* genes from *Synechocystis* sp. and *Anabaena* sp., a cyanophycin content of 17.9% of CDM was obtained in 30 litre cultures, which is claimed as the highest rate among the 30 litre scale [105]. Furthermore the study also concluded that the rate of synthesis is dependant upon extracellular arginine availability, oxygen limitation and the absence of PHA accumulation. Additionally, by using a KDPG-aldolase gene (*eda*)-dependent addiction system to express *cph*A from *Synechocystis* sp. PCC 6803, a tremendous amount of cyanophycin is seen accumulating in *R. eutropha* (host strain), with up to 40% of CDM [106]. The integration of the improved *cph* expression system and the host strain selection has seen a superb advancement in the process of cyanophycin synthesis, with *cph* genes in *Synechocystis* sp. PCC 6803 among the best available system in the field, and thus set the benchmark for others to follow.

2.5.2 Polyhydroxyalkanoate (PHA) and Polyhydroxybutyrate (PHB)

Polyhydroxyalkanoate (PHA) is another type of biopolymer that *Synechocystis* sp. PCC 6803 is capable of making. The most common form of PHA is polyhydroxybutyrate (PHB) (refer to

Figure 2.11), and there are over 140 possible ways of producing monomers of PHB from PHA [22]. The advantages of cheaper production cost and environmental friendly approaches using cyanobacteria as the main PHA production host have attracted much commercial interest [107, 108]. Furthermore, PHB is a biodegradable polymer derived from the bio-renewable resources. It can be used to replace million tonnes of plastic wastes generated each year from the fossil fuel-based polyethylene and polypropylene plastics. *Synechocystis* sp. contains two PHA synthase genes, *pha*C and *pha*E, identified in 1998 [109], followed by other two PHA family genes, encoding by a PHA-specific beta-ketothiolase (*pha*A) and an acetoacetyl-coenzyme A reductase (*pha*B), in 2000 [110]. The formation of PHA is found in almost all bacteria, and is thought to be a cellular response toward non-carbon based, nutrient-limited environments (similar to cyanophycin production) [111]. It has been suggested that PHA (or PHB) serves as an extra carbon reserve in the form of intracellular granules [107]. Yet, there have been observations of widespread variations in the occurrence of PHA synthases amongst cyanobacteria [112], with PHA accumulation can reach as high as 85% of the cell dry weight (CDW) [113]. In some gram-negative bacteria, specifically *R. eutropha*, the intracellular accumulation of PHA is reported at over 90% of CDW [113].

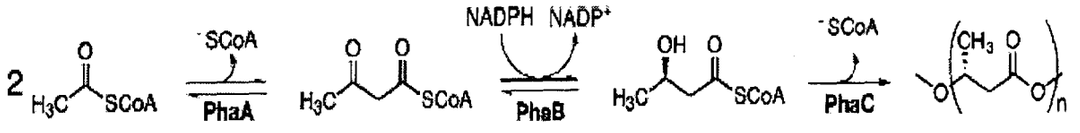


Figure 2.11. The proposed PHB synthesis pathway. This figure is adapted from Stubbe *et al.* [95].

In one of the earlier studies carried out by Wu *et al.*, the PHB content accumulated in cells is only 4.1% of CDW under nitrogen-starved condition, however it is found that high light favours the PHB production, whereas the presence of glucose during mixotrophic growth has a negative effect [7]. Since PHB is a form of carbon storage compound, Wu *et al.* modified the cellular carbon partition by creating an ADP-glucose pyrophosphorylase gene (*agp*) deletion mutant to

enhance PHB production in *Synechocystis* sp. [114]. Indeed the PHB accumulation rises to 14.9% of CDW, with an even higher accumulation of 18.6% in the presence of acetate. The effects of different carbon sources during photoautotrophic growth were further investigated by Sudesh *et al.*, where the study agrees with previous findings that PHA accumulation is higher in acetate supplemented medium at 7% of CDW compared to only 3% without acetate supplementation [8]. The study also concluded that PHA synthase is not the rate limiting step during production, as only a marginal increment is observed in the *Synechocystis* sp. mutant bearing the PHA biosynthetic operon from *R. eutropha* (11%) against the wild-type (7%). A very recent study, in July 2006, used pre-adapted *Synechocystis* sp. cells in glucose-supplemented medium, and subjected them to phosphate-limited conditions in the presence of acetate, resulted in a remarkable 6-fold increase in PHB accumulation against the wild-type, with up to 29% of CDW [9]. Similar effects of phosphate starvation on PHB accumulation has also been observed in *Nostoc muscorum* and *Spirulina platensis* [115]. From a proteomic analysis of recombinant *E. coli* producing PHB under phosphate starvation, the results indicated that the accumulation of PHB under phosphate-limited environment is triggered by a stress response [116]. The study also further verified that stress environments have a direct and positive impact on the PHA synthesis process.

The advancement in PHA synthesis (specifically refer to *pha* system in *Synechocystis* sp.) has grown steadily in recent years. A recently developed high-throughput screening technique using flow cytometry [117] promises to increase the speed and quality of mutant selection and production monitoring experiments, although progress in this field is still relatively slow when compared to the cyanophycin synthesis process as mentioned previously. Unfortunately it is still not possible to accumulate PHA and cyanophycin simultaneously in the same organism [105].

2.6 *Synechocystis* sp. as Drug Factory?

Synechocystis sp. PCC 6803 has fully functional carotenoid and tocopherol pathways which are capable of producing many carotenoid compounds, such as β-carotene, zeaxanthine, echinenone and myxoxanthophyll (refer to Figure 2.12); as well as many other tocopherol (Vitamin E)

compounds [11, 34, 118, 119] (refer to Figure 2.13). The main functional roles of these naturally occurred compounds are thought to be in the protection of lipid peroxidation in cyanobacteria [14]; serving as antioxidants during photosynthesis to prevent cellular damage as a result of free radical formation [11, 34], and other photosynthesis related processes, for example, Photosystem assembly, light harvesting, and photo-morphogenesis [120-122]. Besides the usual role as antioxidants, there is evidence that α-tocopherol is also involved in non-antioxidant functions [123], especially in regulating photosynthetic activity and macro-nutrient homeostasis [15].

The brief review here is purely focused on the carotenoid and tocopherol production from *Synechocystis* sp. The synthesis process leading to the geranylgeranyl diphosphate (GGDP), the key substrate in the carotenoid and tocopherol pathways, will not be discussed in depth. A full review of isoprenoid biosynthesis process is covered elsewhere [124]. Briefly, isopetenylpyrophosphate (IPP) and dimethylallylpyrophosphate (DMPP) are synthesized through methylerythritol 4-phosphate (MEP) pathway using glyceraldehyde-3-phosphate and pyruvate as main substrates. GGDP is then synthesized from IPP and DMPP through *crt*E (encoded by slr0739 in *Synechocystis* sp.). There is another possible route leading to isoprenoid biosynthesis, explicitly cytosolic mevalonic acid pathway and the plastidic mevalonate-independent pathway.

Some of the genes involved in these pathways are not essential during normal growth, except under stress environments, such as high light conditions as reported in a *crt*O deletion mutant [118, 125], in sll0418 methyltransferase [13], and in both tocopherol intermediates slr1736 and slr1737 [14, 126]. The presence of zeaxanthine and myxoxanthophyll, however, is essential in the photosynthetic electron transport and Photosystem II complex [118]. Myxoxanthophyll is also thought to be crucial in S-layer and thylakoid membranes formation [12]. Much effort has focused on the molecular characterisation and identification of the genes involved in the carotenoid and tocopherol biosynthetic pathways in these organisms, and this is summarised in Table 2.1.

Table 2.1. An up to date list of carotenoid and tocopherol production genes identified in *Synechocystis* sp. PCC 6803. Many do not have EC numbers, and a small number remain unannotated.

ORF	EC no.	Gene	Gene description	Reference
Carotenoid pathway				
slr1255	2.5.1.32	*crtB, pys*	phytoene synthase	[127, 128]
slr1254	1.3.99.-	*crtP, pds*	phytoene desaturase	[128, 129]
slr0940	1.14.99.30	*crtQ, zds*	zeta-carotene desaturase	[130]
sll1468	1.14.13.-	*crtR, bhy*	beta-carotene hydroxylase	[131]
slr0088	-	*crtO*	beta-carotene ketolase	[125]
sll0254	-	$crtL^{diox}$	lycopene cyclise and dioxygenase	[132]
slr1293	-	*crtD*	C-3',4' desaturase	[133]
sll0033	-	*crtH*	cis-to-trans carotene isomerase	[134, 135]
Tocopherol pathway				
slr0090	1.13.11.27	*ppd*	4-hydroxyphenylpyruvate dioxygenase	[136]
slr1736	2.5.1.-	-	homogentisate phytyltransferase (HPT)	[126, 137, 138]
sll0418	-	-	2-methyl-6-phytylbenzoquinone (MPBQ) methyltransferase	[139]
slr1737	-	-	tocopherol cyclase, VTE1 ortholog	[140, 141]
slr0089	2.1.1.95	*erg6, sed6*	tocopherol O-methyltransferase	[142]
slr1652	-	-	Phytol kinases	[143]

In 2004, Ryu *et al.* demonstrated that the expression of *crt*B, *crt*P, *crt*Q and *crt*R are light-dependent, activated in the dark through glucose induction, and controlled by cytosolic pH rather than a redox or glucose sensing mechanisms [144]. This was further illustrated when *crt*P and *crt*Q were found to be involved in *psa*AB phylloquinones of the Photosystem I electron transport chain [145]. Lagarde *et al.* used genetic engineering approaches to alter *crt* expression in *Synechocystis* sp., attempting to obtain an optimum yield of zeaxanthine [146]. At least a 50% increase in zeaxanthine yield was noticed in the mutant strain compared to the wild-type, when *crt*P and *crt*B genes were over-expressed. At the same time, a 2.5-fold increment was also noted in over-expression of *crt*R. While Qi *et al.* introduced the *nir*A promoter from

Synechococcus sp. PCC 7942 into the tocopherol system of *Synechocystis* sp. PCC 6803, and observed an increase in tocotrienols by up to 5-fold (which comprised at least 20% of the total tocopherol content) [147].

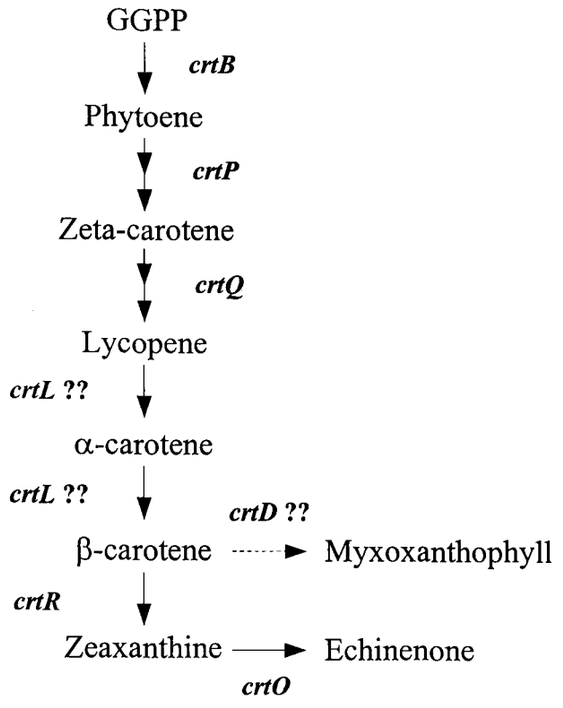


Figure 2.12. The simplified carotenoid biosynthetic pathway in *Synechocystis* sp. PCC 6803. This figure is modified from Lagarde *et al.* [146].

Unlike the progress in cyanophycin and PHB biosynthesis in *Synechocystis* sp., development in the carotenoid and tocopherol area is still in its infancy, with new genes/intermediates still being discovered daily. The proposed mechanism of these pathway are available (refer to Figure 2.12 and 2.13), however relatively fewer studies have been focused on the possibility of enhancing the production of the desired tocopherol or carotenoid compounds. The rates of limiting step(s), controlling enzyme(s), as well as the influence of physiological parameter(s) are still largely unknown.

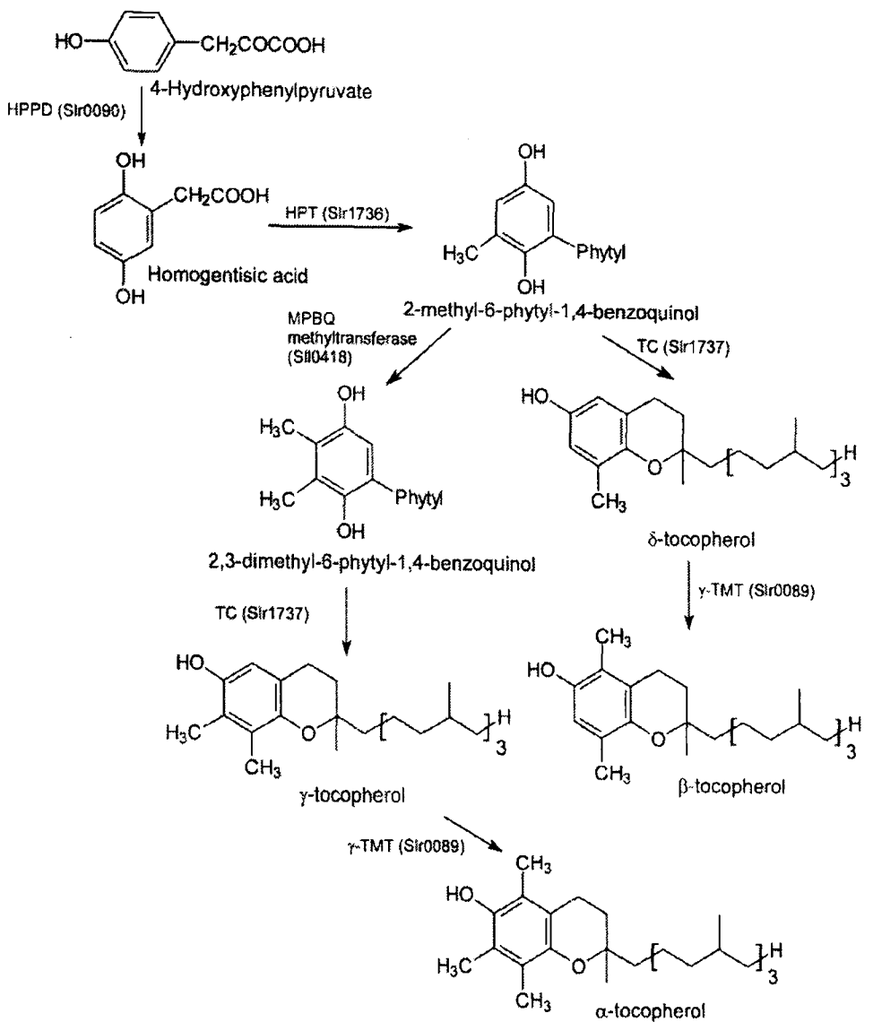


Figure 2.13. The proposed α-tocopherol biosynthetic pathway. This figure is adapted from Sakuragi *et al.* [15].

2.7 All 'Factories' Are Triggered By Stress.

In nature, all living organisms on the planet adapt to the environment by constantly amending their molecular function and structure through certain forms of control mechanism. This form of control can exist in the form of post-transcriptional modifications [148], post-translational modifications [149, 150], or other types of functional machinery. One particular scenario that would trigger a cellular response is stress induction [151].

2.7.1 Environmental stress (light, heat and salt)

The most common stress exerted on cyanobacteria is light stress. Light energy is absorbed and processed through the Photosystem I and II complex located at the thylakoid membrane of the

cell [58]. The transcript level of these photosynthetic genes accumulate in light [152], and are reportedly controlled by the oxygen-dependent acceleration of electron transfer from Photosystem I to $NADP^+$ due to activation of the Calvin cycle, as well as the retardation of electron flow between two Photosystems governed by a trans-membrane proton gradient [153]. However, excessive exposure to high light conditions (reported at more than 1000 $\mu E\ m^{-2}\ s^{-1}$) can cause a total destabilisation in the D1 reaction centre of the Photosystem II complex, which ultimately leads to cell death [154, 155]. The cell is constantly changing its photosynthetic content upon high light acclimation, and it has been suggested that this is carried out by modulating the Photosystem I and II content stoichiometrically [54, 156]. The Photosystem stoichiometry adjustment is thought to be modulated through the down-regulation of electron transfer, rather than maintaining photosynthesis efficiency [157]. A more detailed review of cyanobacteria response toward light is described elsewhere [158].

At least 25% of *Synechocystis* sp. genes exhibit some form of response toward the light, and as a result of this work, a 30 min intervals are recommended to accurately capture their transcription profile [66]. Yet, changes in transcript is reported to occur as fast as 15 min after exposure in genes involved in light absorption and photochemical reactions [65]. Different colours of light also trigger different response from the system. According to Hubschmann *et al.*, red light enhances the gene expression of those genes involved in transcription, translation, and photosynthesis, whereas far red light induces the stress-related genes [159]. On the other hand, blue light stimulates the cAMP in photo-signal transduction in cyanobacteria [160, 161]. The use of very strong UV light, however, is characterised as harmful to *Synechocystis* sp., as the UV inhibits the repair process of the Photosystem complex [162, 163]. A more detailed study to understand the global transcription response under UV light was conducted by Huang *et al.*, and apparently the induction of D1 protein recycling and coupling are synchronised between a decrease in phycobilisome biosynthesis and an increase in phycobilisome degradation [164].

The high thermal energy from high light environments is dissipated through the membrane-bound chlorophyll antenna of Photosystem II, the phycobilisome [122, 165]. Further recently, the water soluble orange carotenoid protein, encoding slr1963 in *Synechocystis* sp., was also identified as being involved in this high energy dissipation process by mediating the amount of light energy from phycobilisome to the Photosystem complex [166]. Under light stress, the Photosystem repair system is mediated through FtsH (cell division protein) complexes [167, 168], with the *isi*AB operon (iron-stress induced chlorophyll-binding protein) acting as the photo-protective agents [169]. The FtsH proteases also reportly plays a key role during heat stress [170]. Among many stress-related genes, *gro*EL, *gro*ES, *dna*K2, *dna*J3, *clp*B1 and *clp*P1, are reported as light-dependent as seen in their enhanced transcript level upon exposure to high light [171]. Cyanobacteria, generally, are well-equipped with many cellular defence systems that help to avoid any serious damage as a result of stress induction. One particular example has been demonstrated previously, the 'detoxification' process carried out by the tocopherol and carotenoid pathways [11, 14, 33]. The expression of superoxide dismutase SodB is also reported as light-dependent [172], and is useful in scavenging the free radical as a result of oxidative stress [173]. Other anti-oxidative stress system includes the type II and 1-Cys peroxiredoxins [174].

Interestingly, light is also reported as a medium in modulating the heat shock proteins during a heat shock response [175]. It is unsurprising since multiple chaperonins play multiple physiological roles under stress conditions [176, 177]. They are important during heat stress, specifically to prevent irreversible protein aggregation and to facilitate protein renaturation [178, 179]. Some common genes show enhanced transcript levels are *gro*ESL1, *gro*EL2, *htp*G, *hsp*A, and *clp*B1 [180]. Heat induction is commonly 'sensed' through the thylakoid membrane before the heat shock proteins can be activated [181]. It is necessary for the heat shock proteins to disassemble before releasing their substrates, in order to achieve oligomeric stability for cellular function [182]. Besides heat response, *hsp*17 is thought to be involved in stabilizing the physical state of the lipid membrane [183-185]. Similarly, *hsp*16.6 can be induced by exposure to salt, sorbitol, hydrogen peroxide, and high light [186]. It is thought some of the heat shock

genes might be regulated through Hik34 (a part of two-component sensor histidine kinase) [187]. This is further verified by Slabas *et al.* where a *hik*34 deletion mutant has caused an elevation in levels of heat shock proteins under both non-heat and heat related shock conditions [188]. SigB, SigE (both are RNA polymerase group 2 sigma factor) and HrcA (heat-inducible transcriptional repressor) are also reported as transcription regulators for a variety of heat shock proteins to various degrees [189].

Salt stress has a distinctive signature in *Synechocystis* sp., with a set of unique genes that are 'attached' to it [190]. When cells are subjected to high salt content, the accumulation of glucosylglycerol [191-194] and sucrose [195, 196] are essential in protecting the cells from excessive osmosis stress. Furthermore, unsaturated fatty acids in the membrane lipids are also suggested as being involved in the tolerance of the photosynthetic machinery to salt stress [197-199]. High salt prevents the *de novo* synthesis of Photosystem II proteins [200, 201]. Apparently iron metabolism is also affected by the high salt concentration, as seen in the over-expression of the *isi*AB operon [202]. Induced mutation of the *gcp* gene, which encodes a putative glycoprotease, has been seen to cause a reduction in salt tolerance of *Synechocystis* sp. PCC 6803, however, this stress also resulted in the accumulation of cyanophycin and carotenoid compounds [100]. That is just another example of substrates induction by environmental stress. Many salt-induced genes/proteins have unknown function, and they tend to appear as either periplasmic proteins [203] or hypothetical proteins [204]. Interestingly, in another study, all periplasmic proteins identified contain a signal peptide, with majority of them with a Sec-dependent signal peptide [205]. Marin *et al.* discovered some of the histidine kinases, Hik16, Hik33, Hik34, and Hik41 are involved in the perception of salt stress as well as the over-expression of some heat shock genes [206, 207]. Among the heat-shock genes, *hsp*A, *gro*EL2, and *dna*K2 are reportedly salt-induced [208]. In a large-scale proteomic investigation on soluble proteins, Fulda *et al.* compared the short and long term effects of salt acclimation in *Synechocystis* sp., and suggested the possibility of some form of post-transcriptional modification involved in the salt-induced genes/proteins [209]. A similar proteomic study was also conducted based on the plasma membrane of *Synechocystis* sp. [210].

2.7.2 Nutrient depletion stress (N, S, Fe, P)

In many cases, gradual nutrient depletion in the growth medium has been used to encourage the target substrate(s) to produce the desired product(s), with some degree of enhancement, but without compromising other parameters required for the cell growth as seen in many examples before (Section 2.4, 2.5 and 2.6). This approach, sometimes, also can be used to investigate the cellular responses toward the depletion environment, when the researchers do not have specific targets in their experiments [211-213]. However, it can, sometimes, produce a negative result. For example, when cells are subjected to Na^+ stress, the oxygen evolving site present in the ion transfer channel, as well as the reaction centre of Photosystem II, are affected and damaged [214].

Furthermore, nitrate and bicarbonate uptake appears to be regulated by the cytoplasmic substrate-binding proteins NrtC and CmpC [215], whereas the control of the global nitrogen regulator, *ntc*A and other nitrogen-related genes (such as *nrt*) are governed by multiple group 2 and group 1 sigma factors [216-218]. It has been suggested that the cell regulates nitrogen metabolism by sensing intracellular 2-oxoglutarate levels [219]. In a separate study, the deletion mutant of *pam*A (a PII-binding protein in *Synechocystis* sp. PCC 6803) affected the transcript level (down-regulation) of the nitrogen-related genes, including *nrt*ABCD [220]. Interestingly, the glycolysis genes, such as *gap*1, *zwf*, and *gnd* also consequently decreased their expression in the *pam*A deletion mutant. There have been suggestions that inorganic carbon and nitrogen metabolisms are co-regulated by the phosphorylation of PII proteins, triggered through the redox state of the cells [221, 222]. During nitrogen starvation, the degradation of pigment proteins (the 'bleaching' effect) is noticeable, specifically the phycocyanin and allophycocyanin subunit [223, 224]. Furthermore, it is believed that cyanobacteria utilise the phycobilisome when a nitrogen source is not readily available [103]. Besides phycobilisome, ribulose-1,5-bisphosphate carboxylase (Rubisco) has been reported as a nitrogen storage compound that is activated during nitrogen starvation [225]. Similarly, two of the glutamine synthetase genes (*gln*A and *gln*N) [226-228] and the $NADP^+$-isocitrate dehydrogenase gene (*icd*) [229] are also

reportedly regulated in response to nitrogen-limited conditions. The effect of nitrogen depletion on the hydrogenase (*hox*) system [84] and cyanophycin production [99, 103] have been discussed in Sections 2.4 and 2.5.

These 'bleaching' effects on the phycobilisome proteins, however, does not occur under the sulphur-deprivation process [230]. The main sulphur transport system in *Synechocystis* sp. is encoded by the *suf* operon, which is composed of *suf*B, *suf*C, *suf*D, and *suf*S [231], with *suf*R as the transcriptional repressor [232]. Despite the existence of the *suf* operon in the genome, relatively little is known about their functional regulation [232]. Furthermore, the transcript level of *nbl*A (for phycobiliproteins degradation) and *sbp*A (sulfate binding periplasmic protein) in *Synechocystis* sp. have been shown to accumulate in response to nitrogen and sulfate limitation [233, 234]. Additionally, sulphur deprivation also has been reported to increase the biological hydrogen production in *Synechocystis* sp. [2]. The correlation between iron and sulphur has been well described in the iron-sulphur centre of the Photosystem I [235-237]. The iron-sulphur protein has been an essential element during cyclic electron flow around Photosystem I, and during photo-heterotrophic growth [238].

Iron-depleted conditions can be measured by the induction of iron stress-induced protein (IsiA) [202, 239, 240], as well as the delayed rate of biogenesis in Photosystem II [241], together with the impairment to photosynthesis apparatus [242]. Interestingly, it has been suggested that *isi*B, which encodes flavodoxin, is not essential during iron limitation [243]. Further, it seems that iron uptake in *Synechocystis* sp. is governed by the (ABC)-type ferric iron transporter, encoded by *fut*A1, *fut*A2, *fut*B, and *fut*C genes [244]. Apparently the natural iron source in the *fut* system is not ferric citrate, as the high level of citrate actually inhibits iron uptake rate [245], which is contradicting to previous findings that suggested otherwise [246]. As much as 50% of the cellular iron content is stored in the bacterioferritin proteins (Brf), and these proteins play an important role in iron homeostasis [247]. Singh *et al.* carried out a genome-wide microarray study of iron-starved *Synechocystis* sp. cells, and found 85 genes with significant regulation under iron stress conditions [213]. Amongst the common iron-dependent genes, the expression

of *cpc*G, *psb*C, *psb*O, *psa*A, *apc*ABC, and *cpc*BAC1C2D increased, while that of *isi*A, *idi*A, *nbl*A, *psb*A, and slr0374 decreased [248]. On the other hand, a *psa*FJ deletion mutant is reported to exhibit similar behaviour as those seen during iron deprivation conditions, and suggests that oxidative stress created by the Photosystem I mutant triggers the induction of iron-dependent gene expression, specifically seen in the *isi*AB operon [249].

There have been relatively fewer investigations focused on the phosphate starvation response in *Synechocystis* sp. PCC 6803. The phosphate sensing system in *Synechocystis* sp. is modulated by the *pho*R (sll0337) and *pho*B (slr0081) regulon [250]. Inactivation of *pho*U, one of the phosphate transport system regulatory proteins, has proven to cause disruption in both inorganic phosphate (Pi) uptake and removal rates [251]. Under phosphate starvation conditions, significant over-expression is also observed in *ppa* and *ppx* genes, which are involved in the hydrolysis of Pi [252]. Suzuki *et al.* carried out one of the most comprehensive studies to date, based on microarray analysis, of the phosphate starvation response in *Synechocystis* sp. [253]. In that study, *pst*1 and *pst*2 genes (which encode the ABC-type phosphate transporter), as well as the alkaline phosphatase (*pho*A) and extracellular nucleus (*nuc*H) were significantly over-expressed. Indeed one of the most common features during phosphate limitation is the up-regulation of *pho* genes, as seen in *Bacillus subtilis* [254, 255] and *E. coli* [256, 257]. There has been suggestion that high light also triggers the expression of the *pho* regulon, since light accelerates phosphate assimilation yet there is limited cell capacity to perform the task [258]. The phosphate starvation response in cyanophycin accumulation has been reported in Section 2.5 and elsewhere [101].

2.8 Proteomic Analysis of *Synechocystis* sp. PCC 6803

2.8.1 2-Dimensional electrophoresis (2DE)

Soon after the complete sequencing efforts of 1996 [51], the first 2-dimensional electrophoresis (2DE) gel was carried out in the Kazusa DNA Research Institute, with a total of 130 proteins successfully identified via this technique when coupled with MALDI-TOF peptide mass fingerprint techniques [259]. 2DE has been described as a powerful whole-cell proteomics tool,

used to visualise protein spot on the gel, based on both isoelectric focusing point (p*I*) and molecular weight (Mw) [260-262]. One of the advantages of 2DE is the ability to detect post-translational modifications [263], for example, phosphorylation [264, 265] and methylation [266]. However there are limitations with this technique. Proteins with extreme size (i.e. less than 10 kDa or bigger than 300 kDa), extreme pI (less than p*I* of 3 or more than 10), and extreme solubility (specifically membrane and high hydrophobicity proteins) create problems for 2DE [267, 268]. Despite many criticisms, most of the proteomic studies in *Synechocystis* sp. have to date been carried out using this technique.

The continuing efforts of the Kazusa DNA Research Institute saw the total number of proteins identified via 2DE increases to 227 from 234 gel spots [269] (refer to Figure 2.14). Although the study also focused on the soluble, insoluble, thylakoid membrane, and secreted protein fractions, relatively fewer proteins were identified from thylakoid membrane. Focusing on the soluble fraction, Simon *et al.* employed narrow-range zoom gels, and successfully identified 105 proteins with 37 novel proteins not previously found using 2DE [270]. On the other hand, Wang *et al.* concentrated the 2DE approach purely on thylakoid membranes, and found 51 previously unknown proteins [271]. Herranen *et al.* also attempted 2DE on membrane complexes, however only 20 proteins were discovered [272]. A combination of sucrose gradient and two-phase partitioning extraction protocols, developed by Norling *et al.* [273], has seen an improvement in the isolation of outer membrane [274] and plasma membrane proteins [275] from *Synechocystis* sp., with 49 and 57 proteins found respectively. Srivastava *et al.* also used the same protocol (sucrose gradient) to extract thylakoid membrane proteins, and successfully identified 76 proteins [276]. Similar efforts also resulted in the discovery of periplasmic proteins in *Synechocystis* sp., while Fulda *et al.* applied 2DE to investigate the proteomic profile of *Synechocystis* sp. under different salt concentrations [205]. Many other studies also concentrate their proteomic approach on using the 2DE technique, such investigations shed light on the proteome response of *Synechocystis* sp. PCC 6803 to acid-stress [277], heat-stress [180, 188], salt-stress [209, 210] and light-dark cycles [67, 278].

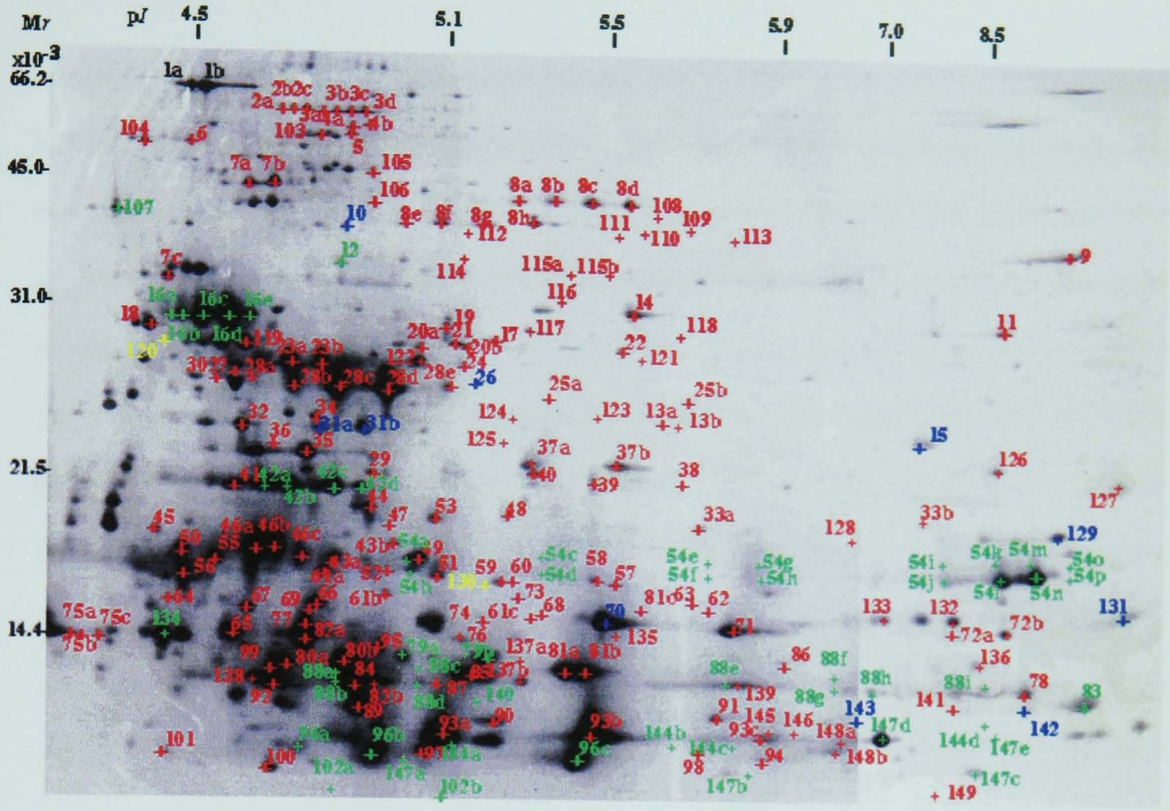


Figure 2.14. The 227 protein spots identified from *Synechocystis* sp. This gel image is extracted from Cyano2Dbase [269].

2.8.2 Non-gel based "shotgun" proteomic

In recent years, there has been an increasing interest in proteomics methods that avoid 2D gel electrophoresis for protein separation, with the goals of improving detection of the more difficult to detect acidic, basic, and hydrophobic proteins, and in order to increase reproducibility and throughput [267, 279-283]. These gel-free ("shotgun") approaches rely on the simplification of a protein or peptide mixture before it is analyzed in a mass spectrometer, with separation on a C18 reverse-phase (RP) liquid chromatography column being typical. To date, a variety of methods have been used in multidimensional separation schemes to fractionate peptides prior to this step. Reported peptide separation methods include strong cation exchange (SCX) chromatography [284, 285], slab-gel isoelectric focusing [286], capillary isoelectric focusing [287], liquid phase isoelectric focusing [288] and capillary electrophoresis [289, 290]. A possible advantage of choosing SCX separation is that it may be implemented in an on-line manner with RP as long as a washing step is used to remove salts.

Although these dual peptide fractionation schemes are effective in simplifying samples before MS analysis, whole-cell protein extracts are complex enough that an additional separation step can significantly increase the number of identified proteins. Rather than performing this separation at the peptide level, researchers have experimented with fractionating proteins prior to digestion with a protease [291, 292]. Towards this end, separations by size exclusion chromatography (SEC) [293-296], hydrophobic interaction chromatography (HIC) [297], RP chromatography [298, 299], Weak (WAX) and Strong (SAX) anion exchange chromatography [300-303], liquid- and gel-phase isoelectric focusing [304-306], and 1-D PAGE [307, 308] have been reported. Use of these multi-level separation schemes can result in the identification of many proteins (e.g., more than 1100 unique proteins identified from a total of 2071 peptides, in the case of human serum [309]); however, there have been few studies [310] in which these various options for protein and peptide separation have been compared. More separation steps have the potential to yield simpler peptide mixtures at the MS inlet, allowing for the deconvolution of 'flying' peptides through the flow streams in mass spectrometers, but also increase workload and sample loss.

As will be discussed later in this thesis (see chapter 3), the only 'shotgun' proteomic study available to date specifically on *Synechocystis* sp. PCC 6803 was conducted by Gan *et al.* in 2005 [36]. In this study, a total of 6 protein and peptide pre-fractionation workflows were compared, with different fractionation techniques compared at each level. SCX and iso-electric focussing (IEF) were used for peptide fractionation, whereas 1D gel, WAX and IEF were employed at the protein level. In all, 776 unique proteins were confidently identified out of 1000+ candidate proteins. The study also concluded that the gel-based fractionation workflow, i.e. IEF and 1D gel, produced a better separation and resolution profile compared to the various chromatography methods investigated. However the processing time (in gel-based techniques) was significantly longer.

2.8.3 Quantitative proteomics

The information derived from qualitative proteomic approaches is considerable, yet limited to knowledge of presence or absence of the proteins of interest [311]. While the processing time spent on qualitative approaches is almost identical in the quantitative experiments, there is no doubt that the gain from protein quantification has significant advantages over the protein identification alone. Together with advancements in mass spectrometry, there are various emerging techniques introduced to quantifying proteins, such as stable isotope labelling of amino acid in cell cultures (SILAC) [312-315], isotope-coded affinity tags (ICAT) [315-319], metabolic labelling (^{15}N or ^{13}C) [320, 321], mass-coded abundance tagging (MCAT) [322], and the use of CyDyes in a two-dimensional electrophoresis (2-DE) methodology [323-325].

One of the most exciting inventions in quantitative proteomics is the introduction of isobaric tags for relative and absolute quantitation (iTRAQ) in 2004 [37]. The principle of iTRAQ focus on the use of amine-reactive isobaric reagents, by tagging them onto the N-terminal and lysine side chains of digested peptides. The iTRAQ reagent is comprised of three main sections, the reporter ion (ranging from 114 to 117 m/z), the balance group (ranging from 28-31 m/z), and the amine specific reactive peptides (refer to Figure 2.15). The tag always has a nominal mass of 145 (from the combination of reporter ion and the balance group), and is thus isobaric. For the reader's information, a detailed review of the shotgun proteomics approaches using iTRAQ is covered elsewhere [326]. The complementary nature of the iTRAQ workflows with cancer [327], neuroscience [328], Alzheimer's disease [329], and in post-translational modification studies [330], has also been discussed elsewhere. Briefly, the protein samples are reduced, alkylated and digested (with trypsin) separately, prior to labelling with iTRAQ reagents. In total, up to 4 (and soon 8) phenotypes can be processed simultaneously. The labelled samples are then subjected to SCX fractionation, before being analysed in the mass spectrometer. When the peptide, which is attached onto the amine specific reactive peptides, enters the collision cell, fragmentation causes the reporter ion to uncouple from the peptide, together with the neutral loss of the balance group. The reporter ion is then used to provide the peptide quantitation ratio, whereas the peptide sequence undergoes peptide identification as illustrated in Figure 2.15.

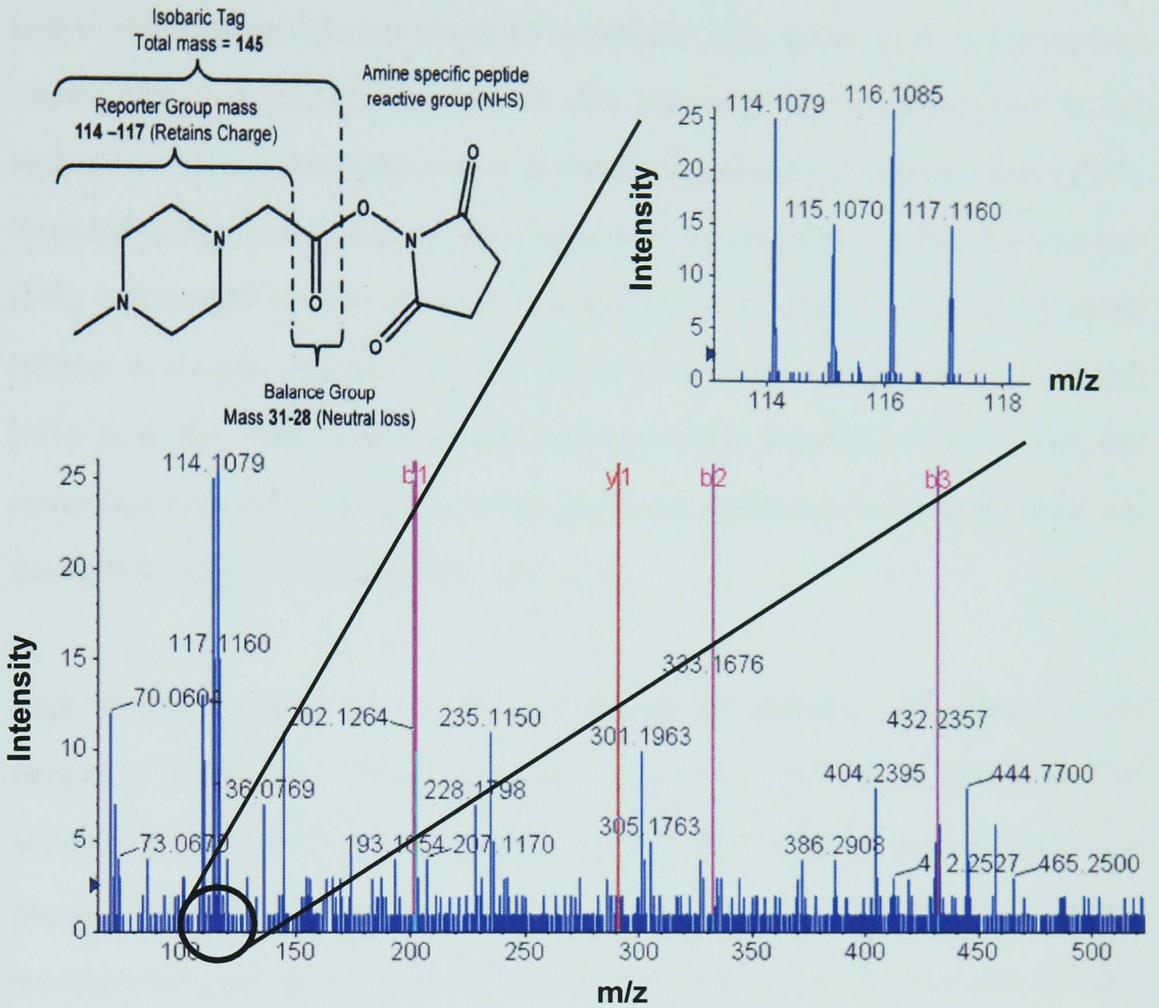


Figure 2.15. One of the MS/MS spectrum examples of iTRAQ sample. This peptide bearing the sequence of GMVLAKPGSITPHTEFEGEVUVLJJ, belongs to the elongation factor TufA protein (sll1099). The figure on the top left hand corner is adapted from Ross *et al.* [37].

This iTRAQ-mediated quantitative proteomic approach has been widely applied to many studies, such as in clinical investigation of pancreatic acinar cells [331], aging systems [332], liver system (kuppfer cells) [333], hematopoietic stem cells [334], lung cancer cells [335], neurodegeneration system (mouse model of cerebellar dysfunction) [336], neural precursor cells (NPCs) [337], proteomics of cerebrospinal fluid (CSF) [338], human bone marrow stromal cell [339], stem cell-like cell line (FDCP-mix) [340], liver peroxisomes [341], cancer markers from endometrial tissues [342], and human parotid saliva [343]. Furthermore, iTRAQ has been the key technique employed in the environmental microbial systems, such as the study of the

centrosomal proteome of *Dictyostelium discoideum* [344], the metabolically engineered *E. coli* used in cis-1,2-dichloroethylene (cis-DCE) degradation [345], nitrate stress in *Desulfovibrio vulgaris* Hildenborough [212], and *E. coli* cells expressing *rhs*A elements [346]. Similar application are also seen in plants, such as chloroplast biogenesis in *Arabidopsis thaliana* [347], the phosphoproteome of *A. thaliana* [348], and the mapping organelles proteome of *Arabidopsis* [349]; and common eukaryote systems, for examples, the investigation of the Snf1p kinase complex in *Candida albicans* [350] and ethanol fermentation in *Saccharomyces cerevisiae* [351]. It is also worth mentioning that post-translational modification studies have also successfully been carried out using iTRAQ approaches, specifically in the identification and quantification of protein phosphorylation [352-354].

Since there is a amount of data generated through this approach, and a complex ratio comparison required for a reliable quantitative result, it is necessary to have a powerful bioinformatics software/tool to perform this. So far, the commercial software available to process iTRAQ data is ProQuant, introduced by Applied Biosystems. However, this program is not compatible with more widely used search engines such as SEQUEST and Mascot, which makes cross-data analysis more complicated. There has been some development in this area such as the introduction of the i-Tracker software [355], a program that easily integrates with other search engines by extracting the reporter ions to a readable format (i.e. from .dta and .mgf files). ProGroup is another programme from Applied Biosystems that allows further analysis of output from ProQuant by reducing the proteins' redundancy based on Occam's razor principle. The recently developed software Multi-Q, is one of the most robust iTRAQ software tools available [356]. This software claims it can accommodate various types of input data formats, as well as carry out multiple tasks such as peak detection, background subtraction, isotope correction, normalization, and finally the calculation of peptide ratios.

The complementary nature of the iTRAQ approach with gel-based experiments has been reported in *B. subtilis*, while gel-free methods provide a higher degree of protein resolution [357]. The evaluation of a non-gel based method using ICAT was reported to produce a median

coefficient of variation (CV) of 18.6% from 4 independent ICAT experiments using *E. coli* as model organism [319]. A recent study by Wu *et al.* which compared three different protein quantitative methods: cleavable isotope-coded affinity tag (cICAT), iTRAQ and fluorescence difference gel electrophoresis (DIGE), found that iTRAQ is more sensitive to quantitation, but more susceptible to errors in precursor ion isolation [358]. This suggests that possible mass spectrometer interference could hinder the iTRAQ reliability issue. Previously, the implication of multiple injections on iTRAQ reproducibility was carried out by Chong *et al.*, and a high compatibility was established with an average CV of 9% across the triplicate analyses [39]. By applying three LC-MS/MS injections, it was possible to reduce the existing MS variation. Another study used Grubb's and Rosner's statistical outlier tests to improve the consistency of the quantitation data and demonstrated protein expression using iTRAQ had a CV of less than 0.24 [359].

Most of the studies carried out employing iTRAQ reagents assume the standard deviation measured from the real-time quantitation window covered the variation caused by either structural or random nuisance effects, however the actual effect of the variation from either technical or biological sources has yet to be established and identified. Gan *et al.* attempted to understand the effects of random technical, experimental and biological variations by applying replicate analysis on 10 different iTRAQ experiments [38]. The study revealed that the percentage of proteins falling within an average variation calculated through replicate analysis does not reflect the true confidence of the protein regulation, instead a higher cut-off point at ± 50% is recommended. This last piece of work also forms the basis for Chapter 4.

2.9 Summary

The use of *Synechocystis* sp. PCC 6803 as a major 'cell factory' is evident from the potential of H_2 fuel to anti-carcinogen production in cancer treatment. Researchers have been exploring this organism from different perspectives, with one common goal, that is to achieve a functional organism capable of unleashing its aptitude to the full. The most common approaches to study a cellular response are through established transcriptomic techniques, such as cDNA microarrays,

northern blotting and quantitative PCR analysis. On the other hand, proteomics based investigation has started to gain ground, with some research groups, such as Norling's in Sweden, Slabas's in (Durham) England, Aro's in Finland, Lindahl's in Spain and Wright's in (Sheffield) England, have been pioneering the proteomics field using *Synechocystis* sp. PCC 6803 as a model organism. The focus of these researches (using proteomics technique) so far has concentrated on gel-based approach specifically 2-DE analysis. However there are limitations in this approach.

The deployment of gel-free proteomic 'shotgun' analysis, in contrast, has been anticipated as being high throughput, with a higher reproducibility factors compared to gel-based method. Furthermore, quantitation can be easily achieved through this method, for examples, cICAT, iTRAQ and SILAC analysis, without the hassles of gels handling and processing. In order to fully exploit the potential of *Synechocystis* sp. as a 'cell factory', it is necessary to advance and develop new research methodologies (especially the 'shotgun'-based procedures), in order to achieve a 'Systems Microbiology' understanding in this organism.

Cellular stress response is thought to be responsible for the production of H_2, cyanophycin, PHA/PHB and carotenoid compounds (as illustrated before in Section 2.4, 2.5. and 2.6). The author believes that an understanding at the cellular level (in this case refer to the proteome level) is crucial in optimising the workflow for the production of these natural products. It must not be forgotten many of the physiological biological processes are most commonly carried out by proteins, which serve as enzymes responsible for catalysing the myriad of metabolic reactions present in the organism. However it is also important not to overlook the power of transcriptomics, because a change in gene expression is as equally important as a change in protein abundance.

To understand an organism, it is vital to recognize the most fundamental process within, in this case, light (and dark) acclimation. At different light alteration (photoperiod), different sets of genes/proteins take over, as well as the changes from photosynthesis to respiration, and vice

versa. The cell must be robust, yet sensitive in order to perform such complex process. In fact, light has been proven to play a key role in many of these naturally produced compounds (refer to Section 2.3, 2.4, 2.5 and 2.6). Similarly, depletion in macro-nutrition (such as phosphate starvation) would trigger a cellular response that might be linked to the accumulation of these natural products. With all these basic yet fundamental ideas in mind, they form the principle foundation for the entire investigation framework in this research work.

Chapter 3

'Shotgun' Proteomic Analysis of

Synechocystis sp. PCC 6803 [π]

[π] The content of this chapter has been published in Proteomics (June 2005), volume 5, issue 9, page 2468-78.

3.1 Abstract

Protein or peptide analysis by gel-free ('shotgun') proteomics relies on the simplification of a peptide mixture before it is analyzed via mass spectrometer. While separation on a reverse-phase (RP) liquid chromatographic column is widely employed, a variety of other methods have been used to fractionate both proteins and peptides before this step. A total of six different protein and peptide fractionation workflows were compared, using *Synechocystis* sp. PCC 6803, a useful model cyanobacterium for potential exploitation to improve its production of hydrogen and other secondary metabolites. Pre-digestion protein separation was performed by strip-based isoelectric focusing (IEF), 1-D PAGE, or weak anion exchange (WAX) chromatography, while pre-RP peptide separation was accomplished by IEF or strong cation exchange (SCX). Peptides were identified using electrospray ionization quadrupole time-of-flight tandem mass spectrometry (qQ-Tof). MS and MS-MS spectra were analyzed using ProID software employing both the *Synechocystis* sp. PCC 6803 and the entire NCBI non-redundant database, and a total of 776 proteins being identified using a stringent set of selection criteria. Method comparisons were made on the basis of the results obtained (number and types of proteins identified), as well as ease of use and other practical aspects. IEF-IEF protein and peptide fractionation prior to RP exhibited the best overall performance.

3.2 Introduction

There has been an increasing interest in proteomics field to develop alternative high throughput protein (and peptide) separation methods that 2D gel electrophoresis cannot achieve, especially for the detection of acidic (p*I* > 3), basic (p*I* >10), and hydrophobic proteins [267, 279-283]. These gel-free ("shotgun") approaches rely on the simplification of a protein or peptide mixture before it is analyzed in a mass spectrometer, with separation on a C18 reverse-phase (RP) liquid chromatography column being employed. Among the common techniques used during protein separation are by size exclusion chromatography (SEC) [293-296], hydrophobic interaction chromatography (HIC) [297], RP chromatography [298, 299], weak (WAX) and Strong (SAX) anion exchange chromatography [300-303], liquid- and gel-phase isoelectric focusing [304-306], and 1-D PAGE [307, 308]. On the other hand, peptide level separation methods include strong cation exchange (SCX) chromatography [284, 285], slab-gel isoelectric focusing [286], capillary isoelectric focusing [287], liquid phase isoelectric focusing [288] and capillary electrophoresis [289, 290]. The combination of these multi-level (protein then peptide) fractionation approaches has been successful and resulting in the identification of many proteins [309]. Each separation technique has its own advantages and disadvantages, yet relatively few studies have compared these various options for protein and peptide separation.

The goal of the research presented here was to compare different multidimensional protein and peptide fractionation methods, including various combinations of IEF (proteins and peptides), 1-D PAGE (proteins), WAX (proteins), and SCX (peptides), all in combination with RP chromatography and ESI-MS/MS analysis (Figure 3.1, Tables 3.1 and 3.2). These shotgun methods were also compared with standard 2-D PAGE and mass spectrometric proteomic reports from the literature [205, 259, 269-272, 274, 275]. The organism used for this study was the cyanobacterium *Synechocystis* sp. PCC 6803, a useful model photosynthetic prokaryote of interest for its potential to produce hydrogen [1], and for coupling carbon dioxide fixation to production of polyhydroxyalkanoic acids [109, 360, 361], as detailed in Chapter 2.

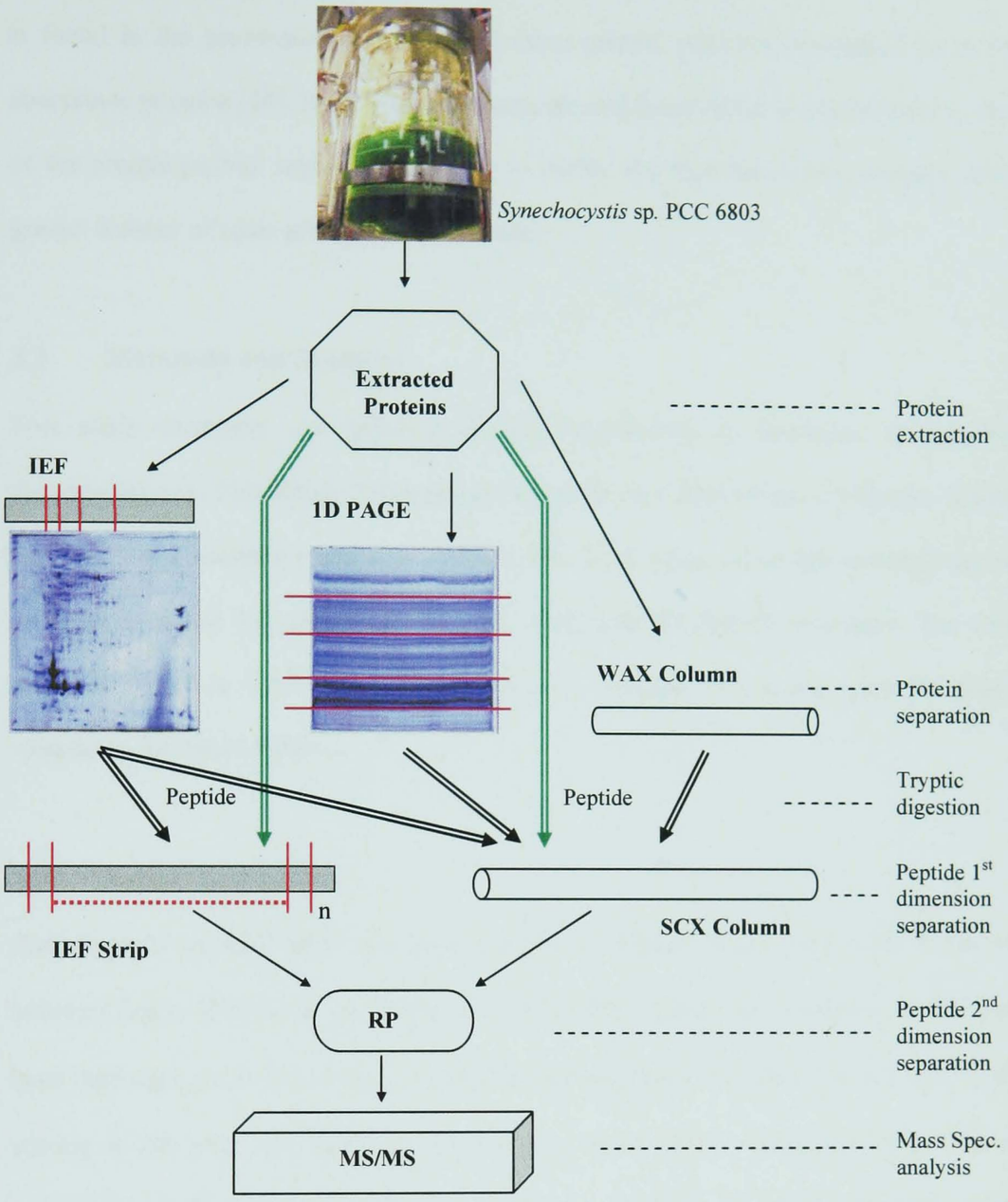


Figure 3.1. Summary representation of the six sample workflows compared in this study. (⟹) Whole cell digested peptide from extracted protein. (——) Cut made to 1D and IEF strips. (⟹) Peptide after tryptic digestion from protein separation.

One of the challenges of cyanobacterial proteomics is the presence of several phycobiliproteins (e.g., phycocyanin and allophycocyanin), which comprise up to 50% of the soluble protein content of *Synechocystis* sp. and other cyanobacteria [362]. These proteins are normally assembled in phycobilisomes, the cyanobacterial light-harvesting apparatus. Although these proteins are important, their abundance can overwhelm a separation scheme and make detection of other proteins difficult. In this sense, cyanobacterial proteomics faces the same challenge as

is found in the proteomics of serum and other protein mixtures dominated by a few high abundance proteins [363-366]. For this reason, an additional factor evaluated here is the ability of the protein/peptide separation schemes to isolate the high-abundance proteins, allowing a greater number of other proteins to be detected.

3.3 Materials and Methods

This study compares six different shotgun approaches to determine the proteome of *Synechocystis* sp. PCC 6803. The approaches include four that involve protein pre-fractionation and two that use peptide separation methods only. In all cases, the samples are introduced to the mass spectrometer following nano-flow LC employing RP peptide separation. The workflows are summarised in Table 3.1 and Figure 3.1. A separate sample was used for each of the workflows described below.

3.3.1 Sample preparation

Synechocystis sp. PCC 6803 was grown in BG-11 medium in a 2 litre (with 1 litre working volume) Braun BioStat B^+ fermentor (Sartorius BBI, Melsungen Germany) at 26°C on a 12 hour light-dark cycle with a light intensity of approximately 90 μEinstein/m^2s with continuous stirring at 200 rpm, and bubbling with pre-humidified (from a humidifier filled with sterile water) 0.2 μm filtered air at 1.0 litre/min. The light intensity was measured using a QSL-2100 Scalar PAR Irradiance Light Sensor (Biospherical Instrument Inc, San Diego, CA). Cells were harvested at mid-exponential phase (OD_{730} of 3.0) by centrifugation at 5,000×g for 15 min at 4°C. Cells were re-suspended in Tris buffer [367], consisting of 40 mM Tris-HCl (pH 8.7), 1 mM ascorbic acid, 5 mM $MgCl_2$, 1%w/v polyvinylpyrrolidone (PVPP), 1 mM DTT, 0.2% Bio-Lyte pH 3-10 (Bio-Rad, Hertfordshire, UK) and 5% Protease Inhibitor Cocktail (Sigma, Cat no. P8465). The suspended cells were incubated for at least 30 min at room temperature prior to protein extraction. Unless noted all chemicals were obtained from Sigma-Aldrich (Gillingham, Dorset, UK).

Table 3.1 Summary of protein-peptide and peptide only separation techniques (all followed by RP nano-flow LC).

Separation method/ designation	**IEF-IEF**	**IEF-SCX**	**1D-SCX**	**WAX-SCX**	**SCX only**	**IEF only**
Initial Protein Concentration	2 mg for each method					
Protein Separation	IPG Strips pH 3-10NL, 2 strips with 1 mg each strip.		1-D PAGE System	PolyWAX LP™ Column	None.	
Protein cuts	Cut into 5 pieces according to the 1D or 2D gel image (see Fig 3.1 for an example).			45 fractions collected, but some fraction combined with other to bring total fraction to 5 only.	None.	
Protein clean-up	None.			Centricon YM-3 Ultrafiltration Cartridge	None.	
Tryptic digestion	In-gel digestion.			Whole cell digestion.		
Peptide clean-up	None.	Discovery® DSC-18 SPE column	None.		Discovery® DSC-18 SPE column	
Peptide separation	IPG Strips pH 3-10L, 2 strips with 1 mg each strip.	PolySULFOETHYL™ A Column				IPG Strips pH 3-10L, 2 strips with 1 mg each strip.
Fractions collected	Each cut to 5 pieces with each piece ca. 3.4 cm to produce a total of 25 fractions.	45 fractions collected for each cut, but some fractions combined with others to bring total fraction to 5 only. Total fractions 25.				Cut to 25 pieces with each piece approximately 0.7cm.
Final clean-up	Discovery® DSC-18 SPE column	None.				Discovery® DSC-18 SPE column
Prior to mass spectrometry.	Dried in a vacuum concentrator.					
On-line nano LC and MS/MS	85 minute LC-MS/MS run using LC Packings Ultimate nano-flow LC (with RP final peptide separations on PepMAP™ 100 C18 column) coupled to Applied Biosystems QStar XL ESI hybrid quadrupole TOF MS/MS.					

Tris-HCl was used as a pH buffer. DTT was a reducing agent to reduce protein disulfide bonds. Bio-Lyte was included as an ampholyte during IEF. It was believed that $MgCl_2$ is able to extract peptidoglycan associated proteins [368]; whereas ascorbic acid was used to increase the amount of acidic proteins [369]. PVPP had been used to extract phenolic compounds as to encourage the total protein identification [370]. The protease inhibitor was added to reduce the possibility of proteins degradation.

3.3.2 Protein extraction

Proteins were extracted using liquid nitrogen combined with mechanical cracking via mortar and pestle. Liquid nitrogen was poured into a mortar before pipetting suspended cells into the liquid nitrogen. The sample was ground into powder using a pestle. The extraction process was repeated twice before recovering the extracted protein in a micro-centrifuge tube. The recovered proteins were centrifuged at 21,000×g for 30 min at 4°C to separate the supernatant proteins dissolved in Tris buffer from the insoluble pellet. The supernatant, which consisted of the dissolved proteins in Tris buffer, was transferred to a new micro-centrifuge tube and stored at -20°C until further use. The pellet, on the other hand, was discarded. The protein concentration in the supernatant was measured using the Bio-Rad RC DC protein assay.

3.3.3 Protein and peptide separation methods

A number of different off-line[σ] separation techniques were compared in this study, with both protein and peptide separations being examined. Workflows starting with protein separation used IEF, 1-D PAGE or WAX prior to peptide separation. Peptide separation steps were carried out using SCX chromatography or IEF. These workflows, each of which began with 2 mg of soluble protein, are summarised in Table 3.1. In all cases, these separations were followed by nano-flow RP chromatography carried out on-line to the mass spectrometer. Details of all separation and analytical methods are provided below.

[σ] with regard to the mass spectrometer

The IEF-IEF and IEF-SCX workflows used IEF as the protein pre-fractionation technique, while 1D-SCX employed a 1-D PAGE approach; in both cases, in-gel tryptic digestion was performed. In the IEF-SCX workflow, it was necessary to remove urea prior to SCX separation. To accomplish this, the digested protein fractions from IEF were cleaned using Discovery® DSC-18 SPE cartridges (100 mg capacity, Supelco, Sigma). In this SPE step, the cartridge was first conditioned with methanol and equilibrated with 0.1% TFA in water according to the manufacturer's instructions. The sample was then loaded on the cartridge, which was then washed with 0.1% aqueous TFA. Finally, peptides were eluted using 0.1% TFA in 80% acetonitrile. Cleaning was not required for the IEF-IEF and 1D-SCX approaches. WAX-SCX used weak-anion-exchange (WAX) chromatography for protein separation. WAX protein fractions were cleaned and concentrated using Centricon YM-3 Ultrafiltration Cartridges (Amicon, Milipore, Watford, UK). The treated fractions were then digested with trypsin prior to SCX peptide separation.

The IEF-only and IEF-IEF workflows used IEF as the means for peptide separation, while the others (1D-SCX, IEF-SCX, WAX-SCX and SCX-only) used SCX chromatography. IEF- and SCX-only did not involve any protein prefractionation methods. Whole-cell lysate samples for the IEF-only and SCX-only methods were digested with trypsin and cleaned as described above using Discovery® DSC-18 SPE cartridges before peptide separation. After peptide separation, peptide fractions from IEF-IEF and IEF-only were cleaned using Discovery® DSC-18 SPE cartridges prior to MS analysis. Cleaning of peptide fractions was not needed in the 1D-SCX, WAX-SCX, IEF-SCX and SCX-only workflows.

3.3.4 Solution-phase tryptic digestion

For the methods that did not use a gel-based protein fractionation step (IEF-only, SCX-only, and WAX-SCX), a standard solution-phase tryptic digestion procedure was used. Reduction and alkylation of the proteins was carried out prior to tryptic digestion. To accomplish this, the sample was reduced with a final concentration of 10 mM DTT and 50 mM ammonium bicarbonate for 1 hour at 56°C, followed by alkylation with a final concentration of 55 mM

iodoacetamide and 50 mM ammonium bicarbonate for 30 min in the dark at 37°C. Sequencing grade trypsin (Promega, Southampton, UK) was prepared according to the manufacturer's protocol. Trypsin was added to the protein mixture in a 1:50 mass ratio. In the case of the WAX protein separation, the mass of protein in each fraction was determined using the Bio-Rad RC DC protein assay. The sample was incubated overnight at 37°C. The following day, the same amount of trypsin was added again and incubated at 37°C for at least another 5 hours. Trypsin activity was inactivated by lowering the pH of solution to below 4 with TFA.

3.3.5 In-gel digestion

In-gel digestion was used for the 1D-SCX, IEF-IEF and IEF-SCX methods. Gel pieces (or IEF strips) were cut into small pieces as described in Table 3.1. Gel pieces were destained twice with 200 mM ammonium bicarbonate in 40% acetonitrile. Subsequently, the gel pieces were dehydrated using acetonitrile for approximately 15 min at 37°C; the liquid phase was then removed and samples were placed in a vacuum concentrator for another 15 min to completely dry the gel pieces. Reduction and alkylation were carried out as previously described above. After alkylation, gel pieces were dehydrated using acetonitrile for 15 min at 37°C. Gel pieces were then conditioned with 50 mM ammonium bicarbonate for 15 min at 37°C prior to a final dehydration using same volume of acetonitrile for 15 min at 37°C. Gel pieces were completely dried in a vacuum concentrator prior adding trypsin in a 1:50 mass ratio. The mass of protein in each IEF or 1-D PAGE fraction was estimated from the distribution of spot intensities in the corresponding dimension of a 2-D PAGE gel (refer to Figure 3.1). To increase the activity of the trypsin, acetonitrile was added to the sample to a final concentration of 10%-40%v/v [371]. Gel pieces were incubated overnight at 37°C. No next-day trypsin was added during the in-gel digestion process. After overnight digestion, the liquid phase was transferred to a new tube. Peptides were extracted by one change of 5% formic acid, two changes of 50% acetonitrile in 5% formic acid and one change of 100% acetonitrile incubating at 37°C for 15 min for each change. All liquid phases were combined and dried in a vacuum concentrator.

3.3.6 IEF for protein separation

Cup loading (Bio-Rad) was employed during protein IEF. Broad range (pH 3-10NL) IPG strips 17cm (Bio-Rad) were rehydrated with Destreak Rehydration Buffer (GE Healthcare, Buckinghamshire, UK) overnight at room temperature. Protein was precipitated using 0.1% TCA and 2 mM tributyl phospine (TBP) in acetone overnight. The next day, the sample was washed with 2 mM TBP in acetone. The protein pellet was then dissolved in Destreak Rehydration Buffer (GE Healthcare) and 0.5% pharmalyte pH 3-10. The protein solution was then loaded into the cup and run using a Protean II IEF Cell (Bio-Rad) for 18.5 hours. The programme was 50 V for 9 hours, 200 V for 1 hour, 1,000 V for 1 hour, 10,000 V for 6 hours and a final 15,000 volt-hours. Total volt-hours were approximately 45,000.

3.3.7 IEF for peptide separation

Cup loading was also used for peptide IEF. Broad range (pH 3-10L) 17cm IPG Strips (Bio-Rad) were rehydrated with 8 M urea overnight at room temperature. Tryptically digested, SPE-cleaned peptides were vacuum dried. The dried pellet was re-dissolved in 8 M urea with 0.5% pharmalyte pH 3-10. To ensure maximum recovery, the sample was sonicated for 10 min prior to cup loading. The IEF program was the same as described for protein IEF. After IEF, acetonitrile was added to the strips to shrink the gel side. Peptides were then extracted as described above.

3.3.8 1-D PAGE

Protein was dried in a vacuum concentrator and re-dissolved in Sample Buffer/Laemmli Buffer (Bio-Rad), to at least 2 volumes of the original sample. The composition of 1D gels (size: 7 cm long × 7 cm width × 0.75 mm thick) was 12.5% for the stacking gel and 4% for the resolving gel. The sample was loaded into the well, and ran in the mini-Protean III electrophoresis system (Bio-Rad) at 50 V for 5 hours. Subsequently, gels were washed with deionised water 3 × 5 minutes. Bio-safe Coomassie Blue (Bio-Rad) was then used to stain the gels for 1 hour, with gels destained overnight with deionised water at room temperature.

3.3.9 SCX separation

Separation was carried out using a PolySULFOETHYL™ A Column (PolyLC, Columbia, MD, USA) 5 μm particle size of 100 mm length × 2.1 mm id, 200 Å pore size, on a BioLC HPLC unit (Dionex, Surrey, UK). Buffer A was composed of 10 mM KH_2PO_4 and 20% acetonitrile, pH 2.97, whereas buffer B was composed of 10 mM KH_2PO_4, 20% acetonitrile and 1 M KCl, pH 2.97. The gradient used was 0% B for 5 min, 7.5%-20% B for 25 min, 20%-50% B for 5 min, 50% B for 5 min and 0% B for 5 min, total of 45 min. Injection volume was 200 μL. The flow rate was maintained at 0.2 ml/min. Separated fractions were collected every minute for 45 min using a Foxy Jr. Fraction Collector (Dionex), whilst the chromatogram was monitored through a UV Detector UVD170U set at 214 nm and Chromeleon Software, version 6.50 (Dionex/LC Packings, The Netherlands). Fractions were later pooled according to the UV/Vis signal at 214 nm (Table 3.1).

3.3.10 WAX separation

Separation was carried out using a PolyWAX LP™ Column (PolyLC) 5 μm particle size of 100 mm length × 4.6 mm id, 1000 Å pore size, on a BioLC HPLC unit (Dionex). Buffer A was composed of 10 mM KH_2PO_4 and 20% acetonitrile, pH 7.20, whereas buffer B was composed of 10 mM KH_2PO_4, 20% acetonitrile and 1 M NaCl, pH 7.20. The gradient was 0% B for 5 min, 0%-5% B for 5 min, 5%-50% B for 25 min, 50% B for 5 min and 0% B for 5 min, total of 45 min. Injection volume was 200 μL. Flow rate was maintained at 0.8 ml/min. Fraction collection and pooling were carried out as described for the SCX separation.

3.3.11 Nano-LC ESI-MS/MS

Prior to mass spectrometric analysis, all samples were dried and re-dissolved in a buffer consisting of 0.1% formic acid and 3% acetonitrile. All analysis was undertaken using a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF-MS/MS (Applied Biosystems, Framingham, MA, USA; MDS-Sciex, Concord, Ontario, Canada), coupled with a nano-LC system comprising a combination of a LC Packings Ultimate Pump, Switchos Column and Famos Autosampler (Dionex/LC Packings). The Switchos Column had a

μ-Precolumn™ Cartridge 5 μm particle size of 300 μm id × 5 mm length for desalting and filtering purposes before the actual C18 capillary column 3 μm particle size of 75 μm id × 15 cm length. Both columns consisted of the RP material PepMAP™ 100 C18 silica-based, 100 Å pore size (Dionex/LC Packings).

The buffers used in the gradient were Ultimate Buffer A, consisting of 0.1% formic acid in 5% acetonitrile and Ultimate Buffer B, consisting of 0.1% formic acid in 95% acetonitrile. The nano-LC gradient was 85 min in length with the first 15 min comprised of 5% B, followed by 57 min ramping from 5% to 40% B, then 8 min of 40% B before a final 5 min of 5% B. The flow rate of the gradient was 300 nL/min. The ESI-MS detector mass range was set at 300-2000 m/z. The MS data acquisition was performed in the positive ion mode. During the scan, peptides with a +2 or +3 charge state were selected for fragmentation.

3.3.12 Database searching and data interpretation

MS/MS data were analysed using ProID software (Applied Biosystems, MDS-Sciex) for protein identification. A total of two databases were used for the search. These were the entire NCBI non-redundant (NR) database (July 2004) and the *Synechocystis* sp. PCC6803 (SYN) single-organism database at NCBI (March 2004). Within ProID, the peptide tolerance was set to 2.0 Da and the MS/MS tolerance was set to 0.8 Da. The quality filter of 80% confidence and score of 10 were used as recommended by Applied Biosystems.

3.4 Results and Discussion

3.4.1 Methods comparison

The results of the six different protein/peptide separation methods used to determine the *Synechocystis* sp. PCC 6803 proteome during the mid-light phenotype are summarised in Table 3.2. Greater proteome coverage was achieved with pre-digestion protein fractionation prior to the shotgun proteomics analysis. If neither protein not peptide fractionation was performed prior to RP-LC-MS/MS, then only seven proteins were identified using the *Synechocystis* sp. PCC 6803 database and required matching of ≥ 2 peptides (row 7 in Table 3.2), compared to the

identification of hundreds of proteins when any of the fractionation workflows were used (rows 1 to 6 in Table 3.2). The introduction of fractionation greatly reduced the complexity of protein/peptide sample. This improved the ability of the MS to detect low abundance peptides/proteins, since ion-suppression effects caused by overlapping signals from high and low abundance ions was reduced [372].

In this study, two different peptide separations approaches, IEF-only and SCX-only, were employed before the final reverse-phase (RP) HPLC separation and MS/MS analysis. The combination of SCX and RP form the most common MDLC method [284], while peptide-level separation using IEF is not as commonly practiced [286, 373]. A total of 11 additional (with ≥ 2 peptide using SYN database) proteins were identified using IEF-only compared to SCX-only peptide separation. During IEF, peptides were retained on the IPG strip, thus minimizing the possibilities of contamination and protein loss. This may contribute to the slight increase in the number of total proteins found.

Use of pre-digestion protein fractionation increased the total number of proteins identified, as detailed in Table 3.2 (row 3 to 6). Considering only proteins with ≥ 2 peptides found in the *Synechocystis* sp. PCC 6803 database, the total number of proteins identified in 1D-SCX and IEF-IEF is approximately double that compared to IEF- or SCX-only peptide separation alone. One of the weaknesses in chromatographic separation found in this study is the recovery rate and separation efficiency. WAX-SCX performed the worst of all the methods, in that it allowed for identification of the lowest number of proteins. More than 80% of the protein was lost during the WAX separation, as determined by a total protein assay. High sample losses during chromatography have also been reported elsewhere [292, 297, 374]. A comparison of the methods tested here suggests that conventional 1-D PAGE and IEF were the best approaches for protein-level separation. Some proteins did not diffuse and focus completely in the strips during IEF. Evidence of this was apparent in the IEF-SCX separation, where fewer proteins were identified than in the 1D-SCX approach. For peptide separation, IEF peptide separation was more effective (compare IEF-only vs. SCX-only and IEF-IEF vs. IEF-SCX in Table 3.2).

Table 3.2 Total unique proteins achieved using different protein-peptide separation methods. Rows 1 to 6: IEF peptides, SCX peptides, 1D-SCX, WAX-SCX, IEF-SCX and IEF-IEF separation. Rows 7 to 8: whole cell digested protein with no additional protein or peptide separation, and SCX peptides separation (40 fractions).

							NCBI Non-Redundant Database				*Synechocystis* sp. PCC6803 Database			
	Protein separation	No of fractions	Peptide separation	No of fractions	Total Processing Time (a)	Total MS samples	Total unique peptides	Total unique Proteins			Total unique peptides	Total unique Proteins		
								≥ 1	≥ 2	≥ 3		≥ 1	≥ 2	≥ 3
1	no	-	IEF	25	3 days	25	232	120	53	20	634	321	171	53
2	no	-	SCX	25	2 days	25	176	99	34	18	612	335	160	49
3	1D	5	SCX	5 per	4.5 days	25	497	211	99	63	1321	394	308	205
4	WAX	5	SCX	5 per	3 days	25	104	41	19	10	406	216	99	32
5	IEF	5	SCX	5 per	4 days	25	336	163	74	35	788	299	204	101
6	IEF	5	IEF	5 per	6 days	25	507	217	106	64	1405	467	344	191
Total unique proteins/peptides in all methods							1122	397	176	106	3700	1107	711	357
7	no	-	no	-	1 day	1	23	10	6	4	36	15	7	6
8	no	-	SCX	40	2 days	40	995	390	221	135	1900	571	453	362

(a) Excluding MS time, which is 25 fractions × 90 min MS time = 1.5 days for each method with 25 fractions.

Depending upon the type of information available for the proteins of interest, different techniques might be chosen for the protein separation process. In this study, 1-D PAGE and IEF separation proved to be the more powerful pre-digestion protein fractionation methods. Proteins that are deposited in-gel avoid the problems of protein solubilisation and low protein recovery rate during chromatographic separation. The IEF-SCX, 1D-SCX, and IEF-IEF separation methods yielded the identification of more than 200 unique proteins (more than 300 in the case of the latter two methods). No particular bias in observed proteins when using the IEF, 1-D or WAX method for protein separation. Almost all the proteins were within the p*I* region of 3 to 10 (Figure 3.2). The theoretical Mw and p*I* from Figure 3.2 were predicted using online JVirGel program [375] (http://www.jvirgel.de/).

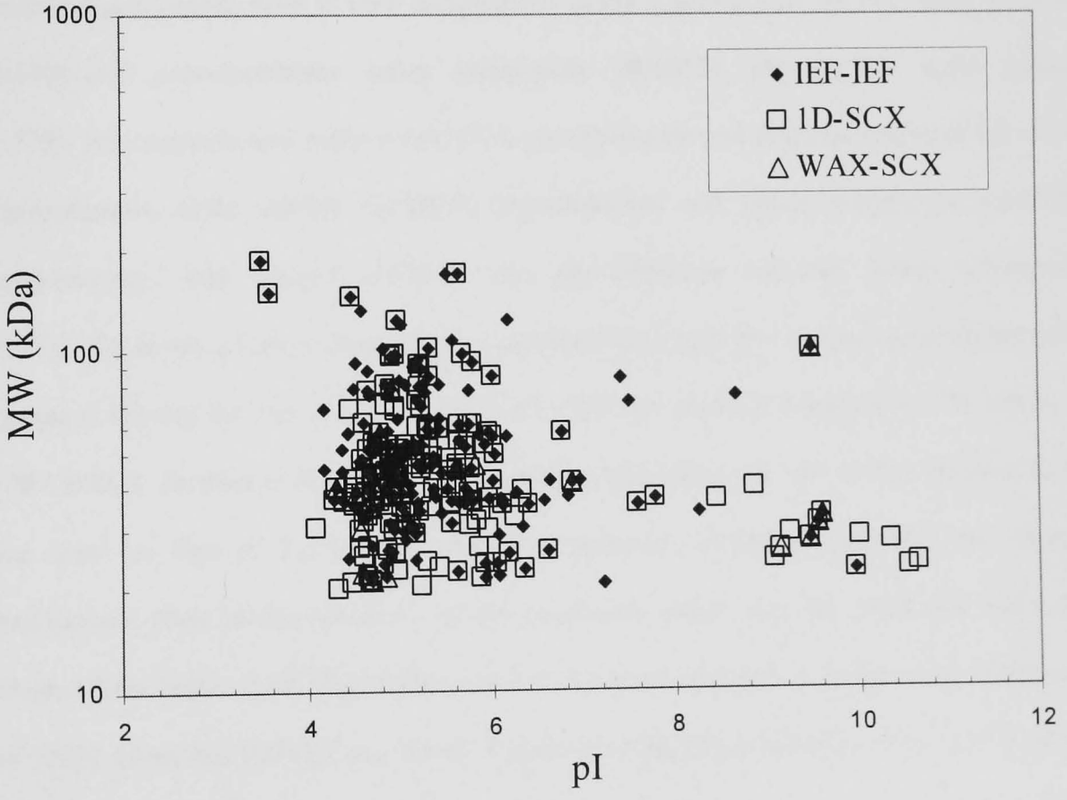


Figure 3.2. 2-D gel plot of unique proteins (with > 1 peptide in NR database) identified in IEF-IEF, 1D-SCX and WAX-SCX separation methods.

In peptide separation, the number of proteins found corresponds to the number of fractions collected. A comparison of the number of fractions obtained using the SCX peptide separation method (row 2 and 8 in Table 3.2) shows that an increase from 25 to 40 fractions increased the total number of proteins identified by 2.8 fold (Table 3.2). While increased fractionation might be appealing, the choice of the number of fractions should be made not only with the ambition of getting the highest number of the total proteins, but also with consideration of the time needed for sample processing and MS analysis.

One of the factors that influenced the outcome of separations in this study was the abundance of several phycobiliproteins (e.g., phycocyanin, allophycocyanin) within the protein mixtures. The top ten hits for the SCX-only, IEF-only and WAX-SCX proteins, respectively (against the SYN database) shared the same set of phycobiliprotein identification, which were (the following numbers in parentheses refer to ORF designations in the *Synechocystis* sp. PCC 6803 genome) phycobilisome core-membrane linker polypeptide (slr0335), phycocyanin alpha subunit (sll1578), phycocyanin beta subunit (sll1577), phycobilisome rod linker polypeptide (sll1579), allophycocyanin alpha subunit (slr2067), phycobilisome rod linker polypeptide (sll1580), allophycocyanin beta subunit (slr1986) and phycobilisome rod-core linker polypeptide (slr2051). These are all phycobiliproteins. Comparatively, only two or three phycobiliproteins appeared in the top ten hits of the IEF-SCX, IEF-IEF and 1D-SCX separations. The failure of the WAX-SCX method to separate the phycobiliproteins from the rest of the proteins in the lysate may be one of the reasons why this approach produced relatively few protein identifications. Due to the similarity of the isoelectric point (p*I*), the molecular mass and hydrophobicity index of the phycobiliproteins, it is not an easy task to isolate them from each other [299]. However, IEF-IEF and 1D-SCX performed the best at isolating these proteins from the rest, as these methods produced greater numbers of identifications (i.e. the extra proteins were not obscured). The gel-based methods were better than the LC methods, in this regard, as it was very clear where these phycobiliproteins were because it could be visualized and separated out fractions accordingly (Figure 3.1).

Interestingly, although membrane proteins were not the main focus here, a total of 123 proteins were found with at least 1 transmembrane helix (TMH), using the TMHMM program from the Centre of Biological Sequence Analysis, Denmark Technical University [376]. Amongst the 123 TMH proteins, more than 70% (89 proteins) were exclusive TMH proteins, i.e. unique proteins only found exclusively in one workflow only (refer to Table 3.3). It is worth mentioning that the high resolving power of gel-based workflows, such as 1D-SCX and IEF-IEF, produced the highest number of TMH proteins, with 43 and 40 proteins respectively. Comparatively, the LC-based methods utilising WAX-SCX and SCX-only only yielded 14 and 26 proteins respectively (Table 3.3). However the question remained: Why were there so many TMH proteins identified? Polyvinylpyrrolidone (PVP) is a water-soluble polymer made from the monomer N-vinyl pyrrolidone. Polyvinylpolypyrrolidone (PVPP), on the other hand, is a cross-linked form of PVP, used to bind impurities from the solutions.

Table 3.3 The total number of proteins found with at least 1 transmembrane helix (TMH) from each workflow. Exclusive TMH proteins are defined as unique proteins with TMHs found exclusively in one method and not appearing in results from any other methods.

Workflows	Exclusive TMH proteins in each workflow	Total TMH proteins in each workflow
SCX-only	12	26
IEF-only	16	31
1D-SCX	20	43
WAX-SCX	8	14
IEF-SCX	10	26
IEF-IEF	23	40
Total exclusive TMH proteins	89	-
Total TMH proteins found in all workflows	123	

The speculation is that PVPP made the extracted supernatant protein more viscous, or denser. Therefore when the sample was subjected to centrifugation at 21,000×g for 30 minutes, the light membrane vesicles that were formed during liquid nitrogen extraction did not go to the insoluble pellet, but instead were suspended in the supernatant. Protein in these light membrane vesicle then became possible to be identified in the different workflows described here.

Similarly Charmont *et al.* also suggested that PVPP trapped the phenolic compounds as to prevent their unspecific binding with other proteins thus to increase the number of total identified peptides per protein [370]. Further details of each TMH protein, such as the number of predicted helices etc, are available in Appendix A.

In addition to the total number of proteins identified, consideration of practical issues such as the total processing time of the samples and the complexity of sample preparation is also important. Cutting IEF strips for peptide analysis took more than 10 min and may have allowed significant peptide diffusion [373]. This would be less of a problem for protein-IEF since proteins have larger sizes compared to peptides, with subsequently lower diffusion in the gel. This diffusion issue was relatively important as the choice of IEF cuts was made to allow for isolation of the highly abundant phycobiliproteins into certain fractions. This would also impact on lower abundance proteins/peptides near the cut margins. Among all the separation workflows, the total sample processing time for SCX-only peptides was the fastest (2 days), followed by IEF-only peptide and WAX-SCX (3 days each). The longest processing time was for IEF-IEF (6 days) and 1D-SCX (4.5 days); although these were the methods that yielded the highest number of proteins found (refer to Table 3.2). The total MS time for each (25 fractions) method was 1.5 days. Chromatographic separations have the advantage of higher capacity and easier operation, which greatly reduce the level of method complexity. In common with the experiences of other researchers, gels (IEF or 1D method) still lag behind in terms of the labour time required, yet these options are the most economical, most effective separation method (in the terms that the most proteins were identified) and produce high quality visualization results.

Multiple injections of the same sample into the MS resulted in different proteins being identified, due to chance auto-selection of peptides for MS/MS analysis [377]. Schaefer *et al.* [378] suggested that five or more injections for sample analysis are required to ensure reproducibility and reliability of the results. One of the samples in the IEF-IEF separation method (fraction 19) was injected twice into the MS, as a test case. Based on the SYN database with proteins $\geq$ 2 peptides, the first injection found one protein whilst the second injection

recovered the same protein with an additional three unique proteins (Table 3.4). The reason for this difference may be the variation in the peptide ion peaks chosen for automatic MS/MS analysis. For better protein coverage (and number) in a sample, multiple injections of the same sample are thus recommended. The reason for this difference may be the variation in the peptide ion peaks chosen for MS/MS analysis. For better protein coverage (and number) in a sample, multiple injections of the same sample are thus recommended (see also, for example, [379] for improved proteome coverage when applying replicate injections). However, this aspect was not the focus of this study, as this is well accepted.

Table 3.4 Total number of unique proteins for multiple repeated injections of the same sample. One of the IEF-IEF separation method samples (Fraction 19) was injected twice and proteins were identified using the *Synechocystis* sp. PCC 6803 database.

Peptide no.	Total unique proteins	
	1st injection	2nd injection
≥ 1	15	37
≥ 2	1	4
≥ 3	0	1

3.4.2 Data analysis

To investigate the reliability of protein identity (ID), all of the MS data files were analysed using two different databases, the entire NCBI non-redundant (NR) and *Synechocystis* sp. PCC 6803 (SYN) amino acids databases (Table 3.2). Both databases were downloaded from the NCBI ftp folder (ftp://ftp.ncbi.nih.gov/). Approximately 1100 and 3700 unique *Synechocystis* sp. PCC 6803 peptides were found using the NCBI non-redundant and *Synechocystis* sp. PCC6803 amino acids databases, respectively.

The NR database consists of fully or partially sequenced genomes/genes of more than 130,000 organisms, and thus to find a unique *Synechocystis* sp. protein is more akin to 'looking for a needle in a haystack'. Thus, with such a large choice, if one obtains a hit against a *Synechocystis* sp. PCC 6803 protein, then this is likely to be credible. Even with a single peptide from a

protein identified, the probability of a false positive is greatly minimised. Therefore, in this study, any protein with more than one peptide, regardless of number of MS/MS hits, was considered as a reliable count when using the NR database.

The use of two databases provided the opportunity to examine the consistency of results from a single genome database, in this case *Synechocystis* sp. PCC6803, versus a considerably larger one (i.e. the NR). Due to the limited number of only 3168 proteins within this database, a hit in this single genome database is more likely, and harder to assign as a false positive. For increased reliability, at least two or more peptides matches are required to validate identifications from the SYN database. The possibility of achieving a true positive identification for a protein on the basis of a single peptide is still unknown and unpredictable.

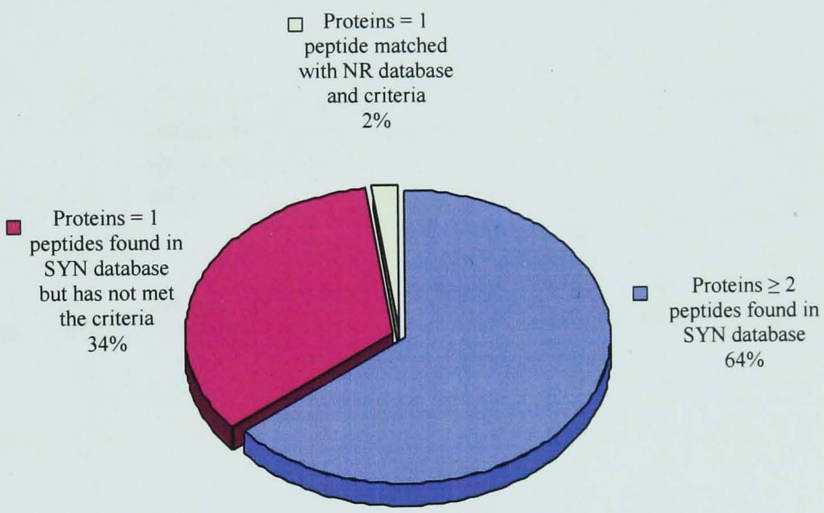


Figure 3.3. Distribution of proteins with different peptide number in the total number proteins found using the SYN database. Proteins with single peptides must be identified in both databases (SYN and NR), and appeared in the same method and same fraction to meet the criteria. A total of 1107 proteins with ≥ 1 peptide were detected using the SYN database.

According to Zhu *et al.* [283], an example of a quality matching criterion is to decide that all proteins must have at least two peptides identified with the exception of two MS/MS hits for single peptide. However, this criterion is very time consuming for studies like this, in which a total of more than 1000 proteins were found, of which 40% were single peptide hits. Without

proper aids from the Database Search Engine, it may be impractical to check the number of MS/MS hits in each of the 400 single peptide proteins.

When the proteins identified on the basis of a single peptide matches to the SYN database were matched with the NR database, 56 out of 396 were identified in both databases, suggesting the strong possibility of those 56 being correctly identified. The 396 proteins were determined by subtraction of 711 proteins with ≥ 2 peptides and 1107 proteins with ≥ 1 peptide using the SYN database (Table 3.2). However, a positive cross-match was only counted for those protein IDs that were found in both databases, with the proviso that IDs appeared in the same method and within the same fraction. With this more stringent criteria, this reduced the number to 46, and finally to 23. This figure corresponds to 2% of the total number of proteins found using the SYN database (Figure 3.3).

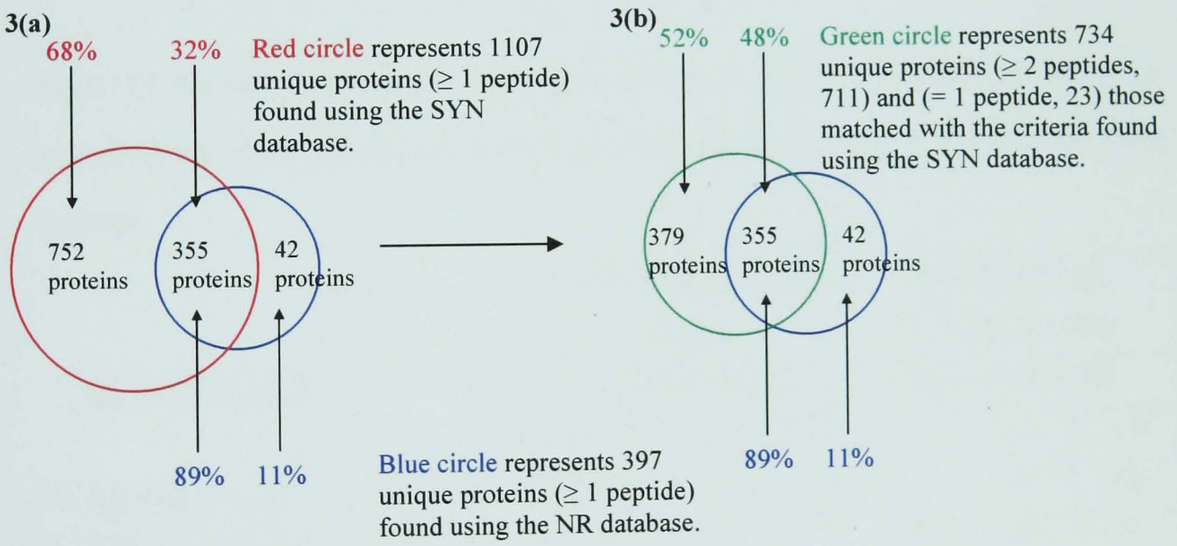


Figure 3.4. (a) Total number of proteins identified in both databases (NR and SYN) before applying the criteria, and (b) reliable proteins after matching with criteria. 89% of the proteins found in NR database can be found in SYN database.

Using the SYN database, 1107 proteins with ≥ 1 peptide were found and using the NR database, 397 proteins were found with ≥ 1 peptide. Among them, 355 were found in both databases (Figure 3.4). Hence, using what is believe to be a conservative identification criteria, the final

total number of unique proteins found in this study is 711 containing ≥ 2 peptides (Table 3.2) from the SYN database, 42 proteins with ≥ 1 peptide from the NR database (Figure 3.4), and the 23 that were identified in both databases, appearing in the same fraction of a method. This then represents a total of 776 unique *Synechocystis* sp. PCC 6803 proteins (Figure 3.4).

A master list of all proteins for the unicellular cyanobacterium, *Synechocystis* sp. PCC6803 was created. From other investigations in the literature (all 2DE-based) [205, 259, 269-272, 274, 275], a total of 344 unique proteins have been identified. Of these, 162 were also found in this study, along with 614 additional unique proteins. Thus, this study has contributed 64% of the total number of unique proteins found in proteomic studies for this organism to date, or approximately 19% of the 3168 predicted genes from CyanoBase (http://www.kazusa.or.jp/cyano/*Synechocystis*/). The master list of proteins is available online at http://wrightlab.group.shef.ac.uk/wrightlab/projects/ and in Appendices A-C.

Table 3.5. Exclusive proteins only found in each method. Exclusive proteins are defined as unique proteins found exclusively in one method and not appearing in results from any other methods.

		NCBI Non-Redundant Database			*Synechocystis* sp. PCC6803 Database		
	Separation method	Exclusive proteins			Exclusive proteins		
		≥ 1	≥ 2	≥ 3	≥ 1	≥ 2	≥ 3
1	IEF peptide only	23	7	3	93	46	15
2	SCX peptide only	13	4	1	126	56	19
3	1D protein - SCX peptide	51	17	20	108	97	84
4	WAX protein - SCX peptide	8	3	0	63	34	9
5	IEF protein - SCX peptide	32	10	6	69	50	17
6	IEF protein - IEF peptide	59	28	24	150	132	79
	Total exclusive proteins	186	69	54	609	415	223
	Total unique proteins in all methods (from Table 3.2)	397	176	106	1107	711	357

Further interpretation of the data identified "exclusive" proteins. "Exclusive" is defined as those proteins only identified in a particular method. Using proteins matched from the SYN database with ≥ 2 peptides, 58% are "exclusive" proteins (Table 3.5). This table suggests that a whole proteome study is necessarily limited if only one separation method is used. At least two pre-fractionation methods, prior to nano-flow LC RP separation and MS/MS, are recommended for a more diverse proteome study on this bacterium. This is in agreement with results on human cilla proteomics [310], for example. Method selection should be based on different protein/peptide characteristics, such as p*I*, size, hydrophobicity etc. Of course this study necessarily represents a snapshot of what can be identified per method. Other factors such as use of multiple injections into the MS of the same sample will alter (and increase) the proteins identified as well ([379], and Table 3.4).

3.5 Conclusion

In the present study, using the stringent identification selection criteria, a total of 776 unique *Synechocystis* sp. PCC 6803 proteins were found. Unsurprisingly, as the number of fractions was increased prior to mass spectrometric analysis, so too did the number of proteins that was identified. In comparison of six protein/peptide separation and pre-fractionation workflows (each with a total of 25 fractions) it was found that 1D and IEF methods are recommended for protein-level separation to obtain the highest number of confident identifications. IEF was found to work well for peptide separation. Furthermore, it was, surprisingly, possible to identify a large number of integral membrane proteins, among the abundant soluble proteins. Therefore this is a very positive finding, since it should allow identification (using some of these work-flows) of most of the integral membrane proteins starting with purified membranes , which are depleted of phycobilisomes, ribosomal proteins, elongation factors and other abundant soluble proteins.

With regard to database interrogation, the criterion used when searching a single genome database (in this case *Synechocystis* sp. PCC 6803) was that a protein with more than two peptides identified was considered to be a confident hit; single peptide protein identifications

could be considered confident only if the proteins were matched in more diverse databases, such as NR, with both identifications appearing within the same method and same fraction. Future work should investigate the 1D-IEF separation approach, since both strategies utilise in-gel separation. As in a number of other studies [378, 379] the use of multiple repeated injections resulted in different proteins being identified, and led to a substantial increase in the total number of proteins identified. Although the actual number of peptides and proteins identified might vary with replicate performances of each workflow, the overall trends can be expected to remain the same. Optimisation of the MS analysis, such as cycle (gradient) times, gas phase fractionation and multiple injections, was not considered in detail here and may become the subject of future work.

Chapter 4

Technical, Experimental and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) [Ω]

[Ω] The content of this chapter has been published in Journal of Proteome Research (February 2007), volume 6, issue 2, page 821-827. This publication is led by CS Gan (the author) with the collaboration of PK Chong (for *Sulfolobus solfataricus* iTRAQ sets) and TK Pham (for *Saccharomyces cerevisiae* KAY446 iTRAQ set).

4.1 Abstract

In this study, the reliability of isobaric-tags for relative and absolute quantitation (iTRAQ) was assessed, based on different types of replicate analyses taking into account technical, experimental and biological variations. In total, 10 iTRAQ experiments were analysed across three domains of life involving *Saccharomyces cerevisiae* KAY446, *Sulfolobus solfataricus* P2 and *Synechocystis* sp. PCC 6803. The confidence of protein expression of iTRAQ analysis increased as the variation tolerance increased. In brief, 88% of proteins would fall within a cut-off point of 50% variation (±0.50) in quantification based on an analysis of biological replicates. Technical replicate analysis produced a higher coverage level of 95% at a lower cut-off point of 30% variation. Experimental or iTRAQ variations exhibited similar behaviour as biological variations, which suggested that most of the measurable deviations came from biological variations. These findings underline the importance of replicate analysis as a validation tool and benchmarking technique in protein expression analysis. By employing multiple injections, the total number of proteins identified increased by 50% from three LC-MS/MS analyses compared to only one analysis. Furthermore the iTRAQ quantification values were found to be highly reproducible across the injections, with an average coefficient of variation (CV) of 0.09 (scattering from 0.14 to 0.04) calculated based on log mean average ratio for all three organisms.

4.2 Introduction

The introduction of iTRAQ reagents has caught the attention of the proteomics community, with the ability to perform relative (or absolute) quantification for up to four phenotypes (and soon eight phenotypes) using an isobaric mass tagging technique [37]. iTRAQ-based proteomics has been widely applied on different systems, ranging from an understanding of micro-organism stress responses [212, 345-347] to evaluating mammalian organelle systems [331, 332, 341, 343, 353]. The hundreds of thousands of spectra generated from this technique contain peptide sequences ranging from false positives to outlier peptides, and most data requires stringent manual checking. Nevertheless, these are relatively costly reagents used to derivatize peptides, a chemical reaction which could, in principle, be a source of variability [380].

The source of variation differs from one experiment to another, and it is a function of time, manpower, instrument, subject, subject condition, and preparation process etc. These sources of variation (measured experimentally or technically) must be minimized or identified. By definition, the variation is a measure of its dispersion, indicating how its possible values are spread around the expected value. In the biological sciences, this can be measured as three different forms: technical, experimental and biological. Typically, a technical replicate assesses the extent to which the measurements of a test subject remain consistent over repeated tests of the same subject under identical conditions. It eliminates the possible errors contributed from the sample preparation, which are also normally known as the sample variance. To assess the biological variability associated with the test subject, it is commonly measured as a biological replicate, by repeating the creation of the test subject under the same conditions, so that the inconsistency/variability related within can be estimated. Experimental replicates are not commonly compared to technical and biological replicates. In fact, there is no experimental variation in the 2-DE system, because the technical replicate is the experimental replicate. Hence in this context, the experimental variation also refers as the iTRAQ variation, i.e. the repetition of an identical iTRAQ experiment.

Extensive studies on protein expression validation have focused on the 2-DE workflow, where cDNA microarray-like normalisation methods were introduced to estimate the biases from different dyes [381]; the development of a partial least squares-discriminant analysis (PLS-DA) for identification of differentially expressed protein spots [382]; comparison between various 2-DE software packages to estimate the staining efficiency [383]; and the consideration of the impact of different replicate types during protein expression experiments [384].

In this study, the idea of validation in the form of replicate analysis into the iTRAQ experiments has been incorporated. Replicate analysis is common in 2-DE [384], and is used to estimate the variation between different phenotypes by the presence or absence or the intensity of the spots. This is the first time a large scale investigation has been carried out on iTRAQ reliability and reproducibility in terms of technical, experimental and biological replicate on the 3 domains of life: eucarya (*Saccharomyces cerevisiae* KAY446), archaea (*Sulfolobus solfataricus* P2) and bacteria (*Synechocystis* sp. PCC 6803). The aim of this study is to emphasize the importance of validation in iTRAQ analyses; with the anticipation to form a basic benchmarking foundation stone that can be widely applied in other quantitative proteomics applications.

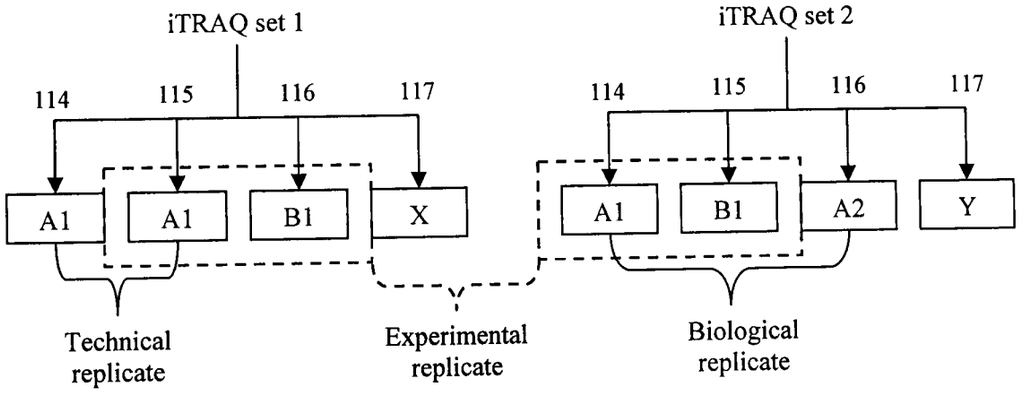


Figure 4.1. Illustration of the relationship between technical, experimental and biological replicates in iTRAQ experiments. A technical replicate has two identical samples from the same biological source in an iTRAQ set; whereas a biological replicate has two distinct biological samples from same condition in an iTRAQ set. An experimental replicate compares the variation of an identical sample in two different iTRAQ sets.

Table 4.1. The list of experiments and their designated role in replicates analysis, adapted from Chong *et al.* [39]. In total, there are 10 experiments with 5 technical replicates ([a-e]), 3 biological replicates ([f-h]) and 3 experimental replicates ([i-k]).

		iTRAQ reagents			
Species	Exp	114	115	116	117
Yeast					
S. cerevisiae	1	12% w/v glucose [a]	12% w/v glucose [a]	21% w/v glucose	30% w/v Glucose
Archaeon					
S. solfataricus	2	Glucose [b]	Glucose [b]	Glucose + 1.0% v/v ethanol	Glucose + 0.5% v/v Acetone
	3	Glucose	0.1% v/v ethanol	0.5% v/v ethanol	1.0% v/v ethanol [i]
	4	Glucose	1.0% v/v ethanol [f,g,i]	1.0% v/v ethanol [c,f]	1.0% v/v ethanol [c,g]
	5	Glucose	0.2% v/v n-propanol [h]	0.2% v/v n-propanol [h]	0.5% v/v n-propanol
Cyanobacterium					
Synechocystis sp.	6	24hr dark [d]	24hr dark [d]	24hr light [e]	24hr light [e]
	7	24hr dark	0.5hr dark	1hr dark [j]	1.5hr dark
	8	24hr dark	1hr dark [j]	6hr dark	11hr dark
	9	24hr dark	12.5hr light	13hr light [k]	13.5hr light
	10	24hr dark	13hr light [k]	18hr light	23hr light

4.3 Materials and Methods

4.3.1 Experimental design.

Ten iTRAQ experiments generated from *S. cerevisiae*, *S. solfataricus* and *Synechocystis* sp. yielded 5 technical replicates (1 set of *S. cerevisiae*; 2 sets of *Synechocystis* sp.; 2 sets of *S. solfataricus*), 3 experimental replicates (2 sets of *Synechocystis* sp.; 1 set of *S. solfataricus*) and 3 biological replicates (all from *S. solfataricus*) as depicted in Table 4.1. The biological purpose of these experiments was to investigate the alcohol metabolism of *S. cerevisiae* under high gravity fermentation; the alcohol metabolism of *S. solfataricus* using different types of alcohols and ketones; and the proteome response of *Synechocystis* sp. to changes in the 12hr light-dark cycle (Chapter 5 and 6). The technical replicates are exp 1 (115:114), exp 2 (115:114), exp 4 (116:117), exp 6 (115:114) and exp 6 (116:117). The experimental replicates are exp 3 (117:114) and exp 4 (115:114); exp 7 (116:114) and exp 8 (115:114); exp 9 (116:114) and exp

10 (115:114). The biological replicates are exp 4 (116:115), exp 4 (117:115) and exp 5 (116:115). The definition of these replicates is further illustrated in Figure 4.1.

4.3.2 Sample preparation.

All chemicals were purchased from Sigma-Aldrich (Gillingham, Dorset, UK) unless otherwise stated. *S. cerevisiae* was grown at 30°C with continuous stirring at 120rpm in a medium consisting of 5g/L yeast extract, 3g/L peptone, 5g/L KH_2PO_4, 1.5g/L NH_4Cl, 0.7g/L $MgSO_4$, 1.7g/L KCl, 5.8g/L casamino acids, 7.2g/L fresh yeast autolysate [385]; and different concentrations of glucose, as shown in Table 4.1. The system was set up to maximise ethanol production in *S. cerevisiae* by altering the glucose concentration. Cells were harvested at the late-exponential phase by centrifugation at 3,000×g for 5 minutes at room temperature.

S. solfataricus was grown at 80°C in pH 4.0 basal medium [386] supplemented with 25µl of Wolfe's vitamins stock [387]. The culture was grown with various concentrations of alcohols and ketones, either with or without 4.0g/L of glucose (Table 4.1) to investigate the ability of *S. solfataricus* in alcohols or ketones conversion. *S. solfataricus* was harvested at the mid-exponential phase by centrifugation at 5,000×g for 15 minutes at room temperature.

Synechocystis sp. was grown in BG-11 medium at 30°C on a 12 hour light-dark cycle with a light intensity of approximately 120 µEinstein/m^2s with continuous stirring at 200 rpm, and bubbling with air (with 3% CO_2) at 1.0 litre/min, for the study of cellular response in light-dark cycle (Chapter 5 and 6). Cells were harvested at the mid-exponential phase at different time points (Table 4.1) by centrifugation at 5,000×g for 5 minutes at 4°C (Detailed experimental workflow is available in Chapter 6).

4.3.3 Protein preparation.

The harvested *S. cerevisiae* cell pellet was washed twice with deionised water, and twice with yeast medium buffer consisting of 50mM 2-Morpholinoethanesulfonic acid (MES), 10mM

EDTA, and 10mM $MgCl_2$ at pH 4.5 [388]. The pellet was then re-suspended in two volumes of a solution consisting of yeast medium buffer and 5% v/v of a protease inhibitor cocktail. Proteins were extracted using one volume of glass beads (425-600 μm) coupled with vigorous vortexing for 9 minutes, with each cycle alternating between 45 seconds vortex and 45 seconds in an ice bath. The disrupted cells were centrifuged at 13,000×g for 5 minutes at room temperature, and the supernatant containing soluble proteins was recovered. After protein quantification (see below), 100μg of protein was precipitated using ice-cold acetone overnight at -20°C, followed by centrifugation at 21,000×g for 5 minutes at 4°C. The washed pellet was then re-suspended in 20μL of 500mM TEAB at pH 8.0, giving a final concentration of 5 μg/μL.

Cells harvested from *Synechocystis* sp. and *S. solfataricus* was re-suspended in 500mM TEAB buffer. *Synechocystis* sp. and *S. solfataricus* proteins were extracted using liquid nitrogen coupled with mechanical cracking [36, 389]. Proteins were recovered from the supernatant by centrifugation at 21,000×g for 30 minutes at 4°C. The total protein concentration from all cell extracts was measured using the RC DC Protein Quantification Assay (Bio-Rad, Hertfordshire, UK) according to the manufacturer's protocol. The extraction buffer used here (500mM TEAB) was different compared to a previous study (Chapter 3), due to the complementary nature of TEAB buffer with the rest of the iTRAQ workflow. TEAB buffer was the "Dissolution Buffer" in the iTRAQ kit, and was recommended by the Applied Biosystems. Furthermore, the previous extraction buffer contained Tris, DTT, Bio-Lyte and a protease inhibitor cocktail which would interfere with the iTRAQ reagent as well as the labelling procedures.

4.3.4 Isobaric peptide labelling.

100μg of protein in 20μL of 500mM TEAB from each phenotype was reduced, alkylated, digested and labelled with iTRAQ reagents according to the manufacturer's (Applied Biosystems) protocol with some modifications. These included a two-day tryptic digestion and two volumes of ethanol used during labelling. All samples were labelled as shown in Table 4.1. After a two-hour incubation period, labelled samples were then combined and dried in vacuum concentrator.

4.3.5 Strong cation exchange (SCX) fractionation.

Due to the complementary nature of SCX separation with the iTRAQ workflow, SCX was chosen ahead of other workflows described in Chapter 3. Furthermore, the labelling step occurred at the peptide level, not the protein level, therefore workflows with protein pre-fractionation, such as 1D-IEF and IEF-IEF, are not compatible. When comparing SCX-only and IEF-only, the total number of proteins identified in these two workflows were about the same (refer to Row 1 and 2 in Table 3.2). If the number of SCX fractions were to increase to 40, the total number of protein identified was the highest (as depicted in Row 8, Table 3.2). Dried peptides re-suspended in 200µL of Buffer A were fractioned using a PolySULFOETHYL™ A Column (PolyLC, Columbia, MD, USA) 5 µm particle size of 100 mm length × 2.1 mm id, 200 Å pore size, on a BioLC HPLC unit (Dionex, Surrey, UK) with a constant flow rate of 0.2ml/min and an injection volume of 200µl. Buffer A consisted of 10mM KH_2PO_4 and 25% acetonitrile, pH 3.0 and Buffer B consisted of 10mM KH_2PO_4, 25% acetonitrile and 500mM KCl, pH 3.0. The 60-minute gradient consisted of 100% A for 5 minutes, 5% to 30% B for 40 minutes, 30% to 100% B for 5 minutes, 100% B for 5 minutes and finally 100% A for 5 minutes. The chromatogram was monitored through a UV Detector UVD170U and Chromeleon Software, version 6.50 (Dionex/LC Packings, The Netherlands). Fractions were collected every minute using a Foxy Jr. Fraction Collector (Dionex), and later were pooled together according to manual manipulation of the chromatogram profile based on the peak intensity. The total pooled fractions for each experiment are shown in Table 4.1. These fractions were dried in vacuum concentrator, and stored at -20°C prior to mass spectrometric analysis.

4.3.6 Mass spectrometric analysis.

Each dried SCX peptide fraction was re-dissolved in 100µL of Switchos buffer (0.1% formic acid and 3% acetonitrile) and then 20µL of sample was injected to the nano-LC-ESI-MS/MS system for each analysis. Mass spectrometry was performed using a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF-MS/MS (Applied Biosystems, Framingham, MA, USA; MDS-Sciex, Concord, Ontario, Canada) coupled with an

online capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands) as described in a previous study [36]. The peptide mixture was separated on a PepMap C-18 RP capillary column (LC Packings) with a constant flow rate of 0.3μL/min. The LC gradient started with 3% Buffer B (0.1% formic acid in 97% acetonitrile) and 97% Buffer A (0.1% formic acid in 3% acetonitrile) for 3 minutes, followed by 3% to 25% or 30% Buffer B for either 60, 90, 120 or 127 minutes, then 90% Buffer B for 7 minutes, and finally 3% Buffer B for 8 minutes. Triplicate independent LC-MS/MS injections were made for all samples. The mass spectrometer was set to perform data acquisition in the positive ion mode, with a selected mass range of 300 – 2000 m/z. Peptides with +2 to +4 charge states were selected for tandem mass spectrometry and the time of summation of MS/MS events was set to be 3 seconds. The two most abundant charged peptides above a 5 count threshold were selected for MS/MS and dynamically excluded for 60 seconds with ±50mmu mass tolerance.

4.3.7 Data analysis.

Peptide identification and quantification was carried out using ProQuant software v1.1 (Applied Biosystems; MDS-Sciex). The search was performed against the Mixed Genome Database consisting of *S. cerevisiae* (6298 ORFs), *S. solfataricus* (2995 ORFs), *Synechocystis* sp. (3264 ORFs), *Escherichia coli* K12 (4237 ORFs) and 222 keratins extracted from human and mouse [39]. The search parameters allowed for peptide and MS/MS tolerance up to 0.15 Da and 0.1 Da respectively; one missed cleavage of trypsin; oxidation of methionine and cysteine modification of MMTS. Only peptides above a 70% confidence were saved for identification and quantification. ProGroup Viewer software v1.0.6 (Applied Biosystems, MDS-Sciex) was used to group and identify proteins with at least 95% confidence. The peptide list obtained from ProGroup Viewer was exported to Microsoft Excel for further analysis. Shared peptides; proteins with 1 MS/MS; alien proteins (proteins identified from organisms other than the test subject itself) and peptides without quantitation were further removed from the peptide list. All peptide ratios were converted to $\log_{10}$ space, and average protein expression was estimated using the following equation adapted from the ProQuant software tutorial:

$$\text{Weighted mean of log ratio} = \frac{\sum_{i=1}^{N}(w_i \times x_i)}{\sum_{i=1}^{N} w_i} \quad (1)$$

Where w_i is 1/(%Error of the peptide)$_i$ and x_i is the $\log_{10}$ (peptide ratio)$_i$. The % Error was calculated by the ProGoup Viewer, representing an estimate of the error in the quantitation that is determined for each ratio. The weighted standard deviation (SD_w) was estimated using the following equations also adapted from the ProQuant software tutorial:

$$\text{Weighted standard deviation, } SD_w = \frac{SD}{b^{0.5}} \quad (2)$$

$$\text{Where SD is unweighted standard deviation and b} = \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2} \quad (3)$$

4.4 Results and Discussion

4.4.1 Quality of the data.

The data used in this study was carefully interpreted and filtered to ensure the accuracy of the measurements in replicate analysis. The use of the Mixed Genome Database has been demonstrated previously on this dataset to be very efficient in eliminating false positive identifications down to less than 3%, and even to below 1% in some organisms (*Synechocystis* sp. and *S. solfataricus*) [39]. The peak area under the reporter ion, which gives the relative (or absolute) protein expression, was limited to those with a peptide confidence of 70% and above, not to mention other criteria pre-set in the ProQuant software, such as the peak area being greater than or equal to 40. Peptides with no quantification value, or having values of 9999 or 0 (the absence of one of the reporter ions during quantification) were removed. Each case was subjected to manual inspection to ensure the loss of one of the reporter ions is not from an intrinsic biological effect. So far no cases support or provide sufficient evidence that a peptide could show an expression value of 9999 or 0. Shared peptides among proteins are common, which sometimes confuses the type of regulation exhibited by their parental proteins. To avoid this scenario, any quantitation contributed by the shared peptide is not considered. In order to

comply with the Paris Consensus, the general guideline for proteomics data publication, detailed in the *Molecular and Cellular Proteomics Journal* (available on the website, http://www.mcponline.org/), single MS/MS identification was not taken into account. However, peptides found repeatedly in different LC-MS/MS injections which contribute to the overall peptide count of more than 1 MS/MS experiment were considered in identification and quantification. Full details regarding the implication of multiple injections was recently reported in previous study [39]. The full list of data in each experiment can be found on the website at http://wrightlab.group.shef.ac.uk/projects/, or in the Appendix D.

The problem of working in linear space during quantitation is the uneven peak distribution at the range of the standard deviation. Hence all peptide ratios obtained were converted into $\log_{10}$ space before compiling them into the calculation for the overall protein ratio, and the estimation of the standard deviation. In fact, the estimated standard deviation of each ratio is an independent assessment of the MS variation. The average of all standard deviations calculated from each protein was then used to determine the MS variance as depicted in Table 4.2. The protein expression ratio was normalised against the percentage error under the peak as shown in equation 1, to remove bias caused by the difference in each label. Detailed calculation steps to derive the relative iTRAQ protein ratio from its corresponding peptide are illustrated in Appendix E.

In total, more than 300 valid proteins quantifications were recorded for *S. cerevisiae* and *S. solfataricus* as shown in Table 4.2. The number of proteins used in this study for *Synechocystis* sp. was significantly less, at approximately 200. This is due to the abundance problem in pigment proteins, where it have previously shown in a qualitative study that only 776 protein identifications were obtained for *Synechocystis* sp. from 6 different shotgun pre-fractionation techniques [36]; unlike its counter-part *S. solfataricus* which does not have the abundance problem, where a total of 1399 proteins identifications were reported with only 2 different shotgun methods [389]. The challenge in protein abundance for cyanobacteria is yet to be

addressed fully. Only proteins found in both sets of experiments in iTRAQ replicates were used for analysis (Table 4.2).

Table 4.2. Summary of the number of proteins used in this analysis, MS variance of individual experiments and the percentage of variation for technical, experimental and biological replicates. For iTRAQ replicates, only proteins found in both sets of experiments were used.

Technical replicates

	Label	No. of proteins	MS variance	% variation
Exp 1	115:114	316	9.0×10^{-04}	10.4
Exp 2	115:114	444	1.4×10^{-03}	12.8
Exp 4	116:117	329	1.2×10^{-03}	12.4
Exp 6	115:114	196	1.1×10^{-03}	10.9
Exp 6	116:117	200	5.8×10^{-03}	7.2
	Average		1.0×10^{-03}	10.7

Experimental (iTRAQ) replicates

	Label	No. of proteins	MS variance	% variation
Exp 3	117:114	252	2.4×10^{-03}	27.2
Exp 4	115:114		2.3×10^{-03}	
Exp 7	116:114	110	2.1×10^{-03}	23.4
Exp 8	115:114		1.6×10^{-03}	
Exp 9	116:114	97	7.8×10^{-03}	19.6
Exp 10	115:114		2.5×10^{-03}	
	Average		2.0×10^{-03}	23.4

Biological replicates

	Label	No. of proteins	MS variance	% variation
Exp 4	116:115	325	1.7×10^{-03}	23.7
Exp 4	117:115	325	1.4×10^{-03}	23.1
Exp 5	116:115	257	3.6×10^{-03}	28.9
	Average		2.2×10^{-03}	25.2

4.4.2 Principle of replicate analysis.

The purpose of replication is to get as precise an estimate as possible of the effect of interest. It seems impossible for a scientist to perform a perfect experiment. The existence of random variations measured from the start, during, after the experiment, either contributed directly from the test subject itself, or indirectly from the method or instrument variance, always poses a threat to the quality of the information obtained. During experimental design, the variation within the layout often over-looked. By simply assuming the variation is small, or negligible, can sometimes cause fatal consequences to the experimental outcome. In this study, some

possible sources of variation in an iTRAQ experiment were identified, measured in terms of technical, experimental and biological variations. The use of technical replicates is important to establish the significance in protein expression level, especially under tightly-regulated environments. The experimental replicate is the actual iTRAQ replicate, the repetition of the same samples in two or more experimental sets, and they must have the same reference point or control. Biological replicates are used to examine the variation of random biological effects.

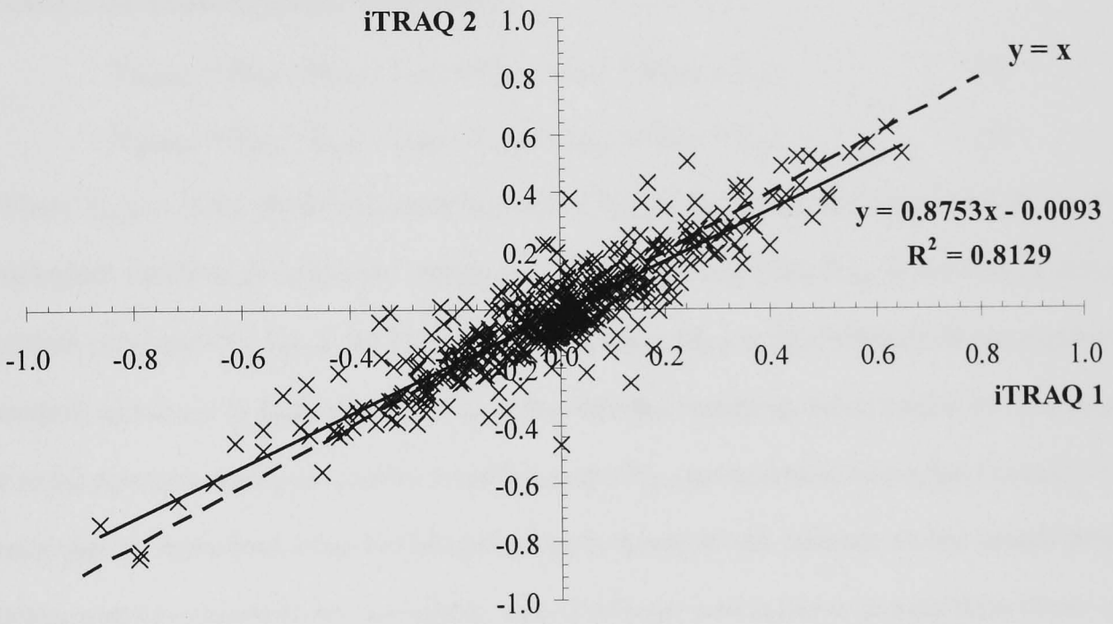


Figure 4.2. The theoretical relationship between an experimental replicate (y = x; dotted line) against the best-fit line (y = 0.8753x – 0.0093; solid line) generated from the experimental data. The slope of the best-fit line should have a value as close to unity as possible, whereas the intercept should cross at 0.

Since replicate analysis involves the use of the same biological sample, either from the same or a different sample pool. Their theoretical relationship should be X = Y, i.e. the slope of the graph should be unity (slope of 1) or in another words, the division between these values should be equal to one. For experimental replicates, the protein expression ratio in both iTRAQ experiments should be similar, or ideally the same (Figure 4.2) whereas for technical and biological replicates, the ratio of protein expression in the iTRAQ experiment should be 1.0. Any variations deviating from unity would allow us to measure the random or structural

nuisance effects from different types of replicates. This assumption holds true provided the test subject comes from the same condition, however the effect from random conditions has yet to be studied, before drawing any conclusions on the influence of replicate variations. Nonetheless, one could assume the variations measured in this study would be indicative of those iTRAQ experiments with random or different phenotypes.

The actual protein expression level is normally hidden by many existing variations. Let us consider the following simple relationships:

$$Y_{protein1} = X_{true} + B_{var1} + L_{eff1} + P_{conc1} + I_{dev1} + M_{diff1} + T_{var1} \quad (4)$$

$$Y_{protein2} = X_{true} + B_{var2} + L_{eff2} + P_{conc2} + I_{dev2} + M_{diff2} + T_{var2} \quad (5)$$

Where $Y_{protein'i'}$ is the observed protein expression; X_{true} is the true protein expression; B_{var} is the biological variation of individual sample; L_{eff} is the labelling bias; P_{conc} is the bias in initial protein concentration; I_{dev} is the MS injection deviation; M_{diff} is the difference in measurement method, technique or instrument; and T_{var} is the technical variations within a replicate. Subscript 1 and 2 represent duplicate samples 1 and 2. L_{eff} and P_{conc} are structural biases that can easily be removed or normalised using bioinformatics tools, however this depends on the experimental design and bioinformatics tools available, since ProQuant is designed to include these biases by taking into considerations of L_{eff} (and indirectly P_{conc}). I_{dev} can be minimised by multiple injections [39]. This leads us to the remaining variations of B_{var}, M_{diff} and T_{var} before X_{true}, the actual protein expression can be assessed, as seen in equations 6 and 7.

$$Y_{protein1} = X_{true} + B_{var1} + M_{diff1} + T_{var1} \quad (6)$$

$$Y_{protein2} = X_{true} + B_{var2} + M_{diff2} + T_{var2} \quad (7)$$

For the study of technical replicates, the same sample from the same culture pool was used, which eliminates the technical variation T_{var}, but the protein expression is still hindered by the biological (B_{var}) and experimental (M_{diff}) differences (equation 8).

$$Y_{protein} = X_{true} + B_{var} + M_{diff} \quad (8)$$

An experimental replicate reduces the effect of M_{diff}, and using biological replicates, B_{var} is reduced. In this study, the estimated experimental (iTRAQ) variation has a smaller observation than the biological variation (Figure 4.4). Therefore it is possible to assume:

$$B_{var} > M_{diff}. \quad (9)$$

Hence, equation 6 and 7 can be simplified to the following form:

$$Y_{protein} = X_{true} + B_{var} \quad (10)$$

This argument is true, provided the variation from the same sample in different iTRAQ experiments exhibits a smaller divergence from the real value as the random biological effect in an iTRAQ experiment. It is not known yet what the combination effects of the biological replicate in two or more set of iTRAQ experiments is, instead of using same sample in different experiments as illustrated for the experimental replicate.

4.4.3 Technical variation.

This variation ranges from protein extraction and protein quantification through to tryptic digestion. One would assume the variations through these steps would be minor, if carried out carefully. Using iTRAQ as the study subject, an average of ±11% variation in protein expression at the technical level was detected, as shown in Table 4.2. This is in agreement with the findings from Redding *et al.*, where the internal error (technical replicate in an iTRAQ experiment) was found to be ±0.12 [212]. A similar technical replicate application employed in an iTRAQ study of lung cancer cells [335] and FDCP-Mix cells [340] also reported a high consistency between the labels. A separate iTRAQ study of human fibroblast revealed a standard deviation in log space ranging from 0.039 to 0.067 [332]. In order to accurately measure the variation at this level, MS variance in each experiment was recorded, and the variance was reported at an average of 0.1% (Table 4.2), which suggests the MS performance measured by triplicate injections does not have significant role in this replicate analysis. Aside from *S. cerevisiae*, each organism contributes an equal amount of data, with *S. solfataricus* exhibiting a slightly higher level of average variation at ±12.6% compared to the rest.

An average of ±11% variation across all three organisms does not indicate the accuracy of the measurement. It only reveals the average dispersion caused by the technical variations. To be able to measure the accuracy and the percentage of data expression that will fall within the acceptable error range, each iTRAQ result is categorised into groups of variation ranging from

10% to 100%. At ±10% variation, 63% of the expression values fall within this range. In other words, 63% of the proteins identified and quantified had a variation of expression of ±10%. In general, the number of proteins covered increases as the allowed variation increases. The coverage level saturated at approximately ±50% variation where more than 99% coverage of the identified proteins occurred (Figure 4.3). In this instance, it is recommended a cut-off point of ±30% variation for a 95% coverage in protein expression, which is well above the average variation of ±11%.

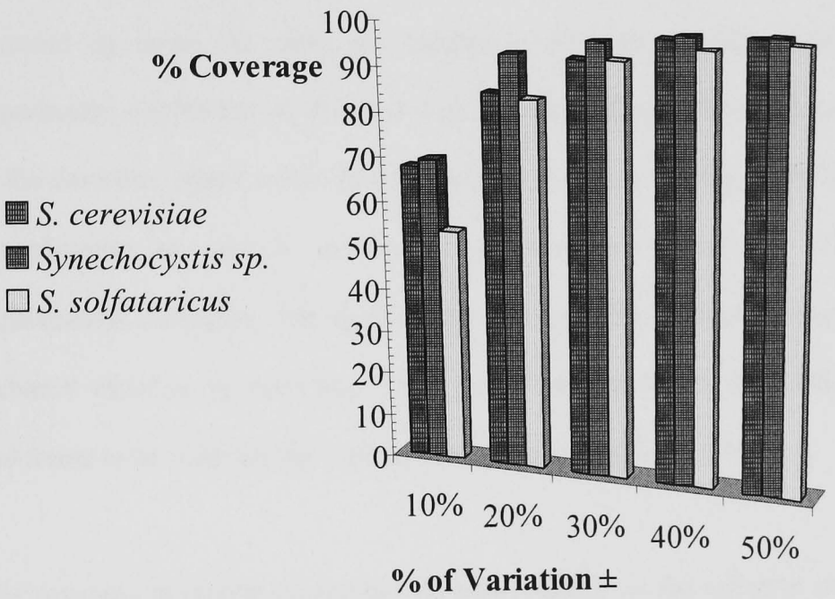


Figure 4.3. The percentage coverage at different variation levels for technical replicates. A total of 5 replicates were used with 2 each from *Synechocystis* sp. and *S. solfataricus* respectively, and 1 set from *S. cerevisiae*.

A similar concept of assessing the associated variations (from sample preparation to image analysis) in a gel-based experiment (2-DE) was also reported by Choe and Lee [390]. In the investigation, at least 95% spots presented in three out of four replicate gels had a cut off CV of 0.52. Furthermore, another study was conducted using the DIGE system, where two identical gels (where an identical protein sample is utilised across the multi-gel experiments) were used to estimate the number of replicates required to produce a target power of 0.8 at different fold

change level [325]. The diversity across three different domains of life (eucarya, archaea, bacteria) has allowed us to form a better understanding of the variations at this level, and this trend should certainly be generally applicable to other organisms, although one would argue the exact proteomic characteristic of each organism is different from each other.

4.4.4 Experimental (iTRAQ) variation.

The ability to measure the discrepancy across the information generated from different experiments can greatly enhance the reproducibility of the subject, and the reliability of the platform used to measure the changes in the subject variation. In this instance, this variation is assessed by using the same test sample in different iTRAQ experiments. These sets of experiments carried out on different days by different operators allow us a better measurement of the deviation which comes directly from these effects. Based on the datasets generated from *Synechocystis* sp. and *S. solfataricus*, an average variation of ±23% was measured in experimental replicates. The significance of experimental variation has increased the existing technical variation by more than 100%. Again the contribution from MS variance was recorded and found to be insignificant with an average of 0.2% (Table 4.2).

The coverage in protein expression increases steadily as the variation increases, and eventually saturated at a variation level of ±100% with 98% coverage (Figure 4.4). This saturation point is significant higher than the technical variation. At ±30% variation, 78% coverage interval in protein quantification was recorded; whereas 89% coverage was noted for ±50% variation (Table 4.2). The increment in coverage above ±50% variation was significantly smaller compared to the variation below 50%; therefore a cut-off point at ±50% variation for experimental variation does seem reasonable and acceptable with a decent coverage interval. With the use of experimental replicates in iTRAQ experiments, it becomes possible to cross-compare the data produced from other research groups, which greatly enhances the procedure for the peer-reviewed process for information validation, and should encourage collaboration for an investigation of the same study subject among researchers. This is not to mention the

divergence in methods and instruments that can easily be assessed using this variation estimation. Figure 4.2 shows a relatively high reproducibility from 3 experimental replicates with a slope of 0.88.

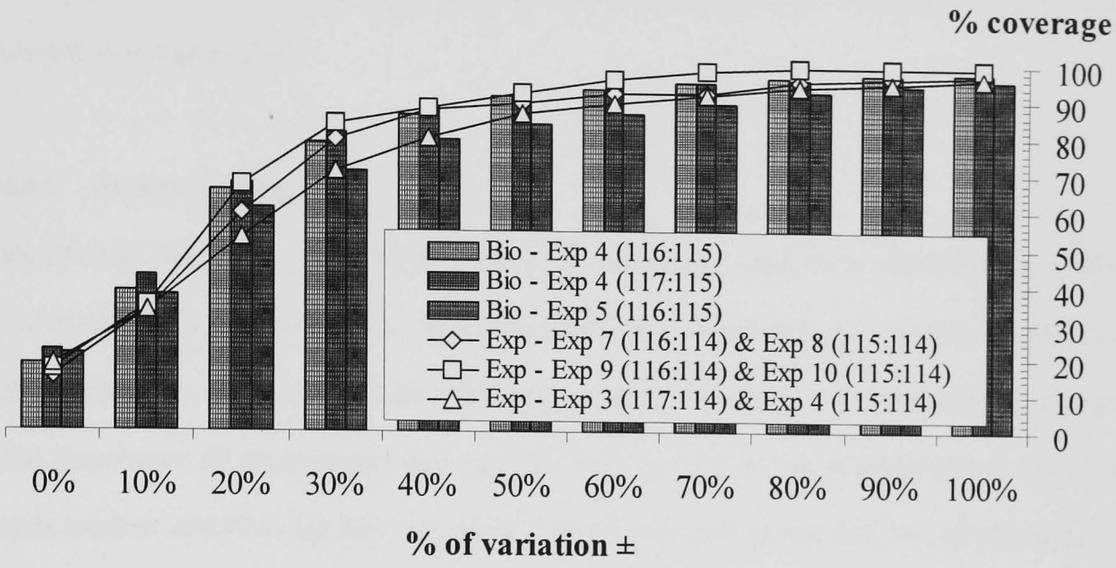


Figure 4.4. The percentage coverage at a different variation level for biological replicates (represented by bars) and experimental replicates (represented by lines). These two types of replicates share a similar distribution across different variation intervals.

4.4.5 Biological variation.

The biological effect measured as the biological replicate is the most likely source of variation. In this study, the biological variation in *S. solfataricus* has been investigated with a total of 3 datasets from 2 different iTRAQ experiments. An average variation of ±25% was identified, sharing close proximity with the experimental variation at ±23% (Table 4.2). At 30% biological variation, 77% coverage level in protein expression was noted with MS variance of 0.2%. The recommended cut-off point is at ±50% variation when the coverage level hit 88% (Figure 4.4).

Among all the replicate types studied here, unsurprisingly, the biological variation has the highest variation with the lower number of proteins covered. This increases the need for the use of biological replicates in protein expression studies as the main variation source, since the technical and experimental (iTRAQ) variations are smaller or at least equivalent to the random

biological effects. A similar observation was noticed in the DIGE system, where the technical variance was low and the main attention was focused on biological replicates [384]. By increasing the cut-off point from ±30% in technical variation to ±50%, there is more certainty that the observed protein expression is the true expression, albeit covering 88% of the proteins identified and quantified.

4.4.6 Sample pooling effect.

Significant efforts have focussed specifically on microarray analysis to address the reliability and quality of the dataset [391, 392]. The proteomics community can clearly draw on the philosophy and tools developed for microarray analysis and apply it in the proteomics content. The importance of experimental design [393, 394], as well as the development of statistical tools (such as ANOVA) for data analysis [395-397], were also among the investigated issues. In a recent study, global gene expression using cDNA microarrays compared results obtained from different laboratories and platforms [398]. The reproducibility of this data is very much depended on the standardized protocol followed during sample preparation until the data normalization point. At least three replicates was recommended for a single microarray experiment [399]. It is common in cDNA microarray applications to pool or mix more than one biological replicate before measurement, in an effort to reduce the impact of biological variation [400-403]. More than 15% of the datasets deposited in the Gene Expression Omnibus database involve pooled RNA samples [404]. It is a possible alternative to biological replicates when sample limitation and financial constraints cause a problem. A pooling strategy in DIGE system was investigated by Karp *et al.*, where consideration of different types of statistical analysis was shown to be crucial when using a mix of replicates, as one-way ANOVA can often cause more false positive identifications compared to nested ANOVA [384]. This pooling strategy was recently applied by Neubauer *et al.* to identify differential protein abundances between invasive ductal carcinoma cells from cryo-preserved ER+/PR+ and ER+/PR- mammary tumor specimens [405]. A similar pooling strategy was also employed in an iTRAQ examination of hematopoietic stem cells, where the protein starting template was generated from a total of 40 mice [334]. The iTRAQ examination of nitrate stress in *Desulfovibrio vulgaris* Hildenborough

was from a combination pool of triplicate controls and test samples [212]. Most biomedical research, which require higher precision in fold-change measured as a community effect, always involves pooling of the starting template, for example as reported in the study assessing the effects of the diurnal variation of human parotid saliva [343].

The random biological variation in one sample pool is generally lower than one set of individual samples, as the biological variation can be normalised by *n* samples before being introduced into the experiment. However, the sample pooling strategy does not remove the technical variation within each replicate T_{var}. This in agreement with the experimental design from Redding *et al.* where a technical replicate was included, despite having a pooled sample [212]. A recent study of microarray experiments using the MCF-7 breast cancer cell line, compared individual samples and pooled samples and found that pooling might not be helpful, and could even decrease one's ability to identify differentially expressed genes where biological variability is expected to be small [406]. Pooling also can result in the loss of information, as reported in a microarray study, approximately 50% more transcript expression changes were observed in the individual sample compared to the pooled sample [407]. On the other hand, a recent DIGE study comparing individual and pooled sample of healthy human liver, reported the CV ranged from 6.4 to 108.5% with the median CV of 19% which indicate the protein expression among each test subject is relatively stable [408]. It is undeniable that there is potential for this strategy for protein expression studies employing iTRAQ. With proper pooling strategy, this approach should be considered more widely for future iTRAQ experimental design, or even for other protein quantification techniques.

4.4.7 Multiple injections [Ψ]

A previous study had reported that the increment of unique proteins per injection will decrease as the injection number increases [389]. The same observation was noted across the studied organisms, where a higher increment of unique proteins was achieved in the second injection compared to the increment in the third injection, except for *S. solfataricus* giving the same unique protein increment in both the injections (refer to Figure 4.5). Based on two injections, a 5% increment in unique proteins was found in *S. cerevisiae*, 14% in *S. solfataricus* and 33% in *Synechocystis sp.* This was followed by a lower unique protein increment of 2% and 13% for *S. cerevisiae* and *Synechocystis sp.* obtained from the third injection. *Synechocystis sp.* had the highest overall increment in the total number of proteins identified at approximately 50%, followed by *S. solfataricus* at 30% and *S. cerevisiae* at 6% (Figure 4.5) via three analyses. However, *Synechocystis sp.* had the least total confident proteins identified, an average of 218 proteins, compared to *S. solfataricus* with an average of 423 proteins and 380 proteins from *S. cerevisiae.*

One of the major challenges in *Synechocystis sp.* is the abundance of the pigment proteins, belonging to the phycobilisomes family used to capture energy from sunlight for photosynthesis [36]. Among all the *Synechocystis sp.* iTRAQ sets, more than five of the top ten listed proteins belong to the phycobilisome protein group, in particular the phycobilisome core-membrane linker polypeptide ApcE (slr0335) being the highest ranked protein, followed by most popular hits of phycocyanin subunit A (sll1577) and phycocyanin subunit B (sll1578), with more than 100 peptide-spectral matches being correctly identified. On the other hand, *S. cerevisiae* shows high reproducibility in protein identification, by giving the highest number of correctly identified spectra, approximately 18,068 (despite the overall lowest spectra increment in three

---

[Ψ] The detailed study of multiple injections using iTRAQ approach has been published in the Journal of Proteome Research (January 2006), volume 5, issue 5, page 1232-40. This is a collaboration work between CS Gan (the author), PK Chong and TK Pham. This publication is led by PK Chong.

analyses, shown in Figure 4.5) compared to an average of 7,222 and 1,900 found in *S. solfataricus* and *Synechocystis sp.* respectively.

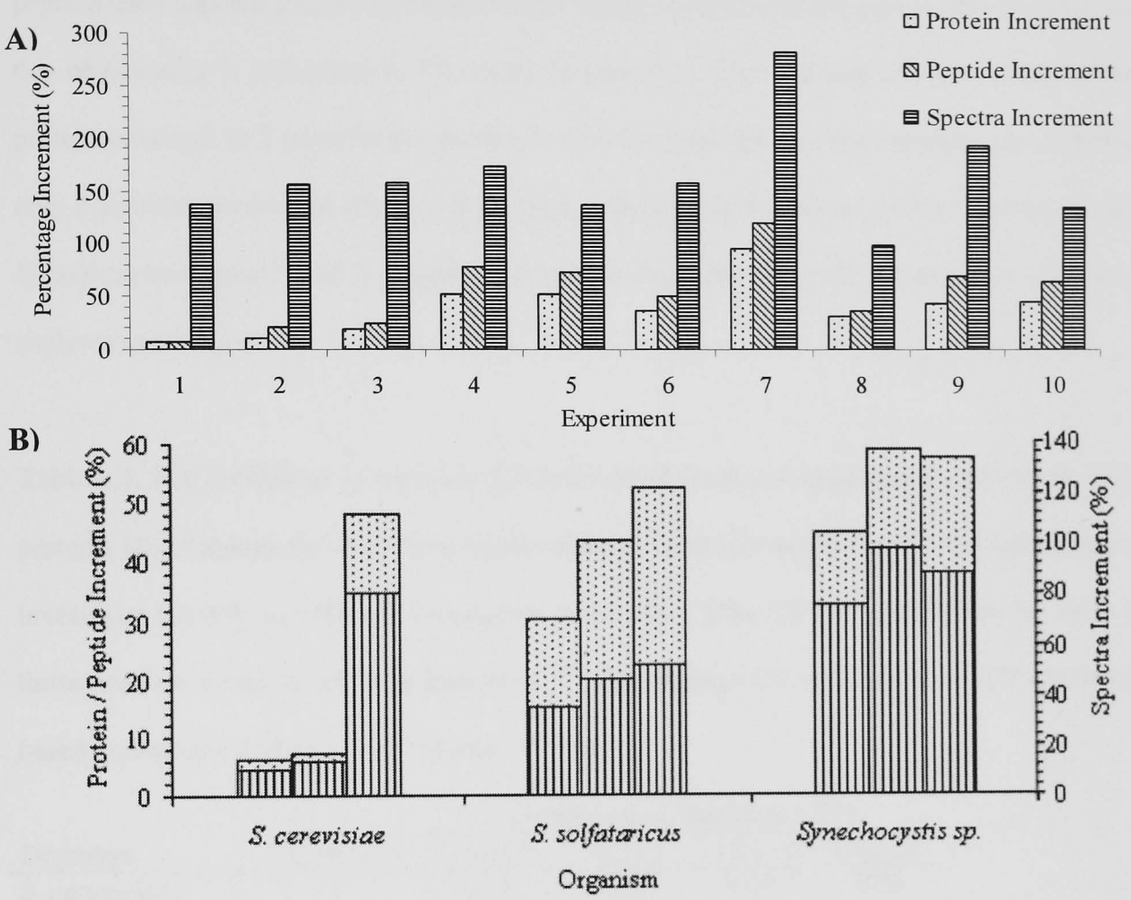


Figure 4.5. (A) The overall percentage increment in the total number of unique proteins, unique peptides and spectra identified for each iTRAQ experiments via three analyses. (B) The average percentage increment in total number unique proteins, unique peptides and spectra identified in (i) two analyses, which was calculated by comparing the result obtained from two injections (1st and 2nd injection) to the first injection; and (ii) three analyses, the comparison between the result obtained in all the three injections to the total result obtained from the first two injections (1st and 2nd injection).

Higher numbers of peptides confidently identified across the replicate analyses ought to lead to greater numbers of proteins identified. Again, *Synechocystis sp.* gave the highest increment in

unique peptides identified across the analyses, with an average at 68%, followed by an average of 50% for *S. solfataricus,* whereas the lowest increment was observed in *S. cerevisiae*, with an average of 7% (Figure 4.5). According to Elias *et al.* [409], the percentage increment at the peptide level was not parallel and significantly higher compared to the gain in protein level, and this observation is consistent in this study (Figure 4.5). The multiple analyses enhanced the protein coverage to 3 peptides per protein in *Synechocystis sp.* via three analyses compared to only 2 peptides per protein obtained from either 1 analysis or 2 analyses, while *S. cerevisiae* and *S. solfataricus* gave 5 and 3 peptides per protein respectively in all the analyses (1, 2 or 3 analyses) carried out.

Table 4.3. The coefficient of variation (CV) calculated for the 10 experiments carried out. Only proteins found repeatedly in all three injections were taken into account in the CV calculation to investigate the MS and iTRAQ variance in this study. ([a])The CV calculated from log ratio of those proteins found in all three injections. ([b]) The average CV obtained from CV calculated based on log ratio 115:114, 116:114 and 117:114.

		Coefficient of Variation (CV)			
Organism	Experiment	115:114[a]	116:114[a]	117:114[a]	Overall[b]
S. cerevisiae	1	0.04	0.08	0.09	0.07
S. solfataricus	2	0.06	0.06	0.05	0.05
	3	0.06	0.05	0.06	0.06
	4	0.13	0.09	0.11	0.11
	5	0.10	0.08	0.09	0.09
	Average	0.09	0.07	0.07	0.08
Synechocystis sp.	6	0.05	0.12	0.12	0.10
	7	0.11	0.08	0.10	0.10
	8	0.09	0.06	0.06	0.07
	9	0.13	0.05	0.14	0.11
	10	0.10	0.10	0.08	0.09
	Average	0.10	0.08	0.10	0.09

For each iTRAQ experiment, the CV calculated for those proteins found repeated across injections scattered from 0.14 to 0.04, for all the species studied (Table 4.3). The average CV of the 3 injections within the experiments (insight into the MS variance) for *S. cerevisiae*, *S. solfataricus* and *Synechocystis sp.* were 0.07, 0.08 and 0.09, whereas the average CV of the experiment (insight into the iTRAQ variance) for *S. solfataricus* and *Synechocystis sp.* were less

than 0.1. The average CV calculated from those proteins found repeated in all three analyses was 0.08 compared to 0.10 for those proteins identified in any two analyses. This consistency suggested the injection replicates have a higher impact in protein ID evidence compared to its quantification counterpart, which further strengthen the use of multiple analyses in iTRAQ.

4.5 Conclusions

The analysis used in this study is simple and straight-forward, without the use of sophisticated statistical tools, although these can clearly be applied. The derivation of differential protein expression in iTRAQ is difficult. It is crucial at this stage not to introduce overly complex validation tools or analysis that would create a more complicated situation. Based on replicate analysis, the significance of protein expression in iTRAQ can be assessed according to different types of replicates. It forms a sturdy validation tool required for iTRAQ experiments.

The study has successfully identified and estimated the variation associated with iTRAQ experiments, either expressed as technical or biological effects. The technical associated deviation was small, but measurable with an average ±11% variation, yet the recommended cut-off point was at ±30% variation where the protein expression coverage reached 95%. Comparing this value to the variation caused by random biological effects, a higher than average variation was detected at ±25% and a higher recommended cut-off point at ±50% where the coverage of identified and quantified proteins was 88%. iTRAQ (experimental) replicates exhibit a similar coverage distribution across the range of variation as seen for biological replicates. The average variation was at ±23% and a cut-off point was at ±50% variation with an 89% coverage interval. This suggests most of the measurable differences in experiments were biologically-associated rather than caused by structural nuisances such as time, operator or even location. The findings in this study prove and strengthen the use of biological replicates in protein expression experiments. Therefore, it is highly recommended to include at least one biological replicate in an iTRAQ experiment.

The study also demonstrated the contribution of MS variance as being insignificant at 0.1-0.2% after triplicate independent MS injections. It must be reiterated that this variance was measured based on the instrument settings, and could vary differently across different platforms. Sample pooling was not investigated in this study, yet it has the potential to reduce the random biological variation by combining more than one biological sample. Since this strategy is well-established in cDNA microarray technology, there is no reason why the pooling would not be applicable in protein expression experiments. Furthermore, since iTRAQ reagents are relatively expensive, as is MS time, pooling samples would definitely help to reduce the experimental cost.

Analysing iTRAQ samples multiple times improves both the protein and peptide identification. The percentage increment in unique protein identification via multiple injections was found to vary between in different organisms, yet a reasonable improvement in proteome coverage was achieved. At least two injection replicates in qualitative and quantitative shotgun proteomics is recommended for higher proteome coverage. The iTRAQ quantification was highly consistent across the multiple injections; with an average CV of 0.09 (spreading across 0.14 to 0.04) obtained from log mean average ratio for all the organisms studied. Thus multiple analyses of iTRAQ samples are strongly encouraged for greater peptide identification and quantification.

Chapter 5

Central Carbon Metabolism Reconstruction of *Synechocystis sp.* PCC 6803 During Light Acclimation

5.1 Abstract

The light-dependent metabolism of *Synechocystis* sp. PCC 6803 was investigated via the use of an iTRAQ-mediated quantitative proteomic approach. Central carbon metabolism – glycolysis and gluconeogenesis together with the carbon fixation pathway (Calvin cycle), showed enhanced protein expression in the light compared to the dark. Similarly, when illuminated, enhanced expression was also observed in proteins involved in photosynthesis. The relative protein expression differed between Photosystems I and II proteins, suggesting the possibility that Photosystem I is not modulated under the same mechanism as Photosystem II during photosynthesis. Cellular stress-response proteins such as GroEL1, GroEL2, GroES (slr2076, sll0416, slr2075) and ClpP4 (slr0164) were also amongst the proteins found to be up-regulated during the light cycle. A set of 23 selected genes from major parts of metabolism were used in quantitative real-time RT-PCR experiments to compare to results obtained at the proteomic level. The results of this comparison reveal that approximately one third were regulated similarly at both the protein and gene level, while the remaining two thirds showed contrasting differential expression changes. GroEL1 showed the most extreme difference, being 2-fold up-regulated in relative protein abundance, whilst an almost 3-fold down-regulation in mRNA abundance was observed under the same shift in conditions. These results generally indicate a likely role for post-translational modifications, or complex system-wide inputs into the regulation of such an important metabolic process within this organism. The association of other light-dependent proteins using transcriptional and translational comparisons also discussed.

5.2 Introduction

All living organisms respond towards a change in their environment through a set of proteins/genes essential for cell adaptation and survival. Light is the main driving-force for photosynthetic organisms such as cyanobacteria, algae and higher plants. It is fundamental to understand this phenomenon which allows these organisms to carry out photosynthesis and other light-associated processes. The ability to assimilate atmospheric carbon dioxide, and potentially turn one of the major contributors to the green-house effect into products (as described in Chapter 2) has been a fascinating topic of research for many years [41, 42, 410]. Towards this end, genome sequencing efforts, and in this case the sequencing of *Synechocystis* sp. PCC 6803 in 1996, has been a main motivator in developing this species as a type strain in this regard [51]. This has enabled researchers to initiate a system-wide analysis, and enhance the understanding of this organism. The understanding of global gene expression changes using DNA microarray techniques during light acclimation in *Synechocystis* sp. PCC 6803 has been studied extensively [65, 66, 411]. However, only recently the focus has switched from analysis at the transcript level to the protein level, where Kurian *et al.* used 2-DE gels to understand differential protein expression between light and dark cycle phenotypes [67]. However, the study only revealed 22 proteins with significant differential regulation based on stain-intensity, which in principle could be a source of variation [267, 268, 380, 412].

In the light, photosynthetic micro-organisms undergo various states of energy harvesting and carbon fixation, ultimately to produce ATP and to regenerate NAD^+ or $NADP^+$ to its reduced form of NADH or NADPH. The involvement of the gluconeogenesis and pentose phosphate pathways during the light-dark cycle has been reported widely [61, 62]. Yet, the TCA cycle is incomplete in most cyanobacteria, including *Synechocystis* sp. [413]. During the light period, the Calvin cycle is vital for carbon fixation from the atmosphere, as well as the generation of glyceraldehyde-3-phosphate, the basic building block from which these micro-organisms can synthesise a wide variety of compounds. The extra energy generated through this process is then converted and stored as glycogen. The sustainability of these micro-organisms in the dark

depends on the amount of stored glycogen, since this is the main energy source for processes such as respiration, and the generation of precursors for amino acid biosynthesis.

Here, *Synechocystis* sp. PCC6803 was maintained in constant light for 24 hrs, either at high or low light. The focus of this chapter is in developing an understanding of the *Synechocystis* sp. proteome during an extended photoperiod utilising iTRAQ-based quantitative proteomics. In this chapter, particular attention was given to central carbon metabolism during the light period of the light-dark cycle. iTRAQ-mediated quantitative proteomics, together with quantitative real time-PCR techniques were employed, in an attempt to re-construct some of the relevant metabolic pathways, and to enhance knowledge of this organism at a number of scales. A comparison of the relative merits of iTRAQ and 2-DE approaches for quantitative proteomics is also provided.

5.3 Materials and Methods

5.3.1 Growth conditions.

A batch culture of *Synechocystis* sp. PCC 6803 was grown in BG-11 media in a 2 litre Braun BioStat B+ fermentor (Sartorius BBI, Melsungen Germany) at 30°C, with a 1.5 litre working volume in constant light. Light was supplied by cool-white fluorescent lamps at an incident intensity of 120 μEinstein/m^2s. The light intensity was measured using a QSL-2100 Scalar PAR Irradiance Light Sensor (Biospherical Instrument Inc, San Diego, CA). The fermentor was stirred continuously at 200 rpm. Pre-humidified (from a humidifier filled with sterile water) 0.2 μm filtered air with 3% CO_2 was supplied at 1.0 litre/min. During the mid-exponential phase (OD_{730} of 3.5), the fermentor was switched to full dark (<3 μEinstein/m^2s) for 24 hours before switching back to full light (120 μEinstein/m^2s). Samples were collected at the end of full dark and full light conditions respectively. All chemicals were purchased from Sigma-Aldrich (Gillingham, Dorset, U.K.) unless otherwise stated.

5.3.2 Protein sample preparation.

Synechocystis sp. cells harvested after 24 hours of light and 24 hours of dark were re-suspended in 500 mM triethylammonium bicarbonate (TEAB) buffer, pH 8.5 and subjected to liquid nitrogen extraction coupled with mechanical cracking using a sterilised mortar and pestle [36]. Soluble proteins were recovered from the supernatant by centrifugation at 21,000 × g for 30 minutes at 4°C. The total soluble protein concentration was measured using the RC DC Protein Quantification Assay (Bio-Rad, Hertfordshire, UK) according to the manufacturer's protocol.

5.3.3 RNA sample preparation.

Cells harvested at 24 hours light and 24 hours dark were added directly to a 10% volume of 95:5 (vol:vol) ethanol (95% purity) and phenol before cooling with liquid N_2 [414]. Chilled cells were immediately harvested by centrifugation at 5,000 × g for 5 minutes at 4°C. Harvested cells were frozen again with liquid N_2 before storage at -80°C. RNA extraction was carried out using the RNeasy Mini Kit (Qiagen, West Sussex, UK) coupled with on-column DNase treatment (Qiagen). A total of 300 ng of RNA was used to synthesize cDNA using the Quantitect Reverse Transcription kit (Qiagen), according to the manufacturer's protocol. Negative controls were carried out to ensure there was no DNA contamination before proceeding to real time PCR analysis.

5.3.4 Quantitative real time-PCR.

Quantitative real-time PCR was performed on an iCycler IQ (Bio-Rad), using SYBR Green Jump Start Taq ReadyMix reagent (Sigma), and primers designed in-house targeting 23 selected genes and a housekeeping gene (23S rRNA) [415]. The primers were designed using DS Gene software version 1.5 (Accelrys Ltd., Cambridge, UK), by employing several rules of thumb in primers design such as ensuring the percentage G+C content was between 50%-55%; the melting temperatures of the primers was set at 55°C-65°C with a temperature difference of 1°C between the forward and reverse primer; and a final primer length of between 18-22 bp. Primers were ordered and synthesized by Operon Technologies (Cologne, Germany). The sequence of these primers is tabulated in Table 5.1.

Table 5.1. Primer sequences for the 24 genes (23 genes representing central metabolism switching points + 1 housekeeping gene) selected for the quantitative real time PCR study, and their corresponding melting temperature (Tm).

Gene	Forward primers	Reverse primers	Tm (°C)
sll0018	TTGCGGATGGCGGAATAGCAAG	TGGCATTCCCGCTTTCAACGTC	63
sll0170	CACTTTATCGCCCAACTCGGTG	ATGTCACTGCTAAAGACCGGGG	63
sll0945	GGGGCTTAATTCCTCCGGTTTC	ACCGGCTATTGCTTTGACCG	60
sll0947	TATTTCCGCAACTGACGGGC	TGATGTTCATCTTTCCGTGGCC	60
sll1194	ACGGTCATCGCCAGAAGTTAGC	TTCGGGATTTTCGTGGTCTGCG	60
sll1260	GGCTTCTAGGGCTTCCAATTCC	ATTGACCTGGTGCAAACCGC	60
sll1342	ACCAAATCCACCACGGAAACG	TGGCATCATCAAAGGCACCATG	60
sll1356	TCGCCGGAATTTCCAACGTAG	CTCAACTGTGACAATCGGGTGG	60
sll1502	ATGACAGGCCCGCATCATAATG	AAACATGCTGGGTTGCCCTG	65
sll1577	CCATGTCACGCAAACAAGCAG	TGTTTCCCAAGCTGATGCTCG	60
sll1626	AATCGTCCACGAGGTCATTGC	GGCCCCCATTCAAAGTCGATTG	65
sll1712	CATTTTGCCTGCGGAGAAAGC	GGTCACGGACAAAGCGAAAGAG	60
sll1746	AACAGCGGCTTTACCACCAG	CGAAGAAGTTCCTGCCGACAAG	60
sll1852	ACAACCAAGGGTTAAGGGTGGC	ATTTGTTCCTCTCCCGTGGTGG	60
sll1994	TCCACTTCTTTCAACGCTTCCC	CAAGCTAATGCTGGGGCAGATG	63
slr0394	GGTGTGGCCCTGTTGGAAAATC	TTCGATGGCTCCTTGGAGGAAC	63
slr0434	TTTGAAGAGGTGAGCATTGCCC	AATAGGGGCACCATTACCTGGG	60
slr1349	CCAAGGGAACATAAAACCCCCC	TTTAAGCAGTTCAGCACCCGG	60
slr1622	TGTTGATCGAAATTCCCGCAGG	TCGCCACCGTCAATCATTTCC	65
slr1834	GCTATGCCTTTGGCGGAGAAAC	TCCCAACCAGATACTTGGCAGG	65
slr1963	TGGTAAGGACGGCAAACGCATC	TGGCACTCTTCCCGGAAAAAGC	63
slr2076	GCCGCTGTGGAAGAAGGTATTG	TCCTTCTCTTTGACTCGCTCGG	60
slr2094	TCGTCCCCGCCACAAAGAATTG	GCATGGCAGCAGCGGAAATTAC	60
23S rRNA	ATGGAGGTCGCAGTGAATAGGC	TGCCCAGATCGTTACGCCTTTC	60

For quantitative purposes in the real time analysis, standard calibration curves were generated from a series of dilutions (dilution factor of 3) of the cDNAs pool for all samples (both control and experimental). All RT-PCR analysis was carried out in triplicate for all samples. The thermal cycling parameters for the real time reactions were set to 95°C for 3 minutes; followed by 50 cycles of 95°C for 30 seconds; Tm of each primer pair (see Table 5.1) for 30 seconds; and 72°C for 30 seconds; before ending with a melting curve analysis with temperatures ranging from 95°C to 55°C.

5.3.5 Isobaric peptide labelling.

100 μg of protein in 20 μL of 500 mM TEAB from each phenotype was reduced with tris-2-carboxyethylphosphine (TCEP) for 1 hr at 60°C; alkylated using methyl methanethiosulfonate (MMTS) for 30 min at room temperature; digested with trypsin at 37°C before being labelled with iTRAQ reagents according to the manufacturer's (Applied Biosystems, Framingham, MA, USA; MDS-Sciex, Concord, Ontario, Canada) protocol with some modifications. These included a two-day tryptic digestion and two volumes of 100% ethanol used during labelling. The full dark sample was tagged with iTRAQ labels 114 and 115, whereas labels 116 and 117 were used for full light sample as illustrated in Table 5.2. After a two-hour incubation period, labelled samples were then combined and dried in a vacuum concentrator.

Table 5.2. Summary of the iTRAQ labelled samples used in this study for the proteome response of *Synechocystis sp.* during light acclimation.

iTRAQ Reagent Label	114	115	116	117
Sample	24-hour dark #	24-hour dark #	24-hour light *	24-hour light *

#, * technical replicates

5.3.6 Strong cation exchange (SCX) fractionation.

Dried peptides re-suspended in 200 μL of Buffer A were fractioned using a PolySULFOETHYL™ A Column (PolyLC, Columbia, MD, USA) 5 μm particle size of 100 mm length × 2.1 mm id, 200 Å pore size, on a BioLC HPLC unit (Dionex, Surrey, UK) with a constant flow rate of 0.2 ml/min and an injection volume of 200 μL. Buffer A consisted of 10 mM KH_2PO_4 and 25% acetonitrile, pH 3.0 and Buffer B consisted of 10 mM KH_2PO_4, 25% acetonitrile and 500 mM KCl, pH 3.0. The 60-minute gradient comprised 100% A for 5 minutes, 5% to 30% B for 40 minutes, 30% to 100% B for 5 minutes, 100% B for 5 minutes and finally 100% A for 5 minutes. The chromatogram was monitored through a UV Detector UVD170U and Chromeleon Software, version 6.50 (Dionex/LC Packings, The Netherlands). Fractions were collected every minute using a Foxy Jr. Fraction Collector (Dionex), and later

were pooled together according to variations in peak intensity. Pooled fractions were dried in a vacuum concentrator, and stored at -20°C prior to mass spectrometric analysis.

5.3.7 Mass spectrometric analysis.

Each dried SCX peptide fraction was re-dissolved in 100 µL of Switchos buffer (0.1% formic acid and 3% acetonitrile), and then 20 µL of sample was injected into the nano-LC-ESI-MS/MS system for analysis. A total of 3 injections were made for each sample. Mass spectrometry was performed using a Q-Star XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF-MS/MS (Applied Biosystems; MDS-Sciex) coupled to an online capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands) as described previously [36]. The peptide mixture was separated on a PepMap C-18 RP capillary column (LC Packings), with a constant flow rate of 0.3 µL/min. The LC gradient started with 3% Buffer B (0.1% formic acid in 97% acetonitrile) and 97% Buffer A (0.1% formic acid in 3% acetonitrile) for 3 minutes, followed by 3% to 25% or 30% Buffer B for either 60 or 90 minutes, then 90% Buffer B for 7 minutes, and finally 3% Buffer B for 8 minutes. The mass spectrometer was set to perform data acquisition in the positive ion mode, with a selected mass range of 300 – 2000 m/z. Peptides with +2 to +4 charge states were selected for tandem mass spectrometry, and the time of summation of MS/MS events was set to 3 seconds. The two most abundantly charged peptides above a 5 count threshold were selected for MS/MS, and dynamically excluded for 60 seconds with ± 50 mmu mass tolerance.

5.3.8 Data analysis.

Protein identification and quantification was carried out using ProQuant software v1.1 (Applied Biosystems; MDS-Sciex). The search was performed against the "mixed genome database" created in-house, including individual genomes of four representative micro-organisms downloaded from NCBI (June 2005): *Escherichia coli* K12, *S. cerevisiae*, *S. solfataricus* and *Synechocystis* sp., encoding 4237, 6298, 2995 and 3264 entries respectively, plus the 222 keratins extracted from humans and mice. The use of this Mixed Genome Database has been demonstrated previously on this dataset to be very efficient in eliminating false positive

identifications down to less than 1% [39]. The search parameters allowed for peptide and MS/MS tolerance were up to 0.15 Da and 0.1 Da respectively; one missed cleavage of trypsin; oxidation of methionine and cysteine modification of MMTS. Only peptides above 70% confidence were saved for identification and quantification. ProGroup Viewer software v1.0.6 was used to identify proteins with at least 95% confidence. The results obtained from ProGroup Viewer were exported to Microsoft Excel for further analysis.

5.4 Results and Discussion

5.4.1 Data quality

In order to ensure the quality of the data, all proteins discussed in this study must have had at least 2 or more MS/MS (peptides) identified with a significant peak area (above 40 counts) and % error (derived from ProQuant software), unless otherwise mentioned. The reference used in this context is tag 114 (full dark); whereas tag 115 is the technical replicate of the full dark sample (as discussed in the previous chapter). Tags 116 and 117 are technical replicates for full light samples. The relative quantitative ratio used is derived from the average of 116:114 and 117:114. The criteria for data validation through multiple injections and replicate analysis were discussed in Chapter 4 and elsewhere [38, 39]. Specifically, up-regulation is deemed to have occurred if the quantitative ratio yields a value of 1.50 or above with an acceptable error margin (SD < log ratio). Similarly, down-regulation occurs when a -1.5-fold expression level decrease is deemed to have occurred. The error factor (EF) is used as an error estimation, and can be expressed as the anti-$\log_{10}$ of the standard deviation. Since the standard deviation is calculated based on the $\log_{10}$ transformation of the peptide ratio not the linear space ratio, it is necessary to convert standard deviations into EF for use in this study. The upper limit of the protein ratio can be calculated by multiplying the protein ratio by the EF; while the lower limit is obtained by dividing the protein ratio by the EF. The list of up- and down-regulated proteins, with their quantitative values, is presented in Table 5.3. The full list of proteins identified in this study is available in Appendix F.

Table 5.3. The list of proteins identified as either up- or down-regulated (≥ ±1.5-fold) with their corresponding relative quantitative ratio and estimated error factor (EF). Hypothetical proteins with homology to another known function organism are in parentheses.

ORF	Protein Name	MS/ MS	Light:Dark (116:114)		Light:Dark (117:114)		Average	
			Ratio	EF	Ratio	EF	Ratio	EF
Amino Acids Metabolism								
sll1981	acetolactate synthase ilvB	2	2.12	1.23	2.26	1.24	**2.19**	**1.24**
sll1931	serine hydroxymethyltransferase glyA	4	2.23	1.17	2.14	1.25	**2.18**	**1.21**
sll0080	N-acetyl-gamma-glutamyl-phosphate reductase argC	6	1.52	1.18	1.67	1.14	**1.60**	**1.16**
Carbohydrate Metabolism								
slr0809	dTDP-glucose 4,6-dehydratase rfbB	2	1.34	1.20	1.87	1.12	**1.61**	**1.16**
Carbon Fixation								
sll1525	phosphoribulokinase prk	28	2.83	1.10	3.05	1.10	**2.94**	**1.10**
slr1839	carbon dioxide concentrating mechanism protein CcmK homolog 4	3	2.14	1.58	2.06	1.62	**2.10**	**1.60**
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	15	1.97	1.21	2.05	1.21	**2.01**	**1.21**
Cellular Processes								
slr1992	glutathione peroxidase-like NADPH peroxidase gpx2	2	1.53	1.22	2.22	1.03	**1.87**	**1.12**
slr1198	antioxidant protein	50	1.72	1.14	1.82	1.13	**1.77**	**1.14**
slr0374	cell division cycle protein	2	2.74	1.68	2.96	1.42	**2.85**	**1.55**
sll1712	DNA binding protein HU	28	2.26	1.15	2.38	1.13	**2.32**	**1.14**
sll1463	cell division protein FtsH	2	-2.69	1.05	-2.49	1.11	**-2.59**	**1.08**
slr1890	Bacterioferritin BrfB	12	1.97	1.08	2.12	1.09	**2.05**	**1.08**
Energy Metabolism								
slr1329	ATP synthase beta subunit atpB	8	1.47	1.25	1.58	1.27	**1.52**	**1.26**
slr1622	soluble inorganic pyrophosphatase ppa	8	1.80	1.22	2.00	1.22	**1.90**	**1.22**
Folding, Sorting and Degradation								
slr0164	ATP-dependent Clp protease proteolytic subunit clpP4	8	2.32	1.23	2.54	1.16	**2.43**	**1.20**
slr2076	60kD chaperonin groEL1	23	1.97	1.13	2.04	1.13	**2.00**	**1.13**
sll0416	60 kDa chaperonin 2 GroEL2	18	1.78	1.20	2.03	1.22	**1.91**	**1.21**
sll0533	trigger factor	4	1.74	1.27	1.90	1.28	**1.82**	**1.27**
slr2075	10kD chaperonin groES	19	1.56	1.10	1.64	1.10	**1.60**	**1.10**
Glycolysis/gluconeogenesis								
slr0752	enolase	3	4.05	1.85	5.32	1.62	**4.69**	**1.73**
sll0018	fructose-bisphosphate aldolase, class II fbaA	24	2.04	1.20	2.17	1.20	**2.11**	**1.20**
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	34	1.89	1.14	2.09	1.13	**1.99**	**1.14**

ORF	Protein Name	MS/ MS	Light:Dark (116:114)		Light:Dark (117:114)		Average	
			Ratio	EF	Ratio	EF	Ratio	EF
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	32	1.60	1.09	1.68	1.09	**1.64**	**1.09**
Hypothetical Proteins								
slr0001	(flavin reductase-like)	4	3.34	1.36	3.82	1.42	**3.58**	**1.39**
sll0359	(transcriptional regulator)	14	3.09	1.20	3.32	1.20	**3.21**	**1.20**
slr0172	(IMP dehydrogenase)	5	2.34	1.23	2.52	1.32	**2.43**	**1.28**
slr0670	(universal stress protein family)	2	2.07	1.28	2.47	1.25	**2.27**	**1.26**
slr0058	(Polyhydroxyalcanoate granule associated protein)	3	2.24	1.02	2.02	1.04	**2.13**	**1.03**
slr0923	(putative 30S ribosomal, Ycf65-like protein)	8	1.78	1.45	2.05	1.41	**1.92**	**1.43**
slr2025	hypothetical protein	2	1.73	1.38	1.90	1.31	**1.82**	**1.35**
slr0455	(Polyhydroxyalcanoate granule associated protein)	3	1.74	1.04	1.85	1.11	**1.80**	**1.08**
ssr1528	hypothetical protein	5	1.57	1.23	1.84	1.17	**1.70**	**1.20**
slr0049	(short chain dehydrogenase)	8	1.55	1.12	1.68	1.12	**1.62**	**1.12**
slr1273	hypothetical protein	2	1.39	1.05	1.75	1.03	**1.57**	**1.04**
sll0272	hypothetical protein	8	1.47	1.04	1.60	1.03	**1.54**	**1.04**
slr1852	hypothetical protein	13	1.54	1.31	1.60	1.33	**1.57**	**1.32**
slr0821	(thiamine Sulfur transferase)	3	1.56	1.14	1.47	1.08	**1.51**	**1.11**
sll0529	(transketolase)	6	1.49	1.31	1.53	1.37	**1.51**	**1.34**
Nucleotide Metabolism								
sll1035	uracil phosphoribosyltransferase	2	2.37	1.11	2.73	1.15	**2.55**	**1.13**
sll1815	adenylate kinase adk	3	1.80	1.25	1.87	1.22	**1.83**	**1.23**
sll1852	nucleoside diphosphate kinase	6	1.53	1.07	1.58	1.05	**1.55**	**1.06**
Other								
sll1784	periplasmic protein	2	2.84	1.04	2.73	1.03	**2.78**	**1.03**
ssl2296	pterin-4a-carbinolamine dehydratase	4	2.02	1.21	2.12	1.18	**2.07**	**1.19**
sll1305	probable hydrolase	2	1.87	1.41	1.86	1.55	**1.86**	**1.48**
slr1719	DrgA protein homolog	4	1.44	1.15	1.56	1.10	**1.50**	**1.12**
Pentose Phosphate Pathway								
slr1734	glucose 6-phosphate dehydrogenase assembly protein opcA	3	1.68	1.24	2.07	1.34	**1.88**	**1.29**
slr0194	ribose 5-phosphate isomerase rpiA	4	1.59	1.08	1.66	1.07	**1.62**	**1.08**
sll1070	transketolase	40	1.47	1.17	1.56	1.17	**1.52**	**1.17**
Photosynthesis								
ssr3383	phycobilisome small core linker polypeptide apcC	5	4.95	1.05	5.35	1.05	**5.15**	**1.05**
sll0199	plastocyanin petE	20	2.85	1.16	3.25	1.20	**3.05**	**1.18**
sll1577	phycocyanin beta subunit cpcB	294	2.80	1.06	2.90	1.06	**2.85**	**1.06**
sll1578	phycocyanin alpha subunit cpcA	389	2.73	1.05	2.82	1.05	**2.78**	**1.05**
sll1398	photosystem II reaction center 13 kDa protein psb28, psbW, psb13, ycf79	3	2.63	2.57	2.73	2.59	**2.68**	**2.58**
slr2067	allophycocyanin alpha subunit apcA	65	2.47	1.18	2.46	1.19	**2.47**	**1.19**
ssl3093	phycobilisome small rod linker polypeptide cpcD	11	2.32	1.13	2.40	1.22	**2.36**	**1.17**
sll1663	phycocyanin alpha phycocyanobilin lyase related protein	2	1.98	1.55	2.17	1.28	**2.08**	**1.42**

ORF	Protein Name	MS/ MS	Light:Dark (116:114)		Light:Dark (117:114)		Average	
			Ratio	EF	Ratio	EF	Ratio	EF
sll0258	cytochrome c550 psbV	20	2.14	1.29	1.97	1.31	**2.06**	**1.30**
slr1986	allophycocyanin beta subunit apcB	170	1.78	1.08	1.89	1.07	**1.83**	**1.07**
sll1194	photosystem II 12 kDa extrinsic protein psbU	19	1.69	1.06	1.89	1.07	**1.79**	**1.06**
sll0928	allophycocyanin-B apcD	42	1.69	1.09	1.86	1.10	**1.78**	**1.10**
sll1579	phycobilisome rod linker polypeptide cpcC2	38	1.68	1.11	1.76	1.11	**1.72**	**1.11**
slr0335	phycobilisome core-membrane linker polypeptide apcE	82	1.62	1.09	1.70	1.09	**1.66**	**1.09**
slr0737	photosystem I subunit II psaD	11	-2.22	1.13	-2.22	1.15	**-2.22**	**1.14**
slr1834	P700 apoprotein subunit Ia psaA	12	-4.33	1.12	-4.36	1.08	**-4.35**	**1.10**
Porphyrin and Chlorophyll Metabolism								
slr0506	light-dependent NADPH-protochlorophyllide oxidoreductase por	4	2.52	1.31	2.99	1.37	**2.76**	**1.34**
Pyruvate Metabolism								
sll1841	pyruvate dehydrogenase dihydrolipoamide acetyltransferase component (E2)	4	3.16	1.18	3.28	1.20	**3.22**	**1.19**
slr1934	pyruvate dehydrogenase E1 component, alpha subunit	7	1.69	1.19	1.87	1.21	**1.78**	**1.20**
sll1721	pyruvate dehydrogenase E1 component, beta subunit	9	1.45	1.27	1.54	1.28	**1.50**	**1.27**
Translation								
sll1802	50S ribosomal protein L2 rpl2	2	2.70	1.23	2.98	1.21	**2.84**	**1.22**
sll1744	50S ribosomal protein L1 rpl1	11	2.19	1.14	2.32	1.15	**2.26**	**1.14**
sll1261	elongation factor TS tsf	7	2.04	1.22	2.09	1.26	**2.06**	**1.24**
sll1743	50S ribosomal protein L11 rpl11	4	1.90	1.08	2.05	1.07	**1.98**	**1.08**
sll1804	30S ribosomal protein S3 rps3	3	1.53	1.05	1.62	1.18	**1.58**	**1.12**
sll1099	elongation factor Tu tufA	60	1.47	1.11	1.64	1.12	**1.56**	**1.12**

To compare translational and transcriptional responses of *Synechocystis* sp. between light and dark, a set of mRNAs were measured by quantitative real-time PCR. In total, 23 genes encoding products representing central metabolism and other major metabolic modules were screened in both light and dark conditions and normalised against the 23S rRNA gene as a housekeeping gene [415]. The relative expression ratios values from light versus dark conditions were calculated with the estimated error in error factor (EF) derived from the triplicate analyses. Since the estimated error via quantitative RT-PCR was also based in $\log_{10}$ space, a similar conversion to EF was carried out. The determination of relative gene expression upper and lower limits was calculated in the same manner as for the protein ratios. A ±1.5-fold change cut-

off point with respect to gene expression was also applied. The results of quantitative RT-PCR are shown in Table 5.4.

Table 5.4. Relative gene and protein expression of the full light phenotype versus the dark phenotype (Light:Dark) for 23 selected genes/proteins. The estimated error for gene and protein expression is determined as an error factor (EF). Gene expression measurements were carried out in triplicate, and selected proteins had at least 2 observed MS/MS spectra in duplicate technical replicates.

ORF	Gene/Protein name	Gene Ratio	EF	Protein Ratio	EF
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	**-1.06**	1.17	**2.11**	1.20
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	**-2.04**	1.15	**1.29**	1.14
sll0945	glycogen synthase glgA	**-1.82**	1.24	**1.15**	1.07
sll0947	light repressed protein A homolog lrtA	**-4.00**	1.79	**1.47**	1.32
sll1194	photosystem II 12 kDa extrinsic protein psbU	**1.60**	1.16	**1.79**	1.06
sll1260	30S ribosomal protein S2 rps2	**-4.30**	1.22	**1.29**	1.11
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	**-1.02**	1.15	**1.99**	1.14
sll1356	glycogen phosphorylase	**-2.18**	1.37	**-1.20**	1.09
sll1502	NADH-dependent glutamate synthase large subunit gltB	**1.16**	1.30	**1.06**	1.04
sll1577	phycocyanin beta subunit cpcB	**-1.25**	1.06	**2.85**	1.06
sll1626	LexA repressor	**-3.21**	1.27	**-1.32**	1.15
sll1712	DNA binding protein HU	**-1.14**	1.16	**2.32**	1.14
sll1746	50S ribosomal protein L12 rpl12	**-4.06**	1.08	**-1.06**	1.20
sll1852	nucleoside diphosphate kinase	**-1.35**	1.12	**1.55**	1.06
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	**1.20**	1.05	**1.24**	1.11
slr0394	phosphoglycerate kinase pgk	**1.30**	1.14	**1.34**	1.19
slr0434	elongation factor P efp	**-1.81**	1.13	**-1.13**	1.03
slr1349	glucose-6-phosphate isomerase	**1.24**	1.41	**1.27**	1.24
slr1622	soluble inorganic pyrophosphatase ppa	**1.25**	1.25	**1.90**	1.22
slr1834	P700 apoprotein subunit Ia psaA	**-1.59**	1.47	**-4.35**	1.10
slr1963	water-soluble carotenoid protein	**1.11**	1.02	**1.03**	1.11
slr2076	60kD chaperonin groEL1	**-2.95**	1.02	**2.00**	1.13
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	**-1.49**	1.22	**1.64**	1.09

In this study, 199 proteins (≥ 2MS/MS) were confidently identified using the stringent criteria as described above. The majority of these proteins were identified as hypothetical proteins with unknown function (15%). This is followed by photosynthesis-related proteins (11%), translation-related proteins (10%), amino acids metabolism proteins (10%), others (9%) and proteins involved in folding, sorting and degradation (8%). Of the 199 proteins, 72 were significantly up-regulated (≥ +1.5-fold), while only 3 proteins were identified as significantly down-regulated (≥ -1.5-fold). In concordance with the functional distribution of identified proteins, the highest number of proteins found to be up-regulated were hypothetical proteins (21%), followed by photosynthesis-related proteins (19%) and translation proteins (8%) (refer to Figure 5.1). During light acclimation, photosynthesis proteins were understandably affected, since photosynthesis process is light-dependent. The increase in relative protein abundance of translation-related proteins may indicate a bias toward regulating the translation process of light-dependent processes. There were also a large number of unknown hypothetical proteins (15 proteins) up-regulated upon illumination.

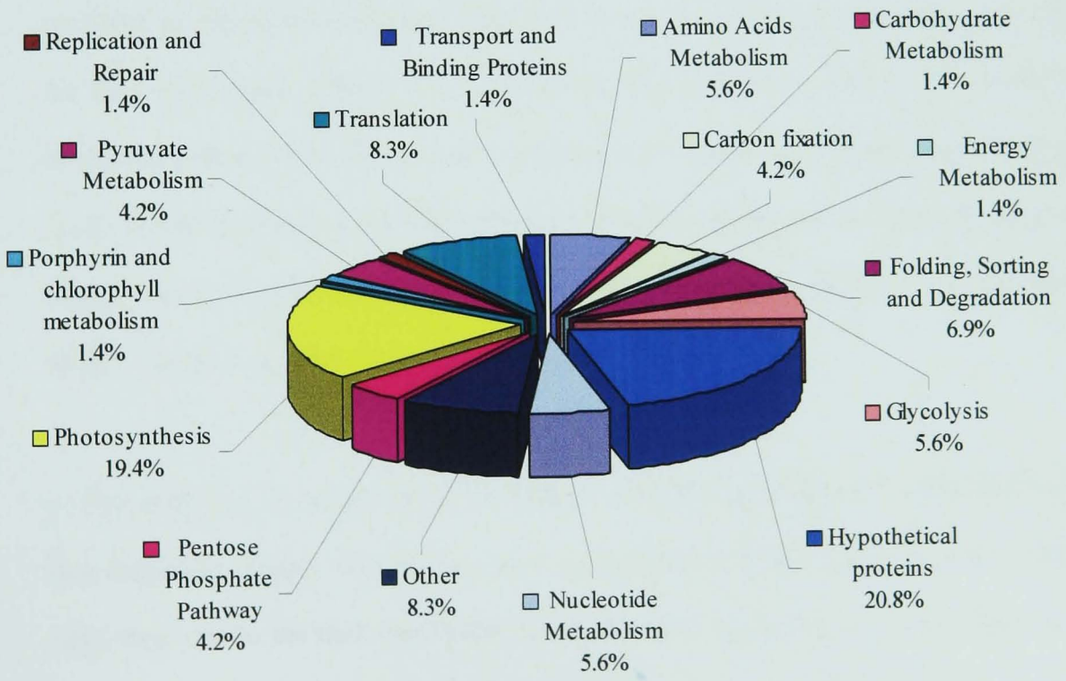


Figure 5.1. The functional distribution of the 72 significantly identified up-regulated proteins (≥ +1.5-fold). Only 3 proteins were significantly down-regulated (≥ -1.5-fold). They are from photosynthesis (2 proteins), and replication and repair (1 protein) processes.

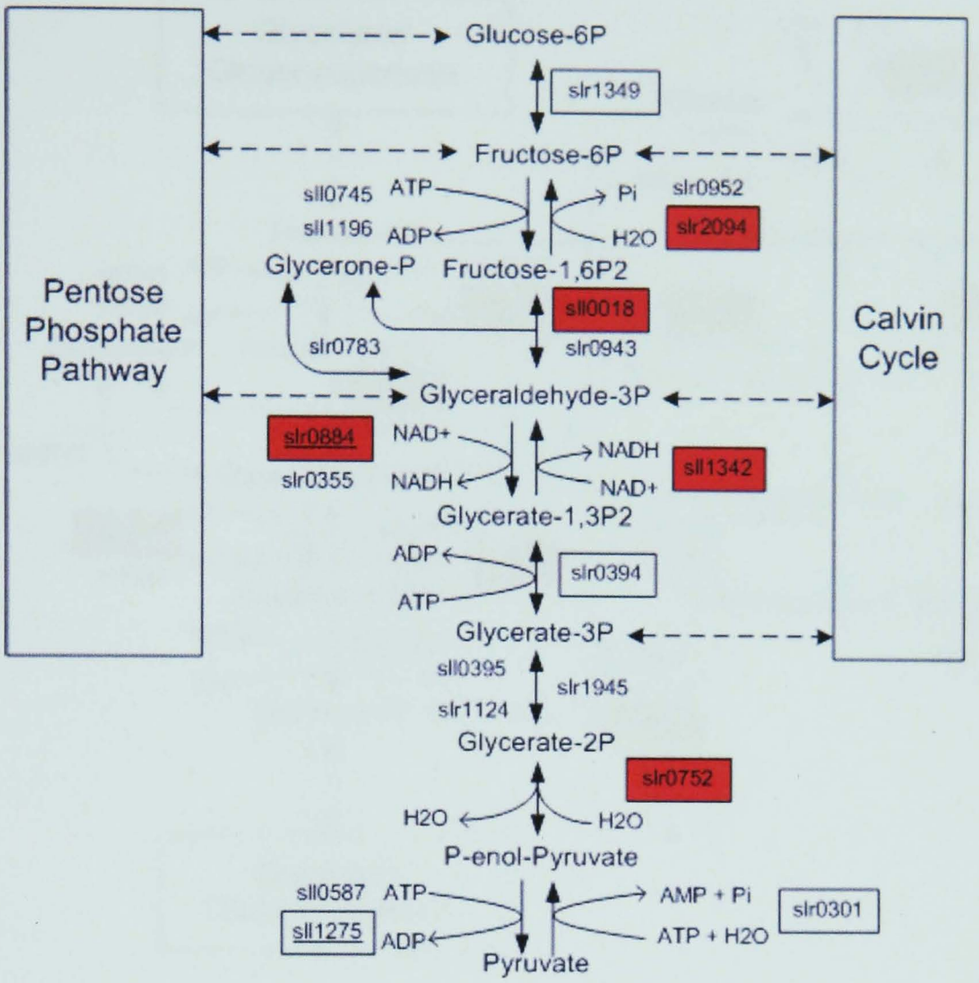
5.4.2 Glucose generation through the Calvin cycle.

As detailed in Chapter 2 (literature review), Calvin cycle proceeds in 3 stages. First stage is the fixation of CO_2 by carboxylation of ribulose-1,5-bisphosphate catalyzed by RuBisCO. Second stage is the reduction of the fixed carbon to begin the hexose synthesis; this is similar to the reactions of gluconeogenesis. Final third stage is the regeneration of the starting compound ribulose-5-phosphate; this is similar to some of the reactions of pentose phosphate pathway, although in the reverse way.

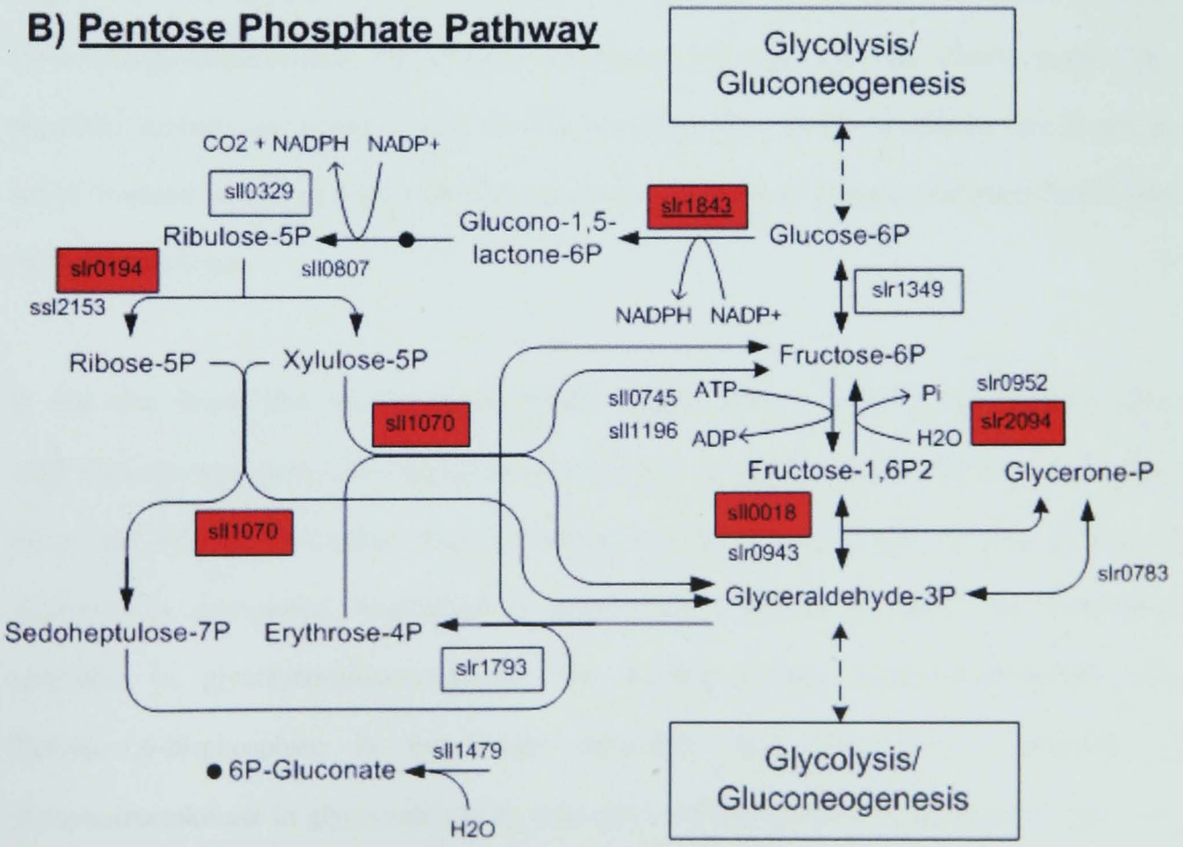
As a whole, an increase in protein expression was observed in enzymes involved in the Calvin cycle during the light phase compared to the dark phase (Figure 5.2C). From the proteomics results, the most important enzyme (in stage 1) responsible for catalysing the fixation of carbon dioxide, ribulose bisphosphate carboxylase (RuBisCO), was up-regulated 2-fold in the light compared to the dark, although some inconsistency was noticed between the expression level of the small subunit (slr0012) (at +2.0-fold) and large subunit (slr0009) (at +1.4-fold) (Table 5.3). It has been reported that the regulation of Calvin cycle activity during exposure to light is mediated by phosphoribulokinase (Prk, sll1525) [416]. Prk was up-regulated near 3-fold over the dark cycle here (Table 5.3). The incremental increase in RuBisCO activity is always associated with the formation of carboxysome by the carbon dioxide concentrating protein, Ccm [417]. CcmK homolog 4 (slr1839) was up-regulated over 2-fold in the light compared to the dark (Table 5.3). However, the relative abundance of others identified Ccm subunits (CcmM, sll1031; CcmK2, sll1028) were unchanged.

In line with the up-regulation of RuBisCO, NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase (Gap2, sll1342) was also up-regulated 2-fold in relative protein abundance in light compared to the dark (see Table 5.4 and Figure 5.2). However no significant change was observed in the relative abundance of the mRNA level of *gap2*. From the literature, Gap2 has a high tendency towards photosynthetic gluconeogenesis, while Gap1 (slr0884, glyceraldehyde 3-phosphate dehydrogenase) is favoured in glucose breakdown during non-photosynthetic glycolysis [418].

A) Glycolysis/Gluconeogenesis



B) Pentose Phosphate Pathway



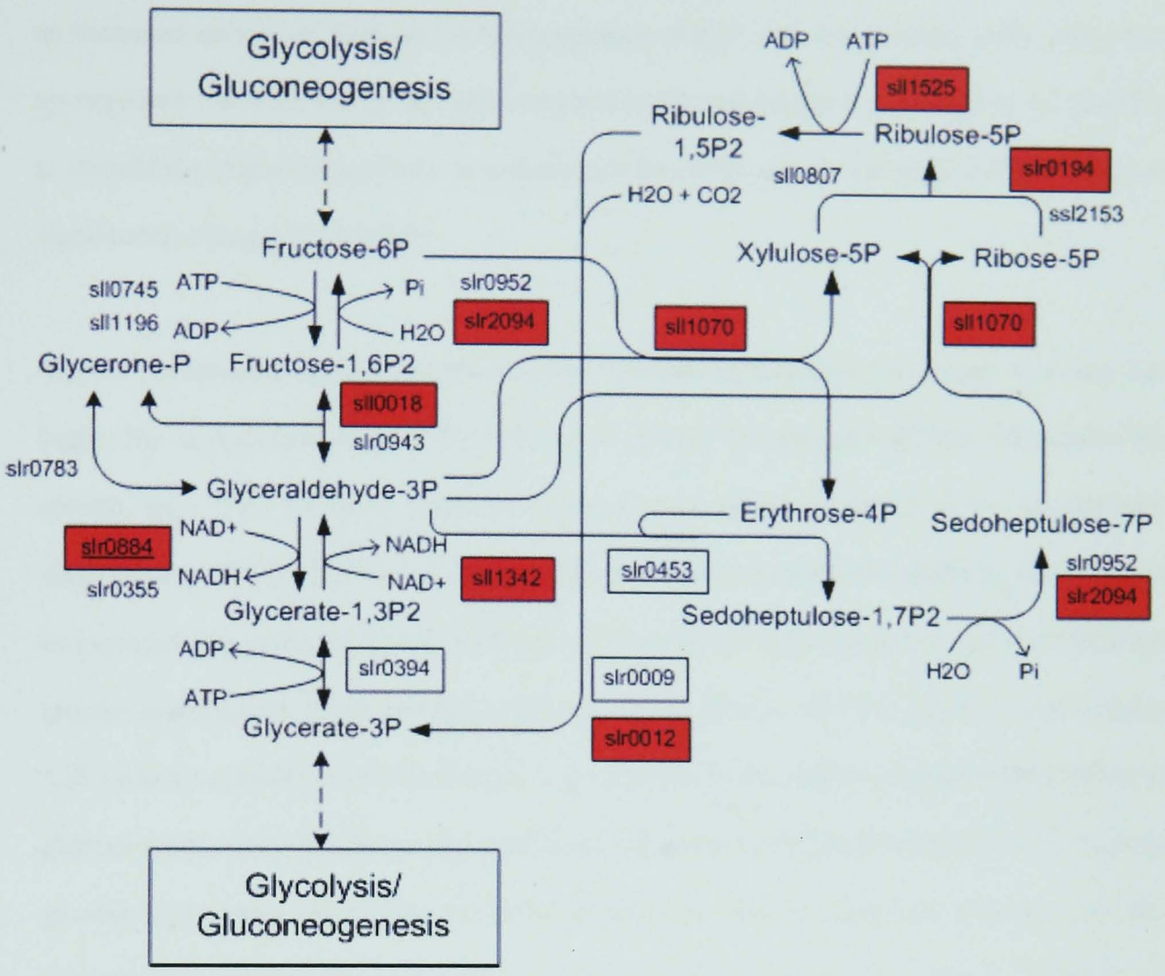


Figure 5.2. Central carbon metabolism of *Synechocystis sp.* during the light, specifically (A) Glycolysis/gluconeogenesis; (B) Pentose phosphate pathway; and (C) Calvin cycle. Up-regulated proteins are boxed in red. Proteins unchanged in relative abundance are boxed in white. Proteins identified with 1 MS/MS spectra are underlined. Proteins not identified in this study are not boxed.

It was also found that fructose-bisphosphate aldolase (FbaA, sll0018) and enolase (Eno, slr0752) were more than 2-fold up-regulated in the light compared to dark. However, since both FbaA and Eno are reversible enzymes acting both in glycolysis and gluconeogenesis, a regulation is unexpected. Regulation of metabolism is mostly performed by irreversible enzymes. In glycolysis/gluconeogenesis the reaction between fructose-6-phosphate and fructose-1,6-bisphosphate is the major regulatory step [58]. It is catalyzed by phosphofructokinase in glycolysis and by fructose-1,6-bisphosphatase in gluconeogenesis. The

activity of these two enzymes therefore sets the metabolic flow in either direction. And indeed an increased amount of fructose-1,6-bisphosphatase (Fbp1) was found in this study. Fbp1 was up-regulated 1.64-fold during the light compared to the dark (Table 5.3 and Figure 5.2), leading to stimulation of gluconeogenesis. In contrast, the transcript relative abundance of *fbp*1 was not significantly changed (Table 5.4).

Pentose phosphate pathway converts by many intervening reactions 5-C sugars to 6- and 3-C sugar. The regeneration stage of the Calvin cycle (Stage 3) uses some of these reactions in the reverse way. Two of these enzymes in the Calvin cycle are found to be up-regulated: transketolase (TktA, sll1070) and ribose 5-phosphate isomerase (RpiA, slr0917), which should be expected. Transaldolase (TalB, slr1793) was however not up-regulated. It was also found that glucose 6-phosphate dehydrogenase assembly protein (OpcA, slr1734) (Table 5.3 and Figure 5.2C) was up-regulated. OpcA is thought to be involved in the oligomerization and activation of glucose-6-phosphate dehydrogenase proteins in the pentose phosphate pathway [419]. Although glucose-6-phosphate dehydrogenase (Zwf, slr1843), in pentose phosphate pathway, was also found to be tentatively up-regulated 2.75-fold, this must be treated with caution since Zwf was identified and quantified by a single MS/MS experiment. The first part of the oxidative pentose phosphate pathway is controlled via the ratio of NADP+ to NADPH [153]. No significant protein relative expression change was detected for 6-phosphogluconate dehydrogenase (Gnd, sll0329) between the light and dark phases (Table 5.3 and Figure 5.2B), suggesting that the regeneration of NADPH remained constant.

5.4.3 Pyruvate channel to Acetyl-CoA.

Relatively higher expression (1.5- to 3.2-fold) of pyruvate dehydrogenase enzymes (sll1841, slr1934, sll1721) was evident during the light compared to the dark (Table 5.3). Among the identified hypothetical proteins, slr0058 and slr0455 showed similarity to the Polyhydroxyalcanoate (PHA) granule associated protein, which was also found to be significantly up-regulated (Table 5.3). PHA is a thermoplastic produced in nature, and has commercial interest [107, 108]. In cyanobacteria, PHA is used as a storage compound for

carbon and energy in addition to glycogen [109]. In the light, acetyl-CoA is the substrate for polymerization to make PHB (one type of PHA compound). Therefore, it was unsurprisingly to find pyruvate dehydrogenase and these hypothetical PHA proteins to be significantly up-regulated for this purpose. No significant up-regulation was detected for any TCA cycle components (see Table 5.3).

In addition to pyruvate dehydrogenase, up-regulation in the light compared to the dark was also detected for acetolactate synthase (IlvB, sll1981) as depicted in Table 5.3. This enzyme is responsible for the first step in the conversion of valine to generate 2-acetolactate from pyruvate, and to also account for the production of 2-aceto-2-hydroxybutanoate in isoleucine with 2-oxobutanoate as a co-product. A total of 19 proteins involved in amino acid biosynthesis were confidently identified with 2 or more MS/MS spectra. Of these, 3 were significantly up-regulated more than 1.5 fold (Table 5.3). Furthermore, serine hydroxymethyltransferase (GlyA, sll1931) was up-regulated 2.18-fold in light compared to the dark. GlyA shares a similar precursor as glycine (which is glycerate-3P), and is responsible for the inter-conversion of serine and glycine to either 5,10-methylene-THF or tetrahydrofolate as the co-substrate/by-product. Another up-regulated protein (> 1.5-fold) was N-acetyl-gamma-glutamyl-phosphate reductase (ArgC, sll0080), which is involved in phosphorylation via the urea cycle by converting N-acetylglutamyl-phosphate and NADPH to N-acetylglutamate semialdehyde and $NADP^+$.

To investigate the relationship between mRNA and protein relative abundances in amino acid metabolism, the NADH-dependent glutamate synthase large subunit (GltB, sll1502) was chosen (since it was not differentially regulated at the proteomic level) for analysis by real time RT-PCR. Comparing the this data with the iTRAQ data resulted in a good correlation between the protein and transcript abundances, with no significant relative expression differences noted between light and dark conditions (Table 5.4). It is worth detailing that all enzymes discussed here, although involved in amino acids biosynthesis, are also responsible for amino acids degradation when a carbon source is not readily available.

5.4.4 Photosynthesis-related proteins.

Photosynthesis related proteins not surprisingly constitute one of the largest functional groups identified, composing 11% of the 199 proteins quantified (≥2 MS/MS). Overall, 14 showed expression increases, whilst 2 proteins in this class were down-regulated (Table 5.3). These two contrasting (down-regulated) proteins, PsaD (slr0737) and PsaA (slr1834), both belong to the Photosystem I group. Two other proteins (albeit identified with only one MS/MS spectrum), PsaB and PsaE (both from Photosystem I), were also found to be marginally down-regulated by 1.5-fold. To analyse this apparent non-conforming behaviour further, the transcriptional product of *psa*A was also analysed via quantitative RT-PCR, yielding results consistent with those observed with iTRAQ at 1.59-fold down-regulation in the light (Table 5.4). The transcript abundance for the *psa*AB operon has been found previously to decrease when cells were grown in complete darkness [420]. Here, however, the Photosystem I reaction centre did not show the same degree of involvement as the Photosystem II reaction centre during extended incubation in light (24hr) compared to dark. This suggests the association of Photosystem I during extended photosynthesis (long incubation in the light) is not modulated in the same way as in Photosystem II. In 2003, Muratmasu and Hihara reported that the Photosystem I response to a high light environment differed at the transcript level as well as the promoter activities when compared with Photosystem II, and this seems important for *Synechocystis* sp. to regulate its Photosystem content in response to changes in photon flux density [156]. Another possible explanation is that proteins accumulate throughout the dark cycle, which may impact on the synthesis and assembly of the Photosystem proteins in the dark [420]. However this was not observed in Photosystem II proteins.

Pigment proteins (see Figure 5.3) such as phycocyanin (CpcA, CpcB, CpcC1, CpcC2, CpcD and CpcG1) and allophycocyanin (ApcA–ApcE), which serve as antenna for light absorption in Photosystem II (besides chlorophyll), were highly relatively expressed (1.6 to 5.0-fold) during the light compared to the dark (Table 5.3). However, quantitative RT-PCR on *cpc*B, in contrast, revealed no significant change in transcript relative abundance in the light versus the dark

(Table 5.4). This is followed by increased expression of the Photosystem II proteins: PsbW (sll1398), PsbU (sll1194) and PsbV (sll0258) in this study.

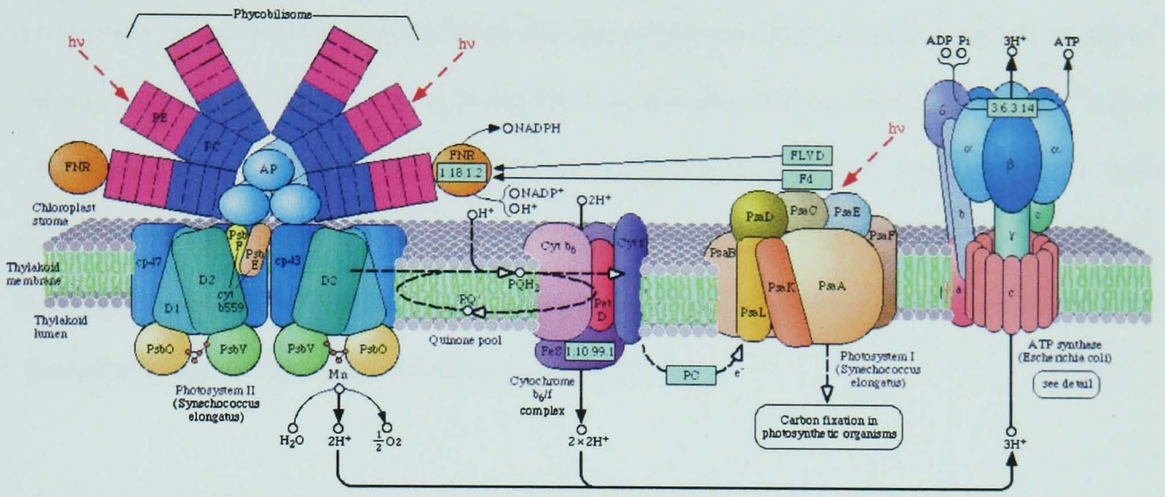


Figure 5.3. The photosynthetic apparatus comprising Photosystem I and II and pigment proteins such as phycocyanin and allophycocyanin is essential in a photosynthetic organism (reproduced from KEGG, http://www.genome.jp/kegg/kegg2.html).

Quantification of *psb*U gene expression, determined via quantitative RT-PCR, also shows an up-regulation (1.6-fold) in the light compared to the dark (Table 5.4). Furthermore, plastocyanin (PetE, sll0199), which acts as an electron transfer agent between cytochrome f of the cytochrome b_6f complex and P700+ in Photosystem I [421, 422], was also found to increase in relative expression (>3-fold) in the light (Table 5.3). Additionally, the light repressed protein A homolog (LrtA, sll0947) exhibited a relative protein abundance change of 1.47-fold using iTRAQ (not considered as up-regulation since it is below 1.5-fold), while a corresponding down-regulation of 4.0-fold was detected in transcript abundance (Table 5.4). Putting these apparent inconsistencies (suggesting possible post-translational modification or cross-regulation) to one side, *lrt*A is known to be dark-induced as the transcript accumulated across the dark period of incubation [423].

In porphyrin metabolism, no significant protein relative abundance changes were observed between the intermediates: HemB (sll1994), HemL (sll0017) and HemF (sll1185), in heme

(hemoproteins) production (Table 5.3). Quantitative RT-PCR analysis also revealed no significant changes in *hem*B transcript relative abundance in the light versus dark cycle (Table 5.4). Hemoproteins are involved in various biochemical pathways, in particular in respiratory complexes in oxidative phosphorylation and the cytochrome b_6f complex involved in photo-phosphorylation, which are crucial in the electron transport leading to ATP synthesis. Only 1 identified protein chlorophyll metabolism was up-regulated in the light compared to the dark. This was the light-dependent NADPH-protochlorophyllide oxidoreductase (Por slr0506), which catalyses the oxidation of protochlorophyllide and NADPH to chlorophyllide a and $NADP^+$, the second last step in chlorophyll a synthesis (Table 5.3).

5.4.5 Translation- and transcription- related proteins.

Amongst the 19 translation-related proteins identified, 6 were up-regulated in the light compared to the dark (Table 5.3). These include the elongation factors TS (Tsf, sll1261) and Tu (TufA, sll1099). Meanwhile, the abundance of elongation factor P (Efp, slr0434) was not significantly altered. However, a down-regulation of 1.80-fold in the corresponding transcript relative concentration was detected (Table 5.4). Elongation factor proteins are involved in protein synthesis, by mediating the binding of amino-acyl tRNA into the A site of the ribosome, and are responsible for the association of ribosome and amino-acyl tRNA complexes by inducing the exchange of GDP to GTP [424]. This is evident by the up-regulation (> 1.5-fold) of a variety of ribosomal proteins, specifically the 50S ribosomal protein L2 (Rpl2, sll1802), 50S ribosomal protein L1 (Rpl1, sll1744), 50S ribosomal protein L11 (Rpl11, sll1743) and 30S ribosomal protein S3 (Rps3, sll1804) (Refer to Table 5.3). The large subunit (50S) of ribosomal proteins Rpl1 and Rpl11 were reported to have an important role in the GTPase function of the ribosome [425]. The gene and protein expression changes of the 50S ribosomal protein L12 (Rpl12, sll1746) and 30S ribosomal protein S2 (Rps2, sll1260) were investigated here using both iTRAQ and quantitative RT-PCR techniques. No significant changes in abundance were observed in protein expression for both proteins; however the transcript relative abundance for both genes showed more than 4-fold down-regulation in the light (Table 5.4).

The LexA repressor (sll1626), which differs from the *Escherichia coli*-type SOS like regulon [81], was recently identified as a transcription regulator, binding to the promoter region of the bidirectional hydrogenase operon (*hox*EFYUH) [78, 79]. Here, this protein did not show any significant change in abundance (-1.31-fold). It transcript abundance was, however, found to be 3.2-fold down-regulated (Table 5.4). The incubation period of 24 hours in both light and dark cycle may be too long to measure the transient effects of the light to dark transfer process. A 30 minute sampling interval was reported to accurately capture the transcriptional dynamics throughout the light/dark transition in *Synechocystis* sp. [66], yet none has been reported at the translational level.

5.4.6 Cellular processes and stress-response proteins.

The long cell adaptation period of 24 hours in both light and dark phase should not generate any differences in protein expression level between heat-shock proteins, since this group of proteins is generally characterized as immediate, intense, and transient activation proteins (as a result of a change in the environmental condition) [175]. Relative abundances of heat-shock protein GrpE (sll0057) and heat-shock protein 70 (DnaK2, sll0170) did not change significantly between 24 hours light and 24 hours dark (Table 5.3). Instead, increased relative expressions were recorded for chaperonin proteins GroEL1, GroEL2, GroES (slr2076, sll0416, slr2075), a trigger factor (sll0533) and an ATP-dependent Clp protease proteolytic subunit (ClpP4, slr0164) (Table 5.3). Mary *et al.* also reported that *gro*EL, *gro*ES, *dna*K2, *dna*J3, *clp*B1 and *clp*P1 were associated with high-light adaptation, and most heat-shock genes were found to be transcribed during the first 3 hours of incubation in high light, followed by a gradual decrease [171]. Clearly 24 hours incubation in the light is not ideal to capture the expression level of these proteins. This is further verified through quantitative RT-PCR analysis, where *dna*K2 gene expression was 2-fold down-regulated, in contrast with both the literature (up-regulated) [171] and the iTRAQ data (no change). Interestingly, quantitative RT-PCR analysis also revealed a down-regulation in *gro*EL1 gene expression, in contrast to the up-regulation observed in protein abundance (Table 5.4). It seems that the half-life of the protein is much longer than the corresponding messenger RNA. On the other hand, it has been reported that light accelerated the

heat induction of *gro*ESL1 and *gro*EL2 in *Synechocystis sp.*, and enhanced the accumulation of GroEL proteins [175]. The control mechanism of heat-shock proteins needs to be robust and specific; therefore the change in transcript level should be instantaneous to counter the change in environment, after the accumulation of the enzyme reaches a plateau, the gene expression levels are reduced as they have completed their task.

The main function of superoxide dismutase (SodB, slr1516) is to protect the cell from the toxic superoxide anion radical O_2^-, by catalysing the formation of H_2O_2 and O_2. It has been reported that SodB is light-dependent, whereby mRNA and enzyme activity are elevated under continuous light conditions compared to continuous dark [172]. In contrast to the literature, the SodB protein relative expression level measured in this study was constant under both light and dark conditions (Table 5.3). Within this study, glutathione peroxidase-like NADPH peroxidise (Gpx2, slr1992), which is involved in reduction of lipid hydroperoxides to their corresponding alcohols and in converting free H_2O_2 to water [426], was found to be up-regulated in the light compared to the dark (Table 5.3). Another protein involved in cellular stress responses, known as an antioxidant protein (slr1198), was up-regulated 1.77-fold in the light compared to the dark. This protein had been previously reported to be a 1-Cys type peroxiredoxin, and the disruption of this gene affected the growth rate compared to the wild-type [174]. Interestingly, one of the hypothetical proteins (slr0670) with similarity to the universal stress protein family was also up-regulated by 2.27-fold (Table 5.3).

On the other hand, the DNA binding protein HU (sll1712) and cell division cycle protein (slr0374) both increased in abundance; whereas the cell division protein (FtsH, sll1463) was down-regulated in the light compared to the dark (Table 5.3). These proteins are involved in cell replication and repair mechanisms. Yet, the 24 hours of continuous light incubation here did not effect the transcript relative abundance of sll1712 (Table 5.4). It has been shown that sll1712 is capable of stabilizing DNA during transcription, thus preventing denaturation as a result of light oxidation [427].

5.4.7 Hypothetical and other proteins.

Of the 30 hypothetical proteins were identified (refer to Table 5.3), 15 were up-regulated by in the light compared to the dark, whilst the relative abundance of the remaining 15 remained unchanged. Thus, none of the identified hypothetical proteins were down-regulated. There is little that can be added in detail regarding the increase in protein expression level for these 15 hypothetical proteins other than their bias toward the light cycle, however it does shed light on the possibility that these proteins are potentially associated with photosynthesis related or light-dependent processes. By searching the protein sequences against the Cyanobacteria Gene Annotation Database (CYROF) (http://cyano.genome.jp/), 10 of the hypothetical proteins showed relatively high similarity (ranging from 40% to 80%) to proteins of known function in other organisms, suggesting a possible functional role in *Synechocystis* sp., as shown in Table 5.3. For example, one of the hypothetical proteins (sll0359) with similarity to a transcriptional regulator AbrB (80% homology to *Crocosphaera watsonii* WH 8501) was 3.21-fold up-regulated in the light (Table 5.3).

Among the 2 bacterioferritins (Brfs) identified here, BrfB (slr1890) was up-regulated in the light, whereas the abundance of BrfA (sll1341) remained unchanged (Table 5.3). These proteins are thought to be responsible for up to 50% of the cellular iron storage, and they are involved in iron homeostasis [247]. Other interesting proteins which were up-regulated in the light compared to the dark are (Table 5.3): DrgA protein homolog (slr1719) and pterin-4a-carbinolamine dehydratase (ssl2296). DrgA is thought to be involved in detoxification of dinoseb via the reduction of the nitro group [428], and possibly participates in the regulation of cytoplasmic NADPH oxidation [429, 430]. *Synechocystis* sp. is known for the capability to produce a pteridine glycoside cyanopterin [431, 432], however it is not proven whether ssl2296 is involved in this synthesis, as metabolites produced were not investigated here. The water-soluble carotenoid protein (slr1963) was subjected to both iTRAQ and quantitative RT-PCR analysis. Both techniques were in agreement, whereby no significant changes in relative abundance were detected in the light compared to the dark (Table 5.4).

5.4.8 Comparing iTRAQ and 2-DE for quantitative proteomics.

So far, most of the proteomics studies employing *Synechocystis* sp. PCC 6803 as the model organism have focused on using 2-DE techniques [67, 180, 209, 276, 277]. Staining intensity with presence and absence of spots on the gels is normally used as the main criterion in determining whether a protein is either up- or down-regulated, which in principle could be a source of variation. The improvement of gel-based techniques using CyDyes differential gel electrophoresis (DIGE) shows promising results compared to conventional densitometry-based 2-DE because it enables the comparison of multiple samples within the same gel, thus reducing the gel-to-gel variation [324, 325]. iTRAQ results derived from this experiment are compared with similar proteomic studies carried out based on the investigation of autotrophic and heterotrophic growth in *Synechocystis* sp., in particular that performed by Kurian *et al.* [67]. The comparison between iTRAQ and 2-DE results is discordant as depicted in Table 5.5.

Table 5.5. A comparison between the iTRAQ results derived in this study and 2-DE results adapted from the literature [67]. * indicates proteins identified by a single peptide.

		Light:Dark	
ORF	**Protein Name**	**2-DE**	**iTRAQ**
sll1525	phosphoribulokinase prk	1.00	2.94
slr0394	phosphoglycerate kinase pgk	1.00	1.34
slr0752	enolase	1.00	4.69
sll0329	6-phosphogluconate dehydrogenase	1.00	1.28
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	1.00	2.11
slr1793	transaldolase	-3.30	1.13
slr0652	phosphorybosilformimino-5-amino- phosphorybosil-4-imidazolecarboxamideisomerase hisA	-5.20	1.47
slr1848	histidinol dehydrogenase hisD	-2.70	**1.51***
sll1626	LexA repressor	-2.70	-1.32
slr0884	glyceraldehyde 3-phosphate dehydrogenase 1 (NAD+) gap1	-2.50	**1.98***
sll1099	elongation factor Tu tufA	-2.50	1.56
slr1963	water-soluble carotenoid protein	-2.30	1.03
sll1908	D-3-phosphoglycerate dehydrogenase serA	-2.20	1.34
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	-2.20	1.99
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	-2.10	1.64
sll1070	transketolase	-1.80	1.52

ORF	Protein Name	Light:Dark	
		2-DE	iTRAQ
sll1028	carbon dioxide concentrating mechanism protein CcmK2	2.40	-1.22
sll1363	ketol-acid reductoisomerase ilvC	2.50	1.04
sll1536	molybdopterin biosynthesis protein moeB	3.20	1.37
slr1198	antioxidant protein	3.30	1.77
slr1852	unknown protein	3.30	1.57
sll0550	flavoprotein	3.40	**3.35***
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	9.50	1.42
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	14.20	2.01

The only similar protein changes, comparing qualitatively and having similar/genuine regulation directions (up- or down-) are slr0012, sll0550, slr1852, slr1198, slr0394 and sll0329. It is not entirely clear why the results derived from both studies with the same objective, investigating light and dark behaviours using proteomics tools, can vary so greatly. Of course there will be some data variation across each experiment, and sometimes this type of data can lead to ambiguous results. For example, the carbon dioxide concentrating mechanism protein CcmK (CcmK2, sll1028), in principle, should be up-regulated in the light phase since the Calvin cycle takes place [66] and clearly the 2-DE study agrees, yet the current iTRAQ results disagree. Similarly Gap2 proteins, reported previously, are associated with the light cycle [278], and the iTRAQ results are in agreement, whereas 2-DE is not. Consequently, it is difficult to judge reliability, based on specific protein observations presented here, as can be shown by comparing different techniques used, despite both studies having similar objectives. This may be explained in part by two particular differences that occur in the sample preparation step (the phenotype), especially for the dark condition. Kurian *et al.* (2006) incubated the cells differently for light and dark cultures, with dark cultures supplemented with 5 mM glucose and a 5 min light pulse of 50 μEinstein/m^2s every 24 hours (versus 120 μEinstein/m^2s in this study) [67]. The addition of glucose (5mM) would stimulate the heterotrophic related expression, while the light would cause similar light related expression. If this cycle was carried out over a long period of time, the speculation is that it is possible to see circadian rhythm start to increase in later gels as *Synechocystis* sp. becomes acclimated to the cycle. In the current study, the light and dark samples were collected from the same fermentor with the only alteration being the light

intensity (light versus dark), and lack of exogenous glucose for any phenotype. It is not possible to discern the overall effect this had on the sample, but clearly the difference in sample preparation and exact phenotype can cause significant changes in the proteome.

5.5 Conclusions

This chapter focussed on the fundamentals of cyanobacterial survival, by attempting to understand the proteome responses as a function of sustained light exposure (24 hours light period). The quantitative proteomics approach presented here successfully unveiled an insight into *Synechocystis* sp., measured as relative protein and mRNA changes in full light against full dark conditions. In particular, and perhaps strikingly, the down-regulation of Photosystem I proteins PsaA (slr1834) and PsaD (slr0737) in the light caught the central attention, with this finding suggesting the possibility that Photosystem I is modulated through a different mechanism compared to Photosystem II during photosynthesis (specifically long incubation in light or dark conditions studied here). Photosystem II and pigment proteins such as phycocyanin and allophycocyanin behave reasonably as expected (i.e. up-regulated) in the presence of light. Quantitative PCR analysis yielded similar results, where *psa*A was found to be down-regulated and *psb*U (Photosystem II) up-regulated in the light cycle compared to the dark cycle.

Gluconeogenesis and the Calvin cycle remain as the core central carbon metabolic pathways during the light phase, as deduced from the overall up-regulation of proteins involved in these processes. These processes are photosynthesis-mediated with light as the main activation source. Some stress-response proteins were also up-regulated during extended exposure to light. Unsurprisingly the chaperonin proteins such as GroEL1, GroEL2, GroES (slr2076, sll0416, slr2075) and Clp protease proteolytic subunit ClpP4 (slr0164) were up-regulated. The harmful superoxide anion radical (O_2^-) must be removed from the cell to prevent serious oxidative stress. Gpx2 (slr1992) was up-regulated for this purpose; however SodB (slr1516) was not up-regulated in the light, an observation which contrasts to findings in the literature.

Often the relative changes in protein abundance do not correlate well with the equivalent gene expression data. Amongst the 23 genes selected throughout the metabolic system of this organism, only 8 (35%) were differentially expressed in the same direction as the corresponding protein, with 14 (60.9%) of them disagreeing, and 1 (4.3%) exhibiting an opposing differential expression change. Further discussion on the correlation between transcription and translation will be carried out in the next chapter. Therefore, without going into detail, clearly there is a disparity between protein and gene expression. It is worth noting that the rate of degradation and accumulation in transcript and translation level vary across the whole spectrum of genes/proteins. Furthermore, these rates depend greatly on the experimental conditions (temperature, pH) and substrate availability. It will be a great challenge in the future to solve this explicit yet complicated relationship.

Chapter 6

Circadian Rhythm Study of *Synechocystis sp.* PCC 6803 Under a 12-hour Dark and Light Cycle

6.1 Abstract

The influence of circadian rhythm in *Synechocystis* sp. PCC 6803 during a recurring 12 hr dark and light cycle was investigated using a quantitative real-time RT-PCR approach coupled with an iTRAQ-mediated proteomic analysis. Out of 26 key genes investigated, all of them showed an oscillatory behaviour in their mRNA abundance, with a periodic half-life of 12 hr, mirroring the cyclic changes in light intensity. Two different types of gene profiles (A- and B-type) were observed based on their unique response toward light. A-type genes were highly sensitive to light activation. In contrast, B-type genes only showed responses toward the dark phase. The effect of circadian rhythm at the protein level was less pronounced compared to the transcript level. Furthermore the transcriptional response was more susceptible to the change in environment compared to the protein expression. Nonetheless, a common feature of a 'shock' response during the first 90 min transition into both light and dark phases was noticeable in over 90% of the total identified proteins. This 'shock' response directly corresponded to the change in the transcript abundance, indicating a close relationship between the transcript and protein during the first 90 min of each (light and dark) phase, and possibly suggesting the translation process of protein biosynthesis is most active during this time range. Comparison between transcriptional and translational behaviours is also discussed.

6.2 Introduction

Cyanobacteria are one of the earliest living organisms on earth, able to carry out carbon dioxide assimilation and produce oxygen, and are thought to be one of the major contributors to the formation of today's biosphere [41]. These bacteria contain plant-like features that allow them to derive energy by carrying out photosynthesis. Light plays a key role in all photosynthetic organisms, and the energy integrated from light is utilised by the cells for various biochemical reactions. The cyanobacteria are also the only bacterial-species that contain a circadian clock [433, 434]. Circadian rhythm is one of the most important features in plants, animals, fungi and cyanobacteria [435, 436]. This clock regulates the heartbeat of animals, controls cell replication and division, initiates gene and protein proliferation, and many other physiological processes [437]. In cyanobacteria, this clock is triggered by light intensity as a result of a change in environment, such as day and night [438]. During the day they photosynthesise, whilst at night, respiration processes take over. It has been reported that the circadian clock in cyanobacteria enhances the adaptive behaviour to the extrinsic factors such as different circadian periods, however this advantage is only valid in the cyclic environment [439].

The association of a circadian clock with various biological processes in *Synechocystis* sp. has been reported widely. For example, significant circadian expressions activated by light were observed in heat shock proteins [440]; folding and sorting mechanism of the *dna*K gene [441] and the Photosystem II *psb*A2 gene [442]. Interestingly, Aoki *et al.* reported that *Synechocystis* sp. exhibited circadian rhythms in complete darkness, and suggested light is not necessarily the only source of the oscillation in gene expression [443]. Similarly microarray analysis has also been applied to examine the global response of circadian gene expression in *Synechocystis* sp. [444].

Synechocystis sp. PCC6803 has also been studied extensively at the proteomic level [36, 188, 209]. However, these reports did not focus on circadian rhythm to the same extent as those studies carried out at the transcriptome level. In general, there are few circadian rhythm studies at the proteomic level in various organisms. Some of the limited examples include work by

Wagner *et al.*, who successfully applied 2-DE to characterize circadian expressed proteins in the green alga *Chlamydomonas reinhardtii* [445]. Similarly, 2-DE proteomic analysis was employed to identify the circadian expressed proteins in the marine dinoflagellate *Lingulodinium polyedrum* [446]. More recently, differential gel electrophoresis using CyDyes (DIGE) was employed to investigate the circadian profile of hepatic proteome in mouse liver [447]. All the studies detailed here used a gel-based proteomics approach, and to date no group has conducted a 'shotgun' analysis of this phenomenon.

Since deeper insight can be obtained by measuring an organism's response at the level of messenger RNA (mRNA) and protein, it is hoped, in this chapter, to characterise circadian rhythm behaviour in *Synechocystis* sp. PCC6803 at both these levels so as to begin to obtain a 'Systems Microbiology' understanding. In this study, both quantitative real-time RT-PCR and an iTRAQ-mediated proteomic analysis were applied on samples obtained temporally from *Synechocystis* sp. PCC6803 over a light-dark cycle. The experiment was designed to capture the transient and prolonged behaviours of the organism during the photoperiod, such that an accurate sampling time can be estimated at both the transcript and protein level. The similarities and differences between the profiles of mRNA and protein abundances as a result of the circadian rhythm are also discussed.

6.3 Materials and Methods

6.3.1 Growth conditions.

The set-up and operation of batch cultures of *Synechocystis* sp. PCC 6803 adapted to a 12-hour dark and light cycle was described in Chapter 5. Briefly, *Synechocystis* sp. PCC 6803 was grown in BG-11 media in a 2 litre Braun BioStat B+ fermentor (Sartorius BBI, Melsungen Germany) at 30°C, with a 1.5 litre working volume on a 12 hour light-dark cycle. Light was supplied by cool-white fluorescent lamps at an incident intensity of 120 μEinstein/m^2s. The fermentor was stirred continuously at 200 rpm and supplied with 0.2 μm filtered air with 3% CO_2 at 1.0 litre/min. During the mid-exponential phase (OD_{730} of 3.5) (refer to Figure 6.2), cells

were collected at different designated time-points as illustrated in Figure 6.1. All chemicals were purchased from Sigma-Aldrich (Gillingham, Dorset, U.K.) unless otherwise stated.

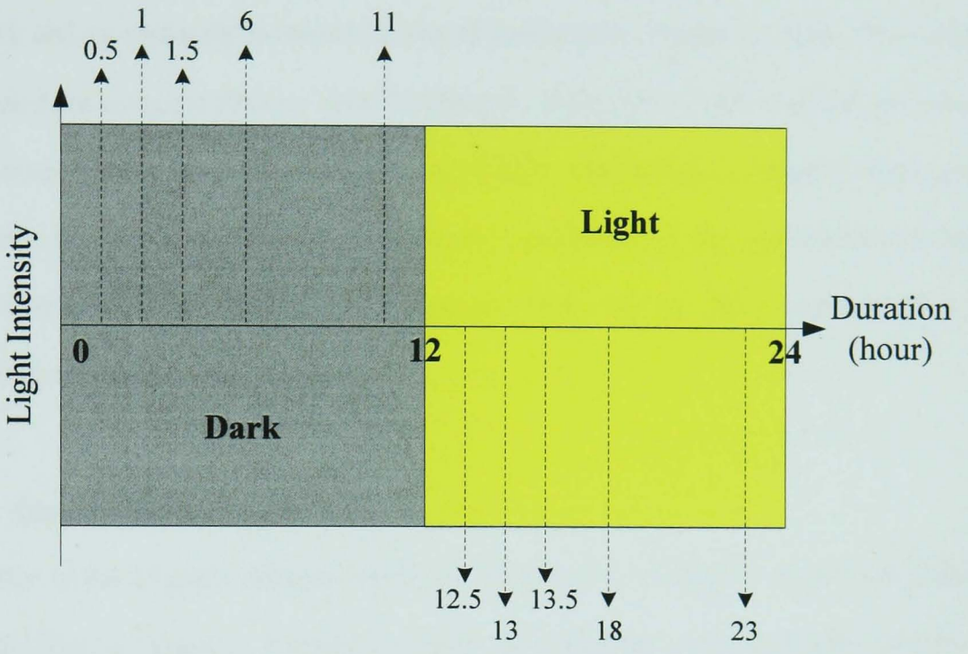


Figure 6.1. The sampling time across 12-hour dark and light period designed to capture the transient and prolonged effects of gene and protein expression. Briefly, samples was taken at an early phase of the cycle at 30-minute intervals up to 90-minutes, and mid-phase at 6-hour and finally during the late-phase at 11-hours into light and dark cycle respectively. A total of 10 time-points were collected in both conditions.

6.3.2 Protein sample preparation.

The detailed protocol for protein sample preparation was described in Chapter 5. Briefly, *Synechocystis* sp. cells were re-suspended in 500 mM triethylammonium bicarbonate buffer (TEAB), pH 8.5 and subjected to liquid nitrogen extraction coupled with mechanical cracking using a sterilised mortar and pestle [36]. The total soluble protein concentrate was then measured using the RC DC Protein Quantification Assay (Bio-Rad, Hertfordshire, UK) according to the manufacturer's protocol.

6.3.3 RNA sample preparation.

The detailed protocol for RNA sample preparation was described in Chapter 5. Cells were added directly to a 10% volume of 95:5 (vol:vol) ethanol and phenol before cooling with liquid N_2 [414]; and immediately harvested by centrifugation prior storage at -80°C. RNA extraction was carried out using the RNeasy Mini Kit (Qiagen, West Sussex, UK) coupled with on-column DNase treatment (Qiagen). A total of 300 ng of RNA was used to synthesize cDNA using the Quantitect Reverse Transcription kit (Qiagen), according to the manufacturer's protocol. Negative controls were carried out to ensure there was no DNA contamination before proceeding to real time PCR analysis.

6.3.4 Quantitative real time-PCR.

In addition to the 23 genes selected previously in Chapter 5, a further 3 genes were added to the list as detailed in Table 6.1. The selected genes are phosphate substrate-binding protein (*pst*S1, sll0680), large subunit bidirectional hydrogenase (*hox*H, sll1226) and cyanophycin synthetase (*cph*A, slr2002). They were selected due to their strategic ability in the phosphate transport system (discussed in detail in Chapter 7), as well as in hydrogen and cyanophycin production. The preparation of cDNA templates and the set-up for quantitative PCR analysis are detailed in Chapter 5. Quantitative real-time PCR was performed on an iCycler IQ machine (Bio-Rad), using SYBR Green Jump Start Taq ReadyMix reagent (Sigma).

Table 6.1. Primer sequences for the 3 genes selected for the quantitative real-time PCR study and their corresponding melting temperature (Tm).

Gene	Forward primers	Reverse primers	Tm (°C)
sll0680	TAAATTCGGGACTCATGGCGGC	CCGCTGGTAGCATTGTTATGGC	60
sll1226	AGGCAAGAATTTCCACCAAGCG	ATCTCTGGGTTATCCCGATGGC	60
slr2002	GAAGCCATCAATGATGTCGGGG	CCGGTTACATGGGCTGGAATTC	63

The sequence of these primers is tabulated in Table 6.1. For quantitative purposes in the real time analysis, standard calibration curves were generated from a series of dilutions (dilution factor of 3) of the cDNAs pool for all samples (both control and experimental). All real-time

RT-PCR analysis was carried out in triplicate for all samples. All time-points were normalised against the 23S rRNA gene [415] in each condition before referencing against the 24-hr dark sample. The quantitation value is expressed as a relative expression against the 24-hr dark sample.

6.3.5 Isobaric peptide labelling.

100 μg of protein from each phenotype were reduced, alkylated, digested and labelled as discussed in Chapter 5. In addition to following the manufacturer's (Applied Biosystems, Framingham, MA, USA; MDS-Sciex, Concord, Ontario, Canada) protocol, a two-day tryptic digestion and two volumes of 100% ethanol used during labelling were also included.

Table 6.2. Summary of the iTRAQ labelled samples in four different experiments.

	iTRAQ Reagent Label			
	114	115	116	117
Experiment 1	24-hour dark	0.5-hour dark	1-hour dark #	1.5-hour dark
Experiment 2	24-hour dark	1-hour dark #	6-hour dark	11-hour dark
Experiment 3	24-hour dark	12.5-hour light	13-hour light *	13.5-hour light
Experiment 4	24-hour dark	13-hour light *	18-hour light	23-hour light

#, * experimental replicates

All ten samples were distributed across four iTRAQ experiments, with a common control/reference sample adapted from Chapter 5, i.e. 24-hour dark sample as depicted in Table 6.2. Experiment 1 (label 116) and 2 (label 115) shared an experimental replicate of 1-hour dark sample, similarly experiment 3 (label 116) and 4 (label 115) shared the 13-hour light sample as an experimental replicate (Table 6.2).

6.3.6 SCX fractionation, mass spectrometric and data analysis.

The fractionation of the combined iTRAQ labelled peptides; the mass spectrometric analysis of the fractionated samples and the MS data analysis were described in detail in Chapter 5. SCX fractionation was carried out using a PolySULFOETHYL™ A Column (PolyLC, Columbia, MD, USA) 5 μm particle size of 100 mm length × 2.1 mm id, 200 Å pore size, on a BioLC

HPLC unit (Dionex, Surrey, UK). Fractions were collected every minute using a Foxy Jr. Fraction Collector (Dionex), and later were pooled together according to variations in peak intensity. Pooled fractions were dried in a vacuum concentrator, and stored at -20°C prior to mass spectrometric analysis.

Mass spectrometry was performed using a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF-MS/MS (Applied Biosystems; MDS-Sciex) coupled to an online capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands) as described previously [36]. A total of 3 injections were made for each sample. The mass spectrometer was set to perform data acquisition in the positive ion mode, with a selected mass range of 300 – 2000 m/z. Peptides with +2 to +4 charge states were selected for tandem mass spectrometry, and the time of summation of MS/MS events was set to 3 seconds. The two most abundantly charged peptides above a 5 count threshold were selected for MS/MS, and dynamically excluded for 60 seconds with ± 50 mmu mass tolerance.

Protein identification and quantification was carried out using ProQuant software v1.1 (Applied Biosystems; MDS-Sciex). The search was performed against the "mixed genome database" created in-house (see Chapters 4 and 5 for further detail). The search parameters allowed for peptide and MS/MS tolerance were up to 0.15 Da and 0.1 Da respectively; one missed cleavage of trypsin; oxidation of methionine and cysteine modification of MMTS. Only peptides above 70% confidence were saved for identification and quantification. ProGroup Viewer software v1.0.6 was used to identify proteins with at least 95% confidence. The results obtained from ProGroup Viewer were exported to Microsoft Excel for further analysis.

6.4 Results and Discussion

6.4.1 The experimental design

In this study, gene and protein expressions were examined using quantitative real-time RT-PCR and an iTRAQ-mediated quantitative proteomic approach respectively. Previously it was

reported that a 30-min sampling time can accurately capture the transcriptional response throughout the dark-light transition [66], hence the transient time-points were designated at 30 min intervals. In total, 6 transient (0.5 hr, 1 hr and 1.5 hr in dark; 12.5 hr, 13-hr and 13.4 hr in light) and 4 prolonged (6 hr and 11 hr in dark; 18 hr and 23 hr in light) time-points were collected for both protein and mRNA studies (refer to Figure 6.1). The term 'transient' used in this context refers to the shorter time of exposure to either light or dark, of less than 90 min in total; whereas 'prolonged' refers to a longer exposure time to these conditions, as defined by 6 hr and 11 hr into the phases. Samples were separated accordingly as depicted in Table 6.2, where the transient time-points were distributed between experiments 1 and 3, followed by prolonged time-points in experiments 2 and 4. The experimental replicates were added to allow inter-iTRAQ comparison, especially between experiments 1 and 2 (dark); and experiments 3 and 4 (light). The dataset in experiment 1 was normalised against the variation caused by the difference between the experimental replicates in experiment 1 and 2. The same philosophy was applied to experiments 3 and 4. The main assumption for this normalisation technique is that the bias across labels (114–117) is small and negligible; therefore the variation derived from any of the four labels is consistent between them.

The sampling strategy is illustrated in Figure 6.1. The sampling was started at the 17^{th} day of the growth and ended at the 18^{th} as depicted in Figure 6.2. The growth was slightly affected by the large volume of sampling (approximately 50mL per time-points). At the end of the period, the growth resumed as illustrated in Figure 6.2, before reaching the stationery phase a week later. During the dark phase, the cells did not grow (as measured by OD_{730}). However they remained viable under dark conditions after examining the cells using a haemocytometer. Furthermore, the cells recovered and grew upon exposure to light. A similar observation was also reported elsewhere [66]. The cells were adapted to a repeating 12 hr dark-light cycle for more than 2 weeks before the sampling process. Therefore, the results obtained here can be taken as a representation of this organism's response to the circadian rhythm and the switch between the photosynthesis and respiration processes.

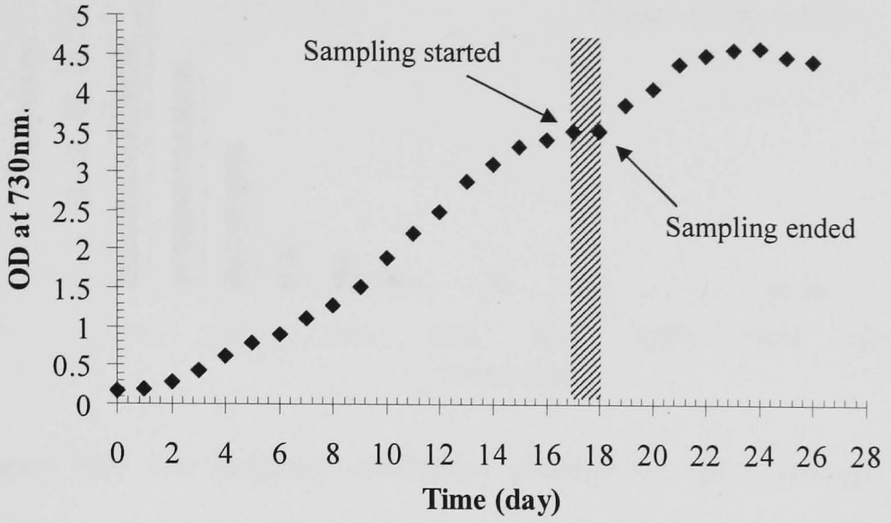


Figure 6.2. The growth rate of *Synechocystis* sp. (at 12 hr light-dark cycle) was measured via OD_{730} everyday at approximately 6 hr into light phase. A total of 2 independent experiments were conducted, and at each experiment at least 2 independent measurements were taken per day. The average OD_{730} is plotted in the graph.

6.4.2 Quality assessment

The relative protein and gene expression (against 24 hr dark sample) are expressed as a $\log_{10}$ ratio, and the associated error is estimated using the standard deviation. The selection and validation of the iTRAQ results were described in Chapter 4 and elsewhere [38, 39]. Briefly, all the proteins identified in this study must have had a confidence of at least 95%, and more than 2 MS/MS experiments. The total number of proteins identified (≥ 2MS/MS) per experiment was 150, except experiment 3 with only 120. Among them, only 82 proteins were found in all 4 experiments (≥ 2MS/MS in each experiment). In order to demonstrate the reliability of these 82 proteins, the percentage variation between the experimental replicates of experiments 1 (116:114) and 2 (115:114) in dark conditions, and experiments 3 (116:114) and 4 (115:114) in light conditions, was calculated (Figure 6.3). From this, over 95% of the proteins showed less than 50% variation, which suggests a sign of good quality control within this tolerance, and consistent with the previous studies described in Chapter 4 and [38].

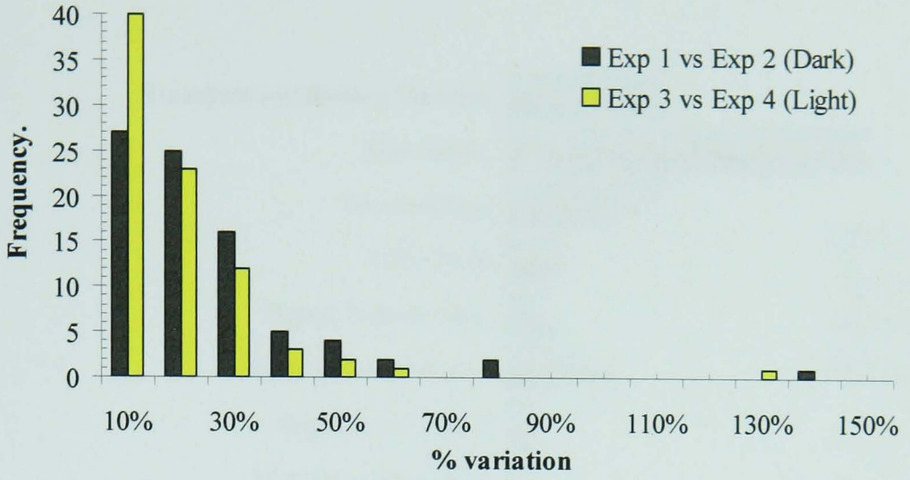


Figure 6.3. The frequency distribution obtained for the percentage variation from the experimental replicates (experiment 1 versus 2 in the dark; experiment 3 versus 4 in the light).

Proteins involved in photosynthesis contributed more than 14% (average) of the total proteins identified in all experiments (Figure 6.4). This is followed by hypothetical proteins and translation-related proteins at approximately 11% and 10% respectively. It is also worth discussing proteins involved in folding, sorting and degradation processes, and amino acids metabolism, at approximately 8% each per category, were also among the top numbers of proteins reliably identified and quantified. However, there is no distinct difference in the distribution of the functional group proteins, between the transient (experiment 1 and 3) and prolonged (experiment 2 and 4) time-points. The full list of proteins identified in this study is available in Appendices G-L.

The criterion for the selected genes in this investigation is dependent on the availability of proteomic data from the iTRAQ-mediated approach. Among the 82 proteins, 26 of them were chosen due to their strategic role in major metabolic pathways (Table 6.3). Most attention was focused on photosynthesis and translation related proteins (since they were the largest group identified in the iTRAQ experiments), whereas glycolysis/gluconeogenesis proteins were selected to investigate the role of central carbon metabolism (see Chapter 5).

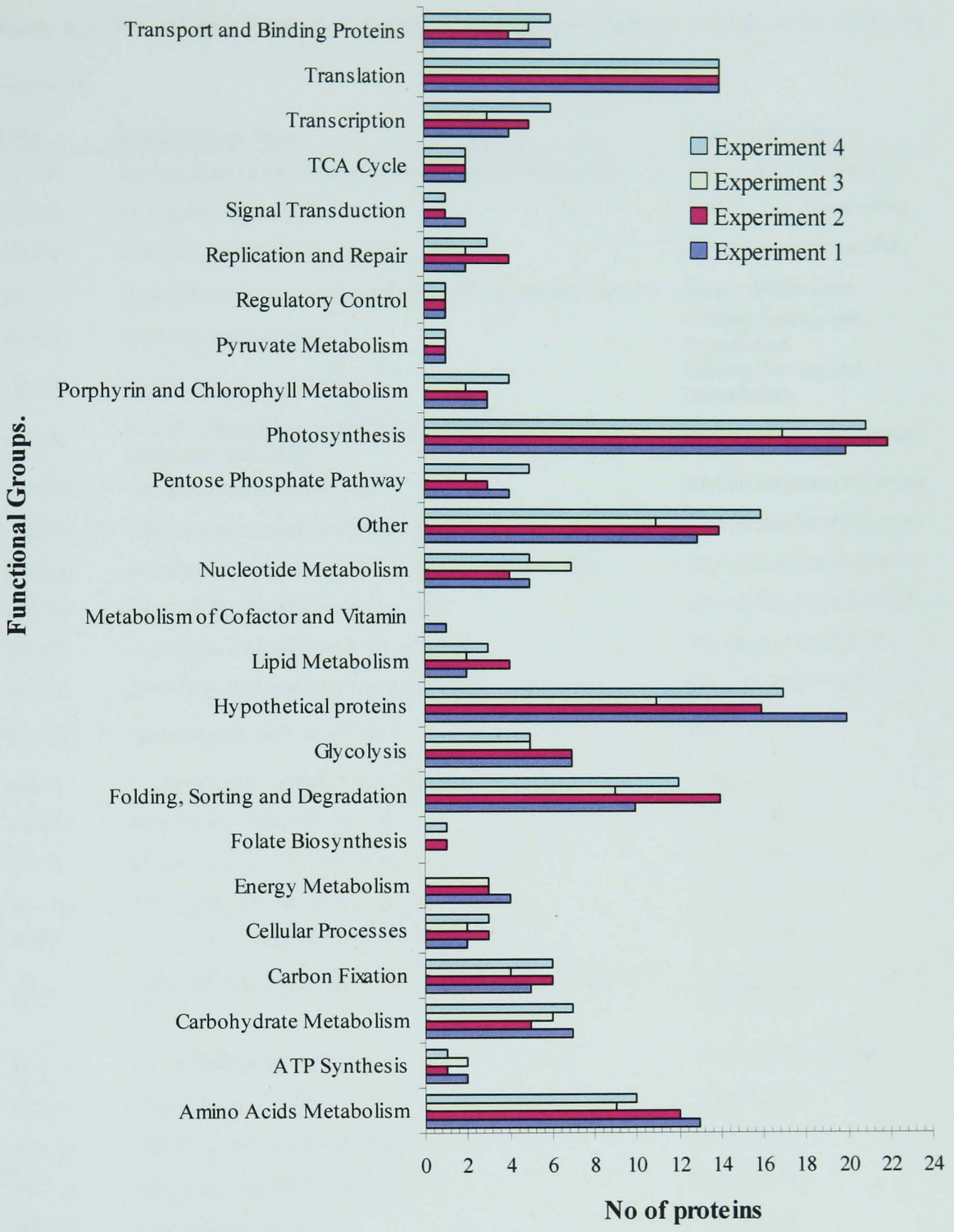


Figure 6.4. The total numbers of confidently identified proteins (≥2 MS/MS) from all 4 experiments, with their corresponding metabolic functions.

Table 6.3. The selected genes and proteins in this study with their corresponding metabolic functions.

ORF	Protein/Gene Name	Functional Groups
sll1502	NADH-dependent glutamate synthase large subunit, gltB	Amino Acids Metabolism
sll0945	Glycogen synthase, glgA	Carbohydrate Metabolism
sll1356	Glycogen phosphorylase, glgP	Carbohydrate Metabolism
sll1226	Hydrogenase subunit of the bidirectional hydrogenase, hoxH	Energy Metabolism
slr2076	60kD chaperonin, groEL1	Folding, Sorting and Degradation
sll0170	Heat shock protein 70, dnaK2	Folding, Sorting and Degradation
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase, gap2	Glycolysis/gluconeogenesis
slr0394	Phosphoglycerate kinase pgk	Glycolysis/gluconeogenesis
slr1349	Glucose-6-phosphate isomerase, gpi	Glycolysis/gluconeogenesis
slr2094	Fructose-1,6-/sedoheptulose-1,7-bisphosphatase, fbpI	Glycolysis/gluconeogenesis
sll0018	Fructose-bisphosphate aldolase, fbaA	Glycolysis/gluconeogenesis
sll1852	Nucleoside diphosphate kinase, ndk	Nucleotide Metabolism
slr1622	Soluble inorganic pyrophosphatase, ppa	Nucleotide Metabolism
slr2002	Cyanophycin synthetase, cphA	Other
slr1963	Water-soluble carotenoid protein, ocp	Other
sll1577	Phycocyanin beta subunit, cpcB	Photosynthesis
sll1194	Photosystem II 12 kDa extrinsic protein, psbU	Photosynthesis
slr1834	P700 apoprotein subunit Ia, psaA	Photosynthesis
sll0947	Light repressed protein A homolog, lrtA	Photosynthesis
sll1994	Porphobilinogen synthase (5-aminolevulinate dehydratase), hemB	Porphyrin and Chlorophyll Metabolism
sll1712	DNA binding protein, hup	Replication and Repair
sll1626	LexA repressor, lexA	Transcription
sll1260	30S ribosomal protein S2, rps2	Translation
sll1746	50S ribosomal protein L12, rpl12	Translation
slr0434	Elongation factor P, efp	Translation
sll0680	Phosphate-binding periplasmic protein, pstS1	Transport and Binding Proteins

6.4.3 Oscillations at the transcript level.

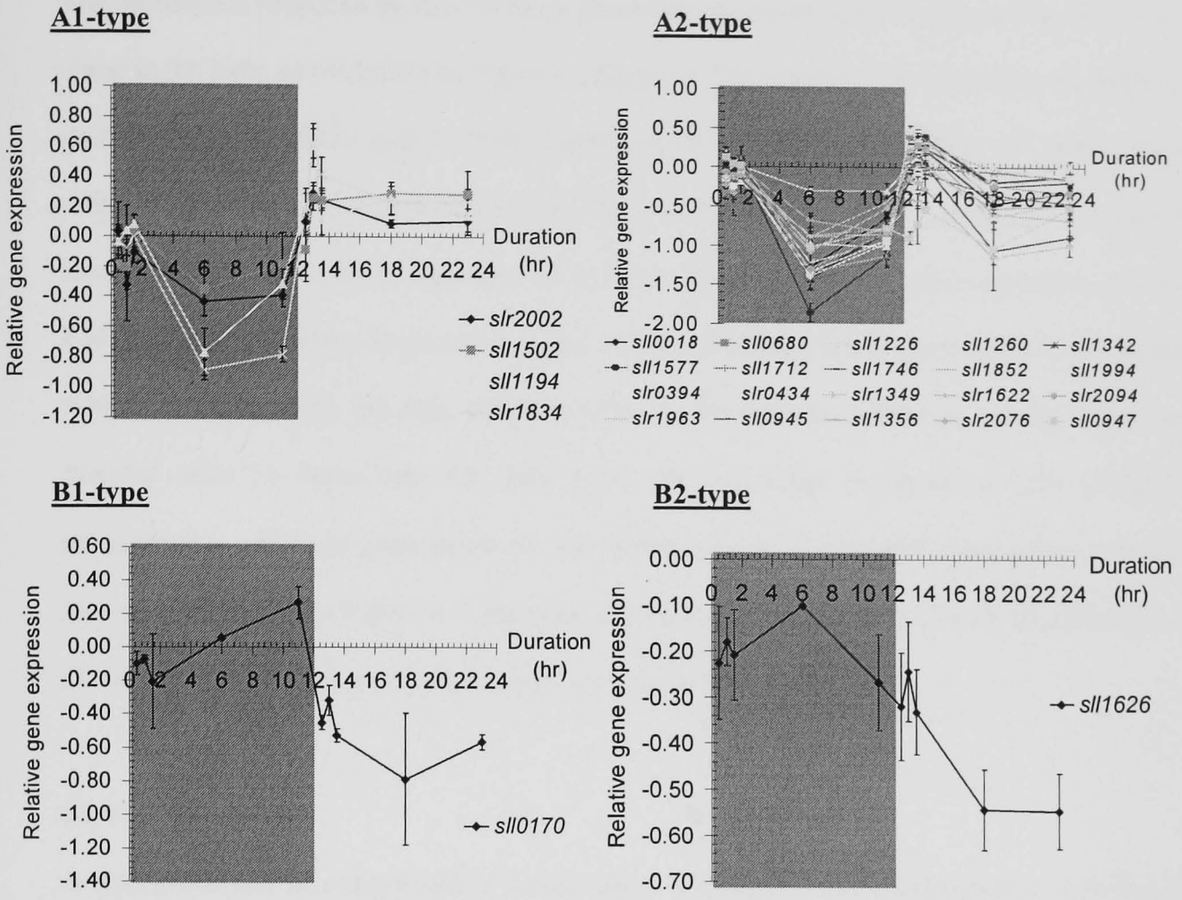


Figure 6.5. The relative expression level of 26 selected genes, with 2 distinct profile types (A- and B-type) and 4 different sub-groups (A1-, A2-, B1- and B2-). The shaded area refers to the dark cycle, while the un-shaded area indicates the light cycle. The y-axis measures the relative gene expression in Log_{10} units relative to the expression of the 24 hr dark time-point.

The transcript relative abundance through the period of the 12-hour dark and light cycle has clearly entrained the circadian rhythm on these set of genes, as illustrated in Figure 6.5. It is worth noting the fermentor was set to a cycling 12 hr dark and light interval, which suggests the change in the environment triggers the initiation of each half-cycle. Two types of gene responses are defined based on their dynamic gene expression profiles across the 12 hr dark and light cycle period: A-type (24 genes) and B-type (2 genes) (Figure 6.5). Generally A-type genes are characterised as highly sensitive to photonic (light) activation; whereas B-type genes are dark-responsive. Furthermore, each gene type is further separated into sub-groups of A1-, B1- (high transcript sustainability) and A2-, B2- (low transcript sustainability) (refer to Figure 6.5).

The oscillations produced by most of these genes have a convex shape in the dark and a concave shape in the light, as illustrated in Figure 6.5 (A-type). The relative abundance of some mRNAs, such as *cph*A (slr2002), *glt*B (sll1502), *psb*U (sll1194) and *psa*A (slr1834) are clearly more stable in the light condition as depicted in Figure 6.5 (A1-type). This is in marked contrast to those transcripts depicted in Figure 6.5 as A2-type, where the relative gene expression was not sustained above the basal level, and declined within the 6 hour period. Conversely, B-type genes exhibited bias towards the dark cycle, specifically the B1-type where an up-regulation was detected after 11 hours into the dark cycle. The transcript response of *lex*A (sll1626), categorised as a B2-type gene, however, only lasted 6 hours into the dark phase before a further decrement in the light (Figure 6.5, B2-type).The individual gene profile for all 26 genes across the whole 12-hour dark and 12-light cycle is available in Appendix M.

6.4.4 A-type genes

A-type genes can be categorised as being highly sensitive to the introduction or removal of light. These can be defined by two broad types, A2 and A1. The A2-type has a response which peaks within the first 90 minutes of exposure to light, reduces to zero thereafter, and is significantly down-regulated during the dark. Meanwhile, the A1-type, whilst having similar features as the A2-type, maintains a more persistent transcript profile upon exposure to light (Figure 6.5, A1-type).

The high transcript sustainability over the period of illumination for A1-type genes has made these candidate genes ideal as "metabolism driving" genes during the light cycle. Among them are the Photosystem I gene, *psa*A (slr1834) and Photosystem II gene, *psb*U (sll1194), both involved in the photosynthesis core reaction centre in cyanobacteria. The relative abundance of the *psb*U transcript was significantly up-regulated in the light from 1.5-fold after 30 min of light exposure to 4 fold after 60 min. Beyond this point, at 90 min, its abundance was maintained at 1.3-fold above the basal level. This is not significant up-regulation according to the defined cut-off of 1.5-fold compare to dark conditions. The transcript profile of *psa*A is comparable to

*psb*U, with a higher gene differential expression (1.7-fold) at the end of the light cycle (refer to Appendix M).

It is worth pointing out that *psa*A was down-regulated during prolonged exposure (24 hr) to the light as reported in Chapter 5, and a different mechanism of modulation is also suggested between Photosystem I and II in extended light conditions. Undoubtedly, the transcript relative abundance of *psa*A from this study disagrees with the findings reported in Chapter 5, however, this might actually support the claim of a different mechanism involved in Photosystems I and II. In this Chapter, the transcript relative abundance clearly shows a consistent performance by changing by more than +1.5-fold for the duration of the 12-hr light phase compared to the dark phase. However, from findings reported in Chapter 5, the transcript relative abundance for *psa*A was lower by more than 1.5-fold after 24 hr of continuous light. The experimental set-up and conditions are the same in both studies, except for the length of exposure to the light, 12 hr (this Chapter) versus 24 hr (Chapter 5). The effect of longer illumination (24 hr), however, does not effect the Photosystem II (refer to Chapter 5). In cyanobacteria, a decrease of Photosystem content is typically observed under high light conditions (over 200 μEinstein/m^2s), and the main target of down-regulation is not Photosystem II but Photosystem I [448, 449]. Since the incident light intensity is the same in both experiments, at 120 μEinstein/m^2s, the speculation is that *Synechocystis* sp. decreases its Photosystem abundance by reducing the Photosystem I content (and maintaining Photosystem II), hence directly causing a shift in the Photosystem stoichiometric ratio between Photosystem I and II upon prolonged exposure to constant light conditions, as to avoid cell degradation and denaturation caused by prolonged light-stress.

From the work described in Chapter 5, *glt*B (sll1502) did not show any significant relative expression differences in the transcript abundance at 24 hr of extended illumination. However, the results obtained here on a 12:12 light-dark cycle are in contrast, since *glt*B shows an approximate two-fold increase in relative abundance throughout the entire light cycle (except at 30 min into the light phase where no significant differential expression was observed). Interestingly, a 60-min lag time response in the light was observed for *glt*B (see Appendix M).

This gene synthesizes the formation of L-glutamate from L-glutamine and 2-oxoglutarate through NADH oxidation [450]. The delay in the response is possibly caused by a lack of substrate abundance, since these compounds are not always in surplus in the cell. It is not yet understood why there is a difference in gene expression in the current study (12-hr light and dark), compared to the previous study (24 hr constant light), although it has been reported previously that *glt*B is non-essential for *Synechocystis* sp. growth, as measured through enzyme assays and western-blot analysis [451]. Since *glt*B is not essential for growth, the results here clearly indicate its involvement in circadian rhythm promoted through the light and dark cycle, rather than as an accessory gene during light acclimation.

Cyanophycin (also known as multi-L-arginyl-poly-L-aspartic acid) is a nitrogen storage polymer found in cyanobacteria [90]. It is synthesized from arginine and aspartate in an ATP-dependent reaction catalyzed by a single enzyme, cyanophycin synthetase [89]. One of the subunits of this enzyme, *cph*A (slr2002) was identified here showing an increase in relative abundance in the light phase, with the highest fold change (+2-fold) after 60 min in the light, with this gradually decreasing in expression but maintaining the transcript relative abundance around basal level (1.3-fold) over the 12 hr incubation in the light (refer to Appendix M). It has been reported that nitrogen incorporation into cyanophycin from external growth medium supplemented with chloramphenicol has a higher intake rate under normal light of 100 μEinstein/m^2s compared to low light at 4 μEinstein/m^2s [5]. The observation here is evidently in agreement with the claim by providing further understanding at the transcript level during cyanophycin proliferation under light and dark conditions.

The selected glycolysis/gluconeogenesis and carbohydrate metabolism genes in central carbon metabolism: *fba*A (sll0018), *gap*2 (sll1342), *gpi* (slr1349), *fbp*I (slr2094), *pgk* (slr0394), *glg*A (sll0945) and *glg*P (sll1356) are all within the A1-type category. Some of these genes are not only involved in glycolysis and gluconeogenesis, but also are key enzymes in the pentose phosphate pathway and Calvin cycle (refer to Figure 5.2 in Chapter 5). As discussed in Chapter 5, light triggers photosynthesis and other light-dependent processes, in particular the Calvin

cycle, where carbon dioxide is fixed from the atmosphere as the main carbon source (Figure 5.2C). During dark incubation, the transcript relative abundance of these genes decreases by over 10-fold compared to the light phase. This is particularly noticeable at the mid-point in the dark cycle. However, during the transition from the dark to the light period, the relative mRNA abundance of these genes increases by more than 2-fold during the first 90 minutes of light exposure. The relative abundance then falls back to the baseline level (or lower) by the time 6 hrs into the light phase is reached (Figure 6.5 and Appendix M). This is in agreement with work reported by Mary *et al.*, where the transcript abundance of stress-associated genes (*clp*, *dna*, *gro* and *hsp*) reached a maximum between 1 to 3 hr when *Synechocystis* sp. cells were subjected to a high light condition from a pre-adapted low light environment [171].

The positive controls chosen for this study were the phycocyanin beta subunit, *cpc*B (sll1577) and the porphobilinogen synthase, *hem*B (sll1994), both involved in the light-harvesting mechanism. The transcript relative abundance of these two genes increased by 2.4- and 3.0-fold respectively during the first 90 min of the light exposure, but this was not sustained over the whole light cycle (refer to Appendix M). Another gene closely related to the photosynthesis process is one of the subunits of the reversible hydrogenase gene complex, *hox*H (sll1226). The transient increase in the abundance of the *hox*H transcript found in this study, is consistent with observations in the literature where transient H_2 production was observed in *Synechocystis* sp. PCC6803 during the first few minutes of illumination [71].

In addition to photosynthesis-related processes, it is also worth examining the genes involved in nucleotide metabolism (*ndk*, sll1852; and *ppa,* slr1622) and translation processes (*rps*2, sll1260; *rpl*12, sll1746; and *efp,* slr0434), which also can be categorised as demonstrating an A-type profile. In exposure to the light, a different set of chemical reactions replace the existing dark cycle reactions, such as respiration. It is unsurprising to observe that the synthesis of protein and nucleotide molecules was accelerated for this purpose to overcome such stress. Other remaining A1-type genes are *gro*EL1 (slr2076), *ocp* (slr1963), *hup* (sll1712), *pst*S1 (sll0680) and *lrt*A (sll0947).

6.4.5 B-type genes

The gene *dna*k2 (sll0170) was only one of two genes measured in this investigation that showed a significant relative increase in expression during incubation in the dark. These genes were categorised as B1-type. In this case, the relative transcript abundance gradually increased with time during the dark until it reached a peak after 11 hr, with a 1.8-fold increase over the abundance observed in the light (Figure 6.5 and Appendix M). This finding agrees with results reported in Chapter 5, where *dna*K2 was reported to be down-regulated by over 2-fold after 24 hr of light exposure (i.e. 2-fold up-regulated in 24 hr dark). This result also further strengthens the observation that none of the relative abundance measurements taken at the various time-points show an up-regulation in the light, hence implying that this gene is not associated with light-dependent processes (at least in constant light conditions), in contrast to literature reports where *dna*K2 was associated with high-light adaptation [171].

The gene *lex*A (sll1626) is thought to be involved as the transcription regulator of the bidirectional hydrogenase operon (*hox*EFYUH) [78, 79]. It is shown above that *hox*H is an A1-type gene, triggered by a change in light intensity; hence their circadian profile must be within close proximity. Yet the result here does not indicate the involvement of *lex*A in the light cycle. As the relative abundance of the transcript reached a maximum at 6 hours into the dark, it gradually decreased below the baseline level at 11 hour into the dark, and decreased even further upon illumination (Figure 6.5 and Appendix M). Much speculation has been made on the involvement of *lex*A as a redox modulator [82], yet the actual functional mechanism of this gene is still largely unknown.

6.4.6 'Shock' response at the protein level.

The change in the system from a light period to a dark period caused perturbations in relative transcript abundances in almost all of the 26 genes tested. It was thus of interest to assess whether these changes in mRNA abundance were reflected in the translation of these messages. Previous studies have shown that there may not be a direct relationship between changes seen in

the transcript, and those seen in the corresponding protein relative abundance [452]. In order to assess the proteome profile across the 12 hr dark and light conditions, *Synechocystis* sp. PCC 6803 was subjected to the iTRAQ-mediated quantitative proteomic experiments. In total, 82 proteins were found in all 4 experiments with ≥ 2MS/MS in each experiment (refer to section 6.4.1 and 6.4.2 for detailed experimental design). Over 90% of the proteins showed a 'shock' response – a spike (either up or down regulation), during the transient time-period within 90 min of entry or exit into either the dark or light phase of the diurnal cycle (refer to Appendix M). After the 90 min time point, the relative expressions measured through the prolonged time-point sampling (between 6 and 11 hr after the perturbation) do not vary greatly in comparison to those seen closer to the transition (30 min, 60 min and 90 min). The influence of the circadian rhythm at the proteome level is less evident compared to the transcript level, but the effect is still noticeable (refer to Appendix M).

6.4.7 Response in photosynthesis-related proteins.

Proteins directly related to photosynthetic processes are the largest functional group identified in this study, with 19 out of the 82 confidently identified proteins (including 2 proteins in chlorophyll metabolism). To further illustrate the dynamic nature of the circadian profile at the proteomic level, more details are provided using photosynthesis-related proteins, since this process has a wide range of profiles. The dynamic profile within this group varies as their functional role diverts, although being categorised under the same family of photosynthesis.

This family consists of core reaction centres – Photosystem I and Photosystem II; light-harvesting apparatus – phycocyanin, allophycocyanin and chlorophyll; and electron transport unit – cytochrome b_6f complex. Each unit carries out a distinctive role during photosynthesis. By dissecting the proteins into each functional unit, a range of dynamic profiles are observed as illustrated in Figure 6.7. A common feature as discussed before, is the 'shock' response during the first 90 min of each light-dark phase transition, is evident in this group of proteins. The presence of light triggers the gene response, which translates message into a protein to perform the designated activity as seen from the transcript relative abundance above (Figure 6.5). Aside

from the phycocyanin alpha subunit, CpcA (sll1578) and beta subunit, CpcB (sll1577) which show significant up-regulation (>1.5-fold) during prolonged light exposure (Figure 6.7C), the remainder of the identified photosynthesis-related proteins remain unchanged in relative abundance in the light.

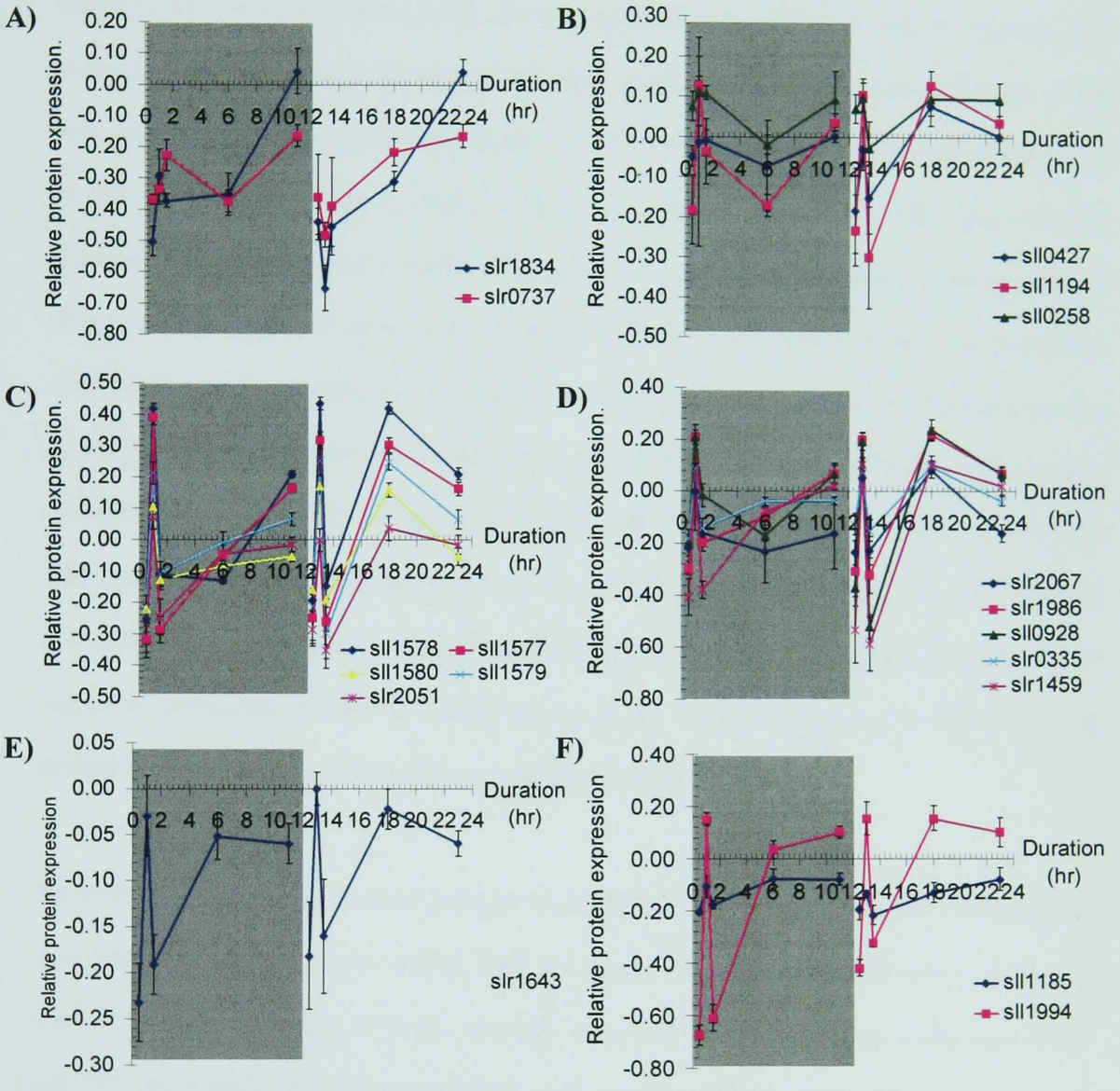


Figure 6.7. Relative expression of photosynthesis-related proteins across the 12-hr dark and light cycle. These proteins are involved in (A) Photosystem I; (B) Photosystem II; (C) phycocyanin; (D) allophycocyanin; (E) cytochrome b_6f; and (F) chlorophyll metabolism. The shaded area refers to the dark cycle while the un-shaded area is the light cycle. The y-axis measures the relative protein expression in Log_{10} units relative to the expression of the 24 hr

dark time-point. sll0947 (LrtA) is not shown in the figure as it does fit into any of the profiles shown.

A particularly interesting phenomenon is the mirror-like effect in Figure 6.7, where proteins showed almost identical profiles in the dark and light phases. This observation is not only visible in this group of proteins, but is also noticeable amongst the other functional groups such as amino acids metabolism, folding, sorting and degradation, and hypothetical proteins. It has been reported that the cell is most effective in a rhythmic environment where the frequency of its internal biological clock is similar to the environmental cycle [439]. After more than 14 circadian cycles (more than 2 weeks of growth), it is possible the organism 'remembers' the time to change phases. In order to be more effective under both photosynthesis and respiration conditions, it produces additional proteins at the interchange phase so that a transition from one cellular activity to the other can be enacted more rapidly. Specifically for proteins involved in photosynthesis, one would imagine that this group of proteins is more susceptible to the light alteration. The finding demonstrates the circadian effect at the translational level (in this case photosynthesis), noticeable at a smaller scale compared to the transcriptal level. After all, the organism synthesizes proteins in abundance and so it is unsurprising to see a smaller scale of protein synthesis process in both light and dark conditions.

All units of photosynthesis-related proteins show different dynamic profiles; however the same functional unit proteins (such as PS1, PSII and phycocyanin) do cluster together, showing a similar fluctuation (Figure 6.7A-F), although there are some exceptional cases like HemF (sll1185) and HemB (sll1994) where their profiles do vary (Figure 6.7F). The relative protein and gene expression profile for the 26 genes selected in this study were plotted, and these figures are available in Appendix M.

6.4.8 Correlation between translation and transcription abundance changes.

Before a protein is made, the corresponding gene has to be induced through a change in the environment, which later translates into the protein to undertake the designated function. This is

the central dogma of biology. If the protein is always translated from the gene, and then a correlation between them is possible by taking into account the gene input, protein accumulation, rate of protein synthesis and degradation, and the probable inhibition factors that might cause a delay during the translation process. Many studies have attempted to elucidate and model this complicated relationship, yet they often do not correlate well [452-454].

This Chapter seeks to provide a rational explanation behind this sophisticated relationship through the observation of the protein and gene expression profiles. By comparing the circadian profile across the dark and light cycle at both the protein and gene levels, the hope is to generate some hypotheses that can form the basis for understanding their interrelationship. Here, the 26 selected genes exhibited five different temporal gene profiles (A and B-type in Figure 6.5), measured using real time RT-PCR, which become the basis of their response toward the light. The corresponding protein relative expression profile, measured via iTRAQ, provides a platform for comparison to the transcript profile, to yield integrated insight into the circadian rhythm responses. A summary of the transcript and protein relative abundance temporal relationship patterns is summarised in Table 6.4.

Table 6.4. Summary of the relationship between the 26 selected genes and their corresponding gene product (protein) based on their temporal response profiles during a 12-hr dark and 12-light cycle.

	Gene-protein expression relationship	**Related ORF(s)**
i)	Transcript and protein change in same direction, but different relative expression magnitude (relative abundance).	sll0947, sll0945, sll1356, slr2076, sll1194, slr2002, sll1502, sll0680, sll1712, sll0170 (in dark)
ii)	Transcript and protein peak within 90 min of light-dark transition, but the transcript abundance change is not sustained over a prolonged period compared to the protein.	slr1963, sll1260, sll1852, slr0434, slr622, sll1577, sll1746, sll1994, sll0018, sll1342, slr2094, slr1349, slr0394
iii)	Transcript and protein relative abundance changes are in different directions.	sll1626, slr1834, sll0170 (in light)
iv)	No protein profile.	sll1226

Twenty two out of 26 (not including sll0170) of the selected genes show some agreement between transcript and protein behaviour (Table 6.4, relationship-types i and ii), with only 3 pairs correlating poorly (Table 6.4, relationship-type iii). It has been reported that mRNA encoding ribosomal proteins, generally, have a longer decay rate compared to other mRNAs [455, 456]. However in this study, two ribosomal genes encoding *rps*2 (sll1260) and *rpl*12 (sll1746) did not show significant differences in their transcript profiles compared to other examined genes (Table 6.4). The cell detects changes in light intensity, and co-ordinates genes to up-regulate relatively rapidly, within the first 90 min of the change, as observed in Figure 6.5 (A1 and A2-type). During this period, proteins were synthesized and used to counter the effects of the perturbation. Once the protein synthesis process was completed to an appropriate level, a feedback signal reduced gene expression strength, which is observed in the relationship-type ii summarised in Table 6.4. Even though the transcript level was sustained, the change in expression may not be the same at the protein level as observed in the relationship-type i depicted in Table 6.4. If a direct link exists between the transcript abundance and the corresponding translated protein abundance, then the translation process is straight-forward from one gene to form one protein, as postulated by the central dogma of molecular biology.

6.5 Conclusions

The nature of a chemical reaction always involves at least a substrate and an enzyme, which ultimately leads to the formation of the product under the right conditions (such as temperature and pH). Before this can take place, the enzyme needs to be translated from mRNA. In this study, changes in transcript relative abundance were found to be more pronounced in response to environmental perturbations than changes observed in the corresponding proteins. The ability for the gene to turnover in a short amount of time is vital for the cell survival and adaptation.

The translated proteins, however, do not show the same kind of tuned circadian response seen in transcript relative abundance, where an oscillatory curve was observed with a periodic half-life of 12 hr in all 26 investigated genes. The dependency of these genes is therefore closely related

to those light-associated processes, in particular to the photosynthesis and central carbon metabolism. Within this group of genes, a total of 2 different types of circadian profiles were observed based on their individual distinctive response to the light (A-type) and dark (B-type) cycle. Furthermore, 2 sub-groups between each type were observed based on their transcript sustainability with A1-, B1- characterised as high; and A2-, B2 as low. A2-type genes respond to a change in the environment within the first 90 min, with their relative transcript abundance falling away beyond this time, in contrast to A1-type genes where the transcript concentrations were more persistent across the entire light period. There was only 1 gene measured, *dna*K2 (sll0170), that exhibited a significant up-regulation (high sustainability) in the dark cycle, and this gene was categorised as a B1-type gene. On the other hand, a B2-type gene, *lex*A (sll1626) showed a response toward the dark cycle. However, this was not sustained across the entire dark cycle period.

The protein responses to the environmental perturbation share a common feature, usually a shock response within the first 90 min of a change from the light and dark phases. This sharp 'spike' corresponds well with that observed in the transcript response to the environmental change, suggesting the possibility that the protein biosynthesis process is most active during this first 90 min period. Another common feature observed in the protein expression changes was the mirror-like effect observed in the dark cycle. Since most of the genes identified here displayed a positive bias (up-regulation) toward the light, the mirror-like effect may have been the influence of the circadian rhythm, which caused the protein to exhibit cyclic behaviour in the dark as a consequence of the previous light cycle.

The use of two different set of sampling times, transient and prolonged time-points, has clearly provided insight into the overall protein and gene responses to the environmental change. It is worth noting that the sampling time could prove to be important in future experimental design, especially for a circadian rhythm study, since the protein and gene expression levels could fluctuate significantly as the sampling time varies.

Chapter 7

Global Proteomic Response of *Synechocystis* sp. PCC 6803 During Phosphate Limitation

7.1 Abstract

Synechocystis sp. PCC 6803 was subjected to phosphate stress by reducing levels of potassium phosphate (K_2HPO_4) in BG-11 medium from 10% to 0.3% of normal. An iTRAQ-mediated proteomic approach was used to measure relative changes in protein expression as a response to these reduced phosphate concentrations. A total of 170 proteins were confidently identified in this study, with 56% of total proteins significantly up-regulated, and 15% significantly down-regulated at the more extreme phosphate-depleted condition (0.3% P-level). The most common response of the phosphate-deprived cells was observable with the in periplasmic proteins, especially with the alkaline phosphatase (sll0654). This protein showed enhanced protein expression from 1.39-fold to a greater than 2-fold increase when the phosphate concentration was reduced. Further measurements using phosphatase enzyme assays also revealed that the alkaline phosphatase activity increased from 0 to 61 nmoles of inorganic phosphate (Pi) released per min per mg of protein as the phosphate concentration reduced. No acid phosphatase activity was detected in all phosphate-reduced samples. The extracellular nuclease NucH (sll0656), which is located up-stream of the sll0654, also increased in relative abundance by more than 4-fold in 0.3% Pi medium compared to the phosphate-replete culture. Two of the phosphate substrate-binding periplasmic proteins (PstS1; sll0680 and PstS2; slr1247) were also identified in this study, although there is some slight variation in their relative protein expressions (sll0680 was down-regulated. whilst slr1247 was up-regulated). The influence of phosphate reduction was also observed in central carbon metabolism, stress-related proteins, amino acids metabolism, nucleotide metabolism and photosynthesis processes.

7.2 Introduction

One of the most essential nutrients in the cellular environment is phosphate, with phosphate being the lowest abundance naturally occurring mineral on this planet [457, 458]. Free phosphate ion in the medium, also called inorganic phosphate (Pi), is most commonly present in the form of adenosine phosphates (ATP, ADP and AMP) in the cell, and found in the main chemistry skeleton of DNA and RNA [58]. The two-component phosphate regulatory system (known as *pho* regulon) in *Escherichia coli* consists of a histidine residue of a specific histidine kinase, namely PhoR, and a specific aspartate residue on the cognate response regulator, known as PhoB [459]. When subjected to phosphate limitation, PhoR is phosphorylated, and the released phosphate group attached to the PhoB which in terms modulates the expression of genes in the Pho regulon that encode alkaline phosphatase, phosphate transporters, and other relevant proteins [460]. In *Synechocystis* sp. PCC 6803, it was found that sll0337 (*sph*S) and slr0081 (*sph*R) encoded proteins are homologous to PhoR and PhoB of *E. coli* respectively [250].

Furthermore, the *pst*SCAB operon is known to control the *pho* regulon in *Escherichia coli* [461]. Interestingly, there are two set of genes for phosphate-specific transport system (designated as Pst1 and Pst2) in the *Synechocystis* sp. PCC 6803 genome [51]. The *pst*1 system includes 6 ORFs (sll0679-sll0684) in the following order: *sph*X-*pst*S1-*pst*C1-*pst*A1-*pst*B1-*pst*B1'. The *pst*2 system includes 4 ORFs (slr1247-slr1250) as follows: *pst*S2-*pst*C2-*pst*A2-*pst*B2. Previously, it has been shown that under phosphate deprivation conditions, the Pho regulatory system in *Synechocystis* sp. responded by increasing its transcript abundance by more than 7-fold compared to the normal Pi replete conditions [253]. Similar responses at the transcript level were also reported for the Pho regulon systems of other bacteria such as *Bacillus subtilis* [255] and *Corynebacterium glutamicum* [462].

One of the earliest phosphate starvation studies was carried out in *Pseudomonas aeruginosa* [463]. The investigation was based on the measurements of optical density, cell counts, the chemical composition of the medium and the alteration from the ribosome pattern. Improved

techniques over the last four decades, stemming from genome sequencing and then post-genomics techniques employing cDNA microarray analysis, and proteomics (facilitated by improvements in mass spectrometry technology) has enabled researchers to focus at a more systems level basis. A genome-wide investigation of *B. subtilis* using a cDNA microarray platform identified that the PhoP and SigB regulon were up-regulated during phosphate starvation conditions, and allowed further genome-wide measurements of the general stress response system [464]. A similar observation of the up-regulation in Pho regulon was also reported from the phosphate-deprived cells of *B. Subtilis* using a 2-DE proteomic analysis [254, 255].

The introduction of isobaric tags for absolute and relative quantification (iTRAQ), using amine-reactive tagging reagents, has provided flexibility in analysing proteomes from up to four (and soon 8) samples simultaneously [37]. Taking advantage of this approach here, phosphate-limited cells from *Synechocystis* sp. PCC 6803 with different degrees of reduced phosphate concentrations (10%, 3% and 0.3% P-level), were subjected to iTRAQ-mediated quantitative proteomic analysis. The reproducibility and reliability of the iTRAQ workflow was investigated and demonstrated previously in Chapter 4 and elsewhere [38, 39]. This is the first time that a quantitative 'shotgun' proteomic approach has been carried out on the phosphate limitation response in any organism. This iTRAQ study forms the basis for understanding cellular responses at the protein level during phosphate limitation in *Synechocystis* sp.

7.3 Materials and Methods

7.3.1 Growth conditions.

Synechocystis sp. PCC 6803 was cultivated in BG-11 media in 1 litre Erlenmeyer flasks with a 0.5 litre working volume with approximately 100 rpm stirring, in a plant growth chamber (Sanyo, model MLR-350H) at 25°C on a 12 hour light-dark cycle. Light was supplied by cool-white fluorescent lamps at an incident intensity of 70 μEinstein/m^2s measured by a QSL-2100 Scalar PAR Irradiance Light Sensor (Biospherical Instrument Inc, San Diego, CA). The Pi-replete *Synechocystis* sp. control culture was grown in BG-11 media consisting of 40 mg/L of

K_2HPO_4 [43]. For phosphate deprivation experiments, the concentration of phosphate was reduced to 10% w/v (4 mg/L), 3% w/v (1.2 mg/L) and 0.3% w/v (0.12 mg/L) of the K_2HPO_4 in BG-11. The cells were washed twice with sterilised water and once with Pi-reduced media, prior to the introduction of Pi-replete cells into the Pi-reduced environments. The growth rate was monitored and recorded using an Ultrospec 2100-Pro spectrophotometer (Biochrom, Cambridge, UK) at an OD of 730nm (Figure 7.2). Cells were harvested on the 60th day of the growth by centrifugation at 5,000 × g for 5 min at 4°C. Cell pellets were stored at -80°C until required. All chemicals were purchased from Sigma-Aldrich (Gillingham, Dorset, U.K.), unless otherwise stated.

7.3.2 Protein sample preparation.

The cell pellets were re-suspended in 500 mM triethylammonium bicarbonate buffer (TEAB) at pH 8.5, and subjected to liquid nitrogen extraction coupled with mechanical cracking using a sterilised mortar and pestle [36]. Soluble proteins were recovered from the supernatant by centrifugation at 21,000 × g for 30 minutes at 4°C. The total soluble protein concentrate was then measured using the RC DC Protein Quantification Assay (Bio-Rad, Hertfordshire, UK) according to the manufacturer's protocol.

7.3.3 Pigment composition.

A wavelength scan of the pigment composition was achieved via methanol (MeOH) extraction protocol as described elsewhere [465]. Briefly, a 1 ml volume of sample was extracted with 90% MeOH at room temperature and incubated at 4°C for overnight in the dark. After overnight incubation, sample was gently vortexed and centrifuged at 13,000×g for 5 minutes at room temperature. The supernatant was measured via an *in vitro* absorption spectra between 300 and 900 nm at room temperature with an Ultrospec 2100-Pro spectrophotometer (Biochrom).

7.3.4 Acid/Alkaline phosphatase assays.

Cells broken by mechanical grinding with liquid nitrogen as described above were subsequently re-suspended in Tris-HCl buffer (pH 8.0) containing 5mM $MgCl_2$. Alkaline phosphatase assays

were performed in duplicate as described by O'Loughlin *et al.* [466] at 30 °C and contained (in μmoles/ml): Tris-HCl (pH 8.0), 50; $MgCl_2$, 20; p-nitrophenyl phosphate, 5 and cell-extract protein, 0.01 - 0.1 mg in a final reaction volume of 1 ml. For acid phosphatase assays, the Tris-HCl buffer was replaced with 50 μmoles/ml Succinate-NaOH buffer (pH 5.0).

Due to the presence of large quantities of chlorophylls in the cell-extract, detection of the of p-nitrophenol production by means of the widely used absorbance measurement at 412nm was not possible. Duplicate reaction tubes were therefore terminated with 0.2 vol of 3M trichloroacetic acid solution following addition of cell extract protein and again 10 min later. The resultant reaction supernatants were cooled on ice for 10 min prior to centrifugation at 20,000 × g for 5 min to remove precipitated protein, following which the phosphate concentration in the samples was determined using the acid molybdate method of Fiske and SubbaRow [467]. 'No substrate' and 'no cell-extract' controls were carried out at the same time as the 'with substrate' assays, and activity was expressed as nano moles of inorganic phosphate (Pi) released per min per mg of cell-extract protein (nmoles/min.mg), taking the control reactions into account. All phosphate determinations were carried out in duplicate.

7.3.5 Isobaric peptide labelling.

100 μg of protein in 20 μL of 500 mM TEAB from each phenotype was reduced with tris-2-carboxyethylphosphine (TCEP) for 1 hr at 60°C; alkylated using methyl methanethiosulfonate (MMTS) for 30 min at room temperature; digested with trypsin at 37°C before being labelled with iTRAQ reagents according to the manufacturer's (Applied Biosystems, Framingham, MA, USA; MDS-Sciex, Concord, Ontario, Canada) protocol with some modifications. These included a two-day tryptic digestion and two volumes of 100% ethanol used during labelling. Samples were tagged with iTRAQ labels as illustrated in Table 7.1. After a two-hour incubation period, labelled samples were then combined and dried in a vacuum concentrator. At the time of experimentation, it was not possible to use the extraction buffer described in Chapter 3, due to incompatibility problems with the iTRAQ reagents, as discussed in Chapter 4.

Table 7.1. Summary of the iTRAQ labelled samples carried out in this study for the proteome response of *Synechocystis sp.* during phosphate limitation.

iTRAQ Reagent Label	114	115	116	117
Sample	Control	10% P-level	3% P-level	0.3% P-level
K_2HPO_4 concentration (mg/L)	40	4	1.2	0.12

7.3.6 Strong cation exchange (SCX) fractionation.

Dried peptides re-suspended in 200 µL of Buffer A were fractioned using a PolySULFOETHYL™ A Column (PolyLC, Columbia, MD, USA) 5 µm particle size of 100 mm length × 2.1 mm id, 200 Å pore size, on a BioLC HPLC unit (Dionex, Surrey, UK) with a constant flow rate of 0.2 ml/min and an injection volume of 200 µl. Buffer A consisted of 10 mM KH_2PO_4 and 25% acetonitrile, pH 3.0 and Buffer B consisted of 10 mM KH_2PO_4, 25% acetonitrile and 500 mM KCl, pH 3.0. The 60 minute gradient consisted of 100% A for 10 minutes, 20% to 25% B for 30 minutes, 25% to 100% B for 5 minutes, 100% B for 5 minutes and finally 100% A for 5 minutes. The chromatogram was monitored through a UV Detector UVD170U and Chromeleon Software, version 6.50 (Dionex/LC Packings, The Netherlands). Fractions were collected every minute using a Foxy Jr. Fraction Collector (Dionex), and were later pooled together according to variations in peak intensity. Pooled fractions were dried in a vacuum concentrator, and stored at -20°C prior to mass spectrometric analysis.

7.3.7 Mass spectrometric analysis.

Each dried SCX peptide fraction was re-dissolved in 100 µL of Switchos buffer (0.1% formic acid and 3% acetonitrile), and then 20 µL of each sample was injected into the nano-LC-ESI-MS/MS system for analysis. A total of 3 injections were made for each sample. Mass spectrometry was performed using a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF-MS/MS (Applied Biosystems; MDS-Sciex) coupled to an online capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands) as described previously [36]. The peptide mixture was

separated on a PepMap C-18 RP capillary column (LC Packings), with a constant flow rate of 0.3 μL/min. The LC gradient started with 3% Buffer B (0.1% formic acid in 97% acetonitrile) and 97% Buffer A (0.1% formic acid in 3% acetonitrile) for 3 minutes, followed by 3% to 25% Buffer B for either 60 or 120 minutes, then 90% Buffer B for 7 minutes, and finally 3% Buffer B for 10 minutes. The mass spectrometer was set to perform data acquisition in the positive ion mode, with a selected mass range of 300 – 2000 m/z. Peptides with +2 to +4 charge states were selected for tandem mass spectrometry, and the time of summation of MS/MS events was set to 3 seconds. The two most abundantly charged peptides above a 5 count threshold were selected for MS/MS, and dynamically excluded for 60 seconds with a ± 50 mmu mass tolerance.

7.3.8 Data analysis.

Protein identification and quantification was carried out using ProQuant software v1.1 (Applied Biosystems; MDS-Sciex). The search was performed against a "mixed genome database" created in-house, including individual genomes of four representative micro-organisms downloaded from NCBI (June 2005): *Escherichia coli* K12, *Saccharomyces cerevisiae, Sulfolobus solfataricus P2* and *Synechocystis* sp., encoding 4237, 6298, 2995 and 3264 entries respectively, plus 222 keratins extracted from humans and mice [39]. The search parameters allowed for peptide and MS/MS tolerance were up to 0.15 Da and 0.1 Da respectively; one missed cleavage of trypsin; oxidation of methionine and cysteine modification of MMTS. Only peptides above 70% confidence were saved for identification and quantification. ProGroup Viewer software v1.0.6 was used to identify proteins with at least 95% confidence. The results obtained from ProGroup Viewer were exported to Microsoft Excel for further analysis.

7.4 Results and Discussion

7.4.1 Data quality

All proteins discussed in this study must have had at least 2 or more MS/MS (peptides) identified with a significant peak area (above 40 counts) and % error (derived from ProQuant software). The use of the Mixed Genome Database has been demonstrated previously to be very efficient in eliminating false positive identifications down to less than 1% in *Synechocystis* sp.

[39]. The criteria for data validation through multiple injections and replicate analysis, as well as the derivation of the protein quantitation ratio from an iTRAQ experiment were reported previously. [38, 39]. Briefly, all peptide ratios were converted into $\log_{10}$ space before compiling them into the calculation for the overall protein ratio and the estimation of the standard deviation (SD). The reference used in this context is label 114 (40mg/L of Pi) and the relative quantitative ratio used was estimated from 115:114 for 10% P-level, 116:114 for 3% P-level and 117:114 for 0.3% P-level. The full list of proteins identified in this study can be found in Appendix L.

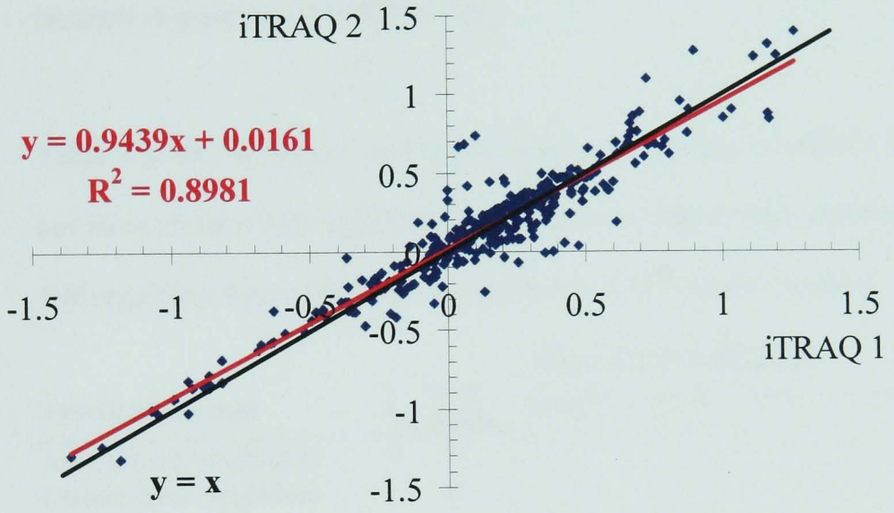


Figure 7.1. The theoretical relationship between an experimental replicate (y = x; black line) against the best-fit line (y = 0.9439x – 0.0161; red line) generated from the experimental replicates. The values plotted in the graph are in $\log_{10}$ space.

In order to assess the reproducibility of the iTRAQ analysis in this study, the iTRAQ experiment was repeated using the biological sample from all four phenotypes to form an experimental replicate [38]. The theoretical relationship between the experimental replicates is Y = X. The best-fit line estimated in this study has a slope of 0.94 and a R^2 value of 0.90, as shown in Figure 7.1. This finding further verifies the consistency of the iTRAQ experiments carried out in this study. To further strengthen the significance of a protein regulation, only those with proteins whose relative abundance increased above 1.5-fold (either increment or

decrement) with an acceptable error margin (SD < log ratio), were considered as significantly changed [38]. The error factor (EF) was used to estimate the variation of a protein ratio from its median value (in linear space). The upper limit of a protein regulation can be calculated by multiplying the protein ratio by the EF; while the lower limit is derived by dividing the protein ratio by the EF.

A total of 170 proteins (≥ 2 MS/MS) were confidently identified in this study (refer to Table 7.2). More than 22% of the total proteins were identified as hypothetical proteins with unknown function. This is followed by photosynthesis-related proteins at 10%, and amino acids metabolism proteins at 9% (Table 7.2).

Table 7.2. The functional distribution of the total number of proteins identified in this study, and those proteins with significant up- or down-regulation. Only proteins with more than ±1.50-fold regulation were considered as significantly differentially regulated.

		Down-regulated Proteins			Up-regulated Proteins		
Functional Groups	**Total proteins**	**10% Pi**	**3% Pi**	**0.3% Pi**	**10% Pi**	**3% Pi**	**0.3% Pi**
Amino Acids Metabolism	15	2	2	1	3	8	6
Carbohydrate Metabolism	6	0	0	0	3	2	2
CO2 Fixation	5	0	0	0	5	5	4
Energy Metabolism	4	1	0	0	0	2	3
Folding, Sorting, Degradation	10	0	0	0	5	9	9
Glycolysis	9	0	0	0	6	7	6
Hypothetical Proteins	38	8	6	6	18	22	25
Lipid Metabolism	6	2	1	0	1	3	3
Metabolism of Cofactors and Vitamins	1	0	0	0	0	0	1
Nucleotide Metabolism	4	1	1	1	2	1	2
Other	14	1	0	1	7	9	10
Pentose Phosphate Pathway	8	0	0	0	5	6	6
Photosynthesis	17	6	5	6	4	4	5
Porphyrin and Chlorophyll Metabolism	1	0	0	0	0	0	1
Replication and Repair	2	0	0	0	0	0	1
Signal Transduction	2	2	2	2	0	0	0
TCA Cycle	2	0	0	0	1	1	1
Transcription	3	1	1	0	0	2	2
Translation	9	0	0	0	2	5	5
Transport and Binding	11	5	5	7	1	3	3
Unknown Proteins	3	0	0	1	2	2	1
Total	**170**	**29**	**23**	**25**	**65**	**91**	**96**

Amongst these 170 proteins, at least 17%, 14% and 15% showed significant down-regulation (≥ -1.5-fold) in 10%, 3% and 0.3% Pi-deprived conditions compared to the Pi-replete condition respectively (Table 7.2). Interestingly the percentage of up-regulated proteins was significantly higher compared to down-regulated proteins. Approximately 38%, 53% and 56% of total proteins showed significant up-regulation (≥ 1.5-fold) in 10%, 3% and 0.3% P-level compared to the phosphate-replete condition respectively (Table 7.2). Among the up-regulated proteins, the majority of them were hypothetical proteins (26%), others (11%), folding, sorting and degradation proteins (9%) and glycolysis proteins (8%). On the other hand, the majority of the down-regulated proteins were hypothetical proteins (26%), transport and binding proteins (22%), photosynthesis proteins (22%), and signal transduction proteins (8%). From the overall picture here, phosphate reduction caused an increase in relative protein abundance for those proteins involved in glycolysis, and folding, sorting and degradation processes, whereas a decrease was observed for those involved in photosynthesis, transport and binding, and signal transduction systems.

7.4.2 Growth rate

In phosphate limited conditions, the growth rate of *Synechocystis* sp. PCC 6803 was significantly impaired, as illustrated in Figure 7.2.

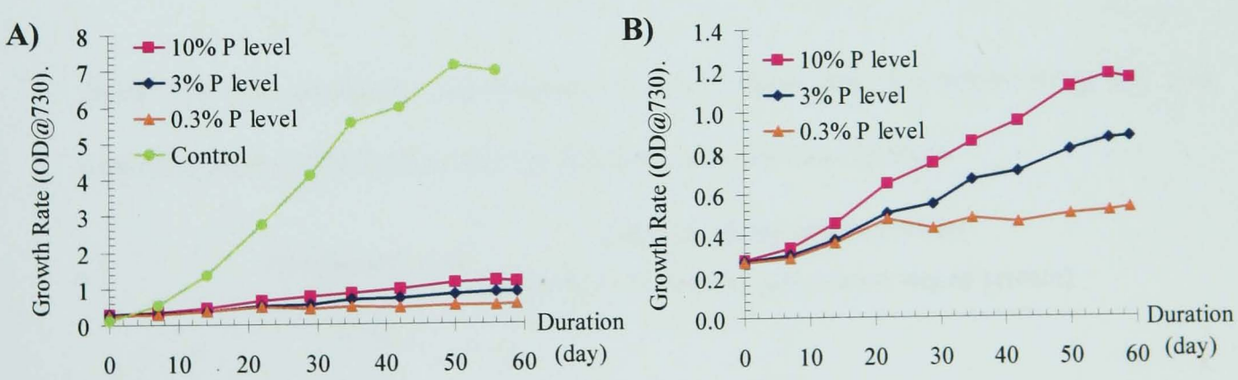


Figure 7.2. The growth rates of Pi-replete *Synechocystis* sp. control culture, and the phosphate-limited cultures: 4 mg/L (10%), 1.2 mg/L (3%) and 0.12 mg/L (0.3%) P-level at 25°C in light incubator.

The control culture (with 100% Pi) reached an OD_{730} of 7 at the end of the 60 days, whereas the phosphate-limited cultures could only achieve the maximum OD_{730} of 1.2 (for 10% P-level), 0.9 (for 3% P-level) and 0.5 (for 0.3% P-level). At 0.12 mg/L (0.3%) of K_2HPO_4, the phosphate-limited culture reached the stationery phase after 3 weeks of growth. This indicates that, at this level, the cell has reached the critical stage as the nutrient (Pi) available in the media was depleted. Unlike in 10% and 3% P-level, the cultures grew steadily, albeit at a lower rate compared to the control culture. Severe growth rate reductions were also observed in *C. glutamicum* [462], *Arabidopsis* [468] and *Sulfolobus acidocaldarius* [469].

7.4.3 Phosphatase response during phosphate limitation.

The most common observation during phosphate limitation is an increase in the activity of the periplasmic enzyme alkaline phosphatase [470]. No acid phosphatase activity was detected in all three phosphate-reduced conditions. The alkaline phosphatase assay, however, yielded interesting results. As the phosphate concentration reduced, the alkaline phosphatase activity increased, as shown in Table 7.3. The maximum activity was recorded as 60.7 nmoles of Pi released per min per mg of protein at 0.3% P-level. This finding corresponds well with alkaline phosphatase activity changes observed in other cyanobacteria such as *Plectonema boryanum* and *Aphanizomenon ovalisporum*, when subjected to long term phosphate limitation [471, 472].

Table 7.3. The phosphate concentration in the samples was determined using the acid molybdate method [467]. Only alkaline phosphatase activity was detected. [ϒ]

Phosphate level	**Alkaline phosphatase activity** (nmoles of Pi released per min per mg of protein)
100%	0
10%	10.2
3%	41.3
0.3%	60.7

[ϒ] Using p-nitrophenylphosphate substrate.

In *Synechocystis* sp. PCC 6803, a total of 3 putative alkaline phosphatase genes exist in the genome, and they are assigned as sll0222, sll0654 and slr0509. However, only the protein expression of the sll0654 was detected by the iTRAQ-mediated proteomic approach here in phosphate limited conditions. The relative protein expression of sll0654 was significantly up-regulated as the phosphate level reduced from 10% to 0.3% as illustrated in Figure 7.3. The effect of phosphate limitation at the 10% P-level (1.58-fold) was not as pronounced as at 3% and 0.3% conditions, where a 2-fold relative increase was observed for both. Interestingly, the increment in the phosphatase assay result was not the same as those obtained in the iTRAQ result. For example, a 400% increment in alkaline phosphatase activity was observed from 10% to 3% P-level in the phosphatase assay, whereas, in contrast, a 33% increase in relative protein abundance was obtained from 10% to 3% P-level using an iTRAQ analysis. This observation clearly indicates the possibility that the remaining unidentified alkaline phosphatases, sll0222 and slr0509, play an important role, as well as contributing most of the phosphatase activity compared to sll0654.

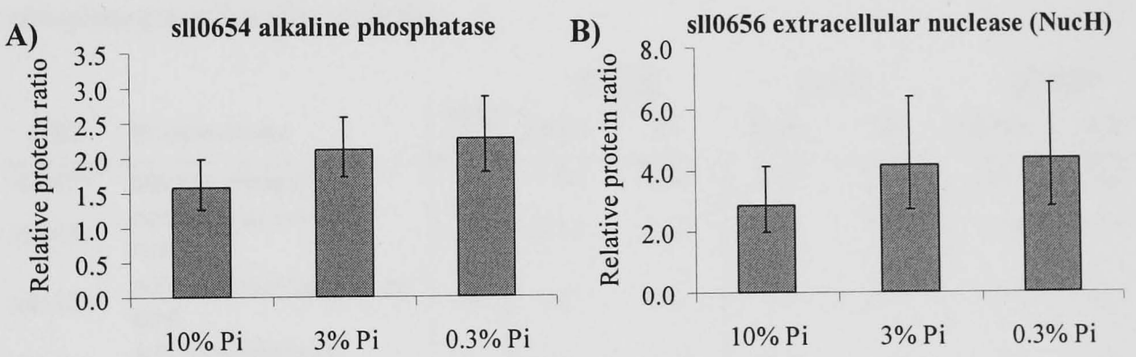


Figure 7.3. The relative protein expressions of extracellular nuclease (A) alkaline phosphatase, sll0654 and (B) NucH, sll0656. The y-axis is in linear space and the expression is relative to the phosphate-replete control.

However, Hirani *et al.* reported that an increase in transcript abundance, determined by northern blotting analysis during phosphate limitation was only observable for the putative alkaline phosphatase sll0654, but none for the other two alkaline phosphatases [250]. It is worth noting that the extracellular nuclease (*nuc*H, sll0656) gene is just located up-stream of the alkaline

phosphatase (sll0654) gene in the annotated sequence database of *Synechocystis* sp. [51]. Interestingly, the protein expression of NucH also showed significant up-regulation of more than 2.5-fold at all phosphate-reduced levels (Figure 7.3 and Table 7.4). It is thought that NucH is useful in scavenging Pi from NTP, and serves a role in liberating Pi from extracellular nucleic acids [473].

On the other hand, among the two identified subunits of ATP synthase in this study, AtpA (sll1326) and AtpB (slr1329) both exhibited up-regulation in relative protein expression by more than 2-fold at the 0.3% P-level compared to phosphate-replete control (Table 7.4). In contrast, the soluble inorganic pyrophosphatase (Ppa, slr1622) demonstrated more than 2-fold decreased relative protein abundance at the 10% P-level compared to the phosphate-replete control, which is in contrast with a previous report, whereby the Ppa abundance was found to increase in both the transcript and protein levels in phosphate deprived conditions [252].

Table 7.4. The proteins involved in either addition (synthase) or removal (phosphatase) of a phosphate group from the substrate.

			10% Pi		3% Pi		0.3% Pi	
ORF	**Protein Name**	**MS/ MS**	**Ratio**	**EF**	**Ratio**	**EF**	**Ratio**	**EF**
sll0654	alkaline phosphatase	7	1.58	1.26	2.12	1.22	2.27	1.26
sll0656	extracellular nuclease nucH	8	2.85	1.45	4.15	1.54	4.38	1.55
sll1326	ATP synthase alpha chain atpA	8	1.43	1.15	1.71	1.15	3.42	1.13
slr1329	ATP synthase beta subunit atpB	6	1.10	1.22	1.27	1.23	2.40	1.26
slr1622	soluble inorganic pyrophosphatase ppa	2	-2.22	1.15	-1.46	1.06	-1.39	1.02

7.4.4 Phosphate limitation induced general stress responses.

When cells encounter a stress environment, such as phosphate-limitation in this study, they tend to respond by inducing a range of general stress proteins available in the cellular system. In this case, the glutathione peroxidase system, which is essential for the removal of lipid hydroperoxides under normal and stress conditions, was induced [474, 475]. Among them, glutathione S-transferase (Gst, sll0067) and the glutathione peroxidase-like NADPH peroxidase

(Gpx2, slr1992) were identified as up-regulated during phosphate limitation (Table 7.5). However, another two of the glutathione peroxidase system proteins: glutathione-dependent peroxidase, type II peroxiredoxin (sll1621) and glutathione peroxidase-like NADPH peroxidase (Gpx1, slr1171) did not exhibit the same protein changes as Gst and Gpx2, instead they were down-regulated more than 1.5-fold at the 10% P-level against the phosphate-replete control (Table 7.5).

This is further supported by changes in relative abundance in an antioxidant protein (slr1198), also known as the 1-Cys type peroxiredoxin in anti-oxidative system protein [174], where significant up-regulation (2-fold) was recorded for all three phosphate-reduced levels (Table 7.5). Another anti-oxidative stress protein, superoxide dismutase (SodB, slr1516), which is well-known for its ability to reduce effects of superoxide radical-mediated oxidative stress [173], was also significantly up-regulated in protein expression by more than 2-fold at the 0.3% P-level compared to the phosphate-replete control (Table 7.5). Clearly, the phosphate reduction process has triggered the general stress defence system, as seen in SodB and in the glutathione peroxidase system.

Thioredoxin (Trx) is capable of reducing protein disulphides to cysteines [476, 477]. Interestingly, among the two thioredoxin proteins identified in this study, TrxA (slr0623) showed more than 2-fold up-regulation in all three phosphate limitation conditions, whereas the relative abundances of TrxB (slr1139) remained unchanged during low phosphate stress, as illustrated in Table 7.5. It has been reported previously that the translated protein of Trx is relatively stable compared to the transcript [476]. TrxA is characterised as m-type thioredoxin while TrxB as x-type thioredoxin, and they appear to have very little in common [477]. These different types of Trx in the same organism suggests that each of them have distinctive physiological roles, and possibly a different subset of target proteins [477]. Thus the difference in the protein expression (of TrxA and TrxB) noticed here might indicate the specificity in their binding targets. Therefore, given this background, it is unsurprising to observe that they respond differently to a phosphate limitation environment.

Table 7.5. The list of proteins characterised as stress-related, and their relative quantitation ratio.

ORF	Protein Name	MS/ MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
sll0067	glutathione S-transferase	3	2.14	1.17	3.54	1.17	4.69	1.28
sll1621	glutathion-dependent peroxidase, type II peroxiredoxin	8	-1.92	1.23	-1.79	1.26	-1.15	1.32
slr1171	glutathione peroxidase-like NADPH peroxidase,glutathione peroxidise gpx1	3	-1.63	1.03	1.00	1.02	1.29	1.05
slr1992	glutathione peroxidase-like NADPH peroxidase gpx2	9	1.09	1.36	2.10	1.38	3.09	1.42
slr1198	antioxidant protein	30	1.98	1.12	2.77	1.13	3.57	1.10
slr1516	Superoxide dismutase sodB	12	-1.12	1.11	1.45	1.13	2.40	1.17
slr0623	thioredoxin trxA	6	2.21	1.33	2.23	1.34	2.72	1.34
slr1139	thioredoxin trxB	2	1.07	1.16	1.26	1.21	1.29	1.21
sll0170	heat shock protein 70, dnaK2	11	1.64	1.26	1.99	1.26	2.49	1.31
sll0416	60 kDa chaperonin 2, groEL-2	12	1.46	1.22	2.36	1.19	3.11	1.18
sll0430	heat shock protein 90, htpG	2	1.43	1.11	3.09	1.45	3.84	1.79
slr2075	10kD chaperonin groES	7	1.49	1.19	1.85	1.25	2.71	1.23
slr2076	60kD chaperonin groEL1	18	1.56	1.11	3.12	1.11	4.13	1.11
slr0164	ATP-dependent Clp protease proteolytic subunit clpP4	2	2.04	1.12	5.15	1.07	7.51	1.05
slr0165	ATP-dependent Clp protease proteolytic subunit clpP3	2	1.13	1.62	2.18	1.80	4.12	1.96
sll1633	cell division protein ftsZ	3	1.01	1.31	1.36	1.34	1.91	1.32
sll1712	DNA binding protein HU	4	1.35	1.59	-1.13	1.59	2.35	1.59

A complementary reaction of increased expression at the protein level was also observed in heat shock proteins under these phosphate limited conditions. All of the identified heat shock proteins exhibited enhanced protein expression in low-phosphate conditions relative to the phosphate-replete control. The increment in their protein abundance changes corresponded well with a decrease in phosphate concentration, with increases of up to 4-fold seen, as summarised in Table 7.5. The heat shock proteins involved are: the 60kD chaperonin (GroEL1, slr2076), heat shock protein 90 (HtpG, sll0430), 60 kDa chaperonin 2 (GroEL2, sll0416), 10kD

chaperonin (GroES, slr2075), and heat shock protein 70 (DnaK2, sll0170). The involvement of these heat shock proteins was reported previously (in *Synechocystis* sp.) for light stress [171, 175], heat stress [180, 189] and salt stress [208, 209], yet none under phosphate limitation. It is probable that the cells respond toward the phosphate stress environment by regulating these heat shock proteins as a general stress response.

ATP-dependent Clp protease proteolytic subunits ClpP3 (slr0165) and ClpP4 (slr0164), although not being categorised as heat shock proteins, their functional role in degrading the irreversibly damaged proteins is especially important under stress conditions [478, 479]. These proteins were also among those highly up-regulated, showing more than 4-fold and 7-fold increases in relative abundance at the 0.3% P-level against the phosphate-replete control for ClpP3 and ClpP4 respectively (Table 7.5). It is also worth mentioned that statistically relevant effects of phosphate limitation were also noticed in proteins involved in cell division and replication processes, especially the cell division protein (FtsZ, sll1633), and the DNA binding protein (Hup, sll1712). These proteins were significantly up-regulated by 2-fold at 0.3% P-level compared to the phosphate-replete control as illustrated in Table 7.5.

7.4.5 Response of periplasmic proteins.

Periplasmic proteins are located between the plasma membrane and the outer membrane. These proteins are involved in numerous biochemical reactions including substrate-binding in uptake reactions, signal transduction, synthesis of peptidoglycan (cell wall), electron transport, extracellular nutrient acquisition and detoxification [480]. The effect of phosphate deprivation significantly impacted on these proteins with unknown functions, especially on ORF slr2144, where an up-regulation of more than 14-fold was detected in 0.3% P-level compared to the phosphate-replete condition (Table 7.6). sll1837 also showed significant relative up-regulation of more than 3-fold in 3% phosphate, whereas sll1380 showed down-regulation of -1.8-fold at the 0.3% P-level compared to the phosphate-replete control (Table 7.6). A similar association of these proteins was observed in high salt environments [205], with the involvement of these proteins during phosphate limitation not previously reported. Therefore, the significant changes

in the protein expression identified here indicate a potential generic role in the cellular stress response system.

Among the annotated periplasmic proteins, sll0680 and slr1247 both have been annotated as phosphate substrate-binding periplasmic proteins, encoded by the *pst*S1 and *pst*S2 gene respectively, also showed changes in their relative protein expression concentration. Interestingly, their relative protein expression levels differed from each other under phosphate reduction conditions, with PstS2 (slr1247) showing slight up-regulation of more than 1.7-fold at the 3% and 0.3% phosphate levels, whereas PstS1 (sll0680) was 1.9-fold down-regulated in 10% and 0.3% phosphate compared to the phosphate-replete control (Figure 7.4).

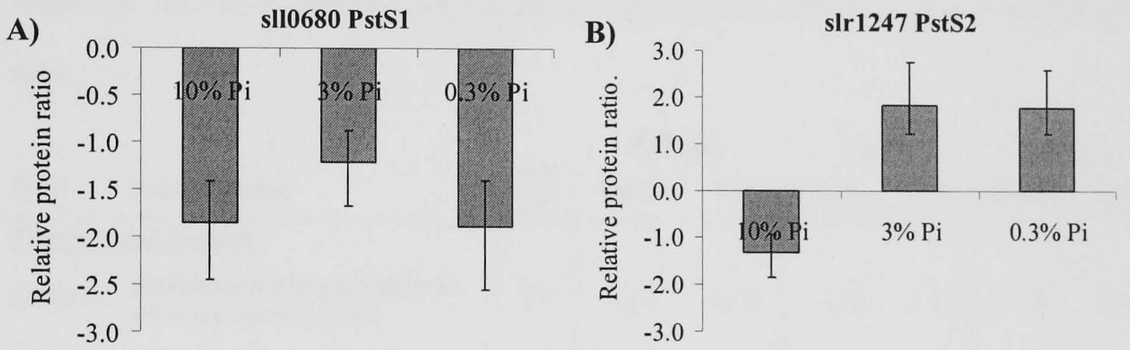


Figure 7.4. The relative protein expression of two PstS proteins: (A) sll0680, PstS1 and (B) slr1247, PstS2. The y-axis illustrated above is in linear space and the expression is relative to the phosphate-replete control.

It was reported previously that the induction of *pst*1 genes (such as PstS1) occurred as fast as 20 min following exposure to the phosphate-depleted condition, while the response of *pst*2 genes (such as PstS2) were relatively slower, reached maximum transcript level after 8 hr of phosphate limitation [253]. The relative transcript expression level of *pst*S2 was 6-fold higher than that for *pst*S1 after 8 hr of exposure to the phosphate limited environment [253]. This corresponds well with the observation here, where the relative protein abundance of PstS2 was at least 3 times higher than PstS1 at the 0.3% P-level.

A similar increase in the transcript relative abundance of the *pst*S gene was also observed in the cyanobacterium *Synechococcus* sp. WH7803 undergoing phosphate limited growth [481]. Further evidence of the involvement of *pst*S genes in phosphate reduction response was also observed in *B. subtilis* and *Streptomyces coelicolor* [482, 483]. Unfortunately, in this study, none of the phosphate transport regulatory system (Pho) proteins were identified through the iTRAQ-mediated proteomic approach; therefore it is not possible to discuss this further at this stage. Other annotated periplasmic proteins, such as D-alanyl-D-alanine carboxypeptidase (slr1924) and FKBP-type peptidyl-prolyl cis-trans isomerase (YtfC, slr1761), were also among those found to be up-regulated under phosphate limitation conditions, as summarised in Table 7.6.

Table 7.6. The list of identified periplasmic-annotated proteins, and their relative quantitation ratio.

			10% Pi		3% Pi		0.3% Pi	
ORF	**Protein Name**	**MS/ MS**	**Ratio**	**EF**	**Ratio**	**EF**	**Ratio**	**EF**
Transport and Binding								
sll0680	phosphate-binding periplasmic protein precursor pstS1	13	-1.86	1.32	-1.22	1.38	-1.89	1.35
slr1247	phosphate-binding periplasmic protein precursor pstS2	7	-1.33	1.40	1.82	1.51	1.77	1.46
sll1450	nitrate/nitrite transport system substrate-binding protein nrtA	4	-1.23	1.10	-1.54	1.09	-2.00	1.17
slr0447	ABC-type urea transport system substrate-binding protein urtA	9	-2.29	1.15	-1.83	1.15	-1.76	1.15
sll1009	iron-regulated protein frpC	2	2.81	1.00	3.28	1.01	2.84	1.05
slr0513	iron transport system substrate-binding protein, futA2	10	-2.23	1.37	-2.17	1.40	-7.71	1.43
slr1295	iron transport system substrate-binding protein futA1	5	-2.33	1.23	-4.30	1.27	-6.67	1.24
slr1452	sulfate transport system substrate-binding protein sbpA	10	-11.86	1.17	-11.34	1.16	-9.89	1.18
ssr2857	mercuric transport protein periplasmic component precursor atx1	2	4.00	1.22	5.77	1.19	7.44	1.25
sll1341	bacterioferritin brfA	6	1.02	1.08	1.19	1.12	-1.14	1.11

ORF	Protein Name	MS/ MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
slr1890	bacterioferritin brfB	4	1.21	1.24	-1.10	1.27	-2.30	1.30
Other Periplasmic Proteins								
slr1761	FKBP-type peptidyl-prolyl cis-trans isomerase, periplasmic protein	2	2.20	2.14	2.55	2.19	2.69	2.08
slr1924	D-alanyl-D-alanine carboxypeptidase, periplasmic protein	2	1.75	1.16	2.44	1.04	-1.36	1.17
sll1380	periplasmic protein, function unknown	2	-1.26	1.13	-1.26	1.11	-1.80	1.05
sll1837	periplasmic protein, function unknown	2	1.85	1.23	3.18	1.02	-1.05	1.24
slr2144	periplasmic protein, function unknown	3	4.69	1.60	7.02	1.65	14.72	1.62

In addition to phosphate substrate-binding periplasmic proteins that were directly affected by the phosphate limited growth environment, other substrate-binding proteins also responded to this change in environment. In particular, proteins involved in the iron transport system, where iron-binding proteins, FutA1 (slr1295) and FutA2 (slr0513) were found to be significantly decreased in their relative protein expression levels as the phosphate concentration reduced from 10% to 0.3% (Table 7.6). In these limited conditions, a decrease of greater than 5-fold and 7-fold in protein relative abundance was observed in FutA1 and FutA2 respectively in 0.3% phosphate compared to the phosphate-replete control. It has been reported that FutA1 and FutA2 play a major role in iron acquisition, since the inactivation of these genes can greatly reduce the ferric iron uptake activity [244].

Bacterioferritin (BrfB, slr1890) is thought to be responsible for as much as 50% of the cellular iron storage in cyanobacteria [247]. In this study, this protein was also found to be down-regulated by near 2-fold in 0.3% phosphate compared to the phosphate-replete control, as depicted in Table 7.6. Another bacterioferritin protein (BrfA, sll1341), however, did not change its relative abundance during phosphate deprivation. This is not surprising, since Bertani *et al.* (1997) found two types of subunits in bacterioferritin from *Magnetospirillum magnetotacticum* that closely resemble these from *Synechocystis* sp., with differences in heme- and metal-binding sites and ferrooxidase activities [484]. Nonetheless, the iron-regulated protein (FrpC, sll1009)

showed more than 2-fold up-regulation in all three phosphate-reduced levels compared to phosphate-replete control, indicating that phosphate deprivation not only caused phosphate limitation, but also iron limitation. With regard to other metal transporters, mercuric transport protein periplasmic component precursor (Atx1, ssr2857), which is also known as copper metallochaperone [485], was up-regulated during the phosphate limitation conditions.

In terms of anion transporters such as the sulfate, nitrate/nitrite and urea transport systems, a number were down-regulated. For example, the sulfate-binding protein (SbpA, slr1452), nitrate-binding protein (NrtA, sll1450) and ABC-type urea-binding protein (UrtA, slr0447) were all down-regulated more than 2-fold during phosphate reduction conditions (Table 7.6). These transport proteins were seemingly affected by the lower growth rate (noticed in Figure 7.2), and consequently a lower metabolic rate. Under phosphate limitation conditions, the cells recycle the cellular component rather than transport new molecules from the extracellular compartment [486], as seen from the general down-regulation of these transport proteins.

7.4.6 Signal transduction under phosphate limitation.

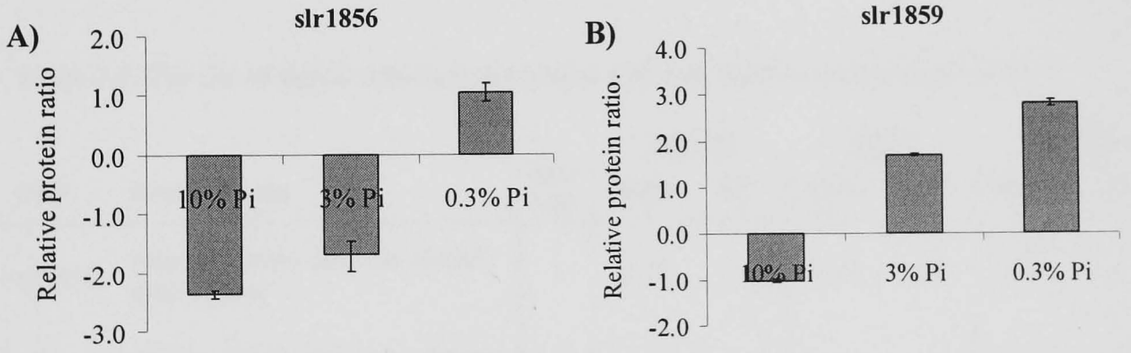


Figure 7.5. The relative protein expression of phosphoprotein substrate (A) slr1856 and (B) slr1859 surrounding the *icf*G gene cluster. The y-axis illustrated above is in linear space and the expression is relative to the phosphate-replete control.

It is predicted that a set of ORFs clustered around the *icf*G gene potentially encode signal transduction proteins resembling the SpoIIE, RsbU, and RsbX protein phosphatases in *B. subtilis* participate in coordination of inorganic carbon and glucose metabolism in *Synechocystis*

sp. [487, 488]. Unfortunately, IcfG (slr1860) was not detected in this study, however the phosphoprotein substrates bracketing the *icf*G gene were identified, and they are slr1856 and slr1859. Both proteins are thought to share a close similarity with the respective phosphoprotein substrates for the *B. subtilis* protein kinases: SpoIIAA, RsbS, and RsbV [488]. The changes in relative protein expression between slr1856 (2-fold down-regulation in 10% P-level) and slr1859 (2-fold up-regulation in 0.3% P-level), however, disagreed quantitatively, but qualitatively their expression directions showed a similar trend; as the phosphate level reduces, the relative expression of both proteins increase as illustrated in Figure 7.5.

Among the proteins involved in signal transduction and the two-component system, the nitrogen regulatory protein P-II (GlnB, ssl0707), and the two-component sensor histidine kinase Hik20 (sll1590) were also significantly affected by the phosphate limited conditions (Table 7.7). Both of these proteins showed significant down-regulation of more than 2-fold. Furthermore, the change in protein expression for GlnB might be a cascaded effect from the down-regulation of NrtA detailed above, since the extra-cellular nitrogen source will need to bind to the substrate protein (such as NrtA) before they can travel through the periplasmic space.

Table 7.7. The list of signal transduction proteins and their relative quantitation ratios.

			10% Pi		3% Pi		0.3% Pi	
ORF	**Protein Name**	**MS/ MS**	**Ratio**	**EF**	**Ratio**	**EF**	**Ratio**	**EF**
slr1856	phosphoprotein substrate of icfG gene cluster	3	-2.38	1.03	-1.71	1.17	1.04	1.16
slr1859	phosphoprotein substrate of icfG gene cluster	4	-1.02	1.04	1.71	1.02	2.81	1.03
sll0271	N utilization substance protein B homolog	2	1.06	1.10	1.54	1.34	2.31	1.33
ssl0707	nitrogen regulatory protein P-II glnB	14	-2.45	1.23	-2.86	1.24	-2.83	1.21
sll1590	two-component sensor histidine kinase hik20	3	-4.38	1.12	-7.33	1.05	-8.80	1.13

In contrast, a 2.3-fold up-regulation in protein expression was noticed for the N utilization substance protein (NusB, sll0271), as tabulated in Table 7.7. As the amount of nitrogen sources

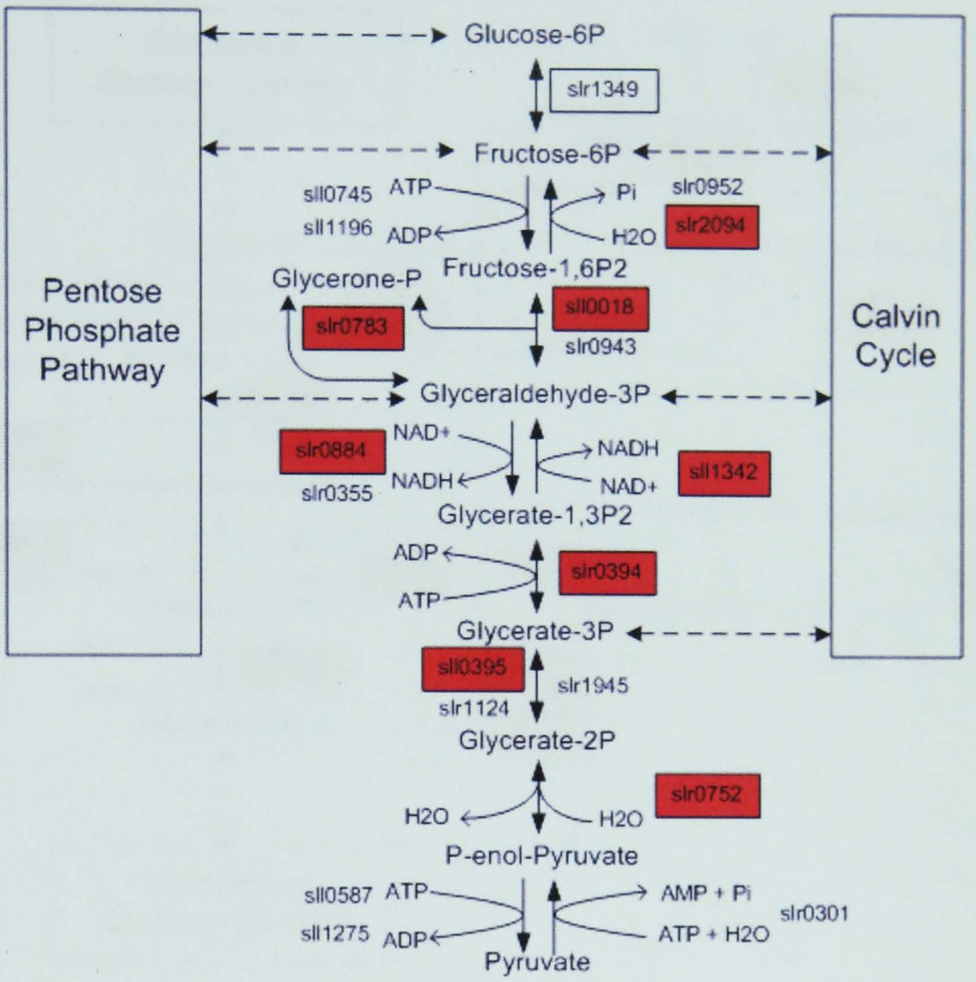
transported through the membrane reduces, the cell counteracts this by increasing the NusB concentration. This is also the first study that shows a direct influence of phosphate deprivation on the Hik20 protein (showing 4- to 7-fold down-regulation). Previously many other Hik proteins were found to be involved in various other conditions such as Hik33 in osmotic and cold stresses [489, 490]; Hik34 in regulating heat shock genes [187]; Hik31 in glucose sensing [491]; and Hik16, Hik33, Hik34, and Hik41 in transducing salt stress signals [206].

7.4.7 Response of central carbon metabolism.

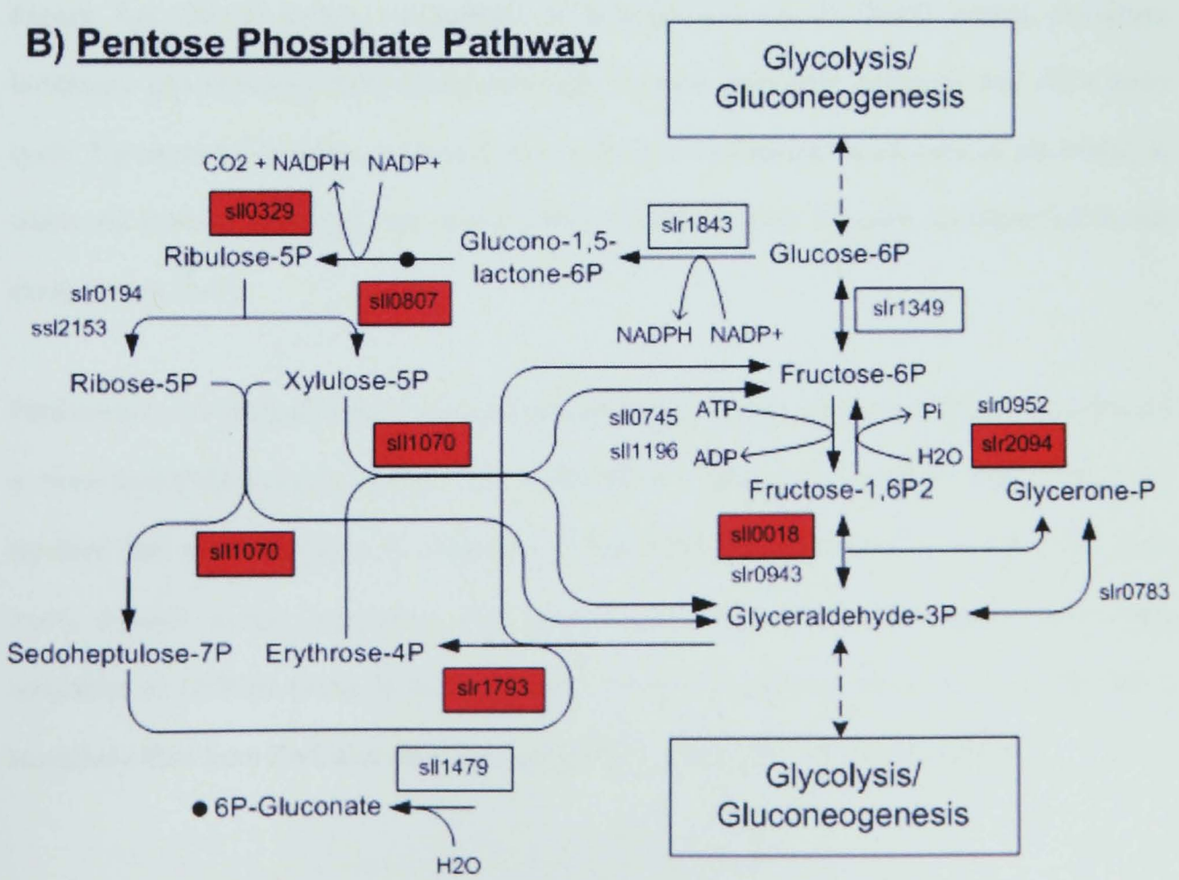
Central carbon metabolism, consisting of glycolysis/gluconeogenesis, the pentose phosphate pathway and Calvin cycle are crucial for carbon and energy derivation (Figure 7.6). The impact of phosphate limitation is significant in this system. The processes where glucose breaks down (in glycolysis), as well as the regeneration of NADPH (in the oxidative pentose phosphate pathway) form a series of phosphorylation and de-phosphorylation reactions. As a consequence, an overall up-regulation was observed in the central carbon metabolism, as measured using the iTRAQ-mediated proteomic approach. Among the identified proteins, only glucose-6-phosphate isomerase (Pgi, slr1349), 6-phosphogluconolactonase (Pgl, sll1479), and glucose 6-phosphate dehydrogenase (Zwf, slr1843), did not show any sign of significant up-regulation (>1.5-fold) in either one of the phosphate-reduced environments. The full metabolic map reconstruction from central metabolism with reference to KEGG (http://www.genome.jp/KEGG) is illustrated in Figure 7.6, and the relative quantitation ratio of these proteins is tabulated in Table 7.8.

The formation of acetyl-CoA involving dihydrolipoamide dehydrogenase (LpdA, slr1096) requires NAD^+ or $NADP^+$ to drive the reaction forward, and indeed this protein (LpdA) increased its relative expression from 2-fold to 4-fold as the phosphate level reduced from 10% to 0.3%, indicating the close relationship of this protein to the abundance of the inorganic phosphate (Pi) in the cellular environment. On the other hand, glucose 6-phosphate dehydrogenase (Zwf, slr1843) though involved in the formation of NADPH from $NADP^+$, showed no significant change in relative protein expression during phosphate limitation.

A) Glycolysis/Gluconeogenesis



B) Pentose Phosphate Pathway



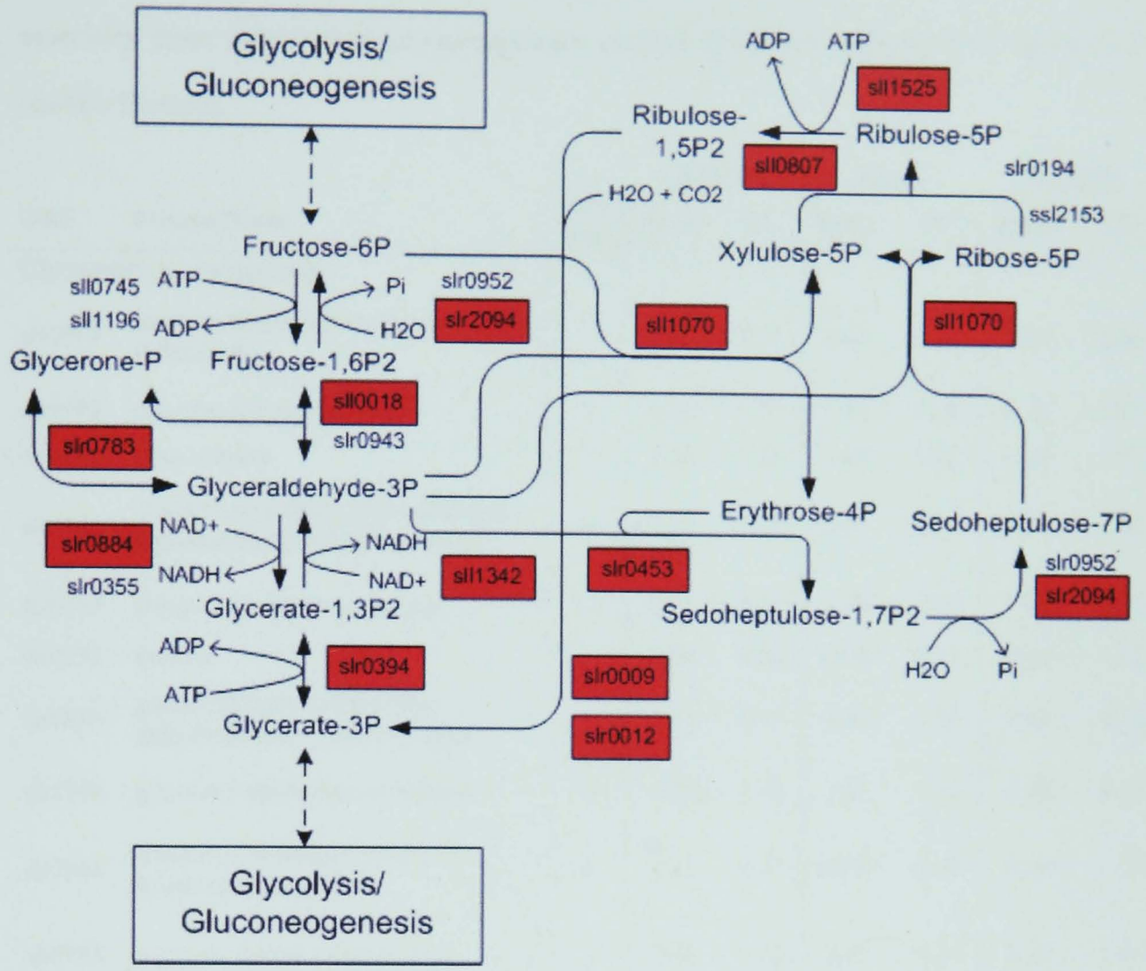


Figure 7.6. Central carbon metabolism of *Synechocystis* sp. PCC6803 during phosphate limitation: (A) Glycolysis/gluconeogenesis; (B) Pentose phosphate pathway; and (C) Calvin cycle. Up-regulated proteins are boxed in red. Proteins unchanged in abundance are boxed in white. All proteins identified here had at least 2 MS/MS spectra. Proteins not identified in this study are not boxed.

Furthermore, the glucose 6-phosphate dehydrogenase assembly protein (OpcA, slr1734) showed a more than 6-fold up-regulation in 0.3% phosphate (Table 7.8). However, Sundaram *et al.* reported that no reduction in the abundance of G6PDH monomers (Zwf) was observed in an *opc*A deletion mutant, suggesting that *opc*A is only involved in the oligomerization and activation of G6PDH [419]. It is likely that 6-phosphogluconolactonase Pgl (sll1479) had a cascaded effect from Zwf, also showing no significant change in protein expression.

Table 7.8. The relative expression of the proteins involved in central carbon metabolism, especially those in glycolysis/gluconeogenesis, pentose phosphate pathway and Calvin cycle (carbon fixation).

ORF	Protein Name	MS/ MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
Glycolysis/gluconeogenesis								
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	2	2.13	1.26	4.84	1.58	3.30	2.08
sll0395	phosphoglycerate mutase	2	-1.01	1.27	1.60	1.28	2.22	1.57
sll0593	Glucokinase	2	1.10	1.50	1.07	1.41	-1.01	1.33
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	9	1.63	1.26	2.21	1.31	2.12	1.32
slr0394	phosphoglycerate kinase pgk	11	2.20	1.18	1.50	1.18	1.07	1.24
slr0752	enolase	5	6.40	1.50	14.89	1.51	14.81	1.52
slr0884	glyceraldehyde 3-phosphate dehydrogenase 1 (NAD+) gap1	5	1.71	1.13	4.53	1.14	4.69	1.15
slr1349	glucose-6-phosphate isomerase	8	-1.08	1.18	1.21	1.15	1.38	1.15
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	8	2.41	1.33	2.36	1.34	1.99	1.36
slr0783	triosephosphate isomerase tpi	2	3.78	1.15	3.39	1.14	3.12	1.04
slr1096	dihydrolipoamide dehydrogenase	2	2.70	1.52	3.17	1.61	4.17	1.62
Carbohydrate Metabolism								
sll0158	1,4-alpha-glucan branching enzyme glgB	2	1.52	1.03	1.48	1.19	1.27	1.12
sll1356	glycogen phosphorylase	5	2.20	1.39	2.29	1.35	1.75	1.45
slr1176	glucose-1-phosphate adenylyltransferase	3	-1.08	1.17	1.22	1.22	1.43	1.36
slr1994	PHA-specific acetoacetyl-CoA reductase phaB	3	1.22	1.16	2.10	1.40	2.97	1.26
Pentose Phosphate Pathway								
sll0329	6-phosphogluconate dehydrogenase	12	1.50	1.16	2.30	1.21	2.20	1.21
sll0807	pentose-5-phosphate-3-epimerase rpe	5	4.83	1.17	4.67	1.16	5.47	1.17
sll1070	transketolase	25	2.04	1.27	2.15	1.30	2.12	1.29
sll1479	6-phosphogluconolactonase	4	-1.12	1.24	1.11	1.19	1.39	1.27
slr0453	putative phosphoketolase	10	1.41	1.13	2.32	1.11	2.12	1.12
slr1793	transaldolase	5	3.37	1.22	5.69	1.25	6.20	1.24
slr1734	glucose 6-phosphate dehydrogenase assembly protein opcA	2	3.16	1.34	5.79	1.30	6.61	1.37

ORF	Protein Name	MS/ MS	10% Pi Ratio	10% Pi EF	3% Pi Ratio	3% Pi EF	0.3% Pi Ratio	0.3% Pi EF
slr1843	glucose 6-phosphate dehydrogenase zwf	3	-1.45	1.10	-1.16	1.10	-1.21	1.09
CO2 Fixation								
sll1029	carbon dioxide concentrating mechanism protein CcmK ccmK1	5	1.80	1.39	1.61	1.42	2.05	1.42
sll1525	phosphoribulokinase prk	12	1.54	1.18	1.59	1.16	1.42	1.19
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	36	2.81	1.16	3.20	1.19	3.38	1.14
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	8	1.54	1.30	1.70	1.25	1.71	1.28

It has been suggested that Polyhydroxyalkanoate (PHA) serves as an extra carbon reserve in the form of intracellular granules [107]. Interestingly, PHA-specific acetoacetyl-CoA reductase (PhaB, slr1994) was 2-fold and 3-fold up-regulated in 3% and 0.3% relative phosphate compared to the control phosphate-replete culture. This is in agreement with the suggestion that phosphate deprivation enhanced the PHA accumulation in *Synechocystis* sp. PCC 6803 [9, 492].

7.4.8 Effect of phosphate limitation on photosynthesis apparatus.

The photosynthesis system is among one of the most significantly affected systems under phosphate limited conditions. Sixteen from nineteen of the identified photosynthesis-related proteins showed some statistically relevant change in their expression as a consequence of the phosphate stress. Photosystem I proteins, PsaD (slr0737) and PsaA (slr1834) showed significant down-regulation in protein expression, whereas Photosystem II proteins, PsbV and PsbU showed significant up-regulation during phosphate limitation, as depicted in Table 7.9. This is not the first time these two sub-groups in the photosynthesis system showed contrasting behaviour, as seen in results described in Chapters 5 and 6. From previous chapters, it was proposed that Photosystems I and II have different sets of modulation control systems, which seems important in regulating the stoichiometric ratio of the Photosystem content in the cell. Once again, the differences in protein expressions observed here during phosphate deprivation provide further evidence to support this hypothesis. It is unclear whether the phosphate

limitation has the same effects on the Photosystem I and II proteins as those conditions discussed in Chapter 5 and 6 (i.e. light stress condition). However, it is undeniable that phosphate limitation significantly reduced the photosynthesis capacity, as observed from the difference in pigment colour (Figure 7.7) as well as the growth rate (Figure 7.2) at the end of the growth cycle.

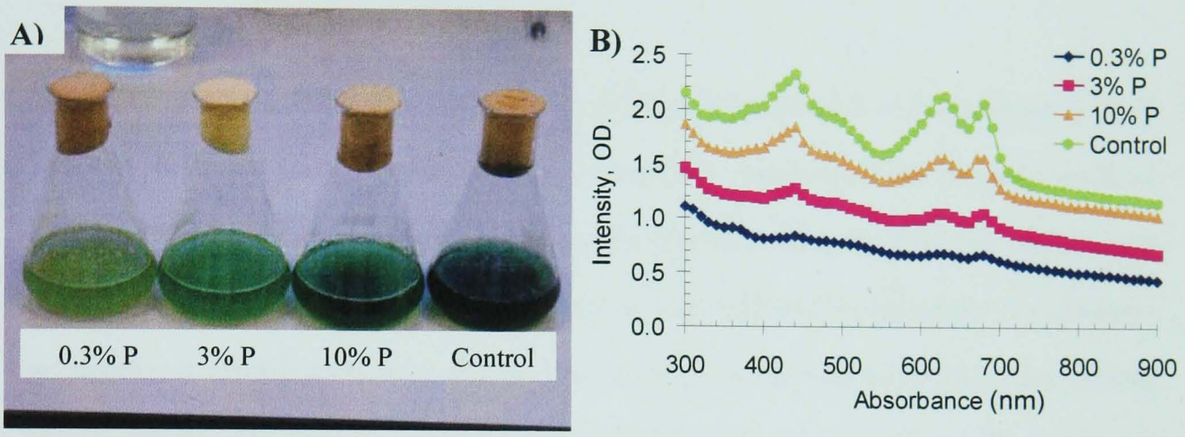


Figure 7.7. (A) *Synechocystis* sp. cultures at the day 60 of the growth cycle in (from left to right) 0.3%, 3%, 10% and 100% (normal BG-11) phosphate media. (B) The wavelength scan (300 nm-900 nm) of the corresponding cultures.

The colour of the *Synechocystis* sp. cells at the end of the 60 day growth period revealed that phosphate reduction has a dramatic impact on the physiological features of the culture. The pigment colour (phycocyanin, allophycocyanin and chlorophyll) changed from a dark green colour to yellowish green colour as the relative phosphate level reduced from 100% to 0.3% (Figure 7.7). This observation is further complemented with a wavelength scan of all phosphate-reduced cultures, from 300nm to 900 nm, which is intended to capture most of the phycobilisomes and chlorophyll contents. Chlorophyll absorb light in the 500-600nm spectral region [58]. This is especially visible in the wavelength scans of the control culture as seen by the two large peaks within that region (Figure 7.7B). One particularly prominent peak is located at approximately 450nm region in Figure 7.7B, and this peak corresponds to the light absorption region (450-650nm) of the antenna (phycobilisomes) proteins [58]. As the phosphate level decreased, the peak intensity reduced, eventually flattening out when the phosphate level

reached 0.3%. This suggests that phosphate-deprived cells are severely limited in the growth (as seen in Figure 7.2) as a consequence of reduced capability in light absorption. Similar observation also noticed in a separate study of phosphate limitation response in *Arabidopsis* using cDNA microarray, where phosphate-deprived cells showed decrease in the transcript level of 29 genes, which were associated with Photosystem I, Photosystem II and chlorophyll binding mechanisms [493].

An overall down-regulation was noted for the phycobilisome proteins as summarised in Table 7.9. This is in general agreement with the 'colour-loss' observed in the phosphate-reduced cultures. There are a few proteins not following this general trend, such as the phycocyanin beta subunit (CpcB, sll1577), which was up-regulated in both 10% and 3% phosphate. Furthermore, no significant relative expression change was observed in the allophycocyanin beta subunit (ApcB, slr1986) and phycobilisome small core linker polypeptide (ApcC, ssr3383) as seen in Table 7.9. In contrast to the relative protein expression changes observed for the phycobilisomes proteins, ferredoxin-NADP oxidoreductase (PetH, slr1643) and plastocyanin (PetE, sll0199) showed enhanced protein expression at all phosphate-reduced levels, with the most significant fold change of more than 5-fold recorded in the 0.3% phosphate environment (Table 7.9). This group of proteins (Pet) is involved in the electron transport chain reaction between the cytochrome b_6f complex from photosystem II, and $P700^+$ from photosystem I [421, 422]. It had been recently reported that PetH and some of the linker proteins in the phycobilisome complex are phosphorylated during exposure to light, and this observation was instrumental in the regulation of assembly or disassembly of phycobilisomes [494].

It is also worth mentioning that the increase in electron transport capacity in cytochrome b_6f complex (as seen in PetH and PetE) also affected the photolysis process. The bidirectional hydrogenase large subunit (HoxH, sll1226) showed more than 2-fold up-regulation in 3% and 0.3% phosphate as tabulated in Table 7.9. An increase in hydrogen production derived from the *hox* operon under sulphur deprivation conditions had been previously reported [2]. However, it

is not known if the actual hydrogen production rate during phosphate limitation would be any higher than the phosphate-replete control, since no measurements were taken.

Table 7.9. The relative protein expression of photosynthesis-related proteins.

ORF	Protein Name	MS/MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
slr0737	photosystem I subunit II psaD	3	-4.78	1.10	-8.44	1.11	-7.40	1.20
slr1834	P700 apoprotein subunit Ia psaA	2	-7.38	1.01	-17.94	1.23	-23.17	1.25
sll0258	cytochrome c550 psbV	4	3.39	2.04	4.00	2.03	3.70	1.97
sll0427	photosystem II manganese-stabilizing polypeptide psbO	7	-1.37	1.17	-1.24	1.16	-1.11	1.13
sll1194	photosystem II 12 kDa extrinsic protein psbU	12	1.22	1.23	1.37	1.19	1.52	1.20
sll1580	phycobilisome rod linker polypeptide cpcC1	5	-2.35	1.16	-4.31	1.46	-3.12	1.43
sll1577	phycocyanin beta subunit cpcB	126	2.63	1.09	2.12	1.09	1.38	1.08
sll1578	phycocyanin alpha subunit cpcA	68	1.20	1.20	-1.19	1.19	-1.52	1.18
slr2051	phycobilisome rod-core linker polypeptide cpcG1	12	-4.90	1.35	-3.91	1.43	-1.10	1.46
sll0928	allophycocyanin-B apcD	7	-1.50	1.12	-1.40	1.18	-1.10	1.17
slr0335	phycobilisome core-membrane linker polypeptide apcE	4	-8.75	1.83	-15.43	1.94	-6.63	1.36
slr1986	allophycocyanin beta subunit apcB	36	1.27	1.17	1.34	1.16	-1.14	1.19
slr2067	allophycocyanin alpha subunit apcA	9	-1.17	1.19	-1.45	1.29	-1.63	1.27
ssr3383	phycobilisome small core linker polypeptide apcC	2	1.12	1.04	1.45	1.07	1.15	1.14
sll0199	plastocyanin petE	15	2.27	1.19	4.11	1.25	5.00	1.30
slr1643	ferredoxin-NADP oxidoreductase petH	3	2.99	1.47	4.08	1.28	7.44	1.41
sll1663	phycocyanin alpha phycocyanobilin lyase related protein	3	-1.04	1.08	1.13	1.05	1.51	1.02
slr1368	precorrin decarbocylase cobL	2	1.02	1.15	1.47	1.13	1.94	1.30
sll1226	hydrogenase subunit of the bidirectional hydrogenase hoxH	2	1.34	1.12	2.39	1.11	2.04	1.13

7.4.9 Effect on protein biosynthesis.

The initiation, elongation and termination processes of a protein involves many types of molecules, including mRNAs, tRNAs, elongation factors and ribosomes [58]. In this study, two of the 50S ribosomal proteins Rpl12 (sll1746) and Rpl29 (ssl3436) showed enhanced relative expressions of more than 15-fold and 5-fold respectively, as the phosphate-level reduced from 10% to 0.3% compared to the phosphate-replete control (Table 7.10). Both proteins had been previously identified in Chapters 5 and 6, with Rpl29 illustrating up-regulation in the light compared to dark, whilst Rpl12 remained unchanged in relative abundance across the light:dark cycle. The significant up-regulation in protein expressions of these sets of proteins suggests the possibility of their involvement in a stress response system, responsible for the enhancement of the protein biosynthesis process when the cell encounters a stress environment, in particular to phosphate deprivation in this context. In a separate study employing cDNA microarray analysis, *rpl2*, *rpl3*, *rpl4*, and *rpl23* genes were found to be strongly enhanced by salt stress, which suggested the stress environment had a knock-on effects on ribosomes, and *de novo* protein biosynthesis might be necessary to maintain the activities of the ribosomes [190].

Table 7.10. Translation related proteins, and their corresponding quantitation ratio.

			10% Pi		3% Pi		0.3% Pi	
ORF	**Protein Name**	**MS/ MS**	**Ratio**	**EF**	**Ratio**	**EF**	**Ratio**	**EF**
sll1099	elongation factor Tu tufA	19	1.13	1.16	1.33	1.20	1.27	1.20
sll1261	elongation factor TS tsf	2	2.14	1.51	2.99	1.39	5.35	1.24
slr0434	elongation factor P efp	4	-1.46	1.08	-1.25	1.07	1.19	1.11
slr1105	elongation factor EF-G fus	2	-1.12	1.51	-1.04	1.41	1.39	1.41
slr1463	elongation factor EF-G fus	2	1.44	1.36	1.46	1.37	1.76	1.48
ssl3436	50S ribosomal protein L29 rpl29	2	1.19	1.14	1.88	1.12	5.58	1.02
sll1746	50S ribosomal protein L12 rpl12	4	1.90	1.55	4.58	1.58	15.81	1.56
sll1425	proline-tRNA ligase proS	2	-1.13	1.47	1.53	1.16	1.82	1.26

Furthermore, this finding is in good agreement with the up-regulation of elongation factor proteins identified in this study, where EF-TS (sll1261) and EF-G (slr1463) showed more than 5-fold and 1.7-fold increases in relative abundance in a 0.3% phosphate environment compared

to the phosphate-replete control respectively. Other elongation factor proteins (sll1099, slr1105 and slr0434) also did not change their relative expressions significantly when shifted from normal to phosphate-deprived medium, as shown in Table 7.10. Only one of the tRNA protein was identified here, a proline-tRNA ligase, (ProS, sll1425). This protein exhibited comparable behaviour to other translation proteins identified in this study, by showing a 1.8-fold up-regulation in the 10% to 0.3% phosphate environments.

7.4.10 Effects on amino acid metabolism.

Phosphate limitation triggered a range of mixed responses on the amino acid biosynthetic apparatus. For example, in isoleucine and valine metabolism (where there was a high proteome coverage) an overall up-regulation was observed (Table 7.11). Other amino acid metabolic proteins also showed significant up-regulation, including glutamate dehydrogenase (GdhA, slr0710), phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (HisA, slr0652), adenosylhomocysteinase (AhcY, sll1234) and glutamate-ammonia ligase (GlnA, slr1756). In contrast, alanine dehydrogenase (sll1682) urease alpha subunit (UreC, sll1750) and dihydrodipicolinate reductase (DapB, sll1058) were down-regulated (Table 7.11), suggesting a more minor role for these amino acids.

Table 7.11 Relative expressions of proteins involved in amino acids metabolism.

			10% Pi		3% Pi		0.3% Pi	
ORF	**Protein Name**	**MS/ MS**	**Ratio**	**EF**	**Ratio**	**EF**	**Ratio**	**EF**
Alanine Metabolism								
sll1682	alanine dehydrogenase	2	-2.04	1.14	-1.57	1.02	-1.40	1.09
Cysteine Metabolism								
slr0676	adenylylsulfate kinase cysC	4	-1.09	1.23	1.35	1.19	1.15	1.26
Glutamate metabolism								
slr0710	glutamate dehydrogenase (NADP+) gdhA	3	1.58	1.21	1.60	1.16	1.33	1.42
slr1756	glutamate--ammonia ligase glnA	8	2.44	1.28	2.60	1.39	2.12	1.39
slr0370	succinate-semialdehyde dehydrogenase (NADP+) gabD	3	1.18	1.82	1.49	1.98	1.19	2.24

ORF	Protein Name	MS/ MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
Histidine Metabolism								
sll1958	histidinol phosphate aminotransferase hisC	2	-1.35	1.01	-1.16	1.08	-1.14	1.07
slr0652	phosphorybosilformimino-5-amino- phosphorybosil-4-imidazolecarboxamideisomerase hisA	2	1.46	1.00	2.28	1.11	2.44	1.02
Isoleucine and valine metabolism								
slr0032	probable branched-chain amino acid aminotransferase ilvE	2	1.26	1.20	1.75	1.15	1.59	1.16
sll1363	ketol-acid reductoisomerase ilvC	3	3.22	1.29	1.77	1.49	2.16	1.13
sll1981	acetolactate synthase ilvB	2	-1.02	1.09	1.52	1.04	1.34	1.01
slr0452	dihydroxyacid dehydratase ilvD	3	1.48	1.19	2.10	1.22	3.23	1.17
Lysine metabolism								
sll1058	dihydrodipicolinate reductase dapB	3	-1.48	1.09	-1.66	1.16	-1.65	1.18
Methionine metabolism								
sll1234	adenosylhomocysteinase ahcY	5	-1.28	1.15	1.62	1.44	1.60	1.47
Urea cycle and arginine metabolism								
sll1750	urease alpha subunit ureC	2	-1.56	1.03	-1.25	1.01	-1.43	1.07
sll0902	ornithine carbamoyltransferase argF	2	-1.32	1.23	1.21	1.15	1.35	1.10

7.4.11 Effects on nucleotide metabolism.

One may expect the effect of phosphate reduction would impact on the nucleotide biosynthesis processes in the cell, since most of the substrate requires inorganic phosphate as well as it being the main building block of many compounds such as ATP, NTP, DNA, RNA etc. Among the 4 proteins identified here, inosine-5'-monophosphate (IMP) dehydrogenase (GuaB, slr1722), adenylate kinase (Adk, sll1815) and dihydroorotase (PyrC, sll1018) showed significant changes in protein expression in the phosphate-reduced cultures, with a more than 2-fold increase (for GuaB and Adk) and a 3-fold decrease (PyrC) respectively, as shown in Table 7.12.

Table 7.12. The relative expression of identified proteins involved in nucleotide metabolism.

ORF	Protein Name	MS/ MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
sll1018	dihydroorotase pyrC	2	-3.09	1.60	-2.97	2.56	-3.02	1.76
sll1815	adenylate kinase adk	7	1.55	1.11	1.25	1.19	2.02	1.20
sll1852	nucleoside diphosphate kinase ndk	4	-1.11	1.19	-1.01	1.19	1.24	1.12
slr1722	inosine-5'-monophosphate dehydrogenase guaB	4	2.25	1.48	2.41	1.62	2.65	1.59

GuaB protein catalyses IMP to xanthosine 5'-phosphate (XMP) with NAD^+ or $NADP^+$ as an acceptor in Purine metabolism, whereas PyrC converts N-carbamoyl-L-aspartate to dihydroorotate during the formation of Uridine 5'-monophosphate (UMP) in Pyrimidine metabolism [58]. The up-regulation in GuaB might be a direct effect from the phosphate stress response, while the down-regulation in PyrC suggests a possible linkage from NrtA and GlnB (both down-regulated as mentioned previously) in substrate-binding and signal transduction system. The main precursor of N-carbamoyl-L-aspartate comes from ammonia/nitrate in nitrogen metabolism, and since nitrogen metabolism showed overall down-regulation during phosphate limitation, it is unsurprising to observe a similar effect in PyrC. Nucleoside diphosphate kinase (Ndk, sll1852), though, relies heavily on ATP or ADP as substrate, and NTP and NDP as an acceptor/donor, no significant change in protein expression was recorded during phosphate limitation (Table 7.12). It is thought that many NDPs can act as acceptors, while many RTP and DTPs can act as donors [58].

7.4.12 Hypothetical proteins related to phosphate limitation.

This group of proteins produced some of the most dramatic changes seen in their protein expression. For example, a relative increase of 18-fold was observed for ORF ssl2501, and a 38-fold decrease was seen in sll1873 during phosphate deprivation conditions, as depicted in Table 7.13. Other hypothetical proteins with more than 10-fold changes in their relative expression ratios are: slr1847, slr1704 and ssl0385. To gain a deeper understanding of these unknown function proteins, they were searched against Cyanobacteria Gene Annotation Database

(CYROF) (http://cyano.genome.jp/), and small amount (9 proteins) of them shows relatively high similarity (in various percentages from 40% to 80%) to proteins of known function.

Table 7.13. The hypothetical proteins identified in this study and their corresponding expression ratio. The proteins with homology to another known function organism are in bracket.

ORF	Protein Name	MS/ MS	10% Pi		3% Pi		0.3% Pi	
			Ratio	EF	Ratio	EF	Ratio	EF
sll0051	hypothetical protein	2	-1.50	1.11	-2.29	1.08	-2.82	1.20
sll0230	hypothetical protein	3	2.48	1.48	3.11	1.46	2.80	1.43
sll0272	hypothetical protein	4	1.60	1.48	1.50	1.63	2.14	1.71
sll0274	(Pentapeptide)	2	1.67	1.00	1.81	1.02	2.03	1.02
sll0529	(Transketolase)	4	2.16	1.12	2.48	1.07	1.89	1.10
sll0588	hypothetical protein	2	-2.81	1.05	-1.69	1.02	-3.30	1.02
sll0822	hypothetical protein	2	-8.36	2.03	-7.86	2.25	-2.52	1.23
sll0995	hypothetical protein	2	1.14	1.21	1.34	1.36	2.31	1.37
sll1106	hypothetical protein	12	1.27	1.22	1.71	1.25	1.32	1.27
sll1118	hypothetical protein	2	-2.17	1.06	-1.72	1.15	-2.16	1.07
sll1130	hypothetical protein	2	-1.06	1.38	1.32	1.55	1.28	1.23
sll1188	(Cupin superfamily)	2	-1.14	5.09	-1.19	6.13	-1.17	5.54
sll1201	(Putative transposase)	2	1.53	1.41	1.71	1.50	2.31	1.57
sll1526	(CheR methyltransferase)	2	-1.25	1.05	-1.25	1.23	-1.25	1.07
sll1873	hypothetical protein	3	-29.97	1.30	-38.47	2.15	-12.65	1.12
sll1891	hypothetical protein	3	1.77	1.43	2.60	1.47	2.70	1.39
slr0001	(Flavin reductase like domain)	5	2.79	1.14	3.84	1.13	4.57	1.14
slr0013	hypothetical protein	2	-2.13	1.52	-1.21	1.61	-1.41	1.83
slr0244	(Universal stress protein family)	2	-3.01	1.05	-1.81	1.10	-1.85	1.01
slr0376	hypothetical protein	2	4.33	1.52	7.24	1.41	7.26	1.55
slr0476	hypothetical protein	2	1.94	1.13	2.77	1.04	4.49	1.07
slr0605	(putative glutathione S-transferase)	2	1.14	1.01	1.15	1.09	1.41	1.07
slr0606	(Glycosyltransferase family)	2	1.52	1.03	2.04	1.02	3.60	1.08
slr0729	hypothetical protein	11	1.54	1.25	2.28	1.24	2.79	1.23
slr0923	hypothetical protein YCF65	3	-1.12	1.22	1.23	1.17	2.57	1.29
slr1034	hypothetical protein YCF41	4	1.90	1.18	2.32	1.20	2.27	1.10
slr1590	hypothetical protein	2	1.50	1.17	1.76	1.09	1.96	1.02
slr1612	hypothetical protein	2	1.43	1.17	5.30	1.51	7.94	1.35
slr1704	hypothetical protein	2	-1.73	1.02	-1.17	1.02	10.10	1.02
slr1780	hypothetical protein YCF54	3	-1.07	2.35	1.00	2.34	1.47	1.75
slr1847	hypothetical protein	2	-1.21	1.10	1.07	1.11	10.86	1.59
slr1852	hypothetical protein	9	1.39	1.59	2.38	1.61	2.95	1.63
slr1855	hypothetical protein	3	-1.13	1.20	1.65	1.16	2.08	1.23
slr2032	hypothetical protein YCF23	4	1.53	1.27	2.18	1.36	1.63	1.47
ssl0385	hypothetical protein	2	3.88	1.37	8.73	1.36	10.89	1.33
ssl2501	hypothetical protein	2	4.62	1.75	13.11	1.76	18.44	1.81
ssr1528	hypothetical protein	3	3.98	1.26	3.70	1.28	2.81	1.27
ssr1853	hypothetical protein	5	1.72	1.55	2.23	1.74	1.95	1.84

Among them, protein slr0244 sharing homology of 70% and 60% to universal stress proteins (UspA) of *Crocosphaera watsonii* WH 8501 and *Anabaena variabilis* ATCC 29413 respectively, demonstrated down-regulation of more than 1.5-fold during phosphate limitation conditions as shown in Table 7.13. This is unusual since, as detailed above, most of the stress-related proteins show increases in their protein relative expression ratios in phosphate limited conditions, rather than a decrease. Protein sll0529 was found to be homologous (70%) with a transketolase protein from *Anabaena variabilis* ATCC 29413, which is one of the key enzymes involved in the pentose phosphate pathway and Calvin cycle. This protein showed good agreement in relative expression change seen here with the annotated transketolase protein TktA (sll1070) in *Synechocystis* sp., by demonstrating more than 2-fold up-regulation during phosphate deprivation conditions (Table 7.13). Other hypothetical proteins with close similarity to known functionally annotated proteins are: slr0605 (putative glutathione S-transferase), sll1201 (Putative transposase), and slr0606 (Glycosyltransferase family). Aside from slr0605, which did not change its protein expression significantly during phosphate limitation, sll1201 and slr0606 both showed relative increases in protein expression of up to 2-fold and 3-fold respectively in the 0.3% phosphate environment compared to the phosphate-replete control (Table 7.13).

7.5 Conclusions

The global proteomic assessment of the phosphate limitation response in *Synechocystis* sp. PCC 6803 has been successfully carried out here. Despite some key proteins not being identified through the iTRAQ-mediated proteomic approach, such as some involved in phosphate transport and the signal transduction two-component system, a good picture of the response was obtained. The quantitative response of the alkaline phosphatase (sll0654) and extracellular nuclease (NucH, sll0656) are in agreement with the expectations from the literature that these periplasmic enzymes increase their protein abundance in phosphate-reduced environment. It is also worth noting that a disparity was observed in the changes in relative protein expression between two of the phosphate substrate-binding periplasmic proteins (PstS) identified in this

study, since sll0680 (PstS1) was down-regulated and slr1247 (PstS2) was up-regulated. The association of PstS in phosphate-deprived conditions has been demonstrated previously, but these studies were normally conducted at the transcriptional level, whereas this study focuses on the interactions at the proteomic level.

As a counter-measure to phosphate reduction, the cells responded by increasing the expression of stress-related proteins, specifically the heat shock proteins and the glutathione peroxidase system. Further evidence was also provided in the up-regulation of superoxide dismutase SodB (slr1516) and an antioxidant protein (slr1198). The greatest impact of phosphate limitation was observed in the response of a number of hypothetical proteins. Central carbon metabolism also showed a strong response to the phosphate deprived conditions, with 16 from 19 enzymes identified showing significant up-regulation. On the other hand, the reduced level of phosphate caused a reduction in the overall photosynthesis capacity, as observed through physiological features such as colour changes in the cultures and a corresponding down-regulation of pigment proteins.

Chapter 8

Conclusions and Recommendation -

What does the Future hold?

8.1 Summary

Investigations tilted towards exacting a 'Systems Microbiology' understanding of a model organism, such as *Synechocystis* sp. PCC 6803 requires a range of assets that can facilitate the prediction of the complete cellular response to a stimulant. An integrated understanding at the gene, protein and metabolite level is vital. However, a great leap forward needs to be a combination of small steps. Old Chinese Proverb: The journey of 1000 miles begins with a single footstep. Researchers must understand the problems from a practical point of view, and answer the questions step by step. A small cross-section of successes does not mean a victory in the overall war, but often that small success can lead to a great discovery. In this context, the author presents this thesis, hoping to demonstrate the complementary nature of the combined techniques of transcriptomics and proteomics, and that the use of these techniques is feasible for scientific investigation. While many people have been sceptical with regard to the utility of proteomics in understanding cellular responses. It must not be forgotten that most biological functions are carried out by proteins and enzymes, and the change in gene expression, as measured through transcriptomics does not necessarily reflect and represent a true cellular action.

8.1.1 To achieve higher proteome coverage, multi-level prefractionation techniques must be considered.

Unlike transcriptomic techniques such as cDNA microarrays which cover almost the entire gene expression profile of a genome, a proteomic approach still faces many limitations. So far, it is impossible to identify the entire protein set from an organism through any of the available proteomic techniques. This limitation causes setbacks when specific information from a group of proteins is needed, especially for those with low abundance, or extreme properties (pI, Mw and hydrophobicity), and invariably these are often the proteins of interest. Many attempts have focused on alternative separation techniques, other than the conventional 2DE approach, in order to increase the percentage of proteome coverage [279-281]. The feasibility of employing multiple stages of orthogonal separation at the 'shotgun' level was examined in Chapter 3 [36]. By comparing the gel-based and gel-free prefractionation methods, gel-based approaches such

as IEF and 1-D PAGE performed the best, with the highest number of the proteins identified. In particular, the IEF protein and IEF peptide workflow gave the best results among the combinations of prefractionation techniques tested. Overall 776 proteins, covering 24% of *Synechocystis* sp. proteome, were confidently identified in the study. However it must be reminded that the 776 proteins were not identified from a single protein-peptide prefractionation workflow, it was a combined effort from all 6 of the prefractionation techniques. This further illustrates the difficulties in readily obtaining a higher percentage of proteome coverage, and the better way to move forward would be a strategy combining different (preferably automated) prefractionation methods, aimed at various experimental and sample properties [495], such as IEF for pI separation, 1-D PAGE for Mw separation and SCX for separation based polarity. On the other hand, chromatographic separation approaches have the advantage of shorter processing times compared to gel-based prefractionation techniques, with a higher number of SCX fractions (from 25 fractions to 40 fractions) significantly altering the total number of proteins identified.

8.1.2 A proteins expression change is significant only when sample variation is accounted for.

The deployment of the iTRAQ-reagent based quantitative proteomic approaches has provided researchers with a robust and high-throughput techniques that can be used to measure the protein expression change across multiple samples within a set of experiments [37]. However such technique can be misleading if the researchers do not take appropriate counter-measurements to qualify the data, such as replicates or other means of method validation, during the experiments. A protein expression change measured as a relative expression to the reference sample, commonly presents the value in linear space which often causes confusion when variation of the error is taken into account. Working in log space, however, has an advantage of even distribution at the mean of the value. In order to assess the reproducibility of an iTRAQ experiment, a total of 10 different iTRAQ set were used, varying in the form of replicates as either: biological, experimental or technical replicate. The average variation of technical, experimental and biological data packages were estimated as ±11%, ±23% and ±25%

respectively. However the percentage of protein distribution at this average variation value was relatively low, which in turn raises doubts regarding the quality of the iTRAQ data if average variation was used. By taking into consideration all three forms of replicate variation, at least 88% of the proteins would fall within the ±50% variation cut-off point, therefore it is recommended to consider the significance of a protein expression/regulation, only if the figure is above ±1.5-fold.

8.1.3 Fundamental understanding comes from the basic cellular response.

To understand the cellular response of an organism, it is essential to understand the most basic and fundamental mechanisms of that organism. In this case, dealing with a cyanobacterium, it would be photosynthesis (during the day) and respiration (during the night). After a lengthy validation process (Chapter 4), employing the iTRAQ-mediated approaches would hopefully gain an insight into the proteome profile of the organism during the light and dark cycle. Quantitative real-time PCR was also employed to examine the consistency between the expression profile of mRNA and protein. As expected (from other studies in the literature), only a third of them shared the same expression profile while the remaining did not correlate well. Amongst the identified Photosystem I (PSI) and Photosystem II (PSII) proteins, PsaA and PsaD (from PSI) proteins showed decreases in protein expression during light acclimation, whereas the PSII proteins, such as PsbW, PsbV and PsbU, showed enhanced expression under the same condition. This finding is further validated from an examination of the *psb*U and *psa*A genes, in which their transcript relative expression ratio agrees with the protein level. It is thought that the PSI and PSII complexes are regulated stoichiometrically during light acclimation by decreasing PSI content in order to maintain the efficiency of the photosynthesis reaction [156]. Among the major metabolic pathways, most of the proteins involved in central carbon metabolism comprising the glycolysis/gluconeogenesis, pentose phosphate and Calvin cycles, were up-regulated during light compared to dark cycle.

8.1.4 Circadian profile is triggered by light and dark conditions.

After developing a basic understanding of cellular (proteome) response to light alteration, it would be of interest if the dynamics of such protein and gene profiles could be captured using both transcriptional and translational techniques. Cyanobacteria are the only prokaryotes that contained a documented circadian clock [438]. It is aimed here, to investigate the circadian influence at the proteome and transcriptome level, by using the light and dark cycle as a trigger. In order to achieve this, the experiments were designed such that the cellular response of long term and short term effects to the light (and dark) acclimation could be captured using a two different time-points strategy: the transient and prolonged time-points. The study successfully unveiled the strong influence of circadian rhythm at the transcript level, whereas the effect at protein level was understandably less severe.

An oscillation with a periodic half-life of 12 hours at the transcript level was observed across all the genes investigated in this study. This is unsurprising, since the ability for the genes to turn-over in a short amount of time is vital for cell survival and adaptation. The disparities observed between the transcript and protein, are not uncommon considering the differences in the rate of synthesis, rate of degradation, as well as the fact that their half-life might actually contribute toward this variation. Nonetheless, according to the gene response toward light alteration, two types of circadian profiles (A- and B-type) were observed. Furthermore, 2 sub-groups between each type were observed based on their transcript sustainability with A1-, B1- characterised as high; and A2-, B2 as low. A-type genes were highly sensitive to light activation, whereas B-type genes only showed responses toward the dark phase. In total, there were 20 genes characterised as A-type (with 4 A1-type genes and 20 A2-type genes). Only 2 genes were categorised as B-type (one each for B1- and B2-type).

8.1.5 The global proteomic response toward phosphate starvation.

Moving on towards more environmental influencing factors, the depletion of macro-nutrients would definitely upset the metabolic balance of an organism and elicit a stress-based cellular response. The intention here was to measure the global proteomic response of *Synechocystis* sp.

when subjected to phosphate reduction stress, and sought to obtain a more in-depth understanding of the methods by which this organism copes with this situation from major metabolic pathways as well as the stress-induced proteins. Despite many missing phosphate-related proteins such as the in the Pho regulon, a very good overview was obtained through the iTRAQ-mediated proteomic approach. Only one (out of three) alkaline phosphatase (sll0654) was identified in the study with enhanced expression in a phosphate starvation environment. The alkaline phosphatase activity was further verified through enzyme assays whereby an increase in the phosphatase activity was noted as the phosphate concentration in the medium was reduced. Furthermore, the extracellular nucleus (sll0656) protein, which is located just upstream of sll0654, was also up-regulated in response to phosphate limitation. Other phosphate associated proteins included phosphate substrate-binding periplasmic proteins, PstS1 and PstS2. Interestingly these two proteins did not show agreement with each other, with PstS1 being down-regulated and vice versa for PstS2. The primary metabolic pathway, central carbon metabolism showed an overall up-regulation, whereas photosynthesis and amino acids metabolism showed a mixed response.

8.2 Future Recommendations

8.2.1 Gel-based prefractionation and multiple injection analysis are the way forward.

The work presented in Chapter 3 [36] has became the benchmark for multi-dimensional shotgun proteomic analysis for the remainder of this body of work and indeed beyond [389, 496], where it was proven that combined prefractionation approaches could aid in large scale protein identification. In particular with the success of (IEF-IEF) gel-based prefractionation techniques, the 1D-IEF workflow is thus recommended. Furthermore, there is a glimmer of hope that analysing the same sample multiple times in the MS might actually increase the total number of proteins and peptides identified. All these recommendations were further proven and verified by subsequent investigations. For example, Chong and Wright carried out a proteomic investigation on *Sulfolobus solfataricus* P2 using the recommended 1D-IEF approach, together with the IEF-IEF workflows [389]. The combined effort of these workflows resulted in the identification of more than 1200 unique proteins (40% proteome coverage), compared to only

432 proteins using conventional 2DE. Moreover, the 1D-IEF workflow showed a slight advantage over the IEF-IEF workflow with an additional 10% more proteins being identified. In this study, multiple injection analysis was also employed and up to 50% more proteins were identified through triplicate analysis compared to single MS analysis, which further validated the suggestions made in Chapter 3 [36]. A completely separate study carried out by Elias *et al.* used multiple injections as a means of validating the peptide and protein confidence further confirmed these findings [409]. This study noted an average of 24% possible error in the protein identification if only one replicate injection was made, compared to less than 2% if triplicate injections were employed. The implementation of multiple injections initiated through Chapter 3 is not limited to qualitative proteomics, it has recently been applied to quantitative proteomic analysis, specifically in iTRAQ-based approaches, and successfully established the importance of information (protein and peptide) validation through this simple yet powerful method [39].

8.2.2 At least one biological replicate and at least ±50% variation necessary for significant protein expression to have occurred.

The ever increasing pace of iTRAQ applications in the proteomics has put the quality of the information obtained through this method on trial. By the authors reckoning, it is necessary to ensure the reliability of the (iTRAQ) dataset by performing some form of validation methods. It has been demonstrated previously that multiple injections can successfully enhance the standard/quality of the iTRAQ data to a CV of 9% [39]. In this context, replicate analysis is recommended. It is advisable to include at least one set of biological replicates in any iTRAQ experiment, since the biological variation was proven to be the greatest contributor amongst all types of replicate. A significant protein regulation change would be that which exceeds the variation present in and between biological samples, as well as dispersion from a sample variance. In Chapter 4 [38], a ±50% cut-off point was recommended, i.e. only proteins with ±1.5-fold regulation can be considered as significantly up- or down-regulation. However, it must be reiterated that this value was obtained through the variation calculated from *Saccharomyces cerevisiae*, *Sulfolobus solfataricus* and *Synechocystis sp.* PCC 6803, thus this figure might vary from organism to organism although it is possible to assume for those

organism investigated here, a ±50% cut-off point is feasible. Another validation strategy employing sample pooling was also discussed. It is unclear if this strategy is applicable to proteomic analysis of this type. However based on simple theoretical assumptions (refer to Chapter 4), it is hypothesized that sample pooling could be used to reduce the sample to sample variation. Since it is a well-established strategy in cDNA microarray technology, the author does not see as reason why it wouldn't work in protein expression experiments.

8.2.3 Ambiguous data requires further validation.

Some ambiguous proteins (those contradicting with the literature reports) still require further validation, either in the form of enzyme assays or western blotting, prior to a more rigorous conclusion, such as the PSI and PSII proteins discussed in Chapters 5 and 6; the PstS1 and PstS2 in phosphate stress study in Chapter 7, for example. Furthermore, although the transcriptional and translational responses do not correlate well in this thesis, it is still advisable to measure both mRNA and protein levels whenever possible in order to increase the understanding of the biological system.

8.2.4 Large-scale studies require computational aid.

It is not the intention of this body of work to find or form a conclusion/solution around the 'holy-grail' question – How to solve the exact quantitative relationship between gene and protein expression? Instead the key is to understand and predict their differences, or at least the factors that would lead to these differences. It is recommended, in the future, to carry out a larger scale investigation targeting more genes and proteins, perhaps employing microarrays (see below), and this is only going to really successful with the aids of computational modelling, whereby gene and protein profiles could be matched in mathematical terms. This modelling could help predict the behaviour of how and when a proteins is made from a transcribed gene, thereby enabling researchers to rationally design a biological system such that the organism could perform the requested task both *in silico* and *in vivo*. For example, Zimmerman proposed a modelling strategy utilising a complex metabolic pathway kinetics models for optimal ethanol production in yeast, by forcing an oscillatory behaviour in the glucose feed [497]. If a similar

computational strategy was applied on *Synechocystis* sp. PCC 6803, it may be possible to enhance cyanophycin, polyhydroxyalkanoate, carotenoids or even biological H_2 production in the near future.

8.2.5 Future cDNA microarray analysis, but is it worth it?

The studies conducted in this thesis (Chapters 5, 6 and 7) would have been further enhanced by the employment of the cDNA microarray analysis, since this would reveal the global transcriptional response, as well as increase the understanding of light acclimation as in Chapter 5, circadian rhythm as in Chapter 6 and phosphate limitation as in Chapter 7. However, with this additional set of transcriptional data, it might well be ended up in discordant results compared to the protein expression changes as measured by the iTRAQ workflow. It would not be surprising since (from previous findings reported in this thesis and in the wider literature) the transcript and protein often do not correlate well with each other. Furthermore the financial cost arises from a genome-wide transcriptional investigation could be significant, especially in temporal studies like that carried out in Chapter 6, if all 10 time-points were to be examined in triplicate. One can only wait until such technology becomes more affordable in the future to attempt this type of blanket transcript study.

8.2.6 Did phosphate starvation trigger biological H_2 production?

One particularly interesting observation from this study (Chapter 7) is the increase in protein relative abundance of the bidirectional hydrogenase protein (HoxH), which might indicate the activation of biological H_2 production under a phosphate stress conditions. The influence of phosphate depletion has been reported in cyanophycin [101] and polyhydroxyalkanoate [9, 115, 116] accumulation, yet never in this instance. It would be of interest if the physiological measurements of H_2 could be taken in the future under phosphate starvation condition, facilitating a direct correlation between phosphate stress and the hydrogen production. Further in-depth investigation is thus recommended. Moreover the study also uncovered a number of hypothetical proteins with significant changes in regulation. The positive (or negative) response of these unknown function proteins yields some small insight into their actual functional

mechanism, in the sense that it appears that they are either being involved in the phosphate transport related system or acting as general stress-induced proteins. Nonetheless this study provides some clues for a future follow-up investigation. When additional information is available for these unknown proteins, the understanding of their response (toward phosphate starvation) will be greatly enhanced.

8.3 Concluding Remarks

The investigations carried out in this thesis are a combination of method developments and biological studies. It is undoubtedly that the employment of new technology, such as iTRAQ, had successfully aided the functional proteomics analysis of *Synechocystis* sp. PCC 6803. The global proteomic views revealed via this approach has provided in-depth understanding of this organism response toward environmental stresses, specifically to the photoperiod (Chapter 5 and 6) and phosphate starvation (Chapter 7). However it is important not to undermine the power of transcriptomics, because the author believes a successful 'Systems Microbiology' study must incorporate all essential tools, including proteomics, transcriptomics as well as metabolomics. Metabolic engineering, though, is not the subject of study here, the information derived from both transcriptional and translational responses will form the basis for future metabolic studies.

References

[1] Schutz, K., Happe, T., Troshina, O., Lindblad, P., Leitao, E., Oliveira, P. and Tamagnini, P. (2004) "Cyanobacterial H(2) production -- a comparative analysis". *Planta.*, **218** (3), 350-9.

[2] Antal, T.K. and Lindblad, P. (2005) "Production of H2 by sulphur-deprived cells of the unicellular cyanobacteria *Gloeocapsa alpicola* and *Synechocystis* sp. PCC 6803 during dark incubation with methane or at various extracellular pH". *J Appl Microbiol*, **98** (1), 114-20.

[3] Antal, T.K., Oliveira, P. and Lindblad, P. (2006) "The bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Int J Hydrogen Energy*, **31** (11), 1439-1444.

[4] Frey, K.M., Oppermann-Sanio, F.B., Schmidt, H. and Steinbuchel, A. (2002) "Technical-scale production of cyanophycin with recombinant strains of *Escherichia coli*". *Appl Environ Microbiol*, **68** (7), 3377-84.

[5] Allen, M.M., Yuen, C., Medeiros, L., Zizlsperger, N., Farooq, M. and Kolodny, N.H. (2005) "Effects of light and chloramphenicol stress on incorporation of nitrogen into cyanophycin in *Synechocystis* sp. strain PCC 6308". *Biochim Biophys Acta*, **1725** (2), 241-6.

[6] Elbahloul, Y., Frey, K., Sanders, J. and Steinbuchel, A. (2005) "Protamylasse, a residual compound of industrial starch production, provides a suitable medium for large-scale cyanophycin production". *Appl Environ Microbiol*, **71** (12), 7759-67.

[7] Wu, G.F., Wu, Q.Y. and Shen, Z.Y. (2001) "Accumulation of poly-beta-hydroxybutyrate in cyanobacterium *Synechocystis* sp. PCC6803". *Bioresour Technol*, **76** (2), 85-90.

[8] Sudesh, K., Taguchi, K. and Doi, Y. (2002) "Effect of increased PHA synthase activity on polyhydroxyalkanoates biosynthesis in *Synechocystis* sp. PCC6803". *Int J Biol Macromol*, **30** (2), 97-104.

[9] Panda, B., Jain, P., Sharma, L. and Mallick, N. (2006) "Optimization of cultural and nutritional conditions for accumulation of poly-beta-hydroxybutyrate in *Synechocystis* sp. PCC 6803". *Bioresour Technol*, **97** (11), 1296-301.

[10] Lagarde, D. and Vermaas, W. (1999) "The zeaxanthin biosynthesis enzyme beta-carotene hydroxylase is involved in myxoxanthophyll synthesis in *Synechocystis* sp. PCC 6803". *FEBS Lett*, **454** (3), 247-51.

[11] Dellapenna, D. and Pogson, B.J. (2006) "Vitamin Synthesis in Plants: Tocopherols and Carotenoids". *Annu Rev Plant Biol*, **57**, 711-738.

[12] Mohamed, H.E., van de Meene, A.M., Roberson, R.W. and Vermaas, W.F. (2005) "Myxoxanthophyll is required for normal cell wall structure and thylakoid organization in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **187** (20), 6883-92.

[27] Scotter, M.J., Castle, L., Croucher, J.M. and Olivier, L. (2003) "Method development and analysis of retail foods and beverages for carotenoid food colouring materials E160a(ii) and E160e". *Food Addit Contam*, **20** (2), 115-26.

[28] Visser, H., van Ooyen, A.J. and Verdoes, J.C. (2003) "Metabolic engineering of the astaxanthin-biosynthetic pathway of *Xanthophyllomyces dendrorhous*". *FEMS Yeast Res*, **4** (3), 221-31.

[29] Higuera-Ciapara, I., Felix-Valenzuela, L. and Goycoolea, F.M. (2006) "Astaxanthin: a review of its chemistry and applications". *Crit Rev Food Sci Nutr*, **46** (2), 185-96.

[30] Singh, D.K. and Lippman, S.M. (1998) "Cancer chemoprevention. Part 1: Retinoids and carotenoids and other classic antioxidants". *Oncology (Williston Park)*, **12** (11). 1643-53, 1657-8; discussion 1659-60.

[31] Smith, T.A.D. (1998) "Carotenoids and cancer: prevention and potential therapy." *British Journal of Biomedical Science*, **55**, 268-275.

[32] Business Communications Company (2005). *The Global Market for Carotenoids*. Business Communications Company Inc., Norwalk, Conn. (Report No. **RGA-110R**).

[33] DellaPenna, D. (2005) "Progress in the dissection and manipulation of vitamin E synthesis". *Trends Plant Sci*, **10** (12), 574-9.

[34] DellaPenna, D. (2005) "A decade of progress in understanding vitamin E synthesis in plants". *J Plant Physiol*, **162** (7), 729-37.

[35] Lee, P.C. and Schmidt-Dannert, C. (2002) "Metabolic engineering towards biotechnological production of carotenoids in microorganisms". *Appl Microbiol Biotechnol*, **60** (1-2), 1-11.

[36] Gan, C.S., Reardon, K.F. and Wright, P.C. (2005) "Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of *Synechocystis* sp. PCC 6803". *Proteomics*, **5** (9), 2468-78.

[37] Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. and Pappin, D.J. (2004) "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents". *Mol Cell Proteomics*, **3** (12), 1154-69.

[38] Gan, C.S., Chong, P.K., Pham, T.K. and Wright, P.C. (2006) "Technical, Experimental and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ)". *J Proteome Res*, 6 (2), 821-7.

[39] Chong, P.K., Gan, C.S., Pham, T.K. and Wright, P.C. (2006) "Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections". *J Proteome Res*, **5** (5), 1232-40.

[51] Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) "Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions". *DNA Res*, **3** (3), 109-36.

[52] Kaneko, T. and Tabata, S. (1997) "Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803". *Plant Cell Physiol*, **38** (11), 1171-6.

[53] Nakamura, Y., Kaneko, T., Hirosawa, M., Miyajima, N. and Tabata, S. (1998) "CyanoBase, a www database containing the complete nucleotide sequence of the genome of *Synechocystis* sp. strain PCC6803". *Nucleic Acids Res*, **26** (1), 63-7.

[54] Herranen, M., Tyystjarvi, T. and Aro, E.M. (2005) "Regulation of photosystem I reaction center genes in *Synechocystis* sp. strain PCC 6803 during Light acclimation". *Plant Cell Physiol*, **46** (9), 1484-93.

[55] Xu, Q., Odom, W.R., Guikema, J.A., Chitnis, V.P. and Chitnis, P.R. (1994) "Targeted deletion of psaJ from the cyanobacterium *Synechocystis* sp. PCC 6803 indicates structural interactions between the PsaJ and PsaF subunits of photosystem I". *Plant Mol Biol*, **26** (1), 291-302.

[56] Thornton, L.E., Ohkawa, H., Roose, J.L., Kashino, Y., Keren, N. and Pakrasi, H.B. (2004) "Homologs of plant PsbP and PsbQ proteins are necessary for regulation of photosystem ii activity in the cyanobacterium Synechocystis 6803". *Plant Cell*, **16** (8), 2164-75.

[57] Summerfield, T.C., Shand, J.A., Bentley, F.K. and Eaton-Rye, J.J. (2005) "PsbQ (Sll1638) in Synechocystis sp. PCC 6803 is required for photosystem II activity in specific mutants and in nutrient-limiting conditions". *Biochemistry*, **44** (2), 805-15.

[58] Stryer, L., 4th. (1995). *Biochemistry,* New York: W. H. Freeman Co.

[59] Berg, M.B., John, L.T. and Stryer, L., 5th. (2002). *Biochemistry,* New York: W. H. Freeman Co.

[60] Iwaki, T., Haranoh, K., Inoue, N., Kojima, K., Satoh, R., Nishino, T., Wada, S., Ihara, H., Tsuyama, S., Kobayashi, H. and Wadano, A. (2006) "Expression of foreign type I ribulose-1,5-bisphosphate carboxylase/ oxygenase (EC 4.1.1.39) stimulates photosynthesis in cyanobacterium *Synechococcus* PCC7942 cells". *Photosynth Res*.

[61] Yang, C., Hua, Q. and Shimizu, K. (2002) "Metabolic flux analysis in *Synechocystis* using isotope distribution from 13C-labeled glucose". *Metab Eng*, **4** (3), 202-16.

[62] Knowles, V.L. and Plaxton, W.C. (2003) "From genome to enzyme: analysis of key glycolytic and oxidative pentose-phosphate pathway enzymes in the cyanobacterium *Synechocystis* sp. PCC 6803". *Plant Cell Physiol*, **44** (7), 758-63.

[75] Boison, G., Bothe, H., Hansel, A. and Lindblad, P. (1999) "Evidence against a common use of the diaphorase subunits by the bidirectional hydrogenase and by the respiratory complex I in cyanobacteria". *FEMS Microbiol Lett*, **174**, 159-165.

[76] Appel, J., Phunpruch, S., Steinmuller, K. and Schulz, R. (2000) "The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis". *Arch Microbiol*, **173** (5-6), 333-8.

[77] Stal, L.J. and Moezelaar, R. (1997) "Fermentation in cyanobacteria". *FEMS Microbiol Rev*, **21**, 179-211.

[78] Oliveira, P. and Lindblad, P. (2005) "LexA, a transcription regulator binding in the promoter region of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803". *FEMS Microbiol Lett*, **251** (1), 59-66.

[79] Gutekunst, K., Phunpruch, S., Schwarz, C., Schuchardt, S., Schulz-Friedrich, R. and Appel, J. (2005) "LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator". *Mol Microbiol*, **58** (3), 810-23.

[80] O'Reilly, E.K. and Kreuzer, K.N. (2004) "Isolation of SOS constitutive mutants of *Escherichia coli*". *J Bacteriol*, **186** (21), 7149-60.

[81] Domain, F., Houot, L., Chauvat, F. and Cassier-Chauvat, C. (2004) "Function and regulation of the cyanobacterial genes lexA, recA and ruvB: LexA is critical to the survival of cells facing inorganic carbon starvation". *Mol Microbiol*, **53** (1), 65-80.

[82] Patterson-Fortin, L.M., Colvin, K.R. and Owttrim, G.W. (2006) "A LexA-related protein regulates redox-sensitive expression of the cyanobacterial RNA helicase, crhR". *Nucleic Acids Res*, **34** (12), 3446-54.

[83] Vignais, P.M., Magnin, J. and Willison, J.C. (2006) "Increasing biohydrogen production by metabolic engineering". *Int J Hydrogen Energy*, **31** (11), 1478-1483.

[84] Sheremetieva, M.E., Troshina, O.Y., Serebryakova, L.T. and Lindblad, P. (2002) "Identification of hox genes and analysis of their transcription in the unicellular cyanobacterium *Gloeocapsa alpicola* CALU 743 growing under nitrate-limiting conditions". *FEMS Microbiol Lett*, **214** (2), 229-33.

[85] Kosourov, S., Tsygankov, A., Seibert, M. and Ghirardi, M.L. (2002) "Sustained hydrogen photoproduction by *Chlamydomonas reinhardtii*: Effects of culture parameters". *Biotechnol Bioeng*, **78** (7), 731-40.

[86] Wang, R., Healey, F.P. and Myers, J. (1971) "Amperometric Measurement of Hydrogen Evolution in *Chlamydomonas*". *Plant Physiol*, **48** (1), 108-110.

[87] Ziegler, K., Diener, A., Herpin, C., Richter, R., Deutzmann, R. and Lockau, W. (1998) "Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin)". *Eur J Biochem*, **254** (1), 154-9.

M.M. (2006) "Effect of nitrogen source on cyanophycin synthesis in *Synechocystis* sp. strain PCC 6308". *J Bacteriol*, **188** (3), 934-40.

[100] Zuther, E., Schubert, H. and Hagemann, M. (1998) "Mutation of a gene encoding a putative glycoprotease leads to reduced salt tolerance, altered pigmentation, and cyanophycin accumulation in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **180** (7), 1715-22.

[101] Stephan, D.P., Ruppel, H.G. and Pistorius, E.K. (2000) "Interrelation between cyanophycin synthesis, L-arginine catabolism and photosynthesis in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Z Naturforsch [C]*, **55** (11-12), 927-42.

[102] Maheswaran, M., Ziegler, K., Lockau, W., Hagemann, M. and Forchhammer, K. (2006) "PII-regulated arginine synthesis controls accumulation of cyanophycin in *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **188** (7), 2730-4.

[103] Li, H., Sherman, D.M., Bao, S. and Sherman, L.A. (2001) "Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria". *Arch Microbiol*, **176** (1-2), 9-18.

[104] Voss, I., Diniz, S.C., Aboulmagd, E. and Steinbuchel, A. (2004) "Identification of the *Anabaena* sp. strain PCC7120 cyanophycin synthetase as suitable enzyme for production of cyanophycin in gram-negative bacteria like *Pseudomonas putida* and *Ralstonia eutropha*". *Biomacromolecules*, **5** (4), 1588-95.

[105] Diniz, S.C., Voss, I. and Steinbuchel, A. (2006) "Optimization of cyanophycin production in recombinant strains of *Pseudomonas putida* and *Ralstonia eutropha* employing elementary mode analysis and statistical experimental design". *Biotechnol Bioeng*, **93** (4), 698-717.

[106] Voss, I. and Steinbuchel, A. (2006) "Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in *Ralstonia eutropha* strain H16". *Metab Eng*, **8** (1), 66-78.

[107] Aldor, I.S. and Keasling, J.D. (2003) "Process design for microbial plastic factories: metabolic engineering of polyhydroxyalkanoates". *Current Opinion in Biotechnology*, **14**, 475-483.

[108] Asada, Y., Miyake, M., Miyake, J., Kurane, R. and Tokiwa, Y. (1999) "Photosynthetic accumulation of poly-(hydroxybutyrate) by cyanobacteria--the metabolism and potential for CO2 recycling". *Int J Biol Macromol*, **25** (1-3), 37-42.

[109] Hein, S., Tran, H. and Steinbuchel, A. (1998) "*Synechocystis* sp. PCC6803 possesses a two-component polyhydroxyalkanoic acid synthase similar to that of anoxygenic purple sulfur bacteria". *Arch Microbiol*, **170** (3), 162-70.

[110] Taroncher-Oldenburg, G., Nishina, K. and Stephanopoulos, G. (2000) "Identification and analysis of the polyhydroxyalkanoate-specific beta-ketothiolase and acetoacetyl

[124] Lichtenthaler, H.K. (1999) "The 1-Deoxy-D-Xylulose-5-Phosphate Pathway Of Isoprenoid Biosynthesis In Plants". *Annu Rev Plant Physiol Plant Mol Biol*, **50**, 47-65.

[125] Fernandez-Gonzalez, B., Sandmann, G. and Vioque, A. (1997) "A new type of asymmetrically acting beta-carotene ketolase is required for the synthesis of echinenone in the cyanobacterium *Synechocystis* sp. PCC 6803". *J Biol Chem*, **272** (15), 9728-33.

[126] Collakova, E. and DellaPenna, D. (2001) "Isolation and functional analysis of homogentisate phytyltransferase from *Synechocystis* sp. PCC 6803 and *Arabidopsis*". *Plant Physiol*, **127** (3), 1113-24.

[127] Martinez-Ferez, I., Fernandez-Gonzalez, B., Sandmann, G. and Vioque, A. (1994) "Cloning and expression in *Escherichia coli* of the gene coding for phytoene synthase from the cyanobacterium *Synechocystis* sp. PCC6803". *Biochim Biophys Acta*, **1218** (2), 145-52.

[128] Fernandez-Gonzalez, B., Martinez-Ferez, I.M. and Vioque, A. (1998) "Characterization of two carotenoid gene promoters in the cyanobacterium *Synechocystis* sp. PCC 6803". *Biochim Biophys Acta*, **1443** (3), 343-51.

[129] Martinez-Ferez, I.M. and Vioque, A. (1992) "Nucleotide sequence of the phytoene desaturase gene from *Synechocystis* sp. PCC 6803 and characterization of a new mutation which confers resistance to the herbicide norflurazon". *Plant Mol Biol*, **18** (5), 981-3.

[130] Breitenbach, J., Fernandez-Gonzalez, B., Vioque, A. and Sandmann, G. (1998) "A higher-plant type zeta-carotene desaturase in the cyanobacterium *Synechocystis* PCC6803". *Plant Mol Biol*, **36** (5), 725-32.

[131] Masamoto, K., Misawa, N., Kaneko, T., Kikuno, R. and Toh, H. (1998) "Beta-carotene hydroxylase gene from the cyanobacterium *Synechocystis* sp. PCC6803". *Plant Cell Physiol*, **39** (5), 560-4.

[132] Mohamed, H.E. and Vermaas, W.F. (2006) "Sll0254 (CrtL(diox)) is a bifunctional lycopene cyclase/dioxygenase in cyanobacteria producing myxoxanthophyll". *J Bacteriol*, **188** (9), 3337-44.

[133] Mohamed, H.E. and Vermaas, W. (2004) "Slr1293 in *Synechocystis* sp. strain PCC 6803 Is the C-3',4' desaturase (CrtD) involved in myxoxanthophyll biosynthesis". *J Bacteriol*, **186** (17), 5621-8.

[134] Masamoto, K., Wada, H., Kaneko, T. and Takaichi, S. (2001) "Identification of a gene required for cis-to-trans carotene isomerization in carotenogenesis of the cyanobacterium *Synechocystis* sp. PCC 6803". *Plant Cell Physiol*, **42** (12), 1398-402.

[135] Breitenbach, J., Vioque, A. and Sandmann, G. (2001) "Gene sll0033 from *Synechocystis* 6803 encodes a carotene isomerase involved in the biosynthesis of all-E lycopene". *Z Naturforsch [C]*, **56** (9-10), 915-7.

[147] Qi, Q., Hao, M., Ng, W.O., Slater, S.C., Baszis, S.R., Weiss, J.D. and Valentin, H.E. (2005) "Application of the *Synechococcus* nirA promoter to establish an inducible expression system for engineering the *Synechocystis* tocopherol pathway". *Appl Environ Microbiol*, **71** (10), 5678-84.

[148] Helm, M. (2006) "Post-transcriptional nucleotide modification and alternative folding of RNA". *Nucleic Acids Res*, **34** (2), 721-33.

[149] Wold, F. (1981) "In vivo chemical modification of proteins (post-translational modification)". *Annu Rev Biochem*, **50**, 783-814.

[150] Han, K.K. and Martinage, A. (1992) "Post-translational chemical modification(s) of proteins". *Int J Biochem*, **24** (1), 19-28.

[151] Murata, N. and Suzuki, I. (2006) "Exploitation of genomic sequences in a systematic analysis to access how cyanobacteria sense environmental stress". *J Exp Bot*, **57** (2), 235-47.

[152] Mohamed, A. and Jansson, C. (1989) "Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium *Synechocystis* 6803". *Plant Mol Biol*, **13** (6), 693-700.

[153] Trubitsin, B.V., Mamedov, M.D., Vitukhnovskaya, L.A., Semenov, A.Y. and Tikhonov, A.N. (2003) "EPR study of light-induced regulation of photosynthetic electron transport in *Synechocystis* sp. strain PCC 6803". *FEBS Lett*, **544** (1-3), 15-20.

[154] Constant, S., Perewoska, I., Alfonso, M. and Kirilovsky, D. (1997) "Expression of the psbA gene during photoinhibition and recovery in *Synechocystis* PCC 6714: inhibition and damage of transcriptional and translational machinery prevent the restoration of photosystem II activity". *Plant Mol Biol*, **34** (1), 1-13.

[155] Allakhverdiev, S.I., Tsvetkova, N., Mohanty, P., Szalontai, B., Moon, B.Y., Debreczeny, M. and Murata, N. (2005) "Irreversible photoinhibition of photosystem II is caused by exposure of *Synechocystis* cells to strong light for a prolonged period". *Biochim Biophys Acta*, **1708** (3), 342-51.

[156] Muramatsu, M. and Hihara, Y. (2003) "Transcriptional regulation of genes encoding subunits of photosystem I during acclimation to high-light conditions in *Synechocystis* sp. PCC 6803". *Planta*, **216** (3), 446-53.

[157] Sonoike, K., Hihara, Y. and Ikeuchi, M. (2001) "Physiological significance of the regulation of photosystem stoichiometry upon high light acclimation of *Synechocystis* sp. PCC 6803". *Plant Cell Physiol*, **42** (4), 379-84.

[158] Mullineaux, C.W. (2001) "How do cyanobacteria sense and respond to light?" *Mol Microbiol*, **41** (5), 965-71.

[159] Hubschmann, T., Yamamoto, H., Gieler, T., Murata, N. and Borner, T. (2005) "Red and far-red light alter the transcript profile in the cyanobacterium *Synechocystis* sp. PCC 6803: impact of cyanobacterial phytochromes". *FEBS Lett*, **579** (7), 1613-8.

[173] Bhattacharya, J., GhoshDastidar, K., Chatterjee, A., Majee, M. and Majumder, A.L. (2004) "*Synechocystis* Fe superoxide dismutase gene confers oxidative stress tolerance to *Escherichia coli*". *Biochem Biophys Res Commun*, **316** (2), 540-4.

[174] Hosoya-Matsuda, N., Motohashi, K., Yoshimura, H., Nozaki, A., Inoue, K., Ohmori, M. and 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". *J Biol Chem*, **280** (1), 840-6.

[175] Asadulghani, Suzuki, Y. and Nakamoto, H. (2003) "Light plays a key role in the modulation of heat shock response in the cyanobacterium *Synechocystis* sp PCC 6803". *Biochem Biophys Res Commun*, **306** (4), 872-9.

[176] Glatz, A., Horvath, I., Varvasovszki, V., Kovacs, E., Torok, Z. and Vigh, L. (1997) "Chaperonin genes of the *Synechocystis* PCC 6803 are differentially regulated under light-dark transition during heat stress". *Biochem Biophys Res Commun*, **239** (1), 291-7.

[177] Kovacs, E., van der Vies, S.M., Glatz, A., Torok, Z., Varvasovszki, V., Horvath, I. and Vigh, L. (2001) "The chaperonins of *Synechocystis* PCC 6803 differ in heat inducibility and chaperone activity". *Biochem Biophys Res Commun*, **289** (4), 908-15.

[178] Basha, E., Lee, G.J., Breci, L.A., Hausrath, A.C., Buan, N.R., Giese, K.C. and Vierling, E. (2004) "The identity of proteins associated with a small heat shock protein during heat stress in vivo indicates that these chaperones protect a wide range of cellular functions". *J Biol Chem*, **279** (9), 7566-75.

[179] Friedrich, K.L., Giese, K.C., Buan, N.R. and Vierling, E. (2004) "Interactions between small heat shock protein subunits and substrate in small heat shock protein-substrate complexes". *J Biol Chem*, **279** (2), 1080-9.

[180] Suzuki, I., Simon, W.J. and Slabas, A.R. (2006) "The heat shock response of *Synechocystis* sp. PCC 6803 analysed by transcriptomics and proteomics". *J Exp Bot*, **57** (7), 1573-8.

[181] Horvath, I., Glatz, A., Varvasovszki, V., Torok, Z., Pali, T., Balogh, G., Kovacs, E., Nadasdi, L., Benko, S., Joo, F. and Vigh, L. (1998) "Membrane physical state controls the signaling mechanism of the heat shock response in *Synechocystis* PCC 6803: identification of hsp17 as a "fluidity gene"". *Proc Natl Acad Sci U S A*, **95** (7), 3513-8.

[182] Giese, K.C. and Vierling, E. (2002) "Changes in oligomerization are essential for the chaperone activity of a small heat shock protein in vivo and in vitro". *J Biol Chem*, **277** (48), 46310-8.

[183] Torok, Z., Goloubinoff, P., Horvath, I., Tsvetkova, N.M., Glatz, A., Balogh, G., Varvasovszki, V., Los, D.A., Vierling, E., Crowe, J.H. and Vigh, L. (2001) "*Synechocystis* HSP17 is an amphitropic protein that stabilizes heat-stressed membranes and binds denatured proteins for subsequent chaperone-mediated refolding". *Proc Natl Acad Sci U S A*, **98** (6), 3098-103.

[196] Desplats, P., Folco, E. and Salerno, G.L. (2005) "Sucrose may play an additional role to that of an osmolyte in *Synechocystis* sp. PCC 6803 salt-shocked cells". *Plant Physiol Biochem*, **43** (2), 133-8.

[197] Allakhverdiev, S.I., Nishiyama, Y., Suzuki, I., Tasaka, Y. and Murata, N. (1999) "Genetic engineering of the unsaturation of fatty acids in membrane lipids alters the tolerance of *Synechocystis* to salt stress". *Proc Natl Acad Sci U S A*, **96** (10), 5862-7.

[198] Allakhverdiev, S.I., Kinoshita, M., Inaba, M., Suzuki, I. and Murata, N. (2001) "Unsaturated fatty acids in membrane lipids protect the photosynthetic machinery against salt-induced damage in *Synechococcus*". *Plant Physiol*, **125** (4), 1842-53.

[199] Sakamoto, T. and Murata, N. (2002) "Regulation of the desaturation of fatty acids and its role in tolerance to cold and salt stress". *Curr Opin Microbiol*, **5** (2), 208-10.

[200] Allakhverdiev, S.I., Nishiyama, Y., Miyairi, S., Yamamoto, H., Inagaki, N., Kanesaki, Y. and Murata, N. (2002) "Salt stress inhibits the repair of photodamaged photosystem II by suppressing the transcription and translation of psbA genes in *Synechocystis*". *Plant Physiol*, **130** (3), 1443-53.

[201] Ryu, J.Y., Suh, K.H., Chung, Y.H., Park, Y.M., Chow, W.S. and Park, Y.I. (2003) "Cytochrome c oxidase of the cyanobacterium *Synechocystis* sp. PCC 6803 protects photosynthesis from salt stress". *Mol Cells*, **16** (1), 74-7.

[202] Vinnemeier, J., Kunert, A. and Hagemann, M. (1998) "Transcriptional analysis of the isiAB operon in salt-stressed cells of the cyanobacterium *Synechocystis* sp. PCC 6803". *FEMS Microbiol Lett*, **169** (2), 323-30.

[203] Fulda, S., Mikkat, S., Schroder, W. and Hagemann, M. (1999) "Isolation of salt-induced periplasmic proteins from *Synechocystis* sp. strain PCC 6803". *Arch Microbiol*, **171** (3), 214-7.

[204] Marin, K., Kanesaki, Y., Los, D.A., Murata, N., Suzuki, I. and Hagemann, M. (2004) "Gene expression profiling reflects physiological processes in salt acclimation of *Synechocystis* sp. strain PCC 6803". *Plant Physiol*, **136** (2), 3290-300.

[205] Fulda, S., Huang, F., Nilsson, F., Hagemann, M. and Norling, B. (2000) "Proteomics of *Synechocystis* sp. strain PCC 6803. Identification of periplasmic proteins in cells grown at low and high salt concentrations". *Eur J Biochem*, **267** (19), 5900-7.

[206] Marin, K., Suzuki, I., Yamaguchi, K., Ribbeck, K., Yamamoto, H., Kanesaki, Y., Hagemann, M. and Murata, N. (2003) "Identification of histidine kinases that act as sensors in the perception of salt stress in *Synechocystis* sp. PCC 6803". *Proc Natl Acad Sci U S A*, **100** (15), 9061-6.

[207] Shoumskaya, M.A., Paithoonrangsarid, K., Kanesaki, Y., Los, D.A., Zinchenko, V.V., Tanticharoen, M., Suzuki, I. and Murata, N. (2005) "Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but

[219] Muro-Pastor, M.I., Reyes, J.C. and Florencio, F.J. (2001) "Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels". *J Biol Chem*, **276** (41), 38320-8.

[220] Osanai, T., Sato, S., Tabata, S. and Tanaka, K. (2005) "Identification of PamA as a PII-binding membrane protein important in nitrogen-related and sugar-catabolic gene expression in *Synechocystis* sp. PCC 6803". *J Biol Chem*, **280** (41), 34684-90.

[221] Hisbergues, M., Jeanjean, R., Joset, F., Tandeau de Marsac, N. and Bedu, S. (1999) "Protein PII regulates both inorganic carbon and nitrate uptake and is modified by a redox signal in *Synechocystis* PCC 6803". *FEBS Lett*, **463** (3), 216-20.

[222] Forchhammer, K. (2004) "Global carbon/nitrogen control by PII signal transduction in cyanobacteria: from signals to targets". *FEMS Microbiol Rev*, **28** (3), 319-33.

[223] Shenderova, L.V. and Venediktov, P.S. (1980) "[Pigment degradation in *Synechocystis aquatilis* under nitrogen starvation with varying illumination]". *Mikrobiologiia*, **49** (6), 906-10.

[224] Duke, C.S., Cezeaux, A. and Allen, M.M. (1989) "Changes in polypeptide composition of *Synechocystis* sp. strain 6308 phycobilisomes induced by nitrogen starvation". *J Bacteriol*, **171** (4), 1960-6.

[225] Duke, C.S. and Allen, M.M. (1990) "Effect of Nitrogen Starvation on Polypeptide Composition, Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase, and Thylakoid Carotenoprotein Content of *Synechocystis* sp. Strain PCC6308". *Plant Physiol*, **94** (2), 752-759.

[226] Reyes, J.C. and Florencio, F.J. (1994) "A new type of glutamine synthetase in cyanobacteria: the protein encoded by the glnN gene supports nitrogen assimilation in *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **176** (5), 1260-7.

[227] Reyes, J.C. and Florencio, F.J. (1994) "A mutant lacking the glutamine synthetase gene (glnA) is impaired in the regulation of the nitrate assimilation system in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **176** (24), 7516-23.

[228] Reyes, J.C., Muro-Pastor, M.I. and Florencio, F.J. (1997) "Transcription of glutamine synthetase genes (glnA and glnN) from the cyanobacterium *Synechocystis* sp. strain PCC 6803 is differently regulated in response to nitrogen availability". *J Bacteriol*, **179** (8), 2678-89.

[229] Muro-Pastor, M.I., Reyes, J.C. and Florencio, F.J. (1996) "The NADP+-isocitrate dehydrogenase gene (icd) is nitrogen regulated in cyanobacteria". *J Bacteriol*, **178** (14), 4070-6.

[230] Richaud, C., Zabulon, G., Joder, A. and Thomas, J.C. (2001) "Nitrogen or sulfur starvation differentially affects phycobilisome degradation and expression of the nblA gene in *Synechocystis* strain PCC 6803". *J Bacteriol*, **183** (10), 2989-94.

[244] Katoh, H., Hagino, N., Grossman, A.R. and Ogawa, T. (2001) "Genes essential to iron transport in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **183** (9), 2779-84.

[245] Katoh, H., Hagino, N. and Ogawa, T. (2001) "Iron-binding activity of FutA1 subunit of an ABC-type iron transporter in the cyanobacterium *Synechocystis* sp. Strain PCC 6803". *Plant Cell Physiol*, **42** (8), 823-7.

[246] Laboure, A.M. and Briat, J.F. (1993) "Uptake of iron from ferric-citrate in the cyanobacteria *Synechocystis* PCC6803". *C R Acad Sci III*, **316** (7), 661-6.

[247] Keren, N., Aurora, R. and Pakrasi, H.B. (2004) "Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria". *Plant Physiol*, **135** (3), 1666-73.

[248] Singh, A.K. and Sherman, L.A. (2000) "Identification of iron-responsive, differential gene expression in the cyanobacterium *Synechocystis* sp. strain PCC 6803 with a customized amplification library". *J Bacteriol*, **182** (12), 3536-43.

[249] Jeanjean, R., Zuther, E., Yeremenko, N., Havaux, M., Matthijs, H.C. and Hagemann, M. (2003) "A photosystem 1 psaFJ-null mutant of the cyanobacterium *Synechocystis* PCC 6803 expresses the isiAB operon under iron replete conditions". *FEBS Lett*, **549** (1-3), 52-6.

[250] Hirani, T.A., Suzuki, I., Murata, N., Hayashi, H. and Eaton-Rye, J.J. (2001) "Characterization of a two-component signal transduction system involved in the induction of alkaline phosphatase under phosphate-limiting conditions in *Synechocystis* sp. PCC 6803". *Plant Mol Biol*, **45** (2), 133-44.

[251] Morohoshi, T., Maruo, T., Shirai, Y., Kato, J., Ikeda, T., Takiguchi, N., Ohtake, H. and Kuroda, A. (2002) "Accumulation of inorganic polyphosphate in phoU mutants of *Escherichia coli* and *Synechocystis* sp. strain PCC6803". *Appl Environ Microbiol*, **68** (8), 4107-10.

[252] Gomez-Garcia, M.R., Losada, M. and Serrano, A. (2003) "Concurrent transcriptional activation of ppa and ppx genes by phosphate deprivation in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Biochem Biophys Res Commun*, **302** (3), 601-9.

[253] Suzuki, S., Ferjani, A., Suzuki, I. and Murata, N. (2004) "The SphS-SphR two component system is the exclusive sensor for the induction of gene expression in response to phosphate limitation in *Synechocystis*". *J Biol Chem*, **279** (13), 13234-40.

[254] Eymann, C., Mach, H., Harwood, C.R. and Hecker, M. (1996) "Phosphate-starvation-inducible proteins in *Bacillus subtilis*: a two-dimensional gel electrophoresis study". *Microbiology*, **142**, 3163-70.

[255] Antelmann, H., Scharf, C. and Hecker, M. (2000) "Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis". *J Bacteriol*, **182** (16), 4478-90.

[270] Simon, W.J., Hall, J.J., Suzuki, I., Murata, N. and Slabas, A.R. (2002) "Proteomic study of the soluble proteins from the unicellular cyanobacterium *Synechocystis* sp. PCC6803 using automated matrix-assisted laser desorption/ionization-time of flight peptide mass fingerprinting". *Proteomics*, **2** (12), 1735-42.

[271] Wang, Y., Sun, J. and Chitnis, P.R. (2000) "Proteomic study of the peripheral proteins from thylakoid membranes of the cyanobacterium *Synechocystis* sp. PCC 6803". *Electrophoresis*, **21** (9), 1746-54.

[272] Herranen, M., Battchikova, N., Zhang, P., Graf, A., Sirpio, S., Paakkarinen, V. and Aro, E.M. (2004) "Towards functional proteomics of membrane protein complexes in *Synechocystis* sp. PCC 6803". *Plant Physiol*, **134** (1), 470-81.

[273] Norling, B., Zak, E., Andersson, B. and Pakrasi, H. (1998) "2D-isolation of pure plasma and thylakoid membranes from the cyanobacterium *Synechocystis* sp. PCC 6803". *FEBS Lett*, **436** (2), 189-92.

[274] Huang, F., Hedman, E., Funk, C., Kieselbach, T., Schroder, W.P. and Norling, B. (2004) "Isolation of outer membrane of *Synechocystis* sp. PCC 6803 and its proteomic characterization". *Mol Cell Proteomics*, **3** (6), 586-95.

[275] Huang, F., Parmryd, I., Nilsson, F., Persson, A.L., Pakrasi, H.B., Andersson, B. and Norling, B. (2002) "Proteomics of *Synechocystis* sp. strain PCC 6803: identification of plasma membrane proteins". *Mol Cell Proteomics*, **1** (12), 956-66.

[276] Srivastava, R., Pisareva, T. and Norling, B. (2005) "Proteomic studies of the thylakoid membrane of *Synechocystis* sp. PCC 6803". *Proteomics*, **5** (18), 4905-16.

[277] Kurian, D., Phadwal, K. and Maenpaa, P. (2006) "Proteomic characterization of acid stress response in *Synechocystis* sp. PCC 6803". *Proteomics*, **6** (12), 3614-24.

[278] Choi, J.S., Kim, D.S., Lee, J., Kim, S.J., Kim, S.I., Kim, Y.H., Hong, J., Yoo, J.S., Suh, K.H. and Park, Y.M. (2000) "Proteome analysis of light-induced proteins in *Synechocystis* sp. PCC 6803: identification of proteins separated by 2D-PAGE using N-terminal sequencing and MALDI-TOF MS". *Mol Cells*, **10** (6), 705-11.

[279] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999) "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags". *Nat. Biotechnol.*, **17** (10), 994-9.

[280] Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R., 3rd (1999) "Direct analysis of protein complexes using mass spectrometry". *Nat. Biotechnol.*, **17** (7), 676-82.

[281] Washburn, M.P., Wolters, D. and Yates, J.R., 3rd (2001) "Large-scale analysis of the yeast proteome by multidimensional protein identification technology". *Nat. Biotechnol.*, **19** (3), 242-7.

isolation of overexpressed proteins and proteome mapping". *Anal. Biochem.*, **258** (2), 349-61.

[295] Gao, J., Opiteck, G.J., Friedrichs, M.S., Dongre, A.R. and Hefta, S.A. (2003) "Changes in the protein expression of yeast as a function of carbon source". *J. Proteome Res.*, **2** (6), 643-9.

[296] Lecchi, P., Gupte, A.R., Perez, R.E., Stockert, L.V. and Abramson, F.P. (2003) "Size-exclusion chromatography in multidimensional separation schemes for proteome analysis". *J. Biochem. Biophys. Methods*, **56** (1-3), 141-52.

[297] Karlsson, K., Cairns, N., Lubec, G. and Fountoulakis, M. (1999) "Enrichment of human brain proteins by heparin chromatography". *Electrophoresis*, **20** (14), 2970-6.

[298] Swanson, R.V. and Glazer, A.N. (1990) "Separation of phycobiliprotein subunits by reverse-phase high-pressure liquid chromatography". *Anal. Biochem.*, **188** (2), 295-9.

[299] Zolla, L. and Bianchetti, M. (2001) "High-performance liquid chromatography coupled on-line with electrospray ionization mass spectrometry for the simultaneous separation and identification of the *Synechocystis* PCC 6803 phycobilisome proteins". *J Chromatogr A*, **912** (2), 269-79.

[300] Butt, A., Davison, M.D., Smith, G.J., Young, J.A., Gaskell, S.J., Oliver, S.G. and Beynon, R.J. (2001) "Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis". *Proteomics*, **1** (1), 42-53.

[301] Kashino, Y. (2003) "Separation methods in the analysis of protein membrane complexes". *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **797** (1-2), 191-216.

[302] Kato, Y., Nakamura, K., Kitamura, T., Tsuda, T., Hasegawa, M. and Sasaki, H. (2004) "Effect of chromatographic conditions on resolution in high-performance ion-exchange chromatography of proteins on macroporous anion-exchange resin". *J. Chromatogr. A*, **1031** (1-2), 101-5.

[303] Linke, T., Ross, A.C. and Harrison, E.H. (2004) "Profiling of rat plasma by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry, a novel tool for biomarker discovery in nutrition research". *J. Chromatogr. A*, **1043** (1), 65-71.

[304] Herbert, B. and Righetti, P.G. (2000) "A turning point in proteome analysis: sample prefractionation via multicompartment electrolyzers with isoelectric membranes". *Electrophoresis*, **21** (17), 3639-48.

[305] Gorg, A., Boguth, G., Kopf, A., Reil, G., Parlar, H. and Weiss, W. (2002) "Sample prefractionation with Sephadex isoelectric focusing prior to narrow pH range two-dimensional gels". *Proteomics*, **2** (12), 1652-7.

[306] Giorgianni, F., Desiderio, D.M. and Beranova-Giorgianni, S. (2003) "Proteome analysis using isoelectric focusing in immobilized pH gradient gels followed by mass spectrometry". *Electrophoresis*, **24** (1-2), 253-9.

profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry". *Mol Cell Proteomics*, **3** (10), 1039-41.

[319] Molloy, M.P., Donohoe, S., Brzezinski, E.E., Kilby, G.W., Stevenson, T.I., Baker, J.D., Goodlett, D.R. and Gage, D.A. (2005) "Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling". *Proteomics*, **5** (5), 1204-8.

[320] Snijders, A.P., de Vos, M.G., de Koning, B. and Wright, P.C. (2005) "A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and 15N-metabolic labelling". *Electrophoresis*, **26** (16), 3191-9.

[321] Snijders, A.P., de Koning, B. and Wright, P.C. (2005) "Perturbation and interpretation of nitrogen isotope distribution patterns in proteomics". *J Proteome Res*, **4** (6), 2185-91.

[322] Cagney, G. and Emili, A. (2002) "De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging". *Nat Biotechnol*, **20** (2), 163-70.

[323] Swatton, J.E., Prabakaran, S., Karp, N.A., Lilley, K.S. and Bahn, S. (2004) "Protein profiling of human postmortem brain using 2-dimensional fluorescence difference gel electrophoresis (2-D DIGE)". *Mol Psychiatry*, **9** (2), 128-43.

[324] Karp, N.A., Kreil, D.P. and Lilley, K.S. (2004) "Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis". *Proteomics*, **4** (5), 1421-32.

[325] Karp, N.A. and Lilley, K.S. (2005) "Maximising sensitivity for detecting changes in protein expression: experimental design using minimal CyDyes". *Proteomics*, **5** (12), 3105-15.

[326] Aggarwal, K., Choe, L.H. and Lee, K.H. (2006) "Shotgun proteomics using the iTRAQ isobaric tags". *Brief Funct Genomic Proteomic*, **5** (2), 112-20.

[327] Overall, C.M. and Dean, R.A. (2006) "Degradomics: systems biology of the protease web. Pleiotropic roles of MMPs in cancer". *Cancer Metastasis Rev*, **25** (1), 69-75.

[328] Tannu, N.S. and Hemby, S.E. (2006) "Chapter 3 Methods for proteomics in neuroscience". *Prog Brain Res*, **158**, 41-82.

[329] D'Ascenzo, M., Relkin, N.R. and Lee, K.H. (2005) "Alzheimer's disease cerebrospinal fluid biomarker discovery: A proteomics approach". *Current Opinion in Molecular Therapeutics*, **7** (6), 557-564.

[330] Zieske, L.R. (2006) "A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies". *J Exp Bot*, **57** (7), 1501-8.

[331] Chen, X., Walker, A.K., Strahler, J.R., Simon, E.S., Tomanicek-Volk, S.L., Nelson, B.B., Hurley, M.C., Ernst, S.A., Williams, J.A. and Andrews, P.C. (2006) "Organellar proteomics: analysis of pancreatic zymogen granule membranes". *Mol Cell Proteomics*, **5** (2), 306-12.

[342] DeSouza, L., Diehl, G., Rodrigues, M.J., Guo, J., Romaschin, A.D., Colgan, T.J. and Siu, K.W. (2005) "Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry". *J Proteome Res*, **4** (2), 377-86.

[343] Hardt, M., Witkowska, H.E., Webb, S., Thomas, L.R., Dixon, S.E., Hall, S.C. and Fisher, S.J. (2005) "Assessing the effects of diurnal variation on the composition of human parotid saliva: quantitative analysis of native peptides using iTRAQ reagents". *Anal Chem*, **77** (15), 4947-54.

[344] Reinders, Y., Schulz, I., Graf, R. and Sickmann, A. (2006) "Identification of novel centrosomal proteins in *Dictyostelium discoideum* by comparative proteomic approaches". *J Proteome Res*, **5** (3), 589-98.

[345] Lee, J., Cao, L., Ow, S.Y., Barrios-Llerena, M.E., Chen, W., Wood, T.K. and Wright, P.C. (2006) "Proteome changes after metabolic engineering to enhance aerobic mineralization of cis-1,2-dichloroethylene". *J Proteome Res*, **5** (6), 1388-97.

[346] Aggarwal, K., Choe, L.H. and Lee, K.H. (2005) "Quantitative analysis of protein expression using amine-specific isobaric tags in *Escherichia coli* cells expressing rhsA elements". *Proteomics*, **5** (9), 2297-2308.

[347] Rudella, A., Friso, G., Alonso, J.M., Ecker, J.R. and van Wijk, K.J. (2006) "Downregulation of ClpR2 Leads to Reduced Accumulation of the ClpPRS Protease Complex and Defects in Chloroplast Biogenesis in *Arabidopsis*". *Plant Cell*, **18** (7), 1704-21.

[348] Jones, A.M., Bennett, M.H., Mansfield, J.W. and Grant, M. (2006) "Analysis of the defence phosphoproteome of *Arabidopsis thaliana* using differential mass tagging". *Proteomics*, **6** (14), 4155-65.

[349] Dunkley, T.P., Hester, S., Shadforth, I.P., Runions, J., Weimar, T., Hanton, S.L., Griffin, J.L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R.B., Dupree, P. and Lilley, K.S. (2006) "Mapping the *Arabidopsis* organelle proteome". *Proc Natl Acad Sci U S A*, **103** (17), 6518-23.

[350] Corvey, C., Koetter, P., Beckhaus, T., Hack, J., Hofmann, S., Hampel, M., Stein, T., Karas, M. and Entian, K.D. (2005) "Carbon source-dependent assembly of the Snf1p kinase complex in *Candida albicans*". *Journal of Biological Chemistry*, **280** (27), 25323-25330.

[351] Pham, T.K., Chong, P.K., Gan, C.S. and Wright, P.C. (2006) "Proteomic Analysis of *Saccharomyces cerevisiae* under High Gravity Fermentation Condition". *J. Proteome Res.*, **5** (12), 3411-19.

[352] Williamson, B.L., Marchese, J. and Morrice, N.A. (2006) "Automated identification and quantification of protein phosphorylation sites by LC/MS on a hybrid triple quadrupole linear ion trap mass spectrometer". *Mol Cell Proteomics*, **5** (2), 337-346.

proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins". *Proteomics*, **3** (7), 1345-64.

[365] Steel, L.F., Trotter, M.G., Nakajima, P.B., Mattu, T.S., Gonye, G. and Block, T. (2003) "Efficient and specific removal of albumin from human serum samples". *Mol. Cell Proteomics*, **2** (4), 262-70.

[366] Maccarrone, G., Milfay, D., Birg, I., Rosenhagen, M., Holsboer, F., Grimm, R., Bailey, J., Zolotarjova, N. and Turck, C.W. (2004) "Mining the human cerebrospinal fluid proteome by immunodepletion and shotgun mass spectrometry". *Electrophoresis*, **25** (14), 2402-12.

[367] Chan, L.L., Lo, S.C. and Hodgkiss, I.J. (2002) "Proteomic study of a model causative agent of harmful red tide, *Prorocentrum triestinum* I: Optimization of sample preparation methodologies for analyzing with two-dimensional electrophoresis". *Proteomics*, **2** (9), 1169-1186.

[368] Moryion, I. and Berman, D.T. (1984) "Interaction of Escherichia coli matrix protein with Brucella abortus peptidoglycan and chitin". *Dev Biol Stand*, **56**, 227-34.

[369] Herbert, B.R., Grinyer, J., McCarthy, J.T., Isaacs, M., Harry, E.J., Nevalainen, H., Traini, M.D., Hunt, S., Schulz, B., Laver, M., Goodall, A.R., Packer, J., Harry, J.L. and Williams, K.L. (2006) "Improved 2-DE of microorganisms after acidic extraction". *Electrophoresis*, **27** (8), 1630-40.

[370] Charmont, S., Jamet, E., Pont-Lezica, R. and Canut, H. (2005) "Proteomic analysis of secreted proteins from Arabidopsis thaliana seedlings: improved recovery following removal of phenolic compounds". *Phytochemistry*, **66** (4), 453-61.

[371] Russell, W.K., Park, Z.Y. and Russell, D.H. (2001) "Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry". *Anal. Chem.*, **73** (11), 2682-5.

[372] Shi, Y., Xiang, R., Horvath, C. and Wilkins, J.A. (2004) "The role of liquid chromatography in proteomics". *J Chromatogr A*, **1053** (1-2), 27-36.

[373] Cargile, B.J., Talley, D.L. and Stephenson, J.L., Jr. (2004) "Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides". *Electrophoresis*, **25** (6), 936-45.

[374] Shefcheck, K., Yao, X. and Fenselau, C. (2003) "Fractionation of cytosolic proteins on an immobilized heparin column". *Anal. Chem.*, **75** (7), 1691-8.

[375] Hiller, K., Schobert, M., Hundertmark, C., Jahn, D. and Munch, R. (2003) "JVirGel: Calculation of virtual two-dimensional protein gels". *Nucleic Acids Res*, **31** (13), 3862-5.

[389] Chong, P.K. and Wright, P.C. (2005) "Identification and characterization of the *Sulfolobus solfataricus* P2 proteome". *J Proteome Res*, **4** (5), 1789-98.

[390] Choe, L.H. and Lee, K.H. (2003) "Quantitative and qualitative measure of intralaboratory two-dimensional protein gel reproducibility and the effects of sample preparation, sample load, and image analysis". *Electrophoresis*, **24** (19-20), 3500-7.

[391] Newton, M.A., Kendziorski, C.M., Richmond, C.S., Blattner, F.R. and Tsui, K.W. (2001) "On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data". *J Comput Biol*, **8** (1), 37-52.

[392] Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C. and Wong, W.H. (2001) "Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects". *Nucleic Acids Res*, **29** (12), 2549-57.

[393] Kerr, M.K. and Churchill, G.A. (2001) "Experimental design for gene expression microarrays". *Biostatistics*, **2** (2), 183-201.

[394] Kerr, M.K. and Churchill, G.A. (2001) "Statistical design and the analysis of gene expression microarray data". *Genet Res*, **77** (2), 123-8.

[395] Kerr, M.K., Martin, M. and Churchill, G.A. (2000) "Analysis of variance for gene expression microarray data". *J Comput Biol*, **7** (6), 819-37.

[396] Kendziorski, C.M., Newton, M.A., Lan, H. and Gould, M.N. (2003) "On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles". *Stat Med*, **22** (24), 3899-914.

[397] Balagurunathan, Y., Wang, N., Dougherty, E.R., Nguyen, D., Chen, Y., Bittner, M.L., Trent, J. and Carroll, R. (2004) "Noise factor analysis for cDNA microarrays". *J Biomed Opt*, **9** (4), 663-78.

[398] Bammler, T., Beyer, R.P., Bhattacharya, S., Boorman, G.A., Boyles, A., Bradford, B.U., Bumgarner, R.E., Bushel, P.R., Chaturvedi, K., Choi, D., Cunningham, M.L., Deng, S., Dressman, H.K., Fannin, R.D., Farin, F.M., Freedman, J.H., Fry, R.C., Harper, A., Humble, M.C., Hurban, P., Kavanagh, T.J., Kaufmann, W.K., Kerr, K.F., Jing, L., Lapidus, J.A., Lasarev, M.R., Li, J., Li, Y.J., Lobenhofer, E.K., Lu, X., Malek, R.L., Milton, S., Nagalla, S.R., O'Malley J, P., Palmer, V.S., Pattee, P., Paules, R.S., Perou, C.M., Phillips, K., Qin, L.X., Qiu, Y., Quigley, S.D., Rodland, M., Rusyn, I., Samson, L.D., Schwartz, D.A., Shi, Y., Shin, J.L., Sieber, S.O., Slifer, S., Speer, M.C., Spencer, P.S., Sproles, D.I., Swenberg, J.A., Suk, W.A., Sullivan, R.C., Tian, R., Tennant, R.W., Todd, S.A., Tucker, C.J., Van Houten, B., Weis, B.K., Xuan, S. and Zarbl, H. (2005) "Standardizing global gene expression analysis between laboratories and across platforms". *Nat Methods*, **2** (5), 351-6.

[399] Lee, M.L., Kuo, F.C., Whitmore, G.A. and Sklar, J. (2000) "Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations". *Proc Natl Acad Sci U S A*, **97** (18), 9834-9.

[412] Gygi, S.P. and Aebersold, R. (1999) "Absolute quantitation of 2-D protein spots". *Methods Mol. Biol.*, **112**, 417-21.

[413] Smith, A.J., London, J. and Stanier, R.Y. (1967) "Biochemical basis of obligate autotrophy in blue-green algae and thiobacilli". *J Bacteriol*, **94** (4), 972-83.

[414] Schmitt, W.A., Jr. and Stephanopoulos, G. (2003) "Prediction of transcriptional profiles of *Synechocystis* PCC6803 by dynamic autoregressive modeling of DNA microarray data". *Biotechnol Bioeng*, **84** (7), 855-63.

[415] Denz, C.R. and Dube, D.K. (2005) "The benefits of 28S rRNA for standardization of reverse transcription-polymerase chain reaction for studying gene expression". *Anal Biochem*, **341** (2), 382-4.

[416] Wedel, N. and Soll, J. (1998) "Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/ glyceraldehyde-3-phosphate dehydrogenase complex dissociation". *Proc Natl Acad Sci U S A*, **95** (16), 9699-704.

[417] Watson, G.M. and Tabita, F.R. (1996) "Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium". *Plant Mol Biol*, **32** (6), 1103-15.

[418] Koksharova, O., Schubert, M., Shestakov, S. and 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 Mol Biol*, **36** (1), 183-94.

[419] Sundaram, S., Karakaya, H., Scanlan, D.J. and Mann, N.H. (1998) "Multiple oligomeric forms of glucose-6-phosphate dehydrogenase in cyanobacteria and the role of OpcA in the assembly process". *Microbiology*, **144 (Pt 6)**, 1549-56.

[420] Smart, L.B. and McIntosh, L. (1991) "Expression of photosynthesis genes in the cyanobacterium *Synechocystis* sp. PCC 6803: psaA-psaB and psbA transcripts accumulate in dark-grown cells". *Plant Mol Biol*, **17** (5), 959-71.

[421] Zhang, L., McSpadden, B., Pakrasi, H.B. and Whitmarsh, J. (1992) "Copper-mediated regulation of cytochrome c553 and plastocyanin in the cyanobacterium *Synechocystis* 6803". *J Biol Chem*, **267** (27), 19054-9.

[422] Zhang, L., Pakrasi, H.B. and Whitmarsh, J. (1994) "Photoautotrophic growth of the cyanobacterium *Synechocystis* sp. PCC 6803 in the absence of cytochrome c553 and plastocyanin". *J Biol Chem*, **269** (7), 5036-42.

[423] Samartzidou, H. and Widger, W.R. (1998) "Transcriptional and posttranscriptional control of mRNA from lrtA, a light-repressed transcript in *Synechococcus* sp. PCC 7002". *Plant Physiol*, **117** (1), 225-34.

[436] Dunlap, J.C. (1999) "Molecular bases for circadian clocks". *Cell*, **96** (2), 271-90.

[437] Golden, S.S., Ishiura, M., Johnson, C.H. and Kondo, T. (1997) "Cyanobacterial Circadian Rhythms". *Annu Rev Plant Physiol Plant Mol Biol*, **48**, 327-354.

[438] Kondo, T. and Ishiura, M. (1999) "The circadian clocks of plants and cyanobacteria". *Trends Plant Sci*, **4** (5), 171-176.

[439] Woelfle, M.A., Ouyang, Y., Phanvijhitsiri, K. and Johnson, C.H. (2004) "The adaptive value of circadian clocks: an experimental assessment in cyanobacteria". *Curr Biol*, **14** (16), 1481-6.

[440] Rensing, L. and Monnerjahn, C. (1996) "Heat shock proteins and circadian rhythms". *Chronobiol Int*, **13** (4), 239-50.

[441] Aoki, S., Kondo, T. and Ishiura, M. (1995) "Circadian expression of the dnaK gene in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **177** (19), 5606-11.

[442] Agrawal, G.K., Asayama, M. and Shirai, M. (1999) "Light-dependent and rhythmic psbA transcripts in homologous/heterologous cyanobacterial cells". *Biochem Biophys Res Commun*, **255** (1), 47-53.

[443] Aoki, S., Kondo, T., Wada, H. and Ishiura, M. (1997) "Circadian rhythm of the cyanobacterium *Synechocystis* sp. strain PCC 6803 in the dark". *J Bacteriol*, **179** (18), 5751-5.

[444] Kucho, K., Okamoto, K., Tsuchiya, Y., Nomura, S., Nango, M., Kanehisa, M. and Ishiura, M. (2005) "Global analysis of circadian expression in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **187** (6), 2190-9.

[445] Wagner, V., Fiedler, M., Markert, C., Hippler, M. and Mittag, M. (2004) "Functional proteomics of circadian expressed proteins from *Chlamydomonas reinhardtii*". *FEBS Lett*, **559** (1-3), 129-35.

[446] Akimoto, H., Wu, C., Kinumi, T. and Ohmiya, Y. (2004) "Biological rhythmicity in expressed proteins of the marine dinoflagellate *Lingulodinium polyedrum* demonstrated by chronological proteomics". *Biochem Biophys Res Commun*, **315** (2), 306-12.

[447] Reddy, A.B., Karp, N.A., Maywood, E.S., Sage, E.A., Deery, M., O'Neill, J.S., Wong, G.K., Chesham, J., Odell, M., Lilley, K.S., Kyriacou, C.P. and Hastings, M.H. (2006) "Circadian orchestration of the hepatic proteome". *Curr Biol*, **16** (11), 1107-15.

[448] Hihara, Y., Sonoike, K. and Ikeuchi, M. (1998) "A novel gene, pmgA, specifically regulates photosystem stoichiometry in the cyanobacterium *Synechocystis* species PCC 6803 in response to high light". *Plant Physiol*, **117** (4), 1205-16.

[449] Murakami, A. and Fujita, Y. (1991) "Regulation of Photosystem Stoichiometry in the Photosynthetic System of the Cyanophyte *Synechocystis* PCC 6714 in Response to Light-Intensity". *Plant Cell Physiol*, **32** (2), 223-230.

[450] Navarro, F., Martin-Figueroa, E., Candau, P. and Florencio, F.J. (2000) "Ferredoxin-dependent iron-sulfur flavoprotein glutamate synthase (GlsF) from the Cyanobacterium

[464] Allenby, N.E., O'Connor, N., Pragai, Z., Ward, A.C., Wipat, A. and Harwood, C.R. (2005) "Genome-wide transcriptional analysis of the phosphate starvation stimulon of *Bacillus subtilis*". *J Bacteriol*, **187** (23), 8063-80.

[465] Burja, A.M. (2003). *Metabolic engineering of marine microbes: the potential of Lyngbya majuscula as a natural products cell factory*. Thesis (Ph.D.). Heriot-Watt University.

[466] O'Loughlin, S.N., Graham, R.L.J., McMullan, G. and Ternan, N.G. (2006) "A role for carbon catabolite repression in the metabolism of phosphonoacetate by *Agromyces fucosus* Vs2". *FEMS Microbiol Lett*, **261**, 133-140.

[467] Fiske, C.H. and SubbaRow, Y. (1925) "The colourimetric determination of phosphorus". *J Biol Chem*, **66**, 375-400.

[468] Ticconi, C.A., Delatorre, C.A. and Abel, S. (2001) "Attenuation of phosphate starvation responses by phosphite in *Arabidopsis*". *Plant Physiol*, **127** (3), 963-72.

[469] Osorio, G. and Jerez, C.A. (1996) "Adaptive response of the archaeon *Sulfolobus acidocaldarius* BC65 to phosphate starvation". *Microbiology*, **142 (Pt 6)**, 1531-6.

[470] Ray, J.M., Bhaya, D., Block, M.A. and Grossman, A.R. (1991) "Isolation, transcription, and inactivation of the gene for an atypical alkaline phosphatase of *Synechococcus* sp. strain PCC 7942". *J Bacteriol*, **173** (14), 4297-309.

[471] Doonan, B.B. and Jensen, T.E. (1980) "Physiological aspects of alkaline phosphatase in selected cyanobacteria". *Microbios*, **29** (117-118), 185-207.

[472] Bacsi, I., Vasas, G., Suranyi, G., M, M.H., Mathe, C., Toth, E., Grigorszky, I., Gaspar, A., Toth, S. and Borbely, G. (2006) "Alteration of cylindrospermopsin production in sulfate- or phosphate-starved cyanobacterium *Aphanizomenon ovalisporum*". *FEMS Microbiol Lett*, **259** (2), 303-10.

[473] Dodd, H.N. and Pemberton, J.M. (1996) "Cloning, sequencing, and characterization of the nucH gene encoding an extracellular nuclease from *Aeromonas hydrophila* JMP636". *J Bacteriol*, **178** (13), 3926-33.

[474] Gaber, A., Yoshimura, K., Tamoi, M., Takeda, T., Nakano, Y. and Shigeoka, S. (2004) "Induction and functional analysis of two reduced nicotinamide adenine dinucleotide phosphate-dependent glutathione peroxidase-like proteins in *Synechocystis* PCC 6803 during the progression of oxidative stress". *Plant Physiol*, **136** (1), 2855-61.

[475] Kobayashi, M., Ishizuka, T., Katayama, M., Kanehisa, M., Bhattacharyya-Pakrasi, M., Pakrasi, H.B. and Ikeuchi, M. (2004) "Response to oxidative stress involves a novel peroxiredoxin gene in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803". *Plant Cell Physiol*, **45** (3), 290-9.

[476] Navarro, F., Martin-Figueroa, E. and Florencio, F.J. (2000) "Electron transport controls transcription of the thioredoxin gene (trxA) in the cyanobacterium *Synechocystis* sp. PCC 6803". *Plant Mol Biol*, **43** (1), 23-32.

[489] Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M. and Murata, N. (2001) "Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*". *Mol Microbiol*, **40** (1), 235-44.

[490] Mikami, K., Kanesaki, Y., Suzuki, I. and Murata, N. (2002) "The histidine kinase Hik33 perceives osmotic stress and cold stress in *Synechocystis* sp PCC 6803". *Mol Microbiol*, **46** (4), 905-15.

[491] Kahlon, S., Beeri, K., Ohkawa, H., Hihara, Y., Murik, O., Suzuki, I., Ogawa, T. and Kaplan, A. (2006) "A putative sensor kinase, Hik31, is involved in the response of *Synechocystis* sp. strain PCC 6803 to the presence of glucose". *Microbiology*, **152** (Pt 3), 647-55.

[492] Panda, B. and Mallick, N. (2006) "Enhanced poly-b-hydroxybutyrate accumulation in a unicellular cyanobacterium *Synechocystis* sp. PCC 6803". *Letters in Applied Microbiology*, doi:10.1111/j.1472-765X.2006.02048.x.

[493] Wu, P., Ma, L., Hou, X., Wang, M., Wu, Y., Liu, F. and Deng, X.W. (2003) "Phosphate starvation triggers distinct alterations of genome expression in *Arabidopsis* roots and leaves". *Plant Physiol*, **132** (3), 1260-71.

[494] Piven, I., Ajlani, G. and Sokolenko, A. (2005) "Phycobilisome linker proteins are phosphorylated in *Synechocystis* sp. PCC 6803". *J Biol Chem*, **280** (22), 21667-72.

[495] Ethier, M. and Figeys, D. (2005) "Strategy to design improved proteomic experiments based on statistical analyses of the chemical properties of identified peptides". *J Proteome Res*, **4** (6), 2201-6.

[496] Barrios-Llerena, M.E., Chong, P.K., Gan, C.S., Snijders, A.P., Reardon, K.F. and Wright, P.C. (2006) "Shotgun proteomics of cyanobacteria--applications of experimental and data-mining techniques". *Brief Funct Genomic Proteomic*, **5** (2), 121-32.

[497] Zimmerman, W.B. (2005) "Metabolic pathways reconstruction by frequency and amplitude response to forced glycolytic oscillations in yeast". *Biotechnol Bioeng*, **92** (1), 91-116.

Appendices

Appendix A

Due to large amount of supporting information available for the study conducted in Chapter 3, all these information will be made available in a soft-copy (CD format), attached to the back of this thesis, rather than as hard-copy.

File name : Appendix A
File type : Microsoft Excel
Description : The master list of total proteins identified.

The attached Excel file contains numerous supplementary documents. Following are the brief description of each sub-document within Appendix A.

	Excel sub-document	**Sub-document Description**
(1)	Total	The list of 776 total proteins found in all 6 protein peptide prefractionation workflows.
(2)	IEF-IEF 4.4 NR	The list of proteins found by injecting the same sample (4.4) twice into the MS. The protein identification is based on non-redundant database (NR).
(3)	IEF-IEF 4.4 SYN	The list of proteins found by injecting the same sample (4.4) twice into the MS. The protein identification is based on *Synechocystis* sp. PCC 6803 database (SYN).
(4)	TMH	The list of 123 TMH proteins, as predicted using the TMHMM program.
(5)	Literature	The list of proteins found from literatures.
(6)	Master List	The list of 958 total proteins found from the literatures and current study.

Appendix C

Due to large amount of supporting information available for the study conducted in Chapter 3. all these information will be made available in a soft-copy (CD format), attached to the back of this thesis, rather than as hard-copy.

File name : Appendix C
File type : Microsoft Excel
Description : The total peptides lists identified in each workflow

The lists of peptides found in each protein-peptide prefractionation workflows, as well as the use of difference search databases are included in the CD. Following are the brief description of each sub-document within Appendix C.

	Excel sub-document	**Workflows used**	**Database used ***
(1)	SCX NR	SCX-only	NR
(2)	SCX SYN	SCX-only	SYN
(3)	IEF NR	IEF-only	NR
(4)	IEF SYN	IEF-only	SYN
(5)	1D-SCX NR	1D-SCX	NR
(6)	1D-SCX SYN	1D-SCX	SYN
(7)	WAX-SCX NR	WAX-SCX	NR
(8)	WAX-SCX SYN	WAX-SCX	SYN
(9)	IEF-IEF NR	IEF-IEF	NR
(10)	IEF-IEF SYN	IEF-IEF	SYN
(11)	Whole NR	Whole-cell lysate	NR
(12)	Whole SYN	Whole-cell lysate	SYN
(13)	SCX 40 NR	SCX-only (40 fractions)	NR
(14)	SCX 40 SYN	SCX-only (40 fractions)	SYN
(15)	Total	All workflows	NR and SYN

Note: * NR refers to non-redundant database
SYN refers to *Synechocystis* sp. PCC 6803 single genome database

Appendix E

Example Calculation of Quantitative iTRAQ Ratio

Note: All the equation and calculation steps here are adapted from the ProQuant Software Tutorial (Applied Biosystems).

Protein name : Photosystem I subunit II, PsaD (slr0737)

Species : *Synechocystis* sp. PCC 6803

Source : Experiment 6 in Chapter 4

The confidently identified PsaD peptides with their quantitation ratio (Ratio 115:114) and the % Error are listed as below.

	Peptide sequence	% Confidence	Locus	Ratio 115:114	% Error
1)	IGQNPEPVTIJ	99	4.1.807.2	0.3638	5.81
2)	IGQNPEPVTIJ	99	5.1.664.2	0.3766	4.73
3)	IGQNPEPVTIJ	96	4.1.728.3	0.3863	9.81
4)	IGQNPEPVTIJ	99	9.1.532.3	0.4020	7.13
5)	IGQNPEPVTIJ	99	5.1.817.2	0.4523	7.77
6)	IGQNPEPVTIJ	90	10.1.596.3	0.6016	8.85
7)	VYPSGEVQYLHPADGVFPEJ	86	6.1.1007.2	0.7468	13.25
8)	VYPSGEVQYLHPADGVFPEJ	95	7.1.922.2	0.7708	8.11

First of all, we need to convert the working space into the log space, such as $\log_{10}$. Here we define **x** as the $\log_{10}$ (peptide ratio).

	Peptide sequence	Ratio 115:114	x
1)	IGQNPEPVTIJ	0.3638	-0.4391
2)	IGQNPEPVTIJ	0.3766	-0.4241
3)	IGQNPEPVTIJ	0.3863	-0.4131
4)	IGQNPEPVTIJ	0.4020	-0.3958
5)	IGQNPEPVTIJ	0.4523	-0.3446
6)	IGQNPEPVTIJ	0.6016	-0.2207
7)	VYPSGEVQYLHPADGVFPEJ	0.7468	-0.1268
8)	VYPSGEVQYLHPADGVFPEJ	0.7708	-0.1131
	Average	**0.5125**	**-0.3097**

Then, we could determine the weighted mean of log ratio, $\mathbf{Y_{log}}$, using the following formula:

$$Y_{log} = \frac{\sum_{i=1}^{N}(w_i \times x_i)}{\sum_{i=1}^{N} w_i}, \text{ where } w_i = 1/\%\ \text{Error} \quad \ldots\ldots\ldots \text{(Equation E1)}$$

And **b** $= (1.0661)^2 / 0.1546$

$= 7.3517$

Using the effective base, **b**, we could determine the weighted standard deviation, $\mathbf{SD_w}$.

Weighted standard deviation, $\mathbf{SD_w}$ $= \frac{SD}{b^{0.5}}$ ..... (Equation E4)

$= 0.1359 / (7.3517)^{0.5}$

$= \underline{0.0501}$

The antilog_{10} of $\mathbf{SD_w}$ is known as error factor, **EF** = Antilog_{10} (0.0501)

$= \underline{1.1223}$

After determining the relative protein ratio and its standard deviation, it is important to bear in mind reporting the value in log space and linear space is different from each other.

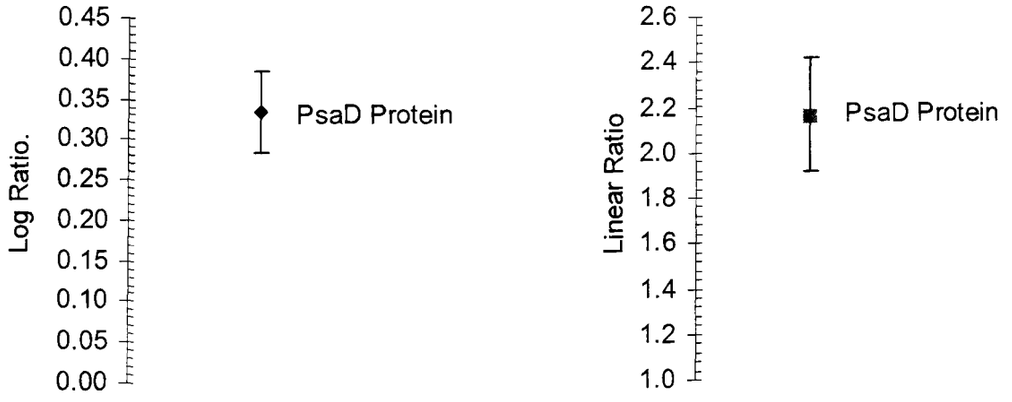
In log space $\Rightarrow$ -0.3336 ± 0.0501

In linear space $\Rightarrow$ -2.16-fold with EF of 1.12, or (-2.42, -1.92)*

* The upper limit of the ratio is determined by -2.16 × 1.1223 = -2.42

* The lower limit of the ratio is determined by -2.16 ÷ 1.1223 = -1.92

Examples of PsaD protein when plotting on the graphs, with its error/standard deviation, in different spaces (log and space).



ORF	Protein Name	MS/MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
slr1839	carbon dioxide concentrating mechanism protein CcmK homolog 4, putative carboxysome assembly protein	3	0.331	0.197	0.314	0.210	0.322	0.203
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	15	0.294	0.082	0.312	0.081	0.303	0.082
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	57	0.138	0.029	0.168	0.030	0.153	0.030
sll1031	carbon dioxide concentrating mechanism protein CcmM, putative carboxysome structural protein ccmM	4	-0.072	0.062	-0.065	0.065	-0.068	0.063
sll1028	carbon dioxide concentrating mechanism protein CcmK ccmK2	26	-0.058	0.027	-0.115	0.020	-0.086	0.023
Cellular Processes								
slr0374	cell division cycle protein	2	0.438	0.227	0.471	0.151	0.454	0.189
sll1712	DNA binding protein HU	28	0.355	0.062	0.376	0.055	0.365	0.058
slr1198	antioxidant protein	50	0.236	0.058	0.259	0.055	0.247	0.056
slr1516	superoxide dismutase sodB	54	-0.029	0.011	-0.014	0.011	-0.021	0.011
slr0757	circadian clock protein KaiB homolog kaiB1	2	0.113	0.004	0.185	0.057	0.149	0.031
ssl0707	nitrogen regulatory protein P-II glnB	19	0.008	0.068	0.002	0.070	0.005	0.069
slr1894	probable DNA-binding stress protein	8	-0.032	0.025	-0.010	0.012	-0.021	0.019
slr0925	single-stranded DNA-binding protein ssb	2	-0.043	0.051	-0.017	0.094	-0.030	0.072
sll1463	cell division protein FtsH ftsH	2	-0.430	0.021	-0.396	0.045	-0.413	0.033
Energy Metabolism								
slr1329	ATP synthase beta subunit atpB	8	0.168	0.096	0.198	0.105	0.183	0.101
sll1326	ATP synthase alpha chain atpA	23	0.000	0.037	0.014	0.036	0.007	0.036
slr0898	ferredoxin--nitrite reductase nirA	6	0.039	0.084	0.059	0.070	0.049	0.077
sll1499	ferredoxin-dependent glutamate synthase glsF	5	-0.001	0.065	0.038	0.060	0.018	0.062
sll1987	catalase peroxidase cpx, katG	10	0.010	0.036	0.019	0.045	0.015	0.040
Folding, Sorting and Degradation								
slr0164	ATP-dependent Clp protease proteolytic subunit clpP4	8	0.365	0.091	0.405	0.065	0.385	0.078
slr2076	60kD chaperonin groEL1	23	0.295	0.052	0.309	0.054	0.302	0.053
sll0416	60 kDa chaperonin 2, GroEL2, molecular chaperone groEL-2	18	0.250	0.079	0.308	0.087	0.279	0.083
sll0533	trigger factor	4	0.241	0.103	0.278	0.107	0.259	0.105

ORF	Protein Name	MS/ MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
slr0172	(IMP dehydrogenase)	5	0.368	0.091	0.402	0.121	0.385	0.106
slr0670	(universal stress protein)	2	0.315	0.106	0.394	0.098	0.354	0.102
slr0058	(Polyhydroxyalcanoate granule associated protein)	3	0.351	0.008	0.306	0.019	0.329	0.013
slr0923	(putative 30S ribosomal, Ycf65-like protein)	8	0.251	0.162	0.312	0.150	0.282	0.156
slr2025	hypothetical protein	2	0.238	0.140	0.279	0.119	0.259	0.129
slr0455	(Polyhydroxyalcanoate granule associated protein)	3	0.241	0.019	0.267	0.045	0.254	0.032
ssr1528	hypothetical protein	5	0.196	0.091	0.264	0.069	0.230	0.080
slr0049	(short chain dehydrogenase)	8	0.190	0.048	0.226	0.050	0.208	0.049
slr1273	hypothetical protein	2	0.143	0.020	0.242	0.012	0.193	0.016
slr1852	hypothetical protein	13	0.186	0.118	0.204	0.123	0.195	0.120
sll0272	hypothetical protein	8	0.168	0.018	0.203	0.013	0.186	0.016
slr0821	(thiamine Sulfur transferase)	3	0.193	0.056	0.167	0.032	0.180	0.044
sll0529	(transketolase)	6	0.173	0.116	0.184	0.136	0.178	0.126
sll1526	hypothetical protein	6	0.137	0.137	0.185	0.133	0.161	0.135
slr0244	hypothetical protein	11	0.141	0.077	0.158	0.073	0.149	0.075
slr0729	hypothetical protein	23	0.109	0.065	0.154	0.062	0.132	0.064
sll1414	hypothetical protein	6	0.067	0.018	0.096	0.023	0.081	0.020
ssr2998	hypothetical protein	9	0.062	0.013	0.098	0.018	0.080	0.016
sll1654	hypothetical protein	2	0.063	0.195	0.054	0.296	0.058	0.246
slr1649	hypothetical protein	7	0.015	0.066	0.065	0.050	0.040	0.058
slr1161	hypothetical protein	9	0.002	0.033	0.028	0.035	0.015	0.034
slr1780	hypothetical protein YCF54 ycf54	5	-0.031	0.028	0.010	0.058	-0.010	0.043
slr0146	hypothetical protein	3	0.009	0.141	-0.043	0.189	-0.017	0.165
slr0147	hypothetical protein	3	-0.038	0.015	-0.012	0.034	-0.025	0.024
sll1188	hypothetical protein	10	-0.063	0.025	-0.054	0.018	-0.059	0.022
sll0781	hypothetical protein	2	-0.108	0.018	-0.051	0.064	-0.080	0.041
sll0051	hypothetical protein	2	-0.107	0.029	-0.059	0.031	-0.083	0.030
slr0605	hypothetical protein	4	-0.128	0.067	-0.073	0.057	-0.101	0.062
Lipid Metabolism								
slr1051	enoyl-acyl-carrier-protein reductase	3	0.131	0.117	0.145	0.128	0.138	0.122
sll1945	1-deoxyxylulose-5-phosphate synthase	2	0.134	0.003	0.137	0.005	0.135	0.004
ssl2084	acyl carrier protein acpP	4	0.104	0.014	0.116	0.063	0.110	0.038
slr1020	sulfolipid biosynthesis protein SqdB sqdB	11	0.102	0.023	0.104	0.040	0.103	0.032
slr2089	squalene-hopene-cyclase shc	13	0.028	0.023	-0.021	0.025	0.003	0.024
Nucleotide Metabolism								
sll1035	uracil phosphoribosyltransferase	2	0.375	0.047	0.436	0.061	0.406	0.054
slr1622	soluble inorganic pyrophosphatase ppa	8	0.256	0.086	0.301	0.087	0.279	0.087
sll1815	adenylate kinase adk	3	0.254	0.096	0.271	0.087	0.263	0.092
sll1852	nucleoside diphosphate kinase	6	0.186	0.031	0.197	0.023	0.192	0.027
sll1043	polyribonucleotide nucleotidyltransferase	10	0.069	0.089	0.067	0.095	0.068	0.092
sll0421	adenylosuccinate lyase purB	4	-0.080	0.039	-0.036	0.033	-0.058	0.036

ORF	Protein Name	MS/MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
sll1663	phycocyanin alpha phycocyanobilin lyase related protein	2	0.298	0.190	0.336	0.108	0.317	0.149
sll0258	cytochrome c550 psbV	20	0.330	0.112	0.295	0.117	0.313	0.115
slr1986	allophycocyanin beta subunit apcB	170	0.250	0.032	0.276	0.030	0.263	0.031
sll1194	photosystem II 12 kDa extrinsic protein psbU	19	0.228	0.024	0.276	0.030	0.252	0.027
sll0928	allophycocyanin-B apcD	42	0.229	0.037	0.269	0.042	0.249	0.040
sll1579	phycobilisome rod linker polypeptide cpcC2	38	0.225	0.044	0.246	0.045	0.235	0.045
slr0335	phycobilisome core-membrane linker polypeptide apcE	82	0.209	0.039	0.231	0.038	0.220	0.039
sll0947	light repressed protein A homolog lrtA	8	0.155	0.117	0.181	0.125	0.168	0.121
sll1580	phycobilisome rod linker polypeptide cpcC1	76	0.142	0.046	0.181	0.048	0.162	0.047
slr2051	phycobilisome rod-core linker polypeptide cpcG1	66	0.134	0.039	0.164	0.041	0.149	0.040
sll0427	photosystem II manganese-stabilizing polypeptide psbO	11	0.108	0.072	0.129	0.082	0.118	0.077
slr1643	ferredoxin-NADP oxidoreductase petH	24	0.096	0.075	0.126	0.078	0.111	0.077
slr1459	phycobilisome core component apcF	9	0.019	0.023	0.048	0.027	0.034	0.025
slr0737	photosystem I subunit II psaD	11	-0.346	0.053	-0.347	0.061	-0.346	0.057
slr1834	P700 apoprotein subunit Ia psaA	12	-0.637	0.048	-0.640	0.033	-0.638	0.040
Porphyrin and Chlorophyll Metabolism								
slr0506	light-dependent NADPH-protochlorophyllide oxidoreductase por	4	0.402	0.116	0.475	0.137	0.439	0.127
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	6	0.076	0.052	0.108	0.042	0.092	0.047
sll0017	glutamate-1-semialdehyde aminomutase hemL	11	0.004	0.035	0.030	0.046	0.017	0.041
sll1185	coproporphyrinogen III oxidase, aerobic (oxygen-dependent) hemF	2	-0.094	0.001	-0.098	0.056	-0.096	0.028
Pyruvate Metabolism								
sll1841	pyruvate dehydrogenase dihydrolipoamide acetyltransferase component (E2)	4	0.499	0.071	0.515	0.080	0.507	0.075
slr1934	pyruvate dehydrogenase E1 component, alpha subunit	7	0.227	0.076	0.272	0.082	0.250	0.079
sll1721	pyruvate dehydrogenase E1 component, beta subunit	9	0.162	0.102	0.189	0.109	0.175	0.105

ORF	Protein Name	MS/ MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
slr1984	nucleic acid-binding protein, 30S ribosomal protein S1 homolog nbp1,rps1b	3	-0.060	0.025	-0.114	0.017	-0.087	0.021
sll1097	30S ribosomal protein S7 rps7	3	-0.118	0.034	-0.069	0.038	-0.094	0.036
sll1816	30S ribosomal protein S13 rps13	3	-0.152	0.046	-0.030	0.067	-0.091	0.057
Transport and Binding Proteins								
slr1890	bacterioferritin	12	0.295	0.032	0.326	0.038	0.311	0.035
sll1341	bacterioferritin	4	0.109	0.048	0.166	0.038	0.137	0.043
sll1762	periplasmic protein, putative polar amino acid transport system substrate-binding protein	12	-0.072	0.031	-0.048	0.031	-0.060	0.031
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein urtA	4	-0.079	0.014	-0.050	0.023	-0.065	0.019

Note: Hypothetical proteins with close similarity to another organism(s) with known functional group(s) were in parentheses.

ORF	Protein Name	MS/MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1393	glycogen (starch) synthase glgA	3	-0.060	0.041	-0.116	0.084	0.038	0.033
slr1176	glucose-1-phosphate adenylyltransferase	13	-0.172	0.058	0.037	0.028	-0.154	0.067
Carbon fixation								
sll1028	carbon dioxide concentrating mechanism protein CcmK ccmK2	10	-0.259	0.039	0.118	0.026	-0.170	0.016
sll1029	carbon dioxide concentrating mechanism protein CcmK ccmK1	7	-0.184	0.092	0.167	0.055	-0.151	0.095
sll1031	carbon dioxide concentrating mechanism protein CcmM, putative carboxysome structural protein ccmM	5	0.037	0.076	0.243	0.050	0.139	0.062
sll1525	phosphoribulokinase prk	14	-0.203	0.097	0.062	0.059	-0.117	0.065
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	44	-0.414	0.053	0.041	0.042	-0.315	0.039
Cellular processes								
slr1198	antioxidant protein	31	-0.120	0.061	0.103	0.074	0.005	0.057
slr1516	superoxide dismutase sodB	34	-0.134	0.017	0.032	0.017	-0.072	0.015
Energy Metabolism								
sll1499	ferredoxin-dependent glutamate synthase glsF	3	-0.244	0.059	0.005	0.028	-0.120	0.065
sll1987	catalase peroxidase cpx, katG	2	-0.179	0.187	-0.139	0.114	-0.064	0.124
slr0898	ferredoxin--nitrite reductase nirA	2	-0.220	0.186	0.003	0.329	-0.076	0.314
ssr0330	ferredoxin-thioredoxin reductase, variable chain ftrV	2	-0.153	0.060	0.051	0.005	-0.019	0.114
Folding, Sorting and Degradation								
sll0020	ATP-dependent Clp protease ATPase subunit	3	-0.309	0.081	0.033	0.066	-0.297	0.039
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	6	-0.150	0.030	0.047	0.039	-0.066	0.050
sll0408	peptidyl-prolyl cis-trans isomerase	2	-0.328	0.025	0.124	0.011	-0.216	0.041
sll1621	AhpC/TSA family protein	32	-0.246	0.028	0.006	0.029	-0.233	0.026
slr0623	thioredoxin trxA	8	-0.263	0.052	0.012	0.041	-0.151	0.038
slr0659	oligopeptidase A	2	-0.070	0.025	0.130	0.098	-0.041	0.027
slr1251	peptidyl-prolyl cis-trans isomerase	3	-0.020	0.049	0.103	0.034	0.037	0.038

ORF	Protein Name	MS/MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Nucleotide Metabolism								
sll0368	uracil phosphoribosyltransferase	2	0.093	0.077	-0.029	0.090	0.062	0.115
sll1043	polyribonucleotide nucleotidyltransferase	6	-0.171	0.037	-0.069	0.035	-0.172	0.023
sll1852	nucleoside diphosphate kinase	11	-0.294	0.047	0.161	0.023	-0.146	0.051
slr1622	soluble inorganic pyrophosphatase ppa	3	-0.323	0.039	0.017	0.040	-0.254	0.033
slr1722	inosine-5'-monophosphate dehydrogenase	5	-0.149	0.094	0.015	0.048	-0.107	0.060
Other								
sll0576	putative sugar-nucleotide epimerase/dehydratease	10	-0.188	0.046	0.019	0.023	-0.210	0.044
sll0588	unknown protein	6	-0.017	0.012	0.248	0.011	0.043	0.014
sll1009	unknown protein	2	0.800	0.044	-0.016	0.006	0.871	0.087
sll1130	unknown protein	3	-0.186	0.036	0.002	0.016	-0.120	0.044
sll1559	soluble hydrogenase 42 kD subunit	7	-0.368	0.077	0.196	0.055	-0.263	0.084
sll1873	unknown protein	12	-0.001	0.099	-0.039	0.048	0.053	0.109
slr1852	unknown protein	4	-0.260	0.027	0.008	0.051	-0.175	0.026
slr1854	unknown protein	2	0.155	0.012	-0.227	0.056	0.153	0.054
slr1855	unknown protein	15	-0.218	0.065	0.001	0.044	-0.188	0.055
slr1963	water-soluble carotenoid protein	7	-0.229	0.132	0.072	0.028	-0.186	0.054
slr2002	cyanophycin synthetase cphA	3	-0.015	0.066	0.086	0.063	0.117	0.053
ssl2296	pterin-4a-carbinolamine dehydratase	2	-0.268	0.147	0.149	0.029	-0.417	0.082
ssr1853	unknown protein	2	0.075	0.066	0.001	0.036	0.145	0.003
Pentose Phosphate Pathway								
sll0329	6-phosphogluconate dehydrogenase	4	-0.201	0.133	0.217	0.062	-0.052	0.166
sll0807	pentose-5-phosphate-3-epimerase rpe	2	-0.411	0.061	0.099	0.047	-0.307	0.021
sll1070	transketolase	8	-0.207	0.046	0.061	0.014	-0.133	0.041
slr1793	transaldolase	6	-0.279	0.099	0.061	0.044	-0.228	0.096
Photosynthesis								
sll0199	plastocyanin petE	3	-0.295	0.092	0.651	0.050	-0.309	0.064
sll0258	cytochrome c550 psbV	12	-0.022	0.036	0.014	0.035	0.011	0.017
sll0427	photosystem II manganese-stabilizing polypeptide psbO	7	-0.001	0.027	0.035	0.012	0.041	0.024
sll0928	allophycocyanin-B apcD	25	-0.292	0.052	0.114	0.046	-0.109	0.043
sll0947	light repressed protein A homolog lrtA	6	-0.141	0.177	-0.142	0.037	-0.137	0.232

ORF	Protein Name	MS/MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr1894	probable DNA-binding stress protein	14	-0.176	0.044	0.041	0.009	-0.109	0.050
Signal Transduction								
sll1590	two-component sensor histidine kinase hik20	3	-0.375	0.077	0.151	0.025	-0.405	0.074
slr2024	two-component response regulator CheY subfamily	2	-0.123	0.043	-0.118	0.055	-0.091	0.092
TCA cycle								
slr0665	aconitate hydratase	2	-0.357	0.109	0.051	0.066	-0.248	0.037
slr1289	isocitrate dehydrogenase (NADP+) icd	8	-0.239	0.033	0.026	0.032	-0.067	0.038
Transcription								
sll1787	RNA polymerase beta subunit rpoB	9	-0.244	0.054	-0.070	0.090	-0.187	0.050
sll1789	RNA polymerase beta prime subunit rpoC2	7	-0.122	0.085	-0.144	0.046	-0.076	0.092
slr1265	RNA polymerase gamma-subunit rpoC1	3	-0.110	0.029	0.156	0.068	-0.176	0.067
slr1856	phosphoprotein substrate of icfG gene cluster	2	-0.431	0.054	0.097	0.015	-0.377	0.021
Translation								
sll0145	ribosome releasing factor frr, rrf	4	-0.072	0.014	0.023	0.022	0.035	0.013
sll1099	elongation factor Tu tufA	45	-0.089	0.055	0.083	0.028	0.007	0.061
sll1260	30S ribosomal protein S2 rps2	10	-0.298	0.109	0.004	0.028	-0.259	0.154
sll1425	proline-tRNA ligase proS	3	-0.215	0.022	-0.073	0.009	-0.076	0.045
sll1745	50S ribosomal protein L10 rpl10	2	-0.147	0.039	-0.035	0.022	-0.058	0.045
sll1746	50S ribosomal protein L12 rpl12	17	-0.373	0.050	0.108	0.024	-0.404	0.053
sll1812	30S ribosomal protein S5 rps5	5	0.036	0.043	-0.033	0.035	0.006	0.022
sll1816	30S ribosomal protein S13 rps13	9	0.046	0.028	-0.175	0.044	0.125	0.019
sll1821	50S ribosomal protein L13 rpl13	3	0.103	0.018	0.160	0.066	0.136	0.027
slr0193	RNA-binding protein rbp3	4	0.003	0.017	-0.191	0.024	0.135	0.015
slr0434	elongation factor P efp	6	-0.259	0.022	0.026	0.022	-0.167	0.014
slr0744	translation initiation factor IF-2 infB	4	0.061	0.015	0.027	0.052	0.051	0.056
slr1356	30S ribosomal protein S1 rps1a	2	-0.118	0.097	-0.166	0.418	-0.033	0.165
slr1463	elongation factor EF-G fus	29	-0.115	0.036	-0.011	0.026	-0.069	0.029

Appendix H

The list of 150 proteins found in Experiment 2 in Chapter 6 via iTRAQ analysis. This set of experiment consists of 'prolonged' time-points derived from the dark cycle: 1 hr, 6 hr and 11 hr against the common reference of 24 hr dark. The ratio is expressed in Log_{10} unit, whereas the error is estimated in standard deviation (SD).

ORF	Protein Name	MS/MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Amino Acids Metabolism								
sll0080	N-acetyl-gamma-glutamyl-phosphate reductase argC	2	0.221	0.015	-0.130	0.009	0.075	0.039
sll0220	L-glutamine:D-fructose-6-P amidotransferase glmS	2	-0.093	0.033	-0.119	0.092	-0.052	0.028
sll1058	dihydrodipicolinate reductase dapB	4	-0.064	0.053	-0.032	0.025	0.001	0.016
sll1363	ketol-acid reductoisomerase ilvC	7	0.011	0.031	0.026	0.022	0.031	0.033
sll1502	NADH-dependent glutamate synthase large subunit gltB	4	-0.047	0.030	-0.075	0.023	-0.008	0.026
sll1750	urease alpha subunit ureC	2	0.088	0.006	-0.121	0.009	-0.004	0.019
sll2001	leucine aminopeptidase	5	-0.020	0.018	-0.037	0.010	-0.031	0.026
slr0032	probable branched-chain amino acid aminotransferase	2	-0.015	0.036	-0.054	0.033	0.052	0.006
slr0288	glutamate--ammonia ligase glnN	9	0.211	0.038	0.181	0.055	0.206	0.035
slr0452	dihydroxyacid dehydratase ilvD	2	-0.119	0.012	-0.102	0.047	-0.053	0.049
slr0543	tryptophan synthase beta subunit trpB	2	0.135	0.001	-0.150	0.053	0.108	0.033
slr1756	glutamate--ammonia ligase glnA	27	0.066	0.046	-0.022	0.023	0.039	0.021
ATP synthesis								
sll1326	ATP synthase alpha chain atpA	10	0.015	0.019	-0.133	0.029	-0.056	0.018
Carbohydrate Metabolism								
sll0158	1,4-alpha-glucan branching enzyme glgB	6	-0.050	0.045	-0.053	0.027	-0.004	0.019
sll0945	glycogen synthase glgA	2	-0.078	0.007	-0.178	0.013	-0.201	0.001
sll1213	GDP-fucose synthetase	2	-0.033	0.018	0.022	0.010	0.022	0.055
sll1356	glycogen phosphorylase	17	-0.090	0.023	-0.072	0.016	-0.056	0.014
slr1176	glucose-1-phosphate adenylyltransferase	17	-0.031	0.042	-0.044	0.025	-0.008	0.028
Carbon fixation								
sll0934	carboxysome formation protein CcmA ccmA	2	-0.120	0.001	-0.253	0.076	0.031	0.035

ORF	Protein Name	MS/MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr0161	twitching motility protein PilT pilT1	2	0.069	0.035	-0.142	0.081	-0.051	0.047
slr0623	thioredoxin trxA	11	0.059	0.050	-0.080	0.060	0.005	0.025
slr1251	peptidyl-prolyl cis-trans isomerase	7	-0.020	0.040	-0.017	0.021	0.010	0.024
slr1641	ClpB protein clpB1	3	-0.077	0.034	-0.105	0.067	-0.031	0.030
slr2075	10kD chaperonin groES	13	0.186	0.036	-0.171	0.040	0.011	0.024
slr2076	60kD chaperonin groEL1	7	-0.048	0.024	-0.045	0.059	-0.128	0.033
Glycolysis								
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	9	0.046	0.082	-0.126	0.029	-0.047	0.032
sll0920	phosphoenolpyruvate carboxylase ppc	3	-0.028	0.016	-0.055	0.019	-0.032	0.059
sll1275	pyruvate kinase 2	3	-0.118	0.018	-0.003	0.010	-0.081	0.024
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	18	-0.074	0.022	-0.193	0.034	-0.142	0.019
slr0394	phosphoglycerate kinase pgk	31	0.021	0.021	-0.059	0.019	-0.018	0.020
slr1349	glucose-6-phosphate isomerase	3	-0.063	0.012	-0.032	0.011	-0.012	0.023
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	12	0.023	0.055	-0.130	0.037	-0.054	0.022
hypothetical proteins								
sll0051	hypothetical protein	6	-0.048	0.013	-0.010	0.011	-0.004	0.011
sll0272	hypothetical protein	4	0.160	0.017	-0.199	0.024	0.123	0.016
sll0359	hypothetical protein	14	0.017	0.075	-0.109	0.036	-0.036	0.022
sll1188	hypothetical protein	6	-0.095	0.044	-0.034	0.049	-0.039	0.078
sll1414	hypothetical protein	3	0.019	0.039	-0.125	0.027	0.024	0.017
sll1526	hypothetical protein	3	-0.087	0.071	-0.052	0.030	-0.095	0.030
slr0038	hypothetical protein	4	0.020	0.055	0.039	0.018	0.075	0.030
slr0110	hypothetical protein	2	-0.092	0.003	-0.159	0.040	-0.132	0.056
slr0244	hypothetical protein	12	-0.141	0.026	-0.130	0.025	-0.125	0.031
slr0729	hypothetical protein	15	-0.032	0.047	-0.195	0.028	-0.100	0.031
slr0923	hypothetical protein YCF65 ycf65	3	0.025	0.012	-0.211	0.016	-0.063	0.049
slr1161	hypothetical protein	5	-0.013	0.020	-0.079	0.026	-0.001	0.025
slr1590	hypothetical protein	2	-0.007	0.025	-0.204	0.024	-0.059	0.030
slr1649	hypothetical protein	5	-0.006	0.039	-0.074	0.023	-0.025	0.053
slr1780	hypothetical protein YCF54 ycf54	7	-0.009	0.025	0.012	0.016	0.037	0.030
ssr2998	hypothetical protein	6	-0.043	0.024	-0.088	0.012	-0.082	0.013
Lipid Metabolism								
slr0015	lipid A disaccharide synthase	2	0.089	0.025	-0.295	0.003	0.007	0.005
slr1020	sulfolipid biosynthesis protein SqdB sqdB	4	0.086	0.028	0.017	0.055	0.057	0.100

ORF	Protein Name	MS/MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1194	photosystem II 12 kDa extrinsic protein psbU	10	0.128	0.023	-0.172	0.028	0.037	0.022
sll1577	phycocyanin beta subunit cpcB	158	0.390	0.024	-0.048	0.011	0.163	0.012
sll1578	phycocyanin alpha subunit cpcA	276	0.418	0.016	-0.128	0.010	0.210	0.010
sll1579	phycobilisome rod linker polypeptide cpcC2	13	0.211	0.035	-0.009	0.035	0.066	0.021
sll1580	phycobilisome rod linker polypeptide cpcC1	57	0.107	0.025	-0.084	0.032	-0.054	0.014
slr0335	phycobilisome core-membrane linker polypeptide apcE	70	0.087	0.020	-0.038	0.016	-0.035	0.014
slr0737	photosystem I subunit II psaD	8	-0.334	0.050	-0.373	0.036	-0.163	0.035
slr0906	photosystem II core light harvesting protein psbB	2	-0.593	0.005	-0.619	0.045	-0.205	0.029
slr1459	phycobilisome core component apcF	8	0.055	0.049	-0.076	0.017	0.022	0.014
slr1643	ferredoxin-NADP oxidoreductase petH	26	-0.029	0.044	-0.052	0.025	-0.060	0.021
slr1834	P700 apoprotein subunit Ia psaA	6	-0.291	0.085	-0.350	0.067	0.043	0.072
slr1986	allophycocyanin beta subunit apcB	84	0.208	0.049	-0.094	0.035	0.070	0.042
slr2051	phycobilisome rod-core linker polypeptide cpcG1	24	0.074	0.042	-0.043	0.026	-0.015	0.023
slr2067	allophycocyanin alpha subunit apcA	28	-0.003	0.147	-0.231	0.122	-0.161	0.136
ssl3093	phycobilisome small rod linker polypeptide cpcD	6	0.240	0.030	-0.050	0.014	0.004	0.013
ssr2831	photosystem I subunit IV psaE	2	-0.227	0.018	-0.274	0.016	-0.176	0.008
ssr3383	phycobilisome small core linker polypeptide apcC	4	0.363	0.051	-0.176	0.031	0.176	0.032
Porphyrin and chlorophyll metabolism								
sll0017	glutamate-1-semialdehyde aminomutase hemL	3	-0.003	0.050	-0.052	0.036	-0.101	0.018
sll1185	coproporphyrinogen III oxidase, aerobic (oxygen-dependent) hemF	4	-0.106	0.043	-0.076	0.036	-0.077	0.023
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	4	0.152	0.024	0.035	0.034	0.102	0.023
Pyruvate Metabolism								
slr1096	dihydrolipoamide dehydrogenase	2	-0.007	0.082	-0.043	0.046	-0.040	0.019
Regulatory control								
ssl0707	nitrogen regulatory protein P-II glnB	10	0.139	0.041	0.006	0.038	0.177	0.031

ORF	Protein Name	MS/ MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr1463	elongation factor EF-G fus	15	-0.068	0.025	-0.065	0.024	-0.056	0.023
slr1984	nucleic acid-binding protein, 30S ribosomal protein S1 homolog nbp1,rps1b	2	-0.051	0.000	-0.135	0.000	-0.052	0.046
Transport and Binding Proteins								
sll0680	phosphate-binding periplasmic protein precursor (PBP)	2	0.118	0.000	0.044	0.028	0.059	0.031
sll1341	bacterioferritin	2	0.063	0.119	-0.328	0.094	0.030	0.037
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein urtA	3	0.100	0.024	-0.136	0.026	0.022	0.021
slr1890	bacterioferritin	8	0.239	0.032	-0.142	0.031	0.098	0.026

ORF	Protein Name	MS/MS	12.5hrL:24hrD (115:114)		13hrL:24hrD (116:114)		13.5hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1031	carbon dioxide concentrating mechanism protein CcmM, putative carboxysome structural protein ccmM	3	-0.063	0.026	0.129	0.038	-0.082	0.037
sll1525	phosphoribulokinase prk	5	-0.423	0.073	-0.094	0.038	-0.468	0.086
Cellular Processes								
slr1198	antioxidant protein	22	-0.282	0.066	-0.024	0.041	-0.360	0.053
slr1516	superoxide dismutase sodB	15	-0.148	0.021	0.009	0.011	-0.201	0.013
Energy Metabolism								
sll1499	ferredoxin-dependent glutamate synthase glsF	2	0.378	0.389	0.043	0.063	0.424	0.370
sll1987	catalase peroxidase cpx, katG	4	-0.575	0.093	-0.113	0.056	-0.542	0.052
slr0898	ferredoxin--nitrite reductase nirA	5	-0.125	0.019	-0.023	0.013	-0.061	0.014
Folding, Sorting and Degradation								
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	4	-0.254	0.024	0.006	0.040	-0.280	0.022
sll0227	peptidyl-prolyl cis-trans isomerase B, periplasmic protein ppiB	3	-0.103	0.058	0.006	0.013	-0.107	0.043
sll0408	peptidyl-prolyl cis-trans isomerase	2	-0.118	0.279	0.021	0.009	-0.127	0.309
sll0416	60 kDa chaperonin 2, GroEL2, molecular chaperone groEL-2	6	-0.017	0.117	-0.047	0.048	-0.018	0.099
sll1621	AhpC/TSA family protein	15	-0.303	0.032	-0.009	0.017	-0.278	0.039
slr0623	thioredoxin trxA	8	-0.392	0.162	0.025	0.030	-0.490	0.192
slr1251	peptidyl-prolyl cis-trans isomerase	5	-0.089	0.028	0.027	0.036	-0.088	0.021
slr2075	10kD chaperonin groES	8	-0.089	0.050	-0.113	0.037	-0.146	0.044
slr2076	60kD chaperonin groEL1	3	0.160	0.036	0.024	0.046	0.196	0.042
Glycolysis								
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	4	-0.544	0.028	-0.069	0.014	-0.587	0.028
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	6	-0.440	0.055	-0.276	0.004	-0.441	0.099
slr0394	phosphoglycerate kinase pgk	11	-0.274	0.029	-0.041	0.032	-0.344	0.047
slr1349	glucose-6-phosphate isomerase	3	-0.156	0.012	0.012	0.004	-0.099	0.028
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	7	-0.135	0.110	-0.054	0.033	-0.129	0.164
Hypothetical Proteins								
sll0051	hypothetical protein	3	-0.085	0.009	-0.030	0.028	-0.077	0.023
sll0359	hypothetical protein	6	-0.259	0.057	-0.159	0.020	-0.203	0.057
sll1526	hypothetical protein	3	-0.142	0.032	-0.021	0.019	-0.120	0.057

ORF	Protein Name	MS/MS	12.5hrL:24hrD (115:114)		13hrL:24hrD (116:114)		13.5hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Photosynthesis								
sll0258	cytochrome c550 psbV	6	0.001	0.036	0.029	0.038	-0.096	0.064
sll0427	photosystem II manganese-stabilizing polypeptide psbO	4	-0.188	0.107	-0.033	0.040	-0.156	0.090
sll0928	allophycocyanin-B apcD	14	-0.572	0.069	-0.006	0.026	-0.721	0.045
sll0947	light repressed protein A homolog lrtA	4	-0.064	0.312	-0.270	0.037	-0.116	0.272
sll1194	photosystem II 12 kDa extrinsic protein psbU	15	-0.325	0.089	0.014	0.016	-0.392	0.128
sll1577	phycocyanin beta subunit cpcB	18	-0.282	0.073	0.283	0.039	-0.295	0.072
sll1578	phycocyanin alpha subunit cpcA	79	-0.457	0.038	0.168	0.011	-0.412	0.041
sll1579	phycobilisome rod linker polypeptide cpcC2	7	-0.334	0.119	0.131	0.064	-0.418	0.081
sll1580	phycobilisome rod linker polypeptide cpcC1	32	-0.266	0.065	0.067	0.015	-0.294	0.083
slr0335	phycobilisome core-membrane linker polypeptide apcE	46	-0.183	0.042	0.021	0.011	-0.206	0.040
slr0737	photosystem I subunit II psaD	3	-0.442	0.137	-0.564	0.027	-0.472	0.158
slr1459	phycobilisome core component apcF	5	-0.561	0.129	0.076	0.043	-0.617	0.104
slr1643	ferredoxin-NADP oxidoreductase petH	14	-0.129	0.058	0.052	0.018	-0.108	0.062
slr1834	P700 apoprotein subunit Ia psaA	8	-0.507	0.043	-0.722	0.048	-0.523	0.064
slr1986	allophycocyanin beta subunit apcB	41	-0.380	0.080	0.130	0.018	-0.396	0.068
slr2051	phycobilisome rod-core linker polypeptide cpcG1	16	-0.262	0.051	0.023	0.056	-0.327	0.057
slr2067	allophycocyanin alpha subunit apcA	24	-0.214	0.048	0.075	0.012	-0.206	0.055
Porphyrin and Chlorophyll Metabolism								
sll1185	coproporphyrinogen III oxidase, aerobic (oxygen-dependent) hemF	3	-0.159	0.039	-0.097	0.025	-0.178	0.036
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	4	-0.465	0.032	0.107	0.013	-0.369	0.016
Pyruvate Metabolism								
slr1096	dihydrolipoamide dehydrogenase	3	-0.138	0.044	0.000	0.035	-0.080	0.029
Regulatory Control								
ssl0707	nitrogen regulatory protein P-II glnB	12	-0.351	0.086	0.124	0.025	-0.280	0.074
Replication and Repair								
sll1712	DNA binding protein HU	5	-0.446	0.071	-0.224	0.016	-0.582	0.087

Appendix J

The list of 150 proteins found in Experiment 4 in Chapter 6 via iTRAQ analysis. This set of experiment consists of 'prolonged' time-points derived from the light cycle: 13 hr. 18 hr and 23 hr against the common reference of 24 hr dark. The ratio is expressed in Log_{10} unit, whereas the error is estimated in standard deviation (SD).

ORF	Protein Name	MS/ MS	13hrL:24hrD (115:114)		18hrL:24hrD (116:114)		23hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Amino Acids Metabolism								
sll1058	dihydrodipicolinate reductase dapB	3	-0.094	0.071	-0.037	0.042	-0.086	0.042
sll1363	ketol-acid reductoisomerase ilvC	2	-0.038	0.001	-0.079	0.019	0.009	0.022
sll1502	NADH-dependent glutamate synthase large subunit gltB	3	-0.038	0.053	-0.004	0.036	0.037	0.046
sll1682	alanine dehydrogenase	2	-0.087	0.021	-0.093	0.014	-0.045	0.025
sll1750	urease alpha subunit ureC	4	0.142	0.073	0.145	0.090	0.167	0.078
sll1931	serine hydroxymethyltransferase glyA	2	-0.039	0.064	0.062	0.044	0.055	0.022
sll2001	leucine aminopeptidase	2	-0.121	0.011	-0.081	0.022	-0.063	0.007
slr0288	glutamate--ammonia ligase glnN	8	0.155	0.018	0.173	0.022	0.215	0.030
slr0543	tryptophan synthase beta subunit trpB	2	0.095	0.030	0.077	0.002	0.127	0.030
slr1756	glutamate--ammonia ligase glnA	15	0.082	0.042	0.120	0.049	0.119	0.040
ATP synthesis								
sll1326	ATP synthase alpha chain atpA	6	-0.064	0.023	0.026	0.046	0.012	0.046
carbohydrate Metabolism								
sll0158	1,4-alpha-glucan branching enzyme glgB	6	-0.013	0.046	-0.020	0.040	0.003	0.044
sll0945	glycogen synthase glgA	3	0.078	0.038	0.047	0.040	0.021	0.044
sll1212	GDP-mannose 4,6-dehydratase	2	0.051	0.114	0.071	0.017	0.061	0.003
sll1213	GDP-fucose synthetase	5	-0.023	0.040	-0.005	0.032	0.014	0.022
sll1356	glycogen phosphorylase	9	-0.075	0.013	-0.102	0.013	-0.069	0.026
slr0984	CDP-glucose 4,6-dehydratase rfbG	3	0.783	0.046	0.714	0.040	0.678	0.062
slr1176	glucose-1-phosphate adenylyltransferase	13	-0.023	0.038	-0.027	0.029	-0.001	0.028
Carbon fixation								
sll1028	carbon dioxide concentrating mechanism protein CcmK2	18	-0.031	0.022	0.010	0.020	-0.002	0.026
sll1029	carbon dioxide concentrating mechanism protein CcmK1	2	0.017	0.089	0.143	0.012	0.172	0.003
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	42	0.113	0.026	0.143	0.029	0.165	0.029

ORF	Protein Name	MS/MS	13hrL:24hrD (115:114)		18hrL:24hrD (116:114)		23hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll0272	hypothetical protein	4	0.197	0.045	0.211	0.026	0.175	0.035
sll0359	hypothetical protein	6	-0.085	0.112	0.002	0.100	0.033	0.113
sll1188	hypothetical protein	8	-0.093	0.033	-0.105	0.031	-0.069	0.032
sll1526	hypothetical protein	5	-0.035	0.024	-0.020	0.073	0.045	0.029
slr0038	hypothetical protein	2	0.038	0.051	-0.104	0.074	0.046	0.019
slr0049	hypothetical protein	2	-0.184	0.213	-0.158	0.199	-0.103	0.246
slr0147	hypothetical protein	3	-0.280	0.050	-0.243	0.019	-0.190	0.041
slr0172	hypothetical protein	3	-0.054	0.039	0.103	0.049	0.103	0.035
slr0244	hypothetical protein	8	-0.261	0.041	-0.202	0.038	-0.142	0.030
slr0455	hypothetical protein	2	-0.065	0.006	0.004	0.022	0.117	0.033
slr0689	hypothetical protein	2	-0.085	0.005	-0.066	0.051	-0.129	0.043
slr0729	hypothetical protein	10	-0.031	0.052	-0.055	0.059	-0.047	0.047
slr1161	hypothetical protein	2	-0.063	0.018	0.038	0.014	0.052	0.015
slr1649	hypothetical protein	10	-0.058	0.024	-0.062	0.026	0.006	0.019
slr1780	hypothetical protein YCF54	3	0.034	0.003	0.035	0.070	0.091	0.076
ssr2998	hypothetical protein	8	-0.159	0.024	-0.067	0.012	-0.010	0.021
Lipid Metabolism								
slr1020	sulfolipid biosynthesis protein SqdB sqdB	3	0.154	0.100	0.154	0.036	0.197	0.098
slr2089	squalene-hopene-cyclase shc	8	0.022	0.033	0.063	0.023	0.039	0.013
ssl2084	acyl carrier protein acpP	5	-0.094	0.011	0.004	0.028	0.028	0.036
Nucleotide Metabolism								
sll0421	adenylosuccinate lyase purB	2	-0.066	0.001	-0.052	0.052	0.023	0.004
sll1043	polyribonucleotide nucleotidyltransferase	5	0.063	0.069	0.028	0.127	0.038	0.060
sll1815	adenylate kinase adk	3	0.231	0.103	0.216	0.143	0.242	0.081
sll1852	nucleoside diphosphate kinase	7	0.123	0.028	0.138	0.025	0.149	0.032
slr1622	soluble inorganic pyrophosphatase ppa	5	0.100	0.046	0.101	0.056	0.063	0.072
Other								
sll0245	probable GTP binding protein	2	-0.093	0.043	-0.067	0.001	-0.071	0.001
sll0576	putative sugar-nucleotide epimerase/dehydratease	9	-0.078	0.030	-0.046	0.048	-0.002	0.038
sll0588	unknown protein	5	0.107	0.007	0.170	0.023	0.172	0.019
sll0617	plasma membrane protein essential for thylakoid formation vipp1	2	-0.011	0.042	0.031	0.007	0.030	0.022
sll0872	unknown protein	2	0.344	0.347	0.387	0.340	0.347	0.335
sll1130	unknown protein	3	0.041	0.138	0.016	0.116	0.025	0.104
sll1305	probable hydrolase	2	0.026	0.030	0.071	0.038	0.070	0.007
sll1559	soluble hydrogenase 42 kD subunit	9	0.175	0.044	0.164	0.048	0.175	0.036
sll1583	unknown protein	2	-0.081	0.208	-0.023	0.287	-0.071	0.261
sll1785	periplasmic protein, function unknown	4	-0.097	0.098	-0.095	0.070	-0.041	0.044
sll1873	unknown protein	15	-0.114	0.026	-0.108	0.015	-0.086	0.018
slr0476	unknown protein	2	-0.087	0.010	-0.049	0.049	0.117	0.084
slr1852	unknown protein	7	-0.132	0.086	-0.067	0.037	-0.030	0.026
slr1855	unknown protein	4	-0.101	0.027	-0.061	0.068	-0.016	0.024

ORF	Protein Name	MS/MS	13hrL:24hrD (115:114)		18hrL:24hrD (116:114)		23hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Porphyrin and chlorophyll metabolism								
sll0017	glutamate-1-semialdehyde aminomutase hemL	7	0.161	0.029	0.089	0.048	0.138	0.039
sll1185	coproporphyrinogen III oxidase, aerobic (oxygen-dependent) hemF	3	-0.133	0.012	-0.130	0.038	-0.113	0.043
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	5	0.155	0.064	0.155	0.047	0.225	0.055
slr0536	uroporphyrinogen decarboxylase hemE	2	0.320	0.142	0.219	0.134	0.204	0.143
Pyruvate Metabolism								
slr1096	dihydrolipoamide dehydrogenase	2	0.004	0.069	0.054	0.100	0.011	0.053
Regulatory Control								
ssl0707	nitrogen regulatory protein P-II glnB	15	0.093	0.021	0.107	0.025	0.105	0.024
Replication and Repair								
sll1463	cell division protein FtsH ftsH	3	-0.054	0.002	-0.069	0.040	-0.038	0.044
sll1712	DNA binding protein HU	9	-0.047	0.059	0.032	0.060	0.139	0.081
slr1894	probable DNA-binding stress protein	6	-0.101	0.019	-0.063	0.013	-0.015	0.016
Signal Transduction								
sll1590	two-component sensor histidine kinase hik20	4	0.079	0.029	0.139	0.033	0.139	0.022
TCA cycle								
sll0401	citrate synthase	2	-0.078	0.121	-0.046	0.146	0.001	0.073
slr1289	isocitrate dehydrogenase (NADP+) icd	6	-0.025	0.050	-0.013	0.056	0.029	0.052
Transcription								
sll1626	LexA repressor	6	-0.085	0.029	-0.098	0.026	-0.079	0.045
sll1787	RNA polymerase beta subunit rpoB	2	0.020	0.168	-0.155	#DIV/0!	0.068	0.004
sll1789	RNA polymerase beta prime subunit rpoC2	2	-0.128	0.041	-0.068	0.003	-0.044	0.053
slr0638	glycyl-tRNA synthetase alpha chain glyQ	2	-0.040	0.003	-0.016	0.016	0.018	0.063
slr1265	RNA polymerase gamma-subunit rpoC1	2	-0.027	0.005	0.040	0.034	0.005	0.023
slr1859	anti-sigma f factor antagonist	5	0.073	0.079	0.117	0.104	0.091	0.068
Translation								
sll0145	ribosome releasing factor frr, rrf	2	-0.010	0.002	0.034	0.004	0.007	0.002
sll0320	probable ribonuclease D	2	0.055	0.012	-0.015	0.024	0.015	0.043
sll1099	elongation factor Tu tufA	23	0.052	0.048	0.089	0.047	0.101	0.040

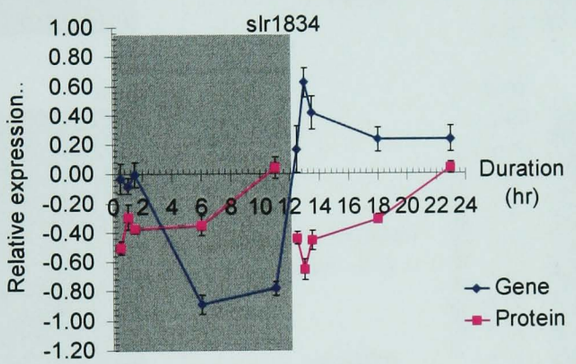
Appendix M

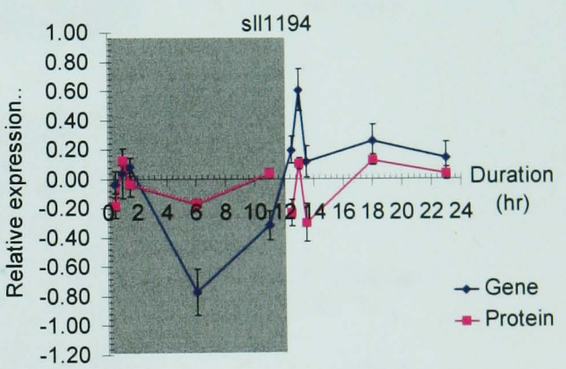
The relative gene and protein expression profile for the 26 selected genes across the 12 hour dark and light cycle. The shaded area refers to the dark cycle while the un-shaded area is the light cycle. The y-axis measures the relative protein expression in Log_{10} unit relative to the expression of the 24 hr dark time-point.

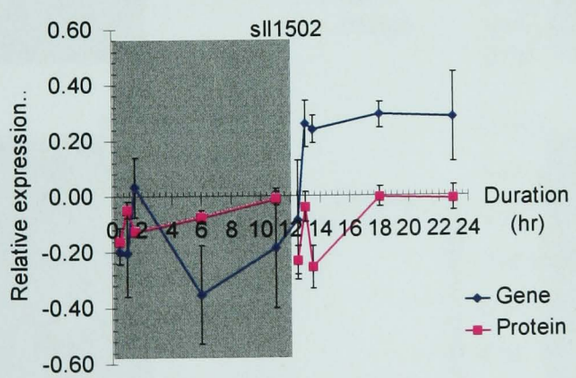
Note:

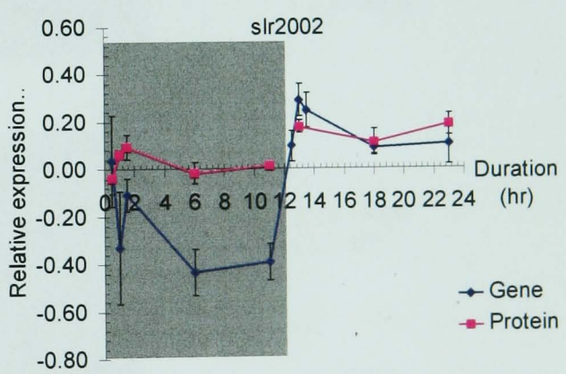
1) The protein time-points at 0.5-hr dark and 1.5-hr dark for sll0018 and sll1626 are unavailable from the iTRAQ experiments.
2) The protein time-points of 12.5-hr dark and 13.5-hr light for slr2002 are unavailable from the iTRAQ experiment.
3) The proteomics profile of sll1226 is not available from the iTRAQ experiment.

A1-type

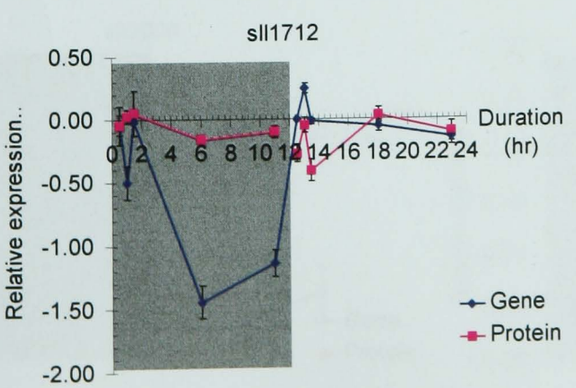


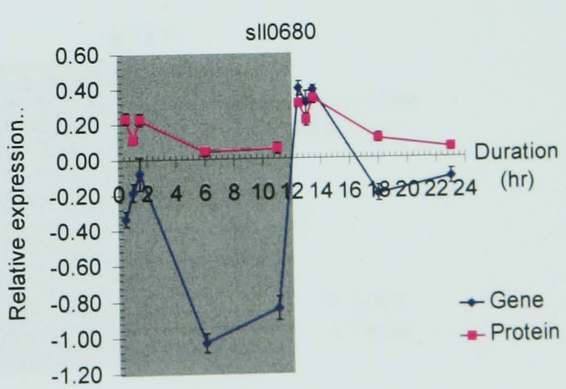


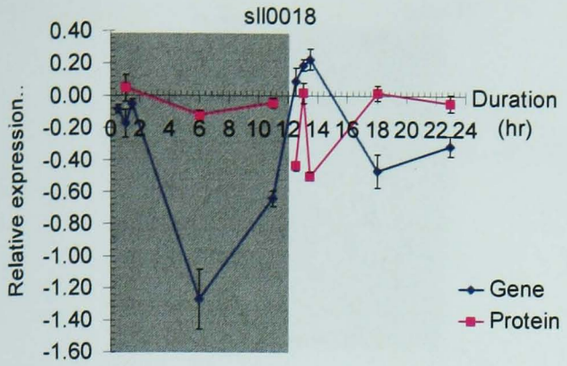


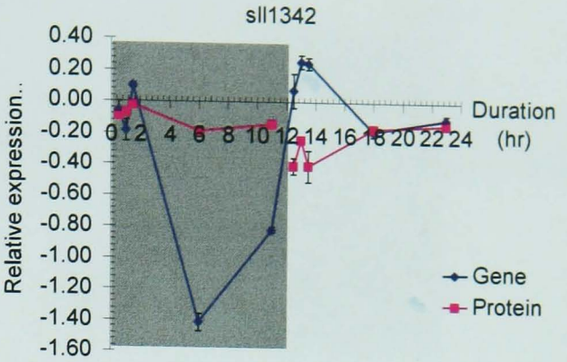


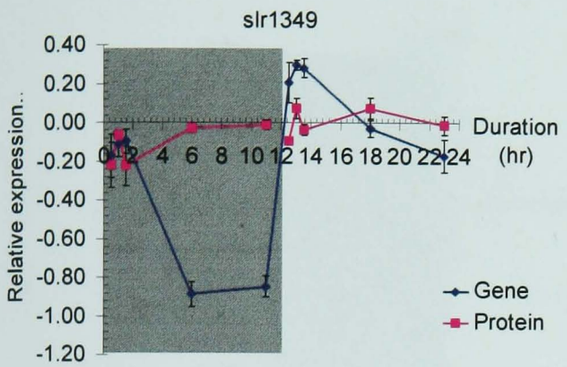
A2-type

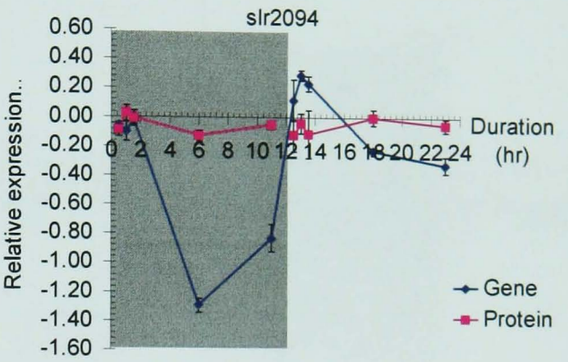


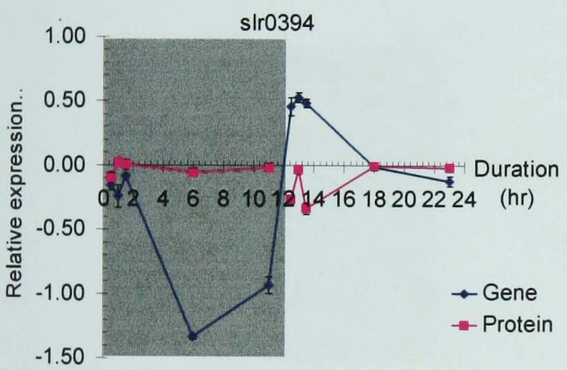


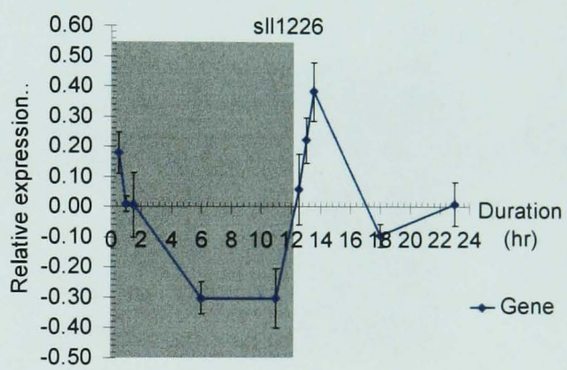

sll0018
Relative expression..
Duration (hr)
Gene
Protein

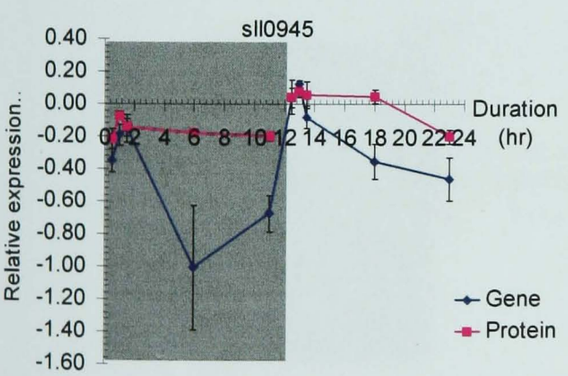

sll1342
Relative expression..
Duration (hr)
Gene
Protein

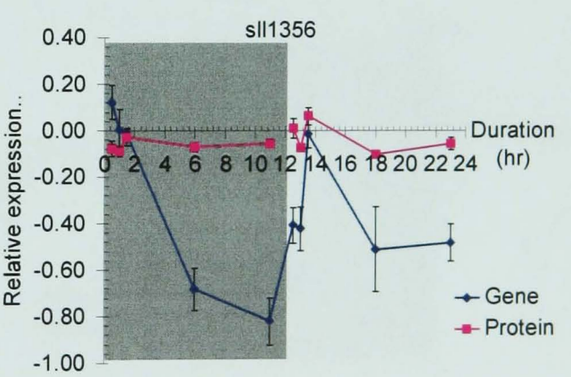

slr1349
Relative expression..
Duration (hr)
Gene
Protein

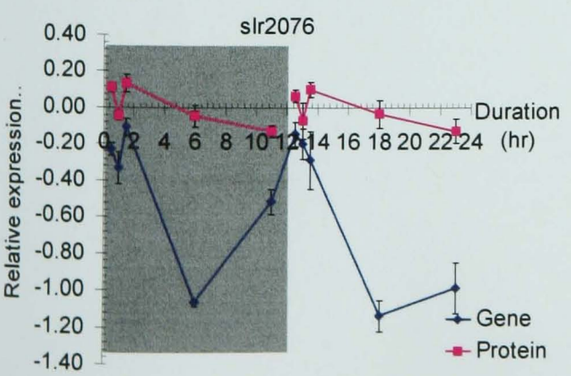

slr2094
Relative expression..
Duration (hr)
Gene
Protein

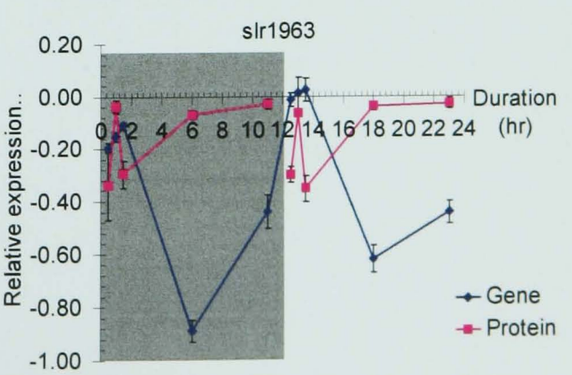

slr0394
Relative expression..
Duration (hr)
Gene
Protein

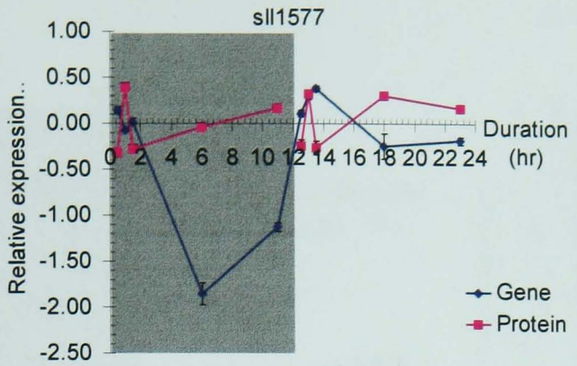

sll1226
Relative expression..
Duration (hr)
Gene

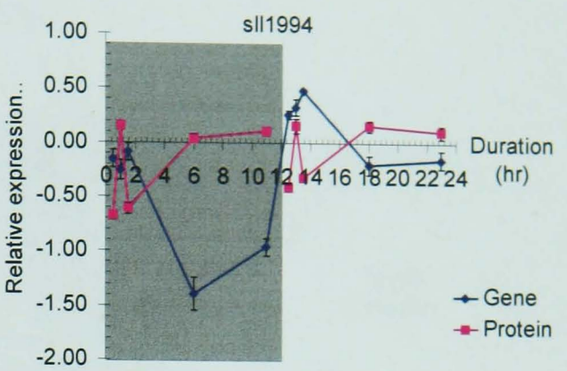

sll0945
Relative expression..
Duration (hr)
Gene
Protein

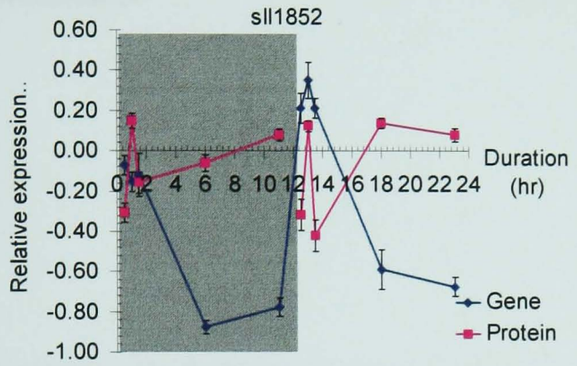

sll1356
Relative expression..
Duration (hr)
Gene
Protein

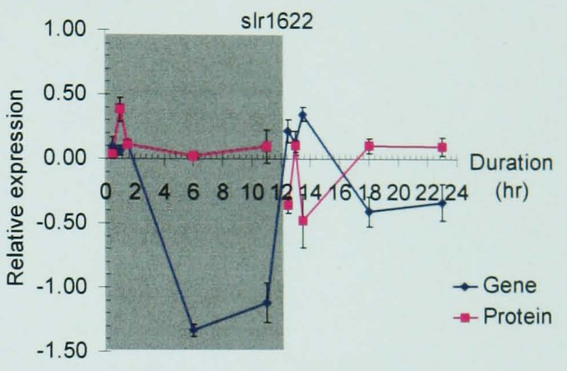

slr2076
Relative expression..
Duration (hr)
Gene
Protein

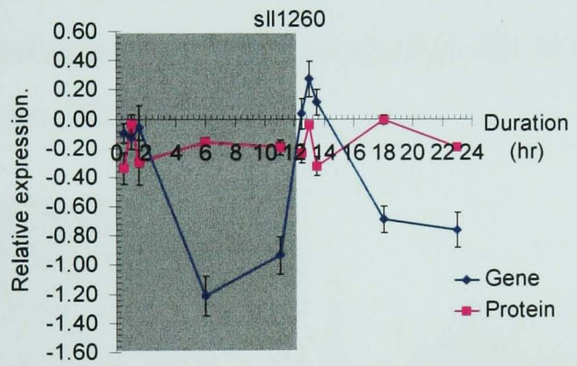

slr1963
Relative expression..
Duration (hr)
Gene
Protein

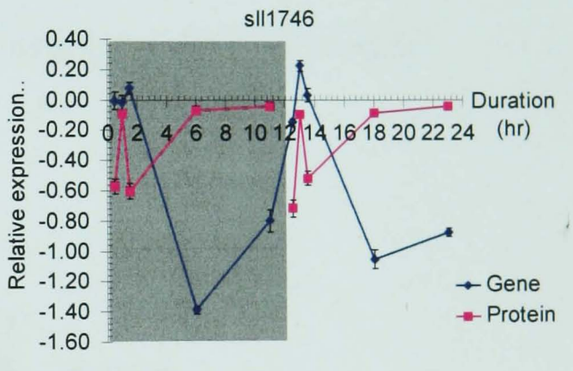

sll1577
Relative expression..
Duration (hr)
Gene
Protein

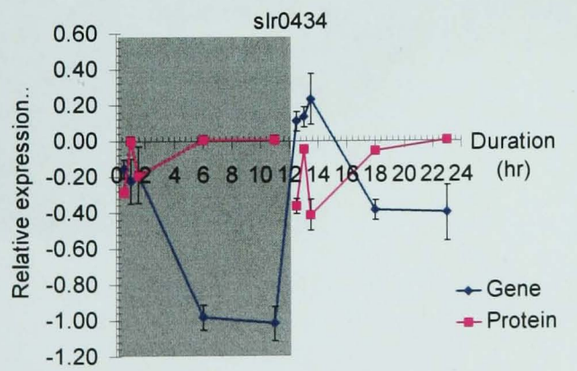

sll1994
Relative expression..
Duration (hr)
Gene
Protein

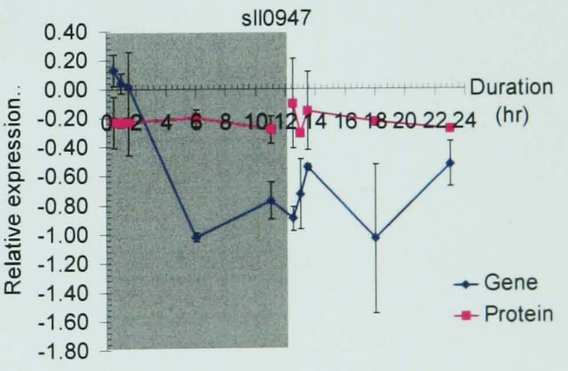

sll1852
Relative expression..
Duration (hr)
Gene
Protein


slr1622
Relative expression
Duration (hr)
Gene
Protein

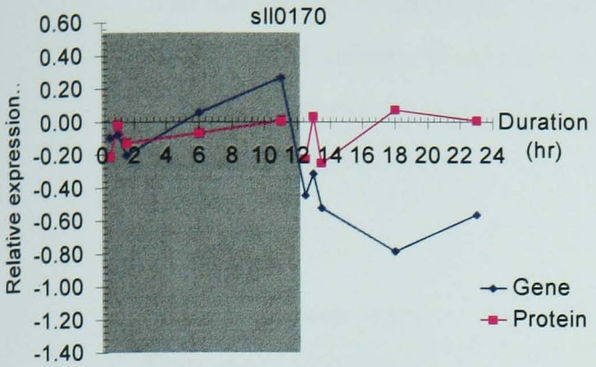

sll1260
Relative expression.
Duration (hr)
Gene
Protein


sll1746
Relative expression..
Duration (hr)
Gene
Protein

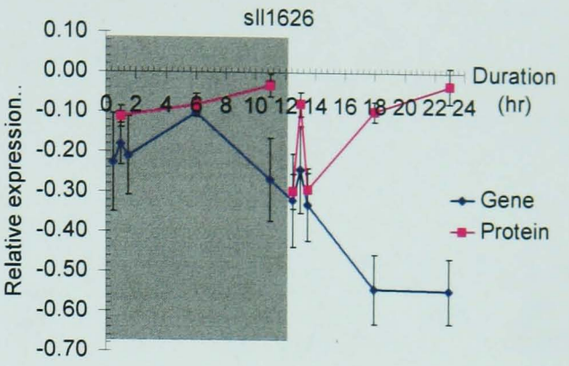

slr0434
Relative expression..
Duration (hr)
Gene
Protein


sll0947
Relative expression..
Duration (hr)
Gene
Protein

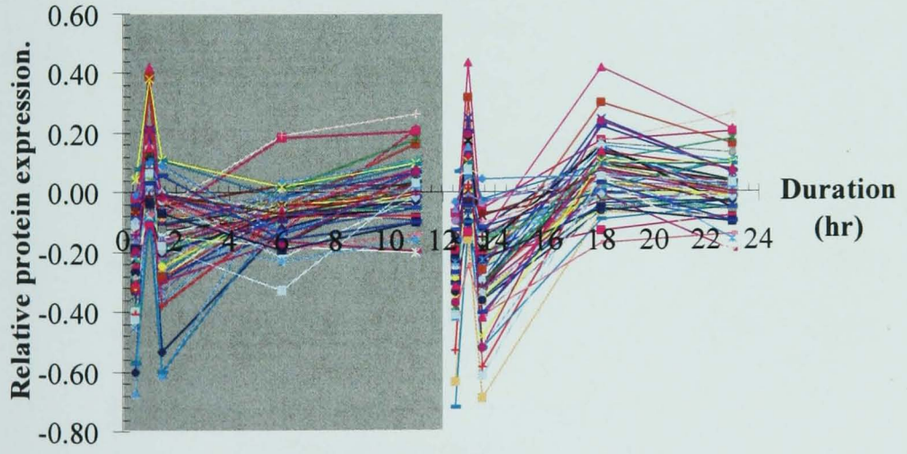
B1-type



B2-type



A very distinctive proteome profile across 12-hr dark and light cycle found in 64 out of 82 selected proteins (78%).



Appendix N

The full list of 170 confidently identified proteins (≥2 MS/MS) in Chapter 7, with their relative abundance at various Pi-reduced levels. The relative protein ratio is shown in $\log_{10}$ space with the error estimation in standard deviation (SD). All relative ratios are against control phosphate-replete culture (100% Pi).

			10% Pi		3% Pi		0.3% Pi	
ORF	**Protein Name**	**MS/ MS**	**Log Ratio**	**SD**	**Log Ratio**	**SD**	**Log Ratio**	**SD**
Amino Acids Metabolism								
sll0902	ornithine carbamoyltransferase argF	2	-0.122	0.088	0.084	0.062	0.130	0.040
sll1058	dihydrodipicolinate reductase dapB	3	-0.171	0.037	-0.220	0.064	-0.219	0.072
sll1234	adenosylhomocysteinase	5	-0.107	0.060	0.210	0.160	0.203	0.169
sll1363	ketol-acid reductoisomerase ilvC	3	0.508	0.112	0.248	0.174	0.334	0.054
sll1682	alanine dehydrogenase	2	-0.309	0.057	-0.196	0.009	-0.147	0.039
sll1750	urease alpha subunit ureC	2	-0.193	0.011	-0.096	0.005	-0.155	0.029
sll1958	histidinol phosphate aminotransferase hisC	2	-0.132	0.006	-0.063	0.033	-0.056	0.031
sll1981	acetolactate synthase ilvB	2	-0.008	0.036	0.183	0.019	0.128	0.004
slr0032	probable branched-chain amino acid aminotransferase	2	0.102	0.079	0.242	0.061	0.202	0.066
slr0370	succinate-semialdehyde dehydrogenase (NADP+)	3	0.073	0.259	0.174	0.297	0.077	0.351
slr0452	dihydroxyacid dehydratase ilvD	3	0.170	0.075	0.323	0.087	0.509	0.069
slr0652	phosphorybosilformimino-5-amino- phosphorybosil-4-imidazolecarboxamideisomerase hisA	2	0.165	0.000	0.357	0.044	0.387	0.010
slr0676	adenylylsulfate kinase cysC	4	-0.037	0.089	0.129	0.075	0.059	0.101
slr0710	glutamate dehydrogenase (NADP+) gdhA	3	0.198	0.083	0.205	0.065	0.125	0.154
slr1756	glutamate--ammonia ligase glnA	8	0.388	0.109	0.415	0.143	0.327	0.142
Carbohydrate metabolism								
sll0158	1,4-alpha-glucan branching enzyme glgB	2	0.182	0.013	0.170	0.075	0.103	0.051
sll1356	glycogen phosphorylase	5	0.342	0.143	0.359	0.129	0.243	0.162
sll1358	putative oxalate decarboxylase, periplasmic protein	3	-0.155	0.108	-0.022	0.125	0.045	0.114
slr1096	dihydrolipoamide dehydrogenase	2	0.431	0.183	0.501	0.207	0.620	0.209
slr1176	glucose-1-phosphate adenylyltransferase	3	-0.032	0.067	0.086	0.085	0.157	0.132
slr1857	isoamylase	2	-0.131	0.079	-0.012	0.067	0.062	0.055
CO2 Fixation								
sll1525	phosphoribulokinase prk	12	0.188	0.070	0.203	0.063	0.154	0.074
sll1029	carbon dioxide concentrating mechanism protein CcmK ccmK1	5	0.255	0.142	0.206	0.151	0.311	0.153

ORF	Protein Name	MS/MS	10% Pi		3% Pi		0.3% Pi	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll0272	hypothetical protein	4	0.203	0.172	0.175	0.212	0.331	0.233
sll0274	(Pentapeptide)	2	0.224	0.002	0.257	0.010	0.309	0.007
sll0529	(Transketolase)	4	0.335	0.050	0.394	0.029	0.276	0.041
sll0588	hypothetical protein	2	-0.449	0.020	-0.228	0.010	-0.519	0.009
sll0822	hypothetical protein	2	-0.922	0.307	-0.895	0.353	-0.402	0.089
sll0995	hypothetical protein	2	0.058	0.081	0.127	0.132	0.364	0.136
sll1106	hypothetical protein	12	0.105	0.086	0.232	0.097	0.122	0.103
sll1118	hypothetical protein	2	-0.336	0.026	-0.237	0.062	-0.334	0.031
sll1130	hypothetical protein	2	-0.023	0.139	0.121	0.191	0.107	0.089
sll1188	(Cupin superfamily)	2	-0.057	0.706	-0.076	0.787	-0.068	0.744
sll1201	(Putative transposase)	2	0.185	0.150	0.232	0.175	0.363	0.195
sll1526	(CheR methyltransferase)	2	-0.097	0.021	-0.096	0.089	-0.098	0.030
sll1873	hypothetical protein	3	-1.477	0.115	-1.585	0.333	-1.102	0.050
sll1891	hypothetical protein	3	0.248	0.156	0.415	0.166	0.431	0.143
slr0001	(Flavin reductase like domain)	5	0.446	0.058	0.584	0.055	0.660	0.057
slr0013	hypothetical protein	2	-0.329	0.181	-0.082	0.207	-0.150	0.262
slr0244	(Universal stress protein family)	2	-0.479	0.023	-0.258	0.041	-0.267	0.005
slr0376	hypothetical protein	2	0.636	0.181	0.860	0.150	0.861	0.191
slr0476	hypothetical protein	2	0.287	0.052	0.443	0.015	0.652	0.028
slr0605	(putative glutathione S-transferase)	2	0.059	0.005	0.059	0.037	0.151	0.031
slr0606	(Glycosyltransferase family)	2	0.182	0.013	0.310	0.010	0.556	0.033
slr0729	hypothetical protein	11	0.187	0.097	0.358	0.093	0.446	0.090
slr0923	hypothetical protein YCF65 ycf65	3	-0.048	0.087	0.091	0.067	0.410	0.112
slr1034	hypothetical protein YCF41 ycf41	4	0.279	0.072	0.366	0.079	0.356	0.043
slr1590	hypothetical protein	2	0.177	0.067	0.246	0.037	0.293	0.007
slr1612	hypothetical protein	2	0.154	0.068	0.724	0.180	0.900	0.131
slr1704	hypothetical protein	2	-0.239	0.007	-0.067	0.010	1.004	0.011
slr1780	hypothetical protein YCF54 ycf54	3	-0.031	0.371	0.001	0.370	0.166	0.244
slr1847	hypothetical protein	2	-0.084	0.041	0.030	0.044	1.036	0.201
slr1852	hypothetical protein	9	0.143	0.201	0.376	0.208	0.469	0.213
slr1855	hypothetical protein	3	-0.052	0.079	0.217	0.066	0.319	0.091
slr2032	hypothetical protein YCF23 ycf23	4	0.185	0.105	0.339	0.134	0.212	0.167
ssl0385	hypothetical protein	2	0.588	0.137	0.941	0.133	1.037	0.124
ssl2501	hypothetical protein	2	0.665	0.243	1.118	0.246	1.266	0.258
ssr1528	hypothetical protein	3	0.600	0.101	0.568	0.108	0.448	0.105
ssr1853	hypothetical protein	5	0.235	0.190	0.349	0.241	0.291	0.265
Lipid Metabolism								
slr1992	glutathione peroxidase-like NADPH peroxidase gpx2	9	0.036	0.134	0.321	0.140	0.490	0.151
sll0067	glutathione S-transferase	3	0.330	0.067	0.549	0.069	0.671	0.106
sll1621	glutathion-dependent peroxidase, type II peroxiredoxin	8	-0.283	0.091	-0.252	0.102	-0.062	0.120
slr0435	biotin carboxyl carrier protein of acetyl-CoA carboxylase accB	3	-0.053	0.026	0.076	0.021	0.004	0.053
slr1994	PHA-specific acetoacetyl-CoA reductase phaB	3	0.088	0.065	0.322	0.147	0.473	0.102

ORF	Protein Name	MS/MS	10% Pi Log Ratio	10% Pi SD	3% Pi Log Ratio	3% Pi SD	0.3% Pi Log Ratio	0.3% Pi SD
Photosynthesis								
sll0199	plastocyanin petE	15	0.356	0.077	0.613	0.095	0.699	0.114
sll0258	cytochrome c550 psbV	4	0.530	0.310	0.602	0.308	0.568	0.294
sll0427	photosystem II manganese-stabilizing polypeptide psbO	7	-0.136	0.070	-0.093	0.065	-0.046	0.053
sll0928	allophycocyanin-B apcD	7	-0.177	0.049	-0.146	0.070	-0.040	0.070
sll1194	photosystem II 12 kDa extrinsic protein psbU	12	0.088	0.089	0.135	0.077	0.183	0.080
sll1577	phycocyanin beta subunit cpcB	126	0.419	0.037	0.326	0.037	0.141	0.034
sll1578	phycocyanin alpha subunit cpcA	68	0.080	0.079	-0.076	0.074	-0.183	0.072
sll1580	phycobilisome rod linker polypeptide cpcC1	5	-0.371	0.065	-0.634	0.163	-0.495	0.154
sll1663	phycocyanin alpha phycocyanobilin lyase related protein	3	-0.017	0.034	0.054	0.020	0.179	0.011
slr0335	phycobilisome core-membrane linker polypeptide apcE	4	-0.942	0.262	-1.188	0.288	-0.822	0.135
slr0737	photosystem I subunit II psaD	3	-0.680	0.040	-0.926	0.047	-0.869	0.080
slr1643	ferredoxin-NADP oxidoreductase petH	3	0.475	0.166	0.611	0.106	0.872	0.148
slr1834	P700 apoprotein subunit Ia psaA	2	-0.868	0.003	-1.254	0.092	-1.365	0.097
slr1986	allophycocyanin beta subunit apcB	36	0.105	0.069	0.127	0.066	-0.057	0.075
slr2051	phycobilisome rod-core linker polypeptide cpcG1	12	-0.690	0.131	-0.592	0.155	-0.040	0.164
slr2067	allophycocyanin alpha subunit apcA	9	-0.067	0.076	-0.161	0.111	-0.211	0.104
ssr3383	phycobilisome small core linker polypeptide apcC	2	0.048	0.017	0.161	0.031	0.060	0.056
Porphyrin and Chlorophyll Metabolism								
slr1368	precorrin decarbocylase cobL	2	0.008	0.061	0.167	0.054	0.287	0.114
Replication and Repair								
sll1712	DNA binding protein HU	4	0.131	0.202	-0.051	0.203	0.370	0.201
slr1894	probable DNA-binding stress protein	11	0.030	0.129	-0.002	0.134	0.023	0.055
Signal Transduction								
sll1590	two-component sensor histidine kinase hik20	3	-0.642	0.047	-0.865	0.020	-0.944	0.053
ssl0707	nitrogen regulatory protein P-II glnB	14	-0.389	0.091	-0.456	0.093	-0.452	0.084
TCA Cycle								
sll0401	citrate synthase	4	0.188	0.123	0.308	0.145	0.214	0.158
slr1289	isocitrate dehydrogenase (NADP+) icd	2	-0.108	0.061	-0.045	0.115	0.038	0.171
Transcription								
sll0271	N utilization substance protein B homolog	2	0.024	0.042	0.186	0.126	0.364	0.122

[13] Backasch, N., Schulz-Friedrich, R. and Appel, J. (2005) "Influences on tocopherol biosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803". *J Plant Physiol*, **162** (7), 758-66.

[14] Maeda, H., Sakuragi, Y., Bryant, D.A. and Dellapenna, D. (2005) "Tocopherols protect *Synechocystis* sp. strain PCC 6803 from lipid peroxidation". *Plant Physiol*, **138** (3), 1422-35.

[15] Sakuragi, Y., Maeda, H., Dellapenna, D. and Bryant, D.A. (2006) "alpha-Tocopherol plays a role in photosynthesis and macronutrient homeostasis of the cyanobacterium *Synechocystis* sp. PCC 6803 that is independent of its antioxidant function". *Plant Physiol*, **141** (2), 508-21.

[16] Lindblad, P. (1999) "Cyanobacterial H2 Metabolism: Knowledge and Potential/Strategies for a Photobiotechnological Production of H2". *Biotecnologia Aplicada*, **16**, 141-144.

[17] Dutta, D., De, D., Chaudhuri, S. and Bhattacharya, S.K. (2005) "Hydrogen production by Cyanobacteria". *Microb Cell Fact*, **4**, 36.

[18] Block, D.L. and Melody, I. (1992) "Efficiency and cost goals for photoenhanced hydrogen production processes". *Int J Hydrogen Energy*, **17**, 853-861.

[19] Levin, D.B., Pitt, L. and Love, M. (2004) "Biohydrogen production: prospects and limitations to practical application". *Int J Hydrogen Energy*, **29**, 173-185.

[20] Obst, M. and Steinbuchel, A. (2004) "Microbial degradation of poly(amino acid)s". *Biomacromolecules*, **5** (4), 1166-76.

[21] Schwamborn, M. (1998) "Chemical synthesis of polyaspartates: a biodegradable alternative to currently used polycarboxylate homo- and copolymers". *Polym. Degrad. Stabil.*, **59**, 39-45.

[22] Steinbuchel, A. and Valentin, H.E. (1995) "Diversity of bacterial polyhydroxyalkanoic acids." *FEMS Microbiol Lett*, **128**, 219-228.

[23] Business Communications Company (2005). *Biodegradable Polymers*. Business Communications Company Inc., Norwalk, Conn. (Report No. **PLS025B**).

[24] Edge, R., McGarvey, D.J. and Truscott, T.G. (1997) "The carotenoids as anti-oxidants - a review". *Journal of Photochemistry and Photobiology B-Biology*, **41**, 189-200.

[25] Vershinin, A. (1999) "Biological functions of carotenoids--diversity and evolution". *Biofactors*, **10** (2-3), 99-104.

[26] Nabae, K., Ichihara, T., Hagiwara, A., Hirota, T., Toda, Y., Tamano, S., Nishino, M., Ogasawara, T., Sasaki, Y., Nakamura, M. and Shirai, T. (2005) "A 90-day oral toxicity study of beta-carotene derived from *Blakeslea trispora*, a natural food colorant, in F344 rats". *Food Chem Toxicol*, **43** (7), 1127-33.

[40] Schopf, J.W., (2000). *The fossil record: tracing the roots of the cyanobacterial lineage,* **In The ecology of cyanobacteria**. by Whitton, B.A. and Potts, M.: Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 13-35.

[41] Paerl, H.W., Pinckney, J.L. and Steppe, T.F. (2000) "Cyanobacterial-bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments". *Environ Microbiol*, **2** (1), 11-26.

[42] Burja, A.M., Dhamwichukorn, S. and Wright, P.C. (2003) "Cyanobacterial postgenomic research and systems biology". *Trends Biotechnol.*, **21** (11), 504-11.

[43] Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M. and Stanier, R.Y. (1979) "Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria". *J. Gen. Microbiol.*, **111**, 1-61.

[44] Castenholz, R.W., (2001). *Phylum BX. Cyanobacteria. General Characteristics of the Cyanobacteria,* **In Bergey's Manual of Systematic Bateriology, Volume 1: The Archaea and the Deeply Branching and Phototropic Bacteria**. by Garrity, G.M. and Boone, D.R.: New York: Springer-Verlag. pp 474-487.

[45] Burja, A.M., Banaigsb, B., Abou-Mansourc, E., Burgessd, J.G. and Wright, P.C. (2001) "Marine cyanobacteria-a prolific source of natural products". *Tetrahedron*, **57** (46), 9347-9377.

[46] Boyd, M.R., Gustafson, K.R., McMahon, J.B., Shoemaker, R.H., O'Keefe, B.R., Mori, T., Gulakowski, R.J., Wu, L., Rivera, M.I., Laurencot, C.M., Currens, M.J., Cardellina, J.H., 2nd, Buckheit, R.W., Jr., Nara, P.L., Pannell, L.K., Sowder, R.C., 2nd and Henderson, L.E. (1997) "Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development". *Antimicrob Agents Chemother*, **41** (7), 1521-30.

[47] Cox, P.A., Banack, S.A. and Murch, S.J. (2003) "Biomagnification of cyanobacterial neurotoxins and neurodegenerative disease among the Chamorro people of Guam". *Proc Natl Acad Sci U S A*, **100** (23), 13380-3.

[48] Namikoshi, M., Choi, B.W., Sun, F., Rinehart, K.L., Evans, W.R. and Carmichael, W.W. (1993) "Chemical characterization and toxicity of dihydro derivatives of nodularin and microcystin-LR, potent cyanobacterial cyclic peptide hepatotoxins". *Chem Res Toxicol*, **6** (2), 151-8.

[49] Kay, R.A. (1991) "Microalgae as food and supplement". *Crit Rev Food Sci Nutr*, **30** (6), 555-73.

[50] Wang, Y., Chang, C.F., Chou, J., Chen, H.L., Deng, X., Harvey, B.K., Cadet, J.L. and Bickford, P.C. (2005) "Dietary supplementation with blueberries, spinach, or spirulina reduces ischemic brain damage". *Exp Neurol*, **193** (1), 75-84.

[63] Yang, C., Hua, Q. and Shimizu, K. (2002) "Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*". *Appl Microbiol Biotechnol*, **58** (6), 813-22.

[64] Anderson, S.L. and McIntosh, L. (1991) "Light-activated heterotrophic growth of the cyanobacterium Synechocystis sp. strain PCC 6803: a blue-light-requiring process". *J Bacteriol*, **173** (9), 2761-7.

[65] Hihara, Y., Kamei, A., Kanehisa, M., Kaplan, A. and Ikeuchi, M. (2001) "DNA microarray analysis of cyanobacterial gene expression during acclimation to high light". *Plant Cell*, **13** (4), 793-806.

[66] Gill, R.T., Katsoulakis, E., Schmitt, W., Taroncher-Oldenburg, G., Misra, J. and Stephanopoulos, G. (2002) "Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **184** (13), 3671-81.

[67] Kurian, D., Jansen, T. and Maenpaa, P. (2006) "Proteomic analysis of heterotrophy in *Synechocystis* sp. PCC 6803". *Proteomics*, **6** (5), 1483-94.

[68] Tamoi, M., Miyazaki, T., Fukamizo, T. and Shigeoka, S. (2005) "The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions". *Plant J*, **42** (4), 504-13.

[69] Yang, C., Hua, Q. and Shimizu, K. (2002) "Quantitative analysis of intracellular metabolic fluxes using GC-MS and two-dimensional NMR spectroscopy". *J Biosci Bioeng*, **93** (1), 78-87.

[70] Asada, Y. and Miyake, J. (1999) "Photobiological hydrogen production". *J Biosci Bioeng*, **88** (1), 1-6.

[71] Cournac, L., Guedeney, G., Peltier, G. and Vignais, P.M. (2004) "Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis* sp. strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex". *J Bacteriol*, **186** (6), 1737-46.

[72] Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., Wunschiers, R. and Lindblad, P. (2002) "Hydrogenases and hydrogen metabolism of cyanobacteria". *Microbiol Mol Biol Rev*, **66** (1), 1-20.

[73] Appel, J. and Schulz, R. (1996) "Sequence analysis of an operon of a NAD(P)-reducing nickel hydrogenase from the cyanobacterium *Synechocystis* sp. PCC 6803 gives additional evidence for direct coupling of the enzyme to NAD(P)H-dehydrogenase (complex I)". *Biochim Biophys Acta*, **1298** (2), 141-7.

[74] Schmitz, O., Boison, G., Salzmann, H., Bothe, H., Schutz, K., Wang, S.H. and Happe, T. (2002) "HoxE--a subunit specific for the pentameric bidirectional hydrogenase complex (HoxEFUYH) of cyanobacteria". *Biochim Biophys Acta*, **1554** (1-2), 66-74.

[88] Joentgen, W., Groth, T., Steinbuchel, A., Hai, T. and Oppermann, F.B. (1998) "Polyaspartic acid homopolymers and copolymers: biotechnical production and use thereof". International patent application WO 98/39090.

[89] Richter, R., Hejazi, M., Kraft, R., Ziegler, K. and Lockau, W. (1999) "Cyanophycinase, a peptidase degrading the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartic acid (cyanophycin): molecular cloning of the gene of *Synechocystis* sp. PCC 6803, expression in *Escherichia coli*, and biochemical characterization of the purified enzyme". *Eur J Biochem*, **263** (1), 163-9.

[90] Aboulmagd, E., Oppermann-Sanio, F.B. and Steinbuchel, A. (2000) "Molecular characterization of the cyanophycin synthetase from *Synechocystis* sp. strain PCC6308". *Arch Microbiol*, **174** (5), 297-306.

[91] Mackerras, A.H., Chazal, N.M.d. and Smith, G.D. (1990) "Transient accumulation of cyanophycin in *Anabaena cylindrica* and *Synechocystis* 6308". *J Gen Microbiol*, **136**, 2057-2065.

[92] Berg, H., Ziegler, K., Piotukh, K., Baier, K., Lockau, W. and Volkmer-Engert, R. (2000) "Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin): Mechanism of the cyanophycin synthetase reaction studied with synthetic primers". *Eur. J. Biochem*, **267**, 5561-5570.

[93] Merritt, M.V., Sid, S.S., Mesh, L. and Allen, M.M. (1994) "Variations in the amino acid composition of cyanophycin in the cyanobacterium *Synechocystis* sp. PCC 6308 as a function of growth conditions". *Arch Microbiol*, **162** (3), 158-66.

[94] Suarez, C., Kohler, S.J., Allen, M.M. and Kolodny, N.H. (1999) "NMR study of the metabolic 15N isotopic enrichment of cyanophycin synthesized by the cyanobacterium *Synechocystis* sp. strain PCC 6308". *Biochim Biophys Acta*, **1426** (3), 429-38.

[95] Stubbe, J., Tian, J., He, A., Sinskey, A.J., Lawrence, A.G. and Liu, P. (2005) "Nontemplate-dependent polymerization processes: polyhydroxyalkanoate synthases as a paradigm". *Annu Rev Biochem*, **74**, 433-80.

[96] Allen, M.M. and Hutchison, F. (1980) "Nitrogen limitation and recovery in the cyanobacterium *Aphanocapsa* 6308". *Arch. Microbiol.*, **128**, 1-7.

[97] Simon, R.D. (1973) "The effect of chloramphenicol on the production of cyanophycin granule polypeptide in the blue green alga *Anabaena cylindrica*". *Arch Mikrobiol*, **92** (2), 115-22.

[98] Allen, M.M., Hutchison, F. and Weathers, P.J. (1980) "Cyanophycin granule polypeptide formation and degradation in the cyanobacterium *Aphanocapsa* 6308". *J Bacteriol*, **141** (2), 687-93.

[99] Kolodny, N.H., Bauer, D., Bryce, K., Klucevsek, K., Lane, A., Medeiros, L., Mercer, W., Moin, S., Park, D., Petersen, J., Wright, J., Yuen, C., Wolfson, A.J. and Allen,

coenzyme A reductase genes in the cyanobacterium *Synechocystis* sp. strain PCC6803". *Appl Environ Microbiol*, **66** (10), 4440-8.

[111] Dawes, E.A. and Senior, P.J. (1973) "The role and regulation of energy reserve polymers in micro-organisms". *Adv Microb Physiol*, **10**, 135-266.

[112] Hai, T., Hein, S. and Steinbuchel, A. (2001) "Multiple evidence for widespread and general occurrence of type-III PHA synthases in cyanobacteria and molecular characterization of the PHA synthases from two thermophilic cyanobacteria: *Chlorogloeopsis fritschii* PCC 6912 and *Synechococcus* sp. strain MA19". *Microbiology*, **147** (Pt 11), 3047-60.

[113] Madison, L.L. and Huisman, G.W. (1999) "Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic". *Microbiol Mol Biol Rev*, **63** (1), 21-53.

[114] Wu, G.F., Shen, Z.Y. and Wu, Q.Y. (2002) "Modification of carbon partitioning to enhance Poly-beta-hydroxybutyrate (PHB) production in *Synechocystis* sp. PCC6803". *Enzyme and Microbial Technology*, **30** (6), 710-715.

[115] Panda, B., Sharma, L. and Mallick, N. (2005) "Poly-beta-hydroxybutyrate accumulation in *Nostoc muscorum* and *Spirulina platensis* under phosphate limitation". *J Plant Physiol*, **162** (12), 1376-9.

[116] Han, M.J., Yoon, S.S. and Lee, S.Y. (2001) "Proteome analysis of metabolically engineered *Escherichia coli* producing Poly(3-hydroxybutyrate)". *J Bacteriol*, **183** (1), 301-8.

[117] Tyo, K.E., Zhou, H. and Stephanopoulos, G.N. (2006) "High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp. strain PCC6803". *Appl Environ Microbiol*, **72** (5), 3412-7.

[118] Schafer, L., Vioque, A. and Sandmann, G. (2005) "Functional in situ evaluation of photosynthesis-protecting carotenoids in mutants of the cyanobacterium *Synechocystis* PCC6803". *J Photochem Photobiol B*, **78** (3), 195-201.

[119] Takaichi, S., Maoka, T. and Masamoto, K. (2001) "Myxoxanthophyll in *Synechocystis* sp. PCC 6803 is myxol 2'-dimethyl-fucoside, (3R,2'S)-myxol 2'-(2,4-di-O-methyl-alpha-L-fucoside), not rhamnoside". *Plant Cell Physiol*, **42** (7), 756-62.

[120] Cuttriss, A.J. and Pogson, B.J., (2006). *Carotenoids*, **The Structure and Function of Plastids**. Wise, R.R. and Hoober, J.K.: Dordrecht, The Netherlands: Springer. 315-34.

[121] Holt, N.E., Fleming, G.R. and Niyogi, K.K. (2004) "Toward an understanding of the mechanism of nonphotochemical quenching in green plants". *Biochemistry*, **43** (26), 8281-9.

[122] Niyogi, K.K. (1999) "Photoprotection revisited: genetic and molecular approaches." *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **50**, 333-59.

[123] Zingg, J.M. and Azzi, A. (2004) "Non-antioxidant activities of vitamin E". *Curr Med Chem*, **11** (9), 1113-33.

[136] Dahnhardt, D., Falk, J., Appel, J., van der Kooij, T.A., Schulz-Friedrich, R. and Krupinska, K. (2002) "The hydroxyphenylpyruvate dioxygenase from *Synechocystis* sp. PCC 6803 is not required for plastoquinone biosynthesis". *FEBS Lett*, **523** (1-3), 177-81.

[137] Schledz, M., Seidler, A., Beyer, P. and Neuhaus, G. (2001) "A novel phytyltransferase from *Synechocystis* sp. PCC 6803 involved in tocopherol biosynthesis". *FEBS Lett*, **499** (1-2), 15-20.

[138] Savidge, B., Weiss, J.D., Wong, Y.H., Lassner, M.W., Mitsky, T.A., Shewmaker, C.K., Post-Beittenmiller, D. and Valentin, H.E. (2002) "Isolation and characterization of homogentisate phytyltransferase genes from *Synechocystis* sp. PCC 6803 and *Arabidopsis*". *Plant Physiol*, **129** (1), 321-32.

[139] Shintani, D.K., Cheng, Z. and DellaPenna, D. (2002) "The role of 2-methyl-6-phytylbenzoquinone methyltransferase in determining tocopherol composition in *Synechocystis* sp. PCC6803". *FEBS Lett*, **511** (1-3), 1-5.

[140] Sattler, S.E., Cahoon, E.B., Coughlan, S.J. and DellaPenna, D. (2003) "Characterization of tocopherol cyclases from higher plants and cyanobacteria. Evolutionary implications for tocopherol synthesis and function". *Plant Physiol*, **132** (4), 2184-95.

[141] Porfirova, S., Bergmuller, E., Tropf, S., Lemke, R. and Dormann, P. (2002) "Isolation of an *Arabidopsis* mutant lacking vitamin E and identification of a cyclase essential for all tocopherol biosynthesis". *Proc Natl Acad Sci U S A*, **99**, 12495-500.

[142] Shintani, D. and DellaPenna, D. (1998) "Elevating the vitamin E content of plants through metabolic engineering". *Science*, **282**, 2098-100.

[143] Valentin, H.E., Lincoln, K., Moshiri, F., Jensen, P.K., Qi, Q., Venkatesh, T.V., Karunanandaa, B., Baszis, S.R., Norris, S.R., Savidge, B., Gruys, K.J. and Last, R.L. (2006) "The *Arabidopsis* vitamin E pathway gene5-1 mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis". *Plant Cell*, **18** (1), 212-24.

[144] Ryu, J.Y., Song, J.Y., Lee, J.M., Jeong, S.W., Chow, W.S., Choi, S.B., Pogson, B.J. and Park, Y.I. (2004) "Glucose-induced expression of carotenoid biosynthesis genes in the dark is mediated by cytosolic ph in the cyanobacterium *Synechocystis* sp. PCC 6803". *J Biol Chem*, **279** (24), 25320-5.

[145] Bautista, J.A., Rappaport, F., Guergova-Kuras, M., Cohen, R.O., Golbeck, J.H., Wang, J.Y., Beal, D. and Diner, B.A. (2005) "Biochemical and biophysical characterization of photosystem I from phytoene desaturase and zeta-carotene desaturase deletion mutants of *Synechocystis* Sp. PCC 6803: evidence for PsaA- and PsaB-side electron transport in cyanobacteria". *J Biol Chem*, **280** (20), 20030-41.

[146] Lagarde, D., Beuf, L. and Vermaas, W. (2000) "Increased production of zeaxanthin and other pigments by application of genetic engineering techniques to *Synechocystis* sp. strain PCC 6803". *Appl Environ Microbiol*, **66** (1), 64-72.

[160] Terauchi, K. and Ohmori, M. (2004) "Blue light stimulates cyanobacterial motility via a cAMP signal transduction system". *Mol Microbiol*, **52** (1), 303-9.

[161] Steglich, C., Futschik, M., Rector, T., Steen, R. and Chisholm, S.W. (2006) "Genome-wide analysis of light sensing in *prochlorococcus*". *J Bacteriol*, **188** (22), 7796-806.

[162] Zsiros, O., Allakhverdiev, S.I., Higashi, S., Watanabe, M., Nishiyama, Y. and Murata, N. (2006) "Very strong UV-A light temporally separates the photoinhibition of photosystem II into light-induced inactivation and repair". *Biochim Biophys Acta*, **1757** (2), 123-9.

[163] Sicora, C., Mate, Z. and Vass, I. (2003) "The interaction of visible and UV-B light during photodamage and repair of Photosystem II". *Photosynth Res*, **75** (2), 127-37.

[164] Huang, L., McCluskey, M.P., Ni, H. and LaRossa, R.A. (2002) "Global gene expression profiles of the cyanobacterium *Synechocystis* sp. strain PCC 6803 in response to irradiation with UV-B and white light". *J Bacteriol*, **184** (24), 6845-58.

[165] Muller, P., Li, X.P. and Niyogi, K.K. (2001) "Non-photochemical quenching. A response to excess light energy". *Plant Physiol*, **125** (4), 1558-66.

[166] Wilson, A., Ajlani, G., Verbavatz, J.M., Vass, I., Kerfeld, C.A. and Kirilovsky, D. (2006) "A soluble carotenoid protein involved in phycobilisome-related energy dissipation in cyanobacteria". *Plant Cell*, **18** (4), 992-1007.

[167] Nixon, P.J., Barker, M., Boehm, M., de Vries, R. and Komenda, J. (2005) "FtsH-mediated repair of the photosystem II complex in response to light stress". *J Exp Bot*, **56** (411), 357-63.

[168] Singh, M., Yamamoto, Y., Satoh, K., Aro, E.M. and Kanervo, E. (2005) "Post-illumination-related loss of photochemical efficiency of Photosystem II and degradation of the D1 protein are temperature-dependent". *J Plant Physiol*, **162** (11), 1246-53.

[169] Havaux, M., Guedeney, G., Hagemann, M., Yeremenko, N., Matthijs, H.C. and Jeanjean, R. (2005) "The chlorophyll-binding protein IsiA is inducible by high light and protects the cyanobacterium *Synechocystis* PCC6803 from photooxidative stress". *FEBS Lett*, **579** (11), 2289-93.

[170] Kamata, T., Hiramoto, H., Morita, N., Shen, J.R., Mann, N.H. and Yamamoto, Y. (2005) "Quality control of Photosystem II: an FtsH protease plays an essential role in the turnover of the reaction center D1 protein in *Synechocystis* PCC 6803 under heat stress as well as light stress conditions". *Photochem Photobiol Sci*, **4** (12), 983-90.

[171] Mary, I., Tu, C.J., Grossman, A. and Vaulot, D. (2004) "Effects of high light on transcripts of stress-associated genes for the cyanobacteria *Synechocystis* sp. PCC 6803 and *Prochlorococcus* MED4 and MIT9313". *Microbiology*, **150** (Pt 5), 1271-81.

[172] Kim, J.H. and Suh, K.H. (2005) "Light-dependent expression of superoxide dismutase from cyanobacterium *Synechocystis* sp. strain PCC 6803". *Arch Microbiol*, **183** (3), 218-23.

[184] Tsvetkova, N.M., Horvath, I., Torok, Z., Wolkers, W.F., Balogi, Z., Shigapova, N., Crowe, L.M., Tablin, F., Vierling, E., Crowe, J.H. and Vigh, L. (2002) "Small heat-shock proteins regulate membrane lipid polymorphism". *Proc Natl Acad Sci U S A*, **99** (21), 13504-9.

[185] Balogi, Z., Torok, Z., Balogh, G., Josvay, K., Shigapova, N., Vierling, E., Vigh, L. and Horvath, I. (2005) ""Heat shock lipid" in cyanobacteria during heat/light-acclimation". *Arch Biochem Biophys*, **436** (2), 346-54.

[186] Fang, F. and Barnum, S.R. (2004) "Expression of the heat shock gene hsp16.6 and promoter analysis in the cyanobacterium, *Synechocystis* sp. PCC 6803". *Curr Microbiol*, **49** (3), 192-8.

[187] Suzuki, I., Kanesaki, Y., Hayashi, H., Hall, J.J., Simon, W.J., Slabas, A.R. and Murata, N. (2005) "The histidine kinase Hik34 is involved in thermotolerance by regulating the expression of heat shock genes in *Synechocystis*". *Plant Physiol*, **138** (3), 1409-21.

[188] Slabas, A.R., Suzuki, I., Murata, N., Simon, W.J. and Hall, J.J. (2006) "Proteomic analysis of the heat shock response in *Synechocystis* PCC6803 and a thermally tolerant knockout strain lacking the histidine kinase 34 gene". *Proteomics*, **6** (3), 845-64.

[189] Singh, A.K., Summerfield, T.C., Li, H. and Sherman, L.A. (2006) "The heat shock response in the cyanobacterium *Synechocystis* sp. Strain PCC 6803 and regulation of gene expression by HrcA and SigB". *Arch Microbiol*, **186** (4), 273-86.

[190] Kanesaki, Y., Suzuki, I., Allakhverdiev, S.I., Mikami, K. and Murata, N. (2002) "Salt stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis* sp. PCC 6803". *Biochem Biophys Res Commun*, **290** (1), 339-48.

[191] Hagemann, M., Richter, S., Zuther, E. and Schoor, A. (1996) "Characterization of a glucosylglycerol-phosphate-accumulating, salt-sensitive mutant of the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Arch Microbiol*, **166** (2), 83-91.

[192] Mikkat, S., Hagemann, M. and Schoor, A. (1996) "Active transport of glucosylglycerol is involved in salt adaptation of the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Microbiology*, **142 (Pt 7)**, 1725-32.

[193] Hagemann, M., Schoor, A., Jeanjean, R., Zuther, E. and Joset, F. (1997) "The stpA gene form *Synechocystis* sp. strain PCC 6803 encodes the glucosylglycerol-phosphate phosphatase involved in cyanobacterial osmotic response to salt shock". *J Bacteriol*, **179** (5), 1727-33.

[194] Marin, K., Huckauf, J., Fulda, S. and Hagemann, M. (2002) "Salt-dependent expression of glucosylglycerol-phosphate synthase, involved in osmolyte synthesis in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **184** (11), 2870-7.

[195] Miao, X., Wu, Q., Wu, G. and Zhao, N. (2003) "Sucrose accumulation in salt-stressed cells of agp gene deletion-mutant in cyanobacterium *Synechocystis* sp PCC 6803". *FEMS Microbiol Lett*, **218** (1), 71-7.

regulate the expression of individual genes to different extents in *Synechocystis*". *J Biol Chem*, **280** (22), 21531-8.

[208] Asadulghani, Nitta, K., Kaneko, Y., Kojima, K., Fukuzawa, H., Kosaka, H. and Nakamoto, H. (2004) "Comparative analysis of the hspA mutant and wild-type *Synechocystis* sp. strain PCC 6803 under salt stress: evaluation of the role of hspA in salt-stress management". *Arch Microbiol*, **182** (6), 487-97.

[209] Fulda, S., Mikkat, S., Huang, F., Huckauf, J., Marin, K., Norling, B. and Hagemann, M. (2006) "Proteome analysis of salt stress response in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Proteomics*, **6** (9), 2733-45.

[210] Huang, F., Fulda, S., Hagemann, M. and Norling, B. (2006) "Proteomic screening of salt-stress-induced changes in plasma membranes of *Synechocystis* sp. strain PCC 6803". *Proteomics*, **6** (3), 910-20.

[211] Hoi le, T., Voigt, B., Jurgen, B., Ehrenreich, A., Gottschalk, G., Evers, S., Feesche, J., Maurer, K.H., Hecker, M. and Schweder, T. (2006) "The phosphate-starvation response of *Bacillus licheniformis*". *Proteomics*, **6** (12), 3582-601.

[212] Redding, A.M., Mukhopadhyay, A., Joyner, D.C., Hazen, T.C. and Keasling, J.D. (2006) "Study of nitrate stress in *Desulfovibrio vulgaris* Hildenborough using iTRAQ proteomics". *Brief Funct Genomic Proteomic*, **5** (2), 133-43.

[213] Singh, A.K., McIntyre, L.M. and Sherman, L.A. (2003) "Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium *Synechocystis* sp. PCC 6803". *Plant Physiol*, **132** (4), 1825-39.

[214] Zhao, J.D. and Brand, J.J. (1988) "Sequential effects of sodium depletion on photosystem II in *Synechocystis*". *Arch Biochem Biophys*, **264** (2), 657-64.

[215] Koropatkin, N.M., Pakrasi, H.B. and Smith, T.J. (2006) "Atomic structure of a nitrate-binding protein crucial for photosynthetic productivity". *Proc Natl Acad Sci U S A*, **103** (26), 9820-5.

[216] Muro-Pastor, A.M., Herrero, A. and Flores, E. (2001) "Nitrogen-regulated group 2 sigma factor from *Synechocystis* sp. strain PCC 6803 involved in survival under nitrogen stress". *J Bacteriol*, **183** (3), 1090-5.

[217] Imamura, S., Tanaka, K., Shirai, M. and Asayama, M. (2006) "Growth phase-dependent activation of nitrogen-related genes by a control network of group 1 and group 2 sigma factors in a cyanobacterium". *J Biol Chem*, **281** (5), 2668-75.

[218] Asayama, M., Imamura, S., Yoshihara, S., Miyazaki, A., Yoshida, N., Sazuka, T., Kaneko, T., Ohara, O., Tabata, S., Osanai, T., Tanaka, K., Takahashi, H. and Shirai, M. (2004) "SigC, the group 2 sigma factor of RNA polymerase, contributes to the late-stage gene expression and nitrogen promoter recognition in the cyanobacterium *Synechocystis* sp. strain PCC 6803". *Biosci Biotechnol Biochem*, **68** (3), 477-87.

[231] Tirupati, B., Vey, J.L., Drennan, C.L. and Bollinger, J.M., Jr. (2004) "Kinetic and structural characterization of Slr0077/SufS, the essential cysteine desulfurase from *Synechocystis* sp. PCC 6803". *Biochemistry*, **43** (38), 12210-9.

[232] Wang, T., Shen, G., Balasubramanian, R., McIntosh, L., Bryant, D.A. and Golbeck, J.H. (2004) "The sufR gene (sll0088 in *Synechocystis* sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria". *J Bacteriol*, **186** (4), 956-67.

[233] Collier, J.L. and Grossman, A.R. (1994) "A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria". *Embo J*, **13** (5), 1039-47.

[234] Laudenbach, D.E., Ehrhardt, D., Green, L. and Grossman, A. (1991) "Isolation and characterization of a sulfur-regulated gene encoding a periplasmically localized protein with sequence similarity to rhodanese". *J Bacteriol*, **173** (9), 2751-60.

[235] Setif, P. and Brettel, K. (1993) "Forward electron transfer from phylloquinone A1 to iron-sulfur centers in spinach photosystem I". *Biochemistry*, **32** (31), 7846-54.

[236] van der Est, A., Bock, C., Golbeck, J., Brettel, K., Setif, P. and Stehlik, D. (1994) "Electron transfer from the acceptor A1 to the iron-sulfur centers in photosystem I as studied by transient EPR spectroscopy". *Biochemistry*, **33** (39), 11789-97.

[237] Leibl, W., Toupance, B. and Breton, J. (1995) "Photoelectric characterization of forward electron transfer to iron-sulfur centers in photosystem I". *Biochemistry*, **34** (32), 10237-44.

[238] Xu, D., Liu, X., Zhao, J. and Zhao, J. (2005) "FesM, a membrane iron-sulfur protein, is required for cyclic electron flow around photosystem I and photoheterotrophic growth of the cyanobacterium *Synechococcus* sp. PCC 7002". *Plant Physiol*, **138** (3), 1586-95.

[239] Singh, A.K. and Sherman, L.A. (2006) "Iron-independent dynamics of IsiA production during the transition to stationary phase in the cyanobacterium *Synechocystis* sp. PCC 6803". *FEMS Microbiol Lett*, **256** (1), 159-64.

[240] Geiss, U., Vinnemeier, J., Kunert, A., Lindner, I., Gemmer, B., Lorenz, M., Hagemann, M. and Schoor, A. (2001) "Detection of the isiA gene across cyanobacterial strains: potential for probing iron deficiency". *Appl Environ Microbiol*, **67** (11), 5247-53.

[241] Liu, X., Zhao, J. and Wu, Q. (2006) "Biogenesis of chlorophyll-binding proteins under iron stress in *Synechocystis* sp. PCC 6803". *Biochemistry (Mosc)*, **71 Suppl 1**, S101-4.

[242] Odom, W.R., Hodges, R., Chitnis, P.R. and Guikema, J.A. (1993) "Characterization of *Synechocystis* sp. PCC 6803 in iron-supplied and iron-deficient media". *Plant Mol Biol*, **23** (6), 1255-64.

[243] Kutzki, C., Masepohl, B. and Bohme, H. (1998) "The isiB gene encoding flavodoxin is not essential for photoautotrophic iron limited growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803". *FEMS Microbiol Lett*, **160** (2), 231-5.

[256] Yagil, E. (1975) "Derepression of polyphosphatase in *Escherichia coli* by starvation for inorganic phosphate". *FEBS Lett*, **55** (1), 124-7.

[257] Van Dien, S.J. and Keasling, J.D. (1998) "A dynamic model of the *Escherichia coli* phosphate-starvation response". *J Theor Biol*, **190** (1), 37-49.

[258] Bhaya, D., Vaulot, D., Amin, P., Takahashi, A.W. and Grossman, A.R. (2000) "Isolation of regulated genes of the cyanobacterium *Synechocystis* sp. strain PCC 6803 by differential display". *J Bacteriol*, **182** (20), 5692-9.

[259] Sazuka, T. and Ohara, O. (1997) "Towards a proteome project of cyanobacterium *Synechocystis* sp. strain PCC6803: linking 130 protein spots with their respective genes". *Electrophoresis*, **18** (8), 1252-8.

[260] Gorg, A., Postel, W. and Gunther, S. (1988) "The current state of two-dimensional electrophoresis with immobilized pH gradients". *Electrophoresis*, **9** (9), 531-46.

[261] Pandey, A. and Mann, M. (2000) "Proteomics to study genes and genomes". *Nature*, **405** (6788), 837-46.

[262] Issaq, H.J., Conrads, T.P., Janini, G.M. and Veenstra, T.D. (2002) "Methods for fractionation, separation and profiling of proteins and peptides". *Electrophoresis*, **23** (17), 3048-61.

[263] Packer, N.H., Pawlak, A., Kett, W.C., Gooley, A.A., Redmond, J.W. and Williams, K.L. (1997) "Proteome analysis of glycoforms: a review of strategies for the microcharacterisation of glycoproteins separated by two-dimensional polyacrylamide gel electrophoresis". *Electrophoresis*, **18** (3-4), 452-60.

[264] Robertson, E.S. and Nicholson, A.W. (1992) "Phosphorylation of *Escherichia coli* translation initiation-factors by the bacteriophage-T7 protein-kinase". *Biochemistry*, **31**, 4822- 4827.

[265] Robertson, E.S., Aggison, L.A. and Nicholson, A.W. (1994) "Phosphorylation of elongation factor G and ribosomal protein S6 in bacteriophage T7-infected *Escherichia coli*". *Mol Microbiol.*, **11**, 1045- 1057.

[266] Ohba, M., Koiwai, O., Tanada, S. and Hayashi, H. (1979) "In vivo methylation of elongation factor Tu of *Escherichia coli*". *J Biochem (Tokyo)*, **86** (5), 1233-8.

[267] Peng, J. and Gygi, S.P. (2001) "Proteomics: the move to mixtures". *J. Mass Spectrom.*, **36** (10), 1083-91.

[268] Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. and Aebersold, R. (2000) "Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology". *Proc. Natl. Acad. Sci. U S A*, **97** (17), 9390-5.

[269] Sazuka, T., Yamaguchi, M. and Ohara, O. (1999) "Cyano2Dbase updated: linkage of 234 protein spots to corresponding genes through N-terminal microsequencing". *Electrophoresis*, **20** (11), 2160-71.

[282] Shen, Y., Tolic, N., Masselon, C., Pasa-Tolic, L., Camp, D.G., 2nd, Lipton, M.S., Anderson, G.A. and Smith, R.D. (2004) "Nanoscale proteomics". *Anal. Bioanal. Chem.*, **378** (4), 1037-45.

[283] Zhu, W., Reich, C.I., Olsen, G.J., Giometti, C.S. and Yates, J.R., 3rd (2004) "Shotgun proteomics of *Methanococcus jannaschii* and insights into methanogenesis". *J. Proteome Res.*, **3** (3), 538-48.

[284] Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. and Gygi, S.P. (2003) "Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome". *J. Proteome Res.*, **2** (1), 43-50.

[285] Vollmer, M., Horth, P. and Nagele, E. (2004) "Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples". *Anal. Chem.*, **76** (17), 5180-5.

[286] Cargile, B.J., Bundy, J.L., Freeman, T.W. and Stephenson, J.L., Jr. (2004) "Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification". *J. Proteome Res.*, **3** (1), 112-9.

[287] Chen, J., Balgley, B.M., DeVoe, D.L. and Lee, C.S. (2003) "Capillary isoelectric focusing-based multidimensional concentration/separation platform for proteome analysis". *Anal. Chem.*, **75** (13), 3145-52.

[288] Xiao, Z., Conrads, T.P., Lucas, D.A., Janini, G.M., Schaefer, C.F., Buetow, K.H., Issaq, H.J. and Veenstra, T.D. (2004) "Direct ampholyte-free liquid-phase isoelectric peptide focusing: application to the human serum proteome". *Electrophoresis*, **25** (1), 128-33.

[289] Figeys, D., Ducret, A., Yates, J.R., 3rd and Aebersold, R. (1996) "Protein identification by solid phase microextraction-capillary zone electrophoresis-microelectrospray-tandem mass spectrometry". *Nat. Biotechnol.*, **14** (11), 1579-83.

[290] Tong, W., Link, A., Eng, J.K. and Yates, J.R., 3rd (1999) "Identification of proteins in complexes by solid-phase microextraction/multistep elution/capillary electrophoresis/tandem mass spectrometry". *Anal. Chem.*, **71** (13), 2270-8.

[291] Righetti, P.G., Castagna, A., Herbert, B., Reymond, F. and Rossier, J.S. (2003) "Prefractionation techniques in proteome analysis". *Proteomics*, **3** (8), 1397-407.

[292] Lescuyer, P., Hochstrasser, D.F. and Sanchez, J.C. (2004) "Comprehensive proteome analysis by chromatographic protein prefractionation". *Electrophoresis*, **25** (7-8), 1125-35.

[293] Opiteck, G.J. and Jorgenson, J.W. (1997) "Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides". *Anal. Chem.*, **69** (13), 2283-91.

[294] Opiteck, G.J., Ramirez, S.M., Jorgenson, J.W. and Moseley, M.A., 3rd (1998) "Comprehensive two-dimensional high-performance liquid chromatography for the

[307] Lee, C.L., Hsiao, H.H., Lin, C.W., Wu, S.P., Huang, S.Y., Wu, C.Y., Wang, A.H. and Khoo, K.H. (2003) "Strategic shotgun proteomics approach for efficient construction of an expression map of targeted protein families in hepatoma cell lines". *Proteomics*, **3** (12), 2472-86.

[308] Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villen, J., Li, J., Cohn, M.A., Cantley, L.C. and Gygi, S.P. (2004) "Large-scale characterization of HeLa cell nuclear phosphoproteins". *Proc. Natl. Acad. Sci. U S A*, **101** (33), 12130-5.

[309] Chan, K.C., Lucas, D.A., Hise, D., Schaefer, C.F., Xiao, Z., Janini, G.M., Buetow, K.H., Issaq, H.J., Veenstra, T.D. and Conrads, T.P. (2004) "Analysis of the Human Serum Proteome". *Clinical Proteomics*, **1** (2), 101-226.

[310] Ostrowski, L.E., Blackburn, K., Radde, K.M., Moyer, M.B., Schlatzer, D.M., Moseley, A. and Boucher, R.C. (2002) "A proteomic analysis of human cilia: identification of novel components". *Mol. Cell Proteomics*, **1** (6), 451-65.

[311] Putz, S., Reinders, J., Reinders, Y. and Sickmann, A. (2005) "Mass spectrometry-based peptide quantification: applications and limitations". *Expert Review of Proteomics*, **2** (3), 381-392.

[312] Thiede, B., Kretschmer, A. and Rudel, T. (2006) "Quantitative proteome analysis of CD95 (Fas/Apo-1)-induced apoptosis by stable isotope labeling with amino acids in cell culture, 2-DE and MALDI-MS". *Proteomics*, **6** (2), 614-22.

[313] Amanchy, R., Kalume, D.E., Iwahori, A., Zhong, J. and Pandey, A. (2005) "Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC)". *J Proteome Res*, **4** (5), 1661-71.

[314] Gruhler, A., Schulze, W.X., Matthiesen, R., Mann, M. and Jensen, O.N. (2005) "Stable isotope labeling of Arabidopsis thaliana cells and quantitative proteomics by mass spectrometry". *Mol Cell Proteomics*, **4** (11), 1697-709.

[315] Everley, P.A., Krijgsveld, J., Zetter, B.R. and Gygi, S.P. (2004) "Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research". *Mol Cell Proteomics*, **3** (7), 729-35.

[316] Ramus, C., Gonzalez de Peredo, A., Dahout, C., Gallagher, M. and Garin, J. (2006) "An optimized strategy for ICAT quantification of membrane proteins". *Mol Cell Proteomics*, **5** (1), 68-78.

[317] Booy, A.T., Haddow, J.D., Ohlund, L.B., Hardie, D.B. and Olafson, R.W. (2005) "Application of isotope coded affinity tag (ICAT) analysis for the identification of differentially expressed proteins following infection of atlantic salmon (*Salmo salar*) with infectious hematopoietic necrosis virus (IHNV) or *Renibacterium salmoninarum* (BKD)". *J Proteome Res*, **4** (2), 325-34.

[318] Yan, W., Lee, H., Deutsch, E.W., Lazaro, C.A., Tang, W., Chen, E., Fausto, N., Katze, M.G. and Aebersold, R. (2004) "A dataset of human liver proteins identified by protein

[332] Cong, Y.S., Fan, E. and Wang, E. (2006) "Simultaneous proteomic profiling of four different growth states of human fibroblasts, using amine-reactive isobaric tagging reagents and tandem mass spectrometry". *Mech Ageing Dev*, **127** (4), 332-43.

[333] Hirsch, J., Hansen, K.C., Choi, S., Noh, J., Hirose, R., Roberts, J.P., Matthay, M.A., Burlingame, A.L., Maher, J.J. and Niemann, C.U. (2006) "Warm ischemia-induced alterations in oxidative and inflammatory proteins in hepatic Kupffer cells in rats". *Mol Cell Proteomics*, **5** (6), 979-86.

[334] Unwin, R.D., Smith, D.L., Blinco, D., Wilson, C.L., Miller, C.J., Evans, C.A., Jaworska, E., Baldwin, S.A., Barnes, K., Pierce, A., Spooncer, E. and Whetton, A.D. (2006) "Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells". *Blood*, **107** (12), 4687-94.

[335] Keshamouni, V.G., Michailidis, G., Grasso, C.S., Anthwal, S., Strahler, J.R., Walker, A., Arenberg, D.A., Reddy, R.C., Akulapalli, S., Thannickal, V.J., Standiford, T.J., Andrews, P.C. and Omenn, G.S. (2006) "Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype". *J Proteome Res*, **5** (5), 1143-54.

[336] Hu, J., Qian, J., Borisov, O., Pan, S., Li, Y., Liu, T., Deng, L., Wannemacher, K., Kurnellas, M., Patterson, C., Elkabes, S. and Li, H. (2006) "Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags". *Proteomics*.

[337] Salim, K., Kehoe, L., Minkoff, M.S., Bilsland, J.G., Munoz-Sanjuan, I. and Guest, P.C. (2006) "Identification of differentiating neural progenitor cell markers using shotgun isobaric tagging mass spectrometry". *Stem Cells Dev*, **15** (3), 461-70.

[338] Abdi, F., Quinn, J.F., Jankovic, J., McIntosh, M., Leverenz, J.B., Peskind, E., Nixon, R., Nutt, J., Chung, K., Zabetian, C., Samii, A., Lin, M., Hattan, S., Pan, C., Wang, Y., Jin, J., Zhu, D., Li, G.J., Liu, Y., Waichunas, D., Montine, T.J. and Zhang, J. (2006) "Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders". *J Alzheimers Dis*, **9** (3), 293-348.

[339] Seshi, B. (2006) "An integrated approach to mapping the proteome of the human bone marrow stromal cell". *Proteomics*, **6** (19), 5169-82.

[340] Unwin, R.D., Pierce, A., Watson, R.B., Sternberg, D.W. and Whetton, A.D. (2005) "Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells". *Molecular & Cellular Proteomics*, **4** (7), 924-935.

[341] Islinger, M., Li, K.W., Loos, M., Luers, G. and Volkl, A. (2005) "Proteomic, quantitative analysis of liver peroxisomes after bezafibrate treatment using iTRAQ labeled probes". *European Journal of Cell Biology*, **84**, 121-121.

[353] Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A. and White, F.M. (2005) "Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules". *Mol Cell Proteomics*, **4** (9), 1240-50.

[354] Sachon, E., Mohammed, S., Bache, N. and Jensen, O.N. (2006) "Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: application to proteins isolated by gel electrophoresis". *Rapid Commun Mass Spectrom*, **20** (7), 1127-34.

[355] Shadforth, I.P., Dunkley, T.P., Lilley, K.S. and Bessant, C. (2005) "i-Tracker: for quantitative proteomics using iTRAQ". *BMC Genomics*, **6**, 145.

[356] Lin, W.T., Hung, W.N., Yian, Y.H., Wu, K.P., Han, C.L., Chen, Y.R., Chen, Y.J., Sung, T.Y. and Hsu, W.L. (2006) "Multi-Q: a fully automated tool for multiplexed protein quantitation". *J Proteome Res*, **5** (9), 2328-38.

[357] Wolff, S., Otto, A., Albrecht, D., Zeng, J.S., Buttner, K., Gluckmann, M., Hecker, M. and Becher, R. (2006) "Gel-free and gel-based proteomics in *Bacillus subtilis* - A comparative study". *Mol Cell Proteomics*, **5** (7), 1183-1192.

[358] Wu, W.W., Wang, G., Baek, S.J. and Shen, R.F. (2006) "Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF". *J Proteome Res*, **5** (3), 651-8.

[359] Choe, L.H., Aggarwal, K., Franck, Z. and Lee, K.H. (2005) "A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells". *Electrophoresis*, **26** (12), 2437-49.

[360] Miyake, M., Takase, K., Narato, M., Khatipov, E., Schnackenberg, J., Shirai, M., Kurane, R. and Asada, Y. (2000) "Polyhydroxybutyrate production from carbon dioxide by cyanobacteria". *Appl Biochem Biotechnol*, **84-86**, 991-1002.

[361] Taroncher-Oldenburg, G. and Stephanopoulos, G. (2000) "Targeted, PCR-based gene disruption in cyanobacteria: inactivation of the polyhydroxyalkanoic acid synthase genes in *Synechocystis* sp. PCC6803". *Appl Microbiol Biotechnol*, **54** (5), 677-80.

[362] Li, H. and Sherman, L.A. (2002) "Characterization of *Synechocystis* sp. strain PCC 6803 and deltanbl mutants under nitrogen-deficient conditions". *Arch Microbiol*, **178** (4), 256-66.

[363] Govorukhina, N.I., Keizer-Gunnink, A., van der Zee, A.G., de Jong, S., de Bruijn, H.W. and Bischoff, R. (2003) "Sample preparation of human serum for the analysis of tumor markers. Comparison of different approaches for albumin and gamma-globulin depletion". *J. Chromatogr. A*, **1009** (1-2), 171-8.

[364] Pieper, R., Gatlin, C.L., Makusky, A.J., Russo, P.S., Schatz, C.R., Miller, S.S., Su, Q., McGrath, A.M., Estock, M.A., Parmar, P.P., Zhao, M., Huang, S.T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N.L., Taylor, J. and Steiner, S. (2003) "The human serum

[376] Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes". *J Mol Biol*, **305** (3), 567-80.

[377] Spahr, C.S., Susin, S.A., Bures, E.J., Robinson, J.H., Davis, M.T., McGinley, M.D., Kroemer, G. and Patterson, S.D. (2000) "Simplification of complex peptide mixtures for proteomic analysis: reversible biotinylation of cysteinyl peptides". *Electrophoresis*, **21** (9), 1635-50.

[378] Schaefer, H., Chervet, J.P., Bunse, C., Joppich, C., Meyer, H.E. and Marcus, K. (2004) "A peptide preconcentration approach for nano-high-performance liquid chromatography to diminish memory effects". *Proteomics*, **4** (9), 2541-4.

[379] Koc, E.C., Burkhart, W., Blackburn, K., Moyer, M.B., Schlatzer, D.M., Moseley, A. and Spremulli, L.L. (2001) "The large subunit of the mammalian mitochondrial ribosome. Analysis of the complement of ribosomal proteins present". *J Biol Chem*, **276** (47), 43958-69.

[380] Cutillas, P.R. (2005) "Principles of Nanoflow Liquid Chromatography and Applications to Proteomics". *Current Nanoscience*, **1**, 65-71.

[381] Kreil, D.P., Karp, N.A. and Lilley, K.S. (2004) "DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results". *Bioinformatics*, **20** (13), 2026-34.

[382] Karp, N.A., Griffin, J.L. and Lilley, K.S. (2005) "Application of partial least squares discriminant analysis to two-dimensional difference gel studies in expression proteomics". *Proteomics*, **5** (1), 81-90.

[383] Fievet, J., Dillmann, C., Lagniel, G., Davanture, M., Negroni, L., Labarre, J. and de Vienne, D. (2004) "Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis". *Proteomics*, **4** (7), 1939-1949.

[384] Karp, N.A., Spencer, M., Lindsay, H., O'Dell, K. and Lilley, K.S. (2005) "Impact of replicate types on proteomic expression analysis". *J Proteome Res*, **4** (5), 1867-71.

[385] Bafrncová, P., Mogroviová, D., Sláviková, I. and Pátková, J.D.Z. (1999) "Improvement of very high gravity ethanol fermentation by media supplementation using *Saccharomyces cerevisiae*". *Biotechnology Letters*, **21** (4), 337-341.

[386] Snijders, A.P., de Vos, M.G. and Wright, P.C. (2005) "Novel approach for peptide quantitation and sequencing based on 15N and 13C metabolic labeling". *J Proteome Res*, **4** (2), 578-85.

[387] Atlas, R.M., (1997). *Handbook of Microbiological Media*, New York: CRC Press.

[388] Trabalzini, L., Paffetti, A., Scaloni, A., Talamo, F., Ferro, E., Coratza, G., Bovalini, L., Lusini, P., Martelli, P. and Santucci, A. (2003) "Proteomic response to physiological fermentation stresses in a wild-type wine strain of *Saccharomyces cerevisiae*". *Biochem J*, **370** (Pt 1), 35-46.

[400] Peng, X., Wood, C.L., Blalock, E.M., Chen, K.C., Landfield, P.W. and Stromberg, A.J. (2003) "Statistical implications of pooling RNA samples for microarray experiments". *BMC Bioinformatics*, **4**, 26.

[401] Kendziorski, C.M., Zhang, Y., Lan, H. and Attie, A.D. (2003) "The efficiency of pooling mRNA in microarray experiments". *Biostatistics*, **4** (3), 465-77.

[402] Shih, J.H., Michalowska, A.M., Dobbin, K., Ye, Y., Qiu, T.H. and Green, J.E. (2004) "Effects of pooling mRNA in microarray class comparisons". *Bioinformatics*, **20** (18), 3318-25.

[403] Zhang, S.D. and Gant, T.W. (2005) "Effect of pooling samples on the efficiency of comparative studies using microarrays". *Bioinformatics*, **21** (24), 4378-83.

[404] Kendziorski, C., Irizarry, R.A., Chen, K.S., Haag, J.D. and Gould, M.N. (2005) "On the utility of pooling biological samples in microarray experiments". *Proc Natl Acad Sci U S A*, **102** (12), 4252-7.

[405] Neubauer, H., Clare, S.E., Kurek, R., Fehm, T., Wallwiener, D., Sotlar, K., Nordheim, A., Wozny, W., Schwall, G.P., Poznanovic, S., Sastri, C., Hunzinger, C., Stegmann, W., Schrattenholz, A. and Cahill, M.A. (2006) "Breast cancer proteomics by laser capture microdissection, sample pooling, 54-cm IPG IEF, and differential iodine radioisotope detection". *Electrophoresis*, **27** (9), 1840-52.

[406] Lusa, L., Cappelletti, V., Gariboldi, M., Ferrario, C., De Cecco, L., Reid, J.F., Toffanin, S., Gallus, G., McShane, L.M., Daidone, M.G. and Pierotti, M.A. (2006) "Questioning the utility of pooling samples in microarray experiments with cell lines". *Int J Biol Markers*, **21** (2), 67-73.

[407] Jolly, R.A., Goldstein, K.M., Wei, T., Gao, H., Chen, P., Huang, S., Colet, J.M., Ryan, T.P., Thomas, C.E. and Estrem, S.T. (2005) "Pooling samples within microarray studies: a comparative analysis of rat liver transcription response to prototypical toxicants". *Physiol Genomics*, **22** (3), 346-55.

[408] Zhang, X., Guo, Y., Song, Y., Sun, W., Yu, C., Zhao, X., Wang, H., Jiang, H., Li, Y., Qian, X., Jiang, Y. and He, F. (2006) "Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis". *Proteomics*, **6** (19), 5260-8.

[409] Elias, J.E., Haas, W., Faherty, B.K. and Gygi, S.P. (2005) "Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations". *Nature Methods*, **2** (9), 667-675.

[410] Xiong, J. and Bauer, C.E. (2002) "Complex evolution of photosynthesis". *Annu Rev Plant Biol*, **53**, 503-21.

[411] Hihara, Y., Sonoike, K., Kanehisa, M. and Ikeuchi, M. (2003) "DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803". *J Bacteriol*, **185** (5), 1719-25.

[424] Wieden, H.J., Gromadski, K., Rodnin, D. and Rodnina, M.V. (2002) "Mechanism of elongation factor (EF)-Ts-catalyzed nucleotide exchange in EF-Tu. Contribution of contacts at the guanine base". *J Biol Chem*, **277** (8), 6032-6.

[425] Schmidt, J., Bubunenko, M. and Subramanian, A.R. (1993) "A novel operon organization involving the genes for chorismate synthase (aromatic biosynthesis pathway) and ribosomal GTPase center proteins (L11, L1, L10, L12: rplKAJL) in cyanobacterium *Synechocystis* PCC 6803". *J Biol Chem*, **268** (36), 27447-57.

[426] Gaber, A., Tamoi, M., Takeda, T., Nakano, Y. and Shigeoka, S. (2001) "NADPH-dependent glutathione peroxidase-like proteins (Gpx-1, Gpx-2) reduce unsaturated fatty acid hydroperoxides in *Synechocystis* PCC 6803". *FEBS Lett*, **499** (1-2), 32-6.

[427] Aitken, A. and Rouviere-Yaniv, J. (1979) "Amino and carboxy terminal sequences of the DNA-binding protein HU from the Cyanobacterium *Synechocystis* PCC 6701 (ATCC 27170)". *Biochem Biophys Res Commun*, **91** (2), 461-7.

[428] Elanskaya, I.V., Chesnavichene, E.A., Vernotte, C. and Astier, C. (1998) "Resistance to nitrophenolic herbicides and metronidazole in the cyanobacterium *Synechocystis* sp. PCC 6803 as a result of the inactivation of a nitroreductase-like protein encoded by drgA gene". *FEBS Lett*, **428** (3), 188-92.

[429] Elanskaya, I.V., Timofeev, K.N., Grivennikova, V.G., Kuznetsova, G.V., Davletshina, L.N., Lukashev, E.P. and Yaminsky, F.V. (2004) "Reduction of photosystem I reaction center in DrgA mutant of the cyanobacterium *Synechocystis* sp. PCC 6803 lacking soluble NAD(P)H:quinone oxidoreductase". *Biochemistry (Mosc)*, **69** (4), 445-54.

[430] Matsuo, M., Endo, T. and Asada, K. (1998) "Isolation of a novel NAD(P)H-quinone oxidoreductase from the cyanobacterium *Synechocystis* PCC6803". *Plant Cell Physiol*, **39** (7), 751-5.

[431] Chung, H.J., Kim, Y.A., Kim, Y.J., Choi, Y.K., Hwang, Y.K. and Park, Y.S. (2000) "Purification and characterization of UDP-glucose:tetrahydrobiopterin glucosyltransferase from *Synechococcus* sp. PCC 7942". *Biochim Biophys Acta*, **1524** (2-3), 183-8.

[432] Lee, S.W., Lee, H.W., Chung, H.J., Kim, Y.A., Kim, Y.J., Hahn, Y., Chung, J.H. and Park, Y.S. (1999) "Identification of the genes encoding enzymes involved in the early biosynthetic pathway of pteridines in *Synechocystis* sp. PCC 6803". *FEMS Microbiol Lett*, **176** (1), 169-76.

[433] Johnson, C.H. and Golden, S.S. (1999) "Circadian programs in cyanobacteria: adaptiveness and mechanism". *Annu Rev Microbiol*, **53**, 389-409.

[434] Kondo, T. and Ishiura, M. (2000) "The circadian clock of cyanobacteria". *Bioessays*, **22** (1), 10-5.

[435] Wilsbacher, L.D. and Takahashi, J.S. (1998) "Circadian rhythms: molecular basis of the clock". *Curr Opin Genet Dev*, **8** (5), 595-602.

Synechocystis sp. PCC 6803: expression and assembly in *Escherichia coli*". *Arch Biochem Biophys*, **379** (2), 267-76.

[451] Navarro, F., Chavez, S., Candau, P. and Florencio, F.J. (1995) "Existence of two ferredoxin-glutamate synthases in the cyanobacterium *Synechocystis* sp. PCC 6803. Isolation and insertional inactivation of gltB and gltS genes". *Plant Mol Biol*, **27** (4), 753-67.

[452] Mehra, A., Lee, K.H. and Hatzimanikatis, V. (2003) "Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations". *Biotechnol Bioeng*, **84** (7), 822-33.

[453] Lee, P.S., Shaw, L.B., Choe, L.H., Mehra, A., Hatzimanikatis, V. and Lee, K.H. (2003) "Insights into the relation between mrna and protein expression patterns: II. Experimental observations in *Escherichia coli*". *Biotechnol Bioeng*, **84** (7), 834-41.

[454] Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999) "Correlation between protein and mRNA abundance in yeast". *Mol. Cell. Biol.*, **19** (3), 1720-30.

[455] Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S. and Cohen, S.N. (2002) "Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays". *Proc Natl Acad Sci U S A*, **99** (15), 9697-702.

[456] Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D. and Brown, P.O. (2002) "Precision and functional specificity in mRNA decay". *Proc Natl Acad Sci U S A*, **99** (9), 5860-5.

[457] Wassen, M.J., Venterink, H.O., Lapshina, E.D. and Tanneberger, F. (2005) "Endangered plants persist under phosphorus limitation". *Nature*, **437** (7058), 547-50.

[458] Wu, J., Sunda, W., Boyle, E.A. and Karl, D.M. (2000) "Phosphate depletion in the western North Atlantic Ocean". *Science*, **289** (5480), 759-62.

[459] Han, J.S., Park, J.Y., Lee, Y.S., Thony, B. and Hwang, D.S. (1999) "PhoB-dependent transcriptional activation of the iciA gene during starvation for phosphate in *Escherichia coli*". *Mol Gen Genet*, **262** (3), 448-52.

[460] Wanner, B.L., 2nd. (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Washington, D. C.: ASM Press. p. 1357-1381.

[461] Ternan, N.G., McGrath, J.W., McMullan, G. and Quinn, J.P. (1998) "Review: Organophosphonates: occurrence, synthesis and biodegradation by microorganisms". *World of Journal of Microbiology & Biotechnology*, **14**, 635-647.

[462] Ishige, T., Krause, M., Bott, M., Wendisch, V.F. and Sahm, H. (2003) "The phosphate starvation stimulon of *Corynebacterium glutamicum* determined by DNA microarray analyses". *J Bacteriol*, **185** (15), 4519-29.

[463] Hou, C.I., Gronlund, A.F. and Campbell, J.J. (1966) "Influence of phosphate starvation on cultures of *Pseudomonas aeruginosa*". *J Bacteriol*, **92** (4), 851-5.

[477] Perez-Perez, M.E., Florencio, F.J. and Lindahl, M. (2006) "Selecting thioredoxins for disulphide proteomics: Target proteomes of three thioredoxins from the cyanobacterium *Synechocystis* sp. PCC 6803". *Proteomics*, **6 Suppl 1**, S186-95.

[478] Clarke, A.K., Schelin, J. and Porankiewicz, J. (1998) "Inactivation of the clpP1 gene for the proteolytic subunit of the ATP-dependent Clp protease in the cyanobacterium *Synechococcus* limits growth and light acclimation". *Plant Mol Biol*, **37** (5), 791-801.

[479] Parsell, D.A. and Lindquist, S. (1993) "The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins". *Annu Rev Genet*, **27**, 437-96.

[480] Mohan, S., Dow, C. and Coles, J.A., 1. (1992). *Prokaryotic Structure and Function: A New Perspective,* Cambridge University Press. 311-339.

[481] Scanlan, D.J., Mann, N.H. and Carr, N.G. (1993) "The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate-binding protein of *Escherichia coli*". *Mol Microbiol*, **10** (1), 181-91.

[482] Sola-Landa, A., Rodriguez-Garcia, A., Franco-Dominguez, E. and Martin, J.F. (2005) "Binding of PhoP to promoters of phosphate-regulated genes in *Streptomyces coelicolor*: identification of PHO boxes". *Mol Microbiol*, **56** (5), 1373-85.

[483] Liu, W., Qi, Y. and Hulett, F.M. (1998) "Sites internal to the coding regions of phoA and pstS bind PhoP and are required for full promoter activity". *Mol Microbiol*, **28** (1), 119-30.

[484] Bertani, L.E., Huang, J.S., Weir, B.A. and Kirschvink, J.L. (1997) "Evidence for two types of subunits in the bacterioferritin of *Magnetospirillum magnetotacticum*". *Gene*, **201** (1-2), 31-6.

[485] Tottey, S., Rondet, S.A., Borrelly, G.P., Robinson, P.J., Rich, P.R. and Robinson, N.J. (2002) "A copper metallochaperone for photosynthesis and respiration reveals metal-specific targets, interaction with an importer, and alternative sites for copper acquisition". *J Biol Chem*, **277** (7), 5490-7.

[486] Cook, A.M., (1988). *Combined carbon and phosphorus or carbon and sulphur substrates,* **In: Mixed and Multiple Feedstocks**. by Hamer, G., Egli, T. and Snozzi, M.: EFB Konstanz.

[487] Beuf, L., Bedu, S., Durand, M.C. and Joset, F. (1994) "A protein involved in co-ordinated regulation of inorganic carbon and glucose metabolism in the facultative photoautotrophic cyanobacterium *Synechocystis* PCC6803". *Plant Mol Biol*, **25** (5), 855-64.

[488] Shi, L., Bischoff, K.M. and Kennelly, P.J. (1999) "The icfG gene cluster of *Synechocystis* sp. strain PCC 6803 encodes an Rsb/Spo-like protein kinase, protein phosphatase, and two phosphoproteins". *J Bacteriol*, **181** (16), 4761-7.

Appendix B

Due to large amount of supporting information available for the study conducted in Chapter 3, all these information will be made available in a soft-copy (CD format), attached to the back of this thesis, rather than as hard-copy.

File name : Appendix B
File type : Microsoft Excel
Description : The total proteins lists identified in each workflow

The lists of proteins found in each protein-peptide prefractionation workflows, as well as the use of difference search databases are included in the CD. Following are the brief description of each sub-document within Appendix B.

	Excel sub-document	Workflows used	Database used *
(1)	SCX NR	SCX-only	NR
(2)	SCX SYN	SCX-only	SYN
(3)	IEF NR	IEF-only	NR
(4)	IEF SYN	IEF-only	SYN
(5)	1D-SCX NR	1D-SCX	NR
(6)	1D-SCX SYN	1D-SCX	SYN
(7)	WAX-SCX NR	WAX-SCX	NR
(8)	WAX-SCX SYN	WAX-SCX	SYN
(9)	IEF-IEF NR	IEF-IEF	NR
(10)	IEF-IEF SYN	IEF-IEF	SYN
(11)	Whole NR	Whole-cell lysate	NR
(12)	Whole SYN	Whole-cell lysate	SYN
(13)	SCX 40 NR	SCX-only (40 fractions)	NR
(14)	SCX 40 SYN	SCX-only (40 fractions)	SYN
(15)	Total NR	All workflows	NR
(16)	Total SYN	All workflows	SYN

Note: * NR refers to non-redundant database
SYN refers to *Synechocystis* sp. PCC 6803 single genome database

Appendix D

Due to large amount of supporting information available for the study conducted in Chapter 4, all these information will be made available in a soft-copy (CD format), attached to the back of this thesis, rather than as hard-copy.

File name : Appendix D
File type : Microsoft Excel
Description : The proteins lists of each replicate analyses

The lists of proteins found in each technical, experimental (iTRAQ) and biological replicates are included in the CD. Following are the brief description of each sub-document within Appendix D.

	Excel sub-document	Replicate types	Experiments	Label
(1)	Tech 1	Technical	1	115:114
(2)	Tech 2	Technical	2	115:114
(3)	Tech 3	Technical	4	116:117
(4)	Tech 4	Technical	6	115:114
(5)	Tech 5	Technical	6	116:117
(6)	Biol 1	Biological	4	116:115
(7)	Biol 2	Biological	4	117:115
(8)	Biol 3	Biological	5	116:115

	Excel sub-document	Replicate types	Replicate 1		Replicate 2	
			Exp	Label	Exp	Label
(9)	iTRAQ 1	Experimental	3	117:114	4	115:114
(10)	iTRAQ 2	Experimental	7	116:114	8	115:114
(11)	iTRAQ 3	Experimental	9	116:114	10	115:114

	Peptide sequence	Ratio 115:114	% Error	w	x × w
1)	IGQNPEPVTIJ	0.3638	5.81	0.1721	-0.0756
2)	IGQNPEPVTIJ	0.3766	4.73	0.2114	-0.0897
3)	IGQNPEPVTIJ	0.3863	9.81	0.1019	-0.0421
4)	IGQNPEPVTIJ	0.4020	7.13	0.1403	-0.0555
5)	IGQNPEPVTIJ	0.4523	7.77	0.1287	-0.0443
6)	IGQNPEPVTIJ	0.6016	8.85	0.1130	-0.0249
7)	VYPSGEVQYLHPADGVFPEJ	0.7468	13.25	0.0755	-0.0096
8)	VYPSGEVQYLHPADGVFPEJ	0.7708	8.11	0.1233	-0.0139
	Σ			**1.0661**	**-0.3557**

Weighted mean of log ratio, $\mathbf{Y_{log}}$ = -0.3557 / 1.0661

= -0.3336

From here, we could antilog$_{10}$ the $\mathbf{Y_{log}}$ to get the linear space quantitation ratio for protein PsaD.

∴Linear space ratio, **Y** = Antilog$_{10}$ (-0.3336)

= 0.4639

However, the **Y** value of 0.4639 can sometimes cause confusion since protein ratio in linear space does not have even distribution below and above 1.00. Therefore it is recommended to express any **Y** values below 1.00 as:

$$-\frac{1}{Y} \quad \text{..... (Equation E2)}$$

PsaD protein ratio thus is -1/0.4639 = -2.1556,

i.e. a 2.16-fold down-regulation was calculated.

Then we need to calculate the effective base, **b**, using the following formula:

$$b = \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}, \text{ where } w_i = 1/\%\ \text{Error} \quad \text{..... (Equation E3)}$$

	Peptide sequence	% Error	w	w^2	x
1)	IGQNPEPVTIJ	5.81	0.1721	0.0296	-0.4391
2)	IGQNPEPVTIJ	4.73	0.2114	0.0447	-0.4241
3)	IGQNPEPVTIJ	9.81	0.1019	0.0104	-0.4131
4)	IGQNPEPVTIJ	7.13	0.1403	0.0197	-0.3958
5)	IGQNPEPVTIJ	7.77	0.1287	0.0166	-0.3446
6)	IGQNPEPVTIJ	8.85	0.1130	0.0128	-0.2207
7)	VYPSGEVQYLHPADGVFPEJ	13.25	0.0755	0.0057	-0.1268
8)	VYPSGEVQYLHPADGVFPEJ	8.11	0.1233	0.0152	-0.1131
	Σ		**1.0661**	**0.1546**	
	Standard deviation, SD				**0.1359**

Appendix F

The list of 199 proteins identified in the 24 hr light against 24 hr dark study conducted in Chapter 5. The average relative protein expression is expressed in Log_{10} unit, whereas the error is estimated in standard deviation (SD) based on 2 technical replicates in an iTRAQ experiment.

ORF	Protein Name	MS/MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
Amino Acids Metabolism								
sll1981	acetolactate synthase ilvB	2	0.326	0.089	0.354	0.094	0.340	0.092
sll1931	serine hydroxymethyltransferase glyA	4	0.348	0.070	0.330	0.097	0.339	0.083
slr1992	glutathione peroxidase-like NADPH peroxidase gpx2	2	0.185	0.088	0.345	0.011	0.265	0.049
sll0080	N-acetyl-gamma-glutamyl-phosphate reductase argC	6	0.183	0.071	0.222	0.056	0.202	0.064
slr0543	tryptophan synthase beta subunit trpB	2	0.181	0.006	0.166	0.040	0.173	0.023
slr0652	phosphorybosilformimino-5-amino- phosphorybosil-4-imidazolecarboxamideisomerase hisA	3	0.156	0.213	0.177	0.198	0.167	0.205
slr1756	glutamate--ammonia ligase glnA	40	0.151	0.076	0.164	0.080	0.158	0.078
sll1908	D-3-phosphoglycerate dehydrogenase serA	10	0.114	0.040	0.142	0.049	0.128	0.044
sll1750	urease alpha subunit ureC	4	0.113	0.053	0.118	0.090	0.115	0.072
slr0288	glutamate--ammonia ligase glnN	7	0.110	0.031	0.105	0.054	0.107	0.042
sll2001	leucine aminopeptidase	6	0.088	0.130	0.101	0.137	0.095	0.133
slr0585	argininosuccinate synthetase argG	2	0.012	0.049	0.056	0.034	0.034	0.042
sll1363	ketol-acid reductoisomerase ilvC	11	0.002	0.081	0.030	0.087	0.016	0.084
sll1502	NADH-dependent glutamate synthase large subunit gltB	3	-0.015	0.010	0.032	0.022	0.009	0.016
slr0452	dihydroxyacid dehydratase ilvD	9	-0.035	0.076	0.011	0.079	-0.012	0.078
slr0370	succinate-semialdehyde dehydrogenase (NADP+)	4	-0.047	0.101	0.022	0.104	-0.013	0.102
sll1172	threonine synthase thrC	2	-0.082	0.026	-0.060	0.023	-0.071	0.024
slr0032	probable branched-chain amino acid aminotransferase	3	-0.144	0.011	-0.029	0.021	-0.087	0.016
slr0963	ferredoxin-sulfite reductase sir	3	-0.115	0.038	-0.091	0.029	-0.103	0.033
Carbohydrate Metabolism								
slr0809	dTDP-glucose 4,6-dehydratase rfbB	2	0.128	0.078	0.272	0.051	0.200	0.065
sll1212	GDP-mannose 4,6-dehydratase	3	0.126	0.190	0.165	0.217	0.146	0.204
sll1213	GDP-fucose synthetase	5	0.061	0.054	0.070	0.054	0.065	0.054
Carbon Fixation								
sll1525	phosphoribulokinase prk	28	0.451	0.041	0.484	0.040	0.468	0.041

ORF	Protein Name	MS/MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
slr0161	twitching motility protein PilT pilT1	2	0.131	0.298	0.186	0.293	0.158	0.296
slr2075	10kD chaperonin groES	19	0.194	0.040	0.215	0.040	0.205	0.040
slr0623	thioredoxin trxA	24	0.151	0.052	0.171	0.055	0.161	0.054
sll0408	peptidyl-prolyl cis-trans isomerase	7	0.111	0.022	0.153	0.028	0.132	0.025
sll0057	heat shock protein GrpE grpE	8	0.117	0.061	0.119	0.065	0.118	0.063
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	19	0.090	0.057	0.128	0.054	0.109	0.056
slr1641	ClpB protein clpB1	6	0.050	0.034	0.061	0.026	0.055	0.030
slr0063	pilus biogenesis protein homologous to general secretion pathway protein E pilB1	2	0.032	0.023	0.047	0.032	0.039	0.028
sll1694	pilin polypeptide PilA1 pilA1	4	0.016	0.021	0.023	0.032	0.019	0.026
sll0020	ATP-dependent Clp protease ATPase subunit	13	-0.013	0.044	-0.011	0.049	-0.012	0.047
slr1251	peptidyl-prolyl cis-trans isomerase	10	-0.055	0.016	-0.026	0.018	-0.041	0.017
sll1621	AhpC/TSA family protein	13	-0.131	0.014	-0.156	0.022	-0.143	0.018
Glycogen Metabolism								
sll0945	glycogen synthase glgA	2	0.015	0.040	0.104	0.020	0.060	0.030
slr1176	glucose-1-phosphate adenylyltransferase	29	-0.020	0.030	0.022	0.029	0.001	0.029
sll0158	1,4-alpha-glucan branching enzyme glgB	8	-0.015	0.122	0.004	0.119	-0.006	0.120
sll1356	glycogen phosphorylase	10	-0.090	0.034	-0.070	0.041	-0.080	0.038
sll1393	glycogen (starch) synthase glgA	4	0.086	0.058	0.147	0.050	0.117	0.054
sll0726	phosphoglucomutase	2	-0.173	0.040	-0.150	0.014	-0.162	0.027
Glycolysis								
slr0752	enolase	3	0.608	0.266	0.726	0.209	0.667	0.237
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	24	0.310	0.080	0.336	0.079	0.323	0.079
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	34	0.277	0.058	0.321	0.054	0.299	0.056
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	32	0.204	0.037	0.226	0.037	0.215	0.037
slr0394	phosphoglycerate kinase pgk	24	0.118	0.075	0.139	0.079	0.128	0.077
slr1349	glucose-6-phosphate isomerase	7	0.087	0.092	0.120	0.094	0.103	0.093
slr0301	phosphoenolpyruvate synthase	4	0.052	0.078	0.056	0.062	0.054	0.070
sll0920	phosphoenolpyruvate carboxylase ppc	3	-0.026	0.011	0.029	0.023	0.001	0.017
Hypothetical Proteins								
slr0001	(flavin reductase-like)	4	0.523	0.135	0.582	0.151	0.553	0.143
sll0359	(transcriptional regulator)	14	0.490	0.079	0.521	0.080	0.506	0.080

ORF	Protein Name	MS/MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
slr1722	inosine-5'-monophosphate dehydrogenase	5	0.029	0.065	0.053	0.074	0.041	0.070
sll0368	uracil phosphoribosyltransferase	3	-0.090	0.032	-0.063	0.028	-0.076	0.030
Other								
sll1784	periplasmic protein, function unknown	2	0.453	0.016	0.436	0.012	0.444	0.014
ssl2296	pterin-4a-carbinolamine dehydratase	4	0.305	0.082	0.326	0.073	0.316	0.077
sll1305	probable hydrolase	2	0.271	0.149	0.269	0.191	0.270	0.170
slr1719	DrgA protein homolog	4	0.157	0.059	0.194	0.042	0.175	0.050
sll0617	plasma membrane protein essential for thylakoid formation vipp1	2	0.156	0.042	0.176	0.012	0.166	0.027
sll1559	soluble hydrogenase 42 kD subunit	15	0.146	0.037	0.178	0.038	0.162	0.037
sll1130	unknown protein	9	0.087	0.046	0.089	0.051	0.088	0.049
sll1830	unknown protein	3	0.025	0.045	0.052	0.038	0.039	0.041
slr1963	water-soluble carotenoid protein	31	0.008	0.046	0.020	0.047	0.014	0.047
sll0872	unknown protein	9	0.020	0.075	0.000	0.057	0.010	0.066
sll0588	unknown protein	4	-0.010	0.010	0.017	0.010	0.004	0.010
sll0576	putative sugar-nucleotide epimerase/dehydratease	23	-0.052	0.018	-0.013	0.017	-0.033	0.017
sll1873	unknown protein	13	-0.084	0.025	-0.034	0.021	-0.059	0.023
slr1855	unknown protein	5	-0.070	0.018	-0.082	0.035	-0.076	0.027
sll1536	molybdopterin biosynthesis MoeB protein moeB	8	0.129	0.036	0.141	0.028	0.135	0.032
Pentose Phosphate Pathway								
slr1734	glucose 6-phosphate dehydrogenase assembly protein opcA	3	0.225	0.092	0.316	0.127	0.271	0.109
slr0194	ribose 5-phosphate isomerase rpiA	4	0.201	0.035	0.220	0.030	0.211	0.032
sll1070	transketolase	40	0.168	0.070	0.193	0.069	0.180	0.070
sll0329	6-phosphogluconate dehydrogenase	7	0.108	0.090	0.108	0.094	0.108	0.092
slr1793	transaldolase	5	0.047	0.061	0.055	0.091	0.051	0.076
Photosynthesis								
ssr3383	phycobilisome small core linker polypeptide apcC	5	0.695	0.022	0.728	0.020	0.711	0.021
sll0199	plastocyanin petE	20	0.454	0.064	0.512	0.078	0.483	0.071
sll1577	phycocyanin beta subunit cpcB	294	0.447	0.025	0.463	0.026	0.455	0.026
sll1578	phycocyanin alpha subunit cpcA	389	0.437	0.022	0.451	0.021	0.444	0.021
sll1398	photosystem II reaction center 13 kDa protein psb28, psbW, psb13, ycf79	3	0.420	0.410	0.436	0.413	0.428	0.412
slr2067	allophycocyanin alpha subunit apcA	65	0.393	0.073	0.392	0.077	0.392	0.075
ssl3093	phycobilisome small rod linker polypeptide cpcD	11	0.365	0.053	0.381	0.086	0.373	0.069

ORF	Protein Name	MS/MS	Light:Dark (116:114)		Light:Dark (117:114)		Ave Log Ratio	Ave SD
			Log Ratio	SD	Log Ratio	SD		
Signal Transduction								
slr0947	response regulator for energy transfer from phycobilisomes to photosystems rpaB, ycf27	5	-0.016	0.102	0.011	0.095	-0.003	0.098
sll1590	two-component sensor histidine kinase hik20	7	-0.086	0.031	-0.101	0.016	-0.093	0.023
TCA Cycle								
sll0401	citrate synthase	8	0.122	0.054	0.141	0.065	0.131	0.059
slr0665	aconitate hydratase	7	0.039	0.042	0.017	0.048	0.028	0.045
slr1289	isocitrate dehydrogenase (NADP+) icd	12	-0.065	0.027	-0.012	0.022	-0.039	0.024
Transcription								
slr1859	anti-sigma f factor antagonist	7	0.121	0.063	0.171	0.044	0.146	0.054
sll1818	RNA polymerase alpha subunit rpoA	9	0.122	0.107	0.158	0.111	0.140	0.109
slr1265	RNA polymerase gamma-subunit rpoC1	11	0.115	0.046	0.125	0.055	0.120	0.051
sll1789	RNA polymerase beta prime subunit rpoC2	3	0.112	0.159	0.101	0.170	0.107	0.165
sll1787	RNA polymerase beta subunit rpoB	7	-0.007	0.030	0.025	0.050	0.009	0.040
ssr1600	similar to anti-sigma f factor antagonist	2	-0.036	0.003	0.006	0.055	-0.015	0.029
sll1626	LexA repressor	9	-0.131	0.065	-0.109	0.056	-0.120	0.061
Translation								
sll1802	50S ribosomal protein L2 rpl2	2	0.431	0.091	0.474	0.082	0.452	0.086
sll1744	50S ribosomal protein L1 rpl1	11	0.340	0.057	0.366	0.059	0.353	0.058
sll1261	elongation factor TS tsf	7	0.310	0.088	0.320	0.101	0.315	0.094
sll1743	50S ribosomal protein L11 rpl11	4	0.280	0.034	0.311	0.030	0.295	0.032
ssl3436	50S ribosomal protein L29 rpl29	2	0.208	0.574	0.216	0.619	0.212	0.597
sll1804	30S ribosomal protein S3 rps3	3	0.185	0.022	0.210	0.073	0.197	0.048
sll1099	elongation factor Tu tufA	60	0.168	0.046	0.216	0.049	0.192	0.048
slr1463	elongation factor EF-G fus	18	0.150	0.063	0.173	0.065	0.161	0.064
slr1356	30S ribosomal protein S1 rps1a	10	0.124	0.100	0.140	0.106	0.132	0.103
sll1807	50S ribosomal protein L24 rpl24	5	0.121	0.041	0.140	0.033	0.131	0.037
sll1260	30S ribosomal protein S2 rps2	10	0.106	0.047	0.115	0.043	0.110	0.045
sll1244	50S ribosomal protein L9 rpl9	3	-0.014	0.045	-0.005	0.069	-0.010	0.057
sll1746	50S ribosomal protein L12 rpl12	41	-0.020	0.078	-0.030	0.082	-0.025	0.080
slr0434	elongation factor P efp	5	-0.063	0.010	-0.039	0.019	-0.051	0.014
slr0193	RNA-binding protein rbp3	8	-0.090	0.026	-0.050	0.020	-0.070	0.023
sll0145	ribosome releasing factor rrf	2	-0.082	0.008	-0.064	0.003	-0.073	0.006

Appendix G

The list of 150 proteins found in Experiment 1 in Chapter 6 via iTRAQ analysis. This set of experiment consists of 'transient' time-points derived from the dark cycle: 0.5 hr, 1 hr and 1.5 hr against the common reference of 24 hr dark. The ratio is expressed in Log_{10} unit, whereas the error is estimated in standard deviation (SD).

ORF	Protein Name	MS/MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Amino Acids Metabolism								
sll0080	N-acetyl-gamma-glutamyl-phosphate reductase argC	2	-0.191	0.036	-0.023	0.039	-0.108	0.065
sll0927	S-adenosylmethionine synthetase	2	-0.187	0.108	0.015	0.004	-0.060	0.125
sll1058	dihydrodipicolinate reductase dapB	3	-0.218	0.061	-0.021	0.053	-0.287	0.055
sll1363	ketol-acid reductoisomerase ilvC	2	0.053	0.222	-0.228	0.278	0.176	0.271
sll1502	NADH-dependent glutamate synthase large subunit gltB	2	-0.099	0.048	0.014	0.015	-0.065	0.008
sll1682	alanine dehydrogenase	2	-0.173	0.042	-0.064	0.004	-0.110	0.007
sll1750	urease alpha subunit ureC	2	-0.519	0.180	0.173	0.000	-0.451	0.084
slr0032	probable branched-chain amino acid aminotransferase	12	-0.164	0.030	-0.015	0.040	-0.121	0.025
slr0288	glutamate--ammonia ligase glnN	4	-0.057	0.041	0.176	0.055	-0.113	0.053
slr0452	dihydroxyacid dehydratase ilvD	5	-0.349	0.049	-0.082	0.010	-0.231	0.021
slr0963	ferredoxin-sulfite reductase sir	3	-0.165	0.109	-0.095	0.043	-0.103	0.051
slr1756	glutamate--ammonia ligase glnA	17	-0.375	0.113	0.147	0.033	-0.286	0.085
slr2088	acetohydroxy acid synthase ilvG	2	0.051	0.000	0.024	0.005	-0.022	0.032
ATP synthesis								
sll1326	ATP synthase alpha chain atpA	4	-0.119	0.027	0.013	0.034	-0.069	0.022
slr1329	ATP synthase beta subunit atpB	6	-0.104	0.036	-0.076	0.047	-0.017	0.014
Carbohydrate Metabolism								
sll0158	1,4-alpha-glucan branching enzyme glgB	7	0.021	0.025	0.028	0.018	0.044	0.035
sll0945	glycogen synthase glgA	5	-0.158	0.063	-0.022	0.013	-0.090	0.054
sll1212	GDP-mannose 4,6-dehydratase	3	-0.328	0.129	0.005	0.046	-0.228	0.079
sll1213	GDP-fucose synthetase	2	-0.040	0.011	0.002	0.019	-0.010	0.036
sll1356	glycogen phosphorylase	12	0.014	0.034	0.003	0.034	0.062	0.036

ORF	Protein Name	MS/MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr1641	ClpB protein clpB1	2	-0.355	0.094	-0.051	0.015	-0.278	0.064
slr2075	10kD chaperonin groES	3	-0.252	0.348	0.292	0.073	0.019	0.286
slr2076	60kD chaperonin groEL1	7	0.174	0.024	0.018	0.027	0.200	0.047
Glycolysis								
sll0593	glucokinase	2	-0.137	0.005	-0.047	0.026	-0.118	0.038
sll0726	phosphoglucomutase	2	-0.066	0.051	-0.038	0.014	-0.064	0.020
sll1275	pyruvate kinase 2	2	-0.247	0.019	-0.015	0.007	-0.245	0.021
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	7	-0.041	0.023	-0.018	0.017	0.029	0.013
slr0394	phosphoglycerate kinase pgk	21	-0.062	0.046	0.056	0.028	0.037	0.037
slr1349	glucose-6-phosphate isomerase	5	-0.087	0.117	0.069	0.060	-0.094	0.101
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	26	-0.063	0.047	0.046	0.031	0.017	0.050
hypothetical proteins								
sll0274	hypothetical protein	4	-0.212	0.085	0.011	0.022	-0.107	0.031
sll0529	hypothetical protein	2	-0.007	0.017	0.120	0.139	-0.134	0.001
sll0735	hypothetical protein	2	-0.074	0.008	-0.081	0.003	0.000	0.004
sll0822	hypothetical protein	2	-0.107	0.151	-0.329	0.260	-0.090	0.138
sll1106	hypothetical protein	8	0.005	0.176	-0.127	0.028	0.109	0.098
sll1188	hypothetical protein	5	-0.104	0.050	-0.092	0.028	-0.113	0.039
sll1526	hypothetical protein	6	-0.196	0.029	-0.015	0.015	-0.106	0.017
slr0001	hypothetical protein	3	-0.631	0.055	-0.157	0.011	-0.594	0.044
slr0038	hypothetical protein	2	0.076	0.151	-0.089	0.101	-0.010	0.114
slr0244	hypothetical protein	9	-0.019	0.011	-0.073	0.016	0.036	0.014
slr0453	hypothetical protein	5	-0.079	0.051	-0.117	0.032	-0.041	0.049
slr0689	hypothetical protein	3	-0.045	0.101	-0.069	0.085	-0.014	0.089
slr0729	hypothetical protein	11	-0.234	0.072	-0.078	0.048	-0.117	0.040
slr0923	hypothetical protein YCF65 ycf65	4	-0.157	0.283	0.103	0.093	-0.126	0.053
slr1220	hypothetical protein	3	0.095	0.093	-0.170	0.076	0.130	0.042
slr1649	hypothetical protein	7	-0.229	0.043	0.026	0.026	-0.199	0.034
slr1780	hypothetical protein YCF54 ycf54	2	0.017	0.013	-0.324	0.144	0.139	0.018
ssl2781	hypothetical protein	2	-0.152	0.135	-0.012	0.018	-0.089	0.154
ssr1528	hypothetical protein	3	-0.332	0.074	0.097	0.076	-0.233	0.067
ssr2998	hypothetical protein	7	-0.230	0.018	-0.064	0.038	-0.105	0.007
Lipid Metabolism								
slr2089	squalene-hopene-cyclase shc	2	-0.245	0.020	0.078	0.006	-0.169	0.062
ssl2084	acyl carrier protein acpP	2	0.032	0.006	0.087	0.024	0.162	0.008
Metabolism of Cofactor and Vitamin								
sll1282	riboflavin synthase beta subunit ribH	4	-0.100	0.031	-0.144	0.045	-0.095	0.055

ORF	Protein Name	MS/MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1194	photosystem II 12 kDa extrinsic protein psbU	17	-0.219	0.086	0.091	0.035	-0.073	0.081
sll1577	phycocyanin beta subunit cpcB	59	-0.200	0.042	0.508	0.038	-0.169	0.042
sll1578	phycocyanin alpha subunit cpcA	138	-0.335	0.044	0.336	0.021	-0.192	0.038
sll1579	phycobilisome rod linker polypeptide cpcC2	14	-0.257	0.045	0.176	0.034	-0.145	0.039
sll1580	phycobilisome rod linker polypeptide cpcC1	44	-0.275	0.043	0.052	0.023	-0.181	0.033
slr0335	phycobilisome core-membrane linker polypeptide apcE	62	-0.191	0.034	0.077	0.021	-0.151	0.035
slr0737	photosystem I subunit II psaD	7	-0.412	0.015	-0.377	0.029	-0.269	0.049
slr1459	phycobilisome core component apcF	4	-0.378	0.071	0.084	0.015	-0.351	0.034
slr1643	ferredoxin-NADP oxidoreductase petH	32	-0.137	0.042	0.066	0.019	-0.096	0.033
slr1834	P700 apoprotein subunit Ia psaA	8	-0.736	0.046	-0.526	0.021	-0.606	0.021
slr1986	allophycocyanin beta subunit apcB	73	-0.320	0.026	0.185	0.027	-0.222	0.032
slr2051	phycobilisome rod-core linker polypeptide cpcG1	28	-0.237	0.060	0.152	0.050	-0.168	0.058
slr2067	allophycocyanin alpha subunit apcA	39	-0.026	0.052	0.189	0.022	0.036	0.052
ssl3093	phycobilisome small rod linker polypeptide cpcD	5	-0.043	0.105	0.246	0.067	0.219	0.125
ssr3383	phycobilisome small core linker polypeptide apcC	7	-0.110	0.109	0.157	0.032	0.172	0.109
Porphyrin and chlorophyll metabolism								
sll1185	coproporphyrinogen III oxidase, aerobic (oxygen-dependent) hemF	5	-0.111	0.009	-0.014	0.022	-0.086	0.015
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	3	-0.480	0.038	0.346	0.022	-0.414	0.049
slr0536	uroporphyrinogen decarboxylase hemE	4	-0.275	0.097	0.143	0.048	-0.157	0.031
Pyruvate Metabolism								
slr1096	dihydrolipoamide dehydrogenase	2	-0.144	0.038	-0.085	0.139	-0.106	0.010
Regulatory control								
ssl0707	nitrogen regulatory protein P-II glnB	18	-0.318	0.075	0.128	0.044	-0.199	0.080
Replication and Repair								
sll1712	DNA binding protein HU	7	-0.052	0.153	0.023	0.141	0.050	0.178

ORF	Protein Name	MS/ MS	0.5hrD:24hrD (115:114)		1hrD:24hrD (116:114)		1.5hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Transport and Binding Proteins								
sll0680	phosphate-binding periplasmic protein precursor (PBP)	6	0.213	0.034	0.096	0.029	0.207	0.035
sll1341	bacterioferritin	4	-0.380	0.059	0.108	0.005	-0.136	0.038
sll1450	nitrate/nitrite transport system substrate-binding protein nrtA	4	-0.244	0.089	-0.089	0.043	-0.124	0.027
sll1762	periplasmic protein, putative polar amino acid transport system substrate-binding protein	2	0.145	0.006	-0.125	0.010	0.151	0.078
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein urtA	9	-0.145	0.047	-0.034	0.044	-0.043	0.049
slr1890	bacterioferritin	4	-0.365	0.052	0.278	0.028	-0.258	0.042

ORF	Protein Name	MS/MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1028	carbon dioxide concentrating mechanism protein CcmK ccmK2	16	0.021	0.047	-0.006	0.047	0.105	0.041
sll1029	carbon dioxide concentrating mechanism protein CcmK ccmK1	3	0.058	0.085	0.016	0.091	0.049	0.109
sll1525	phosphoribulokinase prk	9	0.151	0.060	-0.220	0.031	-0.014	0.036
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	36	0.079	0.021	-0.136	0.016	-0.017	0.019
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	5	0.065	0.097	-0.050	0.045	0.049	0.054
Cellular processes								
sll1633	cell division protein FtsZ ftsZ	2	0.027	0.094	-0.017	0.039	0.070	0.036
slr1198	antioxidant protein	23	0.081	0.066	-0.092	0.034	-0.005	0.042
slr1516	superoxide dismutase sodB	28	0.015	0.016	-0.058	0.015	0.018	0.009
Energy Metabolism								
sll1499	ferredoxin-dependent glutamate synthase glsF	2	0.016	0.044	-0.038	0.035	-0.086	0.132
sll1987	catalase peroxidase cpx, katG	3	-0.143	0.041	-0.116	0.041	-0.027	0.023
slr0898	ferredoxin--nitrite reductase nirA	3	-0.030	0.016	-0.033	0.023	-0.011	0.028
Folate Biosynthesis								
sll1536	molybdopterin biosynthesis MoeB protein moeB	2	0.134	0.141	-0.189	0.003	-0.007	0.138
Folding, Sorting and Degradation								
sll0020	ATP-dependent Clp protease ATPase subunit	9	-0.069	0.033	-0.076	0.025	-0.038	0.021
sll0057	heat shock protein GrpE grpE	2	-0.044	0.016	-0.085	0.022	-0.157	0.033
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	10	-0.020	0.019	-0.072	0.026	0.002	0.017
sll0408	peptidyl-prolyl cis-trans isomerase	3	0.053	0.008	-0.062	0.028	-0.003	0.016
sll0416	60 kDa chaperonin 2, GroEL2, molecular chaperone groEL-2	5	-0.051	0.011	-0.054	0.040	-0.046	0.023
sll1621	AhpC/TSA family protein	26	-0.045	0.017	-0.001	0.011	-0.012	0.020
sll1694	pilin polypeptide PilA1 pilA1	2	0.179	0.039	0.000	0.034	-0.009	0.022
sll1703	protease IV	2	0.174	0.023	0.018	0.062	0.152	0.003

ORF	Protein Name	MS/MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr2089	squalene-hopene-cyclase shc	5	-0.006	0.019	0.004	0.009	0.003	0.018
ssl2084	acyl carrier protein acpP	3	-0.005	0.055	-0.002	0.058	0.051	0.070
Nucleotide Metabolism								
sll0421	adenylosuccinate lyase purB	3	-0.038	0.036	0.019	0.002	0.013	0.047
sll1043	polyribonucleotide nucleotidyltransferase	5	0.026	0.090	-0.054	0.048	-0.012	0.079
sll1852	nucleoside diphosphate kinase	5	0.149	0.038	-0.062	0.041	0.079	0.031
slr1622	soluble inorganic pyrophosphatase ppa	4	0.382	0.095	0.017	0.004	0.097	0.130
Other								
sll0245	probable GTP binding protein	3	-0.149	0.014	-0.060	0.006	-0.039	0.017
sll0446	unknown protein	2	0.002	0.017	-0.175	0.065	-0.139	0.049
sll0576	putative sugar-nucleotide epimerase/dehydratease	5	-0.100	0.018	-0.004	0.017	-0.049	0.014
sll0588	unknown protein	5	0.156	0.015	0.190	0.012	0.264	0.007
sll0872	unknown protein	8	0.200	0.018	0.264	0.013	0.343	0.015
sll1130	unknown protein	4	-0.006	0.092	-0.078	0.021	-0.061	0.048
sll1559	soluble hydrogenase 42 kD subunit	10	0.218	0.039	-0.080	0.040	0.133	0.022
sll1785	periplasmic protein, function unknown	2	-0.180	0.023	-0.074	0.050	-0.088	0.076
sll1873	unknown protein	14	-0.078	0.027	-0.024	0.029	-0.089	0.028
slr1852	unknown protein	6	-0.010	0.035	0.033	0.038	0.030	0.045
slr1855	unknown protein	6	-0.060	0.050	-0.062	0.040	-0.023	0.020
slr1963	water-soluble carotenoid protein	19	-0.038	0.025	-0.071	0.015	-0.029	0.019
slr2002	cyanophycin synthetase cphA	2	0.065	0.014	-0.016	0.045	0.013	0.019
ssl2296	pterin-4a-carbinolamine dehydratase	3	0.123	0.013	-0.200	0.046	-0.024	0.019
Pentose Phosphate Pathway								
sll0329	6-phosphogluconate dehydrogenase	3	0.038	0.026	0.008	0.009	0.113	0.021
sll1070	transketolase	20	0.084	0.043	-0.047	0.016	0.034	0.013
slr1793	transaldolase	3	0.014	0.082	-0.032	0.034	0.068	0.036
Photosynthesis								
sll0199	plastocyanin petE	6	0.537	0.183	-0.098	0.047	0.082	0.063
sll0258	cytochrome c550 psbV	7	0.113	0.086	-0.018	0.024	0.095	0.068
sll0427	photosystem II manganese-stabilizing polypeptide psbO	2	-0.014	0.261	-0.072	0.113	0.002	0.015
sll0928	allophycocyanin-B apcD	46	0.207	0.027	-0.170	0.021	0.069	0.034
sll0947	light repressed protein A homolog lrtA	3	-0.230	0.025	-0.200	0.059	-0.281	0.093

ORF	Protein Name	MS/ MS	1hrD:24hrD (115:114)		6hrD:24hrD (116:114)		11hrD:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Replication and Repair								
sll1463	cell division protein FtsH ftsH	5	-0.088	0.047	-0.158	0.037	-0.179	0.009
sll1712	DNA binding protein HU	7	0.029	0.054	-0.159	0.016	-0.099	0.046
slr1322	putative modulator of DNA gyrase; TldD tldD	3	-0.021	0.033	-0.063	0.030	-0.017	0.034
slr1894	probable DNA-binding stress protein	7	-0.053	0.035	-0.073	0.015	-0.030	0.015
Signal Transduction								
sll1590	two-component sensor histidine kinase hik20	5	0.099	0.016	-0.056	0.042	0.088	0.015
TCA cycle								
slr0665	aconitate hydratase	4	0.035	0.011	-0.039	0.026	-0.031	0.027
slr1289	isocitrate dehydrogenase (NADP+) icd	9	-0.078	0.025	-0.072	0.041	-0.068	0.025
Transcription								
sll1626	LexA repressor	8	-0.112	0.027	-0.084	0.029	-0.033	0.028
sll1787	RNA polymerase beta subunit rpoB	7	-0.072	0.038	-0.064	0.046	-0.035	0.033
sll1789	RNA polymerase beta prime subunit rpoC2	3	-0.101	0.044	0.003	0.046	-0.077	0.020
slr1265	RNA polymerase gamma-subunit rpoC1	2	0.024	0.019	-0.021	0.041	-0.008	0.024
slr1859	anti-sigma f factor antagonist	5	0.084	0.076	-0.189	0.025	-0.046	0.069
Translation								
sll0145	ribosome releasing factor frr, rrf	3	-0.069	0.042	-0.015	0.029	0.011	0.018
sll1097	30S ribosomal protein S7 rps7	2	0.056	0.022	-0.010	0.018	0.042	0.014
sll1099	elongation factor Tu tufA	35	0.049	0.030	-0.040	0.028	0.004	0.018
sll1260	30S ribosomal protein S2 rps2	5	-0.037	0.067	-0.162	0.017	-0.197	0.050
sll1553	phenylalanyl-tRNA synthetase pheT	2	0.160	0.209	-0.204	0.029	0.075	0.108
sll1744	50S ribosomal protein L1 rpl1	2	-0.036	0.026	-0.212	0.022	-0.107	0.023
sll1746	50S ribosomal protein L12 rpl12	37	-0.092	0.014	-0.073	0.017	-0.048	0.011
sll1804	30S ribosomal protein S3 rps3	2	0.060	0.046	-0.104	0.073	0.019	0.036
sll1812	30S ribosomal protein S5 rps5	2	-0.044	0.063	0.036	0.013	0.023	0.004
sll1816	30S ribosomal protein S13 rps13	2	-0.125	0.057	-0.107	0.007	-0.201	0.047
slr0193	RNA-binding protein rbp3	7	-0.129	0.016	-0.094	0.016	-0.132	0.011
slr0434	elongation factor P efp	4	-0.002	0.014	0.007	0.016	0.005	0.010

Appendix I

The list of 120 proteins found in Experiment 3 in Chapter 6 via iTRAQ analysis. This set of experiment consists of 'transient' time-points derived from the light cycle: 12.5 hr, 13 hr and 13.5 hr against the common reference of 24 hr dark. The ratio is expressed in Log_{10} unit, whereas the error is estimated in standard deviation (SD).

ORF	Protein Name	MS/ MS	12.5hrL:24hrD (115:114)		13hrL:24hrD (116:114)		13.5hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
Amino Acids Metabolism								
sll1363	ketol-acid reductoisomerase ilvC	6	-0.355	0.088	0.055	0.020	-0.327	0.075
sll1502	NADH-dependent glutamate synthase large subunit gltB	3	-0.248	0.051	-0.057	0.027	-0.274	0.077
sll1750	urease alpha subunit ureC	2	-0.361	0.018	0.044	0.015	-0.220	0.015
sll2001	leucine aminopeptidase	2	-0.401	0.114	-0.007	0.002	-0.451	0.103
slr0032	probable branched-chain amino acid aminotransferase	2	-0.135	0.050	0.001	0.102	-0.057	0.005
slr0288	glutamate--ammonia ligase glnN	5	-0.057	0.055	0.168	0.036	-0.040	0.063
slr0452	dihydroxyacid dehydratase ilvD	3	-0.505	0.010	-0.081	0.024	-0.561	0.062
slr0963	ferredoxin-sulfite reductase sir	3	-0.207	0.071	-0.034	0.035	-0.139	0.090
slr1756	glutamate--ammonia ligase glnA	17	-0.436	0.064	0.049	0.020	-0.444	0.063
ATP Synthesis								
sll1326	ATP synthase alpha chain atpA	16	0.177	0.019	-0.080	0.019	0.050	0.028
slr1329	ATP synthase beta subunit atpB	3	0.125	0.243	-0.181	0.047	0.079	0.297
Carbohydrate Metabolism								
sll0158	1,4-alpha-glucan branching enzyme glgB	5	-0.027	0.060	0.004	0.040	-0.004	0.068
sll0945	glycogen synthase glgA	3	-0.066	0.106	-0.031	0.042	-0.052	0.080
sll1213	GDP-fucose synthetase	5	-0.071	0.026	0.041	0.015	-0.005	0.023
sll1356	glycogen phosphorylase	10	0.082	0.042	-0.003	0.020	0.135	0.037
slr0528	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase murE	2	-0.023	0.018	-0.040	0.004	0.019	0.049
slr1176	glucose-1-phosphate adenylyltransferase	6	-0.195	0.046	-0.011	0.025	-0.193	0.087
Carbon Fixation								
sll1028	carbon dioxide concentrating mechanism protein CcmK ccmK2	19	-0.289	0.024	0.058	0.014	-0.059	0.024
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	19	-0.421	0.123	-0.067	0.019	-0.416	0.116

ORF	Protein Name	MS/MS	12.5hrL:24hrD (115:114)		13hrL:24hrD (116:114)		13.5hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr0038	hypothetical protein	4	0.191	0.125	-0.047	0.036	0.132	0.146
slr0244	hypothetical protein	3	-0.135	0.033	-0.269	0.058	-0.156	0.007
slr0453	hypothetical protein	4	-0.061	0.086	-0.024	0.047	-0.042	0.079
slr0689	hypothetical protein	2	-0.002	0.002	0.042	0.003	0.065	0.007
slr0729	hypothetical protein	9	-0.267	0.106	-0.075	0.024	-0.271	0.136
slr1590	hypothetical protein	2	-0.027	0.010	-0.051	0.010	0.016	0.027
ssl0352	hypothetical protein	2	0.115	0.048	0.096	0.016	0.145	0.044
ssr2998	hypothetical protein	4	-0.583	0.023	-0.105	0.023	-0.634	0.049
Lipid Metabolism								
slr2089	squalene-hopene-cyclase shc	2	-0.210	0.001	0.098	0.026	-0.291	0.110
ssl2084	acyl carrier protein acpP	3	-0.006	0.021	-0.135	0.020	-0.135	0.022
Nucleotide Metabolism								
sll0368	uracil phosphoribosyltransferase	2	-0.061	0.004	0.073	0.010	0.061	0.029
sll0421	adenylosuccinate lyase purB	3	-0.208	0.075	0.046	0.012	-0.118	0.075
sll1043	polyribonucleotide nucleotidyltransferase	3	-0.331	0.042	-0.007	0.020	-0.377	0.036
sll1852	nucleoside diphosphate kinase	2	-0.355	0.079	0.085	0.022	-0.460	0.077
slr0185	orotate phosphoribosyltransferase	2	-0.152	0.040	-0.123	0.033	-0.100	0.020
slr1622	soluble inorganic pyrophosphatase ppa	2	-0.362	0.065	0.097	0.010	-0.485	0.208
slr1722	inosine-5'-monophosphate dehydrogenase	2	-0.113	0.140	0.009	0.068	-0.098	0.140
Other								
sll0245	probable GTP binding protein	2	-0.173	0.019	-0.095	0.006	-0.152	0.044
sll0576	putative sugar-nucleotide epimerase/dehydratease	6	-0.205	0.021	0.014	0.020	-0.166	0.035
sll0588	unknown protein	6	-0.072	0.026	0.195	0.012	-0.067	0.022
sll0872	unknown protein	2	-0.424	0.190	0.199	0.005	-0.383	0.113
sll1559	soluble hydrogenase 42 kD subunit	3	-0.285	0.132	0.123	0.017	-0.233	0.159
sll1583	unknown protein	3	0.039	0.033	-0.110	0.012	0.080	0.045
sll1785	periplasmic protein, function unknown	3	0.014	0.052	-0.061	0.043	0.013	0.052
sll1873	unknown protein	16	0.077	0.087	-0.047	0.025	0.116	0.091
slr1852	unknown protein	11	-0.035	0.141	-0.022	0.039	-0.029	0.138
slr1855	unknown protein	9	-0.138	0.058	-0.030	0.031	-0.140	0.049
slr1963	water-soluble carotenoid protein	15	-0.253	0.030	-0.019	0.018	-0.307	0.048
Pentose Phosphate Pathway								
sll0329	6-phosphogluconate dehydrogenase	2	-0.120	0.006	0.033	0.004	-0.083	0.054
sll1070	transketolase	8	-0.367	0.026	-0.006	0.017	-0.388	0.029

ORF	Protein Name	MS/MS	12.5hrL:24hrD (115:114)		13hrL:24hrD (116:114)		13.5hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr1894	probable DNA-binding stress protein	9	-0.256	0.109	-0.024	0.015	-0.286	0.085
TCA Cycle								
slr0665	aconitate hydratase	2	-0.045	0.003	-0.014	0.015	-0.090	0.024
slr1289	isocitrate dehydrogenase (NADP+) icd	4	-0.031	0.066	-0.026	0.049	0.018	0.097
Transcription								
sll1626	LexA repressor	14	-0.249	0.044	-0.036	0.034	-0.244	0.041
sll1787	RNA polymerase beta subunit rpoB	3	0.014	0.062	-0.178	0.117	0.010	0.070
sll1818	RNA polymerase alpha subunit rpoA	2	-0.138	0.023	-0.062	0.053	-0.126	0.001
Translation								
sll0145	ribosome releasing factor rrf	6	-0.022	0.029	0.024	0.033	0.005	0.015
sll1097	30S ribosomal protein S7 rps7	9	0.001	0.034	-0.003	0.017	0.062	0.016
sll1099	elongation factor Tu tufA	23	-0.278	0.066	-0.003	0.016	-0.267	0.045
sll1260	30S ribosomal protein S2 rps2	4	-0.262	0.065	-0.069	0.016	-0.349	0.066
sll1261	elongation factor TS tsf	3	-0.151	0.169	-0.001	0.062	-0.153	0.176
sll1744	50S ribosomal protein L1 rpl1	3	-0.592	0.130	-0.205	0.006	-0.619	0.066
sll1746	50S ribosomal protein L12 rpl12	16	-0.626	0.056	-0.004	0.016	-0.428	0.046
sll1812	30S ribosomal protein S5 rps5	3	0.061	0.025	-0.043	0.029	-0.012	0.025
sll1816	30S ribosomal protein S13 rps13	5	0.129	0.020	-0.117	0.017	0.162	0.031
sll1821	50S ribosomal protein L13 rpl13	2	0.076	0.032	-0.113	0.007	0.090	0.014
slr0193	RNA-binding protein rbp3	3	-0.008	0.024	-0.177	0.012	0.076	0.011
slr0434	elongation factor P efp	2	-0.338	0.042	-0.022	0.007	-0.389	0.086
slr1463	elongation factor EF-G fus	18	-0.246	0.026	-0.022	0.013	-0.302	0.039
slr1984	nucleic acid-binding protein, 30S ribosomal protein S1 homolog nbp1,rps1b	3	-0.234	0.119	-0.025	0.029	-0.207	0.138
Transport and Binding Proteins								
sll0680	phosphate-binding periplasmic protein precursor (PBP)	7	0.276	0.022	0.180	0.034	0.303	0.033
sll1341	bacterioferritin	3	-0.509	0.042	-0.021	0.027	-0.711	0.076
sll1450	nitrate/nitrite transport system substrate-binding protein nrtA	4	-0.173	0.114	0.052	0.018	-0.118	0.075
sll1762	periplasmic protein, putative polar amino acid transport system substrate-binding protein	4	-0.082	0.046	-0.150	0.035	-0.067	0.068
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein	5	-0.069	0.015	0.017	0.027	0.007	0.021

ORF	Protein Name	MS/MS	13hrL:24hrD (115:114)		18hrL:24hrD (116:114)		23hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1031	carbon dioxide concentrating mechanism protein CcmM, putative carboxysome structural protein ccmM	5	0.090	0.058	0.088	0.051	0.068	0.050
sll1525	phosphoribulokinase prk	12	0.080	0.038	0.157	0.034	0.224	0.036
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	4	0.084	0.097	0.157	0.067	0.168	0.057
Cellular Processes								
sll1633	cell division protein FtsZ ftsZ	3	-0.063	0.018	-0.034	0.045	0.013	0.028
slr1198	antioxidant protein	28	0.022	0.062	0.068	0.063	0.070	0.050
slr1516	superoxide dismutase sodB	41	-0.008	0.017	-0.005	0.020	0.019	0.018
Folate Biosynthesis								
sll1536	molybdopterin biosynthesis MoeB protein moeB	4	-0.075	0.141	0.005	0.090	0.007	0.087
Folding, Sorting and Degradation								
sll0020	ATP-dependent Clp protease ATPase subunit	8	0.088	0.061	0.027	0.069	0.006	0.061
sll0057	heat shock protein GrpE grpE	2	0.004	0.032	0.057	0.094	0.002	0.040
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	14	0.029	0.032	0.070	0.039	0.076	0.030
sll0408	peptidyl-prolyl cis-trans isomerase	3	-0.005	0.014	0.080	0.057	0.158	0.041
sll1621	AhpC/TSA family protein	25	-0.076	0.010	-0.053	0.011	-0.044	0.014
sll1694	pilin polypeptide PilA1 pilA1	2	0.057	0.028	0.177	0.032	-0.008	0.033
slr0542	ATP-dependent protease ClpP clpP	3	0.042	0.011	0.061	0.035	0.082	0.036
slr0623	thioredoxin trxA	18	-0.002	0.048	0.013	0.041	0.040	0.040
slr1251	peptidyl-prolyl cis-trans isomerase	3	0.011	0.022	0.021	0.028	0.045	0.036
slr1641	ClpB protein clpB1	2	0.068	0.026	0.001	0.030	0.071	0.030
slr2075	10kD chaperonin groES	11	0.066	0.068	0.132	0.061	0.215	0.060
slr2076	60kD chaperonin groEL1	6	-0.071	0.100	-0.035	0.075	0.054	0.066
Glycolysis								
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	11	0.013	0.061	0.016	0.046	0.045	0.052
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	13	-0.245	0.021	-0.173	0.028	-0.029	0.024
slr1349	glucose-6-phosphate isomerase	8	0.074	0.053	0.075	0.052	0.094	0.048
slr0394	phosphoglycerate kinase pgk	27	-0.034	0.032	-0.007	0.034	0.030	0.025
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	14	-0.042	0.066	0.000	0.057	0.051	0.044
hypothetical proteins								
sll0051	hypothetical protein	5	-0.052	0.013	-0.066	0.012	-0.043	0.011

ORF	Protein Name	MS/MS	13hrL:24hrD (115:114)		18hrL:24hrD (116:114)		23hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
slr1963	water-soluble carotenoid protein	19	-0.066	0.014	-0.039	0.017	-0.025	0.021
slr2002	cyanophycin synthetase cphA	2	0.179	0.027	0.113	0.055	0.187	0.047
Pentose Phosphate Pathway								
sll0329	6-phosphogluconate dehydrogenase	4	-0.039	0.062	0.017	0.042	0.033	0.025
sll0807	pentose-5-phosphate-3-epimerase rpe	2	0.016	0.102	-0.065	0.192	-0.051	0.160
sll1070	transketolase	19	0.057	0.058	0.092	0.049	0.099	0.047
sll1479	6-phosphogluconolactonase	3	0.105	0.127	0.120	0.128	0.128	0.101
slr1793	transaldolase	2	0.004	0.139	-0.009	0.161	0.075	0.050
Photosynthesis								
sll0199	plastocyanin petE	6	0.302	0.077	0.280	0.064	0.300	0.029
sll0258	cytochrome c550 psbV	14	0.099	0.036	0.096	0.041	0.094	0.041
sll0427	photosystem II manganese-stabilizing polypeptide psbO	6	-0.029	0.053	0.077	0.049	0.106	0.043
sll0928	allophycocyanin-B apcD	39	0.194	0.034	0.239	0.038	0.288	0.029
sll0947	light repressed protein A homolog lrtA	4	-0.305	0.011	-0.228	0.026	-0.199	0.004
sll1194	photosystem II 12 kDa extrinsic protein psbU	15	0.105	0.041	0.130	0.036	0.165	0.045
sll1577	phycocyanin beta subunit cpcB	141	0.317	0.027	0.302	0.024	0.285	0.022
sll1578	phycocyanin alpha subunit cpcA	242	0.434	0.020	0.420	0.019	0.414	0.019
sll1579	phycobilisome rod linker polypeptide cpcC2	12	0.252	0.047	0.247	0.026	0.241	0.030
sll1580	phycobilisome rod linker polypeptide cpcC1	60	0.173	0.027	0.159	0.023	0.165	0.027
slr0335	phycobilisome core-membrane linker polypeptide apcE	68	0.077	0.021	0.099	0.021	0.118	0.020
slr0737	photosystem I subunit II psaD	11	-0.481	0.040	-0.214	0.043	-0.270	0.036
slr0906	photosystem II core light harvesting protein psbB	2	-0.650	0.168	-0.322	0.095	-0.491	0.095
slr1459	phycobilisome core component apcF	7	0.106	0.025	0.105	0.034	0.143	0.034
slr1643	ferredoxin-NADP oxidoreductase petH	26	0.000	0.018	-0.022	0.022	0.002	0.014
slr1834	P700 apoprotein subunit Ia psaA	11	-0.652	0.072	-0.308	0.031	-0.484	0.041
slr1986	allophycocyanin beta subunit apcB	85	0.203	0.026	0.221	0.026	0.214	0.022
slr2051	phycobilisome rod-core linker polypeptide cpcG1	29	0.000	0.037	0.037	0.038	0.088	0.030
slr2067	allophycocyanin alpha subunit apcA	18	0.054	0.037	0.084	0.030	0.022	0.034
ssl3093	phycobilisome small rod linker polypeptide cpcD	7	0.030	0.018	0.166	0.014	0.270	0.026
ssr2831	photosystem I subunit IV psaE	3	-0.418	0.017	-0.170	0.028	-0.240	0.006

ORF	Protein Name	MS/ MS	13hrL:24hrD (115:114)		18hrL:24hrD (116:114)		23hrL:24hrD (117:114)	
			Log Ratio	SD	Log Ratio	SD	Log Ratio	SD
sll1260	30S ribosomal protein S2 rps2	10	-0.044	0.021	-0.007	0.030	0.025	0.029
sll1744	50S ribosomal protein L1 rpl1	2	-0.028	0.132	0.028	0.144	0.076	0.103
sll1746	50S ribosomal protein L12 rpl12	49	-0.099	0.019	-0.091	0.017	-0.072	0.013
sll1804	30S ribosomal protein S3 rps3	3	0.126	0.014	0.106	0.045	0.164	0.030
sll1812	30S ribosomal protein S5 rps5	4	-0.087	0.062	-0.068	0.076	-0.060	0.069
sll1816	30S ribosomal protein S13 rps13	2	-0.001	0.015	-0.041	0.015	0.000	0.009
slr0193	RNA-binding protein rbp3	8	-0.199	0.032	-0.150	0.018	-0.124	0.015
slr0434	elongation factor P efp	8	-0.046	0.018	-0.058	0.017	-0.015	0.021
slr0744	translation initiation factor IF-2 infB	2	-0.058	0.039	-0.130	0.007	-0.116	0.003
slr1356	30S ribosomal protein S1 rps1a	5	-0.178	0.081	-0.129	0.126	-0.059	0.052
slr1463	elongation factor EF-G fus	8	-0.046	0.021	-0.045	0.021	-0.038	0.024
Transport and Binding Proteins								
sll0680	phosphate-binding periplasmic protein precursor (PBP)	2	0.220	0.036	0.114	0.026	0.107	0.021
sll1341	bacterioferritin	3	0.074	0.225	0.053	0.197	0.042	0.112
sll1450	nitrate/nitrite transport system substrate-binding protein nrtA	7	0.106	0.034	0.076	0.022	0.031	0.053
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein urtA	2	0.055	0.092	0.055	0.045	0.068	0.049
slr1234	protein kinase C inhibitor	2	0.435	0.030	0.351	0.050	0.320	0.033
slr1890	bacterioferritin	7	0.179	0.042	0.178	0.032	0.186	0.042

Appendix K

The relative protein abundance of the 82 proteins found in all 4 (iTRAQ) experiments in Chapter 6 across the period of 12 hr light and 12 hr dark cycles. All time-points must have at least 2 or more MS/MS spectra. The ratio is expressed in Log_{10} unit relative to the common reference of 24 hr dark, whereas the error is estimated in standard deviation (SD).

ORF	Protein Name	Dark cycle (Log Ratio)						Light cycle (Log Ratio)				Dark cycle (SD)						Light cycle (SD)			
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
Amino Acids Metabolism																					
sll1363	ketol-acid reductoisomerase ilvC	0.293	0.011	0.415	0.026	0.031	-0.449	-0.038	-0.420	-0.079	0.031	0.222	0.031	0.271	0.022	0.033	0.088	0.001	0.075	0.019	0.022
sll1502	NADH-dependent glutamate synthase large subunit gltB	-0.160	-0.047	-0.126	-0.075	-0.008	-0.229	-0.038	-0.255	-0.004	-0.008	0.048	0.030	0.008	0.023	0.026	0.051	0.053	0.077	0.036	0.046
sll1750	urease alpha subunit ureC	-0.604	0.088	-0.536	-0.121	-0.004	-0.263	0.142	-0.122	0.145	-0.004	0.180	0.006	0.084	0.009	0.019	0.018	0.073	0.015	0.090	0.078
slr0288	glutamate--ammonia ligase glnN	-0.021	0.211	-0.078	0.181	0.206	-0.071	0.155	-0.053	0.173	0.206	0.041	0.038	0.053	0.055	0.035	0.055	0.018	0.063	0.022	0.030
slr1756	glutamate--ammonia ligase glnA	-0.455	0.066	-0.366	-0.022	0.039	-0.404	0.082	-0.411	0.120	0.039	0.113	0.046	0.085	0.023	0.021	0.064	0.042	0.063	0.049	0.040
ATP Synthesis																					
sll1326	ATP synthase alpha chain atpA	-0.117	0.015	-0.067	-0.133	-0.056	0.192	-0.064	0.066	0.026	-0.056	0.027	0.019	0.022	0.029	0.018	0.019	0.023	0.028	0.046	0.046
Carbohydrate Metabolism																					
sll0158	1,4-alpha-glucan branching enzyme glgB	-0.057	-0.050	-0.034	-0.053	-0.004	-0.044	-0.013	-0.021	-0.020	-0.004	0.025	0.045	0.035	0.027	0.019	0.060	0.046	0.068	0.040	0.044
sll0945	glycogen synthase glgA	-0.214	-0.078	-0.146	-0.178	-0.201	0.044	0.078	0.057	0.047	-0.201	0.063	0.007	0.054	0.013	0.001	0.106	0.038	0.080	0.040	0.044
sll1213	GDP-fucose synthetase	-0.074	-0.033	-0.045	0.022	0.022	-0.135	-0.023	-0.069	-0.005	0.022	0.011	0.018	0.036	0.010	0.055	0.026	0.040	0.023	0.032	0.022
sll1356	glycogen phosphorylase	-0.078	-0.090	-0.030	-0.072	-0.056	0.011	-0.075	0.063	-0.102	-0.056	0.034	0.023	0.036	0.016	0.014	0.042	0.013	0.037	0.013	0.026
slr1176	glucose-1-phosphate adenylyltransferase	-0.241	-0.031	-0.222	-0.044	-0.008	-0.207	-0.023	-0.205	-0.027	-0.008	0.058	0.042	0.067	0.025	0.028	0.046	0.038	0.087	0.029	0.028

ORF	Protein Name	Dark cycle (Log Ratio)										Light cycle (Log Ratio)									
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
Carbon fixation																					
sll1525	phosphoribulokinase prk	-0.114	0.151	-0.028	-0.220	-0.014	-0.249	0.080	-0.294	0.157	-0.014	0.097	0.060	0.065	0.031	0.036	0.073	0.038	0.086	0.034	0.036
sll1028	carbon dioxide concentrating mechanism protein CcmK ccmK2	-0.355	0.021	-0.267	-0.006	0.105	-0.378	-0.031	-0.148	0.010	0.105	0.039	0.047	0.016	0.047	0.041	0.024	0.022	0.024	0.020	0.026
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	-0.376	0.079	-0.277	-0.136	-0.017	-0.242	0.113	-0.236	0.143	-0.017	0.053	0.021	0.039	0.016	0.019	0.123	0.026	0.116	0.029	0.029
Cellular processes																					
slr1198	antioxidant protein	-0.142	0.081	-0.017	-0.092	-0.005	-0.237	0.022	-0.315	0.068	-0.005	0.061	0.066	0.057	0.034	0.042	0.066	0.062	0.053	0.063	0.050
slr1516	superoxide dismutase sodB	-0.150	0.015	-0.089	-0.058	0.018	-0.164	-0.008	-0.217	-0.005	0.018	0.017	0.016	0.015	0.015	0.009	0.021	0.017	0.013	0.020	0.018
Folding, Sorting and Degradation																					
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	-0.217	-0.020	-0.133	-0.072	0.002	-0.232	0.029	-0.257	0.070	0.002	0.030	0.019	0.050	0.026	0.017	0.024	0.032	0.022	0.039	0.030
sll0408	peptidyl-prolyl cis-trans isomerase	-0.399	0.053	-0.287	-0.062	-0.003	-0.144	-0.005	-0.153	0.080	-0.003	0.025	0.008	0.041	0.028	0.016	0.279	0.014	0.309	0.057	0.041
sll1621	AhpC/TSA family protein	-0.297	-0.045	-0.284	-0.001	-0.012	-0.370	-0.076	-0.345	-0.053	-0.012	0.028	0.017	0.026	0.011	0.020	0.032	0.010	0.039	0.011	0.014
slr0623	thioredoxin trxA	-0.216	0.059	-0.103	-0.080	0.005	-0.419	-0.002	-0.517	0.013	0.005	0.052	0.050	0.038	0.060	0.025	0.162	0.048	0.192	0.041	0.040
slr1251	peptidyl-prolyl cis-trans isomerase	-0.143	-0.020	-0.087	-0.017	0.010	-0.105	0.011	-0.104	0.021	0.010	0.049	0.040	0.038	0.021	0.024	0.028	0.022	0.021	0.028	0.036
slr2075	10kD chaperonin groES	-0.358	0.186	-0.087	-0.171	0.011	0.090	0.066	0.033	0.132	0.011	0.348	0.036	0.286	0.040	0.024	0.050	0.068	0.044	0.061	0.060
slr2076	60kD chaperonin groEL1	0.109	-0.048	0.135	-0.045	-0.128	0.065	-0.071	0.101	-0.035	-0.128	0.024	0.024	0.047	0.059	0.033	0.036	0.100	0.042	0.075	0.066
Glycolysis																					
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	-0.097	-0.074	-0.028	-0.193	-0.142	-0.410	-0.245	-0.411	-0.173	-0.142	0.023	0.022	0.013	0.034	0.019	0.055	0.021	0.099	0.028	0.024
slr0394	phosphoglycerate kinase pgk	-0.097	0.021	0.002	-0.059	-0.018	-0.267	-0.034	-0.337	-0.007	-0.018	0.046	0.021	0.037	0.019	0.020	0.029	0.032	0.047	0.034	0.025
slr1349	glucose-6-phosphate isomerase	-0.219	-0.063	-0.226	-0.032	-0.012	-0.094	0.074	-0.037	0.075	-0.012	0.117	0.012	0.101	0.011	0.023	0.012	0.053	0.028	0.052	0.048

ORF	Protein Name	Dark cycle (Log Ratio)					Light cycle (Log Ratio)					Dark cycle (SD)					Light cycle (SD)				
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	-0.086	0.023	-0.006	-0.130	-0.054	-0.123	-0.042	-0.118	0.000	-0.054	0.047	0.055	0.050	0.037	0.022	0.110	0.066	0.164	0.057	0.044
hypothetical proteins																					
sll0588	hypothetical protein	-0.109	0.156	-0.049	0.190	0.264	-0.160	0.107	-0.156	0.170	0.264	0.012	0.015	0.014	0.012	0.007	0.026	0.007	0.022	0.023	0.019
sll1526	(CheR methyltransferase)	-0.269	-0.087	-0.179	-0.052	-0.095	-0.156	-0.035	-0.134	-0.020	-0.095	0.029	0.071	0.017	0.030	0.030	0.032	0.024	0.057	0.073	0.029
sll1873	hypothetical protein	-0.039	-0.078	0.014	-0.024	-0.089	0.011	-0.114	0.050	-0.108	-0.089	0.099	0.027	0.109	0.029	0.028	0.087	0.026	0.091	0.015	0.018
slr0038	(Mitochondrial energy transfer proteins signature)	0.184	0.020	0.099	0.039	0.075	0.276	0.038	0.217	-0.104	0.075	0.151	0.055	0.114	0.018	0.030	0.125	0.051	0.146	0.074	0.019
slr0244	(Universal stress protein family)	-0.086	-0.141	-0.031	-0.130	-0.125	-0.126	-0.261	-0.147	-0.202	-0.125	0.011	0.026	0.014	0.025	0.031	0.033	0.041	0.007	0.038	0.030
slr0729	hypothetical protein	-0.188	-0.032	-0.071	-0.195	-0.100	-0.223	-0.031	-0.226	-0.055	-0.100	0.072	0.047	0.040	0.028	0.031	0.106	0.052	0.136	0.059	0.047
slr1852	hypothetical protein	-0.278	-0.010	-0.193	0.033	0.030	-0.145	-0.132	-0.139	-0.067	0.030	0.027	0.035	0.026	0.038	0.045	0.141	0.086	0.138	0.037	0.026
slr1855	(N-acylglucosamine 2-epimerase)	-0.279	-0.060	-0.249	-0.062	-0.023	-0.209	-0.101	-0.211	-0.061	-0.023	0.065	0.050	0.055	0.040	0.020	0.058	0.027	0.049	0.068	0.024
ssr2998	hypothetical protein	-0.210	-0.043	-0.085	-0.088	-0.082	-0.637	-0.159	-0.688	-0.067	-0.082	0.018	0.024	0.007	0.012	0.013	0.023	0.024	0.049	0.012	0.021
Lipid Metabolism																					
slr2089	squalene-hopene-cyclase shc	-0.329	-0.006	-0.254	0.004	0.003	-0.286	0.022	-0.367	0.063	0.003	0.020	0.019	0.062	0.009	0.018	0.001	0.033	0.110	0.023	0.013
ssl2084	acyl carrier protein acpP	-0.059	-0.005	0.071	-0.002	0.051	0.035	-0.094	-0.095	0.004	0.051	0.006	0.055	0.008	0.058	0.070	0.021	0.011	0.022	0.028	0.036
Nucleotide Metabolism																					
sll1043	polyribonucleotide nucleotidyltransferase	-0.075	0.026	-0.077	-0.054	-0.012	-0.261	0.063	-0.307	0.028	-0.012	0.037	0.090	0.023	0.048	0.079	0.042	0.069	0.036	0.127	0.060
sll1852	nucleoside diphosphate kinase	-0.307	0.149	-0.158	-0.062	0.079	-0.317	0.123	-0.421	0.138	0.079	0.047	0.038	0.051	0.041	0.031	0.079	0.028	0.077	0.025	0.032
slr1622	soluble inorganic pyrophosphatase ppa	0.041	0.382	0.111	0.017	0.097	-0.358	0.100	-0.482	0.101	0.097	0.039	0.095	0.033	0.004	0.130	0.065	0.046	0.208	0.056	0.072
Other																					
sll0576	putative sugar-nucleotide epimerase/dehydratase	-0.308	-0.100	-0.330	-0.004	-0.049	-0.296	-0.078	-0.257	-0.046	-0.049	0.046	0.018	0.044	0.017	0.014	0.021	0.030	0.035	0.048	0.038

ORF	Protein Name	Dark cycle (Log Ratio)						Light cycle (Log Ratio)				Dark cycle (SD)						Light cycle (SD)			
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
sll1559	soluble hydrogenase 42 kD subunit	-0.346	0.218	-0.241	-0.080	0.133	-0.232	0.175	-0.181	0.164	0.133	0.077	0.039	0.084	0.040	0.022	0.132	0.044	0.159	0.048	0.036
slr1963	water-soluble carotenoid protein	-0.339	-0.038	-0.296	-0.071	-0.029	-0.300	-0.066	-0.354	-0.039	-0.029	0.132	0.025	0.054	0.015	0.019	0.030	0.014	0.048	0.017	0.021
Pentose Phosphate Pathway																					
sll0329	6-phosphogluconate dehydrogenase	-0.380	0.038	-0.231	0.008	0.113	-0.192	-0.039	-0.155	0.017	0.113	0.133	0.026	0.166	0.009	0.021	0.006	0.062	0.054	0.042	0.025
sll1070	transketolase	-0.185	0.084	-0.110	-0.047	0.034	-0.303	0.057	-0.325	0.092	0.034	0.046	0.043	0.041	0.016	0.013	0.026	0.058	0.029	0.049	0.047
Photosynthesis																					
sll0258	cytochrome c550 psbV	0.076	0.113	0.109	-0.018	0.095	0.071	0.099	-0.026	0.096	0.095	0.036	0.086	0.017	0.024	0.068	0.036	0.036	0.064	0.041	0.041
sll0427	photosystem II manganese-stabilizing polypeptide psbO	-0.049	-0.014	-0.007	-0.072	0.002	-0.183	-0.029	-0.151	0.077	0.002	0.027	0.261	0.024	0.113	0.015	0.107	0.053	0.090	0.049	0.043
sll0928	allophycocyanin-B apcD	-0.198	0.207	-0.016	-0.170	0.069	-0.371	0.194	-0.521	0.239	0.069	0.052	0.027	0.043	0.021	0.034	0.069	0.034	0.045	0.038	0.029
sll0947	light repressed protein A homolog lrtA	-0.229	-0.230	-0.224	-0.200	-0.281	-0.099	-0.305	-0.151	-0.228	-0.281	0.177	0.025	0.232	0.059	0.093	0.312	0.011	0.272	0.026	0.004
sll1194	photosystem II 12 kDa extrinsic protein psbU	-0.183	0.128	-0.036	-0.172	0.037	-0.233	0.105	-0.300	0.130	0.037	0.086	0.023	0.081	0.028	0.022	0.089	0.041	0.128	0.036	0.045
sll1577	phycocyanin beta subunit cpcB	-0.317	0.390	-0.286	-0.048	0.163	-0.248	0.317	-0.261	0.302	0.163	0.042	0.024	0.042	0.011	0.012	0.073	0.027	0.072	0.024	0.022
sll1578	phycocyanin alpha subunit cpcA	-0.252	0.418	-0.109	-0.128	0.210	-0.191	0.434	-0.146	0.420	0.210	0.044	0.016	0.038	0.010	0.010	0.038	0.020	0.041	0.019	0.019
sll1579	phycobilisome rod linker polypeptide cpcC2	-0.222	0.211	-0.109	-0.009	0.066	-0.213	0.252	-0.297	0.247	0.066	0.045	0.035	0.039	0.035	0.021	0.119	0.047	0.081	0.026	0.030
sll1580	phycobilisome rod linker polypeptide cpcC1	-0.220	0.107	-0.127	-0.084	-0.054	-0.160	0.173	-0.189	0.159	-0.054	0.043	0.025	0.033	0.032	0.014	0.065	0.027	0.083	0.023	0.027
slr0335	phycobilisome core-membrane linker polypeptide apcE	-0.182	0.087	-0.142	-0.038	-0.035	-0.126	0.077	-0.150	0.099	-0.035	0.034	0.020	0.035	0.016	0.014	0.042	0.021	0.040	0.021	0.020
slr0737	photosystem I subunit II psaD	-0.368	-0.334	-0.226	-0.373	-0.163	-0.360	-0.481	-0.389	-0.214	-0.163	0.015	0.050	0.049	0.036	0.035	0.137	0.040	0.158	0.043	0.036
slr1459	phycobilisome core component apcF	-0.407	0.055	-0.380	-0.076	0.022	-0.531	0.106	-0.587	0.105	0.022	0.071	0.049	0.034	0.017	0.014	0.129	0.025	0.104	0.034	0.034

ORF	Protein Name	Dark cycle (Log Ratio)						Light cycle (Log Ratio)				Dark cycle (SD)						Light cycle (SD)			
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
slr1643	ferredoxin-NADP oxidoreductase petH	-0.232	-0.029	-0.191	-0.052	-0.060	-0.182	0.000	-0.160	-0.022	-0.060	0.042	0.044	0.033	0.025	0.021	0.058	0.018	0.062	0.022	0.014
slr1834	P700 apoprotein subunit Ia psaA	-0.501	-0.291	-0.371	-0.350	0.043	-0.438	-0.652	-0.453	-0.308	0.043	0.046	0.085	0.021	0.067	0.072	0.043	0.072	0.064	0.031	0.041
slr1986	allophycocyanin beta subunit apcB	-0.297	0.208	-0.198	-0.094	0.070	-0.306	0.203	-0.323	0.221	0.070	0.026	0.049	0.032	0.035	0.042	0.080	0.026	0.068	0.026	0.022
slr2051	phycobilisome rod-core linker polypeptide cpcG1	-0.315	0.074	-0.246	-0.043	-0.015	-0.286	0.000	-0.351	0.037	-0.015	0.060	0.042	0.058	0.026	0.023	0.051	0.037	0.057	0.038	0.030
slr2067	allophycocyanin alpha subunit apcA	-0.218	-0.003	-0.156	-0.231	-0.161	-0.234	0.054	-0.227	0.084	-0.161	0.052	0.147	0.052	0.122	0.136	0.048	0.037	0.055	0.030	0.034
Porphyrin and chlorophyll metabolism																					
sll1185	coproporphyrinogen III oxidase, aerobic (oxygen-dependent) hemF	-0.203	-0.106	-0.177	-0.076	-0.077	-0.194	-0.133	-0.214	-0.130	-0.077	0.009	0.043	0.015	0.036	0.023	0.039	0.012	0.036	0.038	0.043
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	-0.674	0.152	-0.608	0.035	0.102	-0.417	0.155	-0.321	0.155	0.102	0.038	0.024	0.049	0.034	0.023	0.032	0.064	0.016	0.047	0.055
Pyruvate Metabolism																					
slr1096	dihydrolipoamide dehydrogenase	-0.066	-0.007	-0.029	-0.043	-0.040	-0.135	0.004	-0.077	0.054	-0.040	0.038	0.082	0.010	0.046	0.019	0.044	0.069	0.029	0.100	0.053
Regulatory Control																					
ssl0707	nitrogen regulatory protein P-II glnB	-0.307	0.139	-0.188	0.006	0.177	-0.382	0.093	-0.311	0.107	0.177	0.075	0.041	0.080	0.038	0.031	0.086	0.021	0.074	0.025	0.024
Replication and Repair																					
sll1712	DNA binding protein HU	-0.046	0.029	0.056	-0.159	-0.099	-0.269	-0.047	-0.405	0.032	-0.099	0.153	0.054	0.178	0.016	0.046	0.071	0.059	0.087	0.060	0.081
slr1894	probable DNA-binding stress protein	-0.270	-0.053	-0.203	-0.073	-0.030	-0.333	-0.101	-0.363	-0.063	-0.030	0.044	0.035	0.050	0.015	0.015	0.109	0.019	0.085	0.013	0.016

ORF	Protein Name	Dark cycle (Log Ratio)					Light cycle (Log Ratio)					Dark cycle (SD)					Light cycle (SD)				
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
TCA cycle																					
slr1289	isocitrate dehydrogenase (NADP+) icd	-0.343	-0.078	-0.171	-0.072	-0.068	-0.030	-0.025	0.019	-0.013	-0.068	0.033	0.025	0.038	0.041	0.025	0.066	0.050	0.097	0.056	0.052
Translation																					
sll0145	ribosome releasing factor frr, rrf	-0.164	-0.069	-0.057	-0.015	0.011	-0.056	-0.010	-0.029	0.034	0.011	0.014	0.042	0.013	0.029	0.018	0.029	0.002	0.015	0.004	0.002
sll1099	elongation factor Tu tufA	-0.123	0.049	-0.027	-0.040	0.004	-0.223	0.052	-0.212	0.089	0.004	0.055	0.030	0.061	0.028	0.018	0.066	0.048	0.045	0.047	0.040
sll1260	30S ribosomal protein S2 rps2	-0.338	-0.037	-0.299	-0.162	-0.197	-0.236	-0.044	-0.324	-0.007	-0.197	0.109	0.067	0.154	0.017	0.050	0.065	0.021	0.066	0.030	0.029
sll1746	50S ribosomal protein L12 rpl12	-0.573	-0.092	-0.604	-0.073	-0.048	-0.721	-0.099	-0.523	-0.091	-0.048	0.050	0.014	0.053	0.017	0.011	0.056	0.019	0.046	0.017	0.013
sll1812	30S ribosomal protein S5 rps5	0.025	-0.044	-0.004	0.036	0.023	0.017	-0.087	-0.056	-0.068	0.023	0.043	0.063	0.022	0.013	0.004	0.025	0.062	0.025	0.076	0.069
sll1816	30S ribosomal protein S13 rps13	0.097	-0.125	0.175	-0.107	-0.201	0.245	-0.001	0.279	-0.041	-0.201	0.028	0.057	0.019	0.007	0.047	0.020	0.015	0.031	0.015	0.009
slr0193	RNA-binding protein rbp3	0.065	-0.129	0.197	-0.094	-0.132	-0.030	-0.199	0.054	-0.150	-0.132	0.017	0.016	0.015	0.016	0.011	0.024	0.032	0.011	0.018	0.015
slr0434	elongation factor P efp	-0.286	-0.002	-0.195	0.007	0.005	-0.362	-0.046	-0.413	-0.058	0.005	0.022	0.014	0.014	0.016	0.010	0.042	0.018	0.086	0.017	0.021
slr1463	elongation factor EF-G fus	-0.172	-0.068	-0.127	-0.065	-0.056	-0.271	-0.046	-0.327	-0.045	-0.056	0.036	0.025	0.029	0.024	0.023	0.026	0.021	0.039	0.021	0.024
Transport and Binding Proteins																					
sll0680	phosphate-binding periplasmic protein precursor (PBP)	0.235	0.118	0.228	0.044	0.059	0.316	0.220	0.343	0.114	0.059	0.034	0.000	0.035	0.028	0.031	0.022	0.036	0.033	0.026	0.021
sll1341	bacterioferritin	-0.424	0.063	-0.180	-0.328	0.030	-0.413	0.074	-0.616	0.053	0.030	0.059	0.119	0.038	0.094	0.037	0.042	0.225	0.076	0.197	0.112
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein urtA	-0.011	0.100	0.091	-0.136	0.022	-0.031	0.055	0.045	0.055	0.022	0.047	0.024	0.049	0.026	0.021	0.015	0.092	0.021	0.045	0.049

Appendix L

The relative mRNA (transcript) abundance of the 26 genes investigated in Chapter 6 across the period of 12 hr light and 12 hr dark cycles. All time-points were repeated with at least 3 replicates. The 23S rRNA was used as housekeeping gene, and each ratio is normalised against the 23S rRNA. The relative ratio is expressed against 24 hr dark sample in Log_{10} unit. The error is estimated in standard deviation (SD).

ORF	Gene name	Dark cycle (Ratio)						Light cycle (Ratio)				Dark cycle (SD)						Light cycle (SD)			
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	-0.085	-0.174	-0.054	-1.271	-0.643	0.087	0.189	0.224	-0.469	-0.314	0.025	0.086	0.032	0.188	0.048	0.086	0.035	0.064	0.103	0.061
sll0680	phosphate-binding periplasmic protein precursor (PBP)	-0.333	-0.185	-0.079	-1.034	-0.846	0.401	0.318	0.390	-0.204	-0.111	0.043	0.049	0.091	0.053	0.066	0.040	0.062	0.025	0.044	0.043
sll1226	hydrogenase subunit of the bidirectional hydrogenase hoxH	0.178	0.010	0.007	-0.302	-0.303	0.058	0.222	0.383	-0.095	0.009	0.068	0.026	0.107	0.053	0.099	0.118	0.075	0.098	0.038	0.072
sll1260	30S ribosomal protein S2 rps2	-0.096	-0.125	-0.062	-1.214	-0.936	0.036	0.274	0.114	-0.690	-0.766	0.066	0.080	0.154	0.136	0.128	0.105	0.121	0.089	0.091	0.120
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	-0.061	-0.188	0.102	-1.416	-0.820	0.072	0.256	0.244	-0.187	-0.105	0.062	0.070	0.015	0.056	0.018	0.108	0.043	0.038	0.034	0.026
sll1577	phycocyanin beta subunit cpcB	0.145	-0.075	0.021	-1.855	-1.126	0.112	0.307	0.384	-0.245	-0.184	0.033	0.018	0.030	0.122	0.046	0.036	0.039	0.029	0.133	0.043
sll1712	DNA binding protein HU	-0.037	-0.496	-0.014	-1.445	-1.142	0.002	0.247	-0.015	-0.053	-0.143	0.080	0.131	0.011	0.127	0.108	0.006	0.042	0.025	0.042	0.062
sll1746	50S ribosomal protein L12 rpl12	-0.005	-0.018	0.077	-1.390	-0.803	-0.150	0.223	0.031	-1.060	-0.883	0.055	0.048	0.041	0.026	0.074	0.024	0.037	0.042	0.061	0.028
sll1852	nucleoside diphosphate kinase	-0.071	-0.154	-0.120	-0.876	-0.776	0.211	0.351	0.212	-0.589	-0.676	0.049	0.047	0.107	0.033	0.046	0.075	0.089	0.049	0.099	0.048
sll1994	porphobilinogen synthase (5-aminolevulinate dehydratase) hemB	-0.153	-0.254	-0.089	-1.394	-0.961	0.257	0.325	0.483	-0.203	-0.153	0.080	0.093	0.078	0.151	0.080	0.039	0.079	0.011	0.088	0.081

ORF	Gene name	Dark cycle (Log Ratio)						Light cycle (Log Ratio)				Dark cycle (SD)						Light cycle (SD)			
		0.5	1	1.5	6	11	12.5	13	13.5	18	23	0.5	1	1.5	6	11	12.5	13	13.5	18	23
slr0434	elongation factor P efp	-0.155	-0.222	-0.185	-0.980	-1.015	0.111	0.136	0.233	-0.387	-0.402	0.051	0.125	0.157	0.069	0.095	0.053	0.057	0.142	0.056	0.157
slr0394	phosphoglycerate kinase pgk	-0.154	-0.241	-0.086	-1.335	-0.930	0.456	0.529	0.483	-0.012	-0.125	0.042	0.088	0.018	0.015	0.066	0.072	0.039	0.032	0.021	0.040
slr1349	glucose-6-phosphate isomerase	-0.172	-0.111	-0.096	-0.886	-0.846	0.208	0.296	0.281	-0.031	-0.170	0.111	0.064	0.061	0.065	0.054	0.103	0.027	0.049	0.042	0.086
slr1622	soluble inorganic pyrophosphatase ppa	0.099	0.063	0.121	-1.335	-1.125	0.219	0.125	0.347	-0.411	-0.337	0.067	0.039	0.018	0.048	0.154	0.091	0.092	0.054	0.115	0.143
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbpI	-0.059	-0.100	-0.022	-1.293	-0.832	0.118	0.287	0.231	-0.233	-0.325	0.020	0.064	0.043	0.048	0.094	0.142	0.033	0.051	0.025	0.056
slr2002	cyanophycin synthetase cphA	0.036	-0.329	-0.108	-0.435	-0.392	0.097	0.292	0.249	0.088	0.104	0.190	0.237	0.070	0.097	0.075	0.065	0.067	0.073	0.026	0.085
sll1502	NADH-dependent glutamate synthase large subunit gltB	-0.196	-0.204	0.032	-0.352	-0.186	-0.085	0.259	0.239	0.291	0.283	0.046	0.154	0.108	0.177	0.213	0.215	0.082	0.049	0.047	0.161
sll1194	photosystem II 12 kDa extrinsic protein psbU	-0.035	0.038	0.083	-0.773	-0.318	0.196	0.608	0.118	0.260	0.145	0.092	0.168	0.060	0.156	0.099	0.094	0.145	0.107	0.107	0.107
slr1834	P700 apoprotein subunit Ia psaA	-0.031	-0.082	-0.005	-0.889	-0.783	0.167	0.623	0.418	0.233	0.236	0.101	0.054	0.085	0.065	0.050	0.158	0.098	0.113	0.082	0.089
slr1963	water-soluble carotenoid protein	-0.194	-0.152	-0.110	-0.888	-0.442	-0.014	0.012	0.023	-0.623	-0.445	0.016	0.017	0.015	0.042	0.064	0.026	0.061	0.045	0.051	0.043
sll0945	glycogen synthase glgA	-0.349	-0.192	-0.151	-1.007	-0.674	0.038	0.123	-0.080	-0.353	-0.462	0.072	0.066	0.086	0.382	0.110	0.058	0.018	0.068	0.107	0.132
sll1356	glycogen phosphorylase	0.122	0.003	-0.014	-0.684	-0.823	-0.409	-0.422	-0.013	-0.513	-0.484	0.073	0.091	0.026	0.091	0.101	0.076	0.095	0.063	0.184	0.080
slr2076	60kD chaperonin groEL1	-0.230	-0.332	-0.107	-1.065	-0.515	-0.145	-0.197	-0.287	-1.144	-0.992	0.034	0.088	0.048	0.020	0.070	0.068	0.082	0.159	0.084	0.137
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	-0.100	-0.076	-0.207	0.052	0.266	-0.450	-0.317	-0.526	-0.788	-0.566	0.071	0.026	0.279	0.009	0.098	0.044	0.091	0.037	0.395	0.045
sll0947	light repressed protein A homolog lrtA	0.130	0.043	0.020	-1.023	-0.772	-0.896	-0.726	-0.540	-1.037	-0.520	0.108	0.068	0.240	0.031	0.126	0.082	0.242	0.023	0.515	0.155
sll1626	LexA repressor	-0.227	-0.181	-0.210	-0.102	-0.269	-0.321	-0.244	-0.332	-0.544	-0.548	0.123	0.052	0.098	0.005	0.104	0.116	0.108	0.092	0.087	0.082

ORF	Protein Name	MS/MS	10% Pi Log Ratio	10% Pi SD	3% Pi Log Ratio	3% Pi SD	0.3% Pi Log Ratio	0.3% Pi SD
slr0009	ribulose bisphosphate carboxylase large subunit rbcL	36	0.448	0.064	0.505	0.075	0.529	0.057
slr0012	ribulose bisphosphate carboxylase small subunit rbcS	8	0.189	0.114	0.231	0.096	0.232	0.107
slr0783	triosephosphate isomerase tpi	2	0.578	0.059	0.530	0.056	0.494	0.015
Energy Metabolism								
sll1226	hydrogenase subunit of the bidirectional hydrogenase hoxH	2	0.129	0.050	0.379	0.044	0.309	0.054
sll1326	ATP synthase alpha chain atpA	8	0.155	0.059	0.233	0.063	0.534	0.052
slr1329	ATP synthase beta subunit atpB	6	0.041	0.086	0.104	0.088	0.380	0.101
slr1622	soluble inorganic pyrophosphatase ppa	2	-0.346	0.062	-0.164	0.026	-0.145	0.007
Folding, Sorting, Degradation								
sll0170	DnaK protein 2, heat shock protein 70, molecular chaperone dnaK2	11	0.215	0.099	0.298	0.099	0.396	0.118
sll0416	60 kDa chaperonin 2, GroEL2, molecular chaperone groEL-2	12	0.165	0.088	0.373	0.076	0.493	0.071
sll0430	HtpG, heat shock protein 90, molecular chaperone htpG	2	0.155	0.045	0.490	0.162	0.584	0.253
slr0164	ATP-dependent Clp protease proteolytic subunit clpP4	2	0.309	0.050	0.712	0.031	0.876	0.021
slr0165	ATP-dependent Clp protease proteolytic subunit clpP3	2	0.053	0.210	0.338	0.255	0.615	0.292
slr0623	thioredoxin trxA	6	0.344	0.125	0.348	0.126	0.435	0.127
slr1139	thioredoxin trxB	2	0.031	0.066	0.099	0.082	0.111	0.083
slr1761	FKBP-type peptidyl-prolyl cis-trans isomerase, periplasmic protein	2	0.342	0.329	0.406	0.340	0.430	0.317
slr2075	10kD chaperonin groES	7	0.173	0.077	0.267	0.098	0.433	0.089
slr2076	60kD chaperonin groEL1	18	0.193	0.047	0.493	0.046	0.616	0.045
Glycolysis								
sll0018	fructose-bisphosphate aldolase, class II fbaA, fda	2	0.328	0.101	0.685	0.200	0.519	0.318
sll0395	phosphoglycerate mutase	2	-0.006	0.105	0.204	0.106	0.346	0.195
sll0593	glucokinase	2	0.040	0.175	0.029	0.150	-0.004	0.123
sll1342	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase gap2	9	0.212	0.101	0.345	0.117	0.327	0.122
slr0394	phosphoglycerate kinase pgk	11	0.342	0.070	0.175	0.072	0.028	0.092
slr0752	enolase	5	0.806	0.177	1.173	0.178	1.170	0.181
slr0884	glyceraldehyde 3-phosphate dehydrogenase 1 NAD+ gap1	5	0.234	0.051	0.656	0.058	0.671	0.062
slr1349	glucose-6-phosphate isomerase	8	-0.035	0.073	0.081	0.062	0.141	0.061
slr2094	fructose-1,6-/sedoheptulose-1,7-bisphosphatase fbp1	8	0.382	0.124	0.373	0.125	0.298	0.132
Hypothetical Proteins								
sll0051	hypothetical protein	2	-0.176	0.047	-0.360	0.032	-0.451	0.080
sll0230	hypothetical protein	3	0.395	0.171	0.493	0.164	0.447	0.157

ORF	Protein Name	MS/MS	10% Pi Log Ratio	10% Pi SD	3% Pi Log Ratio	3% Pi SD	0.3% Pi Log Ratio	0.3% Pi SD
slr1171	glutathione peroxidase-like NADPH peroxidase,glutathione peroxidase gpx1	3	-0.213	0.012	0.000	0.010	0.111	0.019
Metabolism of Cofactors and Vitamins								
sll1536	molybdopterin biosynthesis MoeB protein moeB	2	-0.087	0.013	0.163	0.033	0.451	0.024
Nucleotide Metabolism								
sll1018	dihydroorotase pyrC	2	-0.489	0.205	-0.472	0.409	-0.480	0.246
sll1815	adenylate kinase adk	7	0.191	0.047	0.097	0.077	0.306	0.078
sll1852	nucleoside diphosphate kinase	4	-0.047	0.076	-0.006	0.076	0.092	0.050
slr1722	inosine-5'-monophosphate dehydrogenase	4	0.352	0.171	0.382	0.210	0.424	0.201
Other								
sll0617	plasma membrane protein essential for thylakoid formation vipp1	3	-0.254	0.192	-0.169	0.166	-0.175	0.141
sll0654	alkaline phosphatase	7	0.200	0.099	0.326	0.087	0.356	0.102
sll0656	extracellular nuclease nucH	8	0.455	0.162	0.618	0.187	0.642	0.192
sll1009	iron-regulated protein frpC	2	0.449	0.001	0.516	0.006	0.453	0.019
sll1284	esterase	3	0.127	0.056	0.044	0.065	0.064	0.074
sll1559	soluble hydrogenase 42 kD subunit	6	0.142	0.124	0.081	0.139	0.039	0.141
sll1633	cell division protein FtsZ ftsZ	3	0.005	0.119	0.133	0.126	0.280	0.120
sll1825	aklaviketone reductase	4	0.178	0.058	0.390	0.100	0.371	0.089
slr1198	antioxidant protein	30	0.296	0.049	0.443	0.054	0.553	0.042
slr1516	superoxide dismutase sodB	12	-0.048	0.046	0.160	0.052	0.380	0.068
slr1719	DrgA protein homolog	14	0.543	0.057	0.712	0.060	0.780	0.059
slr1853	carboxymuconolactone decarboxylase	2	0.075	0.450	0.509	0.423	0.689	0.287
slr1924	D-alanyl-D-alanine carboxypeptidase, periplasmic protein	2	0.242	0.063	0.387	0.016	-0.134	0.070
slr1963	water-soluble carotenoid protein	5	0.082	0.101	0.197	0.084	0.231	0.105
Pentose Phosphate Pathway								
sll0329	6-phosphogluconate dehydrogenase	12	0.176	0.063	0.361	0.084	0.342	0.084
sll1070	transketolase	25	0.310	0.105	0.333	0.112	0.326	0.112
sll1479	6-phosphogluconolactonase	4	-0.049	0.094	0.044	0.075	0.143	0.104
sll0807	pentose-5-phosphate-3-epimerase rpe	5	0.684	0.068	0.669	0.066	0.738	0.067
slr0453	putative phosphoketolase	10	0.148	0.055	0.366	0.046	0.326	0.050
slr1734	glucose 6-phosphate dehydrogenase assembly protein opcA	2	0.499	0.127	0.763	0.115	0.820	0.136
slr1793	transaldolase	5	0.527	0.086	0.755	0.098	0.792	0.094
slr1843	glucose 6-phosphate dehydrogenase zwf	3	-0.161	0.041	-0.066	0.043	-0.081	0.038

ORF	Protein Name	MS/ MS	10% Pi Log Ratio	10% Pi SD	3% Pi Log Ratio	3% Pi SD	0.3% Pi Log Ratio	0.3% Pi SD
slr1856	phosphoprotein substrate of icfG gene cluster	3	-0.376	0.013	-0.233	0.067	0.016	0.063
slr1859	phosphoprotein substrate of icfG gene cluster	4	-0.009	0.017	0.233	0.007	0.449	0.013
Translation								
sll0227	peptidyl-prolyl cis-trans isomerase B, periplasmic protein ppiB	2	-0.019	0.021	0.265	0.010	-0.048	0.010
sll1099	elongation factor Tu tufA	19	0.051	0.065	0.124	0.079	0.105	0.081
sll1261	elongation factor TS tsf	2	0.331	0.180	0.476	0.142	0.728	0.092
sll1425	proline-tRNA ligase proS	2	-0.054	0.168	0.185	0.065	0.260	0.099
sll1746	50S ribosomal protein L12 rpl12	4	0.279	0.189	0.661	0.199	1.199	0.192
slr0434	elongation factor P efp	4	-0.165	0.033	-0.098	0.028	0.075	0.044
slr1105	elongation factor EF-G fus	2	-0.047	0.179	-0.018	0.149	0.142	0.149
slr1463	elongation factor EF-G fus	2	0.158	0.134	0.164	0.138	0.246	0.171
ssl3436	50S ribosomal protein L29 rpl29	2	0.077	0.056	0.275	0.047	0.747	0.010
Transport and Binding								
sll0680	phosphate-binding periplasmic protein precursor pstS1	13	-0.270	0.120	-0.085	0.141	-0.277	0.131
sll1341	bacterioferritin brfA	6	0.010	0.033	0.076	0.050	-0.058	0.044
sll1450	nitrate/nitrite transport system substrate-binding protein nrtA	4	-0.089	0.042	-0.188	0.036	-0.301	0.069
slr0447	periplasmic protein, ABC-type urea transport system substrate-binding protein urtA	9	-0.360	0.062	-0.263	0.061	-0.245	0.059
slr0513	iron transport system substrate-binding protein, periplasmic protein	10	-0.349	0.138	-0.336	0.146	-0.887	0.157
slr1234	protein kinase C inhibitor	4	0.173	0.062	0.213	0.059	0.387	0.073
slr1247	phosphate-binding periplasmic protein precursor pstS2	7	-0.122	0.145	0.260	0.180	0.249	0.166
slr1295	iron transport system substrate binding protein futA1	5	-0.368	0.091	-0.634	0.104	-0.824	0.094
slr1452	sulfate transport system substrate-binding protein sbpA	10	-1.074	0.068	-1.055	0.065	-0.995	0.072
slr1890	bacterioferritin brfB	4	0.081	0.095	-0.042	0.105	-0.362	0.113
ssr2857	mercuric transport protein periplasmic component precursor atx1	2	0.602	0.088	0.761	0.077	0.872	0.098
Unknown Proteins								
sll1380	periplasmic protein, function unknown	2	-0.099	0.053	-0.102	0.045	-0.255	0.023
sll1837	periplasmic protein, function unknown	2	0.268	0.089	0.503	0.008	-0.021	0.092
slr2144	periplasmic protein, function unknown	3	0.671	0.205	0.846	0.217	1.168	0.210

Note: Hypothetical proteins with close similarity to another organism(s) with known functional group(s) were in parentheses.