The University Of Sheffield.

Proteomics: A tool for understanding adaptation in environmentally significant microorganisms

Jagroop Pandhal

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

Biological and Environmental Systems Group
Department of Chemical and Process Engineering
University of Sheffield

June 2008

Declaration

This is a declaration to state that this thesis is an account of the author's work which was conducted at the University Of Sheffield, U.K. This work has not been submitted for any other degree or qualification.

Acknowledgements

To be given the opportunity to conduct research in an area you are passionate about is a fantastic gift, and for this I thank my sponsors EPSRC and supervisor Dr. Catherine Biggs. I would also like to thank Dr. Biggs and my second supervisor, Professor Phillip Wright, for all their support and guidance throughout the three years.

My parents have always supported me in my sometimes endless mission to pursue a science-based career, and I thank them deeply for their patience, love and trust.

I would also like to thank my research group, and in particular, Ow Saw Yen, James Williamson and Mark Scaife, without who the long days in the laboratory and office would have certainly been less exciting, and I thoroughly enjoyed our scientific discussions, debates and arguments.

Special thanks must also be given to Chee Sian Gan, for his generosity, time and advice, and Martin Barrios-Llerena for his technical help, particularly when I was battling with the mass spectrometer. I also extend my gratitude to Dr. Alireza Fazeli for his guidance during my real-time PCR experiments, Bashir Ashhuby for the provision of *Euhalothece* cells and Agilent for the use of their equipment for the iTRAQ study in Chapter 3.

It is rarely feasible to achieve something great without the help of someone special throughout the journey, and therefore my very heartfelt thanks go to Emma Reynolds. For all those times when the stress was immense, workload overwhelming and when I needed inspiration during difficult times. I would never have got this far without her love and support, nice one Em.

List of Publications

Journal papers

- Pandhal J., Snijders A.P.L., Wright P.C. Biggs C.A. (2008) "A cross-species quantitative proteomic study of salt adaptation in a halotolerant environmental isolate using ^{15}N Metabolic Labelling." *Proteomics* 8 (11): 2266-84.

- Pandhal J., Wright P.C. Biggs C.A. (2008) "Proteomics with a pinch of salt: a cyanobacteria perspective." *Saline Systems* 4 (1)

- Pandhal, J., Wright, P.C., and Biggs, C.A. (2007) "A quantitative proteomic analysis of light adaptation in a globally significant marine cyanobacterium *Prochlorococcus marinus* MED4." *Journal of Proteome Research.*, 6 (3): 996 – 1005.

Conferences

- "Proteomics: A tool for understanding adaptation in halotolerant organisms and halophiles" Presented at *Halophiles*, Colchester, UK 2nd-6th September 2007. (Oral Presentation)

- "Understanding the salt response in *Euhalothece*, a halotolerant cyanobacterium." *British Society for Proteome Research* (BSPR), Cambridge, UK, 10-13th July 2006. (Poster presentation)

- "A quantitative proteomic study of the salt response in an unsequenced extremophile using *in vivo* metabolic labelling." *American Society of Microbiology 106th General Meeting*, Orlando, USA 21st-25th May 2006. (Poster presentation)

- "A quantitative proteomic analysis of environmental adaptation in a globally significant marine cyanobacterium." *American Society of Microbiology 106th General Meeting*, Orlando, USA 21st-25th May 2006. (Poster presentation)

Contents

Abbreviations

1DE	one-dimensional electrophoresis
2DE	two-dimensional electrophoresis
AB	Applied Biosystems
ABC	ATP-binding cassette
AQUA	absolute quantitation
ATP	adenosine triphosphate
BC	before christ
BLAST	Basic Local Alignment Search Tool
bp	base pairs
cDNA	complementary DNA
CHAPS	3-(3-cholamidopropyl)dimethylamonio)-1-propanesulfonate
chl	chlorophyll
CHO	carbohydrate
CID	collision induced dissociation
COG's	Clusters of Orthologous Groups
cps	counts per second
CV	coefficient of variation
Da	Dalton
DIGE	difference gel electrophoresis
DTT	dithiothreitol
emPAI	exponentially modified protein abundance index
ESI-qQ-TOF-MS/MS	electrospray ionisation quadrupole time-of-flight tandem mass spectrometer
GG	glucosylglycerol
GOLD	Genomes Online Database
HL	high light
HLIP	high-light inducible protein
HPLC	high performance liquid chromatography
iCAT	isotope coded affinity tags
ICP-MS	inductively coupled plasma mass spectrometry
IDA	information dependent acquisition
IEF	isoelectric focusing
IPG	immobilised pH gradient
iTRAQ	isobaric tags for relative and absolute quantitation
JGI	Joint Genome Institute
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	liquid chromatography
LL	low light
M	molar
m/z	mass /charge
MALDI	matrix-assisted laser desorption ionisation
min	minute
ml	millilitres
ML	medium light
MMTS	methyl methanethiosulphonate
mM	millimolar
MOPS	3-(N-morpholino)propane sulfonic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
Mw	molecular weight
NCBI	National Centre for Biotechnology Information
OD	optical density

ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAI	protein abundance index
PCC	Pasteur Culture Collection
PCR	polymerase chain reaction
pI	isoelectric point
PIC	protease inhibitor cocktail
PMF	peptide mass fingerprinting
ppm	parts per million
PPP	pentose phosphate pathway
PSI	photosystem I
PSII	photosystem II
PTM	post-translational modification
PVPP	polyvinyl polypyrrolidone
rf	radio frequency
ROS	reactive oxygen species
RP	reverse phase
rpm	revolutions per minute
RT-qPCR	real-time quantitative polymerase chain reaction
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
RV	relative variance
SCX	strong cation exchange
SD	standard deviation
SDS	sodium doecyl sulphate
SEC	size exclusion chromatography
SEM	scanning electron microscope
SILAC	stable isotopes labelling with amino acids in cell culture
TAE	tris acetate ethylenediaminetetraacetic acid
TCA	trichloroacetic acid
TEAB	triethylammonium bicarbonate
TEM	transmission electron microscope
TEMED	N'N'N'N'-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
WWW	world wide web

Summary

Understanding how microorganisms adapt to adverse environments is beneficial for a variety of reasons depending on the conditions under investigation. In this thesis, the ancient life forms, cyanobacteria, form the basis of the research. Research into salt tolerance is useful because agricultural practices are leading to increasing soil salinity, which prevent growth of susceptible crop species. In addition, elucidating adaptive mechanisms in possibly Earth's most abundant photosynthetic microorganism is important, due to its significant role in nutrient cycling and climate control. Proteins are synthesised in cells in response to environmental perturbations and are therefore essential for adaptation, and in this thesis proteomics is the principal method applied. Less pronounced protein expression changes are characteristic of long-term adaptive responses as opposed to immediate shock responses, and therefore techniques were developed and applied which could quantify with high accuracy (25 to 50% changes). Interpreting these changes into biological function required analysis of a sufficient fraction of the proteome, which was achieved through a variety of protein and peptide fractionation techniques. Protein separation through traditional 2DE was also optimised to overcome the detrimental effects of salt. These methods were applied to a well-studied model cyanobacterium, *Synechocystis* sp. PCC6803, which allowed a comparison to be made with present knowledge. Subsequently, an approach to investigate the unsequenced cyanobacterium, *Euhalothece* sp. BAA001, was developed and revealed an unusual response to low salt. A direct comparison of these two cyanobacteria was then made relying on their high degree of proteome homology, which highlighted their shared and contrasting tolerance mechanisms. Finally, the methods were extended to characterise the proteome of *Prochlorococcus marinus* MED4, and increase our understanding of adaptation to varying light intensities characteristic of its oceanic niche.

Chapter 1: Introduction and Hypothesis Statement

1.1 Introduction

This introductory chapter gives a general background to the research described in this thesis and how the findings will contribute to a wider setting. It describes the ultimate goal of the research and states the specific objectives which need to be addressed in order to achieve this. By defining the scope of the research and an outline of the approach implemented, the manner in which the aims are achieved is described. Finally, a summary of the contents in each chapter, and how they fit together, is presented.

1.2 Microbial ecology with proteomics

When attempting to increase our understanding of environmental factors which impact on human life, investigating microorganisms which live in such surroundings can be particularly fruitful. Microorganisms often play vital roles in the complex interactions which govern the state of an environment, and this relationship is investigated in the field of microbial ecology (or environmental microbiology). Environments, which we perceive to be intolerable, are mostly characterised by low species diversity and dominated by microorganisms, described as extremophiles [1].

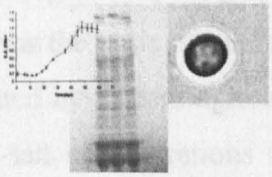
The principal step in a traditional microbial ecology workflow, where the aim is to investigate survival strategies, is to undertake growth studies. Generating pure cultures presents the first challenge, and necessitates a certain degree of imitation of the natural environment, for example, in terms of nutrient content. Further microbiological studies (including characterisation based on morphology, size etc) as well as biochemical studies are further complemented by functional genomics approaches. Discovering which genes play roles in essential cellular processes through techniques such as gene deletions, expression profiling, genome-wide microarrays and systematic genetics, have provided valuable insights into how microorganisms adapt to their natural environments [2-4]. However, in this post-genomic era, the relatively new field of proteomics has presented a new series of opportunities to uncover underlying mechanisms of survival. Proteomics can be considered as a branch of molecular biology, where the focus is shifted to proteins

encoded by the genome, including analysis of protein expression, localisation, and function [5]. By concentrating on the proteins which are synthesised, the major functional entities of living cells are under investigation. Figure 1.1 shows how incorporating global proteomics into a traditional microbial ecology workflow to understand environmental adaptation in organisms, can reveal areas for more directed research using functional genomics or further functional proteomics approaches.

Isolate microorganism of interest from environment

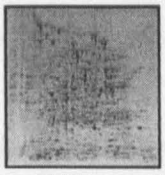
Dilution series,
agar plate streaking

Pure culture in laboratory



16S rRNA identification
Growth studies
Morphology

Characterisation



Identify which gene
products are important

Global proteome differential expression analysis

Functional genomics
e.g. directed mutagenesis, gene knockouts.

Functional proteomics
e.g. protein-protein interaction networks, biochemical pathways, post-translational modifications.

Figure 1.1: A microbial ecology experimental overview of studying environmental adaptation in organisms, with the incorporation of global proteomics.

In this thesis, proteomics will be used to investigate environmental adaptation in three species of cyanobacteria. The strength of this field in uncovering adaptive mechanisms as well as highlighting areas for further research will be demonstrated.

1.3 Environments and related organisms

Two environmental factors are under investigation in this thesis, salinity and light, using three biological systems. Studying the proteome in these organisms requires similar as well as different approaches. This is due to protein identification issues, nature of cell culturing and growth conditions applied. Consequently, the work described in this thesis uses proteomics to increase overall knowledge of cellular and molecular mechanisms in different conditions and biological systems.

1.3.1 Adaptation to salt

Understanding the response of organisms to a saline environment is important from several perspectives, including food spoilage [6] and agricultural problems [7], and in this thesis the latter forms the basis of the research. Salt overloading in soils is a global problem. Exacerbated by current agricultural processes, the ecological risks associated with elevated salt concentrations are predicted to worsen during this century [7-9]. Understanding how crop plants are affected by high salt conditions is therefore an extremely favourable goal. There are a diverse array of studies making direct use of higher plants which are aimed at understanding how they are susceptible to these adverse conditions, and how several species have developed salt tolerance mechanisms [10-14]. However, working directly with plants has disadvantages, including high cost of goods, long life cycle generation times and limited control over genetic manipulations. Therefore, the close relationship between photosynthetic prokaryotes, cyanobacteria, and plant chloroplasts has been exploited, with the model organism *Synechocystis sp.* PCC6803 (henceforth referred to as *Synechocystis* unless otherwise stated) dominating this area of interest.

In the main part of the thesis, the close relationship between cyanobacteria and higher plants is exploited by investigating the salt response, in two contrasting

species, using a proteomics approach. Proteomic techniques were recently applied to the model cyanobacterium *Synechocystis* and have resulted in many interesting findings (as reviewed in Chapter 2 and experimentally tested in Chapter 3). However, the strength of proteomics to reveal adaptation mechanisms has been limited to only semi-quantitative methods, and protein fractionation techniques which do not present a global picture of the proteome. Whilst generating useful information, conclusions are very preliminary in nature due to the relatively few proteins identified which are involved in specific metabolic processes, for example, generating reductive power through the pentose phosphate pathway (PPP). In order to address these limitations, advances in proteomic methodologies and tools are put into practice to increase the global understanding of the salt response in this model organism, whilst concentrating on adaptation rather than immediate shock responses.

The development and application of different but complementary proteomics tools will subsequently be used to understand the salt response in a more challenging system, an unsequenced environmental isolate, which is likely to harbour alternative but equally important adaptive mechanisms. In conjunction, an assessment of the strength of these techniques to elucidate environmental adaptation will be made. The isolate used in this study was identified as a cyanobacterium belonging to the *Halothece* cluster, and named *Euhalothece* sp. BAA001 (henceforth referred to as *Euhalothece*). Isolated from a salt lake in the Sahara, this organism has no sequenced genome and therefore presents a challenge in studying its proteome, and together with the detrimental amounts of salt conditions involved, makes it a particularly unattractive candidate for this type of analysis. The last decade has seen major interest in survival mechanisms of extreme halophiles, which require high internal salt concentrations for their metabolism. However, *Euhalothece* is extremely halotolerant, and adapting to such large variances in salt concentrations appears to be a much more difficult feat, and therefore represents an area which requires further work.

1.3.2 Adaptation to light

Less than 20 years ago, a cyanobacterium, *Prochlorococcus marinus*, was discovered [15]. This marine cyanobacterium was found to be ubiquitous in the oceans in the latitude band of 40°N to 40°S [16]. Its abundance as well as the overall size of the ecological niche, have led scientists to believe that it is the most abundant photosynthetic organism on earth [16]. In other words, this microbe potentially plays more of a significant role in the state of our atmosphere, biogeochemical cycles and nutrient cycling than any other living organism. Furthermore, its impact on global climate change is extremely relevant in these times. Its environment is characterised by very low nutrient levels, relatively high water temperatures and starkly varying light availability [16]. Understanding how this organism adapts to this environment would be an important step in elucidating its overall role in the global processes mentioned above. With rapid advances in computer modelling to predict changes in the earth's climate, an understanding of the fundamental interactions between these microorganisms and their environment is imperative. The genomes of twelve different strains of *P. marinus* have been fully sequenced (January 2008, GOLD version 2.0: http://www.genomesonline.org/gold.cgi) [17], making this organism a more 'suitable' candidate for proteomic studies than *Euhalothece*. However, difficulties associated with culturing this microorganism have been experienced, and before this thesis no previous proteomic studies have been conducted, despite its global significance. Therefore, in this thesis, proteomic tools developed and tested in early chapters are applied to assist in discovering how this organism has evolved as different ecotypes depending on the environment it resides in.

1.4 Aims and scope

In this study, global proteomics will be used as a tool to identify and quantitatively investigate important proteins, which play a key role in environmental adaptation. Subsequently, these findings present areas for further research in complementary fields.

Specifically, the aims of thesis are to:

- *Implement global proteomic tools which can detect with high confidence, proteins differentially expressed by less than 2-fold to uncover environmental adaptation mechanisms*
- *Survey the proteome of an unsequenced environmental isolate*
- *Develop a method to apply two-dimensional electrophoresis (2DE) to high salt samples*
- *Identify possible strategies as to how* Synechocystis *adapts to a hypersaline environment*
- *Identify possible strategies as to how an unsequenced environmental isolate,* Euhalothece, *adapts to large variances in salt concentrations*
- *Identify possible strategies as to how* Prochlorococcus marinus *MED4 adapts to varying light intensities in an oligotrophic ocean*

The techniques implemented are used to provide a further comparison of adaptive mechanisms in evolutionary diverged species, *Synechocystis* and *Euhalothece*.

The intention of this thesis is not to present evidence that proteomics is an exclusive field which can potentially reveal all the mechanisms organisms utilise to adapt to environmental conditions. Rather it aims to identify parts of the cell machinery at the molecular level which play a significant role. Any cell processes which are identified can be labelled as networks and these can be categorised as subnetworks of a much larger and complex global network, depending on the depth of information generated. Whilst the majority of this study is based on protein abundance levels, it is useful to bear in mind that some proteins involved in adaptive mechanisms can be regulated by an array of native cell control systems, including phosphorylations, dephosphorylations, activations, deactivations as well as interactions with other proteins and nucleic acids [18]. The role of differential expression proteomics [19, 20] (applied here) in systems level biology is now being realised as an indispensable tool for identifying cellular responses, but its findings ultimately need to be combined with associated studies, for example, transcriptomics or metabolomics, for a comprehensive and complete understanding [21, 22]. Therefore any proposed mechanisms of adaptation are compared to current knowledge on cell physiology.

1.5 Thesis Outline

An outline of the thesis structure with investigations carried out in the chapters is given in Figure 1.2. An overview of each chapter is summarised below.

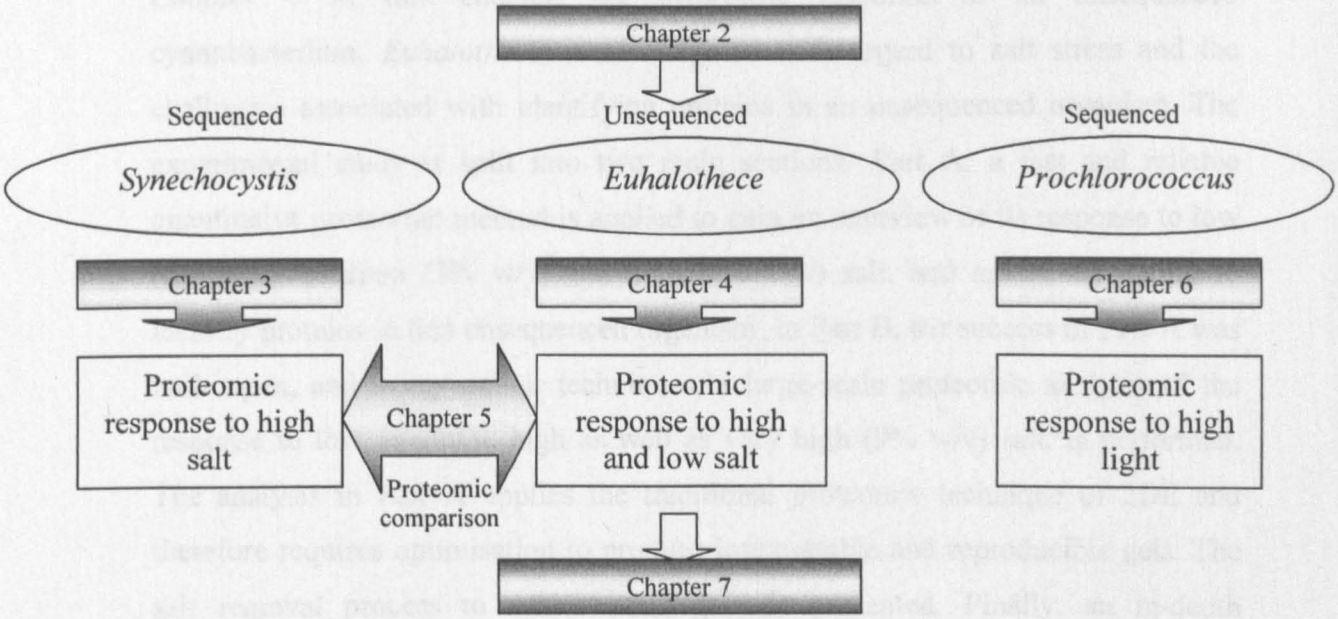


Figure 1.2: An overview of the format of studies in this thesis.

Chapter 2 The principal aim of this chapter is to introduce the reasoning behind the development of the research hypothesis. This is achieved by presenting the background to the main field of proteomics and in particular the techniques which can be applied to address the aims stated in section 1.4. An introduction to the three organisms is then followed by an up-to-date review of research into the relevant environmental perturbations (salt and light stress).

Chapter 3 The model photosynthetic organism for proteomic studies on salt stress is *Synechocystis*, and it forms the basis of research in this chapter. A global proteome survey of how this cyanobacterium adapts to high salt (6% w/v) is given. This chapter is presented here because firstly, it highlights the potential of the techniques employed, for example, the quantitation accuracy and proteome coverage achieved. Secondly, the way in which data is interpreted and related to previous findings is revealed, and finally, the benefits and possibilities of applying the field of proteomics to answering microbial ecology questions are demonstrated. Incorporating quantitative real time polymerase chain reaction (RT-qPCR) here,

also gives a perspective on what can be accomplished with differential protein abundance analysis.

Chapter 4 In this chapter, the proteomic response of an unsequenced cyanobacterium, *Euhalothece*, is investigated with regard to salt stress and the challenges associated with identifying proteins in an unsequenced organism. The experimental study is split into two main sections- Part A: a fast and reliable quantitative proteomic method is applied to gain an overview of its response to low (0% w/v), medium (3% w/v) and high (6% w/v) salt, and assess the ability to identify proteins in this unsequenced organism. In Part B, the success of Part A was built upon, and using similar techniques, a large-scale proteomic analysis of the response to low, medium, high as well as very high (9% w/v) salt, is performed. The analysis in Part A applies the traditional proteomic technique of 2DE and therefore requires optimisation to produce interpretable and reproducible gels. The salt removal process to achieve this goal is presented. Finally, an in-depth discussion of how *Euhalothece* adapts to its changing extreme environment is given with relation to features which make this organism a potentially novel species in its cluster.

Chapter 5 The last two chapters give an indication of how two evolutionary diverged species of cyanobacteria adapt to saline environments and this chapter exploits the closeness of their proteome sequences to make a direct comparison of protein abundances. Similarities and differences are interpreted with respect to protein level responses to salt (0, 3 and 6% w/v NaCl).

Chapter 6 The development of proteomic techniques throughout the thesis is implemented in this chapter to understand how a globally significant marine cyanobacterium adapts to its environment, characterised by ever changing light intensities. Unlike other chapters, the proteome of this organism will be characterised with respect to protein properties and statistics, with particular reference to previously published whole proteome descriptions using gel-free (shotgun) tools in cyanobacteria *Anabaena variabilis* ATCC 29413 and *Synechocystis*.

Chapter 7 The findings and resulting conclusions in the thesis are summarised here, tying together all the studies conducted along with suggestions for future research.

Chapter 2: General Background and Review [1]

[1] This work (in part) has been published in *Saline Systems* (2008), **4**: 1

2.1 Introduction

The 'golden era' for microbiology occurred during the end of the 19th century and into the 20th century, when research demonstrated that microorganisms had the capability to cause disease in humans, plants and animals. It also became apparent that microorganisms played an extraordinary role in the environment, dictating chemical changes in the atmosphere, soil and water. During this time, a platform was established for furthering our understanding of these phenomena, creating scientific disciplines such as immunology, genetics and molecular biology, as well as branches of microbiology such as food and environmental microbiology. Environmental microbiology (also referred to as microbial ecology) is centred on understanding environmental microorganisms which play significant roles in human affairs. In this thesis, the ancient microbes, cyanobacteria form the basis of the research. These fascinating bacteria have an extensive fossil record dating back 2.5 billion years and are responsible for the very oxygen we breathe today. A recent study claimed to have found evidence of cyanobacteria-like creatures as far back as 3.5 billion years in Archean rocks of Western Australia [23], however, the identity of these microorganisms remains somewhat controversial due to poor fossil preservation. Through the process of evolution, cyanobacteria have adapted to an array of varying environments, some extreme, presenting them as a major source of interest in many scientific fields.

Understanding environmentally significant organisms is not exclusive to the field of microbiology, and spans a plethora of subject areas. With the advent of molecular biology techniques, initial analysis of microbial morphology and biochemical characterisation has been accompanied (rather than superseded) by the study of the biology at the molecular level. Genetics is a major component of molecular biology, and has allowed in-depth analysis of microorganisms by manipulating the major molecular components i.e. DNA and RNA. With the arrival of techniques to efficiently unravel the full genetic code of organisms (Sanger's chain termination method [24]), it became possible to extrapolate gene function and the huge amounts of genomic information generated has fuelled major advances in understanding environmental adaptation. It is now considered relatively routine to screen the

expression of thousands of genes in one experiment using the DNA chip or microarray technique [25].

Recently, the importance of studying gene products has been recognized, and therefore we are considered to be in the post-genomic era. Looking at global protein expression in microorganisms is now a major driving force in the subject proteomics. Studying the proteome of environmentally significant microbes depends on several factors and perhaps most significantly the availability of its genetic code or extent of relatedness (or homology) to sequenced organisms. In this chapter, a general background into the field of proteomics is presented, highlighting the implementation of traditional and contemporary tools, and revealing the nature in which the resulting data can be used to infer biological adaptations in microorganisms. In this study, the application of these tools occurs in cyanobacteria, and these fascinating life-forms will be introduced in more detail. The three species investigated in this thesis share common traits but also contrast vastly, most obviously in adaptive abilities they have acquired to differing environments. The problems associated with the varying environmental conditions will be discussed from the perpspective of how far our understanding is presently placed. Finally, the way in which this study fits into this bigger picture is emphasised.

2.2 Proteomics

Studying gene products has long accompanied studying actual gene expression levels and has been performed using traditional techniques such as immunoblotting (western blotting) [26]. However, the main aim of proteomics is to determine gene and cellular function by working at the protein level on a *global* scale [19, 27], and has profited from exceptional progress in areas such as genome sequencing, bioinformatics and analytical techniques [24, 28, 29].

The proteome of an organism refers to the total set of proteins encoded by its genome (protein-coding genes) [5] and therefore the field of proteomics encompasses studying these proteins, specifically the change in abundance in cells, tissues and organelles in response to changing environmental factors. The DNA

sequence and mRNA levels (transcriptomics) give a sound foundation and valuable insight into predicting the workings of an organism, however the functional output of cells are the proteins. Furthermore, it has been demonstrated that expression patterns between mRNA and proteins levels may not correlate well, indicating gaps in our knowledge of how transcription, translation and post-translational modifications (PTM's) interact, and ultimately effect protein levels [30]. By mapping which proteins interact with each other and which proteins play pivotal roles in certain conditions, a more comprehensive understanding of survival traits can be achieved. In effect, genomic information can be decoded into a functional protein interpretation. The task is not simple, and studies using nuclear magnetic resonance imaging have shown that proteins are more active and dynamic than initially predicted, adding complexity to understanding where each protein is located in a cell, when the protein is present and for how long, and with which other proteins it is interacting [31].

Overall, proteomics has the capability to reveal information on the level of expression of proteins, protein isoforms produced from each gene, the extent to which proteins are post-translationally modified and also the cellular and sub-cellular distribution of proteins. Combining this information with transcriptomic and metabolomic (or metabolic flux) data creates a holistic approach to answering biological questions, and forms the basis of the experimental component of integrative systems biology [21, 32]. Investigating protein expression changes in response to external environmental stimuli is particularly useful in understanding adaptive mechanisms, a goal achievable by the field of differential expression proteomics [20, 27] (Figure 2.1).

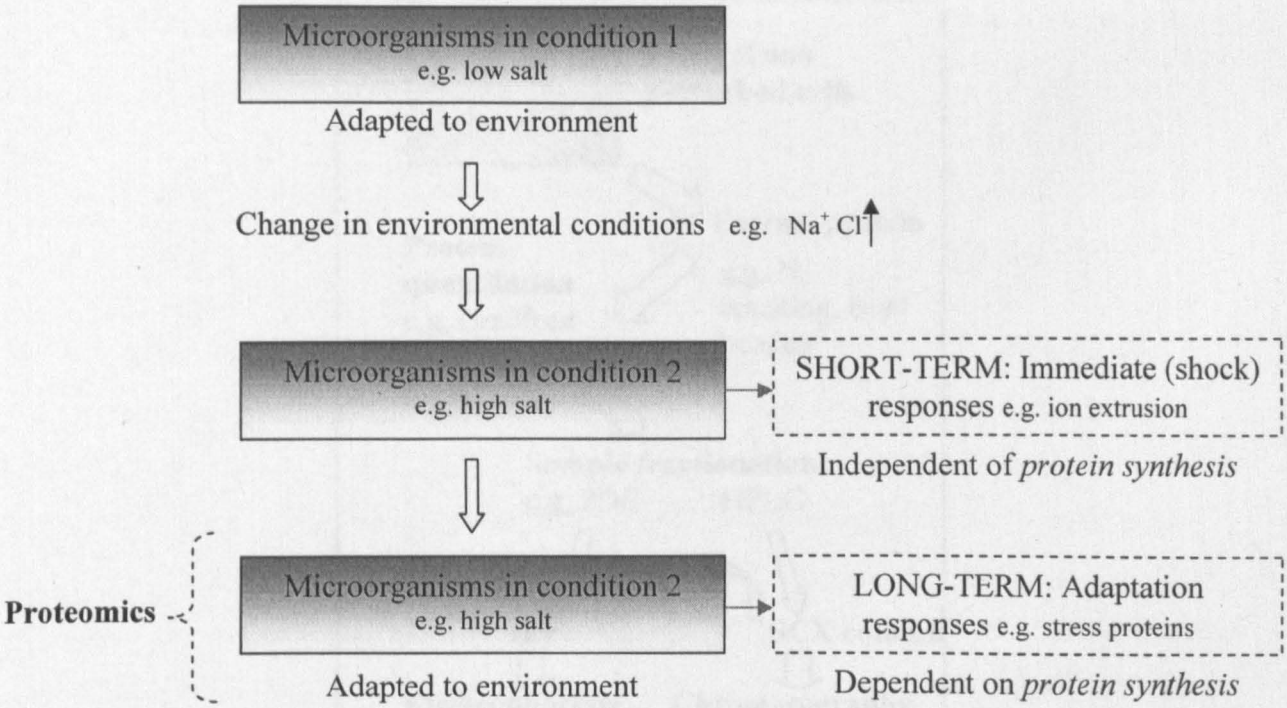


Figure 2.1: Overview of how microorganisms adapt to changes in environmental conditions (condition 1 to condition 2) from the perspective of protein synthesis-independent and protein synthesis-dependant responses. The field of proteomics is mainly concerned with the latter and is highlighted in bold text.

A proteomics experiment involves collaboration of several techniques in order to progress from extracting the proteins and ultimately to their accurate and reliable identification and quantitation. Figure 2.2 illustrates the main steps involved with two contrasting examples of protein sample fractionation.

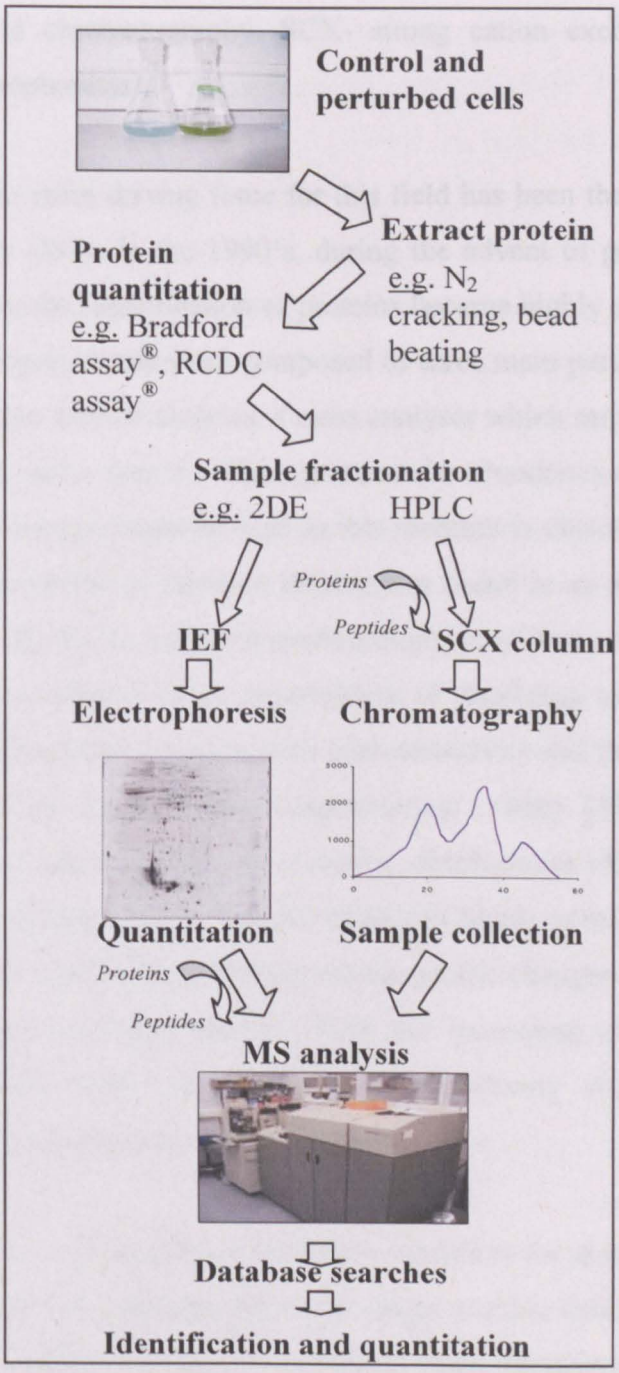


Figure 2.2: A general proteomic workflow containing main steps, cell culture, protein extraction, sample fractionation (traditional 2DE and gel-free chromatography method), mass spectrometer analysis and protein quantitation and identification. Quantitation using 2DE is densitometry-based and implemented with appropriate software (e.g. PDQUEST®, SameSpots®), but can also be undertaken on the mass spectrometer. Identification occurs with database search software (based on peptide mass fingerprinting (PMF) or sequence analysis). Proteins are commonly digested using trypsin. (IEF- isoelectric focusing, HPLC- high

performance liquid chromatography, SCX- strong cation exchange, 2DE- two-dimensional electrophoresis).

Fundamentally, the main driving force for this field has been the progress made in mass spectrometry (MS). In the 1990's, during the advent of genome sequencing and bioinformatics, the identification of proteins became highly sensitive using this instrument. Mass spectrometers are composed of three main parts; an ion source to create gaseous phase ionised analytes, a mass analyser which measures the mass-to-charge ratio (m/z), and a detector which presents the abundance of ions at each m/z value [28]. The mass spectrometer used in this research is discussed in more detail in section 2.2.3. Its ability to interpret information coded in an organism's genome has advanced rapidly due to technical breakthroughs in all three of these parts. Now the technique is coupled to cross examination of databases and has become an intrinsic part of proteomics research, with high sensitivity and throughput. Proteins can be identified by peptide mass fingerprinting (PMF) [33] or tandem MS (MS/MS) [28], and together with the technical development of analysis software, the identification of proteins has become rapid and highly sensitive [34]. This has provided a valuable tool for investigating protein profile changes in microorganisms in response to environmental stimuli. With the increasing number of genome sequencing projects being undertaken, the opportunity for making greater discoveries in the post-genomics field is apparant.

Before selecting the most applicable proteomic workflow for answering a biological question, it is useful to understand the major stages in more detail (Figure 2.2). The following sections describe sample fractionation using the popular method 2DE, as well as more recent non-gel (shotgun) methods. Protein quantitation is tied into these sections, as well as advantages, disadvantages and improvements in these approaches. An overview of MS and analysis methods for interpretation of the generated data is also discussed.

2.2.1 2DE

One of the most commonly implemented methodologies employed in proteomics uses traditional 2DE with MS [35]. In general, the workflow involves separating a

complex protein sample by isoelectric point (pI) and molecular weight (Mw), protein staining and subsequently quantitation is achieved by relative staining intensity. The technique has received a revival of late, much due to technical developments fuelled by expanding numbers of proteomic studies [36]. Recent advances in this technique include the introduction of narrow range and non-linear immobilised pH gradient (IPG) strips, which have significantly improved protein coverage [37-39]. IPG strips are prepared by co-polymerising acrylamide monomers with acrylamide derivatives which contain carboxylic and tertiary amino groups [40]. The buffering groups form a fixed pH in the strip, and upon application of an electrical current, proteins will migrate to the section of the strip (pI), where their net charge is zero. Therefore narrow range strips, for example, pI 4 to 7, allow enhanced protein separation. Automated spot excision systems and mass in-gel trypsin digestion techniques have also reduced the arduous nature of preparing samples from 2DE gels for MS analysis [41, 42], and software packages such as Samespots and Progenesis (Nonlinear Dynamics, Newcastle, U.K.) have enhanced gel analysis capabilities. These and additional methods including pre-2DE fractionation techniques, have been extensively reviewed [43-47]. Another major advantage of 2DE is the ability to detect PTM's, for example phosphorylations [48, 49]. In addition to advances in protein identification and throughput efficiency of 2DE, quantitation methodology has also improved. Fluorescent and radioactive stains can be used prior to protein separation, as well as highly sensitive post separation stains, for example, SYPRO ruby, Coomassie and silver stains [50-52].

Protein mixtures can now be multiplexed and run on the same gel in a technique called differential in-gel electrophoresis (DIGE) [53]. Through the use of multiple fluorescent dyes to label protein samples prior to 2DE, the technique allows multiple samples to be co-separated and visualized on one single gel. By using CyDyes (GE Healthcare, Buckinghamshire, U.K.), not only is the sensitivity and reproducibility enhanced, but also an internal standard is included increasing statistical confidence. Unlu *et al.* [53] initially reviewed the sensitivity and reproducibility of this method and demonstrated that significant fluorescence difference can be reliably detected at nanogram protein levels.

Quantitation can also be improved by using *in vivo* isotope labelling of proteins prior to 2DE separation. This method is particularly attractive because the isotope label is incorporated into the organisms' proteome, with no physiochemical or biological consequences. Potentially all proteins are labelled, and this technique is very applicable and compatible with accompanying techniques. The different peptide isotopes can be resolved by a mass spectrometer and this is where quantitation occurs (based on peptide peak areas) as opposed to staining intensity analysis, which is only semi-quantitative, and requires many replicate gels to be run for statistical relevance. Methods include stable isotope labelling with amino acids in cell culture (SILAC) [54] and elemental labelling [55].

There are several drawbacks of 2DE including their laborious nature and lack of automation. Proteome coverage is a significant problem and the technique struggles to resolve hydrophobic proteins as well as membrane proteins due to their poor solubility [56, 57]. It is also difficult to display proteins with extreme pI values and detect those of low abundance [47, 58, 59]. In addition, using densitometry-based proteomics to assess protein expression changes is only semi-quantitative and relatively large variations in replicate data sets mean larger expression changes are required for confidently identifying protein level changes, leading to potentially higher false negative rates. This also means less significant changes in protein expression (<2 fold) may not be detectable, and smaller changes may be expected in cells living in an adaptive state compared to short-term shocked state (condition 2 in the long-term compared to condition 2 in the short-term in Figure 2.1) [60, 61].

2.2.2 Shotgun methods

The majority of methodological and technical improvements in proteomic techniques have focused on shotgun approaches. Analysing complex peptide samples in shotgun experiments requires simplification of the mixture prior to MS analysis. Reducing complexity of a protein sample prior to MS increases the probability of identifying low abundance candidates and leads to higher confidence in protein identification. Proteins can be separated by characteristics such as size, hydrophobicity or charge, using size exclusion chromatography (SEC) [62], reverse-phase (RP) chromatography [63] and weak/strong anion exchange

chromatography (WAX/SAX) [64, 65], respectively. To provide peak capacity and detection of low abundance proteins, gel-based fractionation techniques have been explored as well as various combinations of protein and peptide separation schemes such as multi-dimensional liquid chromatography [66, 67]. An increasingly popular combination includes strong cation exchange (SCX) with RP separation [68, 69], as well as liquid phase IEF [70] and capillary electrophoresis [71]. A comparative study was conducted by Gan *et al.* [67] using a variety of protein and peptide prefractionation methods. A combination of six fractionation workflows employing techniques which separate proteins via IEF, 1-D PAGE, WAX chromatography as well as peptide separation using IEF or SCX, all prior to RP multi-dimensional liquid chromatography and MS analysis, were compared. Twenty percent of a bacterial proteome was reliably identified, and it was concluded that using two or more combinations of methods would ultimately increase proteome coverage [67].

As mentioned in the previous section, resolving membrane proteins via 2DE remains a considerable challenge due to poor solubility. However, strides have been made to overcome this limitation, ensuring that an increase in the solubility of membrane proteins doesn't compromise compatibility with downstream processes [56, 72]. This can be achieved through the use of high percentage organic solvents and acids. Moreover, the use of strong zwitterionic and non-ionic detergents have been shown to improve the solubilisation of membrane proteins from cyanobacteria [73], where 11.8% of total proteins (254) identified using 2DE, were predicted membrane proteins.

In conjunction with technical advances in this area, the need to add a quantitative dimension into large-scale proteomic studies has been recognised. The use of stable isotopes in shotgun experiments relies on the mass difference separation power of mass spectrometers, which are able to distinguish between pairs of chemically identical analytes of different isotope composition [74]. The resulting signal intensities can be compared between the isotopes and converted into an abundance ratio. There are three ways in which isotope tags are added to proteins. In *metabolic labelling*, heavy elements in amino acids or often in salts are added. ^{15}N and ^{13}C labelling in salts and in SILAC are popular examples [54, 55, 75]. Potentially all peptides can be labelled in this manner and it is the simplest method based on

applicability and compatibility with accompanying techniques. Isotope tags can be added *enzymatically* by transferring ^{18}O to peptides from water [76]. Finally, *chemical reactions* are used where reagents are tagged to the protein or peptides. Examples include isotope coded affinity tags (ICAT) [77, 78], mass coded abundance tagging [79], the AQUA strategy [80] and isobaric tags for relative and absolute quantitation (iTRAQ) [81-84].

The iTRAQ system is implemented in this thesis and has the added advantage of allowing up to four (and very recently eight [85]) different conditions to be analysed in one experiment. The technique is based upon protein digestion to generate peptides, and then chemically tagging the N-terminus of these peptides. The labelled samples (two to eight) are then combined, fractionated by nanoLC and analysed by MS/MS. A database search of the peptide fragmentation data, results in the identification of the labelled peptides, and hence the corresponding proteins. Fragmentation of the tag attached to the peptides generates a low molecular mass reporter ion, which is unique to the tag used to label each of the digests. By measuring the intensity of these reporter ions, the relative quantification of the peptides in each digest and hence the proteins from where they originate, is possible (Figure 2.3). In Figure 2.3C it can be extrapolated that the peptide labelled with the 115.1 reporter tag is the least abundant.

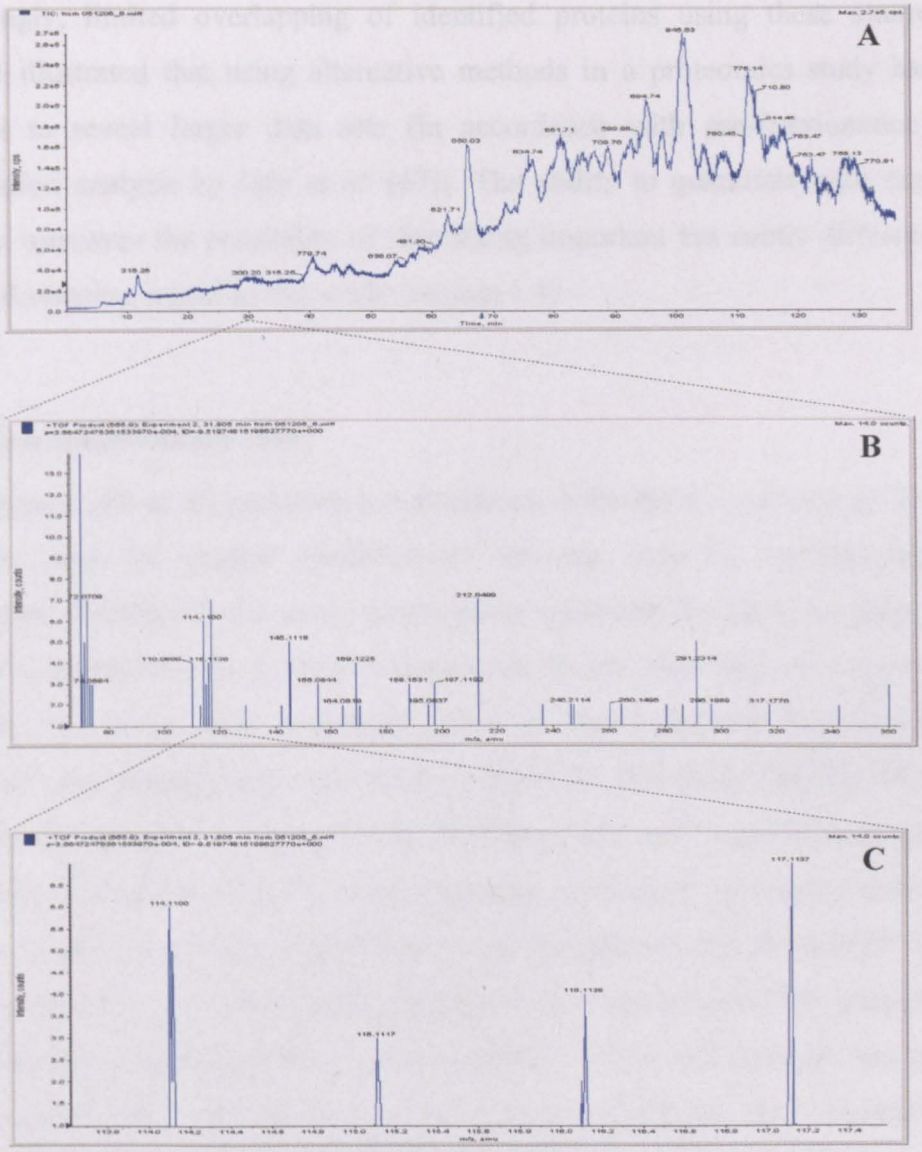


Figure 2.3: iTRAQ methodology by MS. A) Total ion count (TIC) on ESI MS/MS from a peptide fraction. B) MS/MS fragmentation pattern of peptide which eluted after 31.75 minutes. C) Reporter ions (114.1, 115.1, 116.1, 117.1) from iTRAQ reagents which are interpreted into relative quantitations.

A study compared quantitation statistics between a standard 2DE experiment to the shotgun labelling method, iTRAQ, using biological and technical replicates [86]. They found 95% of quantified protein expression ratios had a coefficient of variation (CV) less than 0.81 (81%) and less than 0.53 (53%) using 2DE and iTRAQ, respectively [86]. A more recent study compared improved in-gel methods, DIGE and cleavable isotope-coded affinity tags (cICAT) to iTRAQ [87]. A comparison of quantitation sensitivity found iTRAQ to come on top with DIGE last.

Interestingly, limited overlapping of identified proteins using these alternative methods illustrated that using alternative methods in a proteomics study has the potential to reveal larger data sets (in accordance with pre-fractionation and fractionation analysis by Gan *et al.* [67]). The ability to quantitate with superior accuracy improves the possibility of identifying important but subtly differentially regulated proteins, a goal in this study (section 1.4).

2.2.3 Mass spectrometry (MS)

The common link in all proteomics experiments is the mass spectrometer. It is an important tool for protein identification but can also be implemented for quantitation. Principally the mass spectrometer measures the mass-to-charge ratio of analytes (peptides). Analytes are measured in the gas phase and are ionised most popularly by electrospray ionisation (ESI) or matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) [88, 89]. MALDI-MS is usually coupled to time-of-flight (TOF) analysers and therefore intact peptide masses are measured, and these are subsequently matched to theoretical masses in a database to provide protein identification. This procedure is known as PMF. ESI is mostly coupled to ion traps, triple quadrupole and quadrupole-TOF instruments, which fragment selected precursor ions (peptides) further and generate ion spectra [90] (Figure 2.3B). Utilising an array of different algorithms, this information is matched against protein sequence databases to provide protein identifications and information on peptide sequence is also provided. In this entire study a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer, ESI-qQ-TOF-MS/MS (Applied Biosystems or AB, Framingham, MA; MDS-Sciex, Concord, Ontario, Canada) is implemented. Q refers to a mass-resolving quadrupole and q refers to an r.f.-only (non-mass filtering) quadrupole. It is coupled with an online capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands).

2.2.4 Interpretation of MS data

Presently, a challenge in proteomics is to convert large amounts of collision induced dissociation (CID) spectra from MS/MS analysis into confident and reliable peptide

identifications, and ultimately protein identifications [91]. In PMF, the generated list of experimental peptide masses are compared to the calculated values [33]. However, protein identification using peptide CID spectra provides information about peptide sequence from the peak pattern in the CID spectrum. In MS/MS, an ionised peptide (precursor) is selected during the first MS applied and then cleaved into fragments (products) by collisions with an inert gas (nitrogen) in the second MS. The mass to charge ratio of the resulting products is recorded and then matched to theoretical information in a database using specific algorithms [92, 93]. The product ions produced after fragmentation is dependant on several factors including charge state, internal energy and primary structure [94]. The nomenclature of fragment ions is based on whether the charge (essential for MS detection) is retained on the N-terminal fragment (a, b or c) or C-terminal fragment (x, y or z) (see Figure 2.4).

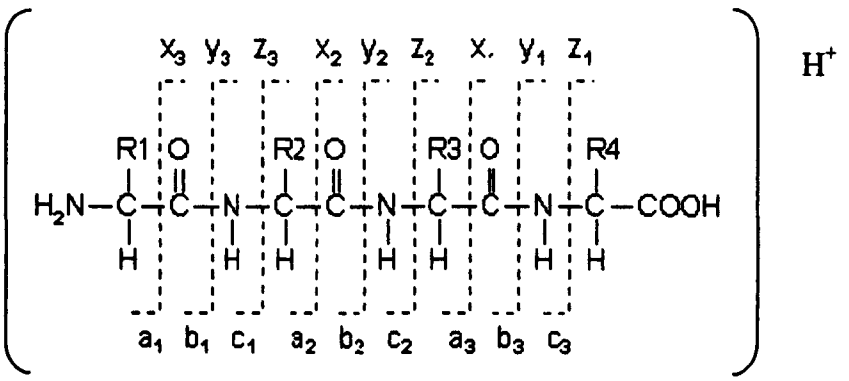


Figure 2.4: The types of fragment ions observed in an MS/MS spectrum [95].

During the advent of proteomics it was possible to manually verify these peptide assignments to spectra as data sets were relatively small. With large data sets this is not a facile task, and despite reports of integrated inspection reports [96], the problem of false positives is enhanced. To address these issues, filtering criteria are often applied based on available data including database search identification scores [69, 97]. It has been recommended that proteins must be identified by 2 or more unique peptides as part of general guidelines which were introduced recently, and are detailed in the Molecular and Cellular Proteomics journal (www.mcponline.org). However, if only one peptide can be found, previous studies have accepted the identification if at least 2 good quality MS/MS spectra are

assigned to it, and the peptide is not *shared* amongst proteins (they are distinct) [84].

A variety of database search strategies have been implemented to give an indication of false positive rates in large scale proteomic studies including reverse database, double database and very recently a target-decoy search strategy [66, 67, 98, 99]. Whichever filtering criteria are used to identify proteins in LC-MS/MS experiments, they have a requirement to be robust and preferably reveal the statistical principles which are employed so that identification probabilities are easy to estimate [66, 100]. Moreover, properties such as Mw in gel electrophoresis and pI in IEF can be used as independent measurements to confirm protein identifications.

2.2.5 Cross-species proteomics

Proteomics has traditionally appealed to the study of an organism with complete or near complete sequencing of its entire nucleotide code, and this is not surprising considering the relative simplicity of identifying proteins in these organisms. Protein identification becomes more of a challenging issue when working with organisms which do not have their genome sequenced. This is an area known as cross-species proteomics. This field has not been fully exploited, and is applicable to this study where characterisation of environmental adaptation in the unsequenced isolate *Euhalothece* is a major goal [101, 102].

Cross-species proteomics is currently being developed, and is at best at an early stage [101, 103-106]. The importance of working with unsequenced and environmentally isolated organisms in proteomics is only becoming apparent since the sheer number and diversity of potentially important microorganisms in the environment, compared to commonly studied laboratory isolates, has been realized. In general, there are two main approaches to identify proteins from unsequenced organisms. Firstly, MS/MS data are searched against the entire protein database of sequenced organisms in a conventional way, where the presence of proteins from evolutionary closely related organisms are likely to yield identical peptides to the

organism in question. Moreover, accompanying the rapid growth in number of fully sequenced genomes is the increasing number of protein sequences present in these databases. This maximises the chance of identifying proteins from an unsequenced organism using this *conventional* method of protein identification. Conventional software, for example Mascot (Matrix Science, London, U.K.) relies on matching of tryptic peptides to theoretical peptides present in databases. With increasing numbers of sequencing projects for evolutionary distinct organisms, the chance of an unsequenced organism producing exactly the same tryptic peptide (ortholog) is correspondingly enhanced. Developments in the area of MS and bioinformatics have also significantly expanded the applicability of *homology-based* proteomics [101, 102, 105, 107, 108], where non-identical peptide sequences can be matched to those present in databases, and can still yield high confidence protein identifications. Mass Spectrometry-driven Basic Local Alignment Search Tool (MS BLAST) is a homology-based database search protocol which uses a list of peptide sequences generated by *de novo* interpretation of MS/MS spectra using ProBLAST software (dove.embl-heidelberg.de/Blast2/msblast.html) [102]. Interpretation of MS/MS spectra involves generating sequence data by measuring mass differences between adjacent fragment ion peaks in a major ion series, for example the b-series (ions with N-terminus) (refer to Figure 2.4). Significant hits in MS BLAST are colour-coded based on a high scoring pairs (HSP's) algorithm. This tool does not allow gaps within individual peptides, whereas gaps between peptides can be of arbitrary length. The actual peptide list used to search the National Centre for Biotechnology Information (NCBI) database is therefore obtained from redundant, degenerate and partially inaccurate peptide sequence data, assembled into a single random searching sequence [102, 105].

More recently, open source software SPIDER uses an alternative algorithm for identification of novel (not derived from sequenced genomes) proteins [109]. Collectively, they rely on the interpretation of MS/MS spectra using *de novo* sequencing, a technique which has also advanced with new powerful software [110, 111]. MS-homology also performs homology-based searches [112], and allows the user to select the maximum number of amino acid substitutions permitted per sequence. Different peptides from a single protein are added to a list and exact and similarity searches are undertaken. This tool also allows mass values to be added

for sections of the sequence. FASTS is another emerging tool used in cross-species proteomics [113]. It searches databases by looking at all possible alternative arrangements of peptide sequences, and therefore does not require them to be entered in any particular order. Most notably, rather than giving a similarity score to matched peptides in databases, an alignment probability is presented. Pevtsov *et al.* [103] recently compared different *de novo* sequencing software tools which resulted in publicly available NovoHMM [114] and PepNovo [115] scoring highest.

Unfortunately, there are limitations to *de novo* sequencing methods, including incomplete sequence information and neutral losses i.e. loss of a water or an ammonia ion [104]. Methods to improve fragmentation during MS include introducing sulfo groups to peptide N-termini to lower amide bond strength [116, 117]. The implementation of sophisticated automated search algorithms is essential for developing more effective tools for high throughput protein identification, and this is particularly prominent in unsequenced organisms. It should be noted here, that if the organism under study is very distantly related to any organism with a sequenced genome, the probability of obtaining a positive protein identification decreases [118]. A 16S rRNA comparison is a useful way of predicting whether an unsequenced environmental isolate is suitable for a cross-species study. An attractive candidate would harbour a similar sequence to a fully sequenced organism. A phylogenetic tree is constructed in this study for this purpose (Chapter 4). Nevertheless, this comparison should be treated with caution, as 16S rRNA divergence may not be indicative of proteome homology.

2.2.6 Incorporating transcriptomics

Global transcriptomics has advanced rapidly since the development of DNA microarrays, or DNA chips, and many studies have utilised this technology to understand adaptation of microorganisms to environmental stresses [119-122]. As organisms respond to environmental stresses, gene expression profiles give an indication of which proteins are required to adapt to the new conditions. However, despite being intricately linked, the flow from transcriptome to proteome is not straightforward. Understanding a system response demands the uncovering of regulatory interactions between transcriptome and proteome, and this requires

simultaneous monitoring of both mRNA and protein levels. Recently, interpretation of protein expression data with reference to transcript level data has become popular [61, 122-126]. The relationship between mRNA and protein expression levels for selected genes in the yeast *Saccharomyces cerevisiae*, were determined by Gygi *et al.* [126]. They concluded that correlation was insufficient to *predict* levels of one characteristic using data from another. Recently, Suzuki *et al.* [122] investigated stress responses using a microarray and 2DE and identified well correlated responses (transcriptional control), as well as those which were mainly regulated at the protein level (translational control). This increased concordance in data may be due to the techniques implemented which give a more global representation of both proteome and genome [127], or because the prokaryotic system studied is less complex to interpret. In this study, reference to transcript level data where available is made to aid interpretation of protein expression data into biological function, and in Chapter 3, transcript level data using RT-qPCR is generated together with proteomic data.

2.3 Cyanobacteria general background

Cyanobacteria, formerly referred to as blue-green algae, are a major diverse group of organisms and are the only known prokaryotes that perform oxygenic photosynthesis [128]. They have colonised a wide range of ecosystems including aquatic systems, soil, air and dry rock [129]. Their ability to thrive in extreme environments, habitats previously believed to be intolerable, has been well documented [21, 130-132], but the strategies adopted are not fully understood at the molecular level. Cyanobacteria are the organisms which form the basis of the investigations in this study, and therefore a general background is given in this section.

For many years, cyanobacteria were classified in the plant kingdom along with algae, but the birth of new biochemical techniques and the utilisation of the electron microscope found them to be prokaryotes, more similar to bacteria than plants. Different species can be red, yellow, brown and green, and the red type is often said to have given the Red Sea its name, attributed to the *Oscillatoria* species, which contains the red photosynthetic pigment phycoerythrin [133]. Cyanobacteria were

often classified according to structural criteria, but this has since been replaced by 16S rRNA classification [134, 135].

Cyanobacteria are able to perform oxygenic photosynthesis for energy, and are therefore believed to be the ancestral origin of higher plant chloroplasts [128]. The process by which cyanobacteria are believed to have integrated into plants' cells is referred to as endosymbiosis, a theory which is supported by various structural and genetic similarities, and predicted to have occurred sometime during the Precambrian [136]. The ability of these microorganisms to split water using light energy to form organic compounds and the bi-product oxygen, is thought to have initiated the oxygenic atmosphere we rely on for survival today, a process believed to have begun up to 3.5 billion years ago [137, 138]. Understanding the mechanisms of this development is of interest to evolutionary biologists. Even differences between certain types of cyanobacteria have revealed novel evolutionary behaviour, including genome reduction to adapt to specific environments [139].

Cyanobacteria have long been known to produce novel compounds to aid their survival. These often-complex compounds are secondary metabolites and have been found to have applications in areas of medicine and industry [21]. As early as 1500 BC cyanobacterial species were used to treat gout and several forms of cancer [140]. However, interest in the production and applications of these novel compounds has increased significantly since the rapid development of tools and techniques allowing us to exploit basic genetic information [21, 22]. Recently the product Cyanovirin 16, a 101 amino acid protein, has been entered into clinical trials for potential anti-viral activity [141, 142]. The compound was isolated from the cyanobacterium *Nostoc* sp., and found to have virion receptor binding ability for the human immunodeficiency virus.

Some species of cyanobacteria are referred to as extremophiles due to their ability to survive in environments previously thought unhabitable. They are classified according to the conditions in which they exist. Thermophiles, hyperthermophiles, psychrophiles, halophiles, acidophiles, alkaliphiles and barophiles are known to live in high temperatures, very high temperatures, cold temperatures, high salt

conditions, acidic conditions, alkaline conditions and high pressure conditions, respectively. The discovery of extremophiles demonstrates the extraordinary adaptability of primitive life forms through the process of evolution, and further raises the prospect of finding at least microbial life elsewhere in the Solar System and beyond [1, 143].

2.4 Cyanobacteria in this study

Three different cyanobacteria, *Synechocystis* sp., *Euhalothece* sp. and *Prochlorococcus marinus*, form the basis of proteomic research in this thesis, and share a variety of features as well as contrasting sharply in others.

Synechocystis is a unicellular and freshwater cyanobacterium. It is the first of its kind to have its genome completely sequenced and annotated [144]. With approximately 3.6 million base pairs (bp), a total of 3,569 protein-coding genes are estimated. Despite more than half of its sequences sharing little or no similarity to genes present in Genbank in 1996 [144], its ability to grow photoheterotrophically and natural transformability all enhance its suitability for biochemical, physiological and genetic studies [144, 145]. Consequently, it is well characterised and is the representative strain of unicellular cyanobacteria in laboratory studies. The large amount of interest in *Synechocystis* stems from several perspectives and is summarised in Table 2.1.

Table 2.1: A summary of how research with the cyanobacteria in this thesis has contributed to human affairs

Organism	Human affairs	Function	Role	Reference
Synechocystis	Energy	Biofuels	H_2 production as alternative to fossil fuels	[146]
	Agriculture	Salt tolerance	Present model organism for salt susceptible crop plants	[60]
	Medical	Drug factory	Carotenoid synthesis	[147]
	Biotechnology	Biopolymer factory	Cyanophycin production as alternative to polyaspartic acid	[148]
Prochlorococcus marinus MED4	Environment	CO_2 sink	Potentially reduce greenhouse effect	[16]
Euhalothece sp. BAA001	Agriculture	Salt tolerance	Possible model organism for salt susceptible crop plants	[149]

composition which is thought to affect their ability to photoacclimatise [154-157]. This trait is an example of environmental adaptation, dependent on whether they reside in surface or deep oceanic waters with high-light (HL) or low-light (LL) conditions respectively. These differing ecotypes coexist in the oceans, allowing the survival of the species as a whole over a broad range of environmental conditions. Despite the arrival of the complete genome sequence for several strains, the proteomes of these globally significant microorganisms have not previously been published.

In this thesis, the proteomic response of *Synechocystis* and *Euhalothece* cells to changes in salt concentrations will be investigated and then compared (refer to section 1.4, 'Hypothesis, aims and scope'). Therefore a review of the findings so far in relation to salt tolerance in cyanobacteria is discussed in the next section. The proteomic response of *Prochlorococcus* to HL is also investigated and the situation regarding our present understanding of the light response is also reviewed, as well as an introduction to this globally significant organism. The salt response in cyanobacteria has been studied more extensively and therefore forms the main body of the review in this chapter.

2.5 Cyanobacteria and salt

Increasing salinity is a major factor impairing worldwide agricultural productivity, and is believed to affect nearly half of the world's irrigated land [158, 159]. Furthermore, the problem is predicted to get considerably worse over the next 30 to 50 years [8]. When plants are exposed to high salinity, major processes such as photosynthesis, respiration, protein synthesis and energy metabolism are affected, and cause reduced productivity or whole plant death. Compounding effects of salinity and associated environmental stresses in plants are not fully understood, particularly at the molecular level.

Understanding how cyanobacteria acclimate to saline environments has been a source of intense interest, fuelled by two main drivers, agriculture and biotechnology. Firstly, comparable to higher plants, cyanobacteria perform oxygen-evolving photosynthesis, and therefore they share many functional pathways. It is

relatively easy to culture many species of cyanobacteria, and additional advantages such as rapid growth rate and low culture costs make these bacteria ideal models for investigating metabolic processes such as photosynthesis and respiration. Cyanobacteria therefore make suitable models for studying the physiology of salt tolerance and this has provided valuable insights into revealing the nature in which salinity prevents crop plant species from using aquatic resources [8, 9]. Indeed, homologues of salt responsive genes in *Synechocystis* are also regulated by salt stress in higher plants [160]. Cyanobacteria have been classified into three groups relating to their salt tolerance, salt sensitive (or stenohaline), moderately halotolerant, and extremely halotolerant [161]. Example species for each group include *Anabaena*, *Synechocystis* and *Aphanothece*, respectively. By increasing our understanding of salt tolerance mechanisms in these microorganisms, their close relationship to higher plants can be exploited, ultimately transferring their survival capability to crop plants and thereby increasing overall world food production [160, 162].

From a non-agricultural perspective, some of the osmotic compounds produced in response to high salt have the potential to play important roles in biotechnology and medicine. Osmotic compounds (also known as compatible solutes, osmoprotectants and osmolytes) are low-molecular mass, uncharged or zwitterionic, hydrophilic molecules that do not interfere with cell metabolism. These compatible solutes help restore osmotic balance with the cell's surroundings, and maintain membrane integrity and protein stability [163, 164]. Cyanobacteria are able to synthesise their own compatible solutes, as well as uptake them from their surroundings. They help stabilise and even enhance protein activity, for example, polyethyl glycols are used to improve the crystallization of proteins [165]. In medicine, trimethylamine N-oxide has the capability of *in vitro* rescue of cystic fibrosis (misfolded) proteins [166]. They are also used to protect cells from stress in the cosmetics industry, for example, ectoine has been shown to protect skin from UVA induced damage [167, 168].

2.5.1 Detrimental effects of salt

It is now clear as to why it is advantageous to fully understand the salt response in cyanobacteria. However, to be able to comprehend what has been discovered to date, it is vital that the problems associated with the presence of salts at the molecular level are apparent. Perhaps obviously, salinity in soils and water refers to the presence of salts, most commonly NaCl. Elevated NaCl concentrations cause both ionic and osmotic stress to cyanobacterial and plant cells. In the first case Na^+ and Cl^- ions enter the cell via the chemical and electrochemical gradient due to ion imbalance, leading to toxic consequences, including affects on enzyme activity, metabolism destabilisation and even irreversible inactivation of oxygen evolving machinery [169, 170]. Low concentrations of Na^+ ions are required in the cells for several cell functions including, pH regulation, ammonia and CO_2 fixation, photosystem II (PSII) water photolysis and even transport of compounds [171, 172]. However, these concentrations are approximately 10 mM, and because Na^+ ions enter cells passively, high external concentrations lead to toxic intracellular conditions [169, 170]. The internal concentration of Cl^- ions is maintained lower than the concentration in the environment, and this is assisted by the negative electric potential of the inner surface of the plasma membrane [170]. However, during salt shock (initial stages), increasing Na^+ ion uptake causes membrane depolarisation therefore causing Cl^- uptake across a chemical gradient (passive transport) [170]. It has been reported that Na^+ and Cl^- ion concentrations can reach 50-90% of the environment in the first few minutes of salt shock in cyanobacteria [161], even though fast Na^+ influx could not be detected in *Synechocystis* using flame photometry [60]. Secondly, a decrease in water potential results in water loss and reduced turgor pressure on the cell, which can cause cells to collapse and further concentration of ions and biomolecules [161]. Ultimately, excess salts can affect all major metabolic pathways [173]. The requirement to counteract or tolerate these detrimental effects is therefore a matter of life or death.

2.5.2 Responses to salt stress

The immediate response by *Synechocystis* cells to salt stress is well understood, and is similar to many other plants and bacteria [61, 160, 174-177]. As mentioned previously, the sudden increase in external NaCl concentration causes an influx of

Na^+ and Cl^- ions into the cytoplasm during turgor collapse. The immediate response is to deal with toxicity of excess Na^+ ions, and these are removed from cells (within 10 to 20 minutes) [129, 161]. Sodium levels in *Synechocystis* have been shown to be under control of Na^+/H^+ antiporters, and analysis of the complete genome sequence suggests at least five Na^+/H^+ antiport enzymes are present [144, 178]. They are embedded in the cytoplasmic membranes and the energy required is believed to be generated from the proton motive force across this membrane, generated by H^+-ATPase and respiratory cytochrome oxidase [179-181]. In most cyanobacteria, Na^+ ions are also removed in exchange for K^+ ions, which are essential for cell survival and the high external Na^+ concentration interferes with K^+ uptake [174]. In salt stress, the K^+/Na^+ transporters start to function with a preference for K^+ ions, and therefore uptake of Na^+ ions is minimised [182]. Under optimum growth conditions, cells maintain high concentrations of K^+ ions and low concentrations of Na^+ ions and the levels of these ions directly effect active functioning of many enzymes, membrane potential generation, and cell volume regulation [13, 183]. Therefore salt stress-induced increase in cytoplasmic Na^+ ions, leads to Na^+ ion export and K^+ uptake, and the stress has been shown to induce sodium and potassium transporters [14].

The next response, which can last over several hours, is to accumulate compatible solutes either by synthesis or uptake [60, 184]. Compatible solutes primarily stabilise proteins and membrane structure via restoring an acceptable osmotic potential with the surroundings [163, 167]. Types of osmolytes include sugars, amino acids, polyols (all with derivatives), betaines, ectoines and peptides. As mentioned previously, the term 'compatible' hints at their ability to not detrimentally affect or inhibit cellular functions. They are uncharged at physiological pH, and have even been described as inert [185]. Ultimately, compatible solutes cause a decrease in the osmotic potential in the cytoplasm preventing further water loss, as well as stabilising 'ionically stressed' proteins [186]. Their molecular modes of function are not entirely clear, and have led to several theories including those summarised below [167];

1) Osmolytes interfere with solute-macromolecule interactions by competing for water molecules, leading to preferential hydration of the protein.

2) Osmolytes have unfavourable interactions with peptide backbones. More of these backbones are hidden in a folded protein compared with a denatured protein and therefore a higher concentration of osmolytes leads to a bias (i.e. makes it more energetically favourable) towards folded proteins. This has been termed the 'osmophobic effect' [187].
3) Solutes induce changes in water structure. Different ions bind to water differently (based on Hofmeister series [188]), chaotropes are weak and displace water from protein surface causing destabilisation, whereas kosmotropes are strong and therefore aid the preservation of the hydration layer around proteins.

Synechocystis can tolerate salt concentrations up to 1.2 M (7.2%) NaCl, and is assisted by accumulation of osmolytes glucosylglycerol (GG) and sucrose [184]. This is achieved by *de novo* synthesis, as well as uptake via ATP-binding cassette (ABC)-type transporters [189]. Indeed, a hypoosmotic shock has been shown to lead to rapid extrusion into the medium [190]. GG is synthesised in a two-step reaction [176] as follows:-

(i) ADP-Glucose + glycerol 3-phosphate $\xrightarrow[\text{(GgpS)}]{\text{GG-phosphate synthase}}$ GG-phosphate (1)

(ii) GG-phosphate $\xrightarrow[\text{GG-phosphate phosphatase}]{}$ GG (2)

Mutant cells for the gene which encodes for GgpS (*ggpS*) were unable to divide in 0.45M (~2.7%) NaCl, and caused cells to almost double in size [174]. GG production has been shown to be activated primarily by a biochemical, salt-dependent mechanism [176] and a hypothetical model for regulation of GG is given by Joset *et al.* [129]. Interestingly, cells shocked by a sudden increase in salt concentration demonstrate no correlation between *ggpS* mRNA and GgpS protein content [191]. However, cells adapted to high salt in the long-term show a linear relationship between external salt concentration, *ggpS* mRNA and GgpS protein content. Therefore biochemical modulation of this enzyme as well as increased gene expression (and increased mRNA stability) may regulate GG production proportionally to stress [191].

The level of salt tolerance harboured by a cyanobacterial strain, depends upon which compatible solute it accumulates and this is summarised in Table 2.2 [184].

Table 2.2: Salt tolerant groups of cyanobacteria with examples (adapted from [184]).

Group	Example species	Maximum NaCl concentrations	Compatible solutes
Salt sensitive	*Synechococcus, Anabaena*	Max 0.7 M (< 4.2%)	Sucrose, trehalose
Moderately halotolerant	*Synechocystis*	Max 1.8 M (< 10%)	Glucosylglycerol
Highest halotolerant	*Spirulina, Aphanothece*	Max 2.7 M (< 15%)	Glycinebetaine, glutamatebetaine

It should be noted here that more than one type of compatible solute can accumulate in one strain in addition to the most abundant one, and this practice is often dependent upon the level of the salt tolerance [129, 161]. Accumulation of disaccharides, for example, trehalose, is advantageous for strains residing in environments with rapidly changing salinity because they are synthesised more rapidly than other compatible solutes, for example GG [192]. The details of trehalose and sucrose synthesis in cyanobacteria remain to be uncovered, however, glycine betaine is known to be synthesised by oxidation of choline in *Synechococcus* [193].

2.5.2.1 Photosynthesis and respiration

Studies have shown that photosynthetic activity in cyanobacteria is reduced in salt-stressed cells [60, 194], (refer to Figure 2.5 for an overview of photosynthesis). The response of photosynthetic pigments to salt stress is particularly interesting, and Schubert *et al.* [195] investigated the effects of different salt concentrations on photosynthesis and pigmentation in *Synechocystis* cells with several interesting outcomes. At 342 mM (~2%) NaCl, cells showed optimum growth rate, photosynthesis rate and pigment (chlorophyll *a* and phycocyanin) content. However, when comparing photosynthesis rate to pigment content, it was demonstrated that absorbed light was more efficiently utilised in photosynthesis in cells with less pigmentation and grown at higher salt concentrations of ~6.2% [195]. Fluorescence emission spectra showed inefficient energy transfer between

the light harvesting complexes (phycobilisomes) and chlorophyll *a* in ~2% NaCl. Moreover, the accumulation of the carotenoids, echinenone, oscillaxanthin and myxoxanthophyll, were detected in high salt (1026 mM) [195]. Carotenoids are anti-oxidants and their increased abundance has been implicated in response to salt stress in cyanobacterial cells [195, 196] and eukaryotic green algae [197]. It is thought that absorbing excess photons prevents oxidative damage, a secondary stress caused by high salt concentrations [195]. A water soluble orange carotenoid has also been associated with state transition in PSI and PSII within the thylakoid membrane of *Synechocystis* [61].

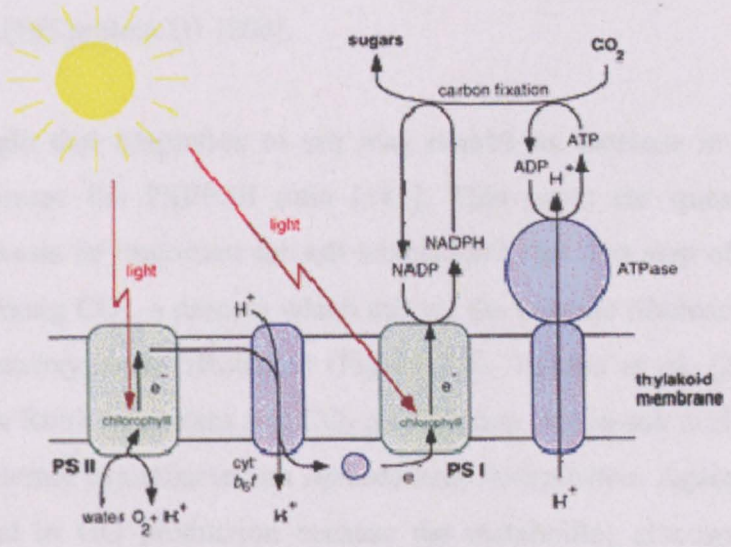


Figure 2.5: Overview of photosynthesis in cyanobacteria (reproduced from ASU, http://photoscience.la.asu.edu/photosyn/education/photointro.html).

In *Synechocystis*, the photosynthetic electron transport is thought to be inhibited by increases in intracellular Na^+ concentrations. The ionic stress causes stabilising proteins to dissociate from PSI and PSII complexes [198]. In *Spirulina platensis*, salt stress has been shown to decrease electron transport efficiency and cause phycobilisome damage [199]. Ionic stress is thought to be the main problem as hyperosmotic stress damage to both photosystems has been shown to be reversible in *Synechococcus sp.* [200]. It is possible that Na^+ ions impair restoration of salt-damaged photosystems. A study on *Synechoccocus* showed that both PSI and PSII were inactivated by a change in K/N ratio, an alteration attributed to changing external salinity [201].

the light harvesting complexes (phycobilisomes) and chlorophyll *a* in ~2% NaCl. Moreover, the accumulation of the carotenoids, echinenone, oscillaxanthin and myxoxanthophyll, were detected in high salt (1026 mM) [195]. Carotenoids are anti-oxidants and their increased abundance has been implicated in response to salt stress in cyanobacterial cells [195, 196] and eukaryotic green algae [197]. It is thought that absorbing excess photons prevents oxidative damage, a secondary stress caused by high salt concentrations [195]. A water soluble orange carotenoid has also been associated with state transition in PSI and PSII within the thylakoid membrane of *Synechocystis* [61].

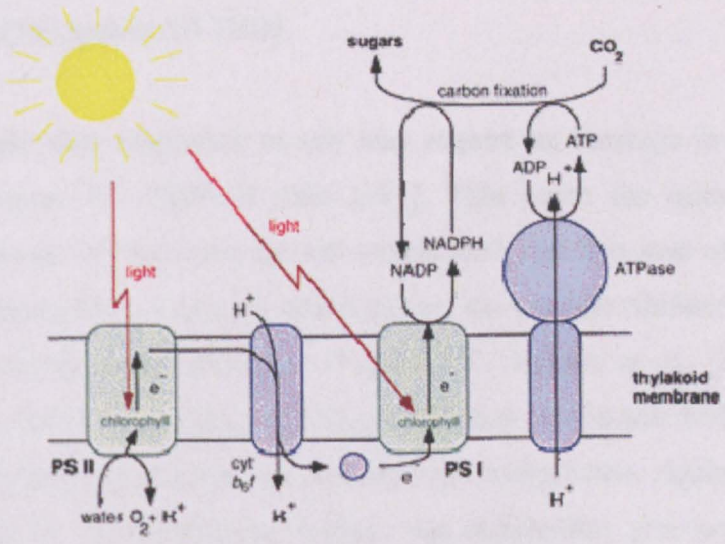


Figure 2.5: Overview of photosynthesis in cyanobacteria (reproduced from ASU, http://photoscience.la.asu.edu/photosyn/education/photointro.html).

In *Synechocystis*, the photosynthetic electron transport is thought to be inhibited by increases in intracellular Na^+ concentrations. The ionic stress causes stabilising proteins to dissociate from PSI and PSII complexes [198]. In *Spirulina platensis*, salt stress has been shown to decrease electron transport efficiency and cause phycobilisome damage [199]. Ionic stress is thought to be the main problem as hyperosmotic stress damage to both photosystems has been shown to be reversible in *Synechococcus sp.* [200]. It is possible that Na^+ ions impair restoration of salt-damaged photosystems. A study on *Synechoccocus* showed that both PSI and PSII were inactivated by a change in K/N ratio, an alteration attributed to changing external salinity [201].

Four percent salt stress is believed to cause functional disconnection of phycobilisomes to the PSII photosynthetic reaction centre in *Synechocystis* cells [195]. Comparably, the addition of salt causes a decrease in energy transfer from allophycocyanin to PSII in the cyanobacterium *Spirulina platensis* [202, 203]. However, no change in chlorophyll *a* content was observed in this cyanobacterium grown under 0.8 M (~4.8%) NaCl [202]. In the same study, an artificial electron donor to PSII was used to illustrate that salt stress may affect the water splitting complex. The target varies in different cyanobacteria and in *S. platensis* the water oxidation complex and PSII reaction centres are directly affected by salt [199]. In *Synechocystis*, salt stress has been shown to inhibit protein synthesis, and in particular PSII protein D1 [204].

It is thought that adaptation to salt may require an increase in PSI activity, and hence increase the PSI/PSII ratio [181]. This poses the question; why would photosynthesis be important for salt adaptation? The first step of the Calvin cycle involves fixing CO_2, a process which utilises the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) (Figure 2.6). Takabe *et al.* [205] measured an increase in Rubisco content and CO_2 assimilation rate in salt acclimated cells from the halotolerant cyanobacterium *Aphanothece halophytica*. Again the Calvin cycle is involved in GG production because the metabolites glucose-6-phosphate and ADP-glucose are precursors for this compatible solute, as discussed previously. Carbon compounds in cyanobacteria are mostly formed from photosynthesis, and these are required for compatible solutes synthesis [206]. This has been demonstrated by rapid accumulation of compatible solutes in the light, whereas in the dark small amounts are synthesised from glycogen or glucose reserves [207]. Therefore, photosynthesis is required for compatible solute synthesis and also an increased requirement for ATP because Na^+ ion extrusion is an active (energy requiring) process. In accordance, cyanobacteria have shown increased electron transport activity in PSI in response to salt stress [181, 199]. Work by Jeanjean *et al.* [181] showed a salt-induced increase in P700 and PSI reaction centres leading to increased cyclic electron transport around PSI. Figure 2.7 illustrates electron flow around PSI in cyclic photophosphorylation. This increase in electron transport is not solely controlled by an increase in electron carrier complexes cytochrome *c* oxidase

or P700, and therefore changes in electron transfer routes most probably occurred [181].

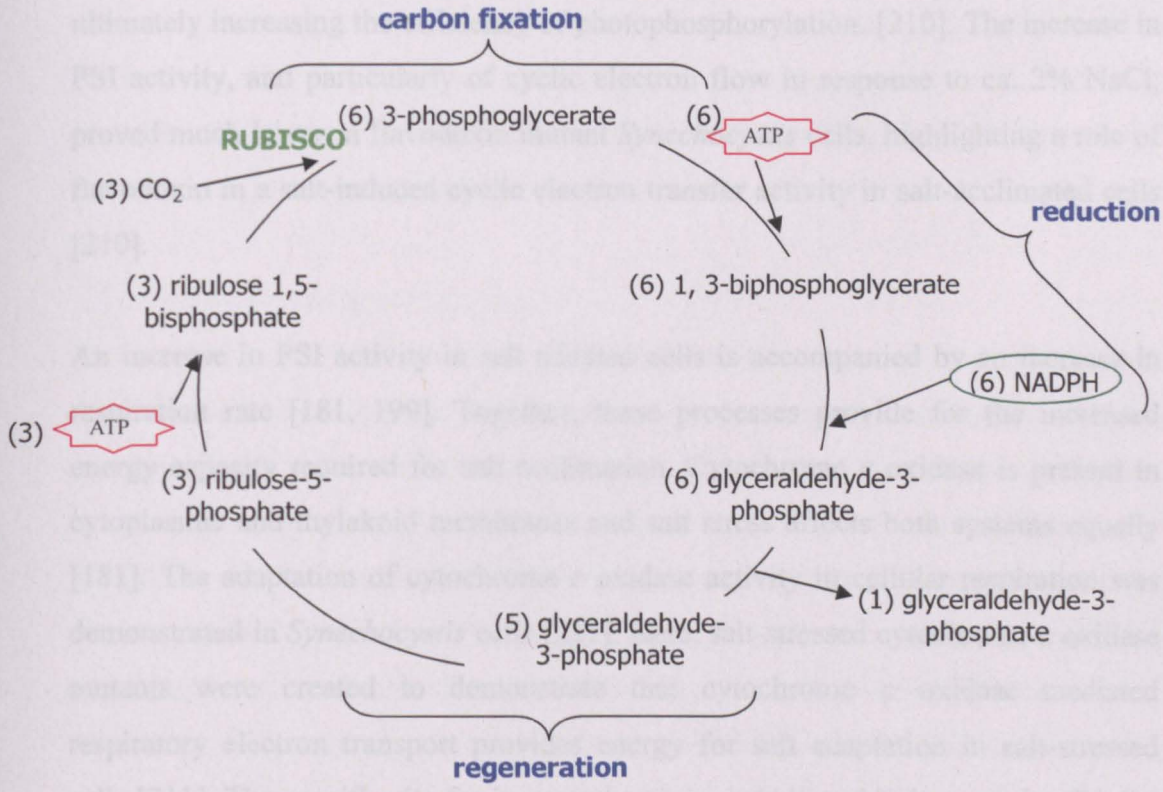


Figure 2.6: Representation of the Calvin cycle or light independent reactions of photosynthesis where CO_2 fixation occurs (in a clockwise direction). The numbers in brackets represent numbers of each molecule. The blue text describes the three main phases and the enzyme Rubisco which binds CO_2, is shown in green [208].

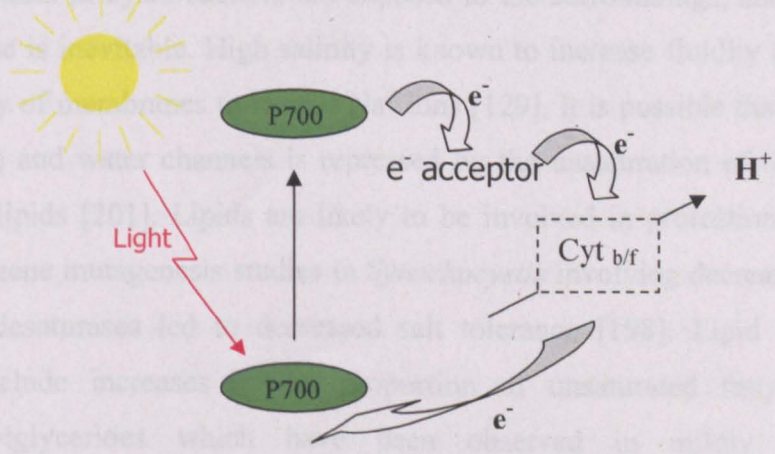


Figure 2.7: Cyclic photophosphorylation in PSI. The electron transferred is not derived from water, but from PSI itself and therefore recycles it back to PSI [209].

Subsequently it was shown that levels of electron flow carrier flavodoxin, an alternative to ferredoxin, increases in salt acclimated cells, and this may lead to the observed enhancement of PSI activity by increased cyclic electron transport, and ultimately increasing the efficiency of photophosphorylation. [210]. The increase in PSI activity, and particularly of cyclic electron flow in response to ca. 2% NaCl, proved much lower in flavodoxin mutant *Synechocystis* cells, highlighting a role of flavodoxin in a salt-induced cyclic electron transfer activity in salt-acclimated cells [210].

An increase in PSI activity in salt adapted cells is accompanied by an increase in respiration rate [181, 199]. Together, these processes provide for the increased energy capacity required for salt acclimation. Cytochrome *c* oxidase is present in cytoplasmic and thylakoid membranes and salt stress affects both systems equally [181]. The adaptation of cytochrome *c* oxidase activity in cellular respiration was demonstrated in *Synechocystis* cells [211]. Here, salt-stressed cytochrome *c* oxidase mutants were created to demonstrate that cytochrome *c* oxidase mediated respiratory electron transport provides energy for salt adaptation in salt-stressed cells [211]. The specific site for increased activity is believed to be complex IV, the terminal part of the cytochrome c oxidase catalysed respiratory chain [211].

2.5.2.2 Membranes

The membranes in cyanobacteria are exposed to the surroundings, and therefore a salt response is inevitable. High salinity is known to increase fluidity and decrease permeability of membranes to H^+ and Na^+ ions [129]. It is possible that the activity of K^+ (Na^+) and water channels is repressed by the unsaturation of fatty acids of membrane lipids [201]. Lipids are likely to be involved in protection against salt stress, and gene mutagenesis studies in *Synechocystis* involving decreased levels of fatty acid desaturases led to decreased salt tolerance [198]. Lipid composition changes include increases in the proportion of unsaturated fatty acids and phosphatidylglycerides which have been observed in mildly halotolerant *Synechococcus* PCC6311 [212], and the latter strategy has also been seen in halophilic *Aphanothece halophytica* [213]. These changes are associated with modifying ion exchange properties of the cells because increasing the polarity of

lipids causes an increase in membrane surface charge [129]. Riviere *et al.* [214] have shown that the ratio of proteins to lipids decreases in the cytoplasmic membrane of *Synechococcus* PCC6301 cells as well as increases in Fe^{2+} and Cu^{2+} ions. These increases can be explained by the increased levels of cytochrome *c* oxidase and H^+/Na^+ antiport activity [214]. The presence of unsaturated fatty acids in thylakoid membranes is also important for the tolerance of photosynthetic machinery to salt stress [215].

2.5.2.3 How do cells sense salt stress?

The cellular changes described above allow *Synechocystis* to tolerate salt concentrations more than twice as much as seawater, but acclimation must involve perception of these stress signals with subsequent transduction. Sensing environmental changes in salt concentration is necessary so that previously unexpressed genes essential for acclimation can be expressed, ultimately leading to stress protein production [120]. *Synechocystis* has its genome completely sequenced, and therefore targeted mutagenesis has been used to understand functioning of potential sensors and transducers. Murata *et al.* [120] performed a systematic analysis of environmental stress perception with histidine kinase and response regulator mutants using DNA microarrays. Initial studies showed that selections of salt-inducible genes and proteins are regulated by sigma factor SigF, as revealed by interposon mutagenesis with 1DE and 2DE [216]. Marin *et al.* [119] implemented a similar method and identified four histidine kinases, Hik16, Hik33, Hik34, and Hik41, which perceived and transduced salt (0.5 M) signals. Again using DNA microarrays, but in combination with RNA slot-blot analysis, three histidine kinase systems, and another type of system were identified in salt stress perception and signal transduction [217]. The marked reduction or elimination of gene expression allowed salt stress genes to be associated with specific two-component response regulator systems. Many salt responsive genes did not experience altered expression in mutant strains, indicating that unknown mechanisms of control still exist. Intracellular concentrations of Ca^{2+} ions have also been observed to increase rapidly in response to salt stress in *Anabaena* sp. PCC7120, and these were postulated to play a role in salt stress signalling [218].

2.5.2.4 Gene expression changes

Microorganisms adapt to different environmental stresses by regulating the expression of various genes, and analysis of these changes enables interpretation of physiological responses. Gene expression is regulated in response to changes in salt concentrations, and using a subtractive hybridisation technique, Apte *et al.* [219] revealed transcription of more than 10% of the genome was affected by salt in *Anabaena torulosa.* More recently, Kanesaki *et al.* [3] used DNA microarray analysis to monitor changes in levels of transcripts in *Synechocystis* in response to 0.5 M (~3%) salt. Upon exposure to salt, 147 (4.8%) of the total genes on the microarray (3079) were found to be up-regulated at least 2-fold, and 228 genes (7.4%) were down-regulated at least 2-fold. Interestingly, half of these genes encoded proteins with unknown function, highlighting the further studies that have to be carried out to elucidate the full salt response [3]. Among the salt-induced genes were those which encode for ribosomal proteins, *rp12, rp13, rp14* and *rp123.* It is suggested that salt stress destabilises ribosomes, and therefore *de novo* synthesis is required to fulfil their active roles in the cell [3]. In addition, genes associated with the D1 protein at the photochemical reaction centre of PSII were also affected. The degradation and regeneration of this protein is essential for PSII activity [220], and genes involved in both were found to be up-regulated [3]. Finally, essential genes for biological membrane structure, the acyl-lipid desaturases, were repressed by salt stress [3].

Marin *et al.* [60] also used the DNA microarray technique to study the response of *Synechocystis* cells to a slightly higher salt concentration of 684 mM (~4%). However, in contrast to the previous study, acclimation was assessed after initial 15 minutes of salt addition, and 30 minutes, 2 hours and 6 hours thereafter. This enabled genome wide profiling of the initial salt shock response. Measurements were also made after 24 hours and 5 days, and therefore allowed analysis of gene expression with cells approaching adaptation to ~4% salt. Overall, 11.7% and 6.5% of total genes on the microarray (3079) were induced and repressed, respectively [60]. Again many of these genes encoded hypothetical proteins. Amongst the salt-induced genes in adapted cells, were those which encode for enzymes involved in GG synthesis and ABC-transporters for compatible solutes. Genes encoding for

proteins in basic carbohydrate metabolism, for example, fructose, 1-6 bisphosphate aldolase, were enhanced after 24 hours but not after 5 days.

In accordance with present understanding of the *Synechocystis* salt response, all but one gene encoding for proteins involved in water and ion export were no longer differentially expressed at 24 hours. A transient decrease in genes involved in photosynthesis activity (both PSI and PSII) was also observed [60]. Perhaps most significantly, despite the early induction of genes for group 2 sigma factors (SigB, SigD) in salt stress, group 3 sigma factors, believed to be involved in expression of salt stress proteins [216], were not. Therefore, it is likely that group 3 sigma factor activity is regulated at the post-transcriptional level [221]. Overall, very few genes (39) remained induced after 5 days acclimation.

2.5.3 Proteomics of salt stress

The work reviewed above has increased our understanding of the strategies cyanobacteria implement to survive in high salinity. The immediate responses (or shock responses), including the accumulation of compatible solutes and active export of inorganic ions, are largely regarded as protein synthesis-independent [129]. However, in order to survive in high salt for extended periods of time, cells must adapt to the new conditions. Adaptation is a long-term process, and is therefore protein synthesis-dependent (see Figure 2.1). It is apparent from the microarray study above [60] that minimal numbers of genes change in expression during long-term adaptation, and therefore it is imperative to compare short-term and long-term responses, particularly at the protein level.

Fulda *et al.* [61] used 2DE in a large-scale proteomics study to address the salt shock (immediate or short-term) response versus salt adaptation (long-term) response in *Synechocystis* cells to approximately 4% NaCl. In order to produce a meaningful comparison, previously generated transcriptomic data and a survey of changes in physiological parameters were used to identify the most revealing times to harvest cells. Short-term shock and long-term adaptation times were at 2 hours and 5 days, respectively, under the growth conditions applied [60, 61]. A pulse ^{35}S-methionine labelling method was used, which allowed clear visualisation of newly

induced proteins from radioactive 2DE-gels. *In vitro* metabolic labelling of cells with radioactively labelled amino acids is routinely performed, but despite advantages such as its relatively higher energy beta emitter potential and high specific activity, any proteins lacking methionine would be undetected [222, 223]. Despite the fewer induced proteins identified compared to genes induced in a microarray study, 90% of these identified proteins were also induced at the transcriptional level [60]. As expected, heat shock proteins (which protect or repair proteins), DNA and RNA-binding stress proteins and antioxidative enzymes all increased in expression in the short and long-term, in both transcriptomic and proteomic studies [3, 60, 61, 224]. However, only truncated versions of the chaperone GroEL1 and the elongation factor-Tu, were identified in the long-term at the protein level, and whether this has functional significance remains unknown [61]. Seven enzymes involved in carbohydrate metabolism also accumulated after 5 days, whereas a similar pattern in gene expression was not observed. In fact, in this study, nearly half of the data for induced genes in acclimated cells (5 days) did not correlate with protein inductions [61]. Therefore, the authors proposed that post-transcriptional modification may play a role in salt adaptation, particularly considering the multiple spots identified in gels corresponding to the same protein. Despite the need for further studies here, the importance of studying protein levels in understanding the salt response is accentuated.

A physiological proteomic study also compared high salt (6% NaCl) shock and adaptation, using 2DE with dual channel imaging, but this time in non-cyanobacterium, *Bacillus subtilis* [225]. Enzymes involved in most of the metabolic pathways were repressed initially after the salt shock, but after a 60 minute period approaching adaptation, growth processes were restarted (the relatively fast generation time of *Bacillis subtilis* means cells adapt relatively fast compared to *Synechocystis* cells). However, some proteins, for example, glycolysis enzymes, were synthesised at a normal rate throughout shock and adaptation [225].

The majority of proteomic studies on salt tolerance in cyanobacteria have been conducted on *Synechocystis* implementing 2DE techniques. Newly expressed or proteins with enhanced expression have been identified and given the term 'stress proteins'. These studies have contributed to the construction of a database called

Cyanobase, containing over three thousand predicted protein-coding genes [226]. Many of these proteins are annotated as 'hypothetical', as sequence similarity to proteins of known function is simply not present. Methods for isolating plasma, thylakoid, periplasmic and outer membranes have been developed, as well as isolating pure plasma and periplasmic space proteins [227-232]. This has allowed analysis of how proteins in different subcellular compartments assist in adaptation of whole cells to high salt, and these findings are reviewed here and summarised in Figure 2.8.

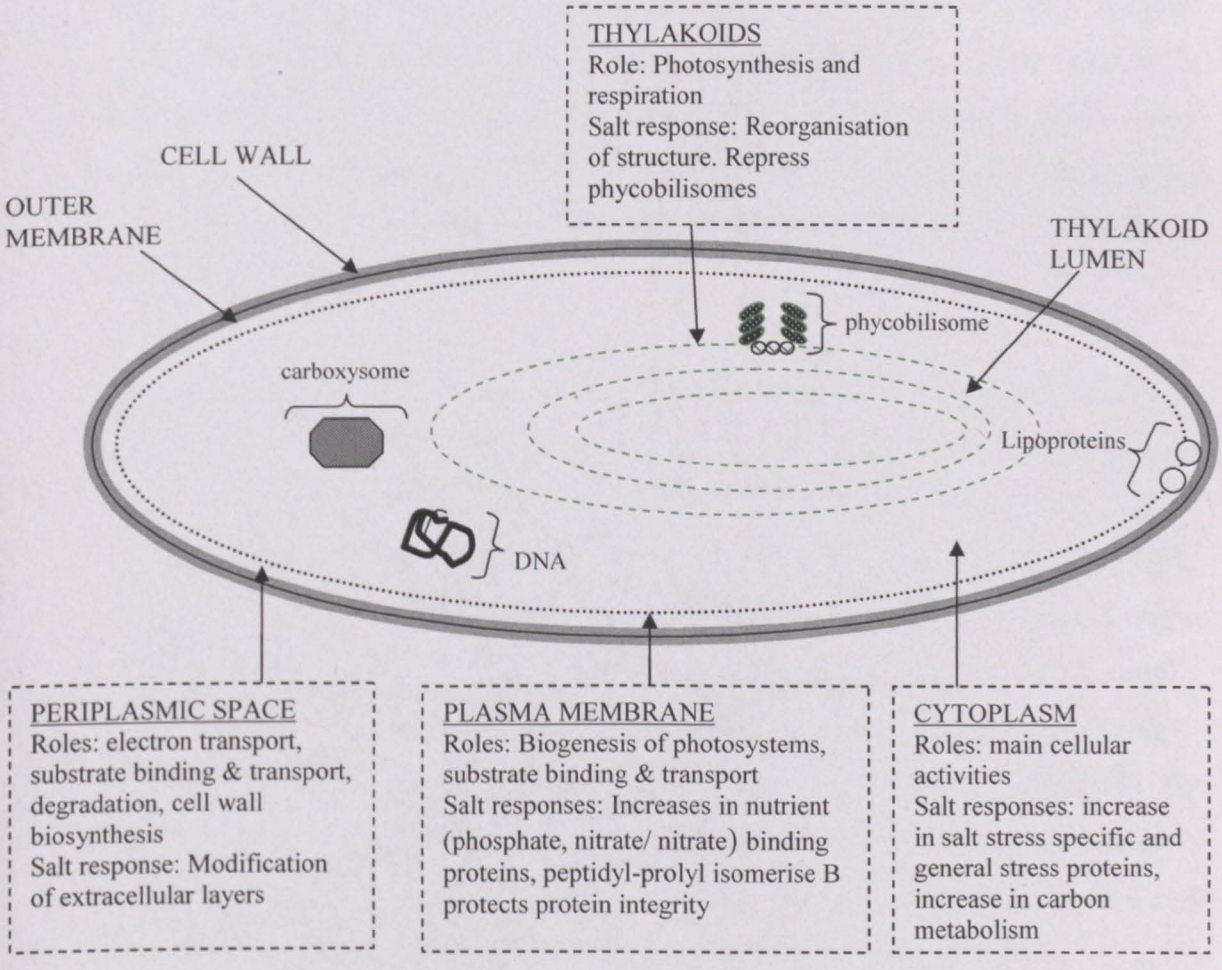


Figure 2.8: A summary of responses to salt stress in a unicellular cyanobacterial cell, as revealed by proteomics studies [61, 231, 233-235].

2.5.3.1 Periplasmic Space

The periplasmic space (Figure 2.8) is situated between the plasma membrane and the outer membrane. No active ion transport mechanisms have been found in the outer membrane in *Synechocystis* cells, but transport proteins involved in the uptake of substrates are present in the periplasmic space, and these could be affected by high salt levels. These reasons inspired Fulda *et al.* [231] to use a cold-osmotic shock technique [236] to isolate and compare the abundance of periplasmic proteins from *Synechocystis* cells grown in low salt (2 mM or 0.01% NaCl) and high salt (342 mM or 2% NaCl). The method of periplasmic protein isolation was unsuitable for looking at adaptation to higher salinities, as the sample fraction becomes contaminated with proteins from other cellular compartments in these conditions. Extracted protein fractions were separated by SDS-PAGE, and two bands of interest were subjected to N-terminal sequencing. Unfortunately, both identified proteins were annotated as hypothetical, sharing no significant similarities to proteins of known function [231].

By implementing 2DE for enhanced resolution with MALDI-TOF MS for improved identification, the experiment was repeated [233]. IPG strips pH 4-7 were used with 23cm by 23cm, 12.5% acrylamide gels. MALDI-TOF MS allows protein identification by soft ionisation methods with a TOF mass spectrometer [89]. Protein fragments generated by protease digestion are fixed in a solid phase matrix and ionised by a laser beam. TOF analysis reveals experimental peptide masses which can then be matched to theoretical masses in a database. Six and three proteins were identified as enhanced and newly expressed, respectively, in the presence of elevated levels of salt [233]. Several of these proteins were thought to be involved in synthesising or modifying extracellular layers in *Synechocystis*, and therefore highlight the importance of changes in cyanobacterial cell walls in adapting to high salinity. For example, a protein showed significant similarity to phosphoglycerate mutases from *Streptococci*, which are involved in polysaccharide synthesis in their capsules [233]. Furthermore, a salt-enhanced protein was found to be homologous to cell surface lipoproteins of *Mycobacteria* spp. [233].

It is apparent from this study, that functional annotation of *Synechocystis* is still at an early stage, and characterisation of these salt-enhanced proteins required

searching for similar functional domains already present in the databases for other source organisms. A number of periplasmic proteins were also observed to have reduced expression in saline media, and these were attributed to the expected reduction in protein synthesis of 'normal' proteins (non-stress proteins) in salt-stressed cells [176, 233].

2.5.3.2 Plasma membrane

A combination of blue native/SDS-PAGE and MALDI-TOF was used to study the composition and dynamics of membrane protein complexes in the same cyanobacterium, perturbed by three different growth modes and varying levels of CO_2, iron, or salt [232]. Blue native is a technique implemented in proteomics in order to separate and study intact protein complexes at high resolution [237, 238]. It enables further elucidation of protein function by analysing protein-protein interactions. A total of 53 protein spots were identified in this study, corresponding to 37 different *Synechocystis* genes [232]. A salt stress of 0.4M (~2.4%) NaCl was applied to cells for 72 hours, but this did not induce any observable changes in the membrane protein complexes, which agreed with previous physiological studies (reviewed by Joset *et al.* [129]) that the plasma membrane and cytoplasm are likely to be the main subcellular compartments which are affected by salinity. The plasma membrane represents a barrier to the surrounding medium, and therefore is likely to be responsive to a salinity change in the environment [234].

Huang *et al.* [224], screened for proteomic changes in the plasma membranes of *Synechocystis* cells in response to salt. These membranes were isolated by sucrose density gradient centrifugation and two-phase partitioning [239]. 106 proteins were identified, corresponding to 66 gene products, using 2DE and PMF with MALDI-TOF MS. 25 proteins changed significantly (>2-fold) in abundance due to salt concentrations of 684 mM (~4.1%). A methanol/chloroform precipitation step was included after protein extraction for desalting and concentrating the protein sample [240]. An advantage of using 2DE to separate and visualise proteins was particularly apparent in this study, where isoform spots were clearly visible. Changes in the abundance of some of the isoforms imply PTM's might play a role in the functioning of the proteins in the stress conditions. Nearly one third of the salt-enhanced proteins identified in this study were substrate-binding ABC

transporters [224]. An increase in GG-binding protein, GgtB, was expected since *Synechocystis* cells can uptake this compatible solute from its environment [241]. The evidence for an interrelation between salt and iron-stress in this organism was supported by the very high accumulation of iron binding lipoprotein, FutA1. It has been previously proposed that FutA1 plays a protective role in PSII under conditions of iron deficiency [242]. Increases in phosphate and nitrite/nitrate binding proteins have been hypothesised to occur in order for cells to overcome salt-induced nutrient deficiency, a problem resulting from structural changes in the plasma membrane [243]. Moreover, an increase in the regulatory protein PII suggests a change in the carbon and nitrogen balance in salt-stressed cells.

Further salt-induced proteins, thought to play significant roles in stress, included vesicle-inducing protein in plastids (Vipp1), a protein involved in thylakoid membrane biogenesis in *Arabidopsis* [244], membrane-bound peptidyl-prolyl isomerase B, which could be involved in maintaining the integrity of proteins in the plasma membrane and CoxB, which supports work done recently suggesting a role in managing photosynthesis in stressed cells [224]. Some of these hypothesised functions in salt tolerance require further characterisation, and this can occur at the protein level, for example, through structural studies, or at the gene level, developing knock-out mutations from the protein-coding genes and observing phenotypic alterations. This is particularly true for hypothetical proteins which have no associated function. In addition to salt regulated proteins, 21 proteins were newly identified [224], and before further studies are undertaken it is possible to predict function to a certain limit. A plethora of open source software tools have been developed [245-247], with algorithms which help to predict cellular functions from conserved domains. Furthermore, characteristics such as hydrophobicity and attributes like lipoprotein or sec- signals, can be used to predict subcellular localisation [248, 249].

Many proteins from the plasma membrane were associated with salt stress in *Synechocystis* for the first time in this study, however, the limitations of using 2DE were recognized in the fact that no integral membrane proteins were identified [224].

2.5.3.3 Cytoplasm

Fulda *et al.* [61] used 2DE with MALDI-TOF to investigate the response of the soluble proteome in *Synechocystis* cells to approximately 4% salt, and 55 out of the 337 identified proteins were induced. In the salt-acclimated cells, induced proteins were organised into four groups; stress proteins which are salt specific, general stress proteins, enzymes involved in basic carbon metabolism and hypothetical proteins.

Specific salt stress proteins included the enzyme ADP-glucose pyrophosphorylase, which is responsible for synthesising the precursor ADP-glucose, for the compatible solute GG. General stress proteins included molecular chaperone GroEL1 and elongation factor-Tu. Increased amounts of soluble electron carriers, flavodoxin and plastocyanin were also observed, and these are thought to play an important role in adjusting to stress-induced changes in electron transfer [61]. Several enzymes involved in carbohydrate metabolism also accumulated, including transketolase, glycogen phosphorylase and phosphoglycerate kinase [61]. The accumulation of the compatible solute GG was proposed to cause this change in carbon metabolism.

2.5.3.4 Thylakoids

A method for isolating pure thylakoid membrane proteins suitable for 2DE analysis was developed nearly ten years ago, implementing aqueous polymer two-phase partitioning in combination with sucrose-density centrifugation [227]. Despite this, a large-scale proteomics experiment assessing the salt response of the photosynthetic membranes has not been performed. By combining traditional protein separation and quantitation methods, SDS-PAGE and western blotting, Sudhir *et al.* [235] investigated the response of thylakoid membrane proteins in the cyanobacterium *Spirulina platensis* to 0.8 M (~4.8%) NaCl. This cyanobacterium was isolated from sodium (150-200 mM Na^+) rich lakes and is cultured in Zarrouk medium [235]. A 40% decrease in protein D1 from the PSII reaction centre was observed. Additional abundance changes in proteins were observed, but these were not identified. The authors undertook additional physiological studies and postulated that a variety of effects occurred on photosynthetic electron transport activities due to the marked alterations in the composition of thylakoid membrane proteins [235].

Vipp1 protein, (refer to *2.5.3.2*), is located in the plasma membranes, but has been associated with thylakoid membrane formation in *Arabidopsis* [244]. It shares sequence similarity to stress-related phage-shock protein PspA [244], and a knockout mutant created in *Synechocystis* led to an inability to form thylakoid membranes. An increase in this protein in salt-acclimated cells may imply a structural reorganisation in these membranes is necessary [224].

This review of salt tolerance studies using proteomic tools illustrates the potential of this field to increase our understanding of the complex mechanisms which cells use to survive in saline environments. However, the major technique employed has been 2DE with densitometry-based (semi-)quantitation, and this method does not have sufficient accuracy to highlight all essential adaptive proteome responses. In previous studies the immediate gene and protein level responses to external stimuli have been shown to be more pronounced than changes which correspond to an adaptive state [60, 61], and therefore a method which can only detect >2-fold changes is likely to miss key expression variations (high false negative candidates). A proteomic method which is more accurate would be ideal to overcome these limitations (refer to section 1.4, 'Hypothesis, aims and scope'). In addition, protein fractionation by 2DE is not ideal for identification of hydrophobic proteins, although progress is being made here [46, 57]. The application of a shotgun method is presented in Chapter 3, and attempts to deal with the drawbacks mentioned above.

Synechocystis can survive in extreme salt conditions, up to ~7% salt. However, proteome or transcriptome responses to more than ~4% NaCl have not been conducted. A possible reason for this trend in proteomics may be the unattractive prospect of working with high salt, which is known to cause detrimental effects in major proteomic techniques including IEF (first dimension separation in 2DE) and MS. Techniques to remove salt are available but many have several disadvantages including protein loss or degradation, increase in final sample volume, time taken or even high implementation cost. Therefore, a facile method to remove salt for 2DE studies would allow more effective analysis of very high salt samples, and this

problem is addressed in Chapter 4 (refer to section 1.4, 'Hypothesis, aims and scope').

2.5.4 Cross-species proteomics of salt tolerance

As mentioned previously, there are several drawbacks associated with working only with model organisms, and techniques need to be developed and evaluated to answer similar questions in environmental isolates. There have been no publications to date on the salt response in any unsequenced cyanobacterium using cross-species proteomics tools. However, a study on the halophilic and unicellular green alga *Dunaliella*, used homology-based proteomics to study salt tolerance, since it was unsequenced [250]. To study this halophile, nanoelectrospray MS, combined with sequence-similarity database-searching algorithms, MS BLAST [102] and MultiTag [118] were applied. Results showed that 61 proteins were up-regulated, including enzymes of central metabolic pathways, such as photosynthesis, energy production, protein synthesis and turnover [250]. However, at the time of study (2004), there were 50 protein entries and 3000 nucleotide entries in the NCBI database for this organism which assisted in protein identification.

A more recent study on the halophilic eubacterium, *Halorhodospira halophila* attempted cross-species proteomics, utilising a chemical derivatization strategy and *de novo* sequence analysis with 2DE and MALDI-TOF/TOF [106]. Three different homology-based search algorithms were implemented, FASTS, MS BLAST and MS-Homology [102, 113, 251]. The study was undertaken to illustrate the applicability of the method in unsequenced organisms and not to elucidate salt adaptation strategies.

Both these studies highlight different proteomic strategies which can be used to improve identification in cross-species proteomics. The applicability of MS BLAST was crucial in both studies. In the *Dunaliella* study, 80% of the selected salt-induced spots were identified, and 42% of protein spots were identified in *H. halophila* [106, 250].

In Chapter 4, the salt response in an extremely halotolerant cyanobacterium which was isolated from a salt lake in the heart of the Sahara, is characterised. Protein identification using *de novo* sequencing and homology-based searches will be implemented, but the use of novel identification confirmation methods is also demonstrated (refer to section 1.4, 'Hypothesis, aims and scope'). Again, adaptation to its hostile environment will be elucidated rather than the immediate shock response, and therefore accurate quantitations as well as reliable identifications are a feature. This work should raise the prospect of utilising proteomic tools in more biologically diverse ecosystem-adapted organisms.

2.6 Cyanobacteria and light

In plants and cyanobacteria, photosynthesis is used to provide energy for survival, and this in turn requires energy from the sun in the form of light. Many organisms obtain nearly all of their carbon and chemical energy for growth through photosynthesis, but the availability or quality of light in the natural environment is never constant [252]. If there are limitations in light availably, growth will concurrently be limited. On the contrary, excess light can cause irreversible damage to cells and is called photoinhibition, unless adaption to the new conditions can be accomplished, termed photoinhibition [253]. Consequently, photosynthetic organisms must adapt to varying light intensities for survival. For example, cyanobacteria residing in water surfaces are constantly confronted with changing light intensities due to water mixing and must have the capacity to respond efficiently to successfully adapt to these environments. If photosynthesis and the components involved are not correctly regulated it has been shown that cellular process can become damaged due to the production of reactive oxygen species (ROS) caused by the imbalance of energy supply and dispersion [254]. Generally, electrons leaking from the photosynthetic electron transport chain generate ROS, leading to oxidative stress [254]. Figure 2.9 shows that excess energy from excited pigments gets transferred to molecular oxygen and thereby produces reactive singlet state oxygen ($^1O_2$). Molecular oxygen can also get reduced at the acceptor side of PSI producing superoxide anions (O_2^-) which subsequently leads to hydrogen peroxide (H_2O_2) and hydroxyl radical ($^{\cdot}OH$) production.

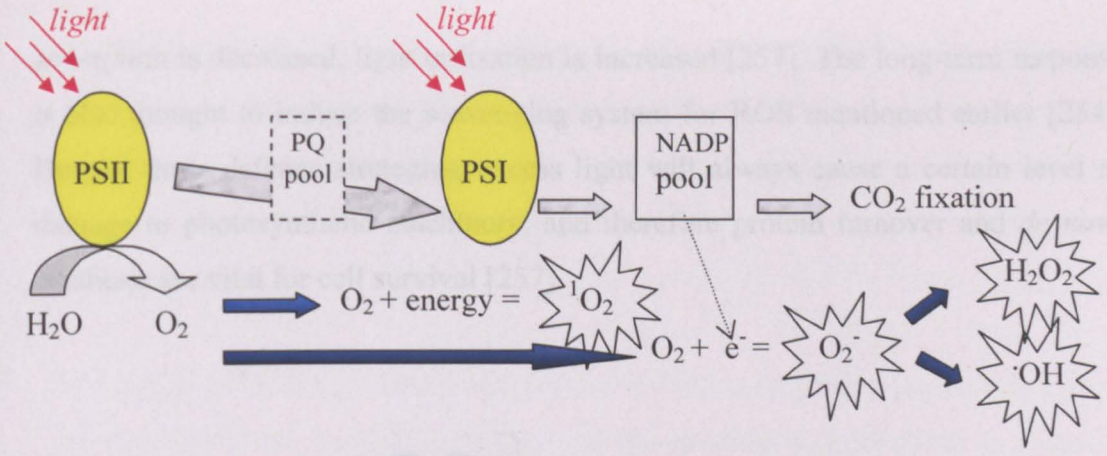


Figure 2.9: Schematic representation of how excess energy causes the formation of ROS in the photosynthetic electron transport chain, which eventually cause oxidative damage to cellular components (PQ-plastoquinone, PSI- photosystem I, PSII- photosystem II, NADP- Nicotinamide adenine dinucleotide phosphate) [254].

There has not been the extent or depth of research into the light response in cyanobacteria that has been undertaken to understand their salt response. What is known to date has mainly focused on plants through traditional biochemical, physiological and molecular biology (forward and reverse genetics) tools [220, 254, 255], and the light response remains relatively poorly understood at the molecular level.

2.6.1 Responses to light stress

Responses to light occur in two stages. There are several immediate (shock) cellular responses to HL, which are independent of gene expression. State transition is an immediate response. It is a regulatory mechanism which controls excitation energy transfer from light harvesting antennae complex to both PSI and PSII [252]. However, protein synthesis dependent adaptation to HL (long-term) occurs over a period of hours, or even days. It is thought to involve adjusting photosynthetic capacity and light absorption by controlling light harvesting antennae size, enhanced repair, changes in energy transfer from the light harvesting complex to photosystem II (PSII) and protective energy dissipation [256-258]. There are also reports of alterations in photosynthetic capacity and photosynthesis apparatus and related components, for example D1 turnover (see Figure 2.10), and although light

absorption is decreased, light utilisation is increased [257]. The long-term response is also thought to induce the scavenging system for ROS mentioned earlier [254]. Despite these defence strategies, excess light will always cause a certain level of damage to photosynthetic machinery, and therefore protein turnover and *de novo* synthesis are vital for cell survival [257].

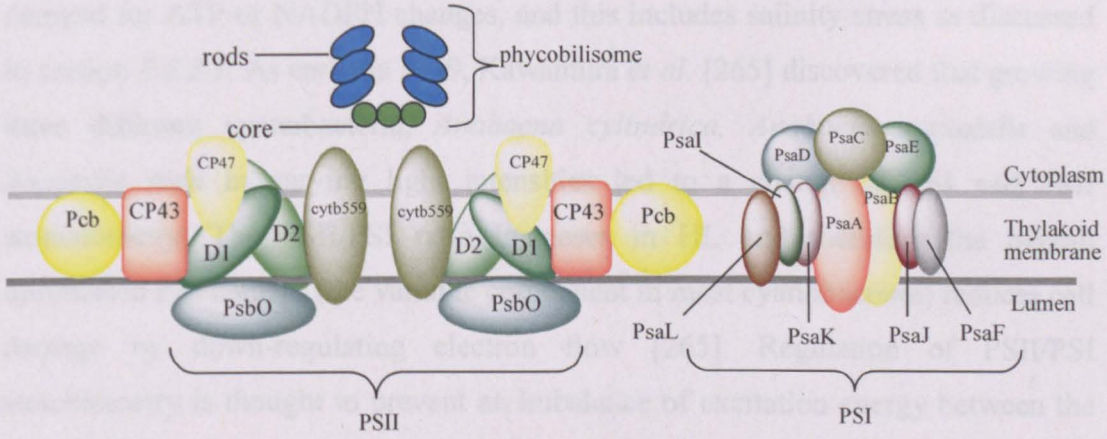


Figure 2.10: A schematic diagram of the major proteins located in the thylakoid membrane of cyanobacteria (organization adapted from Partensky and Garczarek [259]).

To avoid the production of ROS through absorption of excess light energy, both antenna size and photosystem content must be reduced, and the rate of CO_2 fixation must also increase to maintain the energy supply and consumption balance [252]. Most cyanobacteria reduce the length and/or number of phycobilisome complexes (Figure 2.10) upon exposure to HL [260]. Interestingly, Belknap *et al.* [261] discovered that control of phycobilisome content in *Anabaena* PCC sp.7120 was controlled at the post-transcriptional level when mRNA levels increased although overall protein levels decreased. The mechanisms for regulating phycobilisome levels are not certain, although candidate genes have been proposed. The product of the *nblA* gene, found in *Synechococcus* sp. PCC 7942 and *Synechocystis*, may bind to phycobiliproteins and act as a marker for degradation [262].

As mentioned earlier, the freshwater cyanobacterium *Synechocystis* was the first cyanobacterium to have its genome fully sequenced, and has therefore been a representative species for physiological, biochemical, genetic and proteomic studies

[3, 263]. In 1998, Hihara *et al.* [264] investigated the effect of HL (300 μmol photons m^{-2} s^{-1}) on pigment abundance in *Synechocystis* cells. A major reduction in chlorophyll and phycocyanin was observed within 3 hours of exposure, leading to a more subtle decline for up to 12 hours. After this period a gradual recovery was recorded [264].

Regulation of PSI/PSII stoichiometry occurs in cells which must adapt when demand for ATP or NADPH changes, and this includes salinity stress as discussed in section *2.5.2.1.* As early as 1979, Kawamura *et al.* [265] discovered that growing three different cyanobacteria, *Anabaena cylindrica, Anabaena variabilis* and *Anacystis nidu* in varying light intensities led to a change in PSI and PSII stoichiometry. The PSII/PSI ratio increases in HL and therefore the overall diminished PSI content (the variable component in most cyanobacteria) reduces cell damage by down-regulating electron flow [265]. Regulation of PSII/PSI stoichiometry is thought to prevent an imbalance of excitation energy between the two and thus prevent damage, and mutagenesis studies in *Synechocystis* identified the gene *pmgA* as playing a role in regulating the stoichiometry in HL [264].

Excessive light energy causes oxidative damage to the reaction centre of PSII (Figure 2.10) [220]. The overall lasting damage however, is dependent on the recovery process which involves synthesising D1 to replace damaged units. In *Synechocystis* cells, the *psbA* gene encodes for D1, and has been shown to be controlled by transcription rate as well as message stability [266, 267]. Both the rate of production and degradation is controlled by light, but the level of synthesis is also controlled by the level of degradation [268]. Control of D1 and D2 proteins (Figure 2.10) in cyanobacteria is complex, and there is more than one gene which encodes for each protein in cyanobacteria, unlike in the chloroplast genome [252]. Movement to environments with differing light intensity dictates which gene is transcribed, and in the case of *psbA* (encodes for D1) in *Synechococcus* sp. PCC 7942, HL causes the production of Form II over Form I at a rate dependent on light intensity change [269]. After further exposure to HL however, acclimation favours Form I again, implying that Form II type is a transient stress-inducible version, whereas Form I is the acclimated form of D1 [269]. D1 protein appears to be the

primary target for photoinhibition in cyanobacteria, and it is assumed that the exchange between different forms plays an important role in adaptation to varying light intensities [252, 269].

Genetic studies identified the HL inducible protein (HLIP) coding genes [270]. Homologous to eukaryotic chl a/b binding proteins (CABs), mRNA levels of HLIP's are only induced transiently and are thought to work as pigment carrier proteins in HL [252]. Pigment-protein complexes become degraded in HL stress leading to potentially excitable free chls, and these would have to be removed for their energy to be dispersed harmlessly to avoid the inevitable oxidative damage [252]. Another gene related to HL, *clpP1*, was discovered in *Synechococcus* sp. PCC7942 [271]. *ClpP1* encodes for the catalytic subunit of the protease ATP-dependent Clp protease and it may play a role in D1 turnover [252, 271].

These discoveries led to the belief that many light responses are regulated at the level of gene expression, and therefore a DNA microarray analysis of *Synechocystis*, under varying light conditions was conducted [121]. In this study, more than 160 differentially regulated genes were identified during both short-term (15 min) and relatively long-term (15 hours) when shifting from 20 to 300 μmol photons m^{-2} s^{-1} [121]. As expected, many of these changes occurred in genes which encoded thylakoid located complexes. Genes which encoded for subunits of PSII and reaction centre protein D1, increased immediately (15 min) but decreased thereafter. However, transcript levels for reaction centre protein D2, CP47 and CP43 (Figure 2.10) decreased in the short-term and then returned to 50-100% of the original level after 15 hours [121]. Several transcripts encoding for the oxygen-evolving complex *(psbO, psbU, psbV)* were down-regulated almost 2-fold. Almost all PSI subunit related genes were down regulated in HL, and this is similar to findings in the study by Kawamura *et al.*, 22 years earlier [265]. Differences in the regulation of genes encoding for phycocyanin and allophycocyanin were explained by their relative positioning on phycobilisomes, where the former exist in the core and the latter in the rods (Figure 2.10) [121]. Inductions due to HL were also observed in *ndh* genes which encode for subunits of NADPH dehydrogenase, indicating increased CO_2 uptake. CO_2 fixation by Rubisco is a rate-limiting step

during HL saturation, and therefore increasing its abundance helps dissipate excess energy [121]. As predicted, transcripts encoding for stress proteins, for example, superoxide dismutases, heat shock proteins, GroEL and GroES, were up-regulated in HL, some transiently and others during the long-term [121].

The short-term (30 min) and long-term (12 hours) response of several cyanobacteria to light stress was also studied from the perspective of stress associated genes [272]. In this case, a set of genes as highlighted as essential in adaptation to HL in *Synechocystis*, *P. marinus* MED4 and *P. marinus* MIT9313 and included *groEL*, *groES* and *dnaK2*, with the response in *P. marinus* MED4 appearing to be the strongest [272].

2.6.2 Proteomics of light stress

Responses to light intensity have received little attention at the proteome level. However, a very recent proteomic study was conducted on the alga *Chlamydomonas reinhardtii* examining HL (2000 μmol photons m^{-2} s^{-1}) stress. A 2DE approach was used for protein separation and visualisation and PMF for identification [273]. Interestingly, only two proteins (heat shock protein and oxygen-evolving enhancer protein) out of the 32 differentially expressed proteins have been previously associated with HL stress, albeit through transcriptomic studies. Of these 32 proteins, some were carbon-fixation related, molecular chaperones, or belonged to the light harvesting complex [273]. Other proteomic studies have examined the HL response of a specific cellular compartment, for example, the thylakoid proteome in *Arabidopsis* [274] where the role of oxidative stress defence was prominent, or methodology based studies focussed on light-related groups such as light harvesting proteins [275].

2.6.3 *Prochlorococcus marinus*

In 1988, the cyanobacterium *Prochlorococcus* was discovered [15] and found to reside in nutrient poor oceans, which were also thought to be void of living organisms. It is the smallest known photosynthetic organism on earth, with a cell size around 0.5-0.7 μm [16]. Since their discovery in this hostile habitat, the interest

in their ecological impact has increased greatly, mostly because they are believed to contribute to two-thirds of the oceans' primary production [276]. Significantly, half of the global primary production is estimated to occur in the oceans [277], and the substantial contribution from *Prochlorococcus* means they play an important role in nutrient cycling and global climate control. The increasing interest in *Prochlorococcus* has meant that at least twelve different strains have now been fully sequenced [152], assisting genetic studies [16, 155, 278-280] into this microorganism's environmental adaptation strategies.

Prochlorococcus marinus is characterized by a novel complement of pigments [16] including divinyl derivatives of chlorophyll *a* and *b*, referred to as a_2 and b_2. Unlike other cyanobacteria that harvest light using phycobilisomes (extrinsic multi-subunit complexes), its major light harvesting complex is a protein which binds chl a_2 and b_2 and is situated in the thylakoid membrane [281]. Remarkably, it can be found on the surface of the ocean where levels of light are high, and up to 200 metres below, where natural light is in very small quantities. *Prochlorococcus* is able to adapt to these varying light conditions due to different ecotypes that differ in pigment composition [157]. The LL ecotype grows optimally deep below the ocean surface where the amount of light can be as low as 1 μmol quanta m^{-2} s^{-1} and is characterised by a high ratio of chl b_2/a_2 and the possession of multiple antenna genes (*pcb* genes) [282]. The genotype of the most distantly related HL ecotype is only 2.7% genetically distinct in its 16s rRNA sequence [157], but contains a low ratio of chl b_2/a_2 and only one *pcb* gene [282]. The availability of solar light is therefore a major controlling factor for the ecological distribution of *Prochlorococcus* strains [280, 283, 284].

Despite the HL ecotype being dominant near the ocean surface, it can also survive deeper into the ocean where light levels are much lower. However, the LL ecotype is unable to grow at the high irradiances experienced on the oceans' surface [285] and sudden light changes have been shown to damage the PSII reaction centre due to high energy [16]. MED4 is a HL adapted strain of *Prochlorococcus*, and therefore can survive in a range of light intensities. The surface ocean waters are subject to more rigorous mixing than deeper waters, and this could subject cells to

rapidly changing irradiances [286]. Its genome is fully sequenced and annotated, and is the smallest known photosynthetic prokaryote genome at 1.66 Mbp, much smaller than the LL strains [279]. In relation to its preferred environmental conditions, MED4 possess many more HLIP's and photolyases to repair DNA damaged by UV light, compared to its LL counterparts [285]. Interestingly, *P. marinus* is more susceptible to sunlight due to its A+C rich DNA. MED4 has one of the highest A+T contents (69.2%) [278] for bacteria, and this means it is more susceptible to DNA damage caused by UV-induced thymine dimers. Whether cells have specific mechanisms to overcome this sensitivity, or whether the increased likelihood of mutations is an advantage to *P. marinus* strains remains unclear.

Prior to this thesis work, no published studies have examined the proteome in this microorganism, despite several strains having their genomes fully sequenced. Therefore, in Chapter 6, as well as investigating the protein expression changes of strain MED4 which are required to adapt to three different light intensities, the proteome is characterised for the first time (refer to section 1.4, 'Hypothesis, aims and scope').

2.7 Conclusions

In this chapter, a background to the field of proteomics was presented with emphasis on the techniques which are available. Subsequently, areas where improvements have been made were discussed, which have expanded the applicability of the field in answering diverse biological questions. Cyanobacteria are studied in this thesis and therefore a brief introduction to these prokaryotes was presented, with a more detailed review of what is currently understood regarding their ability to adapt to different environmental conditions. The most studied cyanobacterium, *Synechocystis*, is the subject of study in the following chapter (first experimental chapter), and therefore allows direct comparisons of the data and conclusions generated (using the methodology applied) with previous studies. With the addition of transcriptome analysis, the way in which findings fit in with subsequent experimental chapters (Chapters 4, 5 and 6), is made clearer.

Chapter 3: The *Synechocystis* response to high salt using high-throughput proteomics with quantitative real-time PCR

3.1 Introduction

As discussed in Chapter 2, processes associated with modern day agriculture lead to loss of irrigated land for crop plant cultivation, specifically due to increasing salinity levels [7]. Due to their ability to perform oxygenic photosynthesis, cyanobacteria provide favourable models for understanding metabolic processes in higher plants, and therefore cellular responses to salt. The unicellular cyanobacterium *Synechocystis* has the remarkable capability of surviving in non-saline environments and high saline conditions (up to 1.2 M or ~7%NaCl) [161], and has been the representative species for cyanobacterial salt tolerance studies [60, 177, 178, 181, 210, 211, 233, 242]. Therefore, an investigation of the proteomic response in this model organism forms the first experimental part of this thesis. As reviewed in Chapter 2, despite the ability of *Synechocystis* to survive in extreme salt conditions, proteome or transcriptome responses to more than ~4% NaCl have not been conducted, and one possible reason for this trend in proteomics may be the unattractive prospect of working with high salt, which is known to cause detrimental effects on major proteomic techniques including isoelectric focusing and MS. However, extra sample preparation steps can be performed to remove the salt, and proteomic studies have been performed at 6% NaCl in other organisms including food-borne pathogen *Listeria monocytogenes* (using 2DE with MALDI-TOF) [287] and endospore-forming *Bacillis subtilis* (using 2DE with dual channel imaging) [225]. Furthermore, protein separation and quantitation has invariably been performed using 2DE.

It is argued here that studying salinity adaptation in this organism using densitometry-based quantitation with 2DE, does not have sufficient accuracy to highlight the majority of essential adaptive proteome responses. It has been shown previously that immediate changes at the gene and protein level, in response to external stresses, can be significantly different than the changes which exist in the adaptive state, and importantly, changes may even be less pronounced [60, 61]. The accuracy of densitometry-based quantitation means limitations exist in identifying potentially significant relative expression changes, most notably those under 2-fold, and therefore a more quantitatively accurate high-throughput method is desirable. 2DE also limits proteome coverage because it is poor at identifying low abundance

proteins and it is exclusive to soluble proteins within the isoelectric range dictated by IPG strips [46, 56, 72]. Implementing an alternative protein or peptide fractionation method will potentially increase proteome coverage, and elucidate more information.

In this chapter, a large-scale survey of the longer-term (9 days and >1 doubling time) response of *Synechocystis* cells to high (6%) salt using a shotgun-based system is performed. Quantitation is achieved using isobaric tags (iTRAQ). iTRAQ is powerful in that it enables the comparison of several phenotypes in one experiment, permitting the inclusion of biological replicates. Quantitation occurs at the MS/MS stage (Figure 2.3). It has several advantages over densitometry and labelling proteomic techniques [86, 288, 289] (see Chapter 2, section 2.2.2). Samples will be analysed by two different workflows implementing different online chromatography, MS and bioinformatics analysis (for protein quantitation, identification and relative differential expression) systems. The data from the two workflows is first compared and then combined for discussion of salt adaptive strategies.

The differential expression of a selection of proteins is then tested further by designing primers for the protein-coding gene, and measuring relative transcript levels over a time-series. This will allow further understanding of the relationship between shock and adaptation, but mostly between transcript and protein levels, particularly when post-transcriptional control can be postulated. A recent study compared transcriptomic and proteomic data from salt shocked microbial cells and revealed that ca. 40% of the proteomic changes were not detected previously at the RNA level, highlighting the importance in conducting protein-level analyses [61]. Discussions of salt adaptive responses are conducted with reference to existing global transcriptomic and proteomic studies.

The ultimate aim of this chapter is therefore to demonstrate how increased knowledge of the salt response in a well-studied system can be generated using our large-scale directed, proteomic methodology. The applicability of the method to identify more subtle changes in protein abundance and the relationship between proteins and transcript levels are also demonstrated.

3.2 Material and methods

3.2.1 Culturing and cell preparation

Replicate flasks of axenic cells of *Synechocystis* were grown in batch culture, with continuous shaking (220 rpm) in 250 ml flasks at 25°C, 12 hour light-dark cycle at 90 μEinsteins m^{-2} s^{-1}, in BG11 medium [135], buffered to pH 7.4 using 50 mM MOPS. For protein extraction, two biological replicate cultures were grown in BG11 media with no salt added, and 50 ml were harvested during mid-exponential phase (in the mid-light cycle) at 10,000 x *g* for 20 minutes at 4°C (Heraeus Model Multifuge 3S-R, Germany). Another set of biological replicate cultures were grown in BG11 media supplemented with 6% (w/v) salt and 50 ml were harvested during the mid-exponential phase (in the mid-light cycle) for protein extraction.

For RNA extraction, biological replicate cultures were grown in exactly the same conditions as stated in the previous paragraph. Immediately prior to supplementation with 6% (w/v) salt, 50 ml of cells were harvested (calibrator sample) and this was repeated 2 hours, 24 hours and 9 days after the addition of salt. Pelleted cells for RNA extraction were stored in *RNA*later solution (AB, Framingham, MA, USA) at -80°C.

3.2.2 Protein extraction and iTRAQ

Harvested cells were washed 5 times with sucrose based buffer consisting of 50 mM Tris (pH 7.4), 100 mM EDTA and 25% (w/v) sucrose. Cells were resuspended in 1 ml of freshly prepared extraction buffer (40 mM Tris-HCl at pH 8.7, 1 mM ascorbic acid, 5 mM $MgCl_2$, 1.1 g l^{-1} polyvinylpolypyrrolidone (PVPP), 1mM dithiothreitol (DTT) and a 5% (w/v) protease inhibitor cocktail (PIC)). Four rounds of liquid nitrogen freezing with mechanical cracking using a mortar and pestle was performed to extract as much protein in as little volume possible. The crude extract was centrifuged at 21,000 x *g* for 30 min at 4°C, and the supernatant containing soluble proteins was placed into a fresh microcentrifuge tube. Protein quantitation was performed using the RC DC Protein Quantitation Assay (Bio-Rad, Hertfordshire, U.K.). 100μg protein per phenotype was added to a new

microcentrifuge tube and proteins were precipitated in a 1:5 ratio of ice-cold TCA/acetone (with 20mM DTT), overnight at -20 °C. Proteins were recovered by centrifuging at 21,000 x *g* for 30 min at 4°C. The resulting protein pellet was washed with ice-cold acetone (with 20mM DTT). After a short incubation time for drying, the protein pellet was resolubilised in 30 µL of 0.5M triethylammonium bicarbonate (TEAB) buffer at pH 8.5, prior to protein digestion.

The iTRAQ procedure illustrated in Figure 3.1 was followed. In summary, 100µg of protein for each phenotype was reduced and alkylated using reagents accompanying isobaric labels. Trypsin digestion was performed in 1mM HCl solution, overnight in a water bath set to 37°C. Labelling was performed with the AB kit protocol with modifications [83]. Labels 114, 115 and 116 corresponded to phenotypes grown in low salt (0% w/v), high salt (6% w/v) and high salt (6% w/v) (biological replicate), respectively. After labelling, all three samples were combined and vacuum evaporated prior to SCX HPLC separation.

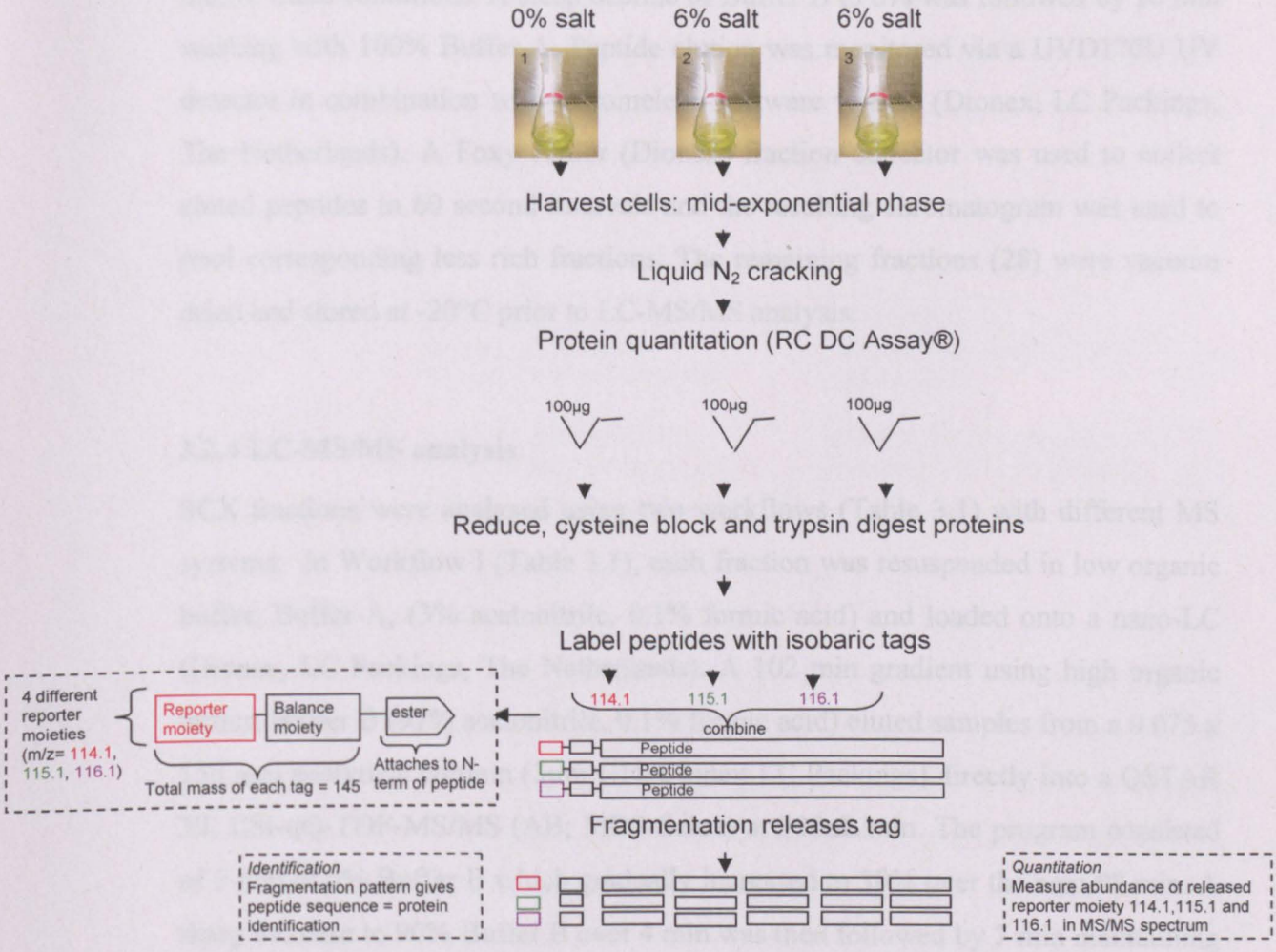


Figure 3.1: Overview of proteomic procedure used in this study with iTRAQ tagging.

3.2.3 Fractionation by SCX HPLC

Fractionation of the complex peptide mixture was achieved using a PolySULFOETHYL A Pre-Packed Column (PolyLC, Columbia, MD) with a 5 µm particle size and a column dimension of 100 mm x 2.1 mm i.d., 200 Å pore size, on a BioLC HPLC (Dionex, Surrey, U.K.). A low (Buffer A) and high ionic (Buffer B) buffer consisted of 10 mM KH_2PO_4 and 25% acetonitrile, 10 mM KH_2PO_4 and 25% acetonitrile and 500 mM of KCl, respectively, both at pH 2.85. The dried combined labelled peptide sample was reconstituted in 200 µL Buffer A, and loaded into the HPLC at a flow rate of 0.2 ml/min. The fractionation program commenced with 5 min 100% Buffer A, followed by a gradual increment to 40% Buffer B over the next 30 min. Buffer B was increased to 100% over the next 10 min, followed by 5

min of these conditions. A steep decline of Buffer B to 0% was followed by 10 min washing with 100% Buffer A. Peptide elution was monitored via a UVD170U UV detector in combination with Chromeleon software v. 6.50 (Dionex, LC Packings, The Netherlands). A Foxy Junior (Dionex) fraction collector was used to collect eluted peptides in 60 second intervals and the resulting chromatogram was used to pool corresponding less rich fractions. The remaining fractions (28) were vacuum dried and stored at -20°C prior to LC-MS/MS analysis.

3.2.4 LC-MS/MS analysis

SCX fractions were analysed using two workflows (Table 3.1) with different MS systems. In Workflow I (Table 3.1), each fraction was resuspended in low organic buffer, Buffer A, (3% acetonitrile, 0.1% formic acid) and loaded onto a nano-LC (Dionex, LC Packings, The Netherlands). A 102 min gradient using high organic buffer, Buffer B (97% acetonitrile, 0.1% formic acid) eluted samples from a 0.075 x 150 mm analytical column (3μm C18 Dionex-LC Packings) directly into a QSTAR XL ESI-qQ-TOF-MS/MS (AB; MDS-Sciex) at 300nL/min. The program consisted of 5 min of 5% Buffer B which gradually increased to 30% over the next 98 min. A sharp increase to 90% Buffer B over 4 min was then followed by 7 min maintaining these conditions. A rapid decline to 5% Buffer B was superseded by 12 min washing in these conditions. Each MS survey scan was performed per second and over an m/z range of 300 to 2000. MS/MS scans were performed on two selected precursors over an m/z range of 65 to 1600. Under manufacturer's instructions, an elevated rolling collision energy range was incorporated to achieve optimal fragmentation (increment of 10eV compared to standard settings). Samples were injected twice to increase proteome coverage and results reliability [82].

In Workflow II (Table 3.1), the exact same SCX fractions were also run on a second different HPLC and MS instrument with different analysis software. The Agilent 1200 Series HPLC-Chip/MS system which combines a trap column with an analytical column (3μm diameter C18 column, 0.075 x 150 mm) on a chip was used with the same buffers as the nano-LC described previously, except trifluoroacetic acid was used in place of formic acid. In addition, the same buffer gradient described previously was used but shortened to 30 min (ca. one third of the time).

This system also used nanospray at 300nL/min for delivery into an Agilent 6510 ESI- Q-TOF mass spectrometer (Agilent Technologies, Waldbronn, Germany). Each MS survey scan was performed per second but MS/MS scans were performed on *three* selected precursors, both over the m/z ranges.

Table 3.1: Two workflows implemented to analyse peptide fractions.

Analysis stage	**Workflow I**	**Workflow II**
HPLC	Nano-LC (Dionex)	1200 Series-Chip (Agilent)
Mass spectrometry	ESI-qQ-TOF-MS/MS (AB)	Agilent 6510
Analysis software	Analyst/Bioanalyst® and ProteinPilot™ Software v 2.0 (AB)	MassHunter and Spectrum Mill Software v A.03.03 (Agilent)

3.2.5 iTRAQ data analysis

In Workflow I, output files from the ESI-qQ-TOF-MS/MS were analysed using ProteinPilot™ Software v2.0 (AB, MDS Sciex), which employs the Paragon Search algorithm [290]. A *Synechocystis* proteome database was used to search all spectra (3264 entries, retrieved from NCBI, March 2007). All spectra were searched against the database using a defined cysteine-fixed modification, mass tolerance of 0.1Da MS, 0.15Da MS/MS and undefined mis-cleavage tolerances so that all possible cleavages would be considered. The results generated included full peptide and protein lists. A cut-off value of 70% was assigned to peptide calculations, and therefore only these would be used to calculate final protein quantitations. Moreover, only proteins with >95% confidence were considered further. Normalisation of data was carried out by adjusting the median ratio of all quantitations towards unity. This is based on the assumption that most proteins would not change in abundance under perturbed conditions. In Workflow II, data acquisition using Agilent 6510 mass spectrometer was performed using Mass Hunter workstation. Spectrum Mill Software version A.03.03 was then used for peptide and protein identification and quantitation using the same settings as above. Normalisation of the median towards unity was performed manually using Microsoft Excel.

There is the possibility that spectra become incorrectly assigned to peptides and therefore result in false positive identifications. It is important when using this type

of analysis, to provide a guide of the rate of these false positive identifications. Several methods of achieving an accurate measure have been proposed recently, including a double database strategy approach [67, 82]. This involves searching a single genome database and also a much larger mixed genome database, such as, a non-redundant database. Therefore, proteins identified in both searches are more likely to be correct. Subsequently, a false identification rate can be estimated by the percentage of proteins which are identified in the mixed genome database, but do not belong to the model organism. Incidentally, an iTRAQ study looking at three different organisms, *Synechocystis*, *Sulfolobus solfataricus* and *Saccharomyces cerevisiae* [82] adopted the double-database search strategy, and reported a false positive percentage very similar to the fraction of proteins that were identified in the single genome database only [82]. However, there are disadvantages to this approach. It is not clear whether all non-overlapping proteins are definitely incorrect, and these would be regarded as false negatives. Additionally, in the larger mixed genome database search, the assignment of proteins to closely related organisms which share a high degree of homology in proteins is very possible and would lead to false positives. An alternative approach to acquire an indication of the false positive identification rate is to use a decoy database [91, 98]. A search against the proteome of a distant model organism can be used [160], followed by manual inspection of the identified peptides. In this study, a decoy database is used but consisting of reversed proteome sequences [98], created using Perl script. Two reversed cyanobacteria proteomes were included (*Nostoc punctiforme* PCC73102, 7672 open reading frames (ORF's) and *Nostoc* sp. PCC7120, 6130 ORF's) as well as reversed *Synechocystis*, 6130 ORFs. All were retrieved from NCBI RefSeq, March 2007. An analysis of variation amongst biological replicate samples was used as a guide to select only significantly differentially regulated proteins (refer to section 3.3.1).

3.2.6 RNA extraction

Total RNA extraction was achieved using a modified version of RNeasy kit (Qiagen, West Sussex, U.K). In order to generate time-series information, 50 ml cells were harvested (centrifugation at 5000 x *g* for 5 min at 4°C) during mid-exponential phase in (i) no salt (t = 0), (ii) 6% salt (t = 2 h after salt addition), (iii)

6% salt (t = 24 h) and finally (iv) 6% salt (t = 9 days). Biological replicate samples were treated in parallel. These cells were further washed three times with sucrose based buffer for 3 min in a cell disruptor (Scientific Industries, NY, USA) followed by 3 min centrifugation at 13,000 rpm. Following the final centrifugation step, cells were resuspended in 350 µL Buffer RLT (with 2-mercaptoethanol included). 200 µg of 0.425-0.600 mm glass beads were added to the sample. Cell disruption was achieved by 1 min on a cell disruptor followed by one min cooling on ice. The cooling step was incorporated to reduce the impact of possible RNases present in cells. This disruption/cooling cycle was performed four times. Samples were then centrifuged for 3 min at 13,000 rpm to pellet cell debris. The resulting supernatant which contains nucleic acids, was removed and placed into a fresh microcentrifuge tube. 250 µL of fresh 100% ethanol was added to the samples and mixed prior to transferring the entire mix to the RNeasy Mini Spin Column. The column was centrifuged briefly (15 sec) at 13,000 rpm and the flow-through discarded. 700 µL of Buffer RW1 was then added to the column and centrifuged under the same conditions and time. Again the flow-through was discarded and a new collection tube added to the spin column. Another two washes were performed using 500 µL of Buffer RPE (allows binding of RNA to silicone membrane) centrifuging at 13,000 rpm firstly for 15 sec, and secondly for 2 min. All flow-through was discarded and another 1 min spin ensured the membrane in the spin column was mostly dry. Following the replacement of the collection tube with a fresh microcentrifuge tube, 30 µL of RNase-free water was added to the spin column membrane. An incubation of 1 min was allowed before a final centrifuge step at 13,000 rpm for 1 min. The extracted RNA (in 30 µL) was stored at -80°C for future DNase treatment, quantitation and quality verification.

3.2.7 DNase treatment, RNA quantitation and quality check

Contaminating DNA in the RNA samples was removed with application of Turbo DNA-*free* kit (Ambion, AB, Framingham, USA). To 30 µL RNA extract, 3 µL of 10 x TURBO DNase buffer and 1 µL of TURBO DNase enzyme were added, and left to incubate at 37°C for 30 min. Following this period, 4 µL of resuspended DNase inactivation reagent was added and the sample was mixed well. 2 min of

incubation at room temperature with occasional mixing was sufficient to inactivate the DNase enzyme. RNA was recovered in the supernatant after centrifuging the sample at 10,000 x *g* for 90 sec. It is imperative to ensure that no DNase inactivation reagent was recovered with the RNA as it can inhibit down-stream processes.

The integrity and size distribution of the DNA-free RNA extract was checked using agarose gel electrophoresis. 1.2% agarose gels were made up using 1 x T.A.E. (Tris-acetate-EDTA) buffer and 0.5 μg ml^{-1} ethidium bromide. A 1 μL aliquot of RNA was removed and added to 2 μL 5 x DNA loading dye (Bioline, London, UK) and 7 μL RNase-free water. This mixture was incubated at 65°C for 5 min, and then immediately placed on ice prior to gel loading. Gels were run at 80 V for 30 min in 1 x T.A.E. buffer. Gel images were captured using a digital camera (Canon, Tokyo, Japan) and UV box (model GVM30, Syngene, Cambridge, UK).

RNA quantitation was performed using RNA-specific fluorescence with the Qubit™ Quant-iT system (Invitrogen, U.K.) using manufacturer's instructions.

3.2.8 cDNA synthesis

Synthesis of cDNA from the total RNA extract was achieved using a Quantitect kit (Qiagen). This kit comprises of a brief DNA elimination step, in the event that undetectable (by agarose gel electrophoresis) amounts of DNA remain in the sample. RNA samples were thawed on ice and a sufficient volume pertaining to 300ng total RNA was added, to 2 μL 7 x gDNA Wipeout Buffer and RNase-free water, to make the total volume up to 14 μL. This mixture was incubated for 2 min at 42°C, and then immediately placed on ice. For cDNA synthesis, 1 μL of Quantiscript Reverse Transcriptase (RT), 4 μL Quantiscript RT Buffer (x 5) and 1 μL RT Primer Mix were added, keeping the mix on ice. Reverse transcription occurred at 42°C for 25 min, and the enzyme was subsequently inactivated by incubation at 95°C for 3 min. Three negative controls were included- (i) cDNA synthesis without RT enzyme, (ii) cDNA synthesis without primers and (ii) cDNA synthesis without RNA. The resulting cDNA was quantified using a

spectrophotometer (UltraSpec 2100 Pro), diluted to 1500 ng μL^{-1} and stored at 4°C for RT-qPCR.

3.2.9 Primer design and RT-qPCR

Primers were designed using Primer Express Software version 3.0 (AB) with optimal annealing temperature set to 58-61°C, and length 50-105 bp. Full sequences are given in Table 3.2 with optimised running conditions. Genes of interest were selected for primer designing following protein quantitation (iTRAQ) analysis and a survey of previous studies. The ribosomal gene 23S (*rrn23Sa*) was used as the endogenous control gene [291], and suitability of this gene (insignificant variance across the test conditions) was checked using the 'Absolute quantitation' mode on the 7500 FAST real-time PCR system (AB). This mode was also used to optimise PCR efficiency for all primer pairs to within 90-110% and ensure non-specific amplification was not occurring, for example, due to primer dimers. The 'Relative quantitation' mode was used to calculate the fold expression changes in genes relative to the calibrator sample (0% salt, t = 0). Power SYBR green (AB) was implemented for RT-qPCR reactions which contains internal ROX dye to normalise any resulting fluorescence in each reaction. The PCR master mix was created using the amounts given in Table 3.3

Table 3.2: Primer (oligonucleotide) sequences designed for each protein-coding gene (T_a = annealing temperature).

gi accession	Protein name	Gene name	Forward sequence	No. of bases	Reverse sequence	No. of bases	T_a	Amplicon length
16329588	16.6 kDa small heat shock protein	*Hsp17*	ACCGGCATCGTAATTAGCTTTAA	23	CCGCCGGGTTATTCCTGTA	19	58	67
16331408	ABC1-like	*UbiB*	TCGCCGGTTTTCAACTTACC	20	TATCTCCGGAGCCCATTGC	19	61	67
16330243	Anti-sigma B factor antagonist	*IcfG*	CGCCAACTCAGCCCAATTA	19	CACGGCCATGGTGTCGAT	18	60	64
16331161	ATP-dependent Clp protease proteolytic subunit	*ClpP*	ACTTGGGAATGCCGCTTTT	19	TCAATGCCGACCTGTTGCT	19	60	55
16332004	Bicarbonate transporter	*CmpA*	GTGCCTCCCGCTGAAACA	18	TCTCCAGTGCTAAAAGCATCCA	22	61	65
16330088	Cell division protein FtsZ	*FtsZ*	AAATGGACGGGTGACAATGC	20	CCGAGGTGGCCAAAGAAAT	19	60	56
16330514	Circadian clock protein KaiC	*KaiC*	GGCCGCCAGGGTAGTTTT	18	CGTGGCAGTAGCATTTTAGTAACC	24	61	60
16330002	Co-chaperonin GroES	*GroES*	CCTACTCCCCTGTGGAAGTCAA	22	GTGCCGGCATACTTGGAATAG	21	59	61
16329356	Elongation factor EF-G	*Fus*	TCTTGGGTTTTGAAGGGTTGA	21	TAGCCGGCACTGGCTTCT	18	59	55
16330539	Flavodoxin	*IsiB*	TCGCCGCCCATTTCTTT	17	CTCAAACCGGCAACACTGAA	20	58	58
16331386	Fructose-bisphosphate aldolase	*CbbA*	ACCACGGGAAGCTTGCAA	18	CATGCAGGCCGCTGATG	17	60	55
16331104	Glutamate--ammonia ligase	*GlnN*	CCGGACGGCTCTTCTTTTC	19	AGCCCCTAGCCTCAAAGGTATC	22	61	58
16330473	Glutathione peroxidise	*GshP*	AGGTGGTGGCTCGCTTTG	18	GCCGCCTTGAGATTAGTGTCA	21	60	58
16330943	Glycerol-3-phosphate dehydrogenase	*GlpA*	GACAATGGCAACGGGTTCTC	20	GCCCCTGGGTAGATGAAGTTT	21	61	62
16332299	Hypothetical protein sll1106	sll1106	CACAACGGCGGCATCTTC	18	GCCCTGGCCAAGCTACAA	18	59	57
16331261	Molecular chaperone DnaK	*DnaK*	ATCCCGAATGGCATTTTCC	19	CCGGGCCAAGTTTGAAGAA	19	58	76
16330429	Periplasmic phosphate binding protein	*PstS*	TTGGTGCCACCGATGCT	17	CCACACCCCGTTTCATCTGT	20	61	60
2499503	Phosphoglycerate kinase	*Pgk*	CCTGCCCACGGACGTAGTAG	20	CAATGTCTAGACCCATCCAACCA	23	60	104
131288	Photosystem II 44 kDa reaction center protein	*PsbC*	CAACCAAGGACCACGGAAGT	20	ACCTGATGCGTTCTCCTTCTG	21	58	80
16331151	Ribosome releasing factor	*Frr*	CGCATCCCGGCGAATAT	17	GCAGGTAAATTGGCGGAAGA	20	60	57
16330362	SOS function regulatory protein	*LexA*	GGACGCAGAGGAAGTGGAA	19	ATGACCCTGAGGGCAAAAAC	20	58	75
16330619	Superoxide dismutase	*Sod*	AGCACCCTGGAGTTCCATCA	20	AGTACCCGCCACTGCATTGT	20	59	75

Table 3.3: PCR master mix per reaction.

Component	Volume in one reaction (µL)	Final concentration
Power SYBR green (2x)	10	1x
Forward primer	variable	50-900 nM
Reverse primer	variable	50-900 nM
Template	2	8ng
Water	variable	-
TOTAL	20	-

The amplification program was run in the FAST mode with the following steps (i) 95 °C for 10 min, (ii) quantitation (x 40 cycles) (95 °C for 15s, and 60 °C for 1 min). A dissociation step was included to verify the specificity of each reaction by displaying a change in fluorescence over an increase in temperature. All reactions included a technical replicate, which enabled standard deviations of quantitations to be calculated, and contamination was assessed by adding no template controls for each gene. Amplification of the endogenous control gene was performed on each plate for extra accuracy and reliability in the quantitations. Data capture was accomplished using FAST system SDS software version 1.4. Relative quantitation of gene expression was achieved using the $\Delta\Delta C_T$ method [292]. Interpretation of the cycle number where fluorescence reaches a particular threshold produced a relative expression of a given gene, which was normalised against the endogenous control. Note that quantitative accuracy using this method relies on high PCR reaction efficiency (90-110%). Furthermore, PCR purity was checked on 1.5% agarose gels

3.3 Results and discussion

Synechocystis cells were successfully cultured in both conditions of low salt (0% w/v) and high salt (6% w/v) with exponential growth rates of 0.26 ± 0.01 and 0.11 ± 0.01 d^{-1} respectively. After addition of 6% salt, mid-exponential growth phase occurred at 9 days and therefore ensured cells were fully adapted to the new conditions. Analysis of molecular function in these cells is presented and discussed in three main parts. Firstly, data from each workflow was analysed and combined. The proteins of interest identified as playing a role in *adaptation* to salt stress are then discussed. Discussions will include published data which has been generated from high-throughput global studies of salt stress in this organism at the transcriptome and proteome level. Finally, the expression of

genes which were selected for analysis in a temporal study will be interpreted in relation to the measured protein levels.

3.3.1 Protein data analysis

3.3.1.1 MS data analysis

Using a 95% protein confidence cut-off, 207 unique proteins were identified by 6563 distinct peptides using Dionex nano-LC and AB ESI-qQ-TOF-MS/MS with ProteinPilot™ (Workflow I). 77% (24,374) of total spectra were successfully matched to peptides at or above the same confidence level. In order to produce accurate protein quantitations, only peptides which were referenced to spectra with ≥ 70% confidence intervals, were used. A bias correction was performed assuming the median of the proteome is towards unity.

Biological and technical variations were calculated and expressed as CV. Firstly, ratios of 115:114 and 116:114 (6% salt compared to 0% salt in both) were compared using $\log_{10}$ ratios and the equation below, where SD is the standard deviation of $\log_{10}$ ratios.

$$CV = \sqrt{((10^{SD^2})-1)} \qquad \text{Equation 3.1}$$

The mean average CV across all 207 proteins was 0.20 (± 20%). This is similar to a study in which 3 data sets across two iTRAQ experiments gave an average biological variation of ± 0.25 [83]. Secondly, it is expected that the ratio of biological replicate samples 115 and 116 would be close to 1. Therefore, the average ratio was calculated by generating the average of the $\log_{10}$ ratios, using the error factor generated using the ProteinPilot™ algorithm as a weight, and thereby obtaining a weighted average. A comparison of 116 (6% salt) to 115 (6% salt) using this method gave an average ratio of 0.98. Finally, the average RV of biological replicates (115/114 versus 116/114) was calculated using Equation 3.2 [84].

$$RV = \text{Absolute deviation of } \frac{Ratio_{xy}}{Ratio_{zy}} \qquad \text{Equation 3.2}$$

Where x and z denote test phenotypes (6% salt) and y denotes control phenotype (0% salt). The absolute difference between the replicate samples (115 and 116) and 114

should be close to unity (1), and therefore deviations here can be used as an estimate of error in the replicate samples. A 4% RV was calculated using this data set.

The programs used for protein identification return the best peptide match for each MS/MS spectrum within the database, and these may not be a perfect match. Moreover, peptide-to-MS/MS spectra matches may be incorrect (false positives), most likely due to coincidental similarity in the fragmentation patterns [293]. To gain a perspective on the rate of false positive measurement, files were run against a decoy database. In total, 10,722 spectra were assigned to valid peptides ($\geq$ 80% confidence interval) in the correct database search and 24 spectra were identified using the decoy database. Using the method described previously by Elias *et al.* [98], the false positive rate was measured to be 0.4%. Therefore, a low percentage (0.4%) of peptide-to-MS/MS spectra matches are possibly incorrect [293].

The combination of Agilent 1200 HPLC and 6510 mass spectrometer with Spectrum Mill version A.03.03 in Workflow II identified 339 proteins. Spectrum Mill implements a pre-processing step which selects only good quality spectra and merges identical spectral information. Autovalidation of results using an in-built algorithm means only peptides which exceed a set score threshold and percentage peak intensity threshold are used to calculate protein quantitations. The bias correction towards unity was calculated manually. The CV was calculated to express biological variations. Ratios of 115:114 and 116:114 were compared using $\log_{10}$ ratios as performed with data generated in Workflow I and the mean average CV across all 339 proteins was 0.32. The same criteria described above were used for setting a threshold value for significantly differentially expressed proteins. The RV between the two calculated ratios of biological replicates (115/114 and 116/114) was 8.3%.

3.3.1.2 Combining results

A comparison was made between the protein identifications and quantitations using in Workflow I and Workflow II. A total of 381 unique protein identifications were gained through both data sets. 165 proteins were identified using both methods, but the majority were unique to workflow II (174). Figure 3.1 shows the distribution of protein identifications.

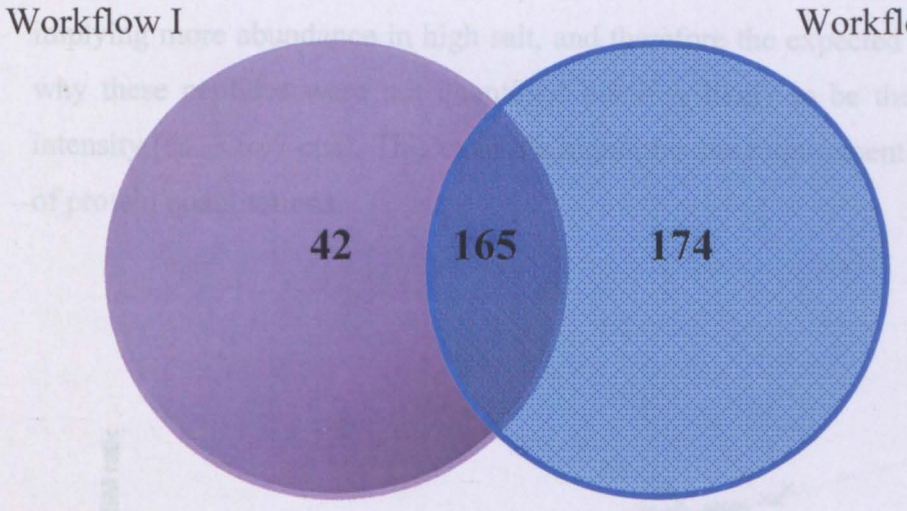


Figure 3.1: Venn diagram to illustrate unique and shared proteins identified using Workflows I and II.

A correlation was made for all proteins which were identified *and* quantified in both workflows (155) using labels 114 and 115 only (Figure 3.2). A Pearson correlation coefficient of 0.74 showed good correlation (where 1 = complete positive correlation and 0 = no correlation). In nine cases (5.8%), quantitations did not correspond well, for example, elongation factor EF-G was deemed to be relatively unchanged in 6% salt compared to 0% salt (1.06-fold) using Workflow I, whereas the same protein was 2.91-fold increased in 6% salt using Workflow II. Upon closer inspection, it is possible to see that the error factor associated with this protein generated in Workflow I is fairly high (4.01) as well as the probability (P)-value (0.70). The quantitation was derived from a single peptide identified in 2 MS/MS experiments. In contrast, 3 unique peptides were identified and 5 quantitations were >2 fold regulated in Workflow II.

In another example, compatible solute synthesis enzyme GG-phosphate synthase was 1.56 and 1.72-fold down-regulated in high salt in Workflow I. This enzyme is expected to increase in high salt because of the requirement for osmotic balance in salt adapted cells, and previous studies as well as results in Workflow II (4.86 and 6.21-fold increase) also suggest this [174, 191]. Inspection at the peptide level reveals that the only peptides identified with high confidence (>70%) were not actually quantified, and therefore the quantitation given for this protein was given for one peptide which had a confidence level

of 60% (below our threshold). Examination of the spectra for the two peptides identified with high confidence reveals a much larger peak for labels 115 and 116 compared to 114, implying more abundance in high salt, and therefore the expected result. It is not certain why these peptides were not quantified but it is likely to be the relatively low signal intensity (ca. 3 to 7 cps). This example highlights the requirement for manual inspection of protein quantitations.

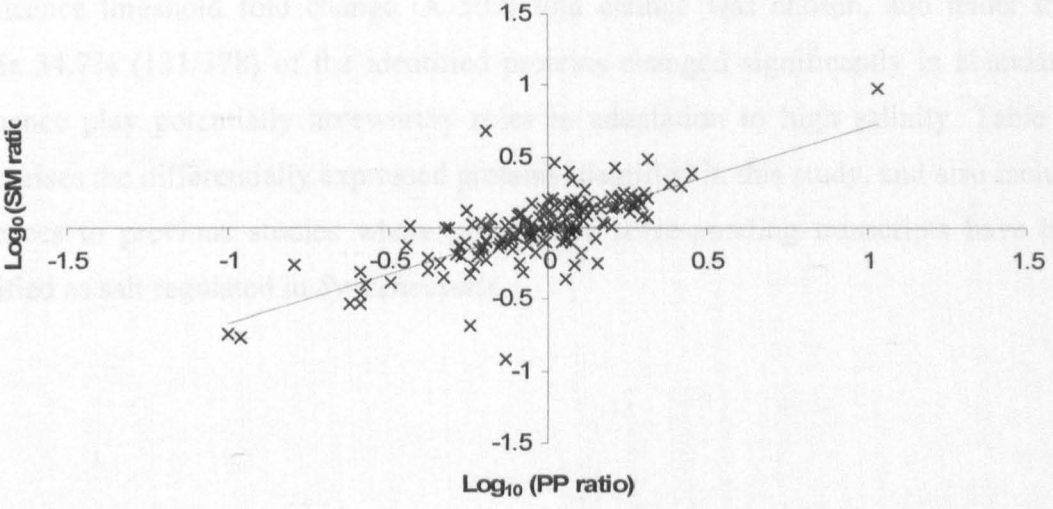


Figure 3.2 Correlation of quantitations for 155 proteins across the two workflows. R = 0.74 (PP- ProteinPilot®, SM- SpectrumMill).

Unique proteins were selected in order to generate a final list of proteins which would be used to interpret biological changes in *Synechocystis* cells in response to salt. Protein identifications and quantitations were only accepted after intense scrutinising by manual inspection of the peptide quantitations, mass spectra intensity and quality, as well as examination of biological replicates. This produced a list of 378 protein candidates for further interpretation, which equates to 11.9% of total predicted protein-coding genes. The full iTRAQ results and this final list are presented in Appendix A.

3.3.2 Cellular adaptations through protein expression

Differentially expressed proteins from a combined list of proteins generated by careful inspection of both workflows were organised into functional groups for further analysis. For proteins to be considered as significantly differentially expressed, a suitable threshold value needs to be assigned. In addition, the statistical values awarded to each quantitation

are important, including the P-value, error factor and the detection of ≥2, or the same peptide identified over more than one MS/MS experiment (as suggested in the General Guideline for Proteomics Data Publication (http://www.mcponline.org/). Moreover, differential regulation amongst both biological replicates in the iTRAQ experiment was a pre-requirement for further analyses (and therefore two ratios are quoted in the discussion). CV and RV calculations presented in section 3.3.1 together with previous iTRAQ studies [83, 84], were all considered for appointing a suitably stringent significance threshold fold change. A 50% fold change was chosen, and under these criteria 34.7% (131/378) of the identified proteins changed significantly in abundance, and hence play potentially noteworthy roles in adaptation to high salinity. Table 3.4 summarises the differentially expressed proteins identified in this study, and also includes references to previous studies where proteins or corresponding transcripts have been identified as salt regulated in *Synechocystis*.

Table 3.4: Differentially expressed proteins in *Synechocystis* cells adapted to 6% salt, quantified using isobaric tags.

gi accession	Name	Biological rep. 1		Biological rep. 2		Functional groups	Seen previously at protein level		Seen previously at transcript level	
		Ratio	CV	Ratio	CV		reference/ no	short/ long term	reference/ no	short/ long term
INCREASED EXPRESSION IN 6% ACCLIMATED CELLS										
16329461	ferredoxin-sulfite reductase	1.82	0.24	1.57	0.24	Amino acid biosynthesis	no		no	
16330927	sulfate adenylyltransferase	1.55	0.28	1.51	0.24	Amino acid biosynthesis	no		no	
16331104	glutamate--ammonia ligase	2.57	0.30	2.72	0.17	Amino acid biosynthesis	no		no	
16331651	cyanate hydratase	1.52	0.32	1.50	0.25	Amino acid biosynthesis	no		no	
16332126	ketol-acid reductoisomerase	1.52	0.21	1.57	0.18	Amino acid biosynthesis	no		no	
16331650	ferredoxin--nitrite reductase	1.86	0.18	1.69	0.28	Amino acid metabolism	[61]	long term only	[60]	short term only
16330763	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase	1.61	0.29	1.64	0.27	Cell envelope	no		no	
16332281	heat shock protein 90	2.24	0.22	1.90	0.20	Cellular processes	no		both	short term only
16330619	superoxide dismutase	2.83	0.13	2.83	0.26	Cellular processes	[61]	short and long term	[3]	short term only
16331261	molecular chaperone DnaK	1.59	0.20	1.83	0.20	Cellular processes	[61]	short and long term	[3]	
16329588	16.6 kDa small heat shock protein	1.75	0.17	2.10	0.35	Cellular processes	[61]	short term only	[60]	short term only
16330002	co-chaperonin GroES	2.61	0.28	2.93	0.35	Cellular processes	[61]	short term only	[60]	short term only
16329392	4-alpha-glucanotransferase	2.78	0.00	3.00	0.00	Central intermediary metabolism	[61]	long term only	no	
16330944	alpha,alpha-trehalose-phosphate synthase	4.86	0.00	6.21	0.13	Central intermediary metabolism	no		no	
16332282	glucose-1-phosphate adenylyltransferase	1.62	0.30	1.66	0.25	Central intermediary metabolism	no		no	
16331386	fructose-bisphosphate aldolase	1.52	0.27	1.55	0.27	Energy metabolism	no		[60]	short term only
16331115	phosphoglycerate kinase	2.68	0.26	2.99	0.24	Energy metabolism	[61]	Long term	no	
16331208	dihydrolipoamide acetyltransferase	2.18	0.31	2.23	0.14	Energy metabolism	no		no	
16330943	glycerol-3-phosphate dehydrogenase	1.95	0.00	1.52	0.00	Fatty acid, phospholipid and sterol metabolism	no		[60]	short and long term
16331180	hypothetical protein sll1863	10.72	0.25	4.74	0.33	Hypothetical	[61]	short and long term	[60]	short and long term
16331737	hypothetical protein slr0846	1.50	0.34	1.70	0.24	Hypothetical	no		[3]	
16329942	hypothetical protein slr1894	1.99	0.19	1.90	0.17	Hypothetical	[61]	short and long term	[60]	short and long term
16330945	hypothetical protein slr1670	4.17	0.27	5.38	0.40	Hypothetical	no		[60]	short and long term
16329773	hypothetical protein slr1039	2.14	0.00	2.67	0.00	Hypothetical	no		no	
16330217	hypothetical protein sll0861	3.57	0.01	4.53	0.05	Hypothetical	no		no	
16330542	hypothetical protein sll0245	1.75	0.36	1.85	0.38	Hypothetical	no		no	

16330922	hypothetical protein slr1161	1.69	0.13	1.70	0.17	Hypothetical	no		no	
16331360	hypothetical protein sll0319	2.41	0.24	2.55	0.24	Hypothetical	no		no	
16332006	hypothetical protein slr0042	1.96	0.39	1.98	0.32	Hypothetical	no		no	
16332072	hypothetical protein slr0551	2.29	0.36	1.74	0.36	Hypothetical	no		no	
16330514	circadian clock protein KaiC	3.23	0.41	3.58	0.44	Other categories	no		[60]	long term
16332149	aconitate hydratase	1.65	0.06	1.66	0.13	Other categories	no		no	
131288	photosystem II 44 kDa reaction center protein	1.82	0.74	1.80	0.32	Photosynthesis and respiration	[224]		no	
16330219	apocytochrome f precursor	1.93	0.19	1.69	0.07	Photosynthesis and respiration	no		no	
16331216	light repressed protein	2.62	0.20	2.52	0.25	Regulatory proteins	[224]		no	
16329957	DNA-directed RNA polymerase beta subunit	1.60	0.35	1.70	0.32	Transcription	no		no	
16329913	30S ribosomal protein S9	2.01	0.20	1.81	0.21	Translation	[61]	short term only	[3]	short term only
16329925	30S ribosomal protein S5	1.74	0.20	1.70	0.17	Translation	[61]	short term only	[3]	short term only
16329927	50S ribosomal protein L6	1.54	0.35	1.67	0.27	Translation	[61]	short term only	[3]	short term only
16329935	30S ribosomal protein S3	1.68	0.25	1.71	0.25	Translation	[61]	short term only	[3]	short term only
16329941	50S ribosomal protein L3	2.99	0.84	2.27	0.34	Translation	[61]	short term only	[3]	short term only
16330008	50S ribosomal protein L7/L12	2.46	0.19	1.88	0.23	Translation	[61]	short term only	[3]	short term only
16329482	translation initiation factor IF-2	2.38	0.26	2.09	0.09	Translation	no		[60]	short term only
16329288	translation initiation factor IF-2	1.61	0.05	1.66	0.15	Translation	no		[60]	short term only
16330189	RNA-binding protein	1.59	0.39	1.75	0.27	Translation	[61]	short term only	[60]	short term only
16332012	RNA-binding protein	2.21	0.11	2.22	0.00	Translation	[61]	short term only	[60]	short term only
16329356	elongation factor EF-G	2.91	0.21	3.17	0.38	Translation	no		no	
16330433	peptidyl-prolyl cis-trans isomerase	2.07	0.17	2.40	0.18	Translation	no		no	
16331631	glutamyl-tRNA synthetase	1.83	0.15	2.32	0.08	Translation	[61]	long term only	no	
16332123	elongation factor EF-2	1.70	0.17	1.79	0.16	Translation	no		no	
16329924	50S ribosomal protein L15	2.58	n/a	1.95	0.36	Translation	no		no	
16331324	50S ribosomal protein L35	2.29	0.28	2.63	0.32	Translation	no		no	
16330084	nitrate transport 45kD protein	1.83	0.14	1.52	0.20	Transport and binding	[224]	long term only	[3]	short term only
16331408	ABC1-like	4.27	0.11	4.93	0.16	Transport and binding	no		no	
16331744	ABC transporter subunit	2.35	0.41	2.66	0.30	Transport and binding	no		no	
DECREASED EXPRESSION IN 6% ACCLIMATED CELLS										
16330243	anti-sigma B factor antagonist	2.70	0.51	2.30	0.22	Unknown	no		no	
16330576	acetolactate synthase	1.70	0.06	1.75	0.27	Amino acid biosynthesis	no		no	
16331281	2-isopropylmalate synthase	2.09	0.14	2.40	0.73	Amino acid biosynthesis	no		no	
16330473	glutathione peroxidase	1.74	0.28	1.99	0.24	Biosynthesis of cofactors, prosthetic groups, and carriers	[61]	opposite	no	

3023789	Probable dihydroneopterin aldolase (DHNA)	1.74	0.31	1.85	0.36	Biosynthesis of cofactors, prosthetic groups, and carriers	no		no	
16331030	molybdopterin biosynthesis protein MoeB	1.64	0.48	1.58	0.26	Biosynthesis of cofactors, prosthetic groups, and carriers	[224]	opposite	no	
16331608	GTP cyclohydrolase I	2.61	0.11	3.54	0.24	Biosynthesis of cofactors, prosthetic groups, and carriers	no		no	
16331297	cell division cycle protein	1.62	0.27	1.77	0.36	Cell division down as is growth	no		no	
16330088	cell division protein FtsZ	1.72	0.21	1.61	0.21	Cellular processes	no		[3]	short term increase
16331493	heat shock protein; GrpE	1.54	0.21	1.87	0.25	Cellular processes	no		no	
16329675	urease alpha subunit	1.64	0.23	1.90	0.12	Central intermediary metabolism	no		no	
16330143	glycogen phosphorylase	2.68	0.21	1.83	0.04	Central intermediary metabolism	[61]		no	
16330360	OxPPCycle gene	1.52	0.44	1.70	0.26	Central intermediary metabolism	no		no	
16331023	inorganic pyrophosphatase	2.56	0.27	2.23	0.15	Central intermediary metabolism	no		no	
16331412	bifunctional GMP synthase/glutamine amidotransferase	1.76	0.26	1.99	0.24	Central intermediary metabolism	no		no	
16331862	polyphosphate kinase	1.70	0.35	2.20	0.76	central intermediary metabolism	no		no	
16331645	hypothetical protein sll0887	2.05	n/a	1.56	0.22	DNA replication, restriction, modification, recombination, and repair	no		no	
2506772	Isocitrate dehydrogenase [NADP]	1.70	0.08	1.61	0.18	Energy metabolism	no		no	
16329656	glutamate decarboxylase	2.79	n/a	1.69	0.18	Energy metabolism	no		no	
16330043	glucose-6-phosphate 1-dehydrogenase	3.74	0.59	2.33	0.42	Energy metabolism	no		no	
16331307	6-phosphogluconate dehydrogenase	2.12	0.14	1.90	0.12	Energy metabolism	no		no	
16331394	ribulose bisphosphate carboxylase small SU	1.90	0.19	2.15	0.19	Energy metabolism	no		no	
16331464	citrate synthase	1.58	0.15	1.57	0.12	Energy metabolism	no		no	
16329546	hypothetical protein slr0670	1.52	0.48	2.36	0.24	Hypothetical	no		no	
16329841	hypothetical protein slr2005	3.84	0.03	2.33	0.03	Hypothetical	no		no	
16330234	hypothetical protein sll1305	2.56	0.26	2.20	0.23	Hypothetical	no		[3]	
16330239	hypothetical protein slr1852	3.55	0.19	3.59	0.08	Hypothetical	no		[3]	
16330242	hypothetical protein slr1855	4.18	0.19	4.68	0.15	Hypothetical	no		[3]	
16330319	hypothetical protein slr2144	3.42	0.29	2.57	0.03	Hypothetical	[61]		no	
16330744	hypothetical protein slr1338	1.97	0.23	1.90	0.11	Hypothetical	no		no	
16330813	hypothetical protein sll1873	1.89	0.26	1.90	0.31	Hypothetical	no		no	
16331085	hypothetical protein sll0272	1.98	0.08	3.22	0.19	Hypothetical	no		no	

16331120	hypothetical protein sll0359	2.21	0.16	2.09	0.15	Hypothetical	no		no	
16332260	hypothetical protein sll0446	1.55	0.42	1.66	0.03	Hypothetical	no		no	
16332268	putative phosphoketolase	3.76	0.17	3.75	0.16	Hypothetical	no		no	
16332299	hypothetical protein sll1106	1.57	0.17	1.51	0.15	Hypothetical	no		[60]	opposite
38505793	hypothetical protein sll8018	1.62	0.11	1.65	0.02	Hypothetical	no		no	
2499211	protein drgA	2.32	0.06	2.51	0.18	Other categories	no		no	
16329322	potential FMN-protein	2.03	0.57	2.06	0.27	Other categories	no		no	
16329908	short chain dehydrogenase	2.28	0.07	2.23	0.08	Other categories	no		no	
16329971	rehydrin	1.60	0.22	1.57	0.18	Other categories	no		no	
16330683	hydrogenase large subunit	1.62	0.19	1.57	0.14	Other categories	no		no	
16330967	soluble hydrogenase 42 kD subunit	1.83	0.05	1.58	0.12	Other categories	no		no	
16332300	zinc-containing alcohol dehydrogenase family	1.87	0.18	2.23	0.35	Other categories	no		no	
16329366	carbon dioxide concentrating mechanism	2.25	0.16	2.63	0.36	Photosynthesis and respiration	no		no	
16329478	allophycocyanin-B	1.81	0.21	2.54	0.10	Photosynthesis and respiration	no		[3]	
16329710	phycobilisome rod-core linker polypeptide	1.85	0.29	1.86	0.24	Photosynthesis and respiration	no		[3]	
16329820	phycocyanin associated linker protein	3.86	0.23	4.66	0.17	Photosynthesis and respiration	no		[3]	
16329821	phycocyanin associated linker protein	3.88	0.34	4.25	0.29	Photosynthesis and respiration	no		[3]	
16329822	phycocyanin associated linker protein	3.79	0.35	4.80	0.41	Photosynthesis and respiration	no		[3]	
16329823	phycocyanin a subunit	3.13	0.21	4.32	0.28	Photosynthesis and respiration	no		no	
16329824	phycocyanin b subunit	2.19	0.29	2.51	0.25	Photosynthesis and respiration	no		no	
16330467	allophycocyanin b	1.56	0.22	1.68	0.20	Photosynthesis and respiration	no		[3]	
16330539	flavodoxin	1.83	0.06	2.06	0.19	Photosynthesis and respiration	[224]	opposite	no	
16331106	cytochrome c550	1.93	0.23	2.16	0.18	Photosynthesis and respiration	no		no	
16331547	chloroplast membrane-assoc 30 kD protein	1.66	n/a	1.63	n/a	Photosynthesis and respiration	no		no	
16332085	NADH dehydrogenase subunit I	1.59	0.16	1.75	0.08	Photosynthesis and respiration	no		no	
16329790	CheA like protein	3.40	n/a	2.23	0.31	Regulatory	no		no	
16330362	SOS function regulatory protein	1.66	0.26	1.61	0.34	Regulatory	no		no	
16331356	transcriptional regulatory protein	1.72	n/a	1.73	n/a	Regulatory	no		no	
16331742	sensory transduction histidine kinase	2.00	0.30	2.96	0.38	Regulatory	no		no	
16329303	peptidyl-prolyl cis-trans isomerase B	1.85	0.05	2.04	0.09	Translation	no		no	
16329479	S-adenosylmethionine synthetase	1.63	0.09	1.52	0.37	Translation	no		no	
16329894	leucyl-tRNA synthetase	1.59	0.15	2.01	0.11	Translation	no		no	
16329909	50S ribosomal protein L25	2.00	0.20	1.83	0.14	Translation	no		no	
16330739	30S ribosomal protein S2	2.09	n/a	3.02	n/a	Translation	no		no	
16331151	ribosome releasing factor	1.55	0.07	1.99	0.11	Translation	no		no	
16331161	ATP-dependent Clp protease proteolytic SU	2.38	0.37	2.29	0.27	Translation	no		no	

16331163	aminopeptidase P	2.28	0.15	1.81	0.57	Translation	no		no	
16331761	threonyl-tRNA synthetase	1.95	0.09	1.79	0.15	Translation	no		no	
16329434	iron transport protein	1.75	0.19	1.99	0.23	Transport and binding	no		no	
16330429	periplasmic phosphate binding protein	10.07	0.21	10.38	0.17	Transport and binding	[224]	opposite	no	
16330540	iron-stress chlorophyll-binding protein	20.65	0.34	29.62	0.20	Transport and binding	no		no	
16331793	periplasmic iron-binding protein	9.28	0.56	11.08	0.54	Transport and binding	[224]	opposite	no	
16332004	bicarbonate transporter	2.09	0.36	3.62	0.24	Transport and binding	[224]	opposite	[61]	opposite
16332097	bacterioferritin	1.61	0.17	1.72	0.14	Transport and binding proteins	no		no	
16330022	poly(3-hydroxyalkanoate) synthase	2.81	n/a	3.47	n/a	Unknown	no		no	

: Functional groups designated by Kazusa (http://bacteria.kazusa.or.jp/cyanobase/Synechocystis/)

* Studies in which protein or transcript levels have previously been identified as differentially expressed (*opposite* labelled proteins refer to proteins where the direction of expression is opposite (i.e. reduced instead of increased expression and vice versa).

In the following sections the differentially expressed proteins are discussed in terms of previous findings, and care is taken to interpret these effectively because of the differences in growth and test conditions (e.g. [3, 60]). Not only do the techniques and equipment employed differ (which all contribute to variation) the salt concentrations investigated, significance thresholds set for differential protein expressions, and the time allowed for cells to adapt to adverse conditions all vary. Before discussion of each group of proteins, a brief summary of the present knowledge of identified cellular responses is given (in bold type), followed by a short hypothesis for this study (italics bold type).

3.3.2.1 Stress proteins

Synechocystis* responds to ionic, osmotic and oxidative stress caused by high salt (up to ~4%) conditions by the synthesis of specific 'stress proteins', for example molecular chaperones or anti-oxidative enzymes [60, 61]. *These proteins are expected to be increased here although not all are expected to be in higher abundance at 9 days adaptation.

In this study, relative quantitations show that several of these proteins are also required in elevated amounts in *Synechocystis* cells adapted (for exactly 9 days) to 6% salinity. A 2.24 and 1.90-fold induction (Table 3.4) was seen in heat shock protein 90 (Hsp90) and this particular stress protein has not been identified as salt-induced in previous cyanobacterial proteomics experiments in the short or long-term. Hsp90 contributes to an array of cellular processes, including protein degradation, protein folding and even signal transduction [294]. Its gene transcript levels have been seen to increase in response to ~3% salt [3] and 4% salt [60] in the short-term. In the latter study by Marin *et al.* [60], transcript levels were not induced after 5 days and therefore post-transcriptional control in *adaptation* to high salt is possible, or the prospect that the extent of the salt stress (6% compared to 4% NaCl) dictates the amount of stress protein required.

Anti-oxidative enzyme superoxide dismutase was up-regulated over 2-fold in all experiments (Table 3.4) and this response has also been seen in the short-term and after 5 days at the protein level [61]. The role of superoxide dismutases is discussed in Chapter 2. A similar scenario was seen for molecular chaperone DnaK, although

fold increases were less than superoxide dismutases. Although the induction in DnaK in response to high salt (4%) has been seen previously at the protein level [61], elevated transcript levels have not been seen in the long-term [60]. It is not possible to extrapolate at this stage whether this gene is transcribed in the long-term and therefore *dnaK* was a candidate for RT-qPCR analysis (section 3.3.3). The stress chaperones, 16.6 kDa small heat shock protein and GroES (co-chaperonin) were both expressed higher in 6% salt (1.75/ 2.10-fold and 1.59/1.83-fold, respectively), however, previous studies have identified transcript and protein levels to be salt responsive in the short-term only [60, 61]. Not all stress proteins were expressed in higher amounts in 6% salt and heat shock protein, GrpE, was reduced by 1.66 and 1.72-fold.

The presence of the majority of these proteins in elevated amounts in 6% salt acclimated cells can be used as proof that the techniques and methodology used here are suitable for identifying adaptation proteins.

3.3.2.2 Carbon and energy metabolism

Enzymes with functions in basic carbon metabolism have been previously associated with 4% salt acclimation due to the requirement of large amounts of compatible solute, GG. Interestingly, the gene transcripts for some of these proteins have only been seen to be induced in the short-term (24 hours) in 4% salt. These changes imply higher rates of processes such as glycolysis, pentose phosphate pathway and CO_2 fixation, with control that may exist at the post-transcriptional level [61]. ***A complex response is predicted here considering the diverse functional roles members of this group have. An increase in carbon requirement for GG synthesis is likely to be sourced from pathways here.***

Two proteins of carbon metabolism pathways, previously identified as induced in 4% salt, were also induced here, fructose-bisphosphate aldolase (1.52 and 1.55-fold) and phosphoglycerate kinase (up to 3-fold). Dihydrolipoamide acetyltransferase which is part of the pyruvate dehydrogenase complex (post-glycolysis) was also over 2-fold up-regulated in high salt. These changes imply increased requirement for energy as carbon is metabolised, as has been reported formerly [61, 176, 186]. As expected, a large increase (4.86 and 6.21-fold) in a

compatible solute synthesis enzyme, GG phosphate synthase was seen, an essential response in salt acclimation. An increase in 4-alpha-glucanotransferase (2.78 and 3-fold) has also been seen previously in the long-term [61], and may play a role in making carbon available for GG synthesis. This may also explain the 1.62 and 1.66-fold increase in glucose-1-phosphate adenylyltransferase. These latter two enzymes are positioned close together in pathways involving the re-arrangement of carbon atoms in glycogen synthesis pathways, and the accumulation of glycogen in salt acclimated cells has been reported previously in *Synechocystis* [196].

Many studies conducted previously fail to identify proteins which have *reduced* expression, and although making analysis more complex, these changes are inherent to drawing conclusions as to what adjustments in metabolic pathways affect adaptation. A reduced amount of glycogen phosphorylase (2.68 and 1.83-fold) indicates a reduced amount of glycogen breakdown, and therefore supports the supposition of glycogen's role. However, Fulda *et al.* [61] reported the opposite control in cells exposed to 4% salt (3.8-fold induced), and postulated that this enzyme could be involved in glycogen synthesis rather than breakdown. We hypothesise here that post-transcriptional/translational control may dictate which direction this protein functions in response to signals related to salt stress. Further work using deletion mutants and techniques to detect PTM's is presently required here.

The cellular response to salt stress can be simply summarised by reporting an increase in energy conversion, however, not all enzymes involved in energy synthesis are necessarily directly related to energy supply in salt stress. A good example of this was reported by Suzuki *et al* [295], where the kinetic activities of enzymes in the glycolytic pathway were investigated in the mangrove tree in response to long-term 150 mM salt stress. Fructose-6-phosphate transferase was thought to play an important role in the logarithmic growth stage where biosynthesis is rapid, however, it is not thought to be directly related to energy supply for the removal of salt from the cytoplasm [295]. The role of the PPP regulatory control protein OxPPCycle gene appears reduced here (1.52 and 1.69-fold), but the former biological replicate sample has a high CV compared to its fold-change (0.44) and therefore this may not confirm an overall lower requirement for

this pathway, the functions of which include generating reducing power (NADPH) for biosynthetic reactions. In support of *overall* reduced PPP requirement, the reduced expression (2.56 and 2.23-fold) of an enzyme responsible for the hydrolysis of pyrophosphate (PPi), which is formed principally as the product of the many biosynthetic reactions that utilise ATP, suggests an overall reduction in biosynthesis, and linking in with the findings above, these are most likely biosynthesis reactions associated with slowed growth rate in 6% salt.

The proteins discussed above are organised into pathways and presented in on-line material such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [296]. However, they are rarely associated with one specific pathway, and therefore several candidates belonging to the same pathways need to change significantly in fold expression, in the same direction, to draw high confidence predictions. The differences in proteins involved in energy conversion, synthesis and central intermediary metabolism (chiefly carbon) here, require careful examination before these predictions can be made. The role of post-transcriptional regulation with these proteins requires further analysis, because no differential expression changes in the long-term were seen in transcript level data apart from the enhanced production of GG phosphate synthase [60]. Furthermore, phosphorylations of proteins have been associated with salt shocked *Synechocystis* cells previously [297].

3.3.2.3 Photosynthesis and respiration

Acclimation to salt (up to 4%) stress in *Synechocystis* involves re-organising of energetic pathways and mainly an increase in electron transport by PSI in the thylakoid membranes, and cytochrome oxidase activity in the plasma membrane during respiration. Pigment levels are reduced but efficiency of electron transfer enhanced (Chapter 2). *A reduction in pigments in 6% salt is also expected and can be visualised from cells prior to harvesting, and an increase in cyclic phosphorylation is expected.*

Fifteen proteins involved in photosynthesis and respiration were differentially regulated, 13 of which decreased in expression. As expected, eight are pigment proteins and therefore 6% salt leads to pigment loss in cells, as does 4% salt [61]. In addition, long-term acclimation does not lead to recovery of these pigments.

Cytochrome c550 or PsbV decreased 1.93 and 2.13-fold and is associated with the oxygen evolving complex in PSII. Reductions in oxygen evolving machinery has been seen previously at the transcript level in both *Synechocystis* cells (4% salt at 24 hrs) and in *in vitro* isolated thylakoid membranes from *Synechococcus* sp. PCC7942 cells [60, 215].

Perhaps the most surprising of the proteins identified with reduced expression was the electron carrier flavodoxin. This protein has previously been associated as an important alternative electron carrier implemented in salt adapted cells of *Synechocystis* (2 to 4% salt) [61, 210]. Its marked decrease here implies this role is no longer applicable in higher salt (6%). Its non-essential role has been demonstrated however, when a flavodoxin mutant displayed reduced cyclic electron flow around PSI, but was still capable of acclimation to 2% salt [210]. In PSII, one subunit, CP43 increased 1.82 and 1.80-fold in expression. Specific salt-induced damage to this subunit may lead to increased synthesis (increased turnover). However, oxidative stress is known to actually *inhibit* repair of salt-damaged PSII protein D1 in cyanobacteria [298]. These results imply variable stoichiometry in the complexes. Vipp1 is a protein previously identified as salt-induced in *Synechocystis*, and is believed to transfer reaction centre proteins to the thylakoid membranes during stress and therefore replacement of damaged subunits occurs [224]. Increased expression of replacement proteins has been observed in light stressed cells previously (Chapter 2, section 2.6.1). Finally, apocytochrome f precursor, part of the cytochrome b_6/f complex, which is involved in electron transfer between both photosystems, increased 1.93 and 1.69-fold in 6% salt. This implies *increased* electron transfer to PSI (non-cyclic photophosphorylation) and therefore enhanced energy synthesis.

NADH dehydrogenase subunit I is part of the larger enzyme complex and involved in oxidative phosphorylation in respiration. Despite its 1.59 and 1.75-fold decrease in expression in high salt, other NADH dehydrogenase subunits were identified and not differentially expressed, for example, beta subunit (1.01 and 0.71-fold change), suggesting a stoichiometric change in this complex. A DNA microarray study found transcript levels for subunits of this complex to increase immediately in response to 0.02M (~1.2%) salt, but decrease in response to 1.0M (6%) salt [204].

3.3.2.4 Transport and binding proteins

Saline environments are thought to cause compositional and structural changes in the plasma membrane, and this is where the majority of transporter proteins are anchored [299]. This has been assumed to affect nutrient uptake in 4% salt acclimated *Synechocystis* cells [224], and therefore lead to an increase in expression of nutrient uptake proteins. *6% salt is expected to cause similar structural changes and therefore rearrangement of transport proteins is also expected, although the extra stress may affect overall energy available for the active transport systems.*

In contrast to the results seen in *Synechocystis* cells adapted to 4% salt (over 6 to 8 days), four nutrient binding proteins experienced large reductions in expression in 6% salt. Relative levels of iron transport protein (1.75 and 1.99-fold), periplasmic phosphate binding protein (10.07 and 10.38-fold), periplasmic iron binding protein (9.28 and 11.08-fold) and a bicarbonate transporter (2.09 and 3.62-fold) were all significantly diminished. Moreover, bacterioferritin which is responsible for storage of as much as 50% of cellular iron was 1.61 and 1.72-fold reduced. It is not certain why cells reduce expression of transporter proteins in high salt (6%). Reduced growth and overall protein synthesis may mean fewer nutrients are required, although further work needs to be done to elucidate the specific reasons for this unexpected response.

Three transport proteins did increase in expression in 6% salt. An ABC-1 like transporter increased 4.27 and 4.93-fold, and is considered to be a putative ubiquinone biosynthesis protein and therefore could be implemented in energy generation. This protein or its coding gene has not previously been identified as differentially regulated in 4% salt. Another ABC transporter also increased 2.35 and 2.66-fold. Finally, a 1.83 and 1.52-fold increase in nitrate transport protein was seen in cells. Interestingly, the uptake of combined nitrogen has previously been suggested to inhibit Na^+ ion influx and therefore a mechanism for the protection against salt stress in freshwater and brackish water cyanobacteria, *Anabaena* sp. strain L-31 and *Anabaena torulosa* [300]. This response however, has not previously been reported in *Synechocystis*.

3.3.2.5 Transcription and Translation

The rates of protein transcription and translation have previously been shown to increase in salt *shocked* cells. Ribosomal proteins can become unstable during the sudden change in external salinity, and their enhanced synthesis is thought to overcome this reaction [61]. *An overall reduction in global translation (6% salt causes a ca. 100% reduction in growth rate) means a reduced expression of related proteins is expected, although increased translation of selected proteins essential for cell survival is expected.*

A 2.91 and 3.17-fold increase in elongation factor EF-G may occur for synthesis of so called 'stress proteins' required for cell survival. Cellular stresses like salt stress which lead to global repression of translation, are often accompanied by increased translation of selected proteins that are required for cell survival. It is possible that increased levels of mRNA coding for stress-proteins, compete more successfully for protein synthesis ability in salt-stressed cells, than transiently expressed mRNA.

Although the response of the translation machinery summarised in bold text is associated with a short-term response to salt stress, 26 proteins involved in these cellular processes were identified as differentially expressed in this study, and 13 of these proteins experienced over a 2-fold change (in at least one replicate). In addition to ribosomal proteins, two translation initiation factors (IF), responsible for protein modification, were also present in higher amounts in acclimated cells. Transcript levels of IF have been seen to increase higher in *Synechocystis* after 24 hours than after 5 days in 4% salt [60], but the data from iTRAQ is collected after a complete acclimation period and therefore it is not possible to corroborate this change.

Proteins were differentially expressed in this long-term adaptation study which had only previously been shown to change in the short-term (at transcript and protein level) [60, 61]), for example, a 2.21 and 2.22-fold increase in RNA-binding protein. It is possible that this protein assists in stabilising nucleic acid structures which are susceptible to damage caused by changes in water potential and ionic strength. Another protein which assists in tackling the detrimental effects of salt in the long-term is peptidyl-prolyl cis-trans isomerase (2.07 and 2.40-fold increase). This

enzyme accelerates protein folding [301]. Finally, the 1.83 and 2.32-fold induction of glutamyl-tRNA synthetase has been seen previously in the long-term [61].

As mentioned previously, is it important to include proteins with reduced expression into the discussion, and 9 proteins belonging to this group were found. Not all proteins involved in transcription and translation were expected to increase in expression in salt adapted cells because cell growth and division is reduced. This is supported by the 1.55 and 1.99-fold decrease in ribosome releasing (or recycling) factor. This protein dissociates ribosomes from mRNA after translation is terminated, and therefore its principal function is in bacterial growth [302].

3.3.2.6 Hypothetical proteins

Hypothetical proteins do not have known associated functions. However, it is possible to relate putative functions to several of these proteins based on coding sequence similarities and many of these proteins play crucial roles in salt acclimation (Chapter 2). ***Adaptation to 6% salt likely requires at least equal numbers of hypothetical proteins and the proteomic methodology applied here are expected to identify more.***

Hypothetical protein sll1863 had the highest induction in salt adapted cells (10.72 and 4.74-fold) and this large increase has been seen previously at the transcript level in cells grown in 4% salt (>200 fold) [60]. It was identified as newly induced in a 2DE based study because the protein spot was not identified in the control gel, and hence revealing a disadvantage of 2DE methodology. Unfortunately there is no function associated with this protein, and perhaps surprisingly, a mutant strain defective in this gene did not show a different tolerance level to salt [60]. It may be useful to make a double mutant by also knocking out the gene for sll1862 considering this gene is co-transcribed with sll1863. Interesting examples include the 1.99 and 1.90-fold increase in hypothetical protein slr1894, which contains a stress associated DNA-binding domain. A 3.57 and 4.53-fold increase in hypothetical protein sll0861 was also seen and this protein harbours a sugar isomerase domain, and therefore possibly controls phosphosugar synthesis as part of compatible solute regulation.

In addition to the 12 induced hypothetical proteins, a further 14 were reduced. Interestingly, only four of the transcript levels for these proteins have previously been identified as experiencing a reduction in expression in salt (~3% salt) [3]. Contrary to this study, hypothetical protein sll1106 was previously identified to be induced in salt [3] and therefore it was chosen as a candidate for RT-qPCR analysis (section 3.4.3).

Only ~10% of the differentially expressed hypothetical proteins identified using iTRAQ have been observed previously in proteomics experiments. 16 of the 26 proteins were differentially expressed over 2-fold and therefore reach the threshold of most proteomics experiments. It is likely that some of these proteins respond to higher salt concentrations (6%), whereas others were simply not identified using techniques which perform poorly at identifying low abundance proteins. Again, the advantage of using the proteomic methodology implemented here is highlighted.

3.3.2.7 Other proteins of interest

This shotgun study identified several proteins which could collectively play an important role to assist a cell's survival in high salt. Cell envelope enzyme UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase (UDP-ligase) is thought to regulate cell shape by controlling the cell wall structure [303]. The cell wall in *Synechocystis* cells and all bacteria protects the cell from the environment by providing rigidity to counteract internal osmotic pressure. This enzyme is specifically involved in peptidoglycan (also known as murein) synthesis, and it is this which provides strength. It is probable that the change in external salt concentration means this enzyme is required in larger amounts to produce more peptidoglycan and help maintain the overall shape of the cell, and confidence in this possibility is increased because of the 2.57 and 2.72-fold increase in enzyme glutamate-ammonia ligase, which functions in the same pathway. Both proteins have not previously been identified in the salt response, however, genes in the same pathway have been shown to be induced in heat-shocked *Synechocystis* cells [304], and an increase in peptidoglycan layers for protection was supposed.

Glycerol-3-phosphate dehydrogenase was also identified as salt-induced (1.95 and 1.52-fold) for the first time using proteomics and is thought to play a role in

compatible solute accumulation. This enzyme-coding gene has been induced in the short-term and long-term [60], and its transcript levels were also investigated in section 3.4.3.

An increase in light-repressed protein has been seen previously at the protein level in a study which specifically looked at plasma membrane proteins [224] (Chapter 2). It was 2.52 and 2.62-fold increased here, and it may be involved in redox state changes due to ionic changes. This illustrates the ability of this technique to identify and quantify proteins from more than one subcellular fraction in a single study.

Anti-sigma B factor antagonist was 2.70 and 2.30-fold less expressed in high salt and despite not being identified previously at the protein or transcript level, this change can be expected. This protein is involved in terminating transcription of the general stress regulon and therefore is *not* required in 'stressed' conditions. Its mode of action has been studied in *Bacillus subtilis* [305], and is illustrated in Figure 3.3. It involves sensing salt concentration changes in the environment which leads to de-phosphorylation of an anti-anti sigma factor protein. In no salt conditions, anti-anti sigma factor is phosphorylated and cannot bind to anti-sigma B factor. Therefore, anti-sigma B factor is free to bind to Sigma B, which subsequently cannot bind to the promoter to initiate transcription of the general stress regulon. In a saline environment, a histidine (response) protein leads to de-phosphorylation of the anti-anti sigma factor which can now bind to anti-sigma B factor (protein identified in iTRAQ) to enable Sigma B to form the transcription complex and ready to initiate transcription. It is therefore assumed that a reduced amount of anti-sigma B factor makes this process more efficient in adapted cells (Figure 3.3).

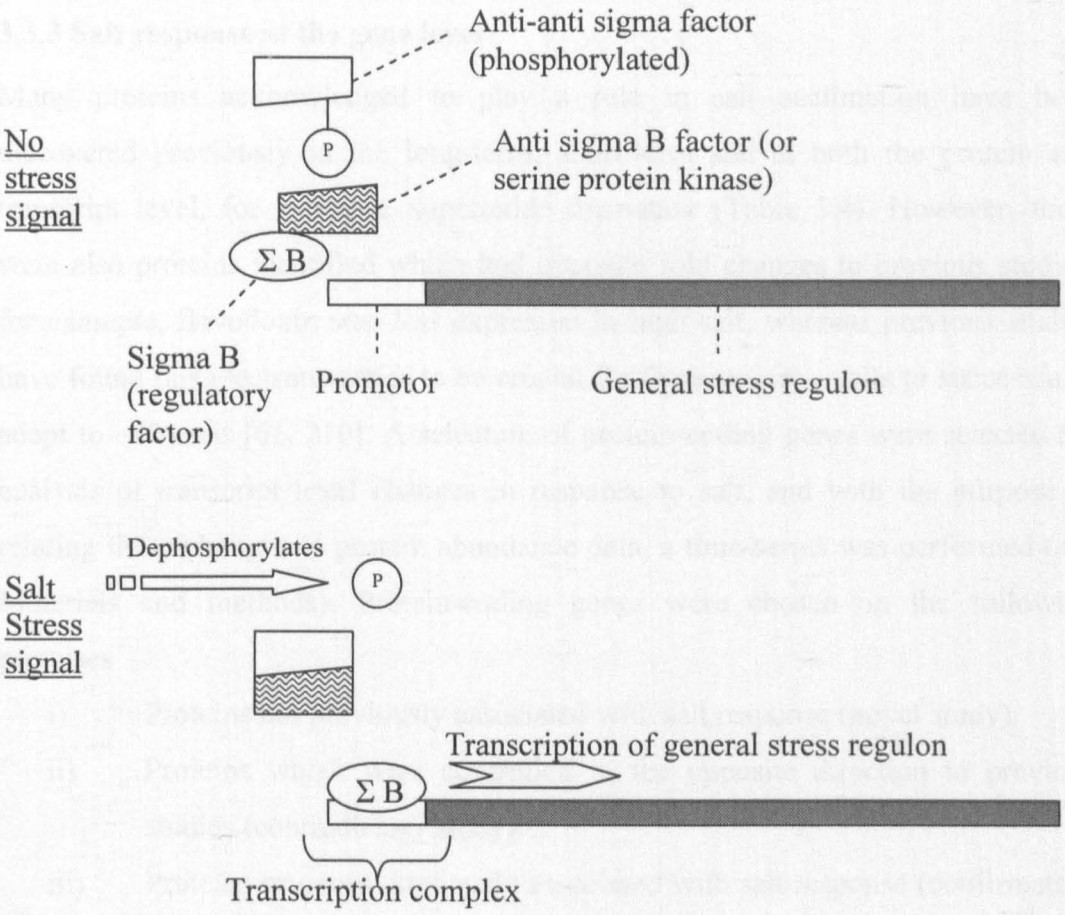


Figure 3.3: The mode of control for transcription of the general stress regulon involving anti-sigma B factor in *Bacillus subtilis* [305], which was 2.7 and 2.3-fold reduced in expression in salt acclimated *Synechocystis* cells.

3.3.3 Salt response at the gene level

Many proteins acknowledged to play a role in salt acclimation have been discovered previously in the long-term, short-term and at both the protein and transcript level, for example, superoxide dismutase (Table 3.4). However, there were also proteins identified which had opposite fold changes to previous studies, for example, flavodoxin was *less* expressed in high salt, whereas previous studies have found this electron carrier to be crucial for *Synechocystis* cells to successfully adapt to ~4% salt [61, 210]. A selection of protein-coding genes were selected for analysis of transcript level changes in response to salt, and with the purpose of relating these changes to protein abundance data, a time-series was performed (see Materials and methods). Protein-coding genes were chosen on the following premises

i) Proteins not previously associated with salt response (novel study).

ii) Proteins which were controlled in the opposite direction to previous studies (contradictory study).

iii) Proteins or genes previously associated with salt response (confirmatory study).

iv) Differences between short-term and long-term responses at both gene and protein level (shock vs. adaptation study)

The RNA quality was checked using agarose gel electrophoresis and an example gel photograph is given in Figure 3.3.

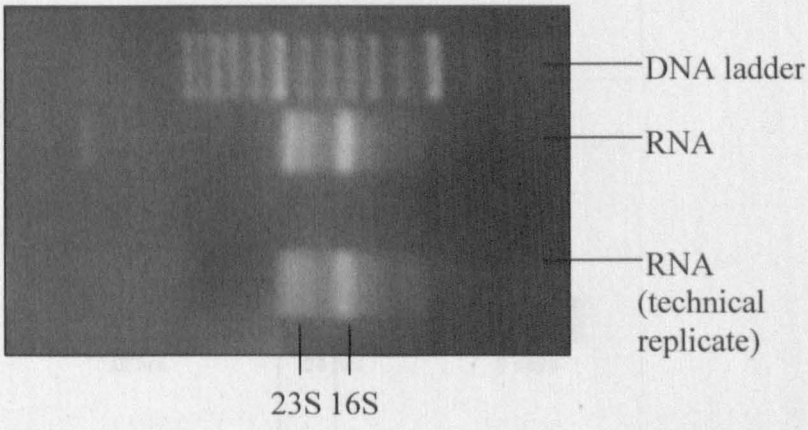


Figure 3.3: DNase treated total RNA extractions (1.2% agarose gel). DNA ladder is Hyperladder I (Bioline).

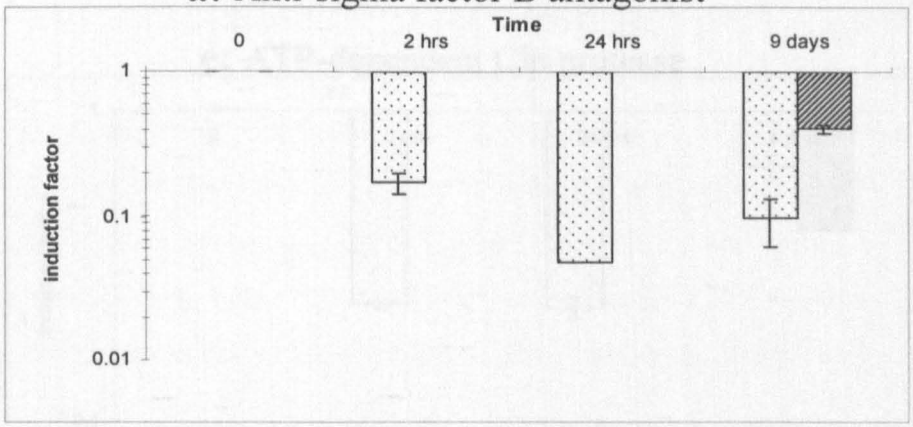
In total, 22 protein-coding genes were quantified across the time-series as with iTRAQ data, quantitative results were only treated as significant if *both* biological replicates exceeded the 1.5-fold threshold for extra stringency. 11 of these had fold inductions (at 9 days) which were in the same direction and above the significance threshold as protein level changes (at 9 days). The results are discussed below.

Figures 3.4 to 3.7 show gene expression changes (dotted bars) after incubation in 6% salt, after 2 hours, 24 hours and 9 days. The dashed bars signify iTRAQ results after only 9 days. The values are all relative to time = 0 or 0% salt, and therefore the first value is recorded as 1 in all cases.

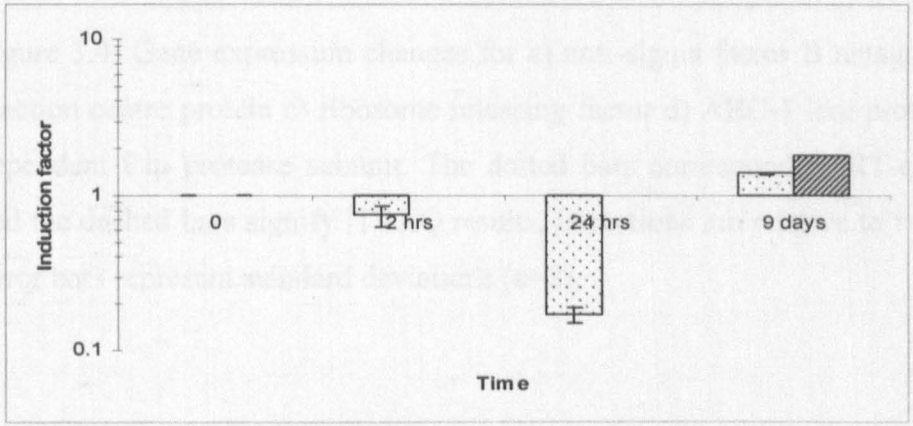
3.3.3.1 Novel protein level changes

RT-qPCR was performed on five protein-coding genes which have not previously been associated with a salt response (Figure 3.4).

a: Anti-sigma factor B antagonist



b: PSII reaction centre



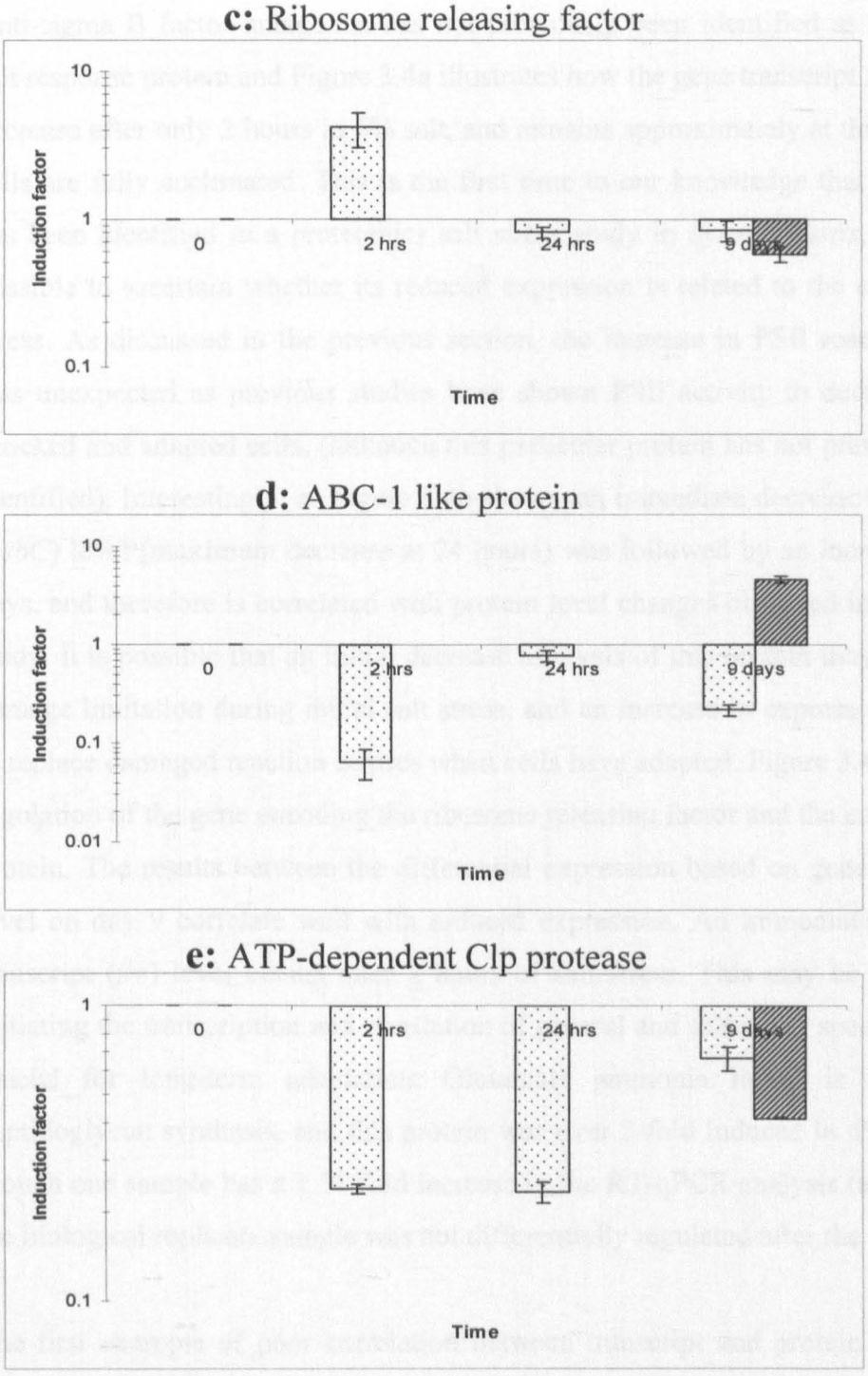


Figure 3.4: Gene expression changes for a) anti-sigma factor B antagonist b) PSII reaction centre protein c) ribosome releasing factor d) ABC-1 like protein e) ATP-dependent Clp protease subunit. The dotted bars correspond to RT-qPCR results and the dashed bars signify iTRAQ results. Inductions are relative to t=0 (0% salt). Error bars represent standard deviations (n=2).

Anti-sigma B factor antagonist has not previously been identified as an essential salt response protein and Figure 3.4a illustrates how the gene transcript levels (*icfG*) decrease after only 2 hours in 6% salt, and remains approximately at this level until cells are fully acclimated. This is the first time to our knowledge that this protein has been identified in a proteomics salt stress study in *Synechocystis*, so it is not possible to ascertain whether its reduced expression is related to the *extent* of salt stress. As discussed in the previous section, the increase in PSII reaction protein was unexpected as previous studies have shown PSII activity to decrease in salt shocked and adapted cells, (although this particular protein has not previously been identified). Interestingly, as Figure 3.4b shows, an immediate decrease in transcript (*psbC*) level (maximum decrease at 24 hours) was followed by an increase up to 9 days, and therefore is correlated with protein level changes observed in the iTRAQ study. It is possible that an initial decrease in levels of this protein may function as damage limitation during initial salt stress, and an increase in expression functions to replace damaged reaction centres when cells have adapted. Figure 3.4c shows the regulation of the gene encoding the ribosome releasing factor and the corresponding protein. The results between the differential expression based on gene and protein level on day 9 correlate well with reduced expression. An immediate increase in transcript (*frr*) level occurs after 2 hours of salt stress. This may be due to cells initiating the transcription and translation of general and salt stress specific proteins crucial for long-term adaptation. Glutamate ammonia ligase is involved in peptidoglycan synthesis, and this protein was over 2-fold induced in 6% salt. Even though one sample has a 1.51-fold increase in the RT-qPCR analysis (after 9 days), the biological replicate sample was not differentially regulated after the same time.

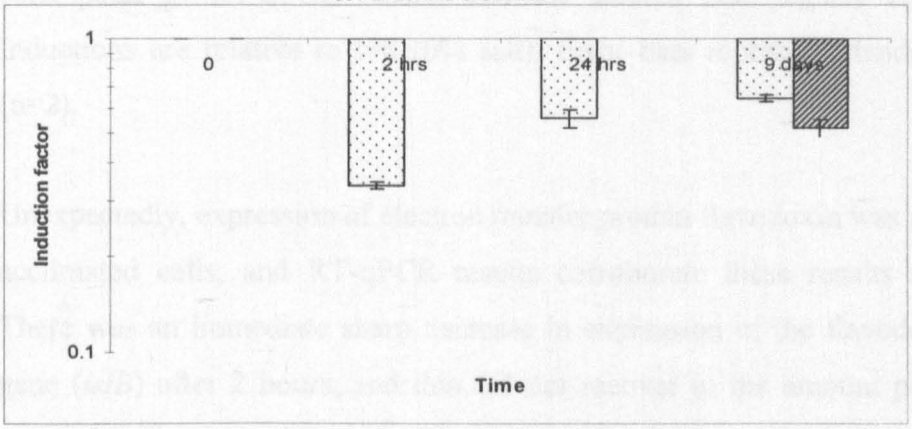
The first example of poor correlation between transcript and protein quantitation occurs with an ABC-1 like protein (*ubiB*) (Figure 3.4d). It was induced 4.27 and 4.93-fold in high salt and therefore postulated to play a significant role in salt acclimation. Transcript data shows an immediate reduction in expression (2 hours). After a recovery to approximately initial levels (24 hours), the expression of the gene after 9 days is again reduced and this final quantitation contrasts vastly with protein levels. A discussion on reasons for incongruent results is given in 3.3.4. ATP-dependent Clp protease proteolytic subunit is involved in degrading protein aggregates during stress and can act as a molecular chaperone [306]. It would

therefore be expected to increase in expression in stressed cells, however, in 6% salt it decreased 2.38 and 2.39-fold. Figure 3.4e shows that transcripts levels for this gene (*clpP*) also decrease immediately in salt-stressed cells, and remain reduced after 9 days, supporting the assumption that this subunit does not play a role in acclimation to 6% salt.

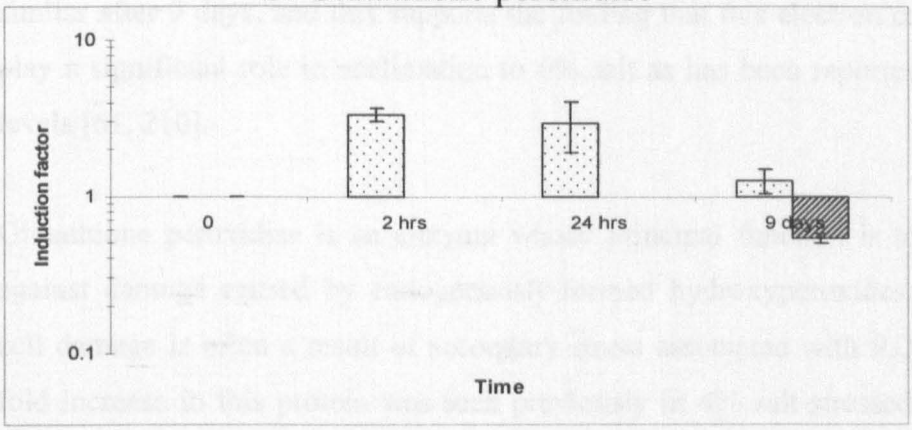
3.3.3.2 Protein level changes in contrast to previous studies

RT-qPCR was performed on four protein-coding genes which were controlled in the opposite direction to previous studies (Figure 3.5).

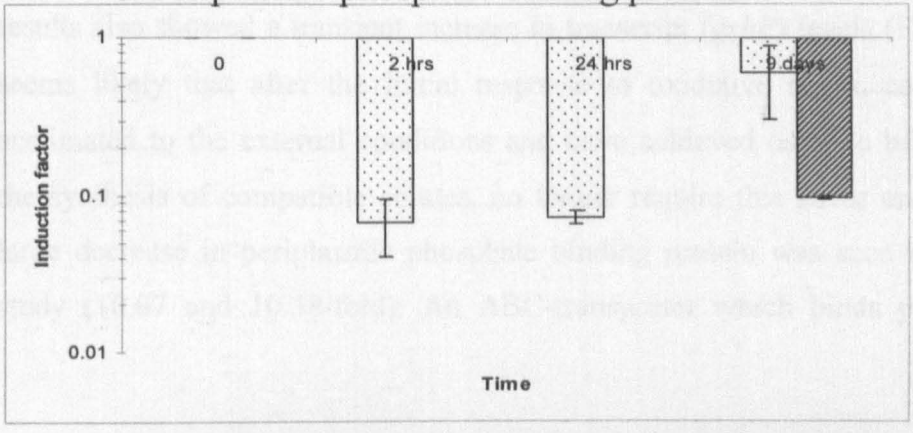
a: Flavodoxin



b: Glutathione peroxidise



c: Periplasmic phosphate binding protein



d: Bicarbonate transporter

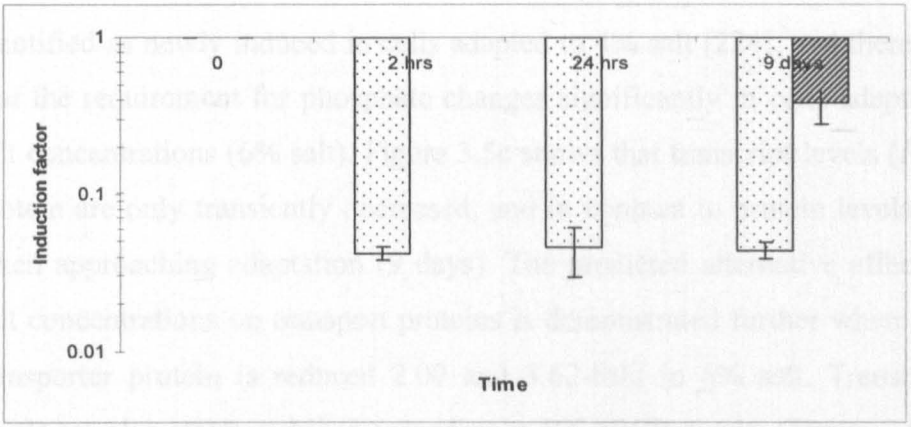


Figure 3.5: Gene expression changes for a) flavodoxin b) glutathione peroxidise c) periplasmic phosphate binding protein d) bicarbonate transporter. The dotted bars correspond to RT-qPCR results and the dashed bars signify iTRAQ results. Inductions are relative to t=0 (0% salt). Error bars represent standard deviations (n=2).

Unexpectedly, expression of electron transfer protein flavodoxin was reduced in salt acclimated cells, and RT-qPCR results corroborate these results (Figure 3.5a). There was an immediate sharp decrease in expression of the flavodoxin encoding gene (*isiB*) after 2 hours, and this did not recover to the amount present in non-stressed cells even at 9 days. The relative level of protein and transcripts were similar after 9 days, and this supports the finding that this electron carrier does not play a significant role in acclimation to 6% salt as has been reported at lower salt levels [61, 210].

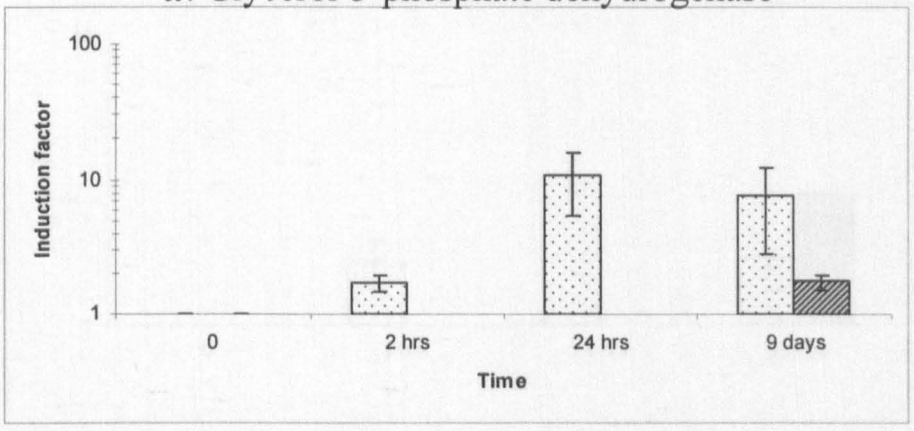
Glutathione peroxidise is an enzyme whose principal function is to protect cells against damage caused by endogenously-formed hydroxyperoxides. This type of cell damage is often a result of secondary stress associated with ROS [195]. A 3-fold increase in this protein was seen previously in 4% salt-stressed cells [61], as well as a transient increase in transcript levels [60]. In agreement, the RT-qPCR results also showed a transient increase in transcript (*gshP*) levels (Figure 3.5b). It seems likely that after the initial response to oxidative stress, cells which are acclimated to the external conditions and have achieved osmotic balance through the synthesis of compatible solutes, no longer require this stress enzyme. A very large decrease in periplasmic phosphate binding protein was seen in the iTRAQ study (10.07 and 10.38-fold). An ABC-transporter which binds phosphate was

identified as newly induced in cells adapted to 4% salt [224], and therefore implies that the requirement for phosphate changes significantly in cells adapted to higher salt concentrations (6% salt). Figure 3.5c shows that transcript levels (*PstS*) for this protein are only transiently decreased, and in contrast to protein levels, it recovers when approaching adaptation (9 days). The predicted alternative effects of higher salt concentrations on transport proteins is demonstrated further where bicarbonate transporter protein is reduced 2.09 and 3.62-fold in 6% salt. Transcript (*cmpA*) levels are also stably reduced according to RT-qPCR results (Figure 3.5d). As with phosphate binding protein discussed above, previous studies show the protein to be newly induced in 4% salt [224], and transcript levels to be over 5-fold induced after 6 hours [60].

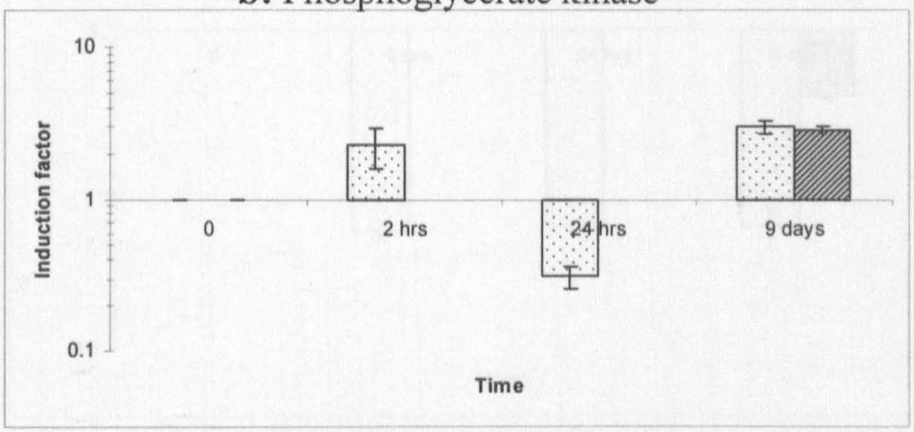
3.3.3.3 Confirmatory protein level changes

RT-qPCR was performed on six protein-coding genes which have previously been associated with salt response (Figure 3.6)

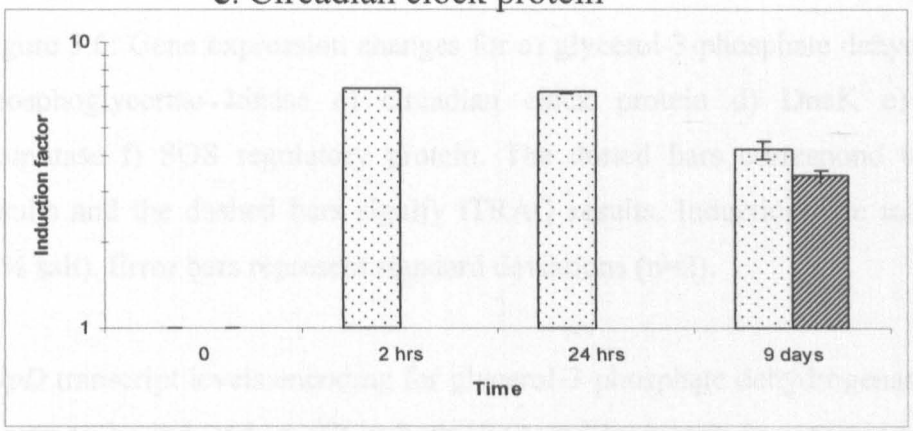
a: Glycerol-3-phosphate dehydrogenase



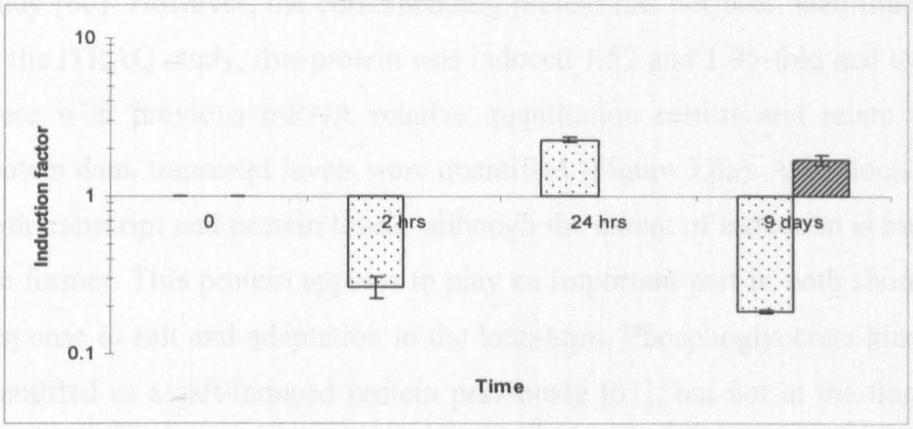
b: Phosphoglycerate kinase



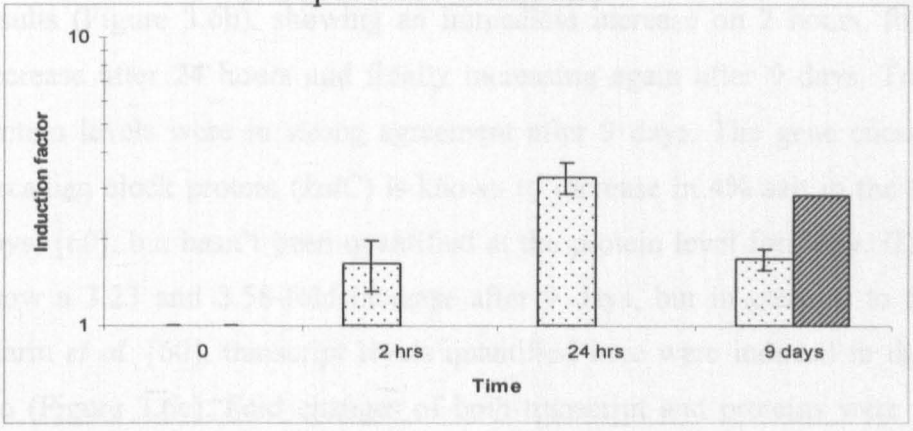
c: Circadian clock protein



d: DnaK



e: Superoxide dismutase



f: SOS regulatory protein

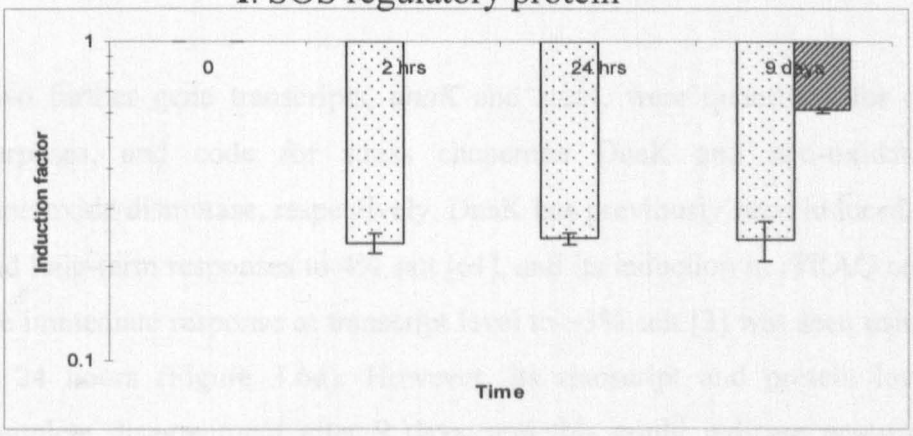


Figure 3.6: Gene expression changes for a) glycerol-3-phosphate dehydrogenase b) phosphoglycerate kinase c) circadian clock protein d) DnaK e) superoxide dismutase f) SOS regulatory protein. The dotted bars correspond to RT-qPCR results and the dashed bars signify iTRAQ results. Inductions are relative to t=0 (0% salt). Error bars represent standard deviations (n=2).

GlpD transcript levels encoding for glycerol-3-phosphate dehydrogenase have been shown to be induced in salt in both short and long-term in a previous microarray study [60]. However, the corresponding protein has not been identified previously. In the iTRAQ study, this protein was induced 1.52 and 1.95-fold and to corroborate these with previous mRNA relative quantitation results and relate these to the protein data, transcript levels were quantified (Figure 3.6a). An induction is seen in both transcript and protein levels, although the extent of induction is much larger in the former. This protein appears to play an important part in both short-term shock response to salt and adaptation in the long-term. Phosphoglycerate kinase has been identified as a salt-induced protein previously [61], but not at the transcript level. The relative quantitation of *pgk* gene expression produced interesting yet complex results (Figure 3.6b), showing an immediate increase on 2 hours, followed by a decrease after 24 hours and finally increasing again after 9 days. Transcript and protein levels were in strong agreement after 9 days. The gene encoding for the circadian clock protein (*kaiC*) is known to increase in 4% salt in the long-term (5 days) [60], but hasn't been quantified at the protein level formerly. iTRAQ results show a 3.23 and 3.58-fold increase after 9 days, but in contrast to the study by Marin *et al.* [60], transcript levels quantified here were induced in the short-term too (Figure 3.6c). Fold changes of both transcript and proteins were very similar after 9 days.

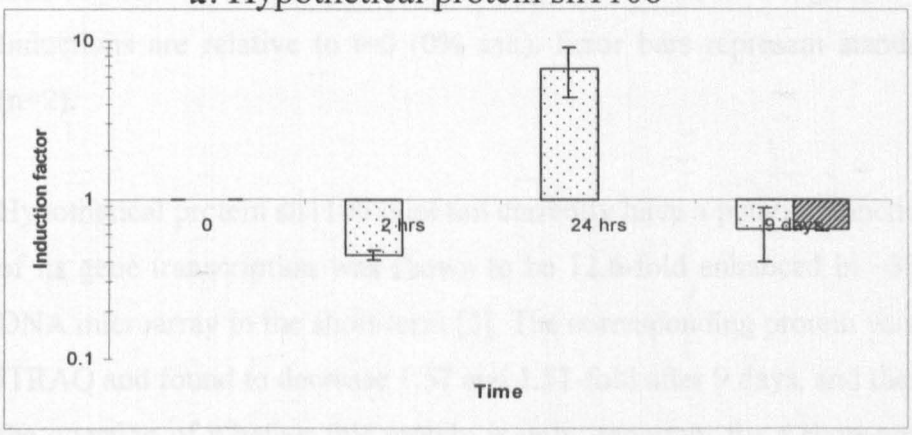
Two further gene transcripts, *dnaK* and *sodA*, were quantified for confirmatory purposes, and code for stress chaperone DnaK and anti-oxidative enzyme superoxide dismutase, respectively. DnaK has previously been induced during short and long-term responses to 4% salt [61], and its induction in iTRAQ concurs. Also, the immediate response at transcript level to ~3% salt [3] was seen using RT-qPCR at 24 hours (Figure 3.6d). However, its transcript and protein levels were in complete disagreement after 9 days, and this could indicate post-transcriptional

control. The same comparison with superoxide dismutase and SOS regulatory protein (LexA protein) produced more complementary results (Figures 3.6e and 3.6f).

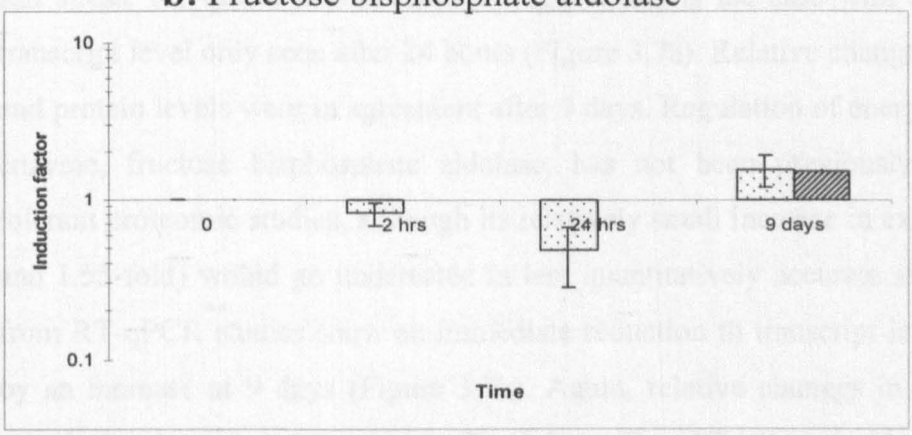
3.3.3.4 Differences in short-term and long-term changes

RT-qPCR was performed on four protein-coding genes where discrepancies arose in regard to short-term and long-term responses between previous work with microarrays, 2DE gels and iTRAQ data presented here (Figure 3.7).

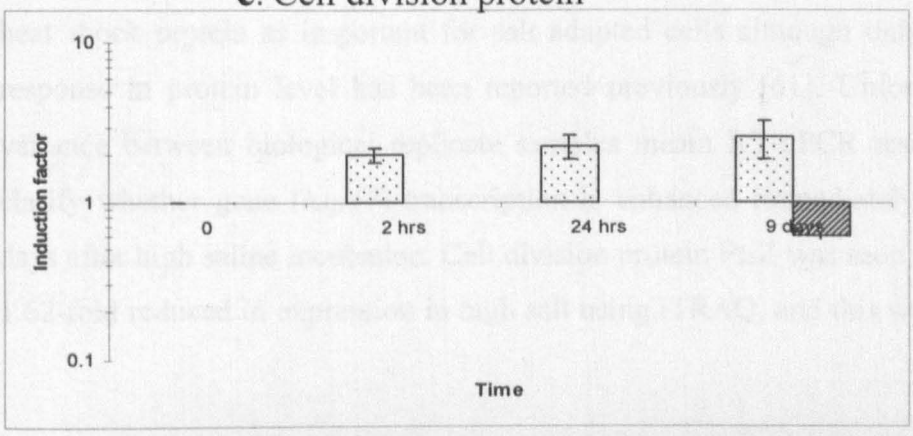
a: Hypothetical protein sll1106



b: Fructose bisphosphate aldolase



c: Cell division protein



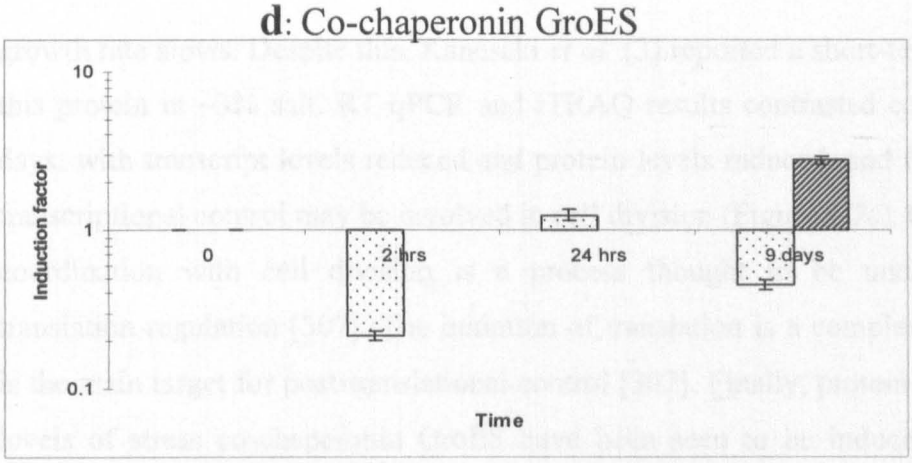


Figure 3.7: Gene expression changes for a) hypothetical protein sll1106 b) fructose bisphosphate aldolase c) cell division protein d) co-chaperonin GroES. The dotted bars correspond to RT-qPCR results and the dashed bars signify iTRAQ results. Inductions are relative to t=0 (0% salt). Error bars represent standard deviations (n=2).

Hypothetical protein sll1106 does not currently have a putative function, but control of its gene transcription was shown to be 12.6-fold enhanced in ~3% salt using a DNA microarray in the short-term [3]. The corresponding protein was quantified by iTRAQ and found to decrease 1.57 and 1.51-fold after 9 days, and these results pose the question of whether this protein is only important for a short-term response to salt stress. RT-qPCR results seem to suggest this is the case with an increase in transcript level only seen after 24 hours (Figure 3.7a). Relative changes in transcript and protein levels were in agreement after 9 days. Regulation of energy metabolism enzyme, fructose bisphosphate aldolase, has not been previously seen in salt tolerant proteomic studies, although its relatively small increase in expression (1.52 and 1.55-fold) would go undetected in less quantitatively accurate studies. Results from RT-qPCR studies show an immediate reduction in transcript levels, followed by an increase at 9 days (Figure 3.7b). Again, relative changes in transcript and protein levels were in agreement after 9 days. The iTRAQ study identified a small heat shock protein as important for salt adapted cells although only a short-term response in protein level has been reported previously [61]. Unfortunately, high variance between biological replicate samples meant RT-qPCR results could not clarify whether gene (*hsp17*) transcription is enhanced immediately, and up to 9 days after high saline incubation. Cell division protein FtsZ was seen to be 1.72 and 1.62-fold reduced in expression in high salt using iTRAQ, and this was expected as

growth rate slows. Despite this, Kanesaki *et al.* [3] reported a short-term increase in this protein in ~3% salt. RT-qPCR and iTRAQ results contrasted completely at 9 days, with transcript levels reduced and protein levels induced, and therefore post-transcriptional control may be involved in cell division (Figure 3.7c). Cell growth in coordination with cell division is a process thought to be under control of translation regulation [307]. The initiation of translation is a complex process, and is the main target for post-translational control [307]. Finally, protein and transcript levels of stress co-chaperonin GroES have been seen to be induced only in the short-term previously [60, 61], whereas iTRAQ results show a 2.61 and 2.93-fold increase after 9 days. Analysis of RT-qPCR data shows a general repression in transcription of the *groES* gene despite a brief return to 'normal' levels after 24 hours (Figure 3.7d). Again, the possibility of post-transcriptional control in expression arises to explain its accumulation after 9 days.

3.3.4 Discrepancies between gene and protein data

The majority of proteins and corresponding measurements of mRNA levels were in accordance, and hence could be used in combination to derive insight into how *Synechocystis* cells acclimate to 6% salt. However, there was strong discordance between protein and transcript fold changes for six of the candidate genes, for example, in the molecular chaperone DnaK (Figure 3.6d). Such large variations imply that protein abundances cannot be predicted by examining gene expression data alone, and recent studies have derived similar statistics [124, 126]. It therefore becomes essential to consider transcriptomics experiments as an indication only of how much functional protein is being synthesised to respond to different external conditions. Reasons for differences here include additional post-transcriptional mechanisms, for example, translation, PTM's, and mRNA stability and degradation. Essentially, cells regulate gene expression and protein abundance separately and this can explain some of the variances. Moreover, one gene rarely translates into just one protein. Furthermore, regulating translation is particularly important when cells are adapting to new conditions because precise changes in protein levels are required [307].

Integrating transcript and protein data was not the aim of undertaking RT-qPCR experiments in this chapter. It was implemented to illustrate how proteomics can reveal the functional state of the cell, whereas transcriptomics more specifically *predicts* this. The rate of corroboration based on the protein-coding genes selected does however reveal more information on the cell system as a whole, for example, revealing where post-transcriptional control of genes is possibly occurring. Table 3.5 contains a selection of studies where mRNA and protein levels in perturbed state cells were compared. The rate is calculated based on actual values of relative fold change as opposed to expression changes which are aimed in the same direction (amplitude instead of direction), and using either Pearson correlation coefficient or Spearman rank correlation [308], values vary from r = 0.21 to 0.61 (where 1 = complete positive correlation and 0 = no correlation). These studies used global transcriptome and proteome identification tools, and further analysis revealed that certain functional groups show higher levels of correlation. For example, Ideker *et al.* [125] found that the correlation between mRNA and protein level changes in metabolic and respiratory pathways was higher than the overall correlation.

Table 3.5: Quantitative studies comparing relative mRNA and protein abundance changes in control and perturbed cells.

System	Transcriptomics tool	Proteomics tool	Correlation results	Reference
Liver tissue	cDNA microarray	ICAT, 2DE DIGE	r = 0.54	[123]
Yeast (*S. cerevisae*)	cDNA microarray	ICAT	r = 0.21	[124]
Yeast (*S. cerevisae*)	cDNA microarray	ICAT	r = 0.61	[125]
Synechocystis	RT-qPCR	iTRAQ	r = 0.30	(this study)

The design of the experiment plays a role in understanding correlation. In this study, only one time point (i.e. 9 days) is used to measure protein levels and therefore no information about immediate protein synthesis is generated. However, the temporal transcript abundance data allows a more detailed examination of how the cells' transcription machinery responded to the external stimuli. The different half-lives of transcripts and proteins (0.1-10 h versus 0.5-500 h, respectively [309]) and the delay of protein synthesis (and therefore maximum abundance compared to transcript levels) will contribute to discordance, but the inclusion of time-series proteome data would assist in improving understanding of their relationship.

3.4 Conclusions and implications for further chapters

In this study, *in vitro* isobaric labelling was applied to gain an insight into how *Synechocystis* cells adapt to high salt (6%) over an acclimation period of 9 days. Using the quantitative power of iTRAQ, protein abundance changes were related to essential functional alterations in the cell system, which enabled survival in these adverse conditions. Many of these changes have been seen before in previous studies using alternative proteomic tools, but novel changes were also discovered. For example, increase in energy demand for active extrusion of salt ions and re-directing control of carbon metabolism enzymes for compatible solute synthesis was observed but with proteins not quantified previously. Newly identified responses were also attributed to the identification strength of shotgun proteomics workflow, for example, the reduced expression of anti-sigma B factor. Moreover, the study was based on acclimation to 6% salt, and no previous *Synechocystis* study (to date) has examined such high salt concentrations. In these salt concentrations, unexpected results included less reliance on alternative electron flow carrier, flavodoxin, and a reduced requirement for several transport and binding proteins.

The two workflows implementing different peptide fractionation systems, mass spectrometers and analysis software, produced complementary results ($r = 0.74$) and a list of proteins identified with high confidence was collated. Unlike previous proteomic studies assessing the immediate salt response in cyanobacteria, concentrating on *adaptation* of cells meant smaller expression changes were expected (50%). An average CV of 0.20 and 0.32 across the two workflows, from all proteins sourced from biologically replicate cells, made this possible. Therefore higher quantitation accuracy, often associated with well designed transcriptome studies, can be achieved with proteomics, and this provides a platform for further studies conducted in this thesis examining similar long-term acclimation strategies.

The relative transcript abundance of 22 protein-coding genes were also temporally quantified to further analyse the functional responses postulated using protein abundance changes. Significant quantitation changes were in agreement for half of these genes. Moreover, at least one replicate in three other cases gave a response

approaching the same conclusion. For example, fructose bisphosphate aldolase increased 1.52 and 1.55-fold in expression 9 days after salt stress, and replicates for transcript levels showed a 1.2 (not significant) and 1.91-fold (significant) increase after the same time period. As a result, there were disagreements in transcript and protein profiles, and this was expected. These differences highlight specific cellular components or functional groups for further investigation focusing on the role of post-transcriptional control.

In this chapter, our understanding of adaptation in a laboratory representative cyanobacterial strain through the use of proteomic tools has been successfully demonstrated. In Chapter 4, a more challenging system will be investigated with the aim of achieving equally promising insights.

Chapter 4: Adaptation of unsequenced *Euhalothece* to varying salt concentrations[2]

[2] This work has been accepted for publication in *Proteomics* (2008): In press

4.1 Introduction

In Chapter 3, the potential to apply global proteomics to identify important cellular processes involved in environmental adaptation, and the ability to highlight areas for future functional genomics/proteomics investigations was demonstrated. An increase in the number of fully sequenced organisms like *Synechocystis* has led to a concurrent increase in the number of proteomic projects. However, this has also meant that unsequenced environmental isolates have been largely ignored. Generally, isolating unsequenced microorganisms from environmental samples requires the ability to obtain a pure culture the organism in the laboratory, and although suitable for microbiological studies, they have remained relatively unattractive candidates for proteomic analysis due to protein identification problems [101].

The vast potential and importance of studying environmental isolates is becoming increasingly recognised. Therefore, recent developments have assisted in proteomic studies of unsequenced organisms, an area referred to as cross-species proteomics, as introduced in Chapter 2. The rapid growth of completed genome sequences means that it is becoming progressively possible to identify proteins of an unsequenced organism using conventional protein identification software like Mascot. Furthermore, developments in the area of MS and bioinformatics have significantly expanded the applicability of cross-species proteomics [101, 102, 105, 107, 108]. Cross-species studies have been successfully applied to a variety of unsequenced organisms to further understand environmental adaptation. Taylor *et al.* [310] studied the impact of a several relevant environmental stresses on a crop plant (Pea) mitochondrial proteome using Mascot searches. Another study was conducted on the partially sequenced halophilic, unicellular green alga *Dunaliella* [250], as discussed in Chapter 2, where nano-ESI MS was employed in combination with sequence-similarity database-searching algorithms, MS BLAST [102] and MultiTag [118].

In this chapter, the applicability of proteomics tools to understand how an unsequenced environmental isolate, *Euhalothece*, adapts to the conditions of its natural habitat (salt lake), was assessed. A method was developed and applied to

gain an overview of the adaptive responses. Subsequently, a large-scale global proteome analysis was conducted to gain a more significant insight into its acclimation processes. Therefore, unlike Chapter 3 where salt adaptation in fully sequenced *Synechocystis* was conducted, identification of proteins is less straightforward here, and this warrants a different proteomic approach. The intention is to achieve this whilst not compromising quantitative accuracy.

Euhalothece was isolated from Qabar-Onn salt lake, situated in the central Sahara in the south of Libya [161]. A morphological insight and 16S rRNA phylogenetic reconstruction analysis was applied to this organism to gain an indication of its applicability for homology-based proteomics, and therefore the potential to perform a large-scale global study. This unsequenced cyanobacterium can adapt to large variances in salt concentration, and unique to this cluster of cyanobacteria, does not require salt for survival. Therefore, it represents an ideal system to study cellular responses to adaptation in both high and low salt environments. Extreme halophiles have generated lots of interest in the last decade [311], however, species which tolerate large variances in salt concentration remain less studied. These halotolerant organisms pose different but relevant questions because achieving acclimation may be much more difficult than maintaining high internal solute concentrations, characteristic of halophiles [311, 312].

This chapter consists of two parts. Firstly, in Part A, an assessment is made of the potential to identify proteins in this unsequenced cyanobacterium through conducting an overview of the proteomic response in *Euhalothece* cells to changing salt conditions. This is achieved by combining the separation and visualisation power of 2DE, with the cross-species identification and quantitation advantages of metabolic labelling [313], where protein identification is performed using conventional (Mascot) and homology-based (MS BLAST) techniques. Despite technical advancements and algorithm development for identifying proteins from unsequenced organisms in complex protein samples, simplifying the proteome (using 2DE here) remains favourable because Mw and pI characteristics can be used as additional identifiers [103]. A major advantage of using metabolic labelling is improved confidence in identification, which is important for unsequenced organisms, as well as accurate quantitation [313]. The difference in mass between

labelled and unlabelled versions of the same peptide provides extra identification evidence, via imposition of an elemental constraint. Using this confirmation method here demonstrates its utility in cross-species proteomics. A major drawback in the majority of present proteomic studies is the time taken to get a meaningful insight into protein abundance changes from initial protein extraction to quantitation and identification data. In Part A, using a pre-screen step means only proteins of interest (differentially expressed) are selected for quantitation on the MS, saving time and cost, whilst increasing quantitative accuracy. Further validation of protein identification and subsequent proof of presence in the *Euhalothece* genome, is also examined through the design of degenerate primers against three parent genes, and by sequencing of the resultant PCR-generated amplicons. A comparison of these sequences with homologous genes in the NCBI non-redundant database also indicates the potential for identifying proteins based on sequence similarity to fully sequenced organisms. Hence, whilst Part A is focused on revealing the potential to perform a more in-depth proteome study on *Euhalothece*, and on the salt adaptation strategies discovered, the techniques and approach presented will provide a valuable protocol for quantitatively screening the proteomes of unsequenced environmental strains in a fast and effective manner.

In Part B, the success of the technique is exploited and developed further to attain a large-scale data set. Increasing the proteome coverage allows a more detailed analysis of exactly how protein groups and metabolic pathways are involved in acclimation to salt stress by looking for similar changes in linked functional pathways. Part B also extends *Euhalothece* adaptation to very high salt (9% w/v).

PART A

4.2 Materials and methods

4.2.1 Culture conditions and cell growth

Axenic cells of *Euhalothece* sp. BAA001 (previously identified using 16S rRNA sequencing, NCBI accession number AY457568) were grown in batch culture in 250 ml flasks at 24°C in modified BG11 medium [135]. Three separate cultures differing in salt (NaCl) concentration were grown, low (0% w/v), medium (3% w/v)

and high salt (6% w/v). Two biological replicate cultures of all three phenotypes were generated. The 3% culture was assigned as the reference culture, where the major nitrogen source, sodium nitrate ($NaNO_3$) was replaced with heavy sodium nitrate ($Na^{15}NO_3$) for the metabolic labelling strategy. *Euhalothece* does not have the ability to fix nitrogen, and this was confirmed using $NaNO_3$ free media, which resulted in cell death due to nitrogen starvation. Ferric ammonium citrate was also replaced with ferric sulphate in order to eliminate the extra source of unlabelled nitrogen. The 3% salt phenotype was chosen as a reference culture because *Euhalothece* grows optimally in this salt concentration (section 4.3.1), and it is represents a mid-point for both remaining salt cultures.

To maximise labelling efficiency of the ^{15}N isotope, bacterial cells were grown in the appropriate media for at least eight doubling times [314]. All cultures were buffered to pH 7.4 with 50 mM MOPS, and illuminated on a 12 hour light-dark cycle at 90 μEinsteins m^{-2} s^{-1} with continuous shaking at 220 rpm. Optical density measurements were used to follow cell growth at a wavelength of 530nm using an UltraSpec 2100 Pro spectrophotometer (Amersham Biosciences, Buckinghamshire, UK). Chlorophyll *a* concentration was measured at the time of harvesting following the method of Mackinney *et al.* [315].

4.2.2 Overview of proteomic workflow

The proteomic workflow followed in Part A is shown in Figure 4.1, with specific material and methods for each step given below. In summary, two cell mixtures were made (1) 0% and 3% salt (2) 3% and 6% salt, with the 3% salt culture as the labelled reference. These mixtures were subsequently run on separate 2DE gels for protein separation, and then compared via densitometry-based quantitation as a pre-screen. Due to equal amounts of the reference phenotype (3% salt) on each gel, densitometry-based comparisons were between the 0% and 6% salt phenotypes. By running two different phenotypes on each gel (Figure 4.1) with a reference phenotype in common, inter-gel variation can be reduced, thus increasing quantitation reproducibility [55, 313]. After gel spot excision, tryptic digestion and MS analysis, a comparison of the intensities of the peptides generated in the mass spectrum is used to achieve accurate relative quantitation [313]. The difference in

mass between heavy and light versions of the same peptide must correspond to the difference in the number of nitrogen atoms present, therefore providing extra identification evidence via imposition of an elemental constraint. Homology-based identification using MS BLAST, was used together with Mascot for protein identification, as described below

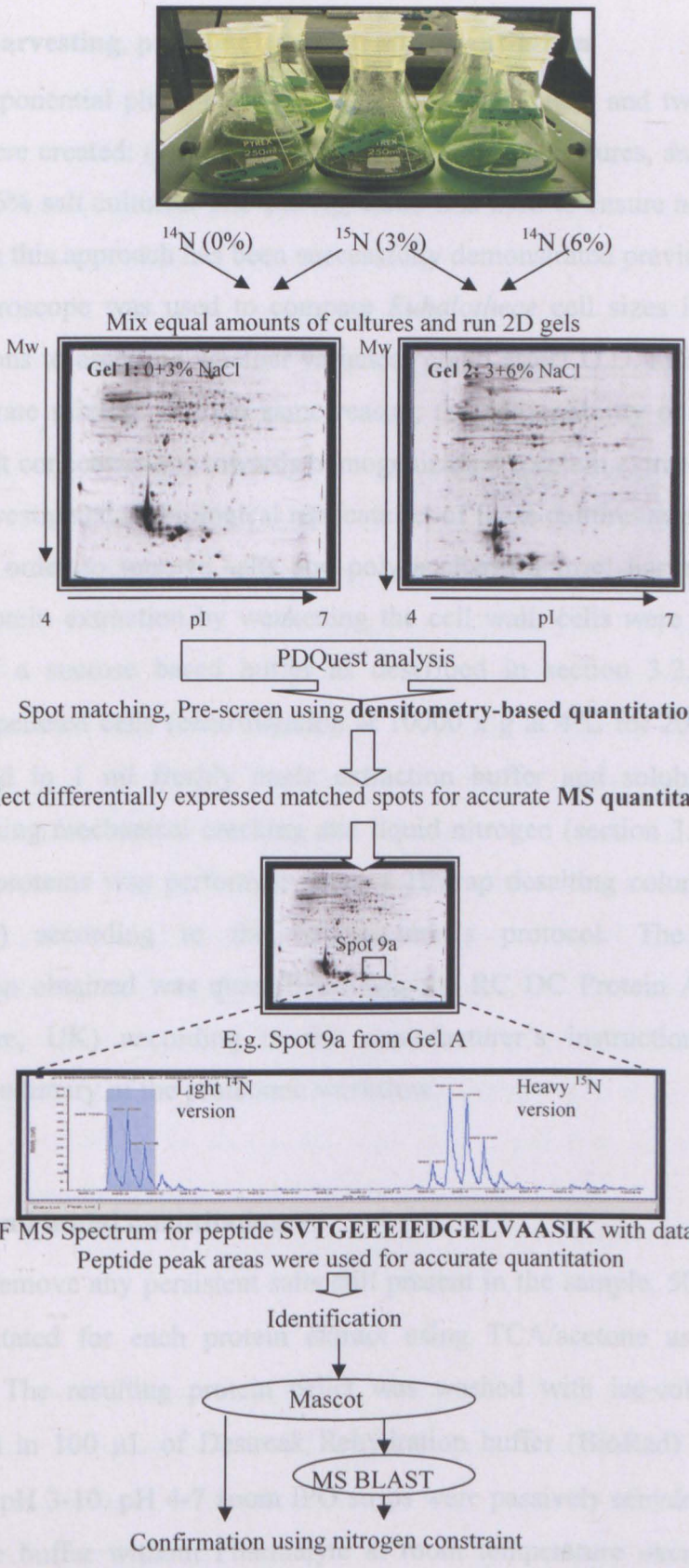


Figure 4.1: Overview of the experimental procedure (NB. this procedure was performed with two biological replicates).

4.2.3 Cell harvesting, protein extraction and quantitation

Mid-late exponential phase ($O.D._{530}$ = 1.3) cells were taken and two different cell mixtures were created: (1) a mixture of 0% and 3% salt cultures, and (2) a mixture of 3% and 6% salt cultures. The $O.D._{530}$ value was used to ensure accurate mixing, and because this approach has been successfully demonstrated previously [313]. An optical microscope was used to compare *Euhalothece* cell sizes in different salt concentrations to ascertain whether variances could affect O.D. measurements and hence accurate mixing. For the same reason, the susceptibility of cells grown in different salt concentrations towards homogenization (protein extraction procedure) was also investigated. A biological replicate set of these cultures was also treated in parallel. In order to remove salts and polysaccharides from harvested cells, and enhance protein extraction by weakening the cell wall, cells were washed with 5 volumes of a sucrose based buffer as described in section 3.2.2. For protein extraction, pelleted cells (centrifugation at 10000 x *g* at 4°C for 20 minutes) were re-suspended in 1 ml freshly made extraction buffer and soluble protein was extracted using mechanical cracking and liquid nitrogen (section 3.2.2). Desalting of soluble proteins was performed using a HiTrap desalting column (Amersham Biosciences) according to the manufacturer's protocol. The total protein concentration obtained was quantified using the RC DC Protein Assay (BioRad, Hertfordshire, UK) according to the manufacturer's instructions. Figure 4.1 provides a summary of the proteomic workflow.

4.2.4 2DE and initial quantitation

To further remove any persistent salts still present in the sample, 500µg of protein was precipitated for each protein extract using TCA/acetone as performed in Chapter 3. The resulting protein pellet was washed with ice-cold acetone and resuspended in 100 µL of Destreak Rehydration buffer (BioRad) containing 2% Pharmalyte pH 3-10. pH 4-7 zoom IPG strips were passively rehydrated in 330 µL of the same buffer without Pharmalyte at room temperature overnight. Proteins were loaded onto the IPG strip (BioRad) using a cup-loading technique during the focusing program [43] on the Protean II IEF cell (BioRad), which had four main steps: (i) an initial 9 hour desalting step at 50V (ii) 200V for 1 hour (iii) 1000V for

1 hour (iv) a 6 hour linear ramp to 10,000V and maintenance of this voltage for a further 1.5 hours [316].

After the first dimension separation, incubation of IPG strips in equilibration buffer for 15 minutes with shaking reduced cysteine residues in proteins. The equilibration buffer consisted of 0.05M Tris at pH 8.8, 6M urea, glycerol 30% (v/v), 2% SDS (w/v) and 2% (w/v) DTT. Alkylation was performed in the dark with shaking using the equilibration buffer with 2.5% (w/v) iodoacetamide. Equilibrated strips were overlaid in a 1% (w/v) agarose solution layer onto a large format 12% SDS-PAGE gel (17 cm x 17 cm x 0.01 cm) and run using Protean II Multicell (BioRad) apparatus. Electrophoresis was carried out with an initial constant current of 16mA for 30 minutes, followed by a constant current of 24mA for 6 hours. Immediately after electrophoresis, gels were stained overnight with Coomassie Brilliant Blue G250 dye and deionised water was used for destaining. Technical (2) and biological replicate (2) gels were run, and biological replicate spots were selected for comparison. A GS-800 Densitometer (BioRad) was used to scan gels for quantitation pre-screening, using a resolution of 100 microns. Gel normalisation, spot detection and statistical analyses were performed using PDQuest® v7.2 Image Analysis software (BioRad). This software identifies differentially expressed proteins based on the intensity of protein staining. The pixel intensity of each spot identified by this software was normalized against the total pixel density arising from *all* the protein spots on the gel. The normalised spot volume is therefore calculated using equation 4.1

$$NV_n = [V_n / V_t] \times 100 \qquad \text{Equation 4.1}$$

Where NV_n = normalised spot volume, V_n = volume of spot n, V_t = total volume of all spots. No internal protein standards of known concentration were included making the experiment semi-quantitative at this stage. Differentially expressed spots (see section 4.3.2) were selected for MS identification and MS quantitation.

4.2.5 Protein isolation and identification

Selected spots were excised from the gels, destained and tryptically digested overnight in a ZipTip®μ-C18 96 well plate (Millipore, Watford, UK). The peptides were extracted and cleaned in sequential steps, as described in the manufacturer's protocol. Each sample was dried in a vacuum centrifuge (Model 5301, Eppendorf, Cambridgeshire, UK) at room temperature, and then stored at -20°C prior to MS analysis. Peptides were initially separated on a Pep-Map C-18 RP capillary column (Famos, Switchos and Ultimate LC system) from Dionex/LC Packings, Amsterdam, Netherlands) interfaced to a QStar XL Hybrid ESI-qQ-TOF-MS/MS (AB, Framingham, MA, USA; MDS- Sciex, Concord, Ontario, Canada). A 35-minute gradient was used, starting with 5% Buffer B (0.1% formic acid in 95% acetonitrile) and 95% Buffer A (0.1% formic acid in 5% acetonitrile) for 3 minutes, followed by a ramping to 50% Buffer B over 30 mins, then 90% Buffer B for 5 min and ending with 5% Buffer B for a further 5 mins. A flow rate of 300 nL/min was used. The MS detector mass range was set to 300 to 2000 m/z, and data acquisition was performed in the positive ion mode using information dependent acquisition (IDA). Peptides with +2 and +3 charge states were selected for MS/MS. IDA data was initially searched using Mascot 2.0 (Matrix Science) in a sequence query type of search based on MS/MS spectra against the regularly updated entire non-redundant database downloaded from NCBI. Search parameters included a 1.2 Da peptide tolerance and 0.6 Da MS/MS tolerance. One mis-cleavage of trypsin was allowed. A carbamidomethyl modification of cysteine was set as a fixed modification, and methionine oxidation as a variable modification. The default setting was used to gain a high confidence Mascot identification, where a MOWSE score of >38 is equal to a significant hit of $p<0.05$ [93]. The database search protocol, MS BLAST, was also used for protein identification using sequences generated by the interpretation of MS/MS spectra (dove.embl-heidelberg.de/Blast2/msblast.html) [102]. Significant identifications in MS BLAST are colour-coded based on high scoring pairs (HSP's) and placed at the top of the output list [101]. If a positive identification was not made using both aforementioned methods for a sample, the peptide mixture was re-run on the MS.

4.2.6 MS quantitation and identification confirmation

The LC-MS reconstruct tool in Analyst Software (version QS 1.1, AB) was used to integrate peptide peak areas in the TOF-MS scans over time. Two sets of peaks appeared in each TOF-MS spectrum, corresponding to the ^{14}N labelled and ^{15}N labelled phenotype, which were co-eluting. The intensities (peak areas) of these heavy and light versions of the same peptide were used for quantitation, by calculating the ratio of peak areas. The ratio of intensities and therefore differential expression of all proteins were calculated. However, proteins were only treated as significantly differentially expressed with fold changes of 1.5 or greater (see section 4.3.7). In addition, the detection of 2 or more peptides was a prerequisite for quantitation significance (which ensured greater confidence in the accuracy of the ratio). Statistical evidence was improved by also subjecting the same spots excised from a separate biological replicate gel to the same workflow and stringent criteria.

The mass difference of the monoisotopic peak for the heavy and light version of the same peptide (as demonstrated in Figure 4.8) could subsequently be calculated and compared to the number of nitrogen atoms present in the identified peptide, using an online Proteomics Toolkit (http://db.systemsbiology.net:8080/proteomicsToolkit/IsotopeServlet.html). A match thus increases confidence in identification due to the imposition of the N-constraint.

4.2.7 Protein data analysis

All identified proteins were organised into functional groups using annotation based on clusters of orthologous groups (COG's) in the major organism used to identify proteins. The cellular location of these proteins was predicted using the program PSORTb v2.0 [248]. Lipoproteins are membrane proteins and cannot be predicted using this software, and therefore LipoP v1.0 [249] was used, which identifies a lipoprotein by a specific signal sequence.

4.2.8 DNA extraction, PCR studies and 16S rRNA phylogenetic tree construction

DNA was extracted from *Euhalothece* cells using the MoBio (Carlsbad, CA, USA) Ultraclean™ Microbial DNA Isolation Kit. Prior to this protocol, bacterial cells were washed three times with 5 volumes of sucrose based buffer. A spectrophotometer (UltraSpec 2100 Pro) was used to measure the concentration of eluted DNA. DNA was also extracted from *Synechocystis* cells to be used as a positive control in PCR experiments. The sequences of primers for cyanobacterial specific 16S rRNA sequences used were forward (CYA106) 5'-CGG ACG GGT GAG TAA CGC GTG A-3', and reverse (CYA781) 5'-GAC TAC TGG GGT ATC TAA TCC CAT T-3' [134]. Degenerate primers were designed to amplify the *Euhalothece lexA* gene using multiple sequence alignments from cyanobacteria, generated using a BLASTx search of the *Synechocystis lexA* gene. The forward primer sequence was 5'-AAG TCG A/T TT A/G GAAC G/A G/T TTACG C-3' and reverse primer 5'-CC C/A T C/T C/A A T/C A/T CTGGC G/A GCGAC-3'. These primers were used in a temperature gradient PCR reaction (TGradient, Whatman Biometra® Goettingen, Germany) with the following program: initial denaturation at 94°C for 4 minutes, followed by 30 cycles of 94°C for 30 seconds, 40°C to 60°C for 30 seconds and 68°C for 30 seconds. The PCR product was run on a 1.5% agarose gel, and the DNA (expected size was ca. 300bp) was extracted using Qiaex II Gel Extraction Kit (Qiagen, West Sussex, U.K.). Cloning was carried out using a Strataclone™ PCR Cloning Kit (Stratagene, Amsterdam, The Netherlands). Positive clones were screened using M-13 vector primers in a PCR reaction (94°C for 10 minutes, followed by 25 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 68°C for 1 minute). Positive clones were grown overnight at 37°C in LB-broth and 50 μg/ml ampicillin. Plasmid DNA was prepared for sequencing using a QIAprep Spin Miniprep Kit (Qiagen). Genetic information was submitted to NCBI (accession number DQ845140).

The same procedure was followed using degenerate primers designed and constructed for PPP enzyme, phosphoribulokinase and anti-oxidative enzyme superoxide dismutase. The forward primer sequence for phosphoribulokinase was 5'- G/T AT C/T GG A/C/T GT A/T GCCGG A/G/T GA C/T C-3' and reverse

primer sequence 5'- C/T TC A/G TA A/G A/G C A/C/T G/T C A/C CA TTT TC - 3', and the forward primer sequence for superoxide dismutase was 5'- TAC G A/C C A/C/T A/T C A/G A/C C/T GCG/T CT C/GGA A/G C -3' and reverse primer sequence 5'- C/T AG GC C/G GC G/T GC C/T A A/G A/G TT A/G GC -3'.

A 16S rRNA phylogenetic tree for *Euhalothece* was constructed using 16S rRNA sequences with Mega3 v1.5 software (see Figure 4.4). The neighbour joining method was implemented and a bootstrap analysis using 1000 repeat samplings was performed [317]. *E.coli* strain K12 was used as an outgroup.

4.3 Results and discussion

4.3.1 Growth studies

Euhalothece was cultured in the laboratory with 90 μEinsteins m^{-2} s^{-1} of incident light at 24°C with continuous shaking, in three salt concentrations, low (0%), medium (3%) and high (6%). It grew optimally in 3% NaCl (0.075 ± 0.001 d^{-1}), with growth rates of 0.068 ± 0.005 d^{-1} and 0.052 ± 0.002 d^{-1} at 0% and 6%, respectively (see Figure 4.2). Not only did cultures grown in 6% salt have a reduced growth rate, but also tended to reach the death phase up to 30% faster than 0% and 3% salt cultures. Furthermore, pigmentation yellowed towards the end of exponential phase for the higher salt concentration.

Euhalothece has an optimum doubling time of 9.24 days under these conditions, and therefore using metabolic labelling is particularly attractive, as two phenotypes are run together on one gel resulting in a reduced biomass requirement for the whole experiment compared to a conventional 2DE proteomics experiment. The ^{15}N incorporation efficiency was calculated using open source software Isopro version 3.0 and by using three peptides from the highly abundant protein, phycocyanin (Gel B, spot 2, Figure 4.5), it was found to be 98.45% ± 0.21. It has been shown previously that the incorporation efficiency is consistent from protein to protein [314]. The $O.D._{530nm}$ was measured to ensure accurate mixing of cultures, since a previous study showed that errors associated with this method of mixing were approximately 1% [318]. O.D. measurements in this study revealed less than 1%

mixing error, and therefore alternative approaches, for example, normalisation against an average ratio calculated from a large data set of $^{14}N/^{15}N$ ratios, was not required.

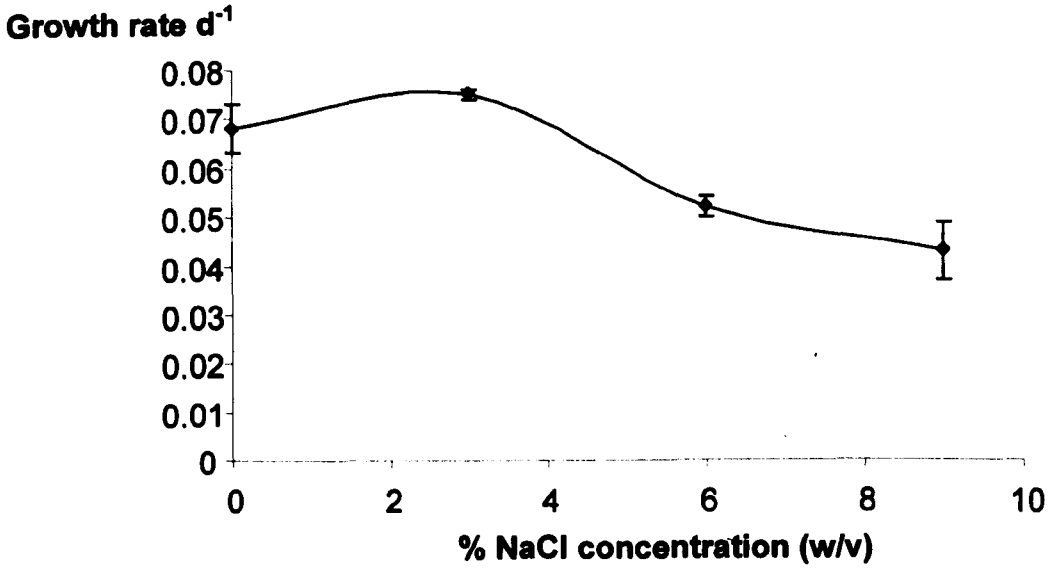


Figure 4.2 Growth rates of *Euhalothece* in varying salt concentrations in BG11 media at pH 7.4 and 24°C.

Although this method of cell mixing was shown to be suitable in terms of cell density, is not certain at this stage whether the protein amounts combined are uniform. For example, if cell size varies across the three salt conditions, it will interfere with protein amounts in both cell mixtures. Significant changes in cell size have been recorded in several *Halothece* strains, but only at 10% salinity and more, and the size can either increase, decrease, or have no significant change [151]. Therefore an optical microscope (Zeiss Axioplan 2, Bernried, Germany, 10x ocular, 100x objective) was used to measure the cell length (diameter) for more than 20 cells (in biological replicate cultures) for four salt concentrations (0, 3, 6 and 9% NaCl), taken during the mid-exponential growth phase. The effects of increasing salt concentration on cell diameter are presented in Figure 4.3. An unpaired *t*-test was conducted to elucidate whether differences in cell diameter were sufficiently significant to affect equal cell protein mixing (0% with 3% salt mixture and 3% salt with 6% salt mixture) and both were found to be statistically insignificant ($p \leq 0.05$).

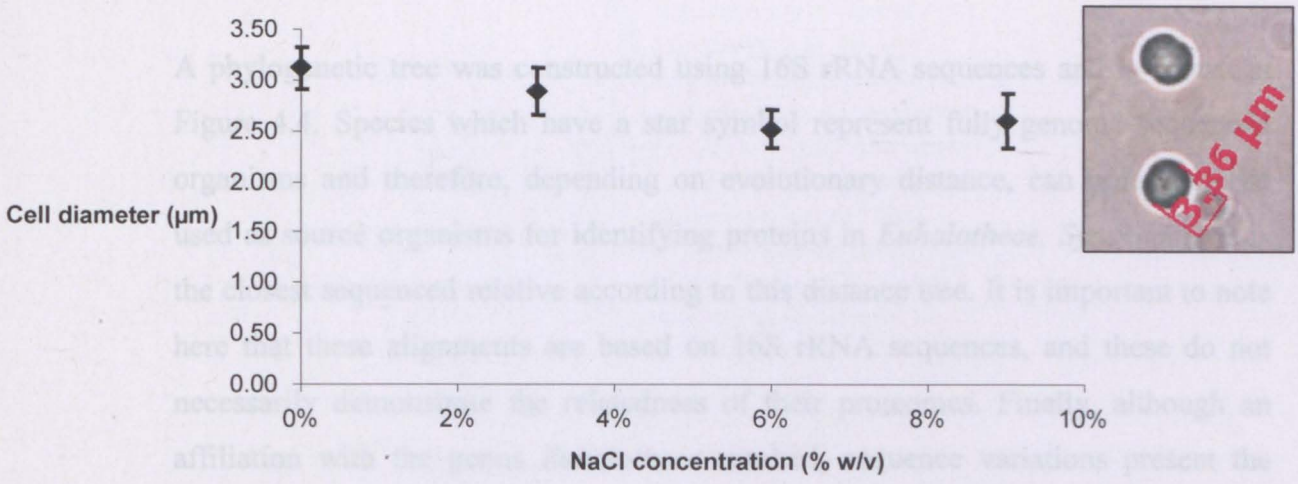


Figure 4.3: *Euhalothece* cell length variation with changing salt concentrations. Inset: photograph of *Euhalothece*, with cell size measurement using Zeiss Axiovision digital imaging software (10 x ocular, 100 x objective).

The method used in this study utilises the capacity of gel analysis software to normalise for minor changes in protein loading and small technical variations. Hence, including this extra verification, the method adopted for quantitation was deemed acceptable.

In addition, to validate the effect of salt on the susceptibility of cells towards homogenization, protein extractions were performed in duplicate (technical replicate) using four salt concentrations (0, 3, 6 and 9% NaCl) and the results are shown in Table 4.1. The total amount of protein extracted for each phenotype was all within a similar range.

Table 4.1: Protein extraction from 50 ml *Euhalothece* cell cultures ($O.D._{530nm}$ 1.3)

NaCl (%)	Protein concentration (mg/ml)	
	Average	Standard deviation
0	2.29	0.39
3	2.50	0.50
6	2.04	0.47
9	2.16	0.13

A phylogenetic tree was constructed using 16S rRNA sequences and is shown in Figure 4.4. Species which have a star symbol represent fully genome sequenced organisms and therefore, depending on evolutionary distance, can potentially be used as source organisms for identifying proteins in *Euhalothece*. *Synechocystis* is the closest sequenced relative according to this distance tree. It is important to note here that these alignments are based on 16S rRNA sequences, and these do not necessarily demonstrate the relatedness of their proteomes. Finally, although an affiliation with the genus *Euhalothece* resulted, sequence variations present the possibility of a new species.

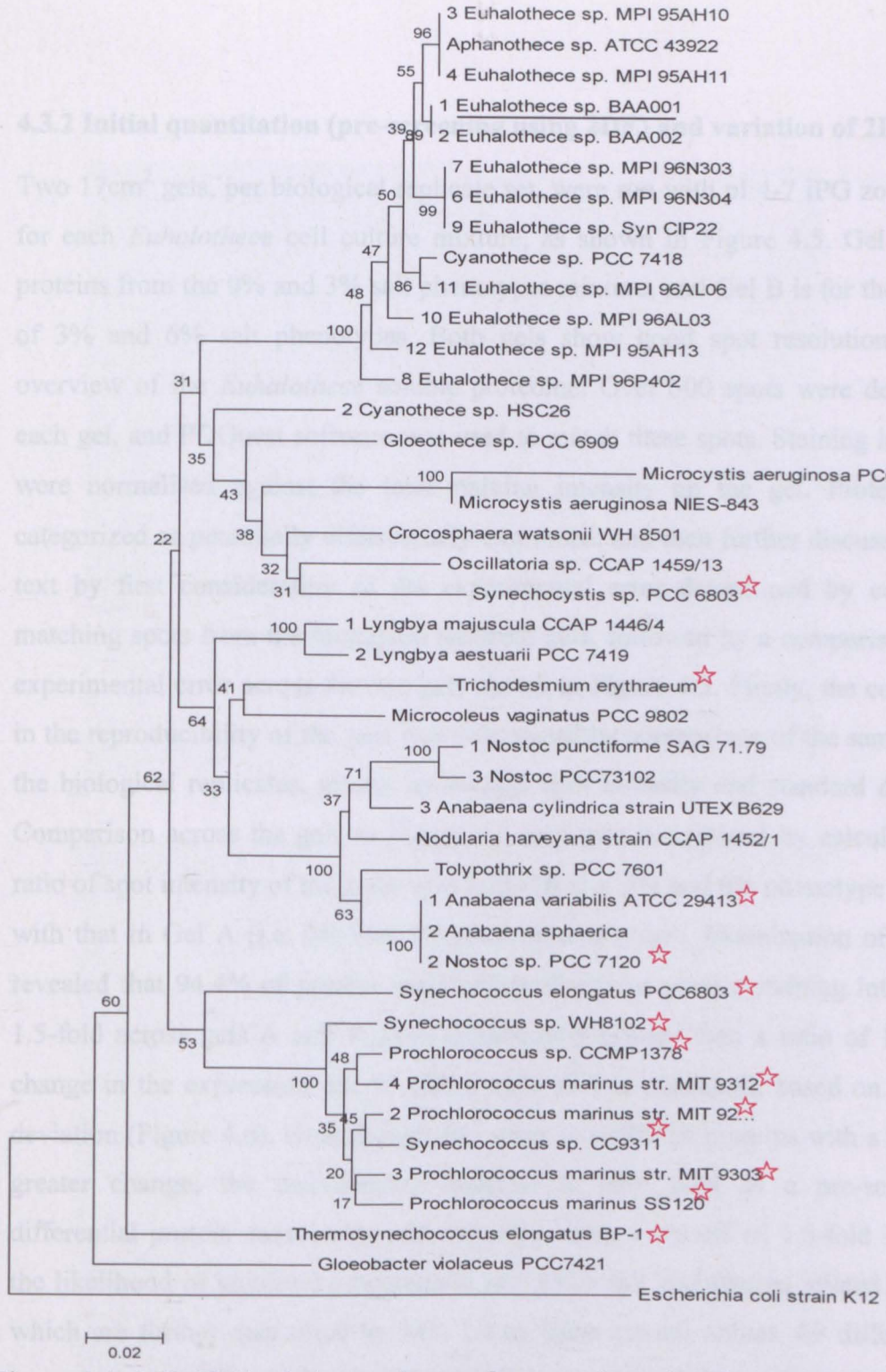


Figure 4.4: 16S rRNA phylogenetic tree for cyanobacteria. The neighbour joining method was used with Tajima-Nei calibration and probability analysis by bootstrap technique using 1000 repeat samplings [317]. *E. coli* K12 was employed as the outgroup. ☆ : symbol denotes species of cyanobacteria which have their genomes fully sequenced and numbers pre-fixed to organisms were added to present different strains of the same species. (*Euhalothece* sp. BAA001 is underlined).

4.3.2 Initial quantitation (pre-screening using 2DE) and variation of 2DE data

Two 17cm^2 gels, per biological replicate set, were run with pI 4-7 IPG zoom strips for each *Euhalothece* cell culture mixture, as shown in Figure 4.5. Gel A is for proteins from the 0% and 3% salt phenotypes mixture, and Gel B is for the mixture of 3% and 6% salt phenotypes. Both gels show good spot resolution, and an overview of the *Euhalothece* soluble proteome. Over 500 spots were detected in each gel, and PDQuest software was used to match these spots. Staining intensities were normalised against the total staining intensity on the gel. Proteins were categorized as potentially differentially expressed, and then further discussed in the text by first consideration of the experimental error determined by comparing matching spots from the biological replicate gels, followed by a comparison of the experimental error across the two gels shown in Figure 4.5. Firstly, the confidence in the reproducibility of the gels was determined by comparison of the same spot in the biological replicates, giving an average spot intensity and standard deviation. Comparison across the gels in Figure 4.5 was then determined by calculating the ratio of spot intensity of the same spot in Gel B (i.e. 3% and 6% phenotype mixture) with that in Gel A (i.e. 0% and 3% phenotype mixture). Examination of the data revealed that 94.4% of protein spots which showed a greater staining intensity of 1.5-fold across gels A and B were significantly greater than a ratio of 1 (for no change in the expression due to salt, a ratio of 1 is expected), based on standard deviation (Figure 4.6). Even though this value is 100% for proteins with a 2-fold or greater change, the densitometry analysis is only used as a pre-screen for differential protein expression, and therefore using a cut-off of 1.5-fold increases the likelihood of identifying potentially important salt acclimation related proteins, which are further quantified by MS. Using these cut-off values, 69 differentially expressed spots from each gel (138 in total) were selected during this pre-screen for identification and further quantitation using isotope labels.

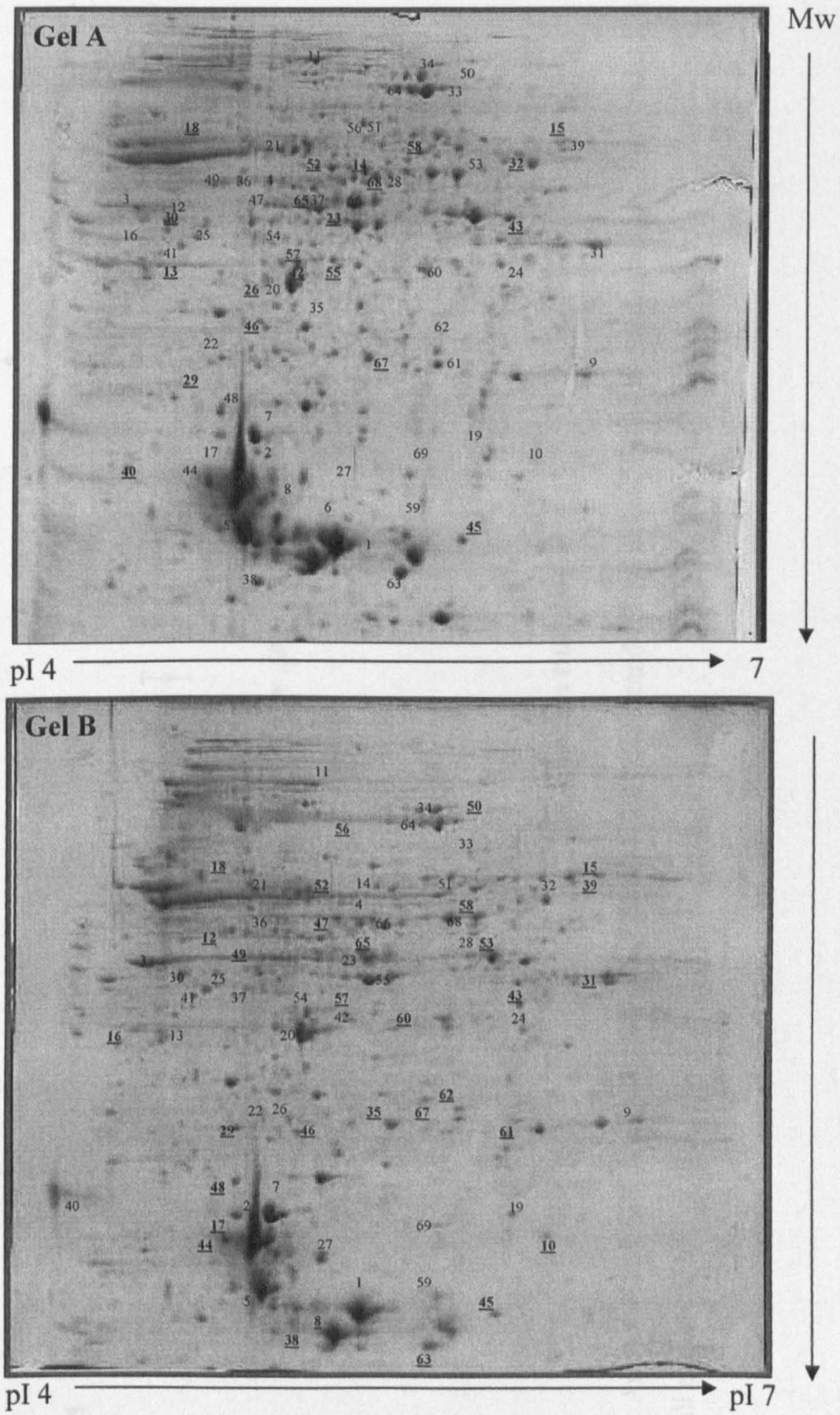


Figure 4.5: Proteome of *Euhalothece*. 17cm^2 gels, pI 4-7, stained with Coomassie Brilliant Blue. Gel A contains isotopically labelled reference culture (3%) with 0% salt culture and Gel B contains isotopically labelled reference culture (3%) with 6% salt culture. Spots which were not identified are shown in **bold** type and underlined.

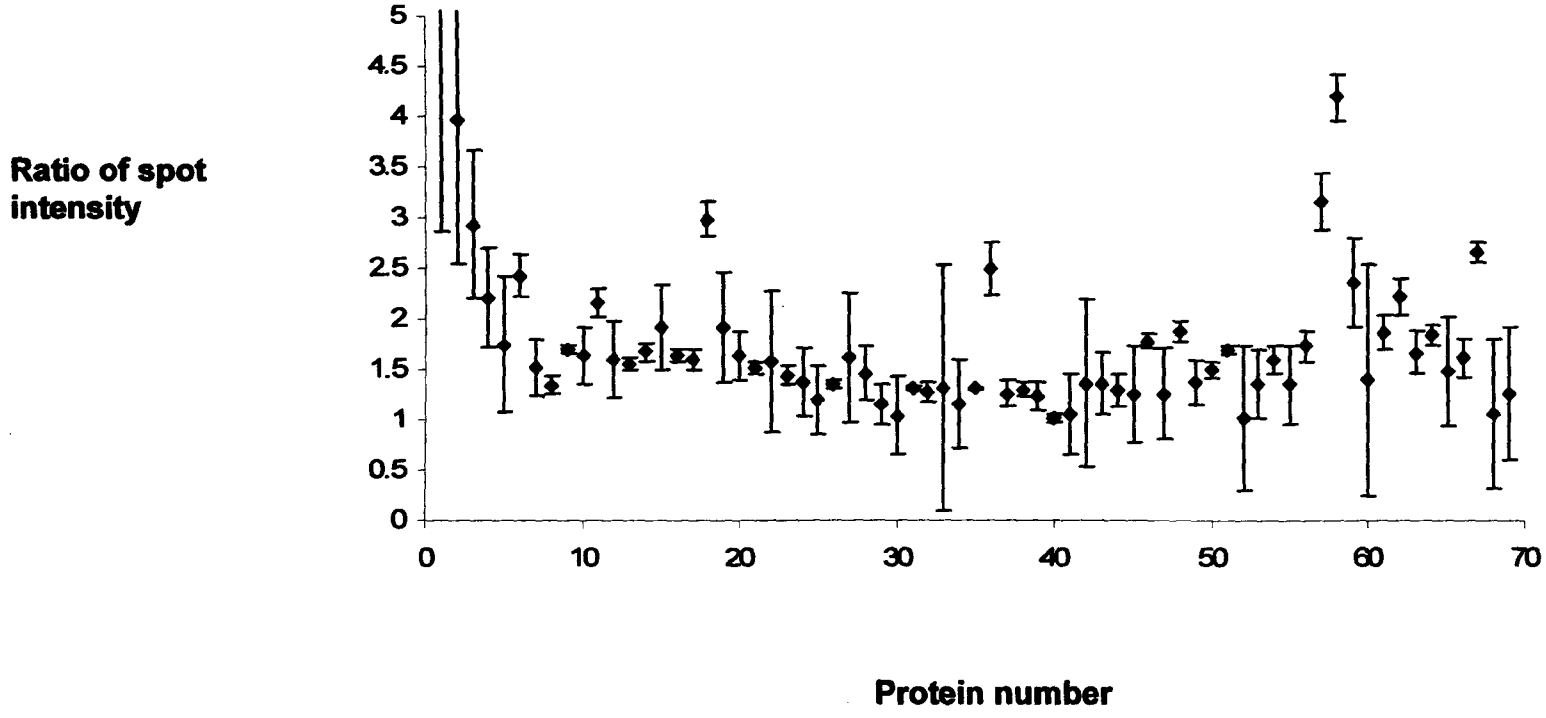


Figure 4.6: An error chart based on ratio of relative quantitations with standard deviations, calculated using densitometry data. Relative ratios are all expressed higher than one.

Table 4.2: Percentage of proteins which are statistically (ratio values with standard deviations which do not cross one) differentially expressed when using a variety of cut-off values in densitometry-based data (calculations based on biological replicates shown in Figure 4.6).

Cut off ratio	No. of proteins do NOT cross 1	Total no. of proteins	% proteins do NOT cross 1
3	5	5	100
2	13	13	100
1.5	34	36	94.4
1.4	38	41	92.7
1.3	44	52	84.6
1.2	50	62	80.6
1.1	51	66	77.3

At this stage, the study followed the work flow of a conventional proteomics experiment, selecting spots which statistically changed in staining intensity for subsequent MS analysis. By comparing the staining intensity between the two gels, differences between the 0% and 6% salt phenotype were investigated, since the reference phenotype (3%) is present in equal amounts on each gel.

4.3.3 Desalting samples for 2DE

Salt did not interfere with the 2DE process, as several desalting steps based on techniques available in the literature [316, 319] and developed in-house, were combined before protein separation and during the first dimension (where high conductivity is used). For example, prior to protein extraction from bacterial cells, external residual salt was removed by thoroughly washing the cells with a sucrose based buffer (refer to section 4.2.3). The size exclusion matrix in the HiTrap® Desalting column (Amersham Biosciences) was then utilised to further remove salt from proteins extracted from cell lysates. A TCA/acetone precipitation also ensured only low salt concentrations would be present in protein samples prior to IEF. During IEF, cup loading was used to introduce proteins, which allows a concentration maximum of 50 mM salt [319]. A linear pH gradient was stabilised across the gel using an increased concentration of Pharmalyte (2%), reducing charge-charge interactions during the IEF program. Finally, a 9-hour, low voltage

(50V) desalting step was introduced into the IEF program [316], which assists focusing by migrating salt ions to the ends of isoelectric focusing strips. A salt removal workflow is presented in Figure 4.7 together with gel images which illustrate visually the effects of salt on gel quality when desalting steps described above were not implemented. Advantages of this method include simplicity, limited samples loss, minimal dilution and low implementation costs compared to alternative methods, for example, dialysis or ultrafiltration [320-322].

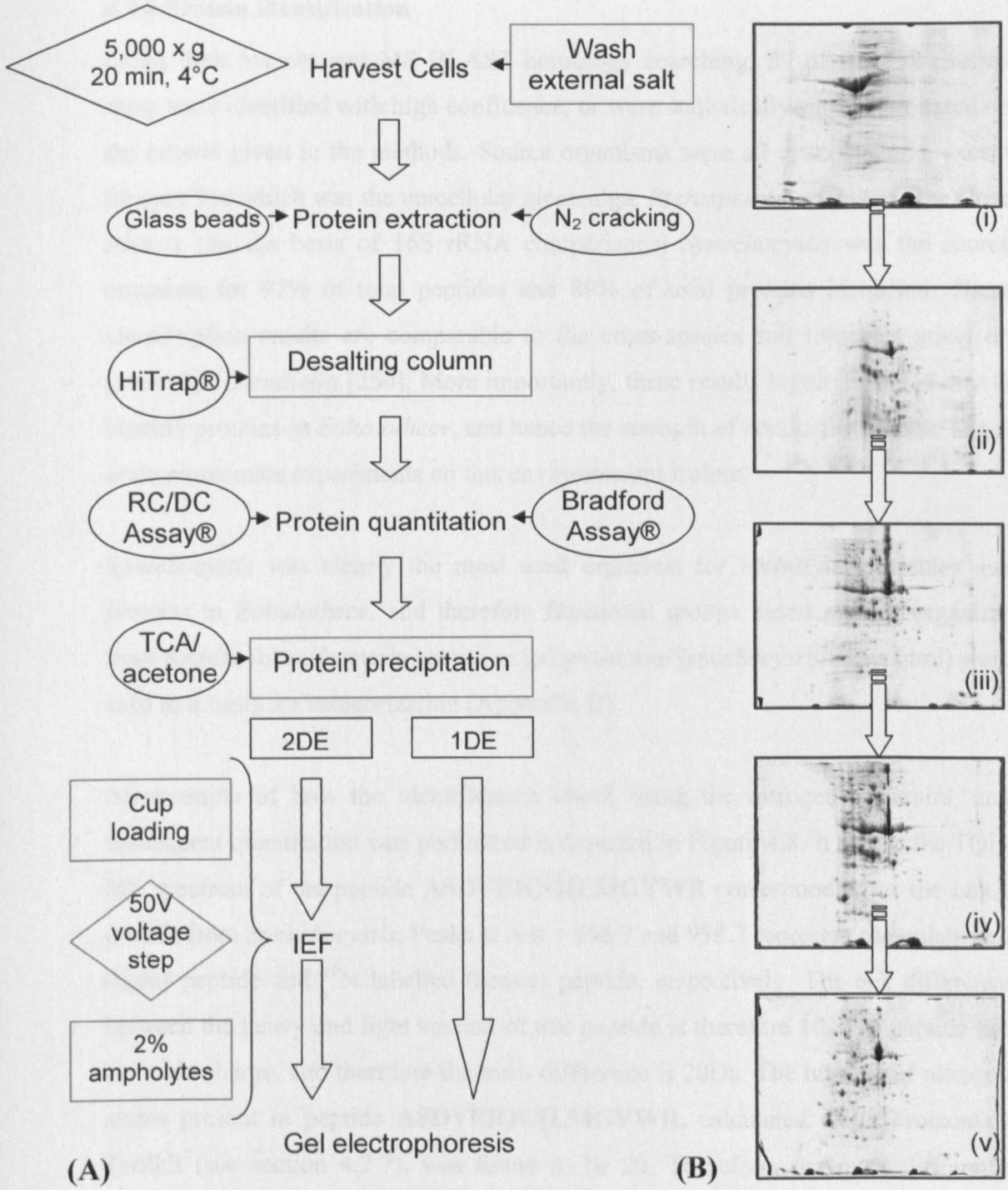


Figure 4.7: (A) A simple salt removal strategy for proteomic experiments using gel electrophoresis for protein fractionation. Desalting steps are in green boxes, example techniques in circular boxes and parameters in diamond boxes. N.B. For halophilic organisms the wash step with no salt buffer must be performed at the protein extraction stage due to cell lyses issues. (B) *Euhalothece* desalting optimisation tests. $7cm^2$ gels, pI 3-10. i) No desalting steps ii) TCA/acetone precipitation only iii) Desalting column not used iv) Washing buffer step not performed v) Complete desalting procedure followed.

4.3.4 Protein identification

Using both Mascot and MS BLAST homology searching, 89 of the 138 excised spots were identified with high confidence, or were statistically significant based on the criteria given in the methods. Source organisms were all cyanobacteria, except for spot 31a which was the unicellular green alga, *Pterosperma cristatum*. The close relative (on the basis of 16S rRNA comparisons) *Synechocystis* was the source organism for 97% of total peptides and 89% of total proteins identified. These identification results are comparable to the cross-species salt tolerance study on green alga *Dunaliella* [250]. More importantly, these results highlight the ability to identify proteins in *Euhalothece*, and hence the strength of conducting further large-scale proteomics experiments on this environmental isolate.

Synechocystis was clearly the most used organism for identifying peptides and proteins in *Euhalothece*, and therefore functional groups based on this organism from Kazusa (http://bacteria.kazusa.or.jp/cyanobase/*Synechocystis*/index.html) were used as a basis for categorization (Appendix B).

An example of how the identification check using the nitrogen-constraint, and subsequent quantitation was performed is depicted in Figure 4.8. It shows the TOF-MS spectrum of the peptide **ASDVEIQGILMGVWR** corresponding to the LexA protein from *Synechocystis*. Peaks at *m/z* = 988.7 and 998.7 represent the unlabelled (light) peptide and ^{15}N labelled (heavy) peptide, respectively. The *m/z* difference between the heavy and light version of this peptide is therefore 10. The peptide has a double charge, and therefore the mass difference is 20Da. The number of nitrogen atoms present in peptide **ASDVEIQGILMGVWR,** calculated using Proteomics Toolkit (see section 4.2.7), was found to be 20. Therefore, the mass difference corresponds exactly to the number of nitrogen atoms present in the peptide, increasing confidence in the I.D. of this peptide, and hence I.D. of the protein. Protein hits in MS BLAST provided peptide sequences, together with a direct link to Analyst Software v1.1 for analysis using the aforementioned procedure (Figure 4.8).

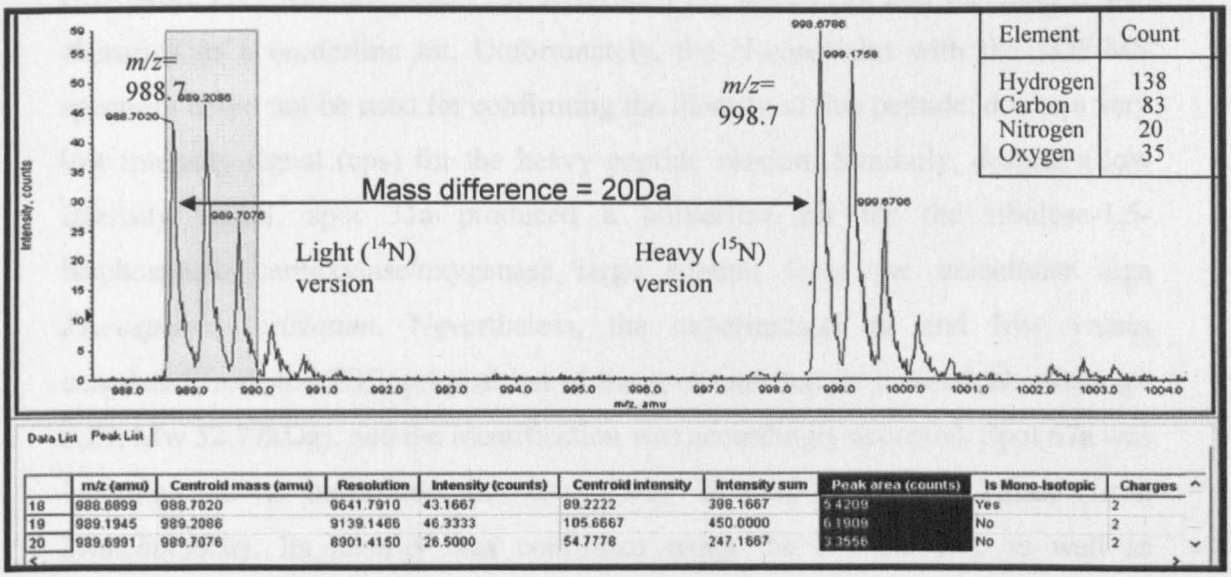


Figure 4.8: TOF-MS spectrum of the peptide **ASDVEIQGILMGVWR** showing both isotopic versions. *m/z* difference is 10; peptide charge is +2; hence mass difference is 20Da. Spectrum data is shown at the bottom of the figure with peak areas (counts) highlighted. The three highlighted peak area values (5.4209, 6.1909 and 3.3556) correspond to the highlighted peaks for the light (^{14}N) version of the peptide. These areas were summed for quantitation of this peptide (Gel A, spot 9, Figure 4.5). Inset: elemental composition of peptide.

Of the 69 proteins pairs selected for further identification and quantitation (using 2DE pre-screen), 32 pairs were either identified as the same protein, or both were unidentified. For the remaining 37 pairs, different proteins were identified on only four occasions (spots 3, 11, 21, 33), and two of these were due to identification of very highly abundant photosynthetic pigments, such as phycocyanin (spot 3) and allophycocyanin (spot 21). Additionally, the positions of these proteins in the gels (spots 3 and 21) did not match their theoretical pI or Mw value, implying a smearing effect in the gel, or protein isoforms. These results show the reliability of protein matching and protein identifications, since on the occasions when both spot pairs were positively identified, the proteins matched 87.9% of the time. MS BLAST was used when either; Mascot could not identify the protein, to confirm Mascot hits where only 1 peptide was found, or finally where both the pI and Mw did not match gel positioning. For example, the homology search identified spot

29b as protein ssr0109 from *Synechocystis*, whereas Mascot was unable to. The HSP score (71) was just below the significant hit score (73), and therefore it was classified as a borderline hit. Unfortunately, the N-constraint with the TOF-MS spectrum could not be used for confirming the identity of this peptide, due to a very low intensity signal (cps) for the heavy peptide version. Similarly, despite a low intensity signal, spot 31a produced a borderline hit for the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit from the unicellular alga *Pterosperma cristatum*. Nevertheless, the experimental pI and Mw values calculated from the 2DE gel position of the spot matched its theoretical values (pI 6.29, Mw 52.77kDa), and the identification was accordingly accepted. Spot 67a was identified as a borderline hit using MS BLAST (uridylate kinase from *Synechocystis*). Its identity was confirmed using the N-constraint, as well as matching its pI and Mw values. MS BLAST was unable to confirm the identity of several spots which were identified by Mascot with only one peptide (6a/b, 16a, 35a, 37b, 55b, 57b, 61a and 64a) (Appendix B).

The identification of the remaining 10 spot pairs was not successful (Appendix B). These proteins were not identified due to either sample failure, very low abundance, or to the lack of similar proteins (i.e. peptide othologs) presently stored in databases. Generally, spot positions corresponded well with predicted pI and Mw values. Some proteins were identified across several spots, and this is likely to be due to the presence of protein isoforms, potential PTM's and truncated proteins. An example of this is transketolase (spots 50a and 64a/b). Spots 34a/b, 39a and 41a/b were all identified as translation elongation factors (EF-Tu) with 6, 6, 1, 2, and 1 peptides detected, respectively. However, the Mw and pI only matched for spots 41a/b.

4.3.5 Verification of protein identification

In cross-species proteomics studies, it is important to be able to verify that the identification of proteins using MS generated data with a database search, is reliable, since it is essentially peptide orthologs to sequenced organisms that are being identified by search methods such as Mascot. One approach is to design degenerate nucleotide primers for the protein-coding gene, using the sequence of

the peptides detected by MS, or alignment files generated from nucleotide sequences from related organisms present in nucleotide databases. Successful sequencing of the gene provides verification of the protein identification, and therefore extra confidence in the identification methods implemented. In this study, degenerate primers were designed to amplify the *lexA* gene in *Euhalothece* (see following section). A temperature gradient PCR revealed that a low annealing temperature (42°C) produced a band at approximately the predicted size (280bp), based on the positive control *lexA* gene from *Synechocystis* (Figure 4.9). An alignment of the *lexA* gene from *Euhalothece* and *Synechocystis* showed 86% similarity (http://www.ebi.ac.uk/clustalw/). A translated BLASTx search using the *Euhalothece lexA* gene nucleotide sequence gave 97% identity to the LexA protein (SOS function regulatory protein) in *Synechocystis* sp. PCC 6803 (e-value $4x10^{-52}$) and 47% identity to the same protein in *Cyanothece* sp. CCY0110 (e-value $1x10^{-25}$). These findings validate the use of cross-species proteomic approaches used here to identify and confirm the presence of differentially expressed proteins in unsequenced, environmental isolates. In this case, peptides detected by the MS are othologs of LexA peptides in *Synechocystis*, which subsequently led to obtaining the near complete nucleotide sequence of the LexA protein in *Euhalothece* (accession number DQ845140). The translated nucleotide sequence of the *lexA* gene produced 122 amino acids. Proteogest [323] was used to tryptically digest this sequence and resulted in 28 peptides, 16 of these were common (orthologous) to *Synechocystis* LexA peptides, and three of these peptides were detected by the MS (see Table 4.3). Unfortunately, the addition of the *Euhalothece* LexA protein to the database used to search the MS/MS data did not produce any additional peptide identifications, possibly due to the small size of this protein.

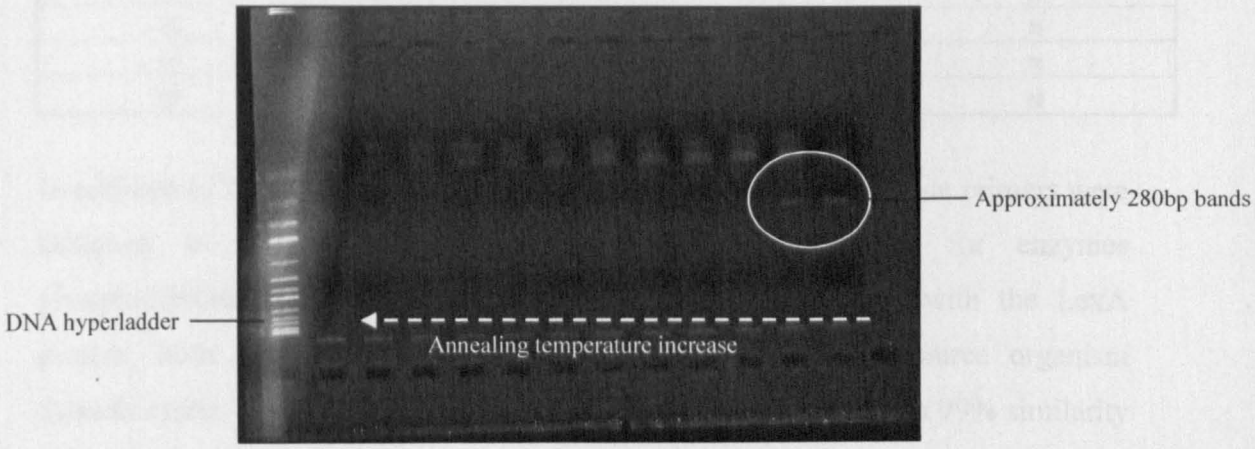


Figure 4.9: Temperature gradient PCR using *lexA* designed degenerate primers in *Euhalothece* (1.5% agarose gel). Bands of the approximate correct size are visible at the lower primer annealing temperatures where less specific binding of primers occurs.

Table 4.3: Tryptic peptides in confirmation sequenced *Euhalothece* LexA protein (**bold type** peptides were identified by MS)

Peptide no.	Tryptic peptides in sequenced *Euhalothece* LexA protein	Tryptic peptides in common with *Synechocystis* sp. PCC 6803 LexA
1	DLVDDFIVEGDMLILRSVTGEEEIEDGELVAAR	N
2	SSVFALRVMSNDLVDDFIVEGDMLILR	Y
3	GVSVIGELKGGELVEADAEEVEK	Y
4	**GGELVEADAEEVEKIDFAPLMK**	Y
5	SVTGEEEIEDGELVAARIDGR	N
6	**VMSNDLVDDFIVEGDMLILR**	Y
7	SVTGEEEIEDGELVAAR	N
8	ILHQKPKGVSVIGELK	Y
9	**GGELVEADAEEVEK**	Y
10	PMWNSPLSRLEHLR	N
11	GYVDWTDGKAR	Y
12	NKGYVDWTDGK	Y
13	TLRILHQKPK	Y
14	IDGRANSTV	N
15	GYVDWTDGK	Y
16	GVSVIGELK	Y
17	PMWNSPLSR	N
18	IDFAPLMKK	Y
19	IDFAPLMK	Y
20	KSSVFALR	Y
21	SSVFALR	Y
22	ILHQKPK	Y
23	LEHLRNK	N

24	ANSTV	N
25	ARTLR	N
26	LEHLR	N
27	IDGR	N
28	TLR	N

In addition to the cloning and sequencing of the *lexA* gene, degenerate primers were designed to amplify, clone and sequence genes coding for enzymes phosphoribulokinase and superoxide dismutase. In conjunction with the LexA protein, both enzymes were identified in *Euhalothece* using source organism *Synechocystis*. The phosphoribulokinase gene in *Euhalothece* shows 99% similarity to *Synechocystis*. A translated BLASTx search using the *Euhalothece* phosphoribulokinase gene nucleotide sequence gave 100% identity (e-value $3x10^{-96}$) to the phosphoribulokinase protein in *Synechocystis* and 87% identity (e-value $1x10^{-88}$) to the same protein in *Cyanothece* sp. CCY0110. 16S rRNA comparisons place *Cyanothece* as a closer relative to *Euhalothece* than *Synechocystis*. *Cyanothece* are freshwater or marine, unicellular cyanobacteria and are capable of fixing atmospheric nitrogen (see Figure 4.4). The superoxide dismutase gene in *Euhalothece* also shows 99% similarity to *Synechocystis*. A translated search gave 98% identity (e-value $9x10^{-98}$) to superoxide dismutase in *Synechocystis*, followed by 81% identity (e-value $4x10^{-82}$) to *Cyanothece* sp. CCY0110.

4.3.6 MS-quantitation using isotopes

69 differentially expressed proteins were detected on each gel using the pre-screen, densitometry-based quantitation. These were selected for further accurate quantitation by MS using the incorporated isotopes. A major advantage of using isotope quantitation here, is that all three phenotypes can be compared, 0% to 3% salt and 3% to 6% salt. Quantitation of protein differential expression was determined by calculating the area under the identified peptide peaks, as shown in Figure 4.8. A comparison of the areas between the labelled and unlabelled peptide was used to produce a quantitation ratio. The confidence in this ratio was calculated using standard deviation based on the ratio of areas under 3 matching peptide peaks (see section 4.3.4). In Figure 4.8, the peaks corresponding to the light (0% salt) peptide (spot 9 in Gel A, Figure 4.5) are smaller than the heavy (3% salt) peptide peaks. For this peptide, the ratio of peak intensity was 1.35, when comparing

growth in 3% salt to 0% salt. It is important to note here that quantitation is performed at the peptide level by MS, and therefore multiple proteins per spot is not a technical problem as it may be when using stain-intensity-based analysis. This is because during densitometry-based quantitation on 2DE gels, if multiple proteins are found per spot, it is not clear which protein is responsible for the change in staining intensity. However, when LC-MS data is used, each peptide will be resolved in time and mass allowing quantitation of each protein. An example of this is in Gel A, spot 7 (Figure 4.5), where both superoxide dismutase and phycocyanin beta chain were identified. The ratio between the heavy (3%) and light (0%) peak for peptide ITGNASAIVSNAAR, belonging to phycocyanin beta chain, was 1.15. The ratio for the peptide RPDYIADFLGK belonging to superoxide dismutase was 1.97 and therefore the change in spot intensity is due to superoxide dismutase.

4.3.7 Variation and normalisation of MS data

To investigate the influence of peptide number on confidence of quantitation, the CV was determined when quantifying 2, 3, 5 and 10 peptides per protein. Peptides with the highest intensity peaks were selected, and results indicate that quantifying 3 peptides produced the lowest average CV value. Average CV's were 0.12, 0.11, 0.12 and 0.16 for 2, 3, 5 and 10 peptides, respectively. Low abundance proteins had the additional problem of noise in the mass spectrum, and the average CV for quantitation of random peptides taken from low abundance proteins was 0.26. To overcome this problem, an increase in amount of protein assayed would be required, achieved by pooling together corresponding spots from different gels. In this study, 3 peptides per protein were quantified for statistical evidence, and peaks with ion intensities below 40 were not included.

The average standard deviation of all protein quantitations was 0.12, and a small value here was important as changes in peptide intensity are potentially dependent upon altered protein digestion caused by trypsin accessibility, and these can be revealed by large standard deviations. Biological replicate gels (2) were also run, and matching spots were excised, tryptically digested and quantified on the MS. A comparison of MS quantitation values gave an overall average CV of 0.09; implying results and therefore experimental procedure were reliable within those tolerances. Although the isotope quantitations show an average 9% variation

between biological replicate spots, differential expressions of at least 50% (1.5 fold) were considered as significant (maintaining consistency with densitometry-based quantitation). A statistical analysis of the errors across biological replicate isotope quantitations is presented in Figure 4.10 and Table 4.4.

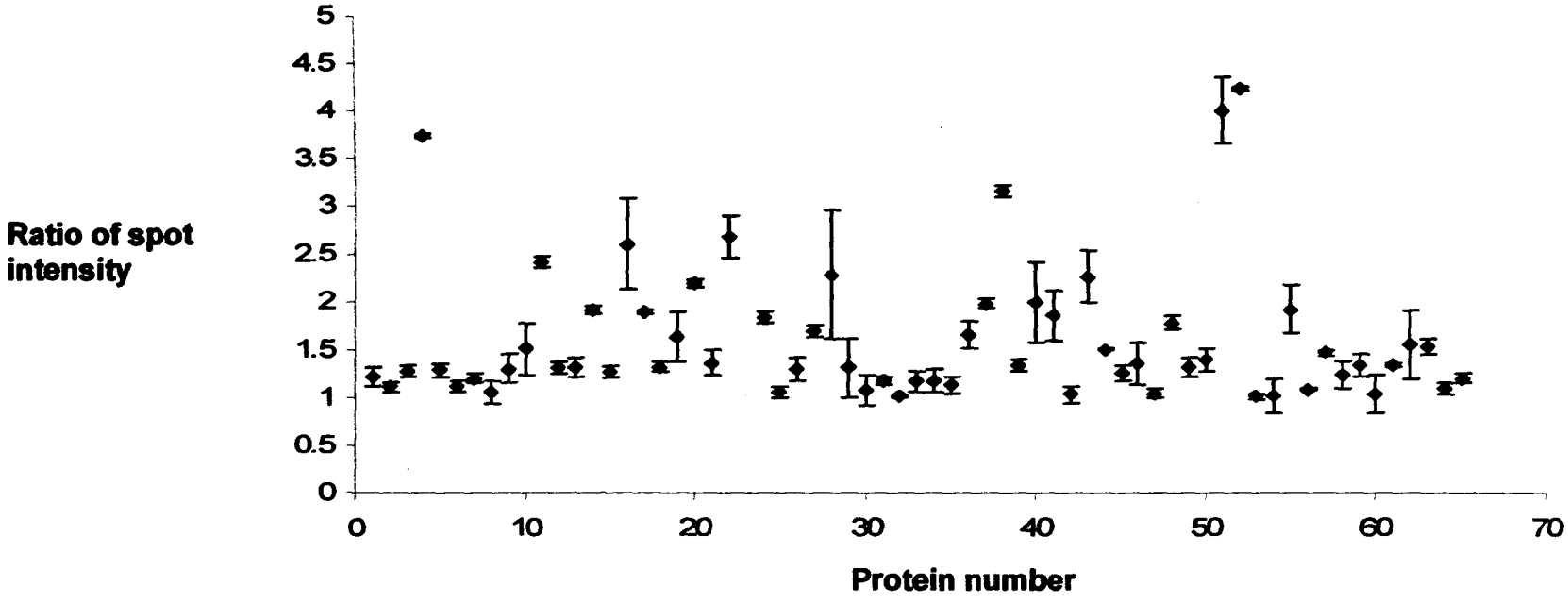


Figure 4.10: An error chart based on ratio of relative quantitations with standard deviations, calculated using isotope quantitation data. Relative ratios are all expressed higher than one.

Table 4.4: Calculations based on biological replicates (Figure 4.10) showing significance of using a variety cut-off values for differential expression in isotope data.

Cut off ratio	No. of proteins do NOT cross 1	Total no. of proteins	% proteins do NOT cross 1
3	5	5	100
2	13	13	100
1.5	26	26	100
1.4	29	29	100
1.3	40	40	100
1.2	51	51	100
1.1	55	55	100

4.3.8 Comparison of 2DE densitometry and MS data

A comparison between the initial quantitation using densitometry, based on the pre-screening exercise with the 2DE gels, and MS-data from the stable isotope labelling is shown in Table 4.5. To compare the changes in differential expression between the two methods of isotope quantitation and 2DE, the ratio calculation for the MS data was performed as shown below. The 3% salt phenotype was common to both gels, and was therefore used as a reference (value 1) for comparing quantitations between 0% and 6% gels. Quantitations for the majority of proteins were in agreement, for example, in the case of ketol-acid reductoisomerase (Figure 4.5, spot 54), average ratios of peptide peak intensity of 1.92 ± 0.25 and 0.92 ± 0.01 were found for Gel A (0% and 3% salt phenotype mixture) and Gel B (3% and 6% salt phenotype mixture) respectively. Therefore, differential expression changes between the gels (Gel B/Gel A), standardised against the reference culture, was calculated as (0.92 +1)/(1.92 +1) = 0.66. Staining intensity in PDQuest (ppm O.D. x area spot quantitation) was similar at 0.61 (Gel B/Gel A). Using 1.5 fold changes for identifying differentially expressed proteins, 6 of the 15 total quantitations compared were not in agreement, and one of these was not considered further (spot 64) because only one peptide was found in Gel A. For the remaining five proteins, three were deemed to be differentially expressed using densitometry-based quantitation only (spots 4, 19 and 51), and two were differentially expressed using isotope intensity only (spots 27 and 28). There were no examples of either methods

showing opposite quantitations (increase and decrease in expression and vice versa).

Table 4.5: A comparison of protein quantitation using the densitometry-based pre-screen image analysis and metabolic-label facilitated MS-quantitation (determining the ratio of peak areas for 3 peptides per protein).

		Image analysis			Isotope ratio		
Spot	I.D.	Gel A*	Gel B*	Densitometry ratio (Gel B/Gel A) or 6%/0%	Gel A (0%/3%)	Gel B (6%/3%)	Isotope ratio (Gel B/Gel A) or 6%/0%
4	Glutamyl-tRNA(Gln) amidotransferase subunit A	1369	3666	2.68	0.95 ± 0.12	1.31 ± 0.15	1.18
7	Superoxide dismutase [Fe]	25991	47125	1.81	0.52 ± 0.04	2.60 ± 0.47	2.37
9	LexA protein	2997	5168	1.72	0.77 ± 0.05	1.65 ± 0.26	1.50
19	Ribosome releasing (recycling) factor (protein sll0145)	1870	2845	1.52	0.94 ± 0.06	1.3 ± 0.11	1.19
20	Hypothetical protein slr1963	30175	45795	1.52	0.81 ± 0.06	1.59 ± 0.67	1.43
22	Hypothetical protein slr1900	2346	3366	1.43	0.93 ± 0.16	1.02	1.05
24	Hypothetical protein sll1213	2959	3712	1.25	1.18 ± 0.11	1.18 ± 0.12	1.00
25	Hypothetical protein sll0408	2245	1786	0.80	0.88 ± 0.09	0.60 ± 0.14	0.85
27	Hypothetical protein slr0001	7180	9970	1.39	0.32 ± 0.06	1.34 ± 0.05	1.77
28	Adenosylhomocysteinase	9523	13113	1.38	0.5 ± 0.41	1.85 ± 0.26	1.90
34	Translation elongation factor EF-Tu	5954	4573	0.77	1.26 ± 0.08	1.36 ± 0.22	1.04
36	DNA-directed RNA polymerase alpha chain	2500	1869	0.75	1.05 ± 0.05	0.56 ± 0.07	0.76
51	Phosphoglycerate dehydrogenase	3442	2157	0.63	0.99 ± 0.04	1.02 ± 0.18	1.02
54	Ketol-acid reductoisomerase	9115	5515	0.61	1.92 ± 0.25	0.92 ± 0.01	0.66
64	Transketolase	8610	4290	0.50	0.63	1.05 ± 0.20	1.26

*Staining intensity (ppm O.D. x area spot quantitation)

4.3.9 Differentially-expressed proteins in salt acclimation

22 identified proteins were significantly differentially expressed (≥1.5 fold) in changing salt concentrations, and were grouped by function [226] (see Appendix B for full details on all spots). Table 4.6 summarises these proteins by the differential expression of protein abundance in 3% compared to 0% salt and 6% compared to 3% salt. This study was designed so that an overview of protein changes in different salt concentrations can be performed, and it is important to note that the relatively small number of proteins discussed below compared to the entire proteome provide the opportunity to only postulate their significance in acclimation.

4.3.8.1 Photosynthesis and Respiration

Four differentially expressed proteins are involved in photosynthesis and respiration. A concerted difference in demand for energy is expected under stress conditions, and can be characterised by changing concentrations of proteins involved in photosynthesis and respiration [181] and this was observed in Chapter 3. The abundant photosynthetic pigment protein, allophycocyanin beta chain experienced a 1.51-fold increase in expression as the cultures reached optimum growth, from 0% to 3% salt (Table 4.6). However, in 6% salt media, the expression of allophycocyanin alpha chain reduced more than 2-fold (2.44-fold) compared to 3% salt, and for phycocyanin an even greater decrease in expression (3.70-fold) was seen. Comparable results have been seen in *Synechocystis* cells, where a reduced photosynthetic pigment content was observed in Chapter 3 at 6% salt, as well as in a previous study by Schubert *et al.* [195], where cells in ~2% salt also had optimum pigment content. Concurrently, chlorophyll *a* measurements in our cultures at the time of harvesting ($O.D._{530nm}$ = 1.3) were highest in the 3% culture, which showed the maximum growth rate. The chlorophyll *a* content in μg/ml were 5.27 ± 0.12, 8.04 ± 0.11 and 2.92 ± 0.10 in the 0%, 3% and 6% cultures respectively. These results give an immediate indication that the effect of 6% NaCl is similar in *Euhalothece* cells and *Synechocystis* cells, as revealed in Chapter 3. A 4.01-fold increase in expression of flavodoxin (spot 40b) in 6% salt cultures compared to 3% salt cultures was observed. This protein may act as an alternative electron flow carrier associated in increasing cyclic electron flow in cells [181, 210, 324]. Interestingly, in Chapter 3, the same concentration of salt caused a decrease in

flavodoxin abundance in *Synechocystis* cells. Therefore adaptation of *Euhalothece* cells to 6% salt appears to require this electron flow carrier, implying that the extent of stress may be significant in determining the role of flavodoxin. Despite the enhanced presence of this protein, the energy synthesis response in *Euhalothece* remains ambiguous because few proteins identified were significantly differentially expressed.

4.3.8.2 General stress proteins

An induction in general stress proteins production was seen in cultures grown in 6% salt compared to 3% salt (Table 4.6). Heat shock protein (P74702) (spot 11b) had a 2.67-fold enhanced expression in 6% compared to 3% NaCl. P74702 is crucial in cellular processes such as protein folding, signal transduction and protein degradation [325]. Similar to Hsp90 (Chapter 3), it plays an important role in protecting newly synthesised proteins from destabilisation and misfolding, the effects of elevated ionic concentrations. The anti-oxidative enzyme, superoxide dismutase (spot 7a/b, 8a, 17b, 33a) has one of the largest inductions with increasing salt concentrations, 1.92-fold from 0% to 3% salt and 2.60-fold from 3% to 6% salt in spots 7a and 7b. It is apparent from these results that long-term adaptation of *Euhalothece* to high salt (6%) requires increased production of general stress proteins, and that these are not only required when cells are shocked.

4.3.8.3 Energy metabolism

The glycolytic enzyme, phosphopyruvate hydratase (enolase) increased 1.56-fold in 6% salt (spot 66a/b), however, the large standard deviation (±0.36) makes it difficult to ascertain the significance of its role in salt acclimation. The PPP (where carbon dioxide is assimilated and energy for reductive biosynthetic reactions is generated) enzyme, 6-phosphogluconate dehydrogenase (Figure 4.5, spot 68b) abundance decreased 1.54-fold in 6% NaCl, but similar to enolase the standard deviation was large compared to the fold change (±0.19). In a study on the halotolerant alga *Dunaliella*, an induction of 6-phosphogluconate dehydrogenase was seen in high salt [250]. It was suggested that this enzyme may play a general role in diverting carbon metabolites under salt stress in plants and algae. This does not seem to be the case in *Euhalothece* in the conditions tested.

Phosphoribulokinase (spot 14b) is an enzyme which also forms a key step in the PPP. This enzyme was 1.85-fold reduced in expression in 6% compared to 3% salt.

4.3.8.4 Hypothetical proteins

Four proteins with no known function (currently annotated as hypothetical) were differentially expressed in higher salt, from spots 20a/b, 25b, 26b, 27a/b and 33b (Table 4.4). Protein, slr0001, was shown to be induced in high salt. Sequence similarity to proteins of known function has allowed elucidation of potential roles in organism protection. A BLASTx search revealed hypothetical protein slr0001 (Figure 4.5, spot 27a/b) harbours a flavin reductase-like domain, with 71% similarity to a protein in the cyanobacterium *Trichodesmium erythraeum* IMS101. It is possible that this protein is used to reduce flavins (water soluble pigments), which can subsequently be used in important biological processes. Its exact role in salt tolerance is not fully understood, however they are co-enzymes to flavoproteins, and when flavoprotein genes were introduced into *Arabidopsis*, improved tolerance to salt and osmotic stress was observed [326]. This protein was induced 3.17-fold in 3% salt compared to 0% salt. Its enhanced synthesis was also recently identified in the response of *Synechocystis* to 5 days in 4% NaCl [61]. A near two-fold decrease (1.96-fold) in hypothetical protein sll0553 (Figure 4.5, spot 26b) as well as a 1.52-fold reduction of hypothetical protein 0529 (Figure 4.5, spot 33b) in 6% compared to 3% salt was also observed.

4.3.8.5 Other Differentially Expressed Proteins

Regulatory protein LexA experienced a 1.65-fold increase in expression in 6% compared to 3% salt cultures (Figure 4.5, spots 9a/b). LexA protein is implicated in DNA repair mechanisms as part of the SOS response in many microorganisms, including *E. coli* [327]. It represses the SOS response, but is inactivated by a second protein RecA. This is interesting, as it seems to increase in expression in higher salt conditions in this organism. However, genetic studies have revealed that LexA does not play a role in the SOS response in cyanobacteria, and its wider function is unknown [328]. This is the first time, to our knowledge, that LexA has been reported to enhance in expression in adaptation to high salt conditions, and Chapter 3 shows that this protein decreased 1.66 and 1.61-fold in *Synechocystis* cells grown in 6% salt compared to 0% salt. RT-qPCR results in Chapter 3 also found the

expression of the *lexA* gene to be reduced (4.82 and 3.64-fold). Kanesaki *et al.* [3] found transcript expression did not change significantly in ~3% NaCl. Recent studies have revealed that LexA is a transcription activator rather than a repressor, and binds in the promoter region of the bi-directional hydrogenase enzyme in the cyanobacterium *Synechocystis* [329, 330]. The action of the bi-directional hydrogenase is of great interest, as it provides a source of renewable energy in the form of naturally produced hydrogen [331]. The presence of this gene encoding LexA in the *Euhalothece* genome was validated, as described above, by DNA sequencing. Its high sequence similarity to the same protein in *Synechocystis* raises the question of what its possible regulatory role is in *Euhalothece* but apparently not in *Synechocystis.*

A near two-fold decrease in expression (1.79-fold) of translation protein DNA-directed RNA polymerase (Figure 4.5, spot 36b) was seen in 6% compared to 3% salt, although there was no change in 3% compared to 0% salt. This enzyme is involved in ribonucleotide polymerisation into a sequence complementary to template DNA, and its decrease in expression may be the result of an overall reduction in protein synthesis in 6% salt.

Table 4.6: List of proteins which were quantified and found to be differentially expressed from 0% to 3% salt and 3% to 6% salt.

Spot no[a]	gI no.[b]	Mascot I.D.[c]	Source organism	Functional group[d]	Mw[e]	pI[f]	Mascot score (matched peptides)	Average Quantitation (CV)	N-constraint match for peptides[g]	Previously identified with salt-response[h]
Proteins: increased in expression in 3% compared to 0% salt										
7a	16330619	Superoxide dismutase [Fe]	*Synechocystis* sp. PCC 6803	Cellular processes	Y	Y	100 (4)	1.92 ± 0.04 (0.02)	Y	[61], [250], C3
28a	6094235	Adenosylhomocysteinase	*Synechocystis* sp. PCC 6803	Energy metabolism	Y	Y	417 (8)	2.00 ± 0.41 (0.21)	Y	N
64a	16329902	Transketolase	*Synechocystis* sp. PCC 6803	Energy metabolism	Y	Y	64 (1)	1.589	Y	[61]
27a	16331382	Hypothetical protein slr0001	*Synechocystis* sp. PCC 6803	Hypothetical	Y	Y	54 (2)	3.17 ± 0.06 (0.02)	Y	[61]
5a	266765	Allophycocyanin alpha chain	*Synechocystis* sp. PCC 6803	Photosynthesis & Respiration	Y	Y	385 (17)	1.51 ± 0.27 (0.18)	Y	[61]
Proteins: increased in expression in 6% compared to 3% salt										
7b	16330619	Superoxide dismutase [Fe]	*Synechocystis* sp. PCC 6803	Cellular processes	Y	Y	287 (12)	2.60 ± 0.47 (0.18)	Y	[61], C3
11b	81671547	Heat shock protein	*Synechocystis* sp. PCC 6803	Cellular processes	Y	Y	870 (22)	2.67 ± 0.13 (0.05)	Y	N
66b	3023705	Phosphopyruvate hydratase (enolase)	*Synechocystis* sp. PCC 6803	Energy metabolism	Y	Y	200 (2)	1.56 ± 0.36 (0.23)	Y	N
40b	415404	Flavodoxin	*Synechocystis* sp. PCC 6803	Photosynthesis & Respiration	Y	Y	200 (3)	4.01 ± 0.35 (0.09)	Y	[61]
9b	16330362	LexA protein	*Synechocystis* sp. PCC 6803	Regulatory functions	Y	N	305 (10)	1.65 ± 0.26 (0.16)	Y	N
9b	16331216	light-repressed protein lrtA	*Synechocystis* sp. PCC 6803	Adaptation	Y	Y	157 (3)	2.20 ± 0.05 (0.02)	Y	C3
Proteins: decreased in expression in 3% compared to 0% salt										
54a	2506910	Ketol-acid reductoisomerase	*Synechocystis* sp. PCC 6803	Amino acid biosynthesis	Y	Y	218 (3)	1.92 ± 0.60 (0.31)	Y	C3 (opposite)
56a	6685569	Dihydroxy-acid dehydratase	*Synechocystis* sp. PCC 6803	Amino acid biosynthesis	Y	Y	130 (2)	1.64	Y	N
48a	16221081	Negative aliphatic amidase regulator	*Synechocystis* sp. PCC 6803	n/a	N	Y	189 (4)	4.17 ± 0.35 (0.08)	Y	N
Proteins: decreased in expression in 6% compared to 3% salt										
68b	1703020	6-phosphogluconate dehydrogenase, decarboxylating.	*Synechocystis* sp. PCC 6803	Energy metabolism	Y	Y	231 (3)	1.54 ± 0.19 (0.12)	Y	[250] opposite, C3

25b	16331452	Hypothetical protein sll0408	*Synechocystis* sp. PCC 6803	Hypothetical	Y	Y	405 (7)	1.67 ± 0.41 (0.25)	Y	N
26b	81818888	Hypothetical protein sll0553	*Synechocystis* sp. PCC 6803	Hypothetical	N	Y	122 (2)	1.96 ± 0.19 (0.10)	Y	N
33b	16332076	Hypothetical protein Sll0529	*Synechocystis* sp. PCC 6803	Hypothetical	N	N	135 (2)	1.52 ± 0.05 (0.03)	Y	[3]
2b	2493300	Phycocyanin beta chain	*Synechocystis* sp. PCC 6803	Photosynthesis & Respiration	Y	Y	489 (13)	3.70 ± 0.28 (0.08)	Y	[61], C3
5b	266765	Allophycocyanin alpha chain	*Synechocystis* sp. PCC 6803	Photosynthesis & Respiration	Y	Y	276 (9)	2.44 ± 0.36 (0.15)	Y	[3]
14b	1653545	Phosphoribulokinase	*Synechocystis* sp. PCC 6803	Photosynthesis & Respiration	Y	Y	713 (17)	1.85 ± 0.21 (0.11)	Y	[250] opposite
36b	2500603	DNA-directed RNA polymerase alpha chain	*Synechocystis* sp. PCC 6803	Transcription	N	Y	296 (6)	1.79	Y	C3 opposite
13b	16331793	Periplasmic iron-binding protein	*Synechocystis* sp. PCC 6803	Transport and binding proteins	Y	Y	188 (3)	1.79 ± 0.03 (0.02)	Y	C3

(a) Spot number (see Figure 4.5). Spot (a) found on Gel A with isotopically labelled reference culture 3% with 0% salt. Spot (b) found on Gel B with isotopically labelled reference culture 3% with 6%
(b) NCBI accession number
(c) Protein name according to NCBI database (http://www.ncbi.nlm.nih.gov/)
(d) Protein functional group from Cyanobase (http://www.kazusa.or.jp/cyanobase/)
(e) Experimental and theoretical molecular weight match (Y-Yes, N-No) (see supplementary table for theoretical values)
(f) Experimental and theoretical pI match (Y-Yes, N-No) (see Appendix B for theoretical values)
(g) N-constraint match for peptides using method described previously in Chapter 4, section 4.3.4, and using Proteomic Toolkit (Y-yes, N-no) (http://db.systemsbiology.net:8080/proteomicsToolkit/IsotopeServlet.html)
(h) Previously identified in proteomic [61], DNA microarray salt study [3], or Chapter 3 (C3) in close relative *Synechocystis*, or similar salt response related proteins identified in the unicellular green alga, *Dunaliella* [250]. A change which has been seen in the opposite direction is also noted.

4.4 Concluding remarks for Part A

In Part A, a cross-species proteome analysis of an extremophilic cyanobacterium was conducted to gain an overview of the salt response. Metabolic labelling was used to improve protein identification, as well as producing reliable quantitation data, using 2DE for protein separation, visualisation and as a pre-screen for differential expression quantitation. Together, implementing the conventional Mascot algorithm, the sequence-similarity search algorithm, MS BLAST, and a nitrogen-constraint, 60% of selected spots were confidently identified.

Euhalothece was successfully cultured in the laboratory, growing optimally in 3% salt conditions. 2DE separation and visualisation was successful for the *Euhalothece* proteome, and optimisation tests revealed a combination of salt removal steps produced a global proteomic representation of over 500 Coomassie-stained spots per gel (Figure 4.5). 69 matched spots were deemed to be differentially expressed ($\geq$1.5-fold) by a pre-screening densitometry analysis. These were selected for identification and more accurate quantitation, and proteins belonging to nine functional groups were found to be differentially expressed in 0%, 3% and 6% salt, including cellular processes, energy metabolism, hypothetical, photosynthesis and respiration, regulatory functions, transcription and translation, and amino acid biosynthesis. PSORTb v2.0 was used to predict the cellular location of identified proteins, and cytoplasmic proteins dominate (51.7%) which was expected because they are soluble proteins and 2DE fractionation cannot sufficiently resolve hydrophobic proteins. Despite this, cytoplasmic membrane protein pyruvate hydrogenase, outer membrane protein phycobilisome linker protein, and periplasmic proteins, iron-binding protein, hypothetical protein sll0408 and negative aliphatic amidase regulator, were all identified (8.62%). The cellular location of the remaining proteins could not be predicted using the algorithm in PSORTb v2.0. Implementing LipoP 1.0 software did not identify any further membrane proteins.

A comparison between the quantitations resulting from the pre-screen (densitometry analysis) and metabolic labelling did reveal several discrepancies (Table 4.5).

Therefore, it is important to note that the pre-screen potentially reduces the quantitative power of the experiment by increasing the rate of false negative and false positive measurements.

Proteins previously associated with salt adaptation were identified, for example, synthesis of flavodoxin increased 4.01-fold in 6% salt compared to 3% salt [210], although it has been shown to be repressed in *Synechocystis* cells grown in 6% salt (Chapter 3). Proteins not previously described in salt adaptation, such as LexA were also identified, and the protein-coding gene presence was confirmed using DNA sequencing. Quantifying three peptides per protein, and comparing this across biological replicate sets, produced sufficiently accurate expression ratios. The average CV of quantitations across biological replicates was 0.09 and therefore confidently detecting a 50% change in protein expression was statistically possible. Utilising metabolic labelling for quantitation and 2DE fractionation with combined protein identification software improved overall accuracy and reliability, as well as overall running time.

It is feasible to conclude from the number of high confidence protein identifications, that a large-scale proteomic study using metabolic labelling with Mascot search engine and MS BLAST, is achievable. In addition, accuracy in quantitation using isotopes over densitometry alone, means proteins involved in long-term adaptation, and therefore with significant but lower differential expressions ($\geq$1.5 fold), can be identified.

PART B

4.5 Introduction to Part B

In Part A, a rapid proteome screening method was applied to gain an overview into the salt adaptive strategies in *Euhalothece*. In Part B, the success of the technique is exploited and developed further to attain a large-scale data set. Increasing the proteome coverage allows a more detailed analysis of exactly how protein groups and metabolic pathways are involved in acclimation to salt stress by looking for similar changes in linked functional pathways. Protein fractionation is to be accomplished by 1D SDS-PAGE which has the advantage of separating and displaying more hydrophobic proteins (membrane and membrane bound). Only five membrane proteins were identified and quantified in Part A. A disadvantage of eliminating 2DE fractionation is that identification confirmation using spot positioning relative to pI and Mw is no longer possible. However, in Part A it was successfully demonstrated that there are many orthologous *Euhalothece* tryptic peptides present in databases, with *Synechocystis* being the dominating source organism (97% total positively identified peptides and 89% of proteins). In addition, using an elemental constraint has been shown to be an effective identification confirmatory technique in cross-species proteomics. Despite its extremely slow growth at higher salinities, *Euhalothece* cells are able to adapt to salt conditions from 0 to 6% salt. In this study, an extra comparison is made between low salt (3% w/v) and very high salt (9% w/v). This very high salt concentration is approaching three times the average salinity of the oceans, and model organism *Synechocystis* is unable to survive in such extreme conditions. Implementing large format (17cm) SDS-PAGE provides excellent protein fractionation as described previously [67]. Eliminating a pre-screen (densitometry-based) step as performed in Part A, means all gel bands will be selected for protein extraction, identification and isotope based quantitation. Protein identification using Mascot and MS BLAST is to be conducted on all gel bands to maximize proteome coverage.

4.6 Materials and methods

4.6.1 Culture conditions and cell growth

Euhalothece cells were cultured as described in Part A (section 4.2.3). A very high salt, 9% (w/v) phenotype was included and biological replicate cultures were also grown. Comparative to Part A, the 3% salt phenotype was chosen as a reference culture.

4.6.2 Cell harvesting, protein extraction and quantitation

50 ml of mid-late exponential phase ($O.D._{530}$ = ~1.0) cells were harvested during the mid-light cycle as described in Part A (section 4.2.3). However, no mixing of cells was performed prior to protein extraction. Despite confirmatory tests conducted in Part A to ensure that cell size or protein extraction efficiency did not affect overall protein concentration, in this study, proteins were extracted and then mixed to eliminate the introduction of even small errors due to any cell size changes. Proteins were extracted and quantified as described via Part A, and then three different protein mixtures were created: (1) a mixture of 0% and 3% salt cultures, (2) 3% and 6% salt cultures, and (3) 3% and 9% salt cultures (50μg of each protein extract). Biological replicates of these cultures were also treated in parallel.

4.6.3 SDS-PAGE

All three protein mixtures (each consisting of 100μg protein) were added to Sample Buffer/Laemmli Buffer (Bio-Rad), to at least 2 volumes of the original sample. The composition of 1D gels (size: 17 cm long × 17 cm width × 0.75 mm thick) was 4% for the stacking gel and 12% for the resolving gel. Technical (2) and biological replicate (3) gels were run. The samples were loaded into wells and using the Protean II Multicell (BioRad) apparatus they were run at 100V for 40 minutes and 280V for 4 hours. Subsequently, gels were washed with deionised water for 5 minutes. Coomassie Brilliant Blue G250 dye (Bio-Rad) was used to stain the gels overnight on an orbital shaker at room temperature. Deionised water was used to de-stain the gels. Gel lanes were cut into small bands in preparation for in-gel digestion. Gel bands were destained twice with 200 mM ammonium bicarbonate in

40% acetonitrile. Subsequently, the gel pieces were incubated in acetonitrile for approximately 15 min at 37°C, and then the liquid phase was removed and samples were placed in a vacuum concentrator (Model 5301, Eppendorf, Cambridgeshire, UK) for another 15 min to completely dry the gel pieces. Reduction was carried out in 50 mM ammonium bicarbonate solution with 10 mM DTT, for 1 hour at 56°C. Alkylation was carried out using a 50 mM ammonium bicarbonate solution with 55 mM iodoacetamide, in the dark for 30 minutes. After alkylation, gel pieces were dehydrated using acetonitrile for 15 min at 37°C followed by conditioning with 50 mM ammonium bicarbonate for 15 min at 37°C and shrunk again using the same volume of acetonitrile for 15 min at 37°C. Gel pieces were dried in a vacuum concentrator prior to digestion of proteins. Trypsin (20 μg/ml), with acetonitrile (20%) was added to the gel pieces, which were incubated overnight at 37°C. 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.

4.6.4 Protein identification

Peptides were separated and run on the QStar XL ESI-qQ-TOF-MS/MS as described in Part A (section 4.2.5). However, due to the increased complexity of protein mixture in a gel band compared to a gel spot, the gradient on the Ultimate LC system was increased to 45 minutes. This started with 5% Buffer B (0.1% formic acid in 95% acetonitrile) and 95% Buffer A (0.1% formic acid in 5% acetonitrile) for 3 minutes, followed by a ramping to 50% Buffer B over 40 mins, then 90% Buffer B for 5 min and 5% Buffer B for a further 5 mins. IDA data was submitted for a Mascot search using Mascot Daemon v.2.1.3 with search parameters as described in Part A, section 4.2.5 (^{14}N), and subsequently against amino acid mass modifications incorporating the ^{15}N isotope. Performing a double search here enabled more efficient selection of co-eluting isotopic peptides for quantitation, as well as increased identification significance. MS BLAST settings were as described previously, and the search was performed for all gel bands. In addition to these criteria, the identified peptide must be present in both isoforms,

with two peptides present per protein, and finally confirmed using the nitrogen constraint. Protein data analysis was performed as described previously, employing PSORTb v2.0 [248] and LipoP v1.0 [249] and gene functional groups from Kazusa (http://bacteria.kazusa.or.jp/cyanobase/*Synechocystis*/index.html).

4.7 Results and discussion

4.7.1 Growth data

The effect of salt concentration on growth rates is represented in Figure 4.2. A yellow-brown colour is typical for natural populations of extremely halotolerant cyanobacteria, and is attributed to the presence of carotenoids in cells [151]. *Euhalothece* cultures showed a little yellow pigmentation in 6% salt, but mostly at 9% salt and reduced green colouration (loss of phycobiliproteins and chlorophyll), which could be explained by the increase in carotenoid production in high salt.

4.7.2 Protein identification and characterisation

In total, 301, 270 and 133 proteins were identified and quantified in data set 1 (combined 0% and 3% salt), data set 2 (combined 3% and 6% salt) and data set 3 (combined 3% and 9% salt) respectively (Appendix C). However, only proteins identified by 2 or more peptides were considered further. There was a noticeable difference in the number of proteins identified and quantified in these data sets, and in particular to data set 3. As part of the identification criteria, both isotopic peptides must be present, and it is likely that overall global protein synthesis is reduced in high salt adapted cells. It is certain that a difference in protein composition occurs in the different conditions [263].

For proteins with 2 or more identified peptides, a Mascot conventional search identified 229, 212 and 96 proteins, and MS BLAST homology search identified a further 32, 30 and 7 proteins in data sets 1, 2 and 3 respectively. In total, 383 unique proteins were identified (with 2 or more peptides) across the two gels. 24 different source organisms were used for identifying these proteins, 20 of which were prokaryotes (see Table 4.7).

Table 4.7: Source organisms for identifying all 383 unique proteins. Where multiple organisms were found to be the source for a particular protein, only the organism with the highest scoring protein was selected for the I.D.

Source Organism	Strain	I.D.'s
Cyanobacteria		
Anabaena variabilis	KCTC AG10273	1
	ATCC 29413	11
Crocosphaera watsonii	WH 8501	13
Nostoc punctiforme	PCC 73102	9
	PCC 7120	1
Prochlorococcus marinus	MIT 9312	3
	CCMP1375	2
Prochlorothrix hollandica		2
Pseudanabaena sp.	PCC 7409	1
Synechococcus sp.	elongatus PCC 6301	1
	CC9605	5
	CC9902	2
	PS677	1
	WH 8102	1
	vulcanus	1
Synechocystis sp.	PCC6714	2
	PCC6803	315
Trichodesmium erythraeum	IMS101	5
Bacteria (non-cyanobacterium)		
Deinococcus radiodurans	R1	1
Thermotoga maritima		1
Eukaryote		
Frankenia pulverulenta		1
Muscari comosum		1
Pinus thunbergii		1
Unknown		1

The close relative (on the basis of 16S rRNA comparisons, see Figure 4.4). *Synechocystis* was the source organism for the majority (82.5%) of total proteins identified, and therefore functional groups from Kazusa were used as a basis for categorization, as in Part A. Figure 4.11 illustrates the different functional groups identified. The most significant fraction were hypothetical proteins (22.8%), which is unsurprising considering 45% of protein-coding genes for *Synechocystis*, were annotated as hypothetical, upon first release [144]. The highest represented and annotated proteins were involved in photosynthesis and respiration (11.9%), energy metabolism (10.1%) and translation (9.8%), very similar to a recent 2DE study in *Synechocystis* [61].

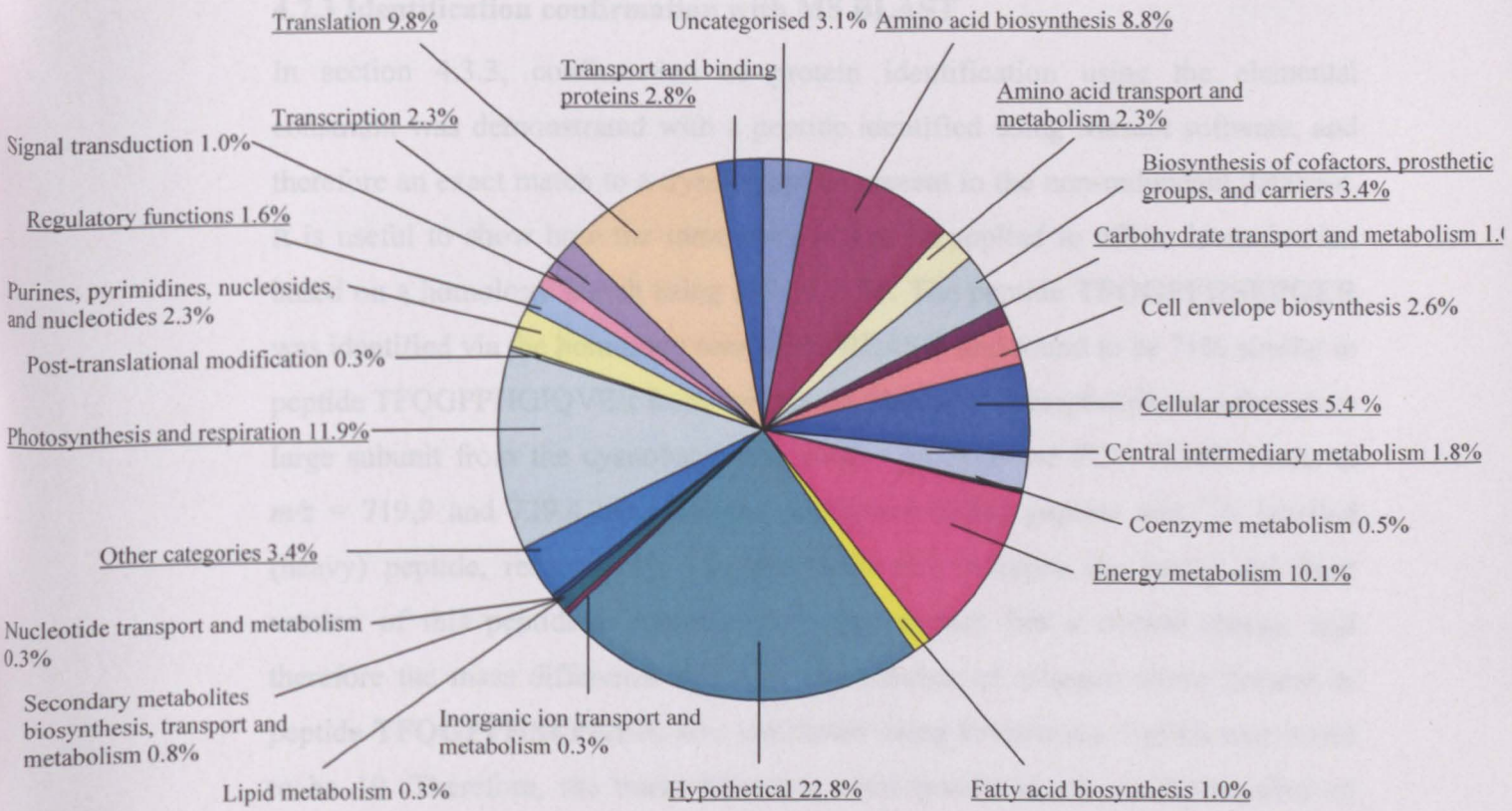


Figure 4.11: Identified proteins categorised into functional groups with percentage of total proteins identified in brackets. Four proteins were placed in more than one functional group (Appendix C) and underlined categories contain proteins which were significantly affected by salinity concentrations.

PSORTb v.2.0 was used to predict the cellular location of identified proteins, and results were remarkably similar to the cellular locations of cyanobacterial proteins predicted in three gel-free studies [83]. All five possible localisation categories were covered (cytoplasmic 48.0%, cytoplasmic membrane 4.63%, extracellular 0.82%, outer membrane 1.63%, periplasmic 2.45%) including 44.6% which were unknown. LipoP v1.0 was implemented to overcome the limitations in PSORTb and predicted 38 membrane lipoproteins. Three and twenty-one of these had been designated cytoplasmic and unknown by PSORTb, respectively (Appendix C). Therefore, 56 membrane proteins in total (14.6%) were identified and quantified in this study.

4.7.3 Identification confirmation with MS BLAST

In section 4.3.3, confirmation of protein identification using the elemental constraint was demonstrated with a peptide identified using Mascot software, and therefore an exact match to a tryptic peptide present in the non-redundant database. It is useful to show how the same method can be applied to affirm identification based on a homology search using MS BLAST. The peptide **TFQGPPHSEPGER** was identified via the homology search MS BLAST, and found to be 71% similar to peptide TFQGPPHGIQVER from the protein ribulose 1,5-bisphosphate carboxylase large subunit from the cyanobacterium *Nostoc punctiforme* PCC 73102. Peaks at *m/z* = 719.9 and 729.4 represent the unlabelled (light) peptide and ^{15}N labelled (heavy) peptide, respectively. The *m/z* difference between the heavy and light version of this peptide is therefore 9.5. The peptide has a double charge, and therefore the mass difference is 19Da. The number of nitrogen atoms present in peptide **TFQGPPHSEPGER,** also calculated using Proteomics Toolkit was found to be 19. Therefore, the mass difference corresponds exactly to the number of nitrogen atoms present in the peptide, increasing confidence in the I.D. of this peptide, and hence I.D. of the protein. As in Part A, all accepted protein identifications were confirmed using this criterion.

4.7.4 Variation and normalisation of MS data

The full set of proteins quantified on the MS would be expected to have an average fold change of 1 (no change), therefore the data was normalised by correcting the ratio towards a median change of unity, before relative quantitation changes were considered significant.

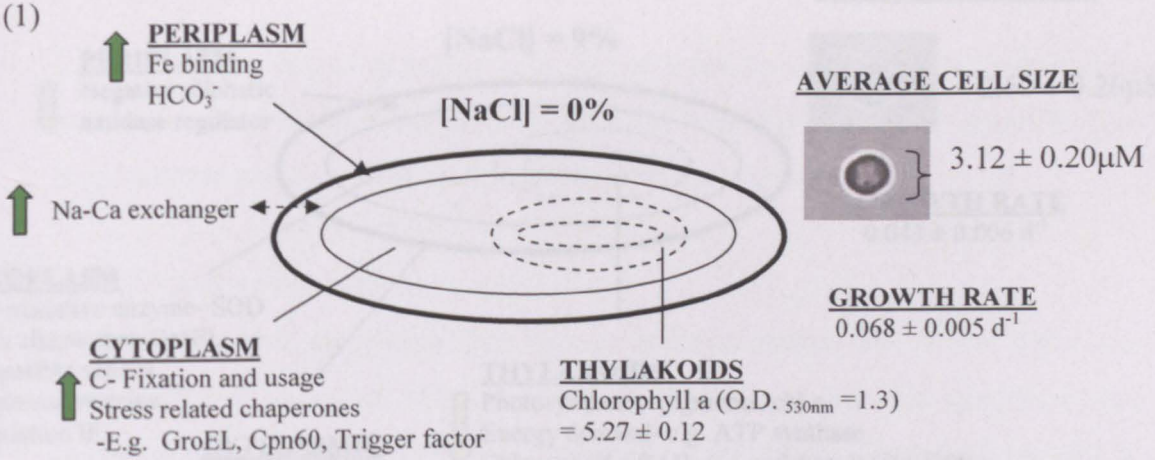
A comparison of quantitations across biological replicate gels (3) gave an overall average CV of 0.11, comparable to Part A (0.09). The standard deviation calculations for the 3 peptides quantified for each protein were also relatively low (average 0.10). The average RV was also determined logarithmically using biological replicate quantitations, as performed previously in Chapter 3. The RV was 0.92 implying an 8% discrepancy for relative quantitation. Therefore fold

expression changes of at least 50% (≥1.5 fold) were again considered significant, important for this adaptation study.

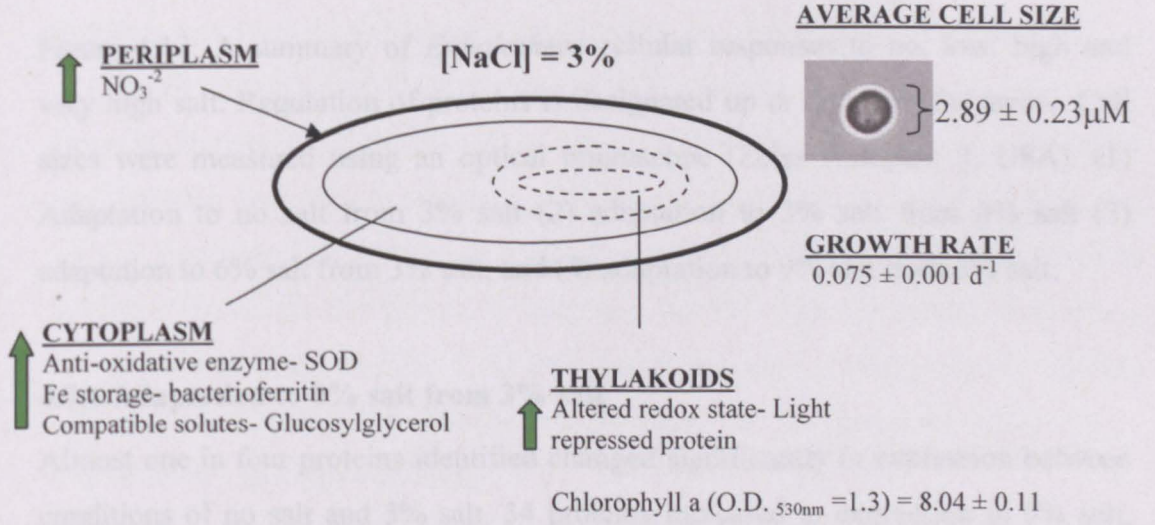
4.7.5 Differentially-expressed proteins in salt acclimation

Proteins which were identified by two or more peptides in both isotopic versions, and changed in expression by equal to or more than 1.5 fold were deemed differentially expressed (see section 4.7.4). 55 proteins were differentially expressed when cells are grown in 0% salt and 3% salt, and 18 proteins had significant changes in abundance in 6% salt compared to 3% salt. A further 35 proteins were affected by a change from 3% to 9% salt (Table 4.8). It is apparent here that *Euhalothece* cells respond much more significantly when cultured in a non-salt environment (0%) compared to a salt environment (3%). Consequently, results were interpreted and discussed in terms of adaptation of cells (i) to no salt from 3% salt, (ii) to 3% salt from no salt (iii) adaptation to 6% salt from 3% salt, and finally (iv) adaptation to 9% salt from 3% salt, and are summarised in Figure 4.12.

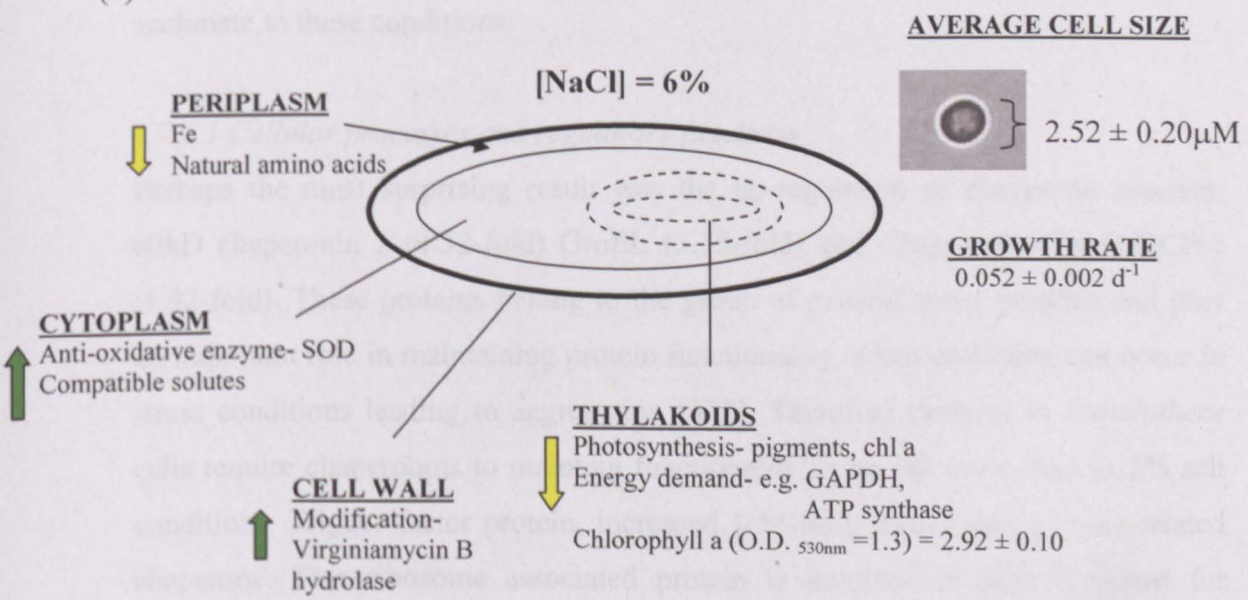
(1)


PERIPLASM
Fe binding
HCO_3^-
[NaCl] = 0%
AVERAGE CELL SIZE
3.12 ± 0.20μM
Na-Ca exchanger
GROWTH RATE
0.068 ± 0.005 d^{-1}
CYTOPLASM
C- Fixation and usage
Stress related chaperones
-E.g. GroEL, Cpn60, Trigger factor
THYLAKOIDS
Chlorophyll a (O.D. 530nm =1.3)
= 5.27 ± 0.12

(2)


AVERAGE CELL SIZE
PERIPLASM
NO_3^{-2}
[NaCl] = 3%
2.89 ± 0.23μM
GROWTH RATE
0.075 ± 0.001 d^{-1}
CYTOPLASM
Anti-oxidative enzyme- SOD
Fe storage- bacterioferritin
Compatible solutes- Glucosylglycerol
THYLAKOIDS
Altered redox state- Light
repressed protein
Chlorophyll a (O.D. 530nm =1.3) = 8.04 ± 0.11

(3)


AVERAGE CELL SIZE
[NaCl] = 6%
PERIPLASM
Fe
Natural amino acids
2.52 ± 0.20μM
GROWTH RATE
0.052 ± 0.002 d^{-1}
CYTOPLASM
Anti-oxidative enzyme- SOD
Compatible solutes
THYLAKOIDS
Photosynthesis- pigments, chl a
Energy demand- e.g. GAPDH,
ATP synthase
CELL WALL
Modification-
Virginiamycin B
hydrolase
Chlorophyll a (O.D. 530nm =1.3) = 2.92 ± 0.10

(4)

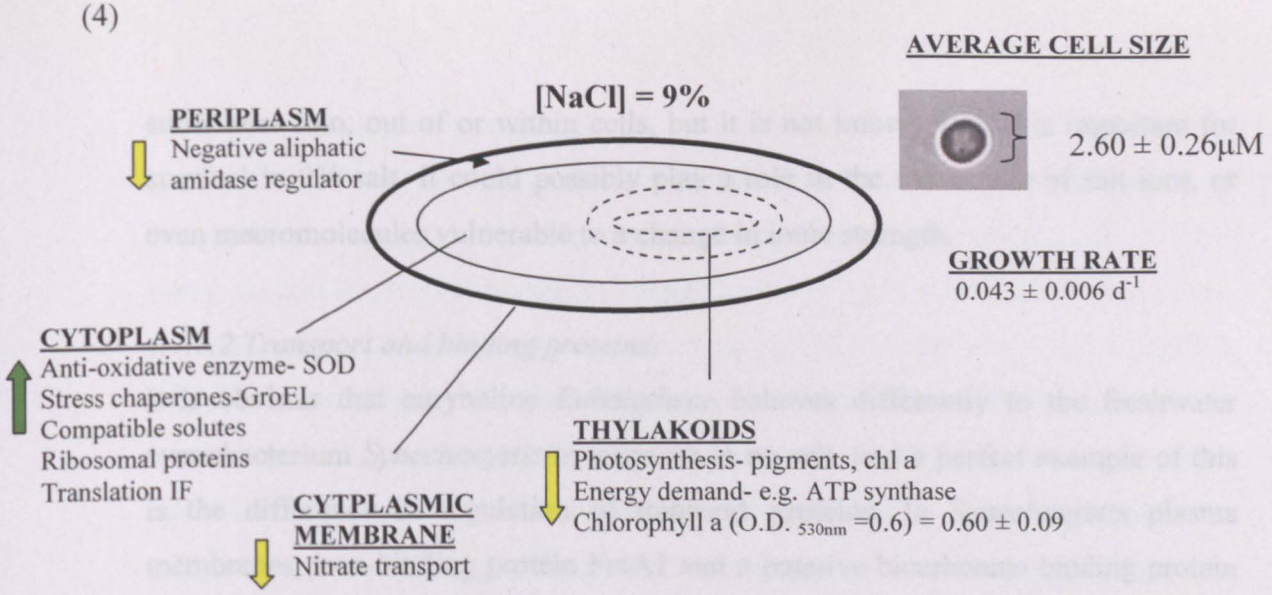


Figure 4.12: A summary of *Euhalothece* cellular responses to no, low, high and very high salt. Regulation of proteins is designated up or down by the arrow. Cell sizes were measured using an optical microscope (Zeiss Axioplan 2, USA). (1) Adaptation to no salt from 3% salt (2) adaptation to 3% salt from 0% salt (3) adaptation to 6% salt from 3% salt, and (4) adaptation to 9% salt from 3% salt.

4.7.6 Adaptation to 0% salt from 3% salt

Almost one in four proteins identified changed significantly in expression between conditions of no salt and 3% salt. 34 proteins increased in expression in 0% salt, highlighting the large number of changes required for *Euhalothece* cells to acclimate to these conditions.

4.7.6.1 Cellular processes and regulatory functions

Perhaps the most surprising result was the up-regulation of chaperone proteins, 60kD chaperonin 2 (4.32-fold) GroEL (3.15-fold) and Chaperonin Cpn60/TCP-1 (1.92-fold). These proteins belong to the group of general stress proteins and play an important role in maintaining protein functionality, when unfolding can occur in stress conditions leading to aggregation [332]. Therefore proteins in *Euhalothece* cells require chaperonins to maintain functionality in no salt more than in 3% salt conditions. Trigger factor protein, increased 1.55-fold, and is also a stress related chaperone. This ribosome associated protein is involved in protein export for maintaining newly synthesized proteins in an open conformation [333]. The twitching mobility protein increased 2.48-fold. It is involved in directed transport of

substances into, out of or within cells, but it is not known why it is important for survival in 0% salt. It could possibly play a role in the movement of salt ions, or even macromolecules vulnerable to a change in ionic strength.

4.7.6.2 Transport and binding proteins

It is obvious that euryhaline *Euhalothece* behaves differently to the freshwater cyanobacterium *Synechocystis* in response to no salt, and a perfect example of this is the difference in regulation of transport proteins. In *Synechocystis* plasma membranes, iron-binding protein FutA1 and a putative bicarbonate binding protein were observed to accumulate in salt-stressed cells [224]. In *Euhalothece* cells, an increase was seen in periplasmic iron-binding protein (1.82-fold) as well as in a sodium-dependant bicarbonate transporter (6.31-fold) in 0% salt (Table 4.8). It isn't clear at this stage what roles these proteins play in acclimation to a non-salt environment. Increases in transport proteins cation-transporting ATPase PacL and Na-Ca exchanger/integrin-beta4 were also observed; interestingly the latter is involved in calcium ion export, in exchange for Na^+ ions. Experimental evidence has previously suggested that Ca^{2+} ions function in salt adaptation in two specific ways. It could facilitate enhanced selectivity for K^+/Na^+ transporters [334], or play a role in stress signal transduction [335].

4.7.6.3 Carbon utilisation

It is possible that the observed increase in the bicarbonate transporter abundance is associated with an increased need for carbon. This postulation is supported by an increase in carbon fixation enzymes fructose-1, 6-biphosphatase (1.71-fold) and glpX (1.64-fold). Furthermore, a 2.16-fold increase in enzyme GTP cyclohydralase was evident. This enzyme performs the first step of tetrahydrofolate synthesis which is used for one-carbon (C_1) metabolism transfer reactions. C_1 metabolism has been associated with adaptation of plants to environmental stresses previously, however, its specific role remains unknown [336]. The increased production of pyruvate (TCA cycle) is demonstrated by a 1.61-fold increase in the malic enzyme, which participates by decarboxylation of malate.

4.7.6.4 Hypothetical proteins

Nearly a third of proteins induced in no salt conditions were designated as hypothetical, and hence have no assigned function. In agreement with three of these, Kanesaki *et al.* [3] observed a repression in gene transcription in slr1852, slr1855 and sll1305 in 0.5M (~3%) salt stress in *Synechocystis* by 7.7, 5.5 and 3.7-fold respectively. Despite having undesignated functions, four hypothetical proteins have conserved domains which give clues to their possible purpose. Interestingly, hypothetical protein sll0408 increased 1.67-fold and is known to exhibit peptidyl-prolyl cis-trans isomerase (PPIase) activity. PPIase accelerates protein folding by catalyzing the cis-trans isomerisation of proline imidic peptide bonds in oligopeptides [301]. This protein supports the change in expression of other chaperones in *Euhalothece* in 0% salt, it is induced in the salt response in *Synechocystis* at the protein level only [61].

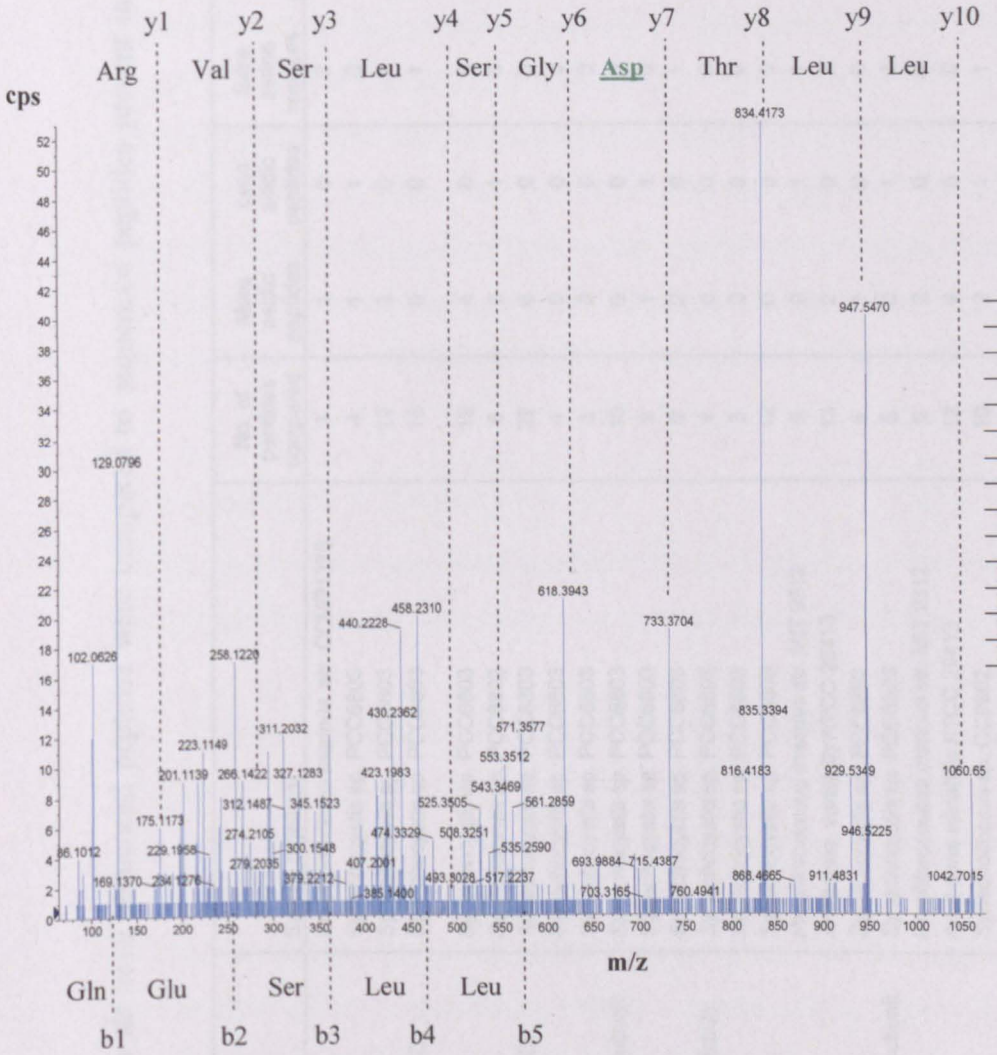
4.7.7 Possible mechanisms for adaptation to non-saline conditions

Interestingly, the requirement of molecular chaperones to maintain protein structure in low salt concentrations is a characteristic of organisms which adopt the 'salt-in' tolerance strategy rather than the low internal ionic concentration (Na^+ extrusion and osmoprotectant) method. These organisms have high intracellular ion concentrations (mostly K^+) maintaining osmotic equilibrium with its surroundings (and therefore less reliance on compatible solutes). This strategy was previously associated with halophilic and halotolerant cyanobacteria [337], but has since proven to be unfounded when it was demonstrated that essential enzymes in halotolerant organisms are dysfunctional in high ion concentrations [338]. In 'salt in' adopting organisms (predominantly haloarcheaea), proteins have additional acidic residues on their surface, thereby increasing their tolerance to salt. These residues have enhanced hydration, enabling a more sophisticated organisation of a hydrated salt ion network [339]. They reduce protein hydrophobicity, and thereby reduce aggregation. Furthermore, strategic salt bridges can form between these acidic residues and basic residues which ultimately provide structural rigidity and stabilisation of the 3D structure [340]. The disadvantage however, is when grown in low or non-salt conditions (low ionic strength), proteins can denature and unfold due to charge repulsion. It is possible that salinity levels in the Qabar-Onn Lake

have provided the driving force for fixations in alternative amino acids which have occurred due to nucleotide substitutions, and which favour stabilisation in salt environments. This is even more likely in view of the fluctuating salinity levels in this lake [149], which leads to an increased probability of fixation of substitutions and are consequently advantageous in one condition over the other. The fact that *Euhalothece* can survive in low salt may be due to the activity of its chaperones. The extent of acidity or reduced hydrophobic interactions in salt adapted proteins has previously been associated with the level of salinity tolerance, using *Salinobacter* as a model [341]. Furthermore, proteins DnaK and GroEL from three moderate halophiles have been shown to be biologically active in low salt, implying they do not require high salt to function [342]. However, an acidified protein is a salt tolerance strategy specific to organisms which have high intracellular ion concentrations, uncharacteristic of most cyanobacteria, except alkaliphiles [343]. Nevertheless, it is possible that the periplasmic space ion concentrations rise considerably higher than in the cytoplasmic space, as this is the part of the cell directly exposed to the external medium. Making direct measurements of ion concentrations in different cellular compartments remains difficult [311], but it is possible that extreme halotolerant and halophilic cyanobacteria maintain relatively high ion concentrations in the periplasmic space or in cytoplasmic membrane systems, and therefore acidic proteins remain functionally active.

An *in silico* comparison was made between *de novo* sequenced peptides in *Euhalothece* (using ProBLAST), and peptides of cyanobacteria which are already present in the database. An enhanced presence of acidic residues aspartic acid (D) and glutamic acid (E), would indicate increased acidity in proteins. In addition, more cases of the amino acid serine could be an indication of salt adapted proteins. Serine is important because of its compact size and borderline hydrophobic-hydrophilic properties [344]. An open source tool employing the EMBOSS pK values model, was also used to estimate the peptide isoelectric point value (http://isoelectric.ovh.org/). Amino acid substitutions were only considered confident if either b or y ions were present in the MS/MS fragmentation (Figure 4.13). Moreover, substitutions must have been observed during separate (technical replicate) runs, and in biological replicate samples. Of the 23 proteins compared, 11 gave the indication of extra acidic residues whereas only 2 suggested they were less

acidic (Table 4.9). Six of the hypothesised extra acidic proteins were membrane proteins, and the remaining 3 membrane proteins tested produced inconclusive results. See Appendix D for a full detailed comparison.



#	residue	b	y
1	Gln	129.0659	1404.7329
2	Glu	258.1084	1276.6743
3	Ser	345.1405	1147.6317
4	Leu	458.2245	1060.5997
5	Leu	571.3086	947.5156
6	Thr	672.3563	834.4316
7	Asp	787.3832	733.3839
8	Gly	844.4047	618.3570
9	Ser	931.4367	561.3355
10	Leu	1044.5208	474.3035
11	Ser	1131.5528	361.2194
12	Val	1230.6212	274.1874
13	Arg	1386.7223	175.1190

Figure 4.13: MS/MS fragmentation of peptide EESLLTDGSLSVR from *Euhalothece*. It is 92% similar to peptide EESLLTNGSISVR from gi|16329822, phycocyanin associated linker protein, *Synechocystis*. The neutral amino acid asparagine (Asn) in *Synechocystis* is substituted for acidic amino acid aspartic acid (Asp) in *Euhalothece*. The table shows all the b and y ion series for this peptide, with those detected in the mass spectrometer highlighted in red (NB. leucine (Leu) and isoleucine (Lle) have isobaric masses and cannot be distinguished in the low collision energy-generated MS/MS spectrum).

Table 4.9: A list of proteins for which *de novo* sequenced peptides were compared to sequenced peptides present in non-redundant NCBI database

Protein no.	Protein I.D.	Source organism	No. of peptides compared	More acidic peptides	Less acidic peptides	Extra serine residues	Less serine residues	Acidity
1	ChlL	*Prochlorococcus marinus* str. CCMP1375	7	1	0	2	0	more acidic
2	Bacterioferritin	*Synechocystis* sp. PCC6803	4	1	1	0	1	same
3	Fat protein	*Synechocystis* sp. PCC6803	12	3	0	0	1	more acidic
4	RNA polymerase beta prime subunit (rpoC2)	*Synechocystis* sp. PCC6803	15	0	0	1	0	same
5	Phycobilisome LCM core-membrane linker polypeptide	*Synechocystis* sp. PCC6803	19	4	0	0	0	more acidic
6	Transketolase	*Synechocystis* sp. PCC6803	6	2	1	0	1	more acidic
7	Phycocyanin associated linker protein (cpcC)	*Synechocystis* sp. PCC6803	22	6	0	0	0	more acidic
8	sll1305 hypothetical protein	*Synechocystis* sp. PCC6803	4	0	0	0	0	same
9	Hypothetical protein sll1762	*Synechocystis* sp. PCC6803	3	0	0	0	0	same
10	Ribulose bisphosphate carboxylase large subunit	*Synechocystis* sp. PCC6803	10	0	0	0	0	same
11	Leukotoxin (LtA)	*Synechocystis* sp. PCC6803	6	1	1	0	0	same
12	3-ketoacyl-acyl carrier protein reductase	*Synechocystis* sp. PCC6803	6	2	0	1	0	more acidic
13	3-oxoacyl-[acyl-carrier protein] reductase (fabG)	*Synechocystis* sp. PCC6803	4	0	0	0	0	same
14	Superoxide dismutase	*Synechocystis* sp. PCC6803	3	0	0	0	0	less acidic
15	Phycocyanin b subunit (cpcB)	*Synechocystis* sp. PCC6803	12	0	0	0	0	same
16	Fatty acid/phospholipid synthesis protein	*Prochlorococcus marinus* str. MIT 9312	9	0	1	1	0	less acidic
17	Allophycocyanin, beta subunit	*Anabaena variabilis* ATCC 29413	13	2	0	2	0	more acidic
18	Hypothetical protein 0359	*Synechocystis* sp. PCC6803	6	1	0	0	0	more acidic
19	Ribulose bisphosphate carboxylase small subunit	*Synechocystis* sp. PCC6803	5	0	1	0	0	less acidic
20	Magnesium-chelatase	*Prochlorococcus marinus* str. MIT 9312	5	2	0	0	0	more acidic
21	Multi-sensor Hybrid Histidine Kinase	*Anabaena variabilis* ATCC 29413	12	4	0	0	0	more acidic
22	Outer membrane autotransporter barrel	*Synechococcus* sp. CC9902	10	2	1	1	0	more acidic
23	Periplasmic iron-binding protein	*Synechocystis* sp. PCC6803	10	0	0	3	1	same

This preliminary analysis is only marginally indicative of an acidic nature to *Euhalothece* proteins. The exact positioning of acidic residues in the protein is more important than the simple presence of extra acidic residues; a more complete analysis would involve analyses of the predicted 3D structure of these proteins of interest, comparison of a larger data set and verification of amino acid substitutions via degenerate primer design and sequencing of *Euhalothece* genes. Therefore, it can only be postulated here that some proteins in this organism may rely on a certain level of salt to maintain their functional configuration, a trait common to extreme halophiles with a 'salt-in' method of salt tolerance as opposed to the ion extrusion/ osmoprotectant strategy. *Euhalothece* may provide an example of an organism which uses the ion extrusion/osmoprotectant strategy, and due to evolutionary tinkering, has replaced amino acid residues with salt-tolerant acidic (or less hydrophobic) residues.

4.7.8 Adaptation to 3% from 0% salt

Twenty one proteins increased significantly in expression in 3% salt, and almost half of these agree with the proteome salt responses observed in *Synechocystis* (Table 4.8).

4.7.8.1 Energy metabolism and central intermediary metabolism

Fructose- bisphosphate aldolase increased 1.56-fold, and is involved in several energy metabolism pathways including glycolysis. Similarly, ribulose phosphate-3-epimerase and transaldolase, involved in the PPP, increased 1.82 and 1.51-fold, respectively. The increase in carbon uptake in no salt compared to 3% salt implies that enhancing central intermediary metabolism could be a mechanism to cope with no salt, however, these results imply that energy metabolism in *Euhalothece* is increased in salt as it is in *Synechocystis* [61]. This increase in 3% salt is almost expected in *Euhalothece* cells, as growth rate is also increased. A 2.37-fold increase in ADP-glucose phosphorylase was also seen. This enzyme synthesises ADP-glucose, a precursor for the biosynthesis of the compatible solute GG. *Euhalothece* accumulates the compatible solute glycine betaine [149], which is characteristic of extremely halotolerant organisms [161]. However, it is also known that accumulation of more than one compatible solute is not uncommon, especially amongst extreme halotolerant species [161]. Finally, glycogen synthesis enzyme, glycogen phosphorylase, increased 1.62-fold [61]. Glycogen accumulation in salt acclimated cyanobacteria has

previously been demonstrated [196], and two enzymes involved were also accumulated in *Synechocystis* cells in 6% salt, as detailed in Chapter 3.

4.7.8.2 Regulatory proteins

A 1.82-fold increase in light-repressed protein was also observed in *Synechocystis* cells in Chapter 3 and a previous study [224]. It has been shown that this regulatory protein is transcribed during the dark phase only in *Synechococcus* sp. PCC7002, and therefore could be controlled by changes in redox state due to a light/dark change. Redox changes also occur in salt acclimated cells [224, 345]. LexA protein increased 1.71 fold in 3% salt, a response discussed in Part A.

4.7.8.3 Other proteins

The increase in anti-oxidative enzyme superoxide dismutase (2.43-fold) can be seen as verification of the methodology employed in this thesis. Amino acid biosynthesis proteins ferredoxin-nitrite reductase and sulfate adenylyltransferase increased 1.62 and 1.72-fold, respectively. The former enzyme increased 2.1-fold in *Synechocystis* cells after 5 days in 684 mM NaCl [61]. Bacterioferritin, responsible for storage of as much as 50% of cellular iron increased 1.51-fold, and this would be expected as iron intake is reduced in salt acclimated cyanobacteria [224, 324]. A 45kDa nitrate transport protein increased 1.84-fold. Compositional and structural changes have been proposed to take place in the plasma membrane of salt acclimated cyanobacterium cells, and this could affect efficient uptake of nutrients such as nitrate resulting in an increased necessity [243].

Table 4.8: Differentially expressed proteins between 0% and 3% salt identified by MASCOT and MS BLAST. Relative peak intensity ratios of both peptide versions (light and heavy) were used for quantitation.

Protein	gi[a]	Source organism[b]	No. of peptides[c]	Score[d]	Fold change[e]	CV[f]	Functional category[g]
ADAPTATION TO NO SALT: Proteins induced in 0% salt compared to 3% salt							
Tryptophan synthase subunit beta	16332038	*Synechocystis* sp. PCC6803	3	202	1.58	0.03	Amino acid biosynthesis
Negative aliphatic amidase regulator	16331081	*Synechocystis* sp. PCC6803	2	167	1.95	0.02	Amino acid transport and metabolism
60kD chaperonin 2	16331442	*Synechocystis* sp. PCC6803	23	1780	4.32	0.04	Cellular processes
GroEL	287463	*Synechocystis* sp. PCC6803	17	1138	3.15	0.01	Cellular processes
Twitching mobility protein	16331035	*Synechocystis* sp. PCC6803	5	272	2.48	0.02	Cellular processes
60kD chaperonin 2	16331442	*Synechocystis* sp. PCC6803	4	266	2.41	0.03	Cellular processes
Chaperonin Cpn60/TCP-1	78185313	*Synechococcus* sp. CC9902	14	531	1.92	0.02	Cellular processes
Trigger factor	16332069	*Synechocystis* sp. PCC6803	10	602	1.55	0.03	Cellular processes
Fructose-1,6- biphosphatase	1753218	*Synechocystis* sp. PCC6803	13	742	1.71	0.02	Energy metabolism
GlpX	75909902	*Anabaena variabilis* ATCC 29413	10	438	1.64	0.01	Energy metabolism
Hypothetical protein slr1762	16330280	*Synechocystis* sp. PCC6803	8	411	2.41	0.03	Hypothetical
Hypothetical protein sll0359	1653614	*Synechocystis* sp. PCC6803	4	285	2.11	0.01	Hypothetical
Hypothetical protein slr1852	16330239	*Synechocystis* sp. PCC6803	5	273	2.00	0.01	Hypothetical
Hypothetical protein slr1855	16330242	*Synechocystis* sp. PCC6803	9	457	1.86	0.02	Hypothetical
Hypothetical protein sll1305	16330234	*Synechocystis* sp. PCC6803	6	292	1.82	0.02	Hypothetical
Hypothetical protein sll1380	16330843	*Synechocystis* sp. PCC6803	2	112	1.78	0.01	Hypothetical
Hypothetical protein sll0408	16331452	*Synechocystis* sp. PCC6803	14	797	1.67	0.02	Hypothetical
Hypothetical protein sll1734	16330025	*Synechocystis* sp. PCC6803	2	145	1.61	0.03	Hypothetical
Malic enzyme	16329255	*Synechocystis* sp. PCC6803	4	316	1.61	0.03	Hypothetical
Hypothetical protein sll0314	16331369	*Synechocystis* sp. PCC6803	6	349	1.55	0.06	Hypothetical
Na-Ca exchanger/integrin-beta4	75703826	*T. erythraeum* IMS101	8	516	1.50	0.09	Hypothetical
Beta transducin-like protein	16330231	*Synechocystis* sp. PCC6803	3	164	2.28	0.01	Other categories
Serine esterase	16329810	*Synechocystis* sp. PCC6803	2	138	1.74	0.02	Other categories
Rehydrin	16329971	*Synechocystis* sp. PCC6803	6	367	1.71	0.01	Other categories
Phycocyanin associated linker protein	16329822	*Synechocystis* sp. PCC6803	8	582	2.93	0.04	Photosynthesis and respiration
Chloroplast membrane associated 30kD	16331547	*Synechocystis* sp. PCC6803	4	348	1.74	0.02	Photosynthesis and respiration
NarL subfamily	16330984	*Synechocystis* sp. PCC6803	2	71	2.00	0.03	Regulatory functions
Transcription antitermination protein	16330012	*Synechocystis* sp. PCC6803	3	118	1.74	0.01	Transcription
Ribonuclease D	16331359	*Synechocystis* sp. PCC6803	5	296	1.71	0.04	Transcription
Bicarbonate transporter	16332004	*Synechocystis* sp. PCC6803	5	302	6.31	0.00	Transport and binding proteins

Periplasmic iron-binding protein	16331793	*Synechocystis* sp. PCC6803	14	911	1.82	0.05	Transport and binding proteins
cation-transporting ATPase PacL	1006591	*Synechocystis* sp. PCC6803	3	131	1.51	0.03	Transport and binding proteins
GTP cyclohydralase	16331608	*Synechocystis* sp. PCC6803	2	101	2.16	0.01	Uncategorised
Molybopterin biosythesis protein MoeB	16331030	*Synechocystis* sp. PCC6803	8	494	2.10	0.01	Uncategorised
ADAPTATION TO 3% SALT: Proteins induced in 3% salt compared to 0% salt							
Ferredoxin-nitrite reductase	16331650	*Synechocystis* sp. PCC6803	4	212	1.62	0.02	amino acid biosynthesis
Sulfate adenylyltransferase	16330927	*Synechocystis* sp. PCC6803	2	142	1.72	0.07	amino acid biosynthesis
1-(5-phophoribosyl)-5-[5-phosphoribosylamino)methylideneaamino]	16332128	*Synechocystis* sp. PCC6803	2	108	1.50	0.12	Amino acid transport and metabolism
Coproporphyrinogen III oxidase	16329455	*Synechocystis* sp. PCC6803	5	231	1.51	0.06	Biosynthesis of cofactors, prosthetic groups, and carriers
Superoxide dismutase	16330619	*Synechocystis* sp. PCC6803	6	455	2.43	0.08	Cellular processes
ADP-glucose pyrophosphorylase	384335	*Synechocystis* sp. PCC6803	3	135	2.37	0.04	Central intermediary metabolism
Glycogen phosphorylase	16330178	*Synechocystis* sp. PCC6803	5	280	1.62	0.05	Central intermediary metabolism
Fructose- bisphosphate aldolase	16331386	*Synechocystis* sp. PCC6803	9	554	1.56	0.05	Energy metabolism
Ribulose phosphate-3-epimerase	16330729	*Synechocystis* sp. PCC6803	3	198	1.82	0.09	Energy metabolism
Transaldolase	16329404	*Synechocystis* sp. PCC6803	4	215	1.51	0.07	Energy metabolism
Fatty acid/phospholipid synthesis protein	78778523	*Prochlorococcus marinus* MIT 9312	5	362	1.55	0.04	Fatty acid, phospholipid and sterol metabolism
Hypothetical protein slr0552	16332073	*Synechocystis* sp. PCC6803	3	135	1.85	0.03	Hypothetical
Hypothetical protein slr1161	16330922	*Synechocystis* sp. PCC6803	2	123	2.27	0.10	Hypothetical
Light repressed protein	16331216	*Synechocystis* sp. PCC6803	4	237	1.82	0.01	Other categories
Phycobilisome rod-core linker polypeptide	16329710	*Synechocystis* sp. PCC6803	2	100	1.74	0.01	Photosynthesis and respiration
Phycocyanin associated linker protein	16329822	*Synechocystis* sp. PCC6803	2	181	3.88	0.13	Photosynthesis and respiration
SOS function regulatory protein	16330362	*Synechocystis* sp. PCC6803	4	296	1.71	0.05	Regulatory functions
Probable cytosol aminopeptidase	8039774	*Synechocystis* sp. PCC6803	4	290	1.57	0.05	Translation
Bacterioferritin	16329907	*Synechocystis* sp. PCC6803	4	296	1.51	0.10	Transport and binding proteins
Nitrate transport 45kDa protein	16330084	*Synechocystis* sp. PCC6803	3	206	1.84	0.03	Transport and binding proteins
Caspase catalytic subunit	75702127	*Anabaena variabilis* ATCC 29413	9	500	1.65	0.04	Uncategorised

(a) gI numbers from NCBI
(b) Source organism from conventional MASCOT search and homology-based MS BLAST search.
(c) No. of unique peptides
(d) MASCOT score
(e) Fold change based on isotope peptide intensity
(f) CV for quantitation
(g) Functional categorisation based on Kazusa (http://bacteria.kazusa.or.jp/cyanobase/*Synechocystis*/index.html)

4.7.9 Adaptation to 6% salt from 3% salt

Less than 10% of the proteins identified changed significantly in expression between conditions of 3% salt and 6% salt, implying cells are able to adapt to double the salt concentration with fewer significant cellular changes (Table 4.10). Despite only ten proteins decreasing in expression when cells were shifted to 6% salt, all proteins are similar to proteins which decrease in expression when cells are grown in 3% salt compared to no salt (i.e. an increase in salt concentration in both).

4.7.9.1 Transport and binding proteins

Iron transport protein and periplasmic iron binding protein decreased 2.92 and 1.81-fold from 3% to 6% salt, respectively. These changes again highlight the relationship between salt acclimation and iron acquisition. It appears that when adapted to salt, *Euhalothece* takes up less iron (in 3% and 6% compared to 0% salt) and stores more in the form of bacterioferritin. An ABC transporter for natural amino acids also decreased 1.78-fold, indicating a reduced requirement for these biomolecules. However, an increase (1.86-fold) in previously designated hypothetical protein sll0224 was observed, and it is now known to harbour a putative amino-acid ABC transporter binding protein.

4.7.9.2 Bioenergetic processes

Five proteins involved in energy requirement, mainly photosynthesis, decreased in 6% salt, in conjunction with a decreased growth rate. Three of these proteins are pigment related. Concurrently, chlorophyll *a* measurements in our cultures at the time of harvesting (Part A, section *4.3.8.1*) were highest in the 3% culture, which showed the maximum growth rate. A reduction in glyceraldehyde-3-phosphate dehydrogenase (NADP+) (1.61-fold) and ATP synthase subunit A (1.5-fold) imply a reduction in carbohydrate degradation (glycolysis) and reduced ATP (energy) requirement, respectively.

4.7.9.3 Cellular processes and regulatory functions

The enzyme GG-phosphate synthase is involved in biosynthesis of GG, and increased 1.82-fold in 6% salt. An expected increase in superoxide dismutase was seen again (1.81-fold), as well as the unexpected increase in LexA protein (1.54-fold).

4.7.9.4 Translation

Elongation factor Tu protein reduced 1.73-fold in 6% salt. It is hard to postulate the function of this protein in association with salt acclimation because it was identified in three additional gel bands with varying expression changes (down 1.35-fold, down 1.16-fold and up 1.4-fold). This protein has a nominal mass of 43.8kDa, and was correctly found in this Mw region in two consecutive bands on the 1DE gel. However, it was also found around 35kDa and 27kDa. This is interesting because truncated versions of elongation factor Tu and GroEL1 have previously been associated with salt stress in *Synechocystis* [61], even though there is a possibility that they are simple breakdown products of the original protein. Glutamyl-tRNA synthetase is involved in attachment of the amino acid to its tRNA molecule, and it increased 1.56-fold. An increase was seen in this protein in *Synechocystis* cells in Chapter 3 and after 5 days in ~4% NaCl, although a gene induction was not reported in a microarray study [60, 61].

4.7.9.5 Hypothetical proteins

Two hypothetical proteins increased in abundance in 6% salt conditions. Virginiamycin B hydrolase increased 1.54-fold, and is a salt response associated with *Synechocystis* [233]. This protein may be involved in the modification of the cell wall or extracellular layers in response to external high salt. Hypothetical protein Sll1130 (2.04-fold) has no known functional domains.

Table 4.10: Differentially expressed proteins between 6% and 3% salt identified by MASCOT and MS BLAST. Relative peak intensity ratios of both peptide versions (light and heavy) were used for quantitation.

Protein	gi[a]	Source organism[b]	No. of peptides[c]	Score[d]	Fold change[e]	CV[f]	Functional category[g]
ADAPTATION TO 6% SALT: Proteins induced in 3% salt compared to 6% salt							
Phosphoglycerate dehydrogenase	16330470	*Synechocystis* sp. PCC6803	3	236	1.54	0.07	Amino acid biosynthesis
Periplasmic binding protein of ABC transporter for natural amino acids	16331678	*Synechocystis* sp. PCC6803	3	129	1.78	0.08	Amino acid transport and metabolism
Phycocyanin associated linker protein	16329821	*Synechocystis* sp. PCC6803	9	574	2.65	0.19	Photosynthesis and respiration
Phycobilisome LCM core-membrane linker polypeptide	16331244	*Synechocystis* sp. PCC6803	14	782	1.81	0.19	Photosynthesis and respiration
Phycobilisome rod-core linker polypeptide CpcG	16329710	*Synechocystis* sp. PCC6803	8	485	1.70	0.07	Photosynthesis and respiration
Glyceraldehyde-3-phosphate dehydrogenase (NADP+)	971587	*Synechocystis* sp. PCC6803	7	376	1.61	0.14	Photosynthesis and respiration
ATP synthase subunit A	16329327	*Synechocystis* sp. PCC6803	4	213	1.50	0.07	Photosynthesis and respiration
Elongation factor Tu	16330913	*Synechocystis* sp. PCC6803	5	329	1.73	0.21	Translation
Iron transport protein	16329434	*Synechocystis* sp. PCC6803	2	83	2.92	0.08	Transport and binding proteins
Periplasmic iron-binding protein	16331793	*Synechocystis* sp. PCC6803	8	344	1.81	0.05	Transport and binding proteins
ADAPTATION TO 6% SALT: Proteins induced in 6% salt compared to 3% salt							
Aspartate aminotransferase	16331463	*Synechocystis* sp. PCC6803	3	202	1.55	0.07	Amino acid biosynthesis
Superoxide dismutase	16330619	*Synechocystis* sp. PCC6803	6	287	1.81	0.05	Cellular processes
Virginiamycin B hydrolase	16331258	*Synechocystis* sp. PCC6803	3	144	1.54	0.31	Hypothetical
Hypothetical protein Sll1130	16330120	*Synechocystis* sp. PCC6803	7	437	2.04	0.05	Hypothetical
glucosylglycerol-phosphate synthase	16330944	*Synechocystis* sp. PCC6803	3	172	1.82	0.00	Other categories
SOS function regulatory protein	16330362	*Synechocystis* sp. PCC6803	3	107	1.54	0.03	Regulatory functions
Glutamyl-tRNA synthetase	16331631	*Synechocystis* sp. PCC6803	2	115	1.56	0.10	Translation
Hypothetical protein sll0224	16329306	*Synechocystis* sp. PCC6803	3	185	1.86	0.04	Transport and binding

(a) gI numbers from NCBI
(b) Source organism from conventional MASCOT search and homology-based MS BLAST search.
(c) No. of unique peptides
(d) MASCOT score
(e) Fold change based on isotope peptide intensity
(f) CV for quantitation
(g) Functional categorisation based on Kazusa (http://bacteria.kazusa.or.jp/cyanobase/*Synechocystis*/index.html).

4.7.10 Adaptation to 9% salt from 3% salt

Despite the combination of 3% and 9% salt phenotypes here, it is useful to interpret the changes in relation to section 4.7.9 (6% compared to 3% salt phenotype). A larger proportion of proteins changed in abundance when comparing 3% salt to 9% salt (35% of total proteins) than 3% to 6% salt (10% of total proteins) (Table 4.11).

4.7.10.1 Photosynthesis and respiration

Six of the proteins which reduced in expression in 9% compared to 3% salt phenotype were pigment related. This overall response was also seen in 6% salt, and the change in colouration of cultures was apparent at time of harvesting, where the chlorophyll *a* content for 9% salt phenotype was only 0.6 ± 0.09 μg ml^{-1} at $O.D._{530nm}$ = 0.6. Three proteins decreased in expression, although they showed no observable change in 6% salt, for example allophycocyanin alpha subunit (1.96-fold) and phycocyanin beta subunit (2.57-fold) implying adaptation to further increases in salt requires more drastic reduction in these pigment-associate proteins. In agreement, the ATP synthase subunit B reduced 1.77-fold in 9% compared to 3% salt, whereas the ATP synthase subunit A relative expression is 1.50-fold down in 6% compared to 3% salt. A 1.77-fold increase in the CO_2 concentrating mechanism protein can be explained by an increase in demand for carbon, most likely for compatible solute production.

4.7.10.2 Cellular processes

Six stress related chaperones increased in expression (Table 4.11). Only superoxide dismutase was identified in data set 2 (i.e. 3 and 6% salt), and therefore it is not possible to conclude whether these stress proteins are required more specifically in very high (9%) salt.

4.7.10.3 Other proteins

An increase in glycogen phosphorylase (1.91-fold) and GG-phosphate synthase (1.64-fold) was expected as these are compatible solute related proteins, as mentioned previously. Interestingly, three hypothetical proteins increased significantly in expression which did not do so in 6% salt. Two (slr1894 and slr1963) have previously been associated with salt stress response in *Synechocystis* at the transcript and protein level [3, 61], and one (sll1130) also increased in 6% salt. Slr1963 is widely accepted

to be a water-soluble carotenoid protein. A BLASTx search revealed significant similarity (82%) to a water soluble orange carotenoid protein from the cyanobacterium *Arthrospira maxima*. The increase in carotenoid proteins has previously been implicated in response to salt stress in cyanobacterial cells [195, 196, 217] and eukaryotic green algae [197], and has also been associated with state transition in PSI and PSII within the thylakoid membrane [61]. Similarly, increases in 30S ribosomal protein S1 (1.87-fold) and a translation initiation factor (1.90-fold) were not seen at 6%, although they have been associated with salt acclimation previously [3, 60, 61]. This indicates that several salt response strategies observed in less tolerant strains, for example in *Synechocystis*, are conserved across cyanobacteria, but may play a role in higher salt concentrations in more halotolerant genera. Finally, an increase in LexA protein (1.90-fold) and elongation factor Tu (1.79-fold) were similar to the responses seen in 6% compared to 3% salt.

Table 4.11: Differentially expressed proteins between 9% and 3% salt identified by MASCOT and MS BLAST. Relative peak intensity ratios of both peptide versions (light and heavy) were used for quantitation.

Protein	gi[a]	Source organism[b]	No. of peptides[c]	Score[d]	Fold change[e]	CV[f]	Functional category[g]
ADAPTATION TO 9% SALT: Proteins induced in 3% salt compared to 9% salt							
Sulfate adenyltransferase	16330927	*Synechocystis* sp. PCC6803	3	203	1.77	0.01	Amino acid transport and metabolism
Negative aliphatic amidase regulator	16331081	*Synechocystis* sp. PCC6803	3	166	1.77	0.01	Amino acid transport and metabolism
slr0364 hypothetical protein	1001461	*Synechocystis* sp. PCC6803	5	1099	1.96	0.00	Hypothetical
Hypothetical protein slr 1841	16330041	*Synechocystis* sp. PCC6803	3	194	1.80	0.01	Hypothetical
Phycocyanin associated linker protein	16329822	*Synechocystis* sp. PCC6803	2	120	7.20	0.06	Photosynthesis and respiration
Phycobilisome rod-core linker polypeptide CpcG	16329710	*Synechocystis* sp. PCC6803	4	176	3.72	0.02	Photosynthesis and respiration
Phycocyanin a subunit	16329823	*Synechocystis* sp. PCC6803	7	546	2.70	0.00	Photosynthesis and respiration
Phycocyanin associated linker protein	16329821	*Synechocystis* sp. PCC6803	7	395	2.57	0.01	Photosynthesis and respiration
Phycocyanin beta subunit	1008532	*Synechocystis* sp. PCC6803	9	888	2.57	0.00	Photosynthesis and respiration
Allophycocyanin alpha subunit	154450	*Synechocystis* sp. PCC6714	9	547	1.96	0.00	Photosynthesis and respiration
ATP synthase subunit B	16330679	*Synechocystis* sp. PCC6803	9	587	1.77	0.00	Photosynthesis and respiration
Nitrate transport 45kDa protein	16330084	*Synechocystis* sp. PCC6803	3	173	1.77	0.01	Transport and binding protein
General secretion pathway protein G	16330343	*Synechocystis* sp. PCC6803	2	118	1.74	0.01	Unknown
ADAPTATION TO 9% SALT: Proteins induced in 9% salt compared to 3% salt							
Phosphoglycerate dehydrogenase	16330470	*Synechocystis* sp. PCC6803	2	95	1.75	0.15	Amino acid biosynthesis
Aspartate aminotransferase	16331463	*Synechocystis* sp. PCC6803	4	206	1.52	0.01	Amino acid biosynthesis
Peroxiredoxin	23126633	*Nostoc punctiforme* PCC 73102	4	176	1.54	0.03	Cellular processes
60kDA chaperonin 2	16331442	*Synechocystis* sp. PCC6803	22	1461	2.07	0.24	Cellular processes
Chaperonin GroEL	16330003	*Synechocystis* sp. PCC6803	19	1219	3.29	0.26	Cellular processes
Superoxide dismutase	16330619	*Synechocystis* sp. PCC6803	5	348	1.75	0.07	Cellular processes
60kD chaperonin 2	16331442	*Synechocystis* sp. PCC6803	6	304	1.54	0.11	Cellular processes
Heat shock protein 90	16332281	*Synechocystis* sp. PCC6803	4	235	1.69	0.08	Cellular processes
Glycogen phosphorylase	16330178	*Synechocystis* sp. PCC6803	13	773	1.91	0.12	Central intermediary metabolism
carbamoyl-phosphate synthase large subunit	1001668	*Synechocystis* sp. PCC6803	8	370	2.11	0.10	Energy metabolism
Hypothetical protein slr1894	16329942	*Synechocystis* sp. PCC6803	5	311	1.52	0.05	Hypothetical
Hypothetical protein slr 1963	16330780	*Synechocystis* sp. PCC6803	8	518	1.53	0.12	Hypothetical
Hypothetical protein slr0006	16331389	*Synechocystis* sp. PCC6803	4	260	2.78	0.06	Hypothetical
Hypothetical protein Sll1130	16330120	*Synechocystis* sp. PCC6803	5	299	1.91	0.08	Hypothetical

glucosylglycerol-phosphate synthase	16330944	*Synechocystis* sp. PCC6803	5	279	1.64	0.04	Other categories
CO2 concentrating mech Ccmk	16329368	*Synechocystis* sp. PCC6803	5	327	1.77	0.07	Photosynthesis and respiration
SOS function regulatory protein	16330362	*Synechocystis* sp. PCC6803	3	137	1.90	0.17	Regulatory
Amino acid adenylation	67921836	*Crocosphaera watsonii* WH 8501	9	336	2.04	0.04	Secondary metabolites biosynthesis, transport, and catabolism
Elongation factor Tu	16330913	*Synechocystis* sp. PCC6803	9	575	1.79	0.11	Translation
30s ribosomal protein S1	16330162	*Synechocystis* sp. PCC6803	2	92	1.87	0.23	Translation
Translation initiation factor IF-2B subunit a	16331195	*Synechocystis* sp. PCC6803	2	127	1.90	0.24	Translation
Nuclear transport factor 2	75703772	*Anabaena variabilis* ATCC 29413	15	611	1.75	0.05	unknown

(a) gI numbers from NCBI
(b) Source organism from conventional MASCOT search and homology-based MS BLAST search.
(c) No. of unique peptides
(d) MASCOT score or MS BLAST (bold text) score
(e) Fold change based on isotope peptide intensity
(f) CV for quantitation
(g) Functional categorisation based on Kazusa (http://bacteria.kazusa.or.jp/cyanobase/*Synechocystis*/index.html).

4.8 Conclusions to Part B

In Part B, a large-scale cross-species proteome analysis of an extremophilic cyanobacterium was conducted in response to no salt and very high salt. Large format 1D SDS-PAGE was implemented for protein separation and visualisation as opposed to 2DE in Part A. Together, using the conventional Mascot algorithm, the sequence-similarity search algorithm, MS BLAST, and a nitrogen-constraint, 383 unique proteins were confidently identified. *Synechocystis* was the major source organism (82.5% of the total unique proteins) even though 22 others organisms were used. The average CV of quantitations across biological replicates was 0.10, and protein abundance change (50%) relevant to long-term adaptation was deemed acceptable.

The response to 0% salt is interesting and seems to suggest that even though *Euhalothece* cells can grow in non-saline conditions, they require the production of molecular chaperones. Cells appear to have similar mechanisms to cope with high salt concentrations (6% and 9%) as the freshwater cyanobacterium *Synechocystis* responds to slightly lower concentrations (2-4%), including increases in stress chaperones and antioxidative enzyme superoxide dismutase, alterations to cell wall, as well as a decreased pigment production. Additionally, similar proteome responses were observed when changing from a non-saline to a 3% saline environment. A possible enhancement of energy metabolism and central intermediary metabolism included an increase in synthesis of compatible solute GG. Major differences were seen when *Euhalothece* was cultured in 0% compared to 3% salt. A non-salt medium induced several molecular chaperones to be produced, including a hypothetical protein with PPIase activity. There is also a probable increase in demand for carbon and iron in 0% salt, supported by enhanced abundance of carbon uptake and fixation proteins and periplasmic iron binding protein. Proteins not previously described in salt adaptation, such as LexA (also seen in Part A) were also identified. The exact function of this regulatory protein, in acclimatising *Euhalothece* cells to salt, remains to be uncovered.

Finally, Part B demonstrates that utilising 1D-SDS PAGE with metabolic labelling, and a combination of protein identification software, improves accuracy in

quantitation and identification of proteins in this unsequenced organism, and provides valuable insights into its adaptation to a hostile environment.

4.9 Comparisons of data sets in Part A and B, and concluding remarks

In Part A, an overview of the salt response in *Euhalothece* was achieved. The results indicate that salt responsive strategies are shared amongst cyanobacteria. It was also concluded that implementing a pre-screen step for identifying proteins of interest reduced overall experiment time, and made it possible to make an overall assessment of the ability to investigate the *Euhalothece* proteome using the techniques employed. Additionally, there were 22 proteins exclusively identified and quantified in Part A (not found in Part B), and this can be attributed to the superior fractionation compared to SDS-PAGE alone, enabling analysis of proteins in relatively lower abundance. Despite the added advantage of confirming a proteins' identity using pI and Mw, using densitometry-based semi-quantitation also led to potentially higher rates of false positive and false negative candidates as proteins of interest (Table 4.5). In addition, using 2DE limited the study to mainly soluble proteins with a pI range of 4 to 7.

Part B was undertaken with the intention of gaining a more complete data set on the *Euhalothece* proteome and was achieved with a larger range of salt concentrations (up to 9% w/v salt). Prefractionation of proteins using SDS-PAGE meant limitations associated with 2DE discussed above, were reduced. Metabolic labelling of proteins was chosen as the quantitation method due to the advantages extrapolated in Part A. This analysis led to the discovery that *Euhalothece* cells elicit a 'stress' response in no salt conditions, a reaction not perceived with the substantially smaller data set in Part A.

39 proteins spots matched across the 2DE gels (Part A) were also quantified in Part B, and represent 25 unique protein species. 29 quantitations were in concordance (up, down or non-regulated). Interestingly, in 9 of the 10 cases where quantitations were not in agreement, a significant quantitation change ($\geq$1.5 fold) was seen in Part A, whereas no change was seen in Part B. This could be due to several factors which

constitute intrinsic differences in the methodologies employed. Differences in sample preparation included mixing of cells with approximation of cell numbers using optical density measurements (Part A), compared to mixing extracted protein (Part B). Technical differences included an extra separation dimension in Part A. Data handling differences are also apparent, for example, normalisation of data in Part A occurs at the densitometry analysis stage (pre-screen) and is based on the pixel intensity of each spot identified being normalised against the total pixel density arising from all the protein spots on the gel (equation 4.1). Contrastingly, in Part B, normalisation was achieved by correcting the ratio towards a median change of unity. There is also the argument for biological variance, but it should be noted that in both studies, biological replicate analysis was incorporated. Finally, in no cases did a quantitation in Part A have a complete opposite result in Part B (up regulated in one and down regulated in the other, and vice versa).

It can therefore be concluded that the study in Part A was useful to assess the ability to subsequently generate a larger data set in this unsequenced environmental isolate, as well as giving an insight into it salt adaptive behaviour. Moreover, despite the extra workload associated with Part B, results from Part A were enhanced and led to discovery of new findings.

Chapter 5: A comparative proteomics study of salt tolerance between a non-sequenced extremely halotolerant cyanobacterium and its freshwater relative using *in vivo* metabolic and *in vitro* isobaric labelling[3]

[3] This work has been submitted for publication in *Journal of Proteome Research*

5.1 Introduction

In the last two chapters, large-scale proteomics studies have been conducted to elucidate the salt response in a moderately halotolerant (*Synechocystis*) and an extremely halotolerant (*Euhalothece*) cyanobacterium. Both studies employed different protein fractionation, identification and quantitative techniques, but revealed that many of the strategies to adapt to high salt are shared between the microbes. Importantly, the proteome sequence similarities between the organisms meant that peptide sequences in *Synechocystis* from the public database provided identification evidence for 82.5% of the proteins confidently identified in *Euhalothece* (Chapter 4). Moreover, the similarities between *Synechocystis* and *Euhalothece* are demonstrated in several characteristics including 16S rRNA sequence (80%), morphology and size (see Table 5.1). The intrinsic similarities therefore offer two distinct advantages that aid a rounded proteomic comparison of salt adaptation strategies. Firstly, a direct comparison between cyanobacteria, which have varying levels of tolerances to salt can be readily weighted. Secondly, with the reliance on putative peptide sequence orthology to provide cross-species protein identification and quantitation, a direct relative comparison of this nature has the potential to reveal distinctive proteomic patterns in *Synechocystis* cells compared to *Euhalothece* cells, over a range of salinity concentrations.

Table 5.1: A general comparison between *Synechocystis* and *Euhalothece* cells [135, 149, 161].

	Euhalothece	*Synechocystis*
Cell shape	Coccoid/bacilloid (some sheathed)	Coccoid (non-sheathed)
Cell type	unicellular	unicellular
Cell size Ø	2.25-3.4μM	2-4μM
Pigments	Chl a/b	Chl a/b
Tolerance group	High (max ~15%)	Moderate (max ~7%)
Compatible solute	Glycinebetaine, glutamatebetaine glucosylglycerol	Glucosylglycerol

Taji *et al.* (2004) carried out a comparative cDNA microarray–based study on salt tolerance in the higher plants, *Arabidopsis* and Salt Cress (*Thellungiella halophile*) [12]. Salt Cress can tolerate high salt concentrations, and is therefore referred to as a halophyte, whereas *Arabidopsis*, like most plants, is a glycophyte. The cDNA microarray approach was possible because the similarity of the nucleotide sequences between the two organisms is >90% [12], and they share several other similar characteristics. The basis for the comparison is the theory that nearly all salt tolerance genes are present in both highly salt sensitive and highly salt tolerant plants, suggesting differences in gene regulation are significant [12, 346]. Interestingly, the results indicate that salt stress tolerance of Salt Cress may be due to constitutive overexpression of many genes that function in stress tolerance under normal growth conditions, and these genes are stress inducible in *Arabidopsis* [12]. Consequently, cross-species studies do not solely seek to reveal novel salt responsive mechanisms, but rather highlight the *level* of control associated with adaptation strategies, and this initiative is applied here.

A proof of concept has been demonstrated by Snijders *et al.* [347], that shared tryptic peptides from proteins which are encoded in different species are suitable for relative quantitation across the species boundary. By metabolically labelling one proteome (^{15}N) and mixing it with the non-labelled proteome (^{14}N), the relative abundance of proteins can be calculated by MS stage for those which share tryptic peptides. However, for this to be successful with an unsequenced organism, a significant number of shared peptides are necessary. In Chapter 4, the Mascot algorithm was used for protein identification, which requires an ortholog match within the search database. In *Euhalothece*, there were sufficient orthologous peptides within the nrNCBI database to result in sufficient meaningful identifications. In total, 95% of identifications came from peptides originating from *Synechocystis* (Chapter 4). This demonstrates that the extent of homology between these genera of cyanobacteria is suitable for a relative quantitation study of proteins across the species boundary.

This cross-species approach is applied here to compare the relative abundance of proteins through sequence identity of shared peptides in conditions of low (0%

w/v), medium (3% w/v) and high (6% w/v) salt, in *Synechocystis* and *Euhalothece*. By assessing protein abundance differences between the organisms, implications for adaptation mechanisms to high salt conditions are discussed. Proteins will be fractionated by SDS-PAGE as performed in Chapter 4, Part B, due to the benefits of such an approach, including superior proteome coverage. Furthermore, this separation method did not to lead to differential elution of heavy and light forms of the same protein [74, 348], highlighting its robustness. In conjunction with metabolic labelling, an iTRAQ experiment is designed to complement proteomic findings on *Euhalothece* and *Synechocystis* cells cultured in 6% salt. This presents the first application, at least to the author's knowledge, of relative protein quantitation, across the species boundary, using iTRAQ-based quantitation. Therefore, both quantitative techniques applied in Chapters 3 and 4 are used in parallel here, and the quantitation and identification outcomes are gauged for both confirmatory and complementary purposes.

5.2 Methods and materials

An overview of the experiment programme conducted in this study is provided in Figure 5.1.

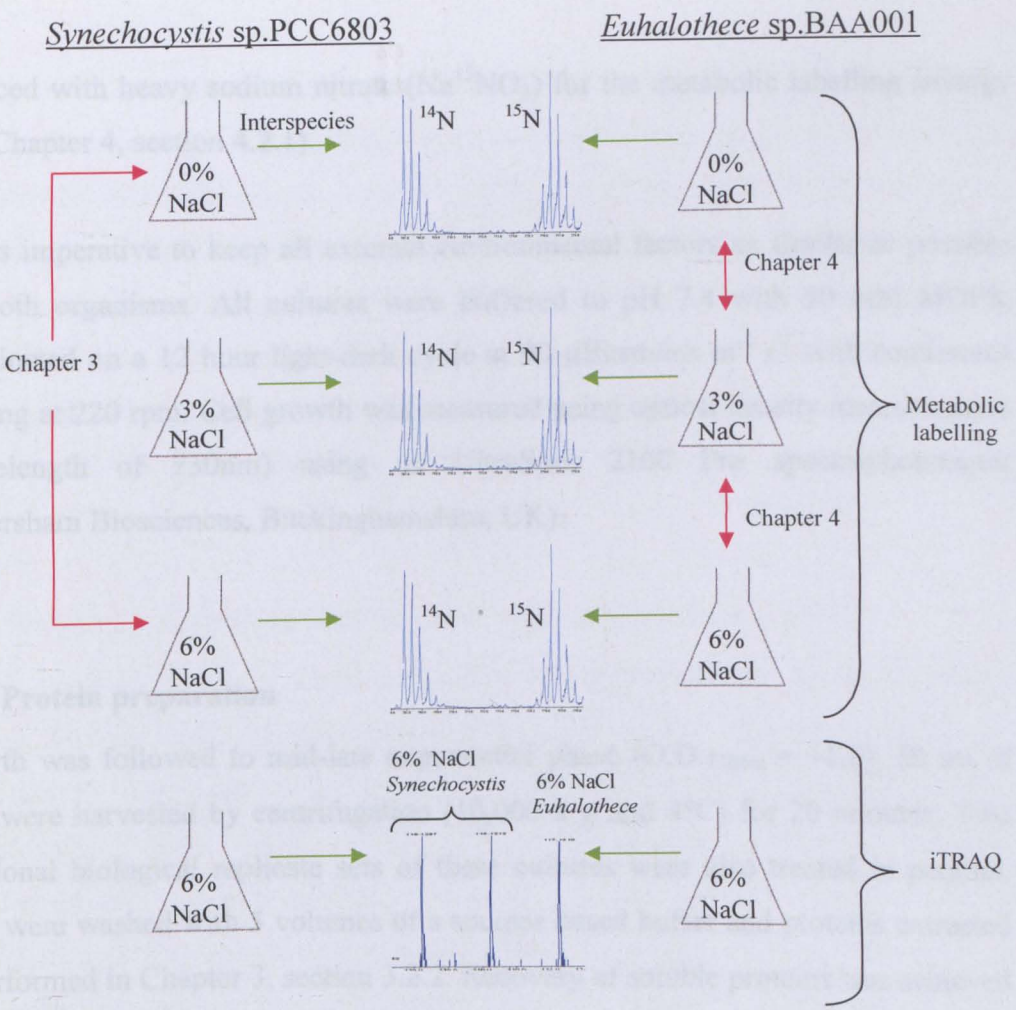


Figure 5.1: Overview of the experiment design. *Euhalothece* is cultured in ^{15}N-BG11 media and mixed in equal amounts with *Synechocystis* cells cultured in ^{14}N-BG11 media, in three salt concentrations (0, 3 and 6%), in triplicate. Shared tryptic peptides are used for quantitation. iTRAQ is performed using biological replicate cultures in BG11 media. Green lines represent interspecies comparisons and red lines represent intraspecies comparisons.

5.2.1 Cell culture preparations and growth

Euhalothece and *Synechocystis* cells were grown separately in batch culture in 250 ml flasks at 24°C in modified BG11 medium [135]. Three separate cultures were grown in triplicate, differing in salt (NaCl) concentration, low (0% w/v), medium (3% w/v) and high salt (6% w/v). All *Euhalothece* cultures were grown in modified BG11 medium, where the major nitrogen source, sodium nitrate ($NaNO_3$) was

replaced with heavy sodium nitrate ($Na^{15}NO_3$) for the metabolic labelling strategy (see Chapter 4, section 4.2.1).

It was imperative to keep all external environmental factors as similar as possible for both organisms. All cultures were buffered to pH 7.4 with 50 mM MOPS, illuminated on a 12 hour light-dark cycle at 90 μEinsteins m^{-2} s^{-1} with continuous shaking at 220 rpm. Cell growth was measured using optical density measurements (wavelength of 730nm) using an UltraSpec 2100 Pro spectrophotometer (Amersham Biosciences, Buckinghamshire, UK).

5.2.2 Protein preparation

Growth was followed to mid-late exponential phase ($O.D._{730nm}$ = ~1.0). 50 ml of cells were harvested by centrifugation (10,000 x *g* and 4°C) for 20 minutes. Two additional biological replicate sets of these cultures were also treated in parallel. Cells were washed with 5 volumes of a sucrose based buffer and proteins extracted as performed in Chapter 3, section 3.2.2. Recovery of soluble proteins was achieved by centrifugation (21,000g at 4°C) for 30 minutes. A HiTrap desalting column (Amersham Biosciences) was implemented according to the manufacturer's protocol. The total protein concentration obtained was quantified in triplicate using a range of dilutions (to maximise accuracy) with the RC DC Protein Assay (BioRad, Hertfordshire, UK).

5.2.3 Protein fractionation

As detailed in Figure 5.1, three different protein mixtures were created: (1) a mixture of 0% salt *Synechocystis* protein extract (^{14}N) and 0% salt *Euhalothece* protein extract (^{15}N), (2) a mixture of 3% salt *Synechocystis* protein extract (^{14}N) and 3% salt *Euhalothece* protein extract (^{15}N), and (3) a mixture of 6% salt *Synechocystis* protein extract (^{14}N) and 6% salt *Euhalothece* protein extract (^{15}N) (50μg of each protein extract).

Each protein mixture was combined with Sample Buffer/Laemmli Buffer (Bio-Rad), to 2 volumes of the original sample. 4% stacking gel and 12% resolving gel was used-for SDS-PAGE gels (size: 17 cm long × 17 cm width × 0.75 mm thick). Technical (n=2) and biological replicate (n=3) gels were run. The samples were loaded into wells and using the Protean II Multicell (BioRad) apparatus, they were run at 100V for 40 minutes and 280V for 4 hours. Afterwards, gels were washed and stained and after band excision, proteins were extracted and digested following the procedure detailed in Chapter 4, section 4.6.3.

5.2.4 Protein identification with peptide fractions from SDS-PAGE

Peptide separation was accomplished on a Pep-Map C-18 RP capillary column (Famos, Switchos and Ultimate LC system from Dionex/LC Packings, Amsterdam, Netherlands) interfaced to a QStar XL Hybrid ESI-qQ-TOF-MS/MS (AB, Framingham, MA, USA; MDS- Sciex, Concord, Ontario, Canada). Chromatography and MS settings were used as described in Chapter 4, section 4.6.4. IDA data was searched using Mascot Daemon (Matrix Science) version 2.1.3 in a sequence query type of search based on MS/MS spectra against the *Synechocystis* protein database (3264 ORF's, retrieved from NCBI Refseq, March 2007). Search parameters included a 1.2 Da peptide tolerance and 0.6 Da MS/MS tolerance. One mis-cleavage of trypsin was allowed. Modifications were set as carbamidomethyl of cysteine (fixed), and oxidation of methionine (variable). Only high confidence Mascot identifications were considered further, where a MOWSE score of >38 is equal to a significant hit of $p<0.05$ [93].

5.2.5 Protein quantitation using MSQuant

Protein quantitation was achieved using MSQuant software (http://msquant.sourceforge.net/) with default settings for ^{15}N labelling quantitation. This software requires raw data files from the mass spectrometer together with its search result file (HTML) generated using the Mascot Daemon search engine. It processes spectra and presents peptide quantitations which are further used to calculate protein quantitations. All quantitations were manually inspected. Proteins were only treated as significantly different in abundance between the species with

fold changes of 2 or greater (refer to section 5.3.2). In addition, the detection of 2 or more peptides was a prerequisite for quantitation significance (which ensured greater confidence in the accuracy of the ratio). Statistical evidence of quantitation was improved by also subjecting the same spots excised from two separate biological replicate gels (3 biological replicates in total) to the same workflow.

5.2.6 iTRAQ labelling and peptide fractionation

100μg of protein from (i) *Euhalothece* in 6% salt, (ii) *Synechocystis* in 6% salt and (iii) *Synechocystis* 6% salt (biological replicate) were precipitated using TCA/acetone at -20°C. This was performed as described in Chapter 3, section 3.2.2. The resulting pellet was washed with ice-cold acetone, and resuspended in 30 μL of 0.5M TEAB buffer. Each sample was reduced, alkylated, digested with typsin, and labelled with iTRAQ reagents according to the manufacturer's (AB) protocol with modifications. iTRAQ tags 115, 116 and 117 were used for *Synechocystis* in 6% salt, *Synechocystis* in 6% salt (biological replicate) and *Euhalothece* in 6% salt, respectively. After labelling, samples were combined and dried in a vacuum centrifuge (Model 5301, Eppendorf, Cambridgeshire, UK) at room temperature.

Fractionation of samples was achieved using strong cation exchange as described in Chapter 3, section 3.2.3. A UV detector UVD170U and Chromeleon Software v6.50 (Dionex/LC Packings, The Netherlands), was used to monitor the chromatogram as fractions were collected every minute using a Foxy Jr. Fraction Collector (Dionex). Pooling of fractions was performed depending on the peak intensities in the UV chromatogram and produced 20 samples. Samples were dried in a vacuum concentrator, and stored at -20°C prior to MS analysis.

5.2.7 Protein identification and data analysis using iTRAQ

Dried peptide samples (20) were resuspended in aqueous Buffer A (3% acetonitrile, 0.1% formic acid) and injected into a capillary liquid chromatography system coupled to a QStar XL Hybrid ESI Q-TOF-MS/MS. Using a flow rate of 0.3 μL/min, peptides were separated on a PepMap C-18 RP capillary column with the same parameters described in Chapter 3. Data acquisition in the mass spectrometer

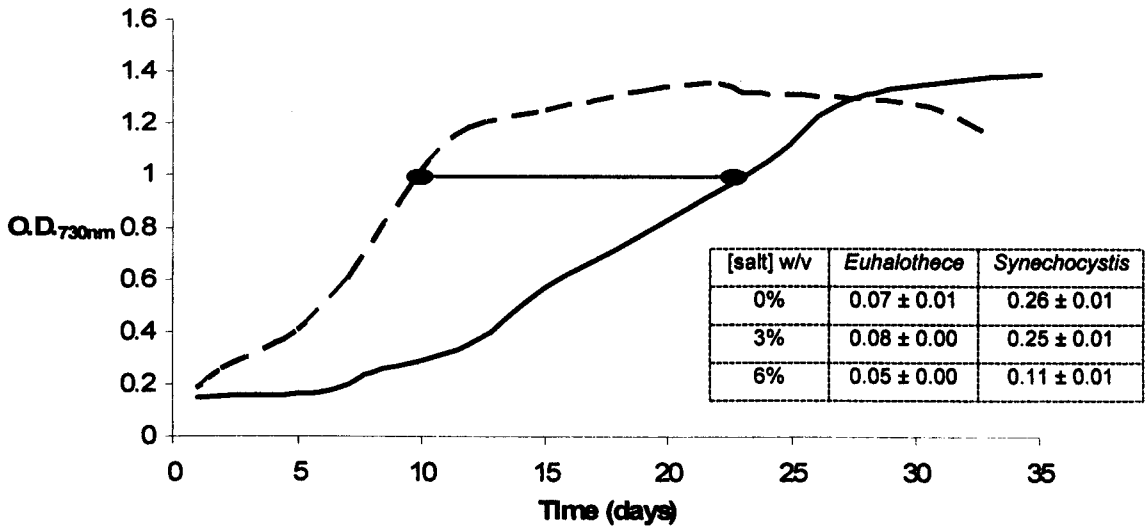
was set to the positive ion mode, with a selected precursor mass range of 300-2000 *m/z* over an accumulation period of 1 second. Tandem fragmentation of 2 dynamically selected precursors was performed over an extended scan from 65-1600 over 3 second accumulation. An elevated collision energy range was also implemented to overcome the stabilizing effects of the isobaric tags [85]. MS/MS was performed preferably on peptide ions with +2 and +3 charge states. Sample fractions were re-injected twice to increase coverage confidence in identification and quantitation.

A pooled list of acquisition data were analysed for protein identification and quantitation using ProteinPilot™ Software v2.0 (AB, MDS Sciex) which employs the Paragon Search algorithm [290]. The search was performed using the *Synechocystis* protein database (3264 ORFs) downloaded from NCBI (Ref Seq, March 2007). Customised search parameter settings have been employed to include cysteine-fixed modification, mass tolerance and mis-cleavages. Protein identifications were accepted as positive with a probability filter cut-off of 95% (Prot Score of $\geq$ 2.0) and P-value. False positive rate was calculated using a decoy interrogator search database [91, 98, 349], as performed in Chapter 3. The decoy database strategy using reversed concatenated proteome sequences has been shown to allow an accurate estimation of the theoretical error associated with false determination rate (FDR) measurements [98]. To achieve a rounded estimation for false positive determination, reversed proteomes of three cyanobacteria which are fully sequenced, and closely related to *Euhalothece* (based on 16S rRNA analysis, *Nostoc punctiforme* PCC73102, 78% and *Nostoc* sp. PCC7120, 75%) were employed for the estimation of FDR. Sequence reversal and concatenation was performed using a basic pre-programmed Perl script. FDR search was subsequently performed by re-analysing MS data files with identical parameters though ProteinPilot™ Software version 2.0.

5.3 Results and discussion

5.3.1 Growth analysis

The growth rate of *Euhalothece* is slower than *Synechocystis* in all three salt concentrations (Figure 5.2). Therefore, to allow a meaningful comparison of protein abundance data, cell harvesting and protein extraction were performed in each species during the *same* growth phase when cell density (based on O.D.) was as similar as possible (Figure 5.2). The mid-exponential growth phase (prior to phase 2 in Figure 5.3) was used.



[salt] w/v	*Euhalothece*	*Synechocystis*
0%	0.07 ± 0.01	0.26 ± 0.01
3%	0.08 ± 0.00	0.25 ± 0.01
6%	0.05 ± 0.00	0.11 ± 0.01

Figure 5.2: Growth curves of *Synechocystis* (dashed line), *Euhalothece* (complete line). The lines represent averaged figures of three biological replicates grown in 0% salt. Inset: growth rates (d^{-1}) for both organisms in the differing salt concentrations. Cells were harvested in the same growth phase, mid-late exponential (connector line).

Figure 5.3 visually demonstrates how *Euhalothece* and *Synechocystis* cells tolerate the effects of high salt. During phase 2, complete pigment loss (chlorosis) occurs in the *Synechocystis* cells in 6% salt concentration and higher. The maximum salt concentration *Synechocystis* can tolerate is ca.7% [161], but clearly the observations show cells must be revived in fresh media to maintain survival. Conversely, *Euhalothece* cells can tolerate all five salt concentrations but begin to show signs of chlorosis in 12% salt during mid-stationary phase (Figure 3, phase 3). It is therefore

recommended to refresh the media for this organism before this time period elapses. The loss of green colouration is due to a decrease in chlorophyll levels, but these high salt conditions have also been associated with the production of carotenoids in extremely halotolerant cyanobacteria, and this could explain the yellow/brown pigmentation [151].

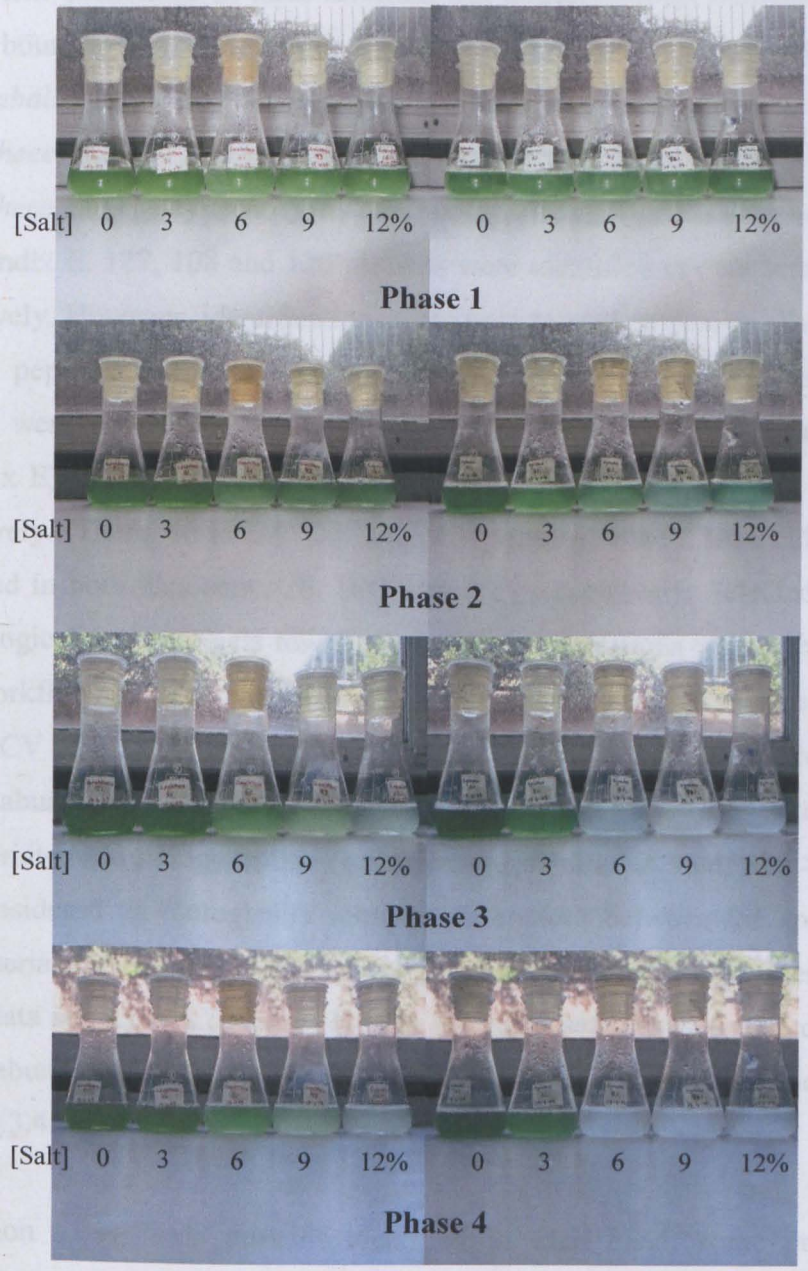


Figure 5.3: Effect of increasing salt concentration on the pigmentation of *Synechocystis* (right hand side) and *Euhalothece* (left hand side) cells. Phase 1 - early exponential phase, Phase 2- late exponential phase, Phase 3- mid-stationary phase, Phase 4-late stationary/death phase.

5.3.2 Metabolic labelling data analysis

The ^{15}N incorporation efficiency was computed using three peptides from the highly abundant protein phycocyanin, as performed in Chapter 4, and was found to be 98.05% ± 0.21.

The protein abundance in three different salt concentrations, compared across the species boundary, produced three data sets. These were denoted data set A (0% NaCl *Euhalothece* and 0% *Synechocystis* NaCl comparison), data set B (3% NaCl *Euhalothece* and 3% *Synechocystis* NaCl comparison) and data set C (6% NaCl *Euhalothece* and 6% *Synechocystis* NaCl comparison). The full data list is available in Appendix E. 127, 108 and 120 proteins were identified in data sets A, B and C, respectively. However, identifications were only treated as statistically relevant if 2 or more peptides per protein were identified. This meant a total of 203 unique proteins were confidently identified across all three conditions (summarised in Appendix E) and 87, 67 and 93 proteins were quantified in data sets A, B and C, respectively. There were 34, 30 and 22 proteins which were identified and quantified in both data sets A/B, B/C and A/C, respectively. Selected bands from two biological replicate gels for all three salt concentrations were subjected to the same workflow and a comparison of protein quantitation data gave a log mean average CV of 0.15. Therefore, for sufficient statistical stringency, proteins with a higher abundance (note here that up-regulation and down-regulation is inappropriate terminology in this cross-species quantitation study) by >2-fold only, were considered as biologically relevant differences between the species. Using these criteria, 26, 16 and 17 proteins had differential abundance between the species in 0% (data set A), 3% (data set B) and 6% (data set C) salt (refer to Table 5.2). Protein abundance differences are interpreted in respect to salt concentration in section 5.3.4.

In addition to multiple proteins identified in each band, some proteins were identified in more than one band. In the majority of cases these bands were adjacent to each other, however, for situations where this was not the case, only proteins which corresponded to gel bands of the correct molecular weight range were reported. Protein positioning which did not correctly match theoretical molecular

weight may be caused by damage or truncation. In all duplicate cases, quantitations were checked manually for consistency.

When using tryptic peptides as identifiers for proteins, they are often of sufficient length to be used as a unique identifier for the protein, referred to as distinct peptides. However, in many cases they may not be unique, and could be sourced from more than one protein and termed shared peptides. This makes it increasingly difficult to obtain meaningful quantitative data, and several studies have dealt with this protein inference problem with specific algorithms [350, 351]. Here, prefractionation using SDS-PAGE means protein size can be used as an additional identifier ensuring peptides are correctly assigned to a protein, and are unlikely to be *shared* peptides from a paralogous protein. It is important to note here that an unsequenced organism is a subject in this study, and it is not possible to elucidate whether peptides are *shared* or *distinct* within its proteome. Therefore avoiding misleading quantitations requires high confidence protein identifications.

Inconsistent quantitations occurred once in the 3% salt gel, where ribulose biphosphate carboxylase was identified in two gel bands. The following three peptides were used for quantitation, LTYYTPDYTPK, TFQGPPHGITVER, and LEDIRFPVALIK, giving different results (0.35 and 1.22-fold changes). Although these peptides are *not* shared in *Synechocystis* and are sourced from Ribulose biphosphate carboxylase only, it is not certain at this stage which protein quantitation applies to it. Therefore, by elucidation of the molecular weight in the gel (~50kDa), we can see that the peptides quantified as 0.35 (^{14}N and ^{15}N) are sourced from ribulose biphosphate carboxylase (~53kDa).

In this study, peptides from orthologous proteins in *different* organisms are used for quantitation and these proteins will have different amino acid sequences. This will affect the physicochemical properties of the protein and hence potentially change the accessibility of trypsin to its cleavage sites, C-terminal to arginine and lysine. Ultimately this would lead to varying peptide intensities for the same protein and lead to large standard deviation values. This does not seem to pose a problem using the in-gel digestion strategy implemented here, because the standard deviations are low with an average 0.09 ± 0.10 across all proteins quantified.

5.3.3 iTRAQ data analysis

An iTRAQ experiment was run with biological replicate cultures comparing *Euhalothece* and *Synechocystis* cells adapted to 6% salt. As noted, this was conducted firstly as an investigation to gauge the applicability of *in vitro* isobaric tagging in cross-species proteomics, and secondly as a corroborative technique to compare with findings obtained using *in vivo* metabolic labelling technique (i.e. a comparison between quantitative techniques employed in Chapter 3 and 4). Naturally, the limitations on proteomic identifications and quantitations for metabolic labelled cross-species studies also apply for the iTRAQ technique, where quantitation of proteins across all three labels could only occur where peptides of identical sequences exist across both species of interest. Using dynamic filters for protein confidence at a cut-off of 95%, a total of 207 unique proteins were reliably identified with MS/MS.

A measure of false positive identifications is invaluable in the case of handling a large proteomic dataset, hence the rate of false positive identifications was calculated using a well accepted methodology outlined by Elias *et al.* [98]. A reverse database was created using *Synechocystis*, *Nostoc punctiforme* PCC73102 and *Nostoc* sp. PCC7120 proteomes in order to assign a 'decoy' to identify falsely determined spectra using the same algorithm. Statistically, 24 falsely determined spectra were identified using the decoy database above the confidence interval of 80% and 10,722 spectra were assigned to valid peptides above the same level of confidence interval in the 'true' database search. Thus, as described by Elias *et al* [98], the resulting false positive rate determined should be in the region of 0.4%.

In an attempt to statistically relate the observations to biological variation in these systems, two biological replicate samples (115 and 116 tag labelled peptides) were used to estimate the CV, using calculations described previously [82] and in Chapter 3. The measure provided an average variation of 30%. Upon estimating a measure for biological variation, there were 20 proteins which were similarly determined to be significantly different (>2 fold increase) in relative abundance across *both* biological replicates.

Using iTRAQ labelling in cross-species proteomics is successfully demonstrated here, with the identification (using sufficiently stringent criteria) and quantitation of over 200 proteins. Of course, the overall success of similar cross-species proteomics studies using isobaric tags will always be dependant on the number of *shared* tryptic peptides between species, or between the unsequenced organism and sequences currently available in databases. These criteria equally apply to cross-species quantitation using metabolic labelling. Snijders *et al* [347] provided an *in silico* analysis of the using shared tryptic peptides for cross-species quantitation in several organisms (spanning archaea, eubacteria and eukaryotes). They revealed that approximately one third of all proteins shared >40% of peptides with at least one other protein in a related species, showing the strong application potential of this technique. The study in this chapter is the first application of this method, to the author's knowledge, in comparing biological adaptation in different organisms (see Appendix F for full iTRAQ results).

5.3.4 Protein abundance differences

Differences in protein abundance were interpreted in three parts relating to the salt concentrations under investigation (Table 5.2).

5.3.4.1 0% salt

The greatest differences in protein abundance (26 proteins) quantified using metabolic labelling was observed in 0% salt (see Table 5.2). This was expected because the study in Chapter 4, Part B, showed that *Euhalothece* potentially belongs to a unique cluster in its genera, and exhibits a 'stress' response in no salt. In contrast, *Synechocystis* is essentially a fresh water cyanobacterium [135, 161]. 9 proteins were more abundant in *Synechocystis* (i.e. with a quantitation ratio $^{14}N/^{15}N \geq 2$) and 17 were more abundant in *Euhalothece* (i.e. with a quantitation ratio $^{14}N/^{15}N \leq 0.5$) (Table 5.2). 5 proteins involved in energy metabolism as well as energy synthesis were more abundant in *Euhalothece*. This could be in response to the stress this extremophile experiences in no salt, a reaction not seen in *Synechocystis* in these conditions. Three stress related proteins were also more abundant in *Euhalothece*, including the hypothetical protein slr1894 (2.03-fold more abundant in *Euhalothece* i.e. with a quantitation ratio $^{14}N/^{15}N$ of 0.49 in Table

5.2) which is a putative DNA-binding stress protein. Rehydrin (2.08-fold more abundant in *Euhalothece*) harbours peroxidise activity (anti-oxidative) and methionine sulfoxide reductase A (2-fold more abundant in *Euhalothece*) has been shown to function as an antioxidant repair enzyme in *Arabidopsis* when ROS cause the oxidation of methionine residues in cells experiencing cold, drought or salinity stress [349]. However, a higher abundance (over 3-fold) of anti-oxidative enzyme superoxide dismutase was seen in *Synechocystis* cells when compared to *Euhalothece*. Two proteins involved in iron acquisition, iron transport protein and periplasmic iron binding protein, displayed the largest abundance differences in *Euhalothece* cells at 3.49 and 3.96-fold more than in *Synechocystis*, respectively. An up-regulation in periplasmic iron binding protein was seen in *Euhalothece* cells in Chapter 4, Part B, in 0% salt compared to 3% salt, suggesting they play a vital role in cells adaptation to no salt. In salt-stressed *Synechocystis* cells, proteins involved in protecting PSII under iron deficiency (FutA1 and Vipp1) have been induced [224] implying a co-relationship between proteins involved in iron and salt 'stress'. The regulatory protein LexA was also more abundant (2.02-fold) in *Euhalothece* cells compared to *Synechocystis* cells in these conditions.

5.3.4.2 3% salt

Euhalothece cells can tolerate much higher salt concentrations than *Synechocystis* cells [135, 149]. The differences in relative abundance of proteins in 3% salt support this, as a stress response is more evident in *Synechocystis* in comparison to *Euhalothece*. 13 proteins were more abundant in *Synechocystis*, whereas only 2 proteins were more abundant in *Euhalothece* (Table 5.2). *Synechocystis* cells appear to be more stressed in these conditions, and a larger energy requirement is postulated by more abundant ATP synthase subunits A (5.56-fold) and B (7.14-fold) and PPP enzyme phosphoribulokinase (2.08-fold). The latter enzyme is ultimately involved in CO_2 assimilation [352]. In addition, higher abundances of central intermediary (carbon utilisation) and energy metabolism proteins (glycogen phosphorylase, phosphoglucomutase, fructose-1,6-biphosphatase, ribulose biphosphate carboxylase) may imply the additional requirement of carbon for the synthesis of the compatible solute, GG; an increase in all these proteins has previously been observed at the protein level in acclimation of *Synechocystis* to ca. 4% salt compared to no salt [61]. *Synechocystis* cells also harbour relatively higher

amounts of nitrogen regulatory protein PII and GDP-D-mannose dehydratase compared to *Euhalothece*. The latter enzyme is responsible for modifying extracellular polysaccharides [353], which may be in response to external conditions. Increased levels of anti-oxidant superoxide dismutase (2.04-fold) implied extra stress in *Synechocystis*, together with hypothetical protein slr1963 (2-fold). As discussed in Chapters 3 and 4, hypothetical protein slr1963 shares similarity to an orange carotenoid protein, and these proteins are involved in cell protection during acclimation to salt environments [195, 196, 217]. Interestingly, in Chapter 4, this protein was found to be up-regulated in *Euhalothece* cells in response to 9% salt compared to 3% salt, but not in a lower salt concentration (6% compared to 3% salt), and this may explain the relatively lower abundance of this protein in *Euhalothece* in 3% salt in comparison to the amount observed in *Synechocystis*. Phycocyanin subunits *a* and *b* were the only proteins quantified to be less abundant in *Synechocystis* cells. Reduced pigment levels are associated with the salt stress response in both organisms, but the higher stress levels which *Synechocystis* cells appear to experience may explain its lower quantity [195].

5.3.4.3 6% salt

As with 3% salt phenotypes, 6% salt induces more of a stress response in *Synechocystis* cells. All proteins selected using metabolic labelling (17) were present in higher amounts in *Synechocystis* cells (Table 5.2). Three stress related chaperones, DnaK, GroEL and 60kDa chaperonin 2 were all present in higher abundances. Moreover, hypothetical protein slr0244 was 2.04-fold higher, and shares homology with proteins from the universal stress protein family. Similar to the 3% salt comparison, four proteins involved in energy generation (through photosynthesis) and carbon utilisation were present in higher quantities in *Synechocystis*. For example, the redox enzyme ferredoxin-NAPH oxidoreductase, which catalyses the final step of the photosynthetic electron transport chain, is 2.1-fold more abundant in *Synechocystis*. Decreased energy synthesis in *Euhalothece* cells adapted to high salt was reported previously in Chapter 4. Protein synthesis was also predicted to be higher in *Synechocystis* because 4 of the 17 more abundant proteins are involved in translation and two in amino acid biosynthesis.

Table 5.2: Significant protein abundance differences calculated using metabolic labelling of proteins (^{14}N is *Synechocystis* and ^{15}N is *Euhalothece*).

Protein	gi number	No. of peptides	Score	Quantitation[1] $^{14}N/^{15}N$	CV	Functional category
Abundance differences in 0% salt						
Glutamate-ammonia ligase	620123	2	96	0.35	0.69	Amino acid biosynthesis
Negative aliphatic amidase regulator	16331081	3	219	0.50	0.15	Amino acid trans.& metabolism
Transketolase	16329902	12	925	0.46	0.33	Energy metabolism
Phosphoglycerate kinase	2499503	13	915	0.44	0.39	Energy metabolism
Fructose- bisphosphate aldolase	16331386	4	280	0.21	0.38	Energy metabolism
Hypothetical protein slr1894	16329942	2	132	0.49	0.14	Hypothetical
Hypothetical protein slr 1841	16330041	7	443	0.48	0.29	Hypothetical
Hypothetical protein slr1762	16330280	3	200	0.47	0.09	Hypothetical
Hypothetical protein sll0529	16332076	4	307	0.32	0.48	Hypothetical
Rehydrin	16329971	2	139	0.48	0.17	Other categories
Phosphoribulokinase	585368	3	213	0.24	0.18	Photosynthesis and respiration
Ribulose bisphosphate carboxylase SSU	16331394	2	157	0.24	0.71	Photosynthesis and respiration
Uracil phosphoribosyltransferase	16329360	2	128	0.44	0.12	Purines/pyrimidines/nucleotides
SOS function regulatory protein	16330362	3	205	0.50	0.05	Regulatory
Methionine sulfoxide reductase A	16329407	7	343	0.50	0.11	Translation
Iron transport protein	16329434	4	224	0.29	0.67	Transport and binding proteins
Periplasmic iron-binding protein	16331793	7	455	0.25	0.38	Transport and binding proteins
Hypothetical protein sll0314	16331369	8	515	4.00	0.02	Hypothetical
Phosphate binding periplasmic protein prec.	16331543	4	237	3.85	0.03	Transport and binding proteins
Superoxide dismutase	16330619	7	448	3.23	0.06	Cellular processes
Elongation factor Tu	16330913	6	464	2.70	0.06	Translation
6-phosphogluconolactonase	2829619	3	161	2.56	0.02	Energy metabolism
Phycobilisome rod-core linker polypeptide	16329710	3	146	2.56	0.05	Photosynthesis and respiration
Hypothetical protein slr2018	16329857	4	263	2.44	0.09	Hypothetical
Putative phosphoketolase	16332268	2	68	2.38	0.06	Hypothetical
Hypothetical protein sll1785	16330236	3	182	2.17	0.06	Hypothetical
Abundance differences in 3% salt						
PII protein	1403577	3	205	2.27	0.03	Amino acid trans.& metabolism
GDP-D-mannose dehydratase	16329177	3	156	2.44	0.09	cell envelope
Superoxide dismutase	16330619	5	289	2.86	0.05	Cellular processes
Glycogen phosphorylase	16330178	5	253	3.33	0.23	Central interm. metabolism
Phosphoglucomutase	16332219	3	151	2.86	0.14	Central interm. metabolism
Fructose-1,6- biphosphatase	1753218	7	332	5.26	0.03	Energy metabolism
Hypothetical protein sll0529	16332076	8	386	2.22	0.03	Hypothetical
Hypothetical protein slr0006	16331389	3	122	2.04	0.25	Hypothetical
Hypothetical protein slr 1963	16330780	12	811	2.00	0.08	Hypothetical
ATP synthase subunit B	16330679	9	683	7.14	0.11	Photosynthesis and respiration
ATP synthase subunit A	16329327	9	531	5.56	0.09	Photosynthesis and respiration
Phosphoribulokinase	16331050	12	702	2.08	0.14	Photosynthesis and respiration
Methionine sulfoxide reductase A	16331392	5	262	2.86	0.12	Translation
Phycocyanin beta subunit	1008532	9	613	0.44	0.38	Photosynthesis and respiration
Phycocyanin a subunit	16329823	5	374	0.38	0.75	Photosynthesis and respiration
Abundance differences in 6% salt						
molybopterin biosythesis protein MoeB	16331030	2	143	2.78	0.02	Unknown
Protein confer. resistance to acetazolamide	16329873	4	227	2.70	0.02	Unknown
Tryptophan synthase subunit beta	16332038	2	87	2.17	0.06	Amino acid biosynthesis
Phosphoglycerate dehydrogenase	16330470	5	243	2.17	0.03	Amino acid biosynthesis
Trigger factor	16332069	5	234	3.33	0.06	Cellular processes
60kD chaperonin 2	16331442	20	1007	3.33	0.08	Cellular processes
Molecular chaperone DnaK	16331261	7	314	2.22	0.12	Cellular processes
Chaperonin GroEL	16330003	13	684	2.22	0.08	Cellular processes

Transaldolase	16329404	5	253	2.56	0.02	Energy metabolism
Hypothetical protein slr0244	16329299	3	160	2.04	0.09	Hypothetical
Ferredoxin-NAPH oxidoreductase	16331051	5	227	4.00	0.21	Photosynthesis and respiration
Phycocyanin associated linker protein	16329821	10	572	3.85	0.05	Photosynthesis and respiration
ATP synthase subunit A	16329327	6	252	3.45	0.06	Photosynthesis and respiration
Carbon dioxide concentrating mechanism prot.	16329366	6	250	3.03	0.09	Photosynthesis and respiration
Elongation factor Tu	16330913	17	1021	5.26	0.05	Translation
Translation initiation factor IF-2	16329288	9	402	4.00	0.08	Translation
50S ribosomal protein L7/L12	16330008	2	119	3.33	0.03	Translation
ATP-dependant Clp protease regulatory SU	16331384	12	533	2.70	0.17	Translation
Elongation factor G	46483	10	492	2.08	0.05	Translation

Note: [1]A quantitation ratio > 2 or < 0.5 is deemed to be a significant difference.

20 proteins were identified as being present in significantly different amounts between the two species using iTRAQ tagging. Similar to the metabolic labelling results, all were present in higher abundance in *Synechocystis* (Table 5.3). In agreement with the metabolic labelling results, photosynthesis and respiration seem to be more prominent processes in *Synechocystis* in 6% salt. Four proteins were identified here, as well as fructose-1,6-biphosphatase, involved in energy metabolism. Concurrently, 3 proteins are involved in protein synthesis (transcription and translation). The increased amount of FKBP-type peptidyl-prolyl cis-trans isomerise (2.18 and 2.59-fold) could be attributed to the increased amount of ionic stress experienced by proteins in *Synechocystis* cells. This enzyme is known to accelerate protein folding by catalysis of cis-trans isomerisation in specific oligopeptides [301]. The higher accumulation of glycogen phosphorylase (2.59 and 2.6-fold) implies the accrual of glycogen in cells of *Synechocystis* (a salt response reported previously in this organism in Chapter 3 and by Zuther *et al.* [196] as being more important in this cyanobacterium in these conditions. The largest abundance difference was seen in hypothetical protein sll0597 (6.83 and 11.10-fold). When searched using Blastp, it is possible this protein harbours permease activity. Permeases are a class of transmembrane proteins which facilitate the movement of molecules in or out of cells. Hypothetical protein sll0597 shares 45% homology to a predicted permease in *Nostoc punctiforme* PCC 73102 (E value $4x10^{-71}$).

Table 5.3: Significant protein abundance differences calculated using iTRAQ methodology in 6% salt (label 115 and 116 are *Synechocystis*, label 117 is *Euhalothece*)

Protein	gi number	115:117 ratio	CV	116:117 ratio	CV	Functional category
Ferredoxin-dependent glutamate synthase	16332153	2.45	0.23	3.28	0.29	Amino acid biosynthesis
Carboxyl-terminal protease	16330090	2.67	0.32	3.79	0.39	Cell envelope biogenesis
Glycogen phosphorylase	16330143	2.59	0.26	2.60	0.28	Central intermediary metabolism
Fructose-1,6-bisphosphatase	16331010	2.32	0.30	2.18	0.24	Energy metabolism
Hypothetical protein sll0597	16331809	6.83	0.58	11.10	0.57	Hypothetical
Hypothetical protein sll1863	16331180	4.41	0.56	2.20	0.32	Hypothetical
Hypothetical protein slr1963	16330780	2.15	0.24	2.93	0.28	Hypothetical
Hypothetical protein slr2144	16330319	2.79	0.29	3.35	0.34	Hypothetical
FKBP-type peptidyl-prolyl cis-trans isomerase	16329650	2.18	0.32	2.59	0.28	Other catergories
Apocytochrome f precursor	16330219	2.14	0.24	2.12	0.22	Photosynthesis and respiration
Chloroplast membrane-associated 30 kD protein	16331547	2.06	0.26	2.37	0.22	Photosynthesis and respiration
Photosystem II complex extrinsic protein precursor	16329439	2.33	0.24	2.30	0.27	Photosynthesis and respiration
Plastoquinol--plastocyanin reductase	16330220	2.90	0.22	2.19	0.21	Photosynthesis and respiration
Isopenicillin N epimerase	16330317	2.17	0.27	2.68	0.30	Posttranslational modification, protein turnover, chaperones
DNA-directed RNA polymerase beta subunit	16329957	2.08	0.29	2.35	0.33	Transcription
50S ribosomal protein L15	16329924	2.21	0.25	2.23	0.28	Translation
Leucyl aminopeptidase	16330631	2.41	0.24	2.19	0.40	Translation
Phosphate-binding periplasmic protein precursor	16331543	2.59	0.23	2.47	0.23	Transport and binding proteins
Extracellular nuclease	16329547	3.29	0.41	3.66	0.49	unknown
General secretion pathway protein G	16330343	2.64	0.15	2.40	0.19	unknown

5.3.5 Comparison of metabolic labelling and iTRAQ findings

Metabolic labelling and iTRAQ analysis successfully identified 340 unique proteins. In addition to previous studies [67, 87] these results illustrate that different techniques have inherent bias towards the identification of different proteins, thus highlighting the complementary nature of the various methods. It is evident that future proteomic strategies will ultimately benefit by the association of different workflows, and this will enhance the proteome coverage.

A total of 47 proteins were identified and quantified by both labelling strategies in the 6% salt phenotype comparison. An analysis of variance (ANOVA) was made between the relative quantitations of these proteins and the scatter plot is shown in Figure 5.4. This figure shows that no significant correlation was observed (Pr = 0.21 using a linear variance model) and that the relative abundance of a majority of proteins (~70%) were less than the necessary 2-fold change needed for the change in protein abundance to be deemed significant beyond biological variation (as

discussed in the sections 'Metabolic labelling data analysis' and 'iTRAQ data analysis). Although the experiment in this study was not specifically designed to statistically compare the two quantitation techniques, for abundance ratios > 2-fold, metabolic labelling appears to produce higher expression ratios (Figure 5.4). A previous study using both iTRAQ and cICAT techniques to search for cancer markers, analogous to this study, found several proteins that were differentially expressed in cICAT, had a smaller than critical (2-fold) change by iTRAQ labelling [354].

The direction of fold change (higher or lower relative abundance) was the same for 91.5% of proteins (see Appendix E). In addition, both methods functioned hand in hand to provide complementary information as to how *Synechocystis* and *Euhalothece* differ in their response to 6% salinity. Cumulatively, 18.1% and 11.3% of the proteins identified and quantified in the 6% salt comparison using metabolic labelling and iTRAQ respectively, were in higher abundance in *Synechocystis*. Similarities in overall relative abundance were found to extend across proteins involved in energy generation (photosynthesis and respiration), energy metabolism and protein synthesis, where all were present in greater abundances in *Synechocystis*.

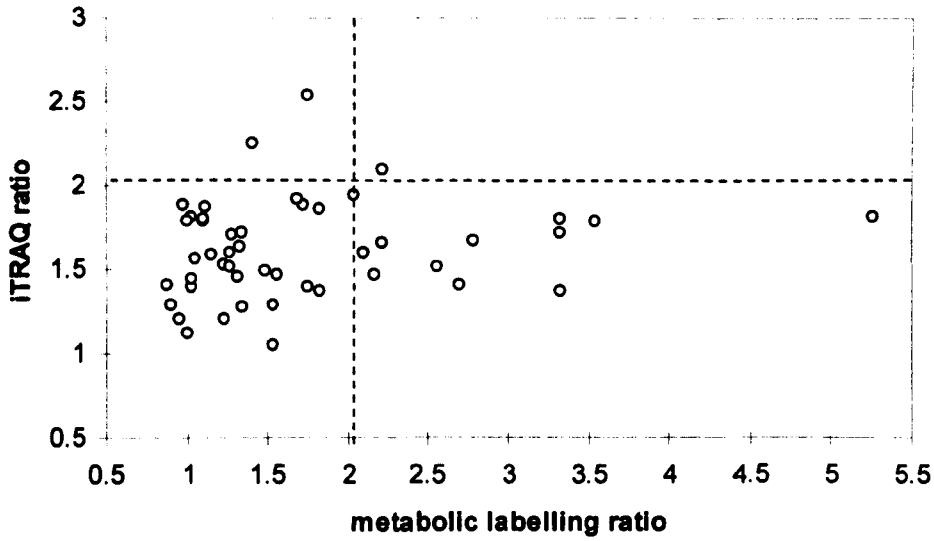


Figure 5.4: Scatter plot of the relative protein abundance ratios of protein quantitations for the 47 spots identified and quantified by both methods (metabolic labelling and iTRAQ) in 6% salt. Dashed line denotes the ratio (i.e. fold ratio of 2) where differential protein abundance is deemed to be insignificant due to technical, experimental and biological variation (refer to 'Metabolic labelling analysis' and 'iTRAQ data analysis').

5.4 Concluding remarks

A comparison of protein abundances in two contrasting cyanobacteria was undertaken by exploiting shared tryptic peptides. Low standard deviation calculations of isotope abundance differences (0.09 ± 0.10) imply in-gel digestion using trypsin was similar between orthologous proteins and 15% log mean average CV across biological replicates (0.15) ensured high confidence in differentially expressed proteins which were identified.

Although limited to proteins which are orthologous to both species, essential differences in cell behaviour in three differing salt concentrations between the species could be proposed. As differences in protein abundance were made in cells cultured in the same conditions, a direct comparison of evolutionary divergence between the species was possible. For example, differences in the requirement of molecular chaperones were used to postulate specific stress levels experienced. *Euhalothece* cells display stress characteristics in 0% salt, unlike *Synechocystis*

cells. Salt concentrations of 3% and 6% lead to a similar stress response, but this time dominating in *Synechocystis* cells, although reduced energy synthesis was a mechanism associated with *Euhalothece* salt-adapted cells in Chapter 4. This makes it possible to assume that many salt adaptation strategies involving protein synthesis are shared by both cyanobacteria, and their quantitative expression affects acclimation. The results imply that lower salinity levels (3% and 6% NaCl) lead to these characteristic responses in *Synechocystis* cells, when compared to *Euhalothece* cells.

It is probable that these differences have evolved at different stages, and could include alterations in gene control by transcription factors, differences in protein activity and functionality, changes in the way cells respond to external signals through adjustment of signal transduction, and response regulators or even generation of paralogs of essential or previously regarded as non-essential genes for salt acclimation [355, 356]. Ultimately, further work is required to understand exactly how *Euhalothece* cells have higher tolerance levels, more specifically, how this organism is able to survive in 6% salt by expressing salt adaptive protein-coding genes at lower levels compared to its fresh water relative.

Chapter 6: Proteome characterisation of the globally significant marine cyanobacterium, *Prochlorococcus marinus* MED4, and analysis of adaptation to light[4]

[4] This work was published (in part) in *Journal of Proteomic Research* 6 (3) pp 996-1005 (2007)

6.1 Introduction

The previous three chapters have successfully demonstrated how differential expression proteomics can reveal insights into how two contrasting cyanobacteria adapt to low and high salinity. In this chapter, the organism of interest is also a cyanobacterium, but the effects of changing light intensity are investigated. This study is an important addition to this thesis because it demonstrates the applicability of the tools and techniques in investigating a different environmental factor, and also due to the significance of the organism studied (this is the first proteomic characterisation of this species).

Previous studies (Chapter 2) have shown that there are many shared responses in bacteria acclimated to salt and light stress. Most notably, pigment structures must be re-organised because of the resulting effects on photosynthesis, the stoichiometry of PSI and PSII is altered, stress related molecular chaperones are produced to rescue misfolded proteins and anti-oxidative enzymes are synthesized to limit secondary stress damage [254, 256-258]. Despite these related acclimation strategies, the stresses are fundamentally different, with the main issue in salt stress being ionic and osmotic stress, whereas excess absorbed energy is the main detrimental factor in light stress.

The cyanobacterium in this study, *Prochlorococcus marinus*, is very different to all cyanobacteria, including *Synechocystis* and *Euhalothece*. *P. marinus* was discovered in 1988 [15], and found to reside in nutrient poor oceans, which, similar to salt lakes, were thought to be void of living microorganisms. Since their discovery in this hostile habitat, the interest in their ecological impact has increased greatly, mostly because they are believed to be the most abundant known photosynthetic organisms on earth [16], and therefore substantial contributors to primary production [276]. In contrast to other cyanobacteria, they contain no phycobilisome structures, and instead harbour a major light harvesting complex protein in their thylakoid membranes, which binds divinyl derivatives of chlorophyll *a* and *b* (chl a_2 and b_2) [16, 281]. Furthermore, these pigments are not found in other cyanobacteria, and these unique characteristics make a study on light acclimation in *P. marinus* particularly novel and significant.

P. marinus is an autotrophic organism, and therefore obtains all its carbon and chemical energy required for growth through photosynthesis. It resides in oceans, where neither quality nor quantity of light is constant, and therefore it must adapt to the changing conditions. Different strains of *P. marinus* have acquired mechanisms to enable survival on the ocean surface, characterised by HL, and up to depths of 200 metres, where very little light is distributed (LL). Several cellular features play a role in adaptation of these differing ecotypes [157, 280, 283, 284]. Strains adapted to the ocean surface are referred to as HL ecotypes, and contain a low ratio of chl b_2/a_2 and possess one antenna gene (*pcb*) [282]. Contrastingly, strains adapted to deeper ocean depths contain a high ratio of chl b_2/a_2, and the possession of multiple antenna genes [282].

The HL adapted strain MED4 is the subject of study here. It has its genome fully sequenced and annotated and at 1.66 Mbp, it is much smaller than the largest sequenced *P. marinus* strain MIT9313 (2.4 Mbp) [279]. Unlike LL ecotypes, HL ecotypes can adapt to both conditions (HL and LL) and the nature of rigorous surface water mixing compared to deeper water, means they represent interesting species for investigating strategies of acclimation to changing irradiances [286]. It is not certain how it is able to achieve this, although comparisons to LL ecotypes reveal the possession of many more HL inducible proteins (HLIP's) and photolyases to repair DNA damaged by UV light [285].

MED4, as well as all photosynthetic organisms, must adapt to varying light intensities for survival. If acclimation cannot be achieved the production of ROS ultimately leads to cell death [254]. The immediate (independent of gene expression) and long-term cellular response to HL were discussed in Chapter 2, and include state transition and photosynthetic apparatus alterations, respectively [256-258]. However, it is rarely feasible to use the nucleotide sequence of an organism to predict ecological function and environmental adaptation, without information on which proteins are synthesised in specific conditions. Therefore, the field of proteomics is potentially useful here and before this thesis work, no high-throughput proteomics studies have been published which aim to understand light acclimation. 2DE has been applied to understand HL stress in the alga

Chlamydomonas reinhardtii [273]. Quantitative comparisons on 444 proteins based on gel staining intensity, identified 105 (24%) which were significantly differentially expressed. Some of these proteins have potentially novel functions in photo-protection, for example, the role played by NAB1 protein in post-transcriptional regulation of light harvesting complex protein [273]. In addition, many proteins involved in carbon dioxide fixation and molecular chaperones were induced in HL.

Prior to this thesis, no published work has been carried out investigating the proteome of *P. marinus*, and therefore an additional aim of this chapter is to characterise the proteome of *P. marinus* MED4. Characterisation will be performed with specific comparisons to previously published shotgun studies in cyanobacteria. The isobaric tagging technique, successfully demonstrated in Chapter 3 and 5, will be employed here to investigate adaptation to three varying levels of light intensity, imitating natural oceanic conditions. However, subsequent bioinformatics analysis will be performed using ProQuant and ProGroup software for protein quantitation, identification and relative differential expression. Essentially, through this chapter, the ultimate goals are to characterise the proteome of a globally abundant marine cyanobacterium, whilst furthering understanding of its mechanisms of photoacclimation.

6.2 Material and methods

6.2.1 Culture conditions and cell growth

P. marinus MED4 was provided by the Roscoff Culture Collection (RCC), and grown in seawater based Pro99 medium [357]. It was originally isolated from the northwest Mediterranean Sea in 1989 from surface waters, and cultured in PCR-S11 medium. The seawater used for the media was obtained locally from the coastline of Hull, UK (see 'Seawater comparison', section 6.2.2). 250 ml flasks were used to culture the microorganism at 24°C, with 12 hour light/dark cycles and continuous shaking (220 rpm). Three separate cultures differing in light intensity were grown, low (20 μEinsteins m^{-2} s^{-1}), medium (60 μEinsteins m^{-2} s^{-1}) and high (100 μEinsteins m^{-2} s^{-1}). These were sub-cultured three times in late-exponential phase

(ca. 30 days) to ensure adaptation to the different conditions. Light measurements were made using a Loggerhead 2100 light probe (Biospherical Instruments, San Diego, CA, USA). Optical density measurements were used to follow cell growth at the same time of day at a wavelength of 446 nm using an UltraSpec 2100 Pro spectrophotometer (Amersham Biosciences, Buckinghamshire, UK). As in previous chapters, growth was measured at the same time of day because the cell cycle is tightly linked to the light/dark phase [156].

6.2.2 Seawater comparison

Levels of major and minor trace elements in local seawater (Hull, UK) and English Channel seawater (Roscoff, France) were compared using inductively coupled plasma mass spectrometry (ICP-MS) (Perkin Elmer, Elan 400). Levels of 18 trace elements and 5 major elements (Ca, Fe, K, Mg, Na) were compared. A detailed composition of seawater [358] was used to estimate the standard addition ranges of a multi-element standard Merck VIII (Merck, Darmstadt, Germany) required for accurate abundance measurements. For both seawater samples two dilution sets were made, a 500:1 and a 1000:1 dilution. There was either- no standard added, 100 ppb standard added, 300 ppb, 500 ppb, 1000 ppb, 3000 ppb and 5000 ppb of the standard added. These standard additions were used for measuring major elements. For trace elements, the same dilutions of seawater (500:1 and 1000:1) were used with again no standard added, with 0.1ppb standard added and 0.3 ppb, 0.5 ppb, 1.0 ppb, 3.0 ppb and 5.0 ppb standard added. A linear relationship between the concentration of element added and the measured signal (cps) was used to calculate the concentration of the elements (full results in Appendix G).

6.2.3 Protein extraction and iTRAQ

50 ml of cells were harvested at mid-exponential phase during the mid-light cycle, and centrifuged at 10,000 g for 20 minutes at 4°C. Cells were washed with 5 volumes of a sucrose-based buffer as described in Chapter 3, section 3.3.2. For protein extraction, pelleted cells were re-suspended in 1 ml freshly made extraction buffer consisting of 9 M urea, 4% (w/v) CHAPS, 1% (w/v) DTT, 0.8% (w/v) Pharmalyte, 5% (w/v) protease inhibitor cocktail and 1.1 g l^{-1} PVPP. Proteins were

extracted using mechanical cracking and liquid nitrogen and recovered by centrifugation at 21,000 x *g* for 30 minutes at 4°C (refer to Chapter 3, section 3.3.2). The total protein concentration obtained was quantified using the RC DC Protein Assay (Bio-Rad). 100μg of protein from all three light conditions including a HL duplicate (technical replicate) were precipitated using TCA/acetone at -20°C, as described in Chapter 3, section 3.3.2, and subsequently resuspended in 30 μL of 500 mM TEAB buffer. Each sample was reduced, alkylated, digested, and labelled with iTRAQ reagents according to the manufacturer's (AB) protocol with modifications [82] (Table 6.1). After labelling, samples were combined and dried in a vacuum centrifuge (Model 5301, Eppendorf, Cambridgeshire, UK) at room temperature.

Table 6.1: Different phenotypes labelled with iTRAQ label reagents. Technical replicates are shown in **bold** type.

	iTRAQ label reagent			
Experiment	114	115	116	117
1	**HL**	LL	**ML**	**HL**
2	LL	**ML**	**HL**	**HL**

Note:
HL- High light at 100 μEinsteins m^2 s^{-1}
ML- Medium light at 60 μEinsteins m^2 s^{-1}
LL- Low light at 20 μEinsteins m^2 s^{-1}

6.2.4 SCX fractionation

Peptides were resuspended in 200 μL of aqueous Buffer A (defined in Chapter 3, section 3.2.3). Separation was performed on a BioLC HPLC unit (Dionex, Surrey, U.K.) using a PolySULFOETHYL A column (PolyLC, Columbia, MD) 5 μm of 100 mm length × 2.1 mm i.d. and 200 Å pore size. The 60 minute gradient consisted of 100% Buffer A for 5 minutes, 8 to 24% Buffer B (defined in Chapter 3) for 40 minutes, 24 to 100% Buffer B for 5 min, 100% Buffer B for 5 min, and finally 100% Buffer A for 5 min. A flow rate of 0.2 ml/min was maintained with an injection volume of 200 μL. A UV detector UVD170U and Chromeleon Software v6.50 (Dionex/LC Packings, The Netherlands), was used to monitor the chromatogram as fractions were collected every minute using a Foxy Jr. Fraction

Collector (Dionex). Fractions were pooled, depending on the peak intensities in the chromatogram and then dried in a vacuum centrifuge. Samples were stored at -20°C prior to MS analysis.

6.2.5 Mass spectrometry and protein identification

MS was performed using a ESI-qQ-TOF-MS/MS (AB, Framingham, MA; MDS-Sciex, Concord, Ontario, Canada) coupled with an online capillary LC system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands). 55 µL of Buffer A (0.1% formic acid and 3% acetonitrile) was used to resuspend dried samples ready for the MS and 20 µL was injected to the nano-LC-ESI-MS/MS system for each analysis. Initial separation took place on a PepMap C-18 RP capillary column (LC Packings) with a constant flow rate of 0.3 µL/min. Buffer B was made up as 97% acetonitrile, 0.1% formic acid. The gradient was started as 97% Buffer A and 3% Buffer B for 3 minutes, followed by 3 to 25% Buffer B for 120 minutes, 90% Buffer B for 7 minutes and finally 97% Buffer A for 7 minutes. Data acquisition in the mass spectrometer was set to the positive ion mode, with a selected mass range of 300-2000 *m/z*. MS/MS was performed on peptides with +2 and +3 charge states. ProQuant software v1.1 (AB, MDS-Sciex) was used for protein quantitation and identification. Identifications were made using the *P.marinus* protein database downloaded from JGI in November 2005 (http://genome.jgi-psf.org/finished_microbes/prom4/prom4.home.html). Search parameters were set at a mass tolerance of 1.2Da, MS/MS tolerance of 0.6Da, one missed cleavage of trypsin, oxidation of methionine and cysteine modification of methyl methanethiosulphonate (MMTS). Peptides with above 80% confidence were used for identification and quantitation, and ProGroup Viewer software v1.0.6 was used to identify proteins with >95% confidence. The expression ratios were calculated using equations adapted from the ProQuant software tutorial [83]. The entire iTRAQ experiment was conducted twice (experimental replicate) using different cultures (biological replicate) to gain statistical evidence for differential expression of proteins in changing light intensities.

6.2.6 Protein data analysis

All identified proteins were organised into functional groups using annotation based on clusters of orthologous groups, COG's (http://genome.jgi-psf.org/finished_microbes/prom4/prom4.home.html). The cellular location of these proteins was predicted using the program PSORTb v2.0 [248], and lipoproteins were detected using LipoP v1.0 [249] (refer to Chapter 4). The hydropathic character of all the proteins identified was expressed using the GRAVY index [359], and was used to support the predictions of any integral membrane proteins. Finally, the abundance of each protein was estimated using the exponentially modified protein abundance index (emPAI) [360]. EmPAI is exponentially related to PAI which is defined as the number of observed unique peptides per protein normalised by the number of peptides obtained by *in silico* digestion. *In silico* digested peptides were generated using Proteogest software [323].

6.3 Results and discussion

The results are presented and discussed in two parts. Firstly the proteome is characterised with comparisons to previous proteome studies in cyanobacteria and secondly, the findings in relation to photoacclimation are considered.

6.3.1 Growth studies

P. marinus MED4 was cultured in the laboratory in three different light intensities, low (20 μEinsteins m^{-2} s^{-1}), medium (60 μEinsteins m^{-2} s^{-1}) and high (100μEinsteins m^{-2} s^{-1}). The growth rate increased with increasing light intensity. In LL the growth rate was 0.13 ± 0.01d^{-1}, in medium light (ML) 0.22 ± 0.02 d^{-1} and HL 0.57 ± 0.05 d^{-1}. These growth rates are comparable to those seen in cultured isolates by Moore and Chisholm [156], where the growth rates of ten *Prochlorococcus* strains were measured by enumerating cells with a FACScan flow cytometer (Figure 6.1). The growth rate comparisons between both sources of seawater (Hull UK and Channel), also shown in Figure 6.1, show that *Prochlorococcus* MED4 grows at remarkably similar rates, despite the fact that Hull seawater contains more than 250% more iron, and higher concentrations of limiting elements (see Appendix G for full ICP-MS comparison of the different seawater).

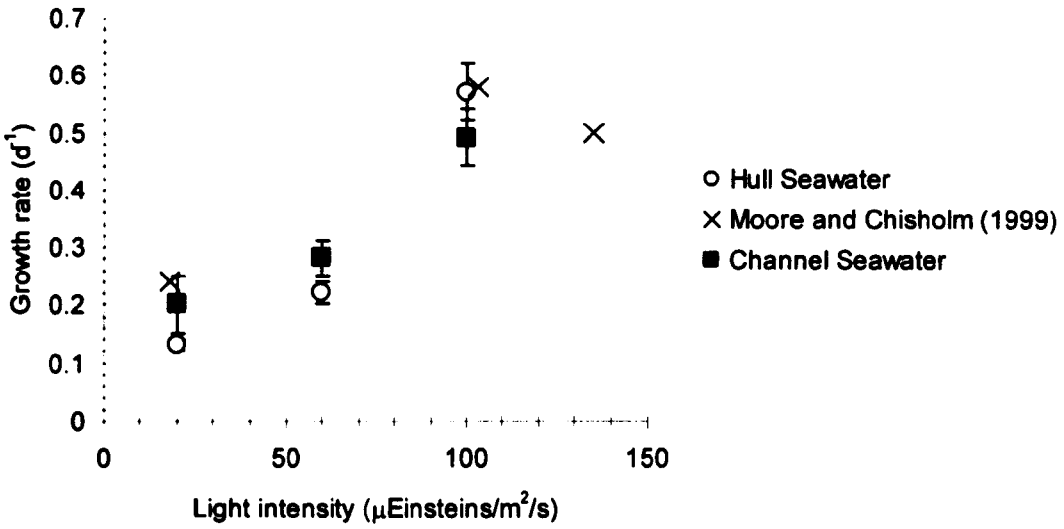


Figure 6.1: Growth rate as a function of light intensity. The growth rate of *P. marinus* MED4 was compared between cultures grown in Pro99 medium made with seawater from Hull, UK and Channel seawater (conducted in this study), and a previous study conducted by Moore and Chisholm [156]. The results from this study are presented as the average ± standard deviation based on triplicate measurements.

6.3.2 Overview and characterisation of proteome

The proteins identified in this study were characterised, and compared to findings in similar proteomic experiments conducted in cyanobacteria. Two iTRAQ experiments were conducted, and will be referred to as iTRAQ1 and iTRAQ2. A total of 184 proteins were identified; 53 proteins were identified in both experiments, 94 unique to iTRAQ 1 and 37 unique to iTRAQ 2. This total corresponds to ca. 11% of total putative protein-coding genes. This is a similar proteome coverage seen in Chapter 3 (11.9%), and previous studies using a typical shotgun approach in cyanobacteria such as *A. variabilis* ATCC 29413 (ca. 13%) [73], and *Synechocystis* [67] (ca. 11%). Approximately 40% of proteins were identified with 2 or more peptides. Where only 1 peptide was identified, 38% of them were identified multiple times (in different MS/MS experiments).

Every functional category based on COG [361] was represented (groups C to S), apart from group N (cell motility and secretions), which only consists of two proteins (Table 6.2). The most significant fraction of identified proteins (18.4%) could not be classified into COGs, and requires further annotation. The highest represented and annotated proteins were involved in translation, ribosomal structure and biogenesis (12.2%). 33.7% of proteins were involved in metabolism, which is surprisingly similar to the 33% involved in cellular metabolism observed in *A. variabilis* ATCC 29413 [73].

Table 6.2: Number and percentage of total identified proteins grouped by Clusters of Orthologous Groups (COG).

Code	COG group	Identified proteins	% of total
	Metabolism		
C	Energy production and conversion	16	8.2
G	CHO transport and metabolism	12	6.1
E	AA transport and metabolism	17	8.7
F	Nucleotide transport and metabolism	6	3.1
H	Coenzyme metabolism	9	4.6
I	Lipid metabolism	4	2.0
Q	Secondary metabolite biosynthesis, transport and catabolism	2	1.0
	Information storage and processing		
J	Translation, ribosomal structure and biogenesis	24	12.2
K	Transcription	10	5.1
L	DNA replication, recombination and repair	7	3.6
	Cellular processes		
D	Cell division and chromosome partitioning	5	2.6
O	PTM, protein turnover, chaperones	17	8.7
M	Cell envelope biogenesis, outer membrane	5	2.6
N	Cell motility and secretion	0	0.0
P	Inorganic ion transport and metabolism	7	3.6
T	Signal transduction mechanisms	5	2.6
R	General function prediction only	7	3.6
S	Function unknown	7	3.6
	No COG	36	18.4

The proteins observed experimentally were mapped against the theoretical proteome in relation to Mw and pI in Figure 6.2. Of the total proteins observed, 73.3% were acidic and 26.7% basic. This was an unexpected result, because the

theoretical proteome of *P. marinus* MED4 is 60% basic, implying that the techniques used are biased towards acidic proteins. Whether this is the result of sample preparation prior to or during iTRAQ, or the iTRAQ procedure itself, is not currently known. If the acidic proteome is only considered in this study, coverage would be 18.5% of the total protein-coding genes. In Chapter 3, 78.8% of proteins identified in *Synechocystis* were acidic, however, its proteome is biased towards an acidic pI [362]. 71.1% of identified proteins had a pI value between 4 and 7, the lowest pI was 3.72 (hypothetical protein PMM0790) and highest 12.12 (30S ribosomal protein S11, a ribosomal structure protein). This is very similar to the range seen in other cyanobacteria such as *A. variabilis* ATCC 29413 [73] (3.72 to 12.02), *Synechocystis* in a study by Gan *et al.* [67] (3.44 to 11.81) and in Chapter 3 (3.19 to 12.53). The observed Mw range in *P. marinus* MED4 was 4.2 kDa (hypothetical protein PMM0465) to 168.1 kDa (glutamate synthase in amino acid transport and metabolism), comparable to *A. variabilis* ATCC 29413 (4.3 KDa to 171.3 kDa) [73].

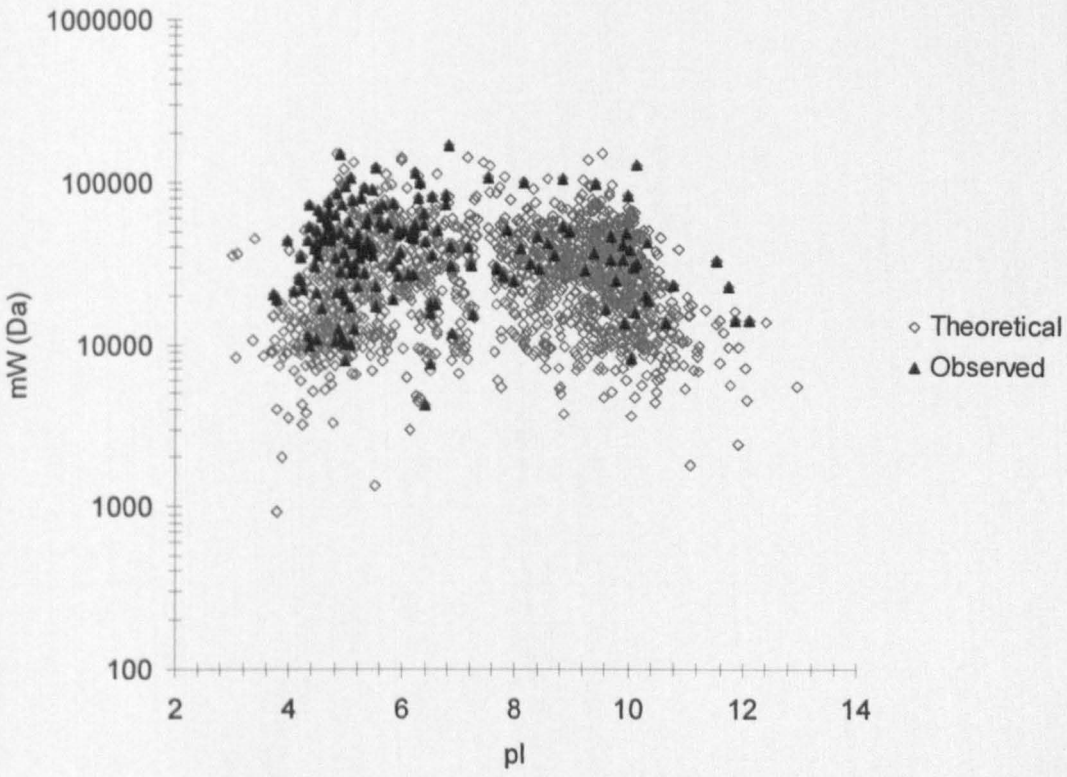


Figure 6.2: A protein map of all theoretical proteins in *P.marinus* MED4 and proteins observed in both iTRAQ experiments. Mw and pI values generated from JVirGel v.2.0 (http://www.jvirgel.de/).

PSORTb was used to predict the cellular location of identified proteins, and these were compared to the cellular locations of proteins identified in the cyanobacteria *Synechocystis* (in a previous study [67] and in Chapter 3) and *A. variabilis* ATCC 29413 [73] (Table 6.3). Despite differences in sample preparation (mainly the extraction buffers implemented), the cellular distribution of identified proteins is very similar. Cytoplasmic proteins dominate, with percentages ranging from 39.2% to 49.6% across the four studies. A higher percentage of cytoplasmic membrane proteins were observed in *P. marinus* MED4 (11.8%) compared to both *A. variabilis* (6.8%) and *Synechocystis* (2.6% and 7.14%).

Table 6.3: A comparison between the number and percentage of proteins identified in different cellular locations in *P. marinus* MED4, *A. variabilis* ATCC 29413, *Synechocystis* (two separate studies) and *Euhalothece*.

	P. marinus MED4		*A. variabilis*		*Synechocystis*				*Euhalothece*	
					Gan *et al.*		Chapter 3		Chapter 4	
Cellular location	**Observed**	**% of total**	**Observed**	**% of total**	**Observed**	**% of total**	**Observed**	**% of total**	**Observed**	**% of total**
Cytoplasmic	84	49.12	253	39.16	194	49.62	176	47.18	176	47.96
Cytoplasmic membrane	24	14.04	44	6.81	10	2.56	27	7.14	17	4.63
Extracellular	0	0.00	3	0.46	0	0.00	0	0.00	3	0.82
Outer membrane	3	1.75	6	0.93	1	0.26	6	1.59	6	1.63
Periplasmic	4	2.34	13	2.01	11	2.81	14	3.7	9	2.45
Unknown	68	39.77	327	50.62	175	44.76	155	41.01	171	46.59
Total	183		646		391		378		382	

Note:

[a]Cellular location was predicted using PSORTb v.2.0 software (http://www.psort.org/psortb/)

[b]Total number of proteins observed

[c]Percentage of total proteins observed

A limitation of PSORTb is the inability to detect membrane lipoproteins, and therefore LipoP v1.0 was implemented to achieve this. Twelve lipoproteins were predicted in total, all containing signal peptidase I. PSORTb designated three of these proteins as periplasmic proteins (PMM0090, PMM0710, PMM0970), one cytoplasmic membrane protein (PMM0226) and one outer membrane protein (PMM0708). The cellular location of the remaining seven predicted lipoproteins was unknown using PSORTb. This increased the overall membrane protein count in *P. marinus* MED4 from 28 to 35.

The hydrophobicity of proteins was calculated using the GRAVY index, and seven proteins with values above 0.3 were identified and regarded as hydrophobic (Appendix H). This is congruent to the five hydrophobic proteins predicted using the GRAVY index in *Synechocystis* [73]. Of these seven proteins, PSORTb predicted six to be cytoplasmic membrane proteins, and the highest index scoring protein (0.74) was a 30S ribosomal protein.

The protein abundance in both iTRAQ experiments was measured using emPAI [360]. In both iTRAQ1 and iTRAQ2, all emPAI values were below 1.0, which contrasts with 84.8% in *A. variabilis* ATCC 29413 and 91.3% in *Synechocystis* in Chapter 3 and 76.4% in a previous study [73]. The two most abundant proteins were PSI protein PsaE (0.36) and thioredoxin peroxidase (0.35) in iTRAQ1, and carboxysome shell protein (0.67) and PsbO (0.57) in iTRAQ2. Thioredoxins are involved in energy metabolism, and were also identified as highly abundant in *A. variabilis* ATCC 29413 [73]. It is apparent from this analysis using emPAI values that there are less abundant proteins in *P. marinus* MED4 compared to other cyanobacteria. This can be advantageous in detecting lower abundance proteins as, for example, there are no phycobilisome proteins in *Prochlorococcus,* and these have been singled out as a major interference in protein detection in cyanobacterial cells [73].

6.3.3 Technical variation and false positive rate

The variation between technical replicates was assessed by calculating the CV for both HL phenotypes against the LL phenotype (reference phenotype). The $\log_{10}$ ratios

were compared for these two sets of data using all proteins identified in both iTRAQ experiments. In iTRAQ1, $\log_{10}$ ratios 114:115 and 117:115 (Table 6.1) were compared for 147 proteins and the mean average CV was 0.11. For iTRAQ2, $\log_{10}$ ratios of 116:114 and 117:114 (Table 6.1) were compared for 95 proteins and gave a higher mean average CV of 0.18. In theory, the ratio of the technical replicates (HL phenotypes) in the same experiment should be equal to 1, and therefore the average ratio was calculated for all the proteins together, for both iTRAQ experiments. This was achieved by calculating the average of the $\log_{10}$ ratios, using the error factor as a weight, to obtain a weighted average. In iTRAQ1, the weighted average of the $\log_{10}$ ratio of the labels corresponding to both HL phenotypes was 0.01 or a ratio of 1.03. This illustrates that there is little technical variation in quantitation between the same HL phenotype samples. For experiment 2, the ratio of both HL phenotypes was similar at 0.91.

As discussed in Chapter 3, the analysis of iTRAQ data requires a measure of the extent of false positive protein identification. The same method was applied here [98] and found to be comparable at 0.5%.

6.3.4 Differentially regulated proteins

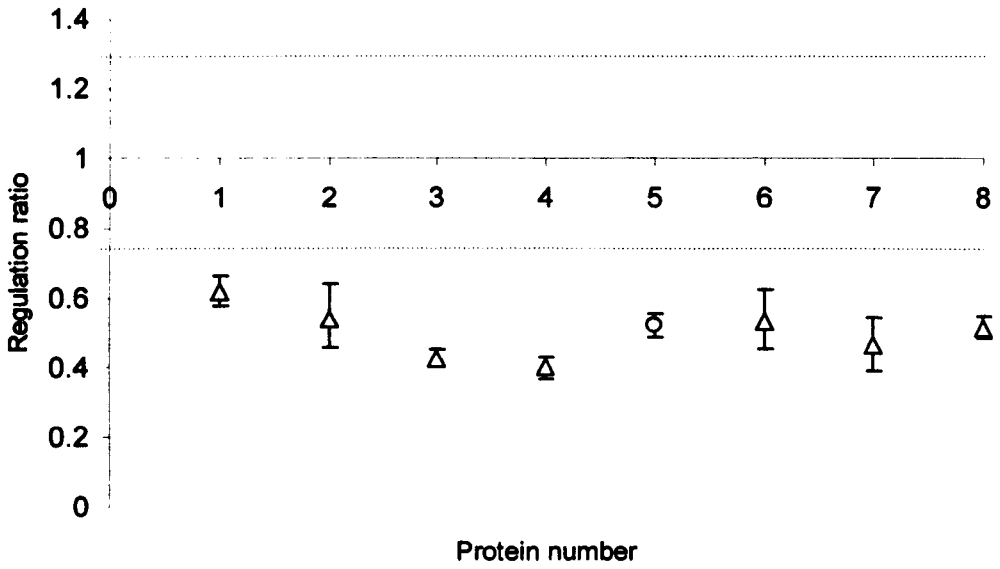
A total of 15 proteins were considered as significantly differentially expressed. These proteins were selected based on the p-value assigned for the quantitation, the magnitude of the error factor, and the RV (as well as the detection of more than 1 peptide, or the same peptide identified over more than one MS/MS experiment (as stated in Chapter 3, section 3.3.2). The ratio of ML to LL was compared between iTRAQ1 and 2 for all 53 proteins which were identified in both experiments, and because they represent biological replicates, the expected ratio was 1.0. The average RV was 1.05 (determined logarithmically), which implies there was a 5% discrepancy for relative quantitation. The RV was also calculated for all HL to LL combinations of the 53 shared proteins (114:115 in iTRAQ1 compared to 116:114 and 117:114 in iTRAQ2 and 117:115 in iTRAQ1 compared to 116:114 and 117:114 in iTRAQ2) and produced values of 2%, 7%, 2% and 3%, respectively. Calculations of CV's between biological replicates (across iTRAQ experiments, HL to LL and ML to LL) gave an average 0.21. However, to ensure that only proteins which express a genuine

differential response were selected for further analysis, only ratios of 1.25 and 0.75 (which fall clearly out of the RV boundaries) were deemed to be differentially regulated. Protein quantitations were calculated using modified equations from ProQuant Software v1.1 (AB, MDS-Sciex) [83].

6.3.4.1 Thylakoid complexes

Nine proteins were identified as belonging to PSI (Psa) PSII (Psb), including the sole light harvesting complex protein (Pcb). Eight proteins were treated as statistically significantly down-regulated in HL (Figure 6.3). There are several light harvesting complex genes in LL ecotypes of *Prochlorococcus,* and they are believed to be important for survival in extremely low irradiances by enabling cells to use the few available photons which reach the deeper ocean [282]. However, there is only a single copy of the *pcb* gene in HL adapted *P. marinus* MED4. This thesis data shows that this sole antennae protein present in MED4 is induced in LL (protein 5 in Figure 6.3). It is therefore likely that the abundance of one specific Pcb protein in MED4, as well as the multiple light harvesting complexes in LL adapted ecotypes, allows *Prochlorococcus* to adapt to LL. The multiple light harvesting complexes in LL adapted *Prochlorococcus* are more sensitive to photodamage by HL conditions present at the ocean surface, demonstrated at 100 μEinsteins m^{-2} s^{-1} light intensity in the laboratory [157], and it may be for this reason why MED4 harbours and increases the expression of only one Pcb protein. *PsbB, psbC* and *psbD* encode CP47, CP43 and D2 respectively, and transcript levels for all were found to be reduced 50 to 100% in *Synechocystis* after 15 hours of exposure to HL [121]. Despite the observed down regulation of several thylakoid located proteins, the PSII reaction centre core protein D1 (coupled to core protein D2), encoded by the *psbA* gene, has been shown to increase in response to HL in MED4 in the short-term, subsequently returning to normal levels [363]. Transcript levels of this gene were measured using RNA blots. However, the LL adapted *P. marinus* SS120 strain maintains high level *psbA* transcripts, and it is not certain whether this difference is an important factor in the ability of MED4, unlike SS120, to acclimate to rapidly changing irradiances. However, the presence of only one copy of the D1 protein in *Prochlorococcus* strains compared to several copies found in all other examined cyanobacteria [364, 365], is thought to contribute to the reduced repair efficiency of its PSII [366]. Unfortunately,

the D1 protein was not identified by either iTRAQ experiment. Despite the 2.2 and 1.82-fold reduction of the peptide AFINNLPFYR belonging to PSI protein PsaL, the regulation was not confirmed because only one peptide was identified in only one MS/MS experiment.



1-PsbO; 2-PsbB; 3-PsbD; 4-PsbC; 5-Pcb; 6-PsaA; 7-PsaE; 8-PsaB

Figure 6.3: Differential regulation of photosystem related proteins in high light (HL) compared to low light (LL). Proteins 1-4: Photosystem II (PsbO, PsbB, PsbD, PsbC). Protein 5: light harvesting complex protein. Proteins 6-8: Photosystem I (PsaA, PsaE, PsaB). Data is shown from the iTRAQ experiment in which the quantitation value has the lowest p-value (derived in ProGroup) and hence is the most likely to be correct (circle: iTRAQ1, triangle: iTRAQ2). The area between the dashed lines represents the RV boundary (between ratios 1.25 and 0.75). The regulation ratio and standard deviation are reported in linear space.

Figure 6.4 shows a schematic representation of thylakoid proteins with confirmed differential regulation in HL. Approximately 50% of the major proteins involved in photosynthetic activity were identified with high confidence. Overall, it appears that down-regulation of both PSI and PSII proteins in HL, is a response *P. marinus* adopts to reduce the susceptibility of cells to HL-induced damage.

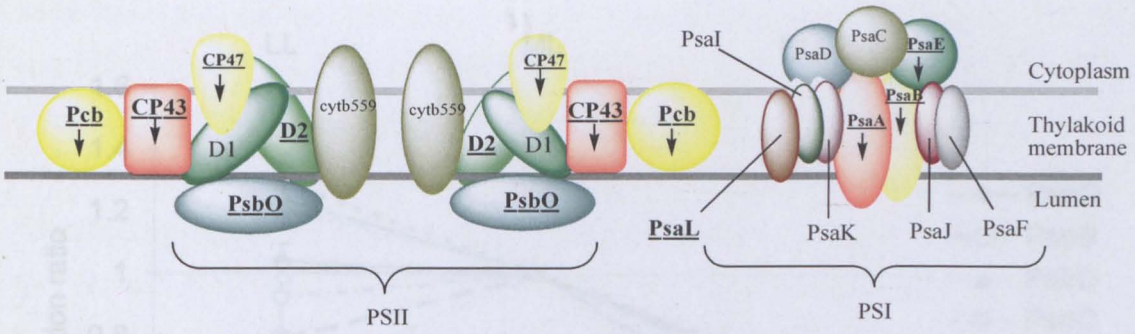


Figure 6.4: A schematic diagram of the major proteins located in the thylakoid membrane in MED4. Proteins in **bold** and underlined were identified in our study, and differential regulation in high light (HL) compared to low light (LL) is indicted by the arrow. The organisation is adapted from Partensky and Garczarek [259].

A ML phenotype was included in both iTRAQ experiments to help understand the light dependent regulation of proteins. Five of the photosynthetic proteins (PsbO, PsbC, PsbD, PsaE and PsaL) experienced very little differential regulation from LL to ML, and then decreased in expression in HL. PsbB, and the light harvesting protein (Pcb) increased in abundance in ML, then decreased in HL. This change indicates that ML conditions may encourage the light harvesting capacity of cells, rather than being directly detrimental. Both major reaction centre subunits in PSI, PsaA and PsaB, were reduced in expression in increasing light (from LL to ML to HL) as demonstrated previously [121]. The ratio of these proteins in HL and LL in relation to ML is represented in Figure 6.5.

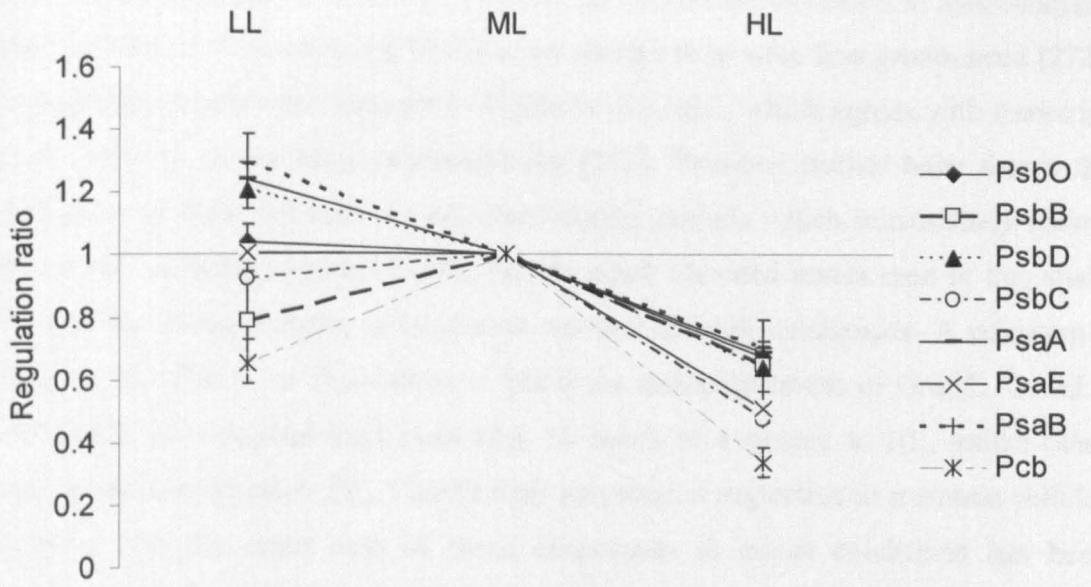


Figure 6.5: The relationship between medium light (ML) phenotype and high light (HL) and low light (LL) phenotypes for photosystem related proteins: Photosystem II (PsbO, PsbB, PsbD, PsbC), Photosystem I (PsaA, PsaE, PsaB) and light harvesting complex protein (Pcb). The value for ML is always 1 because the ratio of ML to ML is 1. Only the most probable (lowest p-value) quantitations are shown. The regulation ratio and standard deviation are reported in linear space.

6.3.4.2 General Stress Proteins

Stress associated proteins are vital in bacteria which suffer in adverse environmental conditions [367]. Proteins can often lose their function as polypeptides become misfolded or unfolded, and it is the job of stress related proteins to prevent these accumulating in the cell, and hence prevent cellular damage. Six stress related proteins were identified in *P. marinus* in this experiment, GroEL, GroEL2, DnaK2, HtpG, ClpC and GroES. GroEL, HtpG and GroES were deemed to be significantly induced in HL (Figure 6.6). GroEL had the highest induction (2.42-fold) in iTRAQ2, whereas GroEL2 had the smallest recorded marginal increase (1.36-fold), also in iTRAQ2. Transcript levels of both these proteins were both up-regulated in *P.marinus* MED4 in a study using RT-qPCR [272]. GroEL transcript levels rose steadily upon exposure to 200 μEinsteins m^{-2} s^{-1}, and peaked after 6 hours with an induction ratio of 10.11 ± 5.01 with respect to the LL culture [272]. GroEL2 transcript levels were comparable, however, levels continued to rise up to 12 hours after HL exposure. The

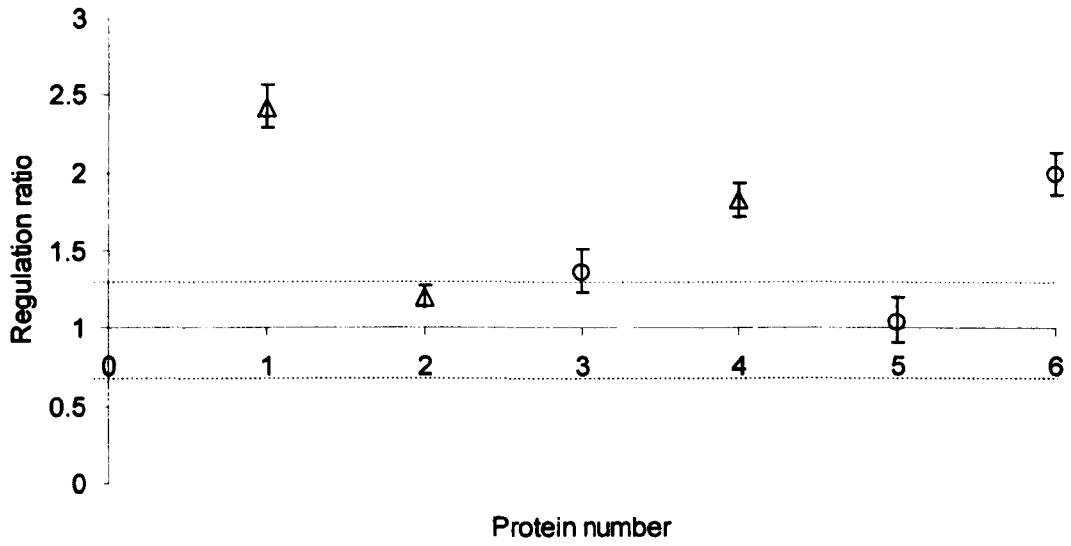
same study also showed transcript levels of these proteins increased in *Synechocystis* and LL adapted *P. marinus* MIT9313, even though they were less pronounced [272]. In this study, GroES was induced 1.99-fold in iTRAQ2, which agrees with transcript level induction in the same microorganism [272]. Previous studies have shown the abundance of these proteins are increased during periods which immediately follow the 'shock' of being exposed to HL, but the small elevated levels seen in this study are possibly a requirement of long-term survival at high irradiances. A microarray study on HL effects on *Synechocystis* found the transcript levels of GroEL, GroEL2 and GroES all remained high even after 15 hours of exposure to HL, unlike other stress associated genes [121]. Clearly their presence is important to maintain cellular function, and the exact role of these chaperones in stress conditions has been described previously in the cyanobacterium *Anabaena variabilis* [332].

Chaperone DnaK2 is a highly conserved protein amongst cyanobacteria [272], and it has been shown to play a significant role in adaptation to HL [121, 272]. However, the protein does not seem to be differentially expressed in either iTRAQ experiment (Figure 6.6 and Table 6.4), and hence the role of DnaK2 in long-term adaptation to HL is questionable. Interestingly, the regulation changes were slightly more evident when comparing ML to LL with increases of 1.58-fold and 1.32-fold in iTRAQ1 and 2, respectively (Table 6.4).

The heat shock protein HtpG increased 1.82-fold in HL, as were transcripts for this protein in *P. marinus* MED4 [272] and *Synechocystis* [121, 272]. This protein has been shown to play a very important role in cyanobacterial cell survival in thermal [368] and oxidative stress [369]. In the latter study, a *Synechococcus* sp. PCC7942 *htpG* mutant had a reduced survival rate in oxidative stress caused by high irradiance, and led to photoinhibition of PSII [369]. This study shows that HtpG is important for the survival of *P. marinus* MED4 in extreme HL conditions on the ocean surface.

ClpC belongs to a family which contains ATP-binding proteins which function as molecular chaperones, and regulate ATP-dependent proteolysis. They are thought to protect cells from stress by controlling aggregation and denaturation of proteins, and transcript levels have been shown to increase more than ten-fold in *P. marinus* MED4 upon exposure to HL, but not in LL ecotypes [272]. The constitutive level of *clpC* in

Synechococcus also increases considerably under conditions of rapid growth, for example, exposure to HL [370]. It is not certain whether protein levels were regulated between LL and HL cultures in this study. It was identified in iTRAQ 1, and the technical replicates produced ratios of 1.04 and 1.22, suggesting little or no role in photoadaptation (Figure 6.6).



1-GroEL; 2-DnaK2; 3-groEL2; 4-HtpG; 5-ClpC; 6-GroES

Figure 6.6: Regulation of stress related proteins in high light (HL) compared to low light (LL). Proteins 1-6: GroEL, DnaK2, GroEL2, HtpG, ClpC and GroES. Data is shown from iTRAQ experiment in which the quantitation value has the lowest p-value (derived in ProGroup) and hence is the most likely to be correct (circle: iTRAQ1, triangle: iTRAQ2). The area between the dashed lines represents the RV boundary (between ratios 1.25 and 0.75). The regulation ratio and standard deviation are reported in linear space.

Figure 6.7 shows the regulation ratios of stress associated proteins in HL and LL in relation to ML. Levels of GroEL and HtpG did not change significantly in ML, and this implies that unlike HL, ML does not initiate stress in *P. marinus* MED4 cells. There was no significant relative change in GroEL2 and ClpC across the three different light intensities, increasing the likelihood that these proteins are not required in photoadaptation in this cyanobacterium.

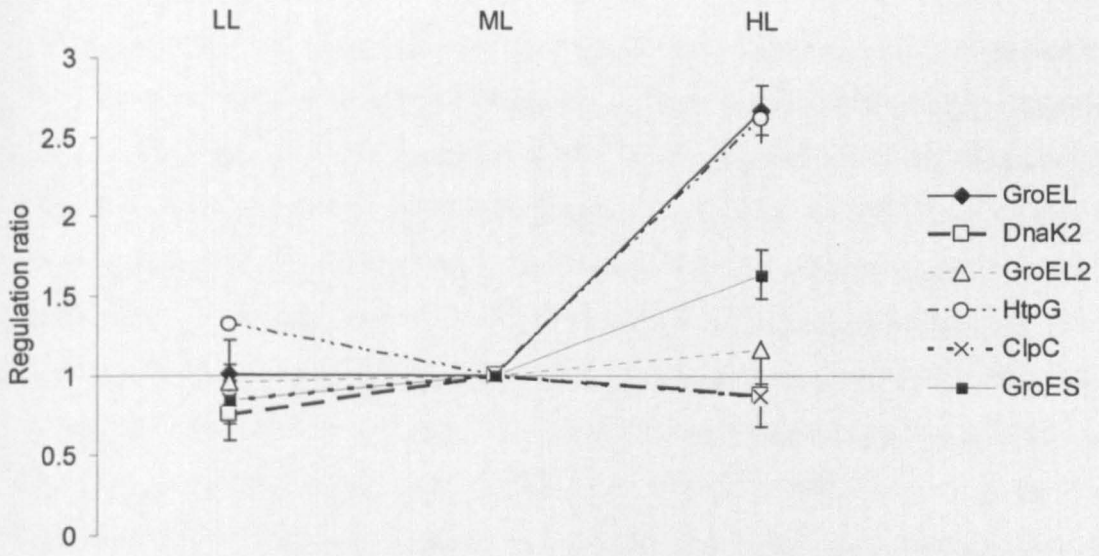


Figure 6.7: The relationship between ML (medium light) phenotype and HL (high light) and LL (low light) phenotypes for stress related proteins: GroEL protein (chaperonin cpn60), molecular chaperone DnaK2 (heat shock protein), GroEL2 protein (chaperonin cpn60 2), HtpG heat shock protein, molecular chaperone ClpC, and GroES protein (chaperonin cpn10). The value for ML is always 1 because the ratio of ML to ML is 1. Only the most probable (lowest p-value) quantitations are shown. The regulation ratio and standard deviation are reported in linear space.

6.3.4.3 Other differentially regulated proteins

Proteins not associated with the thylakoid membrane or stress-related proteins were also found to be altered by HL. The growth rate of *P. marinus* MED4 increased as the amount of light increased, and this is reflected by an increase in the FtsZ cell division protein. This protein initiates cell division by assembling into a cytokinetic ring on the cytoplasmic membrane where division occurs [371]. This protein was induced in HL in iTRAQ1 and iTRAQ2, and its expression also increased in ML for both experiments (Table 6.4). The increase of the putative septum site-determining protein MinD (PMM0321) in HL provides extra evidence for increased cell division, and hence growth rate. This protein ensures the septum is correctly placed during cell division, and was identified in iTRAQ2 and found to increase 2.03-fold.

The carboxysome is an inclusion body found in all cyanobacteria, and several other microorganisms [372]. It is important for creating an elevated local CO_2 concentration [373], and contains the enzyme ribulose 1,5 bisphosphate carboxylase/oxygenase (Rubisco), which converts inorganic carbon to an organic form. Carbon concentrating in the carboxysome is an energy consuming process, and so would be expected to be down-regulated in LL. The structure and components of carboxysomes are poorly understood, and it is not clear how *Prochlorococcus* even transports inorganic carbon [278], but there are three carboxysome shell proteins in *P. marinus* MED4, CsoS1, CsoS2 and CsoS3. CsoS2 was identified in both iTRAQ experiments, and CsoS1 was identified in iTRAQ2 only. The regulation of CsoS2 could not be calculated in iTRAQ1 because only one peptide in one MS experiment was detected, and the technical replicates in iTRAQ2 produced contrasting results. There was a similar case in iTRAQ2 for CsoS1, however the p-value of one technical replicate was zero and gave a 3.02-fold increase. In conclusion, the expression of these proteins remains uncertain. The expression of the enzyme Rubisco was not significantly different in the varying irradiances (Table 6.4) and therefore, it was concluded that carbon fixation rates did not alter. In *Synechocystis*, transcript levels for genes encoding components of the carboxysome including Rubisco, increased even after 15 hours in HL [121].

ABC transporter protein (PMM0710) is part of the ABC-type phosphate transport system, and this protein was 1.47 and 1.49-fold down-regulated in HL in iTRAQ1, and 3.57 and 2.94-fold down-regulated in HL in iTRAQ2. This transporter protein is encoded by *pstS*, and this gene is known to be induced under phosphate limiting conditions in many cyanobacteria, including *P. marinus* MED4 [374]. The non-induction of PhoH-like phosphate starvation-inducible protein (PMM1284) in iTRAQ1 implied that phosphate starvation was not a factor in any of the cultures. Therefore, how regulation of PMM0710 levels is related to irradiance is not presently understood, and further investigations need to be conducted to elucidate its role.

Carotenoids are photosynthetic pigments that play several important roles in the cyanobacterial cell. Not only do they trap light energy in the light harvesting apparatus, but they also dissipate excess light energy helping to protect the cell from radiation damage (Chapter 2). They also provide protection from free radicals formed in the cell from a variety of stress inducers. The major carotenoids in

Prochlorococcus are α-carotene, zeanthanin and a type of phycoerythrin [375, 376]. The enzyme phytoene desaturase in *Prochlorococcus* converts phytoene to ζ-carotene in the zeaxanthin biosynthetic pathway. This enzyme (PMM0144) was identified in iTRAQ1, and was slightly increased in HL, 1.35 and 1.49-fold. An increase in abundance of this enzyme suggests carotenoids play a role in adaptation to HL in *P. marinus* MED4, however, further investigations will have to be undertaken to confirm its specific function.

Table 6.4: Relative regulation ratios of additional important proteins

Protein name	Accession no.	Conditions	iTRAQ 1		iTRAQ 2	
			Regulation ratio	CV	Regulation ratio	CV
		HL/LL	1.34	0.11	1.06	0.17
		HL/LL	1.42	0.06	1.2	0.05
DnaK2	PMM1704	ML/LL	1.58	0.17	1.32	0.06
		HL/LL	1.04	0.19		
ClpC	PMM1088	HL/LL	1.22	0.15		
		HL/LL	1.68	0.19	1.67	0.2
		HL/LL	1.62	0.14	1.89	0.18
FtsZ	PMM1309	ML/LL	1.91	0.4	1.24	0.01
MinD	PMM0321	HL/LL			2.03	0.14
CsoS1	PMM0549	HL/LL			3.02	0.25
		HL/LL	0.98	0.16	1.18	0.04
Rubisco	PMM0550	HL/LL	0.89	0.13	0.99	0
ABC		HL/LL	0.68	0.09	0.28	0.1
transporter	PMM0710	HL/LL	0.67	0.11	0.34	0.04
Phytoene		HL/LL	1.35	0.08		
desaturase	PMM0144	HL/LL	1.49	0.03		

Note:
HL- High light at 100 μEinsteins m^2 s^{-1}
ML- Medium light at 60 μEinsteins m^2 s^{-1}
LL- Low light at 20 μEinsteins m^2 s^{-1}

6.4 Conclusions

In this chapter, the proteome of a globally significant photosynthetic bacterium was characterised for the first time. 184 proteins corresponding to approximately 11% of the complete theoretical proteome were identified with high confidence, representing 16 out of total 17 COG functional categories. Similar to previous studies, further

annotation was required for a large percentage of these proteins [67, 73]. 60% of the theoretical proteome in MED4 is basic, but in this study ~73% of proteins identified were acidic. This could be a result of the sample preparation, but further studies need to be undertaken. A higher percentage of membrane proteins identified in this study compared to previous studies [67, 73] (including chapter 3) may be a result of the urea-based extraction buffer implemented, which can solubilise less hydrophilic proteins. It is interesting to see that out of 184 proteins identified, only 53 were shared between the two experiments. In addition, emPAI value analysis reveals that relatively low abundance proteins were detected in the workflow. Therefore, without interference from highly abundant phycobilisome proteins, similar studies on the proteome of MED4 have the potential to increase overall proteome coverage.

In addition to proteome characterisation, a global proteomic perspective on the adaptability of *P. marinus* MED4 to the varying light intensities it faces in vast oceanic niches was sought. Increasing light intensity from 20 to 100 μEinsteins m^{-2} s^{-1} more than quadrupled the growth rate of *Prochlorococcus* cells from 0.13 ± 0.01 d^{-1} to 0.57 ± 0.05 d^{-1}. The average RV was calculated to be 5% from biological replicates across two separate iTRAQ experiments, and the mean CV was 0.15 and 0.21. For technical replicates, the mean CV was 0.11 and 0.15. This meant that 15 differentially expressed proteins were identified with a high degree of confidence (≥25% change).

In contrast to the proteomic studies in previous chapters, an even smaller significance threshold (≥25% change) was set to identify differentially expressed proteins. This value was generated using RV and CV calculations across two iTRAQ experiments. By conducting two separate iTRAQ experiments using biological replicate samples, and technical replicates within each experiment, sufficient statistical significance was gained to have confidence in this threshold.

The initial response of *P. marinus* MED4 cells to HL is to protect photosystem related proteins (ca. 50% were identified), including the Pcb light harvesting complex by reducing their expression. This potentially avoids the absorption of excess light energy. Differential regulation occurred in ML, and this could be due to the role of changing stoichiometry of PSI and PSII in different light conditions, which has

previously been described at the genetic level in the cyanobacterium *Synechocystis* [121]. The increased growth rate requires elevated amounts of cell division proteins, for example the average 1.72-fold increase in FtsZ cell division protein. The stress related chaperone GroEL exhibited the largest induction in HL (2.4-fold), demonstrating that despite the increased growth rate in HL, constant protection is required against the detrimental effects.

This study was an important addition to the thesis because it highlights the applicability of the methodologies developed in earlier chapters (particularly Chapter 3) to a different biological system. Its ability to photosynthesise and predicted high abundance means it plays a significant role in global nutrient cycling and climate control, both which warrant investigations of its environmental adaptive capabilities. Furthermore, *P. marinus* MED4 is a unique cyanobacterium which prior to this thesis work, has not had its proteome investigated, and by harbouring an extremely small genome (1.66 Mbp) it represents a promising candidate species in which a relatively high percentage of its proteome can be covered and quantitatively investigated in future.

Chapter 7: Conclusions and Future Work

In this thesis, proteomic techniques have been developed and applied to help answer microbial ecology questions, and thereby uncover the molecular adaptation mechanisms of three microorganisms in different environments. Proteomics was demonstrated to be a particularly useful tool to answer these questions because adaptation is a predominantly protein synthesis-dependant process. RT-qPCR experiments were also carried out for extra elucidation of acclimation strategies, but mainly to illustrate the significance of data generated at the protein level. The discoveries have added new insight into understanding the microbial responses to varying salt and light conditions, as well as uncovering a potentially unique halotolerant organism. In addition, the ability to reveal areas which require further research (Chapter 1, Figure 1.1) was demonstrated. In this chapter, the progress made in proteomic methodology in this thesis is summarised, and subsequently the adaptation discoveries made in the three different organisms is reviewed. Finally a discussion on future directions in this area is presented.

7.1 Proteomics methodology

7.1.1 Protein fractionation

Protein extraction and fractionation methods employed in this thesis aimed to provide an overview of predominantly the soluble proteome. Sample preparation and separation using the isobaric tagging procedure (SCX) was particularly biased towards hydrophilic proteins, although membrane and membrane-attached proteins were identified (Chapters 3 and 6). In Chapter 6, it was also observed that this technique was predisposed to the acidic proteome. The percentage of the proteome covered in both Chapters 3 and 6, was similar to previous studies (11.9 and 11%, respectively) [67, 73]. An alternative protein extraction buffer incorporating urea was used in Chapter 6, and resulted in a higher percentage of membrane protein identification. 2DE in Chapter 4, Part A, was limited to non-hydrophobic proteins with a pI range of 4 to 7 and therefore the method applied in Part B using 1D SDS-PAGE provided superior coverage. However, 2DE did provide superior fractionation (one to several proteins per spot) and was required to perform a fast and efficient overview of the proteomic response in *Euhalothece* in Chapter 4, Part A. To enable successful visualisation of proteins using 2DE, a low-cost and easy-to-apply salt-

removal strategy was developed and applied to the proteomic workflow (refer to section 1.4, 'Hypotheses, aims and scope' and Figure 4.7).

7.1.2 Differential protein expression threshold

The biological questions asked in this thesis consisted of revealing environmental adaptation mechanisms in cyanobacteria, and it has been shown that immediate changes to perturbations can be significantly different to changes which exist in the adaptive state, and importantly, changes may even be less pronounced [60, 61]. Therefore, quantitative techniques were required to identify important proteins which were differentially expressed at levels lower than the traditionally used 2-fold (100%) changes (refer to section 1.4, 'Hypotheses, aims and scope'). The iTRAQ technique was applied, and statistical analyses of protein quantitations were performed to calculate which proteins could be deemed to be differentially expressed with high confidence. The significance threshold was set by calculating the CV and RV of quantitations across biological replicate samples. In Chapter 3, one biological replicate sample was included in the experiment (115 and 116 isobaric labels), and using both workflows implemented gave an average CV of 20% and 32%. The RV's were 4% and 8.3%, and therefore a differential expression threshold of 50% was used. In Chapter 6, the biological replicate samples were run in two separate labelling experiments, and therefore values were calculated using only the proteins identified in both studies (n=53). CV's were lower at 15% and 21% and RV varied from 2-7%. Technical replicates were also run in each experiment in Chapter 6 and CV's were 11% and 18%. Technical variance was expected to be lower than biological variance [83]. This reduced variation in Chapter 6, as well as increased confidence in results due to incorporation of biological, technical and experimental (an identical iTRAQ experiment) replicates meant an even smaller threshold change was used in this chapter ($\geq$25%). Similar comparisons across biological replicate experiments were conducted in Chapter 4, where comparisons of differentially resolved, isotopic peptide intensities were used for quantitations. The average CV's across biological replicates were 9% and 10% in Parts A and B, respectively, with a RV of 8% in Part B.

7.1.3 Proteomics of an unsequenced environmental isolate

The homology-based proteomics approach in Chapter 4 successfully demonstrated that *in vivo* metabolic labelling together with *de novo* sequencing of MS data combined with conventional protein identification software was sufficient to identify proteins in unsequenced *Euhalothece*. The confirmation of identifications using the elemental constraint as well as protein characteristics based on fractionation methodology (Mw or pI) increased overall confidence (refer to section 1.4, 'Hypotheses, aims and scope').

7.1.4 A fast and efficient 2DE methodology

Although the studies in this thesis were directed towards achieving a global view of the proteome, the methodology applied in Chapter 4, Part A, demonstrated a fast and efficient way to gain an overview of an organism's proteome. Importantly, the quantitative accuracy was not renounced and metabolic labelling allows quantitation of proteins with 50% differential expression.

7.1.5 Transcriptomics

Studies in this thesis were carried out to show how information on protein abundance in cells can be interpreted into biological function. However, in addition to proteomics, RT-qPCR analysis was carried out in Chapter 3. Transcriptomics has been used extensively in the last decade to predict how much protein is present in certain conditions, however, previous reports [123-125] and the results presented here have shown that outcomes can conflict with proteomics data and therefore, measurements of proteins, the functional output of transcription and translation activity, is particularly useful.

7.2 Adaptation discoveries

The quantitative data generated in the large-scale experiments in the three organisms was interpreted by grouping differentially expressed proteins into functional groups, and therefore significant changes in the same direction could be assigned to specific networks. For example, the identification of a 1.80 and 1.61-fold increase in two

enzymes involved in murein synthesis in *Synechocystis* cells adapted to 6% salt (Chapter 3), increased confidence that cell wall biosynthesis was an important acclimation process. Due to improved quantitation accuracy and proteome coverage, novel discoveries were made, for example, the reduction in synthesis of anti-sigma B factor in *Synechocystis* cells adapted to 6% salt. Its mode of action has been described previously in *B.subtilis* [305]. The main discoveries are highlighted in this section in reference to the hypotheses described in Chapter 1, Section 1.4.

7.2.1 *Synechocystis* and adaptation to salt

Approximately 35% of identified proteins (378) changed significantly in expression in cells adapted to 6% salt compared to 0% salt, implying mass change in cell control is required at the protein level. To date, this level of salt stress has not been investigated in *Synechocystis* cells. Overall, in addition to producing general and salt specific stress proteins, *Synechocystis* cells require major diverting of carbon metabolism enzymes in 6% salt, and as reported previously [61], this is most likely to do with compatible solute (GG) synthesis. A general increase in energy demand is likely to be related to active mechanisms for eliminating ionic stress. A reduced amount of photosynthetic pigments and water-oxidising enzyme in PSII was seen, however, increased levels of the precursor to cytochrome b_6/f complex, apocytochrome f, implies electron flow from PSII to PSI is more active in high salt, and therefore formation of the transmembrane electrochemical proton gradient for energy synthesis is enhanced. From these results, variable stoichiometry in these complexes is evident. A general reduction in biosynthetic capacity seen through reduced expression of PPP enzymes was also predicted here. The relationship between iron usage and salt stress was further exemplified by reduced iron binding and transport proteins in 6% salt. Transcript level measurements matched protein fold differences (at 9 days) in half of the comparisons. The role of post-transcriptional control, predicted in previous studies [61, 297], was also postulated here, including new predictions, for example, the control of cell division.

7.2.2 *Euhalothece* and adaptation to salt

Euhalothece was an interesting organism to study, because it is able to survive in an environment with wide variances in salinity [149], and the mechanisms employed by these organisms are currently under-researched. Applying global proteomic tools led to new discoveries and some which require further research. In 3% salt, *Euhalothece* adopts similar mechanisms to *Synechocystis*, to enable survival in 6% salt, including enhanced synthesis of energy through processes such as glycolysis, and use of enzymes to produce compatible solutes. However, distinct to *Synechocystis*, increased generation of reducing power through PPP was revealed. In 6% salt, similarities with *Synechocystis* were again highlighted, with reduced pigment levels (chlorophyll *a*, phycocyanin), reduced iron uptake and reliance on increased GG synthesis. However, in contrast to its freshwater relative in 6% salt, energy synthesis appeared to decrease with reduced expression of ATPase and glyceraldehyde-3-phosphate dehydrogenase. When grown in the highest salt concentrations investigated (9% NaCl), similar functional groups changed in expression, with reduced pigments and energy synthesis. CO_2 uptake increased, which could be related to GG synthesis and elevated carotenoid production possibly to limit oxidative stress. Surprisingly, the role of regulatory protein LexA was implicated in the salt response in this microorganism, and a homolog for this protein in *Synechocystis* does not behave in the same way. This implies alternative regulation sites for this protein in *Euhalothece*. Perhaps most interesting, was the 'stress-like' response, which cells showed when grown in 0% salt compared to 3% salt. These conditions led to increased production of molecular chaperones, and indicate that the resulting intracellular conditions at 0% salt are not favourable for some/all *Euhalothece* proteins. A brief survey of the *de novo* sequenced peptides in this organism compared to less halotolerant cyanobacteria, indicates the possibility of an acidically biased proteome, which in turn enables survival in high salt environments using alternative mechanisms. The increase in energy synthesis in *Synechocystis* cells in 6% salt is thought to drive the active extrusion of salt ions and this energy requirement is not reciprocated in *Euhalothece* cells in 6% and 9% salt, and may further imply a different intracellular environment to its freshwater relative.

7.2.3 *Synechocystis* vs. *Euhalothece* (adaptation to salt)

These cyanobacteria share many common tryptic peptides as discovered in Chapters 4 and 5. Protein abundances were compared in cells grown in the same conditions, 0, 3 and 6% w/v salt. Although comparisons of the responses to 6% salt could be derived from Chapters 3 and 4, using this methodology with the exact same saline conditions meant significant variances would be due to evolutionary divergence, and therefore allowing a direct comparison of adaptive mechanisms. As expected, *Euhalothece* accumulated stress-related chaperone proteins in higher abundance in 0% salt. In 3% and 6% salt, the opposite was true, and it appears that the enhanced accumulation of these proteins in *Synechocystis* means this organism is more stressed than *Euhalothece*. The greater importance of energy synthesis in *Synechocystis* cells in 6% salt, was also observed.

7.2.4 *Prochlorococcus* and adaptation to high light

Strain MED4 is the smallest known photosynthetic prokaryote, and resides in a particularly nutrient deprived environment. Therefore it is not unexpected that subtle changes are the preferential response this organism performs to adapt to varying light intensities. Only ~8% of the identified proteome in MED4 changed significantly in expression in HL compared to LL. It adapts to a shift to HL by reducing expression of photosystems-related proteins, and this is presumably to reduce absorption of excess energy. The increase in growth rate illustrates that surplus energy is being produced and consumed, despite the attempt to reduce overall energy capture. The dissipation of energy may occur through increased carotenoid production. Even though cells eventually adapt to the HL conditions, they continue to produce elevated amounts of molecular chaperones to protect against damage caused by excess light.

7.3 Future directions

7.3.1 Cellular mechanisms for further research

A major aim in this study was to identify proteins which change significantly in expression between 'normal' and adverse conditions, and interpreting this information by categorising protein biological properties in the context of the perturbation

(summarised in 7.2.1 to 7.2.4). However, another aim of this study was to demonstrate the potential of applying proteomic techniques to environmental organisms in a traditional workflow (Chapter 1, Figure 1.1), and thereby highlighting areas for further research.

Although many differentially expressed proteins were identified as playing important roles, many changes could not be interpreted into biological adaptation mechanisms and therefore require further investigations. For example, hypothetical protein sll1863 increased over 10-fold in the 6% salt acclimated *Synechocystis* cells, but its function remains unknown. Differentially expressed hypothetical proteins highlight the capability of global proteomics as a heuristic method to identify important cellular components, which consequently require additional exploration. An intriguing example is the regulatory role of the LexA protein. This is currently being intensely investigated due to its recent association with controlling the bi-directional hydrogenase gene in *Synechocystis* [329, 330]. It has also been implicated as an essential protein for *Synechocystis* cells facing inorganic carbon starvation [328]. However, its reduced expression in high salt in *Synechocystis* cells is contrary to its increased expression in salt acclimated *Euhalothece* cells, and although sharing 86% nucleotide sequence similarity, a clear difference in functional role is apparent. It would be useful to look for binding sites for this protein in the *Euhalothece* genome, as well as investigating the phenotypic response of *lexA* mutant cells in this organism.

In Chapter 4, *Euhalothece* cells required elevated amounts of molecular chaperones in no salt, and the reasons for this remain unknown. Initial investigations using *de novo* sequencing data imply this may be due to a salt-requiring intracellular system, but further investigations are required. For example, metabolite measurements (compatible solutes) will reveal how reliant this organism is on restoring its osmotic balance, and what its actual osmotic balance is. Information on the relationship between internal and external solute concentrations, and on the state of cell-associated ions within the cytoplasm would also be useful.

7.3.2 Proteome coverage

The methodologies applied here were validated for quantitative accuracy and found to be consistent and reliable. Although comparable to similar studies in cyanobacteria, only 11 and 11.9% of the proteome in *Synechocystis* and *Prochlorococcus* cells were identified, respectively. Improvements in protein fractionation methodologies have improved coverage [44, 67, 377], and the ability to combine multi-level protein and peptide separation techniques to an iTRAQ or metabolic labelling experiment has the potential to reveal more detail on a cells adaptive state. iTRAQ relies on solubilising proteins in TEAB buffer, but if alternative chemicals are used [43, 73], which are confirmed not to interfere with downstream techniques and be adaptable with the peptide tagging technology, increasing protein coverage through analysis of presently insoluble proteins can be achieved. The studies conducted in this thesis were particularly biased towards the cytoplasmic soluble proteome and therefore employing unique fractionation and solubilisation techniques means insights into the response in different cellular compartments, including membranes and even extra-cellular proteins can be revealed.

7.3.3 Post-translational modifications (PTM's)

It is important to note that protein abundance changes can only be used as an indicator of cellular function, and that alternative factors may be involved. In particular, PTM's can determine whether a protein is active, its cellular location, turnover rate and interactions with other proteins and nucleic acids [18]. For example, the control of anti-sigma B factor schematically presented in Chapter 3, Figure 3.3, requires kinase cascades which are either active or inactive depending on the addition or removal of a phosphate group. Therefore, mapping which proteins are modified requires isolation and characterisation, and is achievable using MS [18]. Incorporating this information with protein abundance data will reveal further insight into how complex control mechanisms aid environmental adaptation.

7.3.4 Bioinformatics recommendations

When dealing with large datasets of MS/MS generated data in high-throughput techniques such as iTRAQ, equally high-throughput analysis software must be

implemented to carry out protein identifications and quantitations, and three different types, ProteinPilot™, Spectrum Mill and ProQuant software, were used in this thesis. Using different quantitative algorithms will lead to varied interpretation, for example, in Chapter 3, using two different software algorithms would have contributed to experimental variability, although overall correlation was good (Pearson correlation, r =0.74). With the increasing number of algorithms used to quantify large data sets, it is important that the methods used to calculate the final stated relative ratio is transparent to users. Not only is this essential to verify quantitation manually, but also to allow integration of multiple data sets, or simpler comparisons of inter-laboratory data [309].

All protein identifications must be verified and statistically unacceptable candidates (low score, one peptide) must be discarded. Manual inspection of quantitations including reporter ion intensity in the MS/MS scans are also performed to reveal any erroneous calculations, which often lead to unreliable results (with large CV's). Improvements would include easier interpretation of results files, and therefore reduce the time taken to ensure correct data mining and verification. Recently, the Paragon™ algorithm implemented in ProteinPilot™ has attempted to remove the variability which can affect data processing, by introducing compulsory probability algorithms [290]. This way, less control is in the users' hands.

In Chapter 4, quantitation of peptides required identifying high intensity, co-eluting isotopic peptides, and subsequent quantitation involved integrating peptide peak areas in TOF-MS scans over time and manually calculating peptide abundance. Together this is a very time consuming process. The recently released open source software, MSQuant, was applied in Chapter 5 and performs protein quantitation on isotopically labelled peptides. However, it still requires verification for each peptide pair and to date, contains a plethora of software glitches. In addition, this tool cannot be applied when proteins were identified in *Euhalothece* using *de novo* sequencing software combined with BLAST searches (MS BLAST). Therefore high throughput software which can successfully identify and quantify proteins from unsequenced organisms would be a major advancement for cross-species proteomics.

7.3.5 Systems biology

In systems biology, a systematic investigation of all cellular components and their interactions is undertaken to gain a global understanding of physiology [21, 22]. Unravelling a systems complexity requires the integration of various data sets pertaining to the fields of transcriptomics, metabolomics and proteomics. The quantitative information on metabolites (metabolic flux analysis), mRNA and proteins, can be incorporated into models with the aid of bioinformatics tools, and thereby allow predictions of cellular responses to environmental conditions [22]. The proteomics data generated in this thesis has successfully presented a global perspective on how cells respond to perturbations in the environment, but ultimately they form an invaluable contribution to the bigger picture, where this technology would have to be complemented with associated information for a holistic view. Ultimately, the modelling of salt tolerance strategies can be applied to more complex systems, like higher plants. By modifying cellular processes, salt-susceptible plant species can have higher tolerance levels. Similarly, modelling the MED4 system will have profound implications for predicting (and modelling) climate change. Present-day changes in environmental conditions due to climate change, for example, increasing oceanic temperatures and acidity, will have an impact on *Prochlorococcus* abundance, distribution and behaviour. Elucidating these alterations will produce significant contributions in predicting future climate changes, and could even provide a method for reducing atmospheric carbon dioxide levels through generating marine carbon sinks.

References

1. Pikuta EV, Hoover RB, Tang J: **Microbial extremophiles at the limits of life.** *Crit Rev Microbiol* 2007, **33**:183-209.
2. Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ: **Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*.** *The Plant Journal* 2005, **44**:826-839.
3. Kanesaki Y, Suzuki I, Allakhverdiev SI, Mikami K, Murata N: **Salt Stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis* sp. PCC 6803.** *Biochem Biophys Res Commun* 2002, **290**:339-348.
4. Murata N, Suzuki I: **Exploitation of genomic sequences in a systematic analysis to access how cyanobacteria sense environmental stress.** *J Exp Bot* 2006, **57**:235-247.
5. Wilkins M, Sanchez J, Gooley A, Appel R, Humphery-Smith I, Hochstrasser D, Williams K: **Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it.** *Biotechnol Genet Eng Rev* 1995, **13**:19-50.
6. Abee T, Wouters JA: **Microbial stress response in minimal processing.** *Int J Food Microbiol* 1999, **50**:65-91.
7. Rhoades JD, Loveday J: *Salinity in irrigated agriculture.* American Society of Agronomists; 1990.
8. Hart BT, Lake PS, Webb JA, Grace MR: **Ecological risk to aquatic systems from salinity increases.** *Aust J Bot* 2003, **51**:689-702.
9. James KR, Cant B, Ryan T: **Responses of fresh water biota to rising salinity levels and implications for saline water management: A review.** *Aust J Bot* 2003, **51**:703-713.
10. Gao X, Ren Z, Zhao Y, Zhang H: **Overexpression of SOD2 Increases Salt Tolerance of Arabidopsis.** *Plant Physiol* 2003, **133**:1873-1881.
11. Parida AK, Das AB: **Salt tolerance and salinity effects on plants: a review.** *Ecotoxicol Environ Saf* 2005, **60**:324-349.
12. Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu JK, Shinozaki K: **Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray.** *Plant Physiol* 2004, **135**:1697-1709.
13. Serrano R, Rodriguez-Navarro A: **Ion homeostasis during salt stress in plants.** *Curr Opin Cell Biol* 2001, **13**:399-404.
14. Halfter U, Ishitani M, Zhu JK: **The *Arabidopsis* SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3.** *Proc Natl Acad Sci U S A* 2000, **97**:3735-3740.
15. Chisholm SW, Olson RJ, Zettler E, Goericke R, Waterbury JR, Welschmeyer NA: **A novel free-living prochlorophyte abundant in the oceanic euphotic zone.** *Nature* 1988, **334**:340-343.
16. Partensky F, Hess WR, Vaulot D: ***Prochlorococcus*, a marine photosynthetic prokaryote of global significance.** *Microbiol Mol Biol Rev* 1999, **63**:106-127.

17. Bernal A, Ear U: **Genomes online database (GOLD): a monitor of genome projects worldwide.** *Nucleic Acids Res* 2001, **29:**126–127.
18. Mann M, Jensen ON: **Proteomic analysis of post-translational modifications.** *Nat Biotechnol* 2003, **21:**255-261.
19. Anderson N, Anderson N: **Proteome and proteomics: new technologies, new concepts, and new words.** *Electrophoresis* 1998, **19:**1853-1861.
20. Monteoliva L, Albar JP: **Differential proteomics: an overview of gel and non-gel based approaches.** *Brief Funct Genomic Proteomic* 2004, **3:**220-239.
21. Burja AM, Dhamwichukorn S, Wright PC: **Cyanobacterial postgenomic research and systems biology.** *Trends Biotechnol* 2003, **21:**504-511.
22. Kitano H: **Systems biology: a brief overview.** *Science* 2000, **295:**1662–1664.
23. Taylor TN, Taylor EL: *The Biology and Evolution of Fossil Plants.* New Jersey, New York: Prentice Hall; 1993.
24. Sanger F, Coulson AR: **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.** *J Mol Biol* 1975, **94:**441-448.
25. Schena M, Shalon D, Davis R, Brown P: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270:**467-470.
26. Burnette WN: **"Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A.** *Anal Biochem* 1981, **112:**195-203.
27. Pandey A, Mann M: **Proteomics to study genes and genomes.** *Nature* 2000, **405:**837-846.
28. Aebersold R, Mann M: **Mass spectrometry-based proteomics.** *Nature* 2003, **422:**198-207.
29. Blueggel M, Chamrad D, Meyer HE: **Bioinformatics in proteomics.** *Curr Pharm Biotechnol* 2004, **5:**79-88.
30. Nie L, Wu G, Zhang W: **Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: A multiple regression to identify sources of variations.** *Biochem Biophys Res Commun* 2006, **339:**603-610.
31. Wand AJ: **Dynamic activation of protein function: a view emerging from NMR spectroscopy.** *Nat Struct Biol* 2001, **8:**926-931.
32. Kersey P, Apweiler R: **Linking publication, gene and protein data.** *Nat Cell Biol* 2006, **8:**1183-1189.
33. Pappin DJC, Hojrup P, Bleasby AJ: **Rapid identification of proteins by peptide-mass fingerprinting.** *Curr Biol* 1993, **3:**327-332.
34. Chalmers MJ, Gaskell SJ: **Advances in mass spectrometry for proteome analysis.** *Curr Opin Biotechnol* 2000, **11:**384-390.
35. O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins.** *J Biol Chem* 1975, **250:**4007-4021.
36. Anderson NG, Anderson NL: **Twenty years of two-dimensional electrophoresis: past, present and future.** *Electrophoresis* 1996, **17:**443-453.
37. Righetti PG: *Immobilized pH Gradients: Theory and Methodology,.* Amsterdam.: Elsevier; 1990.
38. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D: **The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences.** *Electrophoresis* 1993, **14:**1023-1031.

39. Simon WJ, Hall JJ, Suzuki I, Murata N, Slabas AR: **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* 2002, **2:**1735-1742.
40. Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W: **Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications.** *J Biochem Biophys Methods* 1982, **6:**317-339.
41. Jahn O, Hesse D, Reinelt M, Kratzin HD: **Technical innovations for the automated identification of gel-separated proteins by MALDI-TOF mass spectrometry.** *Anal Bioanal Chem* 2006, **386:**92-103.
42. Lopez MF: **Better approaches to finding the needle in a haystack: Optimizing proteome analysis through automation.** *Electrophoresis* 2000, **21:**1082-1093.
43. Barry RC, Alsaker BL, Robison-Cox JF, Dratz EA: **Quantitative evaluation of sample application methods for semipreparative separations of basic proteins by two-dimensional gel electrophoresis.** *Electrophoresis* 2003, **24:**3390-3404.
44. Gao M, Hong J, Yang P, Zhang X: **Chromatographic prefractionation prior to two-dimensional electrophoresis and mass spectrometry identifies: Application to the complex proteome analysis in rat liver.** *Anal Chim Acta* 2005, **553:**83-92.
45. Link AJ: **Multi-dimensional peptide separtations in proteomics.** *Trends Biotechnol* 2002, **20:**S8- S13.
46. Molloy MP: **Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients.** *Anal Biochem* 2000, **280:**1-10.
47. Rabilloud T: **Solubilization of proteins in 2-D electrophoresis: An outline.** *Methods Mol Biol* 1999, **112:**9-19.
48. Packer NH, Pawlak A, Kett WC, Gooley AA, Redmond JW, Williams KL: **Proteome analysis of glycoforms: a review of strategies for the microcharacterisation of glycoproteins separated by two-dimensional polyacrylamide gel electrophoresis.** *Electrophoresis* 1997, **18:**452-460.
49. Robertson ES, Nicholson AW: **Phosphorylation of *Escherichia coli* translation initiation factors by the bacteriophage T7 protein kinase.** *Biochemistry (Mosc)* 1992, **31:**4822-4827.
50. Berggren K, Steinberg TH, Lauber WM, Carroll JA, Lopez MF, Chernokalskaya E, Zieske L, Diwu Z, Haugland RP, Patton WF: **A luminescent ruthenium complex for ultrasensitive detection of proteins immobilized on membrane supports.** *Anal Biochem* 1999, **276:**129-143.
51. Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF: **A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling.** *Electrophoresis* 2000, **21:**3673-3683.
52. Shevchenko A, Wilm M, Vorm O, Mann M: **Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.** *Anal Chem* 1996, **68:**850-858.
53. Unlu M, Morgan ME, Minden JS: **Difference gel electrophoresis: a single gel method for detecting changes in protein extracts.** *Electrophoresis* 1997, **18:**2071-2077.

54. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.** *Mol Cell Proteomics* 2002, **1**:376-386.
55. Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, Plasterk RH, Heck AJ: **Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics.** *Nat Biotechnol* 2003, **21**:927-931.
56. Santoni V, Molloy M, Rabilloud T: **Membrane proteins and proteomics: un amour impossible?** *Electrophoresis* 2000, **21**:1054-1070.
57. Santoni V, Rabilloud T, Doumas P, Rouquie D, Mansion M, Kieffer S, Garin J, Rossignol M: **Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels.** *Electrophoresis* 1999, **20**:705-711.
58. Lopez JL: **Two-dimensional electrophoresis in proteome expression analysis.** *J Chromatogr B* 2007, **849**:190-202.
59. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.** *Proc Natl Acad Sci U S A* 2000, **97**:9390-9395.
60. Marin K, Kanesaki Y, Los DA, Murata N, Suzuki I, Hagemann M: **Gene expression profiling reflects physiological processes in salt acclimation of *Synechocystis* sp. strain PCC 6803.** *Plant Physiol* 2004, **136**:3290-3300.
61. Fulda S, Mikkat S, Huang F, Huckauf J, Marin K, Norling B, Hagemann M: **Proteome analysis of salt stress response in the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *Proteomics* 2006, **6**:2733-2745.
62. Opiteck GJ, Jorgenson JW: **Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides.** *Anal Chem* 1997, **69**:2283-2291.
63. Swanson RV, Glazer AN: **Separation of phycobiliprotein subunits by reverse-phase high-pressure liquid chromatography.** *Anal Biochem* 1990, **188**:295-299.
64. Kashino Y: **Separation methods in the analysis of protein membrane complexes.** *J Chromatogr B Analyt Technol Biomed Life Sci* 2003, **797**:191-216.
65. Kato Y, Nakamura K, Kitamura T, Tsuda T, Hasegawa M, Sasaki H: **Effect of chromatographic conditions on resolution in high-performance ion-exchange chromatography of proteins on macroporous anion-exchange resin.** *J Chromatogr A* 2004, **1031**:101-105.
66. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP: **Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome.** *J Proteome Res* 2003, **2**:43-50.
67. Gan CS, Reardon KF, Wright PC: **Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of *Synechocystis* sp. PCC 6803.** *Proteomics* 2005, **5**:2468-2478.
68. Wolters DA, Washburn MP, Yates JR, 3rd: **An automated multidimensional protein identification technology for shotgun proteomics.** *Anal Chem* 2001, **73**:5683-5690.
69. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR, 3rd: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* 1999, **17**:676-682.

70. Davidsson P, Puchades M, Blennow K: **Identification of synaptic vesicle, pre- and postsynaptic proteins in human cerebrospinal fluid using liquid-phase isoelectric focusing.** *Electrophoresis* 1999, **20:**431-437.
71. Kuhn R, Hoffstetter-Kuhn S: *Capillary electrophoresis: principle and practice.* Berlin, Heidelberg: Springer-Verlag; 1993.
72. Rabilloud T: **Membrane proteins ride shotgun.** *Nat Biotechnol* 2003, **21:**508-510.
73. Barrios-Llerena ME, Chong PK, Gan CS, Snijders AP, Reardon KF, Wright PC: **Shotgun proteomics of cyanobacteria--applications of experimental and data-mining techniques.** *Brief Funct Genomic Proteomic* 2006, **5:**121-132.
74. Beynon RJ, Pratt JM: **Metabolic labeling of proteins for proteomics.** *Mol Cell Proteomics* 2005, **4:**857-872.
75. Goshe MB, Smith RD: **Stable isotope-coded proteomic mass spectrometry.** *Curr Opin Biotechnol* 2003, **14:**101-109.
76. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C: **Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus.** *Anal Chem* 2001, **73:**2836-2842.
77. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17:**994-999.
78. Moseley MA: **Current trends in differential expression proteomics: isotopically coded tags.** *Trends Biotechnol* 2001, **19:**S10-16.
79. Cagney G, Emili A: **Mass-coded abundance tagging for protein identification and relative abundance determination in proteomic experiments.** In *The Proteomics Protocols Handbook.* 2005: 193-206
80. Kirkpatrick DS, Gerber SA, Gygi SP: **The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications.** *Methods* 2005, **35:**265-273.
81. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al: **Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents.** *Mol Cell Proteomics* 2004, **3:**1154-1169.
82. Chong PK, Gan CS, Pham TK, Wright PC: **Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections.** *J Proteome Res* 2006, **5:**1232-1240.
83. Gan CS, Chong PK, Pham TK, Wright PC: **Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ).** *J Proteome Res* 2007, **6:**821-827.
84. Stensjo K, Ow SY, Barrios-Llerena ME, Lindblad P, Wright PC: **An iTRAQ-based quantitative analysis to elaborate the proteomic response of Nostoc sp. PCC 7120 under N2 fixing conditions.** *J Proteome Res* 2007, **6:**621-635.
85. Ow SY, Cardona T, Taton A, Magnuson A, Lindblad P, Stensjo K, Wright PC: **Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-Plex isobaric peptide tags.** *J Proteome Res* 2008.
86. Choe LH, Franck KAZ, Lee KH: **A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells.** *Electrophoresis* 2005, **26:**2437-2449.

87. Wu WW, Wang G, Baek SJ, Shen RF: **Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D Gel- or LC-MALDI TOF/TOF.** *J of Prot Res* 2006, **5**:651-658.
88. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM: **Electrospray ionization for mass spectrometry of large biomolecules.** *Science* 1989, **246**:64-71.
89. Karas M, Hillenkamp F: **Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.** *Anal Chem* 1988, **60**:2299-2301.
90. Chernushevich IV, Loboda AV, Thomson BA: **An introduction to quadrupole-time-of-flight mass spectrometry.** *J Mass Spectrom* 2001, **36**:849-865.
91. Elias JE, Haas W, Faherty BK, Gygi SP: **Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations.** *Nat Methods* 2005, **2**:667-675.
92. MacCoss MJ, Wu CC, Yates JR, 3rd: **Probability-based validation of protein identifications using a modified SEQUEST algorithm.** *Anal Chem* 2002, **74**:5593-5599.
93. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *Electrophoresis* 1999, **20**:3551-3567.
94. Roepstorff P, Fohlman J: **Proposal for a common nomenclature for sequence ions in mass spectra of peptides.** *Biomed Mass Spectrom* 1984, **11**:601.
95. Biemann K: **Contributions of mass spectrometry to peptide and protein structure.** *Biomed Environ Mass Spectrom* 1988, **16**:99-111.
96. Chen Y, Kwon SW, Kim SC, Zhao Y: **Integrated approach for manual evaluation of peptides identified by searching protein sequence databases with tandem mass spectra.** *J Proteome Res* 2005, **4**:998-1005.
97. Washburn MP, Wolters D, Yates JR, 3rd: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nat Biotechnol* 2001, **19**:242-247.
98. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.** *Nat Methods* 2007, **4**:207-214.
99. Moore RE, Young MK, Lee TD: **Qscore: an algorithm for evaluating SEQUEST database search results.** *J Am Soc Mass Spectrom* 2002, **13**:378-386.
100. Keller A, Nesvizhskii AI, Kolker E, Aebersold R: **Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.** *Anal Chem* 2002, **74**:5383-5392.
101. Liska AJ, Shevchenko A: **Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications.** *Proteomics* 2003, **3**:19-28.
102. Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, Standing KG: **Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching.** *Anal Chem* 2001, **73**:1917-1926.
103. Grossmann J, Fischer B, Baerenfaller K, Owiti J, Buhmann JM, Gruissem W, Baginsky S: **A workflow to increase the detection rate of proteins from**

unsequenced organisms in high-throughput proteomics experiments. *Proteomics* 2007, 7:4245-4254.

104. Grossmann J, Roos FF, Cieliebak M, Liptak Z, Mathis LK, Muller M, Gruissem W, Baginsky S: **AUDENS: a tool for automated peptide de novo sequencing.** *J Proteome Res* 2005, **4**:1768-1774.
105. Habermann B, Oegema J, Sunyaev S, Shevchenko A: **The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches.** *Mol Cell Proteomics* 2004, **3**:238-249.
106. Samyn B, Sergeant K, Memmi S, Debyser G, Devreese B, Van Beeumen J: **MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from Halorhodospira halophila a bacterium with unsequenced genome.** *Electrophoresis* 2006, **27**:2702-2711.
107. Lester PJ, Hubbard SJ: **Comparative bioinformatic analysis of complete proteomes and protein parameters for cross-species identification in proteomics.** *Proteomics* 2002, **2**:1392-1405.
108. Liska AJ, Shevchenko A: **Combining mass spectrometry with database interrogation strategies in proteomics.** *TrAC, Trends Anal Chem* 2003, **22**:291-298.
109. Han Y, Ma B, Zhang K: **SPIDER: software for protein identification from sequence tags with de novo sequencing error.** *Proc IEEE Comput Syst Bioinform Conf* 2004:206-215.
110. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G: **PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.** *Rapid Commun Mass Spectrom* 2003, **17**:2337-2342.
111. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J: **OLAV: towards high-throughput tandem mass spectrometry data identification.** *Proteomics* 2003, **3**:1454-1463.
112. Clauser KR, Baker P, Burlingame AL: **Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching.** *Anal Chem* 1999, **71**:2871-2882.
113. Mackey AJ, Haystead TA, Pearson WR: **Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences.** *Mol Cell Proteomics* 2002, **1**:139-147.
114. Fischer B, Roth V, Roos F, Grossmann J, Baginsky S, Widmayer P, Gruissem W, Buhmann JM: **NovoHMM: a hidden Markov model for de novo peptide sequencing.** *Anal Chem* 2005, 77:7265-7273.
115. Frank A, Pevzner P: **PepNovo: de novo peptide sequencing via probabilistic network modeling.** *Anal Chem* 2005, **77**:964-973.
116. Dongre AR, Somogyi A, Wysocki VH: **Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides.** *J Mass Spectrom* 1996, **31**:339-350.
117. Keough T, Lacey MP, Youngquist RS: **Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry.** *Rapid Commun Mass Spectrom* 2000, **14**:2348-2356.
118. Sunyaev S, Liska AJ, Golod A, Shevchenko A, Shevchenko A: **MultiTag: Multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry.** *Anal Chem* 2003, **75**:1307-1315.

119. Marin K, Suzuki I, Yamaguchi K, Ribbeck K, Yamamoto H, Kanesaki Y, Hagemann M, Murata N: **Identification of histidine kinases that act as sensors in the perception of salt stress in *Synechocystis* sp. PCC 6803.** *PNAS* 2003, **100**:9061-9066.
120. Murata N, Suzuki I: **Exploitation of genomic sequences in a systematic analysis to access how cyanobacteria sense environmental stress.** *J Exp Bot* 2006, **57**:235-247.
121. Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M: **DNA microarray analysis of cyanobacterial gene expression during acclimation to high light.** *Plant Cell* 2001, **13**:793-806.
122. Suzuki I, Simon WJ, Slabas AR: **The heat shock response of *Synechocystis* sp. PCC 6803 analysed by transcriptomics and proteomics.** *J Exp Bot* 2006, **57**:1573-1578.
123. Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, et al: **Integrated genomic and proteomic analyses of gene expression in Mammalian cells.** *Mol Cell Proteomics* 2004, **3**:960-969.
124. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R: **Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae.** *Mol Cell Proteomics* 2002, **1**:323-333.
125. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 2001, **292**:929-934.
126. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.
127. Chan E, Biosoftware R: **Integrating transcriptomics and proteomics.** In *Drug Discovery & Development* vol. 6. pp. 20-26; 2006:20-26.
128. Bryant DA: **The cyanobacterial photosynthetic apparatus: comparisons to those of higher plants and photosynthetic bacteria.** In *Photosynthetic picoplankton. Volume* 214. Edited by Li TPaWKW: Can. Bull. Fish. Aquat. Sci.; 1986: 423-500
129. Joset F, Jeanjean R, Hagemann M: **Dynamics of response of cyanobacteria to salt stress: deciphering the molecular events.** *Physiol Plant* 1996, **96**:738-744.
130. Fernandez TA, iyer V, Apte SK: **Differential responses of nitrogen fixing cyanobacteria to salinity and osmotic stresses.** *Applied Environmental Microbiology* 1993, **59**:899-904.
131. Samsonoff W, MacColl R: **Biliproteins and phycobilisomes from cyanobacteria and red algae at the extremes of habitat.** *Arch Microbiol* 2001, **176**:400.
132. Thomas DN: **Photosynthetic microbes in freezing deserts.** *Trends Microbiol* 2005, **13**:87-88.
133. Echlin P: **The blue-green algae.** *Sci Am* 1966, **214**:75-81.
134. Nubel U, Garcia-Pichel F, Muyzer G: **PCR primers to amplify 16s rRNA genes from cyanobacteria.** *Appl Environ Microbiol* 1997, **63**:3327-3332.
135. Rippka R, Deruelles J, Waterbury J: **Generic assignments, strain histories and properties of pure cultures of cyanobacteria.** *J Gen Microbiol* 1979, **111**:1-61.

136. Hodgson D: **The ecology of cyanobacteria: Their diversity in time and space.** *Journal of Paleolimnology* 2002, **28**:383-384.
137. De Las Rivas J, Balsera M, Barber J: **Evolution of oxygenic photosynthesis: genome-wide analysis of the OEC extrinsic proteins.** *Trends in Plant Science* 2004, **9**:18-25.
138. McGowan SG, Britton E, Haworth BM: **Ancient blue–green blooms.** *Limnol Oceanogr* 1999, **44**:436-439.
139. Dufresne A, Garczarek: **Accelerated evolution associated with genome reduction in a free-living prokaryote.** *Genome Biology* 2005, **6**:R14.
140. Pietra F: *Secret world: Natural products of marine life.* Basel, Switzerland Birkhäuser; 1990.
141. Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR: **Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus.** *Antiviral Res* 2003, **58**:47-56.
142. Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, Gulakowski RJ, Wu L, Rivera MI, Laurencot CM, et al: **Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development.** *Antimicrobial Agents Chemistry* 1997, **41**:1521–1530.
143. Seckbach J: *Search for life in the universe with terrestrial microbes which thrive under extreme conditions.* Milan, Italy: Editrice Compositori; 1997.
144. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, et al: **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 (supplement).** *DNA Res* 1996, **3**:185-209.
145. Grigorieva G, Shestakov SV: **Transformation in the cyanobacterium *Synechocystis* sp. PCC6803.** *FEMS Microbiol Lett* 1982, **13**:367-370.
146. Asada Y, Miyake J: **Photobiological hydrogen production.** *J Biosci Bioeng* 1999, **88**:1-6.
147. Martinez-Ferez I, Fernandez-Gonzalez B, Sandmann G, Vioque A: **Cloning and expression in *Escherichia coli* of the gene coding for phytoene synthase from the cyanobacterium *Synechocystis* sp. PCC6803.** *Biochim Biophys Acta* 1994, **1218**:145-152.
148. Schwamborn M: **Chemical synthesis of polyaspartates: a biodegradable alternative to currently used polycarboxylate homo- and copolymers.** *Polymer Degradation and Stability* 1998, **59**:39-45.
149. Ashhuby BA: **Biofouling studies on reverse osmosis desalination of hypersaline waters PhD Thesis.** University of Sheffield, Chemical and Process Engineering; 2007.
150. El-Saedi: **A survey of Qabar Onn Lake.** University of Sebha, Department of Environmental Studies; 2002.
151. Garcia-Pichel F, Nubel U, Muyzer G: **The phylogeny of unicellular, extremely halotolerant cyanobacteria.** *Arch Microbiol* 1998, **169**:469-482.
152. **GOLD** [http://www.genomesonline.org] Last accessed 03/08.
153. Tomitani A, Okada K, Miyashita H, Matthijs HC, Ohno T, Tanaka A: **Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts.** *Nature* 1999, **400**:159-162.

154. Bhaya D, Dufresne A, Vaulot D, Grossman A: **Analysis of the hli gene family in marine and freshwater cyanobacteria.** *FEMS Microbiol Lett* 2002, **215**:209-219.
155. Mary I, Vaulot D: **Two-component systems in *Prochlorococcus* MED4: Genomic analysis and differential expression under stress.** *FEMS Microbiol Lett* 2003, **226**:135.
156. Moore LR, Chisholm S: **Photophysiology of the marine cyanobacterium *Prochlorococcus:* Ecotypic differences among cultured isolates.** *Limnol Oceanogr* 1999, **44**:628-638.
157. Moore LR, Rocap G, Crisholm SW: **Physiology and molecular phylogeny of coexisting *Prochlorococcus* types.** *Nature* 1998, **393**:464-467.
158. Boyer J: **Plant productivity and environment.** *Science* 1982, **218** 443-448
159. Nelson DE, Shen B, Bohnert HJ: *Salinity tolerance mechanisms, models and the metabolic engineering of complex traits: Principles and Methods.* New York: Plenum Press; 1998.
160. Bohnert HJ, Ayoubi P, Borchert C, Bressan RA, Burnap RL, Cushman JC, Cushman MA, Deyholos M, Galbraith DW, Hasegawa PM, et al: **A genomics approach towards salt stress tolerance.** *Plant Physiol Biochem* 2001, **39**:295-311.
161. Reed RH, Stewart WDP: *The responses of cyanobacteria to salt stress.* Oxford: Oxford Science Publisher; 1988.
162. Gale M: **Applications of molecular biology and genomics to genetic enhancement of crop tolerance to abiotic stress** Consultative Group on International Agricultural Research Science Council; 2003.
163. Galinski EA: **Osmoadaptation in bacteria.** *Adv Microb Physiol* 1995, **37**:272-328.
164. Yancey PH: **Water stress, osmolytes and proteins.** *Amer Zool* 2001, **41**:699-709.
165. Jeruzalmi D, Steitz TA: **Use of organic cosmotropic solutes to crystallize flexible proteins: application to T7 RNA polymerase and its complex with the inhibitor T7 lysozyme.** *J Mol Biol* 1997, **274**:748-756.
166. Brown CR, Hong-Brown LQ, Biwersi J, Verkman AS, Welch WJ: **Chemical chaperones correct the mutant phenotype of the delta F508 cystic fibrosis transmembrane conductance regulator protein.** *Cell Stress Chaperones* 1996, **1**:117-125.
167. Roberts MF: **Organic compatible solutes of halotolerant and halophilic microorganisms.** *Saline Systems* 2005, **1**:5.
168. Desmarais D, Jablonski PE, Fedarko NS, Roberts MF: **2-Sulfotrehalose, a novel osmolyte in haloalkaliphilic archaea.** *J Bacteriol* 1997, **179**:3146-3153.
169. Hagemann M, Erdmann N: *Cyanobacterial nitrogen metabolism and environmental biotechnology.* New Dehli: Springer; 1997.
170. Niu X, Bressan RA, Hasegawa PM, Pardo JM: **Ion homeostasis in NaCl stress environments.** *Plant Physiol* 1995, **109**:735-742.
171. Dimroth P: **Sodium ion transport decarboxylases and other aspects of sodium ion cycling in bacteria.** *Microbiol Rev* 1987, **51**:320-340.
172. Espie GS, Miller AG, Canvin DT: **Characterization of the Na-requirement in cyanobacterial photosynthesis.** *Plant Physiol* 1988, **88**:757-763.
173. Parida AK, Das AB: **Salt tolerance and salinity effects on plants: a review.** *Ecotoxicology and Environmental Safety* 2005, **60**:324.

174. Ferjani A, Mustardy L, Sulpice R, Marin K, Suzuki I, Hagemann M, Murata N: **Glucosylglycerol, a compatible solute, sustains cell division under salt stress.** *Plant Physiol* 2003, **131:**1628-1637.
175. Gabbay-Azaria R, Tel-Or E: *Mechanisms of salt tolerance in cyanobacteria.* Boca Raton: CRC Press; 1993.
176. Hagemann M, Erdmann N: **Activation and pathway of glucosylglycerol synthesis in the cyanobacterium *Synechocystis* sp. PCC 6803** *Microbiology* 1994, **140:**1427–1431.
177. Elanskaya I, Karandashova I, Bogachev A, Hagemann M: **Functional analysis of the Na^+/H^+ antiporter encoding genes of the cyanobacterium *Synechocystis* PCC 6803.** *Biochemistry (Mosc)* 2002, **67:**432-440.
178. Hamada A, Hibino T, Nakamura T, Takabe T: **Na+/H+ antiporter from *Synechocystis* species PCC 6803, homologous to SOS1, contains an aspartic residue and long C-terminal tail important for the carrier activity.** *Plant Physiol* 2001, **125:**437-446.
179. Blumwald E, Wolosin JM, Packer L: **Na+/H+ exchange in the cyanobacterium *Synechococcus* 6311.** *Biochem Biophys Res Commun* 1984, **122:**452-459.
180. Molitor V, Erber W, Peschek GA: **Increased levels of cytochrome oxidase and sodium-proton antiporter in the plasma membrane of *Anacystis nidulans* after growth in sodium-enriched media.** *FEBS Lett* 1986, **204:**251-256.
181. Jeanjean R, Matthijs H, Onana B, Havaux M, Joset F: **Exposure of the cyanobacterium *Synechocystis* PCC6803 to salt stress induces concerted changes in respiration and photosynthesis** *Plant Cell Physiol* 1993, **34:**1073-1079.
182. Inaba M, Sakamoto A, Murata N: **Functional expression in *Escherichia coli* of low-affinity and high-affinity $Na^+(Li^+)/H^+$ antiporters of *Synechocystis*.** *J Bacteriol* 2001, **183:**1376-1384.
183. Mulet JM, Leube MP, Kron SJ, Rios G, Fink GR, Serrano R: **A novel mechanism of ion homeostasis and salt tolerance in yeast: the Hal4 and Hal5 protein kinases modulate the Trk1-Trk2 potassium transporter.** *Mol Cell Biol* 1999, **19:**3328-3337.
184. Reed RH, Borowitzka LJ, Mackay MA, Chudek JA, Foster R, Warr SRC, Moore DJ, Stewart WDP: **Organic solute accumulation in osmotically stressed cyanobacteria.** *FEMS Microbiol Lett* 1986, **39:**51-56.
185. Karandashova IV, Elanskaia IV: **Genetic control and mechanisms of salt and hyperosmotic shock resistance in cyanobacteria.** *Genetika* 2005, **41:**1589-1600.
186. Hagemann M, Schoor A, Jeanjean R, Zuther E, Joset F: **The stpA gene form *Synechocystis* sp. strain PCC 6803 encodes the glucosylglycerol-phosphate phosphatase involved in cyanobacterial osmotic response to salt shock.** *J Bacteriol* 1997, **179:**1727-1733.
187. Cioni P, Bramanti E, Strambini GB: **Effects of sucrose on the internal dynamics of azurin.** *Biophys J* 2005, **88:**4213-4222.
188. Baldwin RL: **How Hofmeister ion interactions affect protein stability.** *Biophys J* 1996, **71:**2056-2063.
189. Mikkat S, Hagemann M, Schoor A: **Active transport of glucosylglycerol is involved in salt adaptation of the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *Microbiology* 1996, **142 (Pt 7):**1725-1732.

190. Fulda S, Hagemann M, Libbert E: **Release of glucosylglycerol from the cyanobacterium *Synechocystis* spec. SAG 92.79 by hypoosmotic shock.** *Arch Microbiol* 1990, **153**:405-408.
191. Marin K, Huckauf J, Fulda S, Hagemann M: **Salt-dependent expression of glucosylglycerol-phosphate synthase, involved in osmolyte synthesis in the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *J Bacteriol* 2002, **184**:2870-2877.
192. Reed RH, Stewart WDP: **Osmotic adjustment and solute accumulation in unicellular cyanobacteria from freshwater and marine habitats.** *Marine Biology* 1985, **88**:1-9.
193. Deshnium P, Los DA, Hayashi H, Mustardy L, Murata N: **Transformation of *Synechococcus* with a gene for choline oxidase enhances tolerance to salt stress.** *Plant Mol Biol* 1995, **29**:897-907.
194. Sudhir P, Murthy SDS: **Effects of salt stress on basic processes of photosynthesis.** *Photosynthetica* 2004, **42**:481.
195. Schubert H, Fulda S, Hagemann M: **Effects of adaptation to different salt concentrations on photosynthesis and pigmentation of the cyanobacterium *Synechocystis* sp. PCC 6803.** *J Plant Physiol* 1993, **142**:291–295.
196. Zuther E, Schubert H, Hagemann M: **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* 1998, **180**:1715-1722.
197. Murthy KNC, Vanitha A, Rajesha J, Swamy MM, Sowmya PR, Ravishankar GA: **In vivo antioxidant activity of carotenoids from *Dunaliella salina* -- a green microalga.** *Life Sciences* 2005, **76**:1381.
198. Allakhverdiev SI, Nishiyama Y, Suzuki I, Tasaka Y, Murata N: **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* 1999, **96**:5862-5867.
199. Lu C, Vonshak A: **Characterization of PS II photochemistry in salt-adapted cells of cyanobacterium *Spirulina platensis*.** *New Phytol* 1999, **141**:231-239.
200. Allakhverdiev SI, Sakamoto A, Nishiyama Y, Murata N: **Inactivation of photosystems I and II in response to osmotic stress in *Synechococcus*. Contribution of water channels.** *Plant Physiol* 2000, **122**:1201-1208.
201. Allakhverdiev SI, Sakamoto A, Nishiyama Y, Inaba M, Murata N: **Ionic and osmotic effects of NaCl-induced inactivation of photosystems I and II in *Synechococcus* sp.** *Plant Physiol* 2000, **123**:1047-1056.
202. Verma K, Mohanty P: **Changes of the photosynthetic apparatus in *Spirulina* cyanobacterium by sodium stress.** *Z Naturforsch [C]* 2000, **55**:16-22.
203. Verma K, Mohanty P: **Alterations in the structure of phycobilsomes of cyanobacterium *Spirulina platensis* in response to enhanced sodium level.** *World J Microbiol Biotechnol* 2000, **16**:795-798.
204. Allakhverdiev SI, Nishiyama Y, Miyairi S, Yamamoto H, Inagaki N, Kanesaki Y, Murata N: **Salt stress inhibits the repair of photodamaged photosystem II by suppressing the transcription and translation of *psbA* genes in *Synechocystis*.** *Plant Physiol* 2002, **130**:1443-1453.

205. Takabe T, Incharoensakdi A, Arakawa K, Yokota S: **CO_2 fixation rate and RuBisCO content increase in the halotolerant cyanobacterium, *Aphanothece halophytica*, grown in high salinities.** *Plant Physiol* 1988, **88:**1120-1124.
206. Mackay M, Norton RS: **^{13}C nuclear magnetic resonance study of the biosynthesis of glucosylglycerol by cyanobacterium under osmotic stress.** *J Gen Microbiol* 1987, **130:**2177-2191.
207. Erdmann N, Berg C, Hagemann M: **Missing salt adaptation of *Microcystis firma* (cyanobacterium) in the dark.** *Arch Hydrobiol* 1989, **114:**521-530.
208. Calvin M: **The path of carbon in photosynthesis.** *Harvey Lect* 1950, **Series 46:**218-251.
209. Raven P, Johnson G: *Biology.* 4 edn: William C Brown Pub; 1996.
210. Hagemann M, Jeanjean R, Fulda S, Havaux M, Joset F, Erdmann N: **Flavodoxin accumulation contributes to enhanced cyclic electron flow around photosystem I in salt-stressed cells of *Synechocystis* sp. strain PCC 6803.** *Physiol Plant* 1999, **105:**670-678.
211. Ryu JY, Suh KH, Chung YH, Park YM, Chow WS, Park YI: **Cytochrome c oxidase of the cyanobacterium *Synechocystis* sp. PCC 6803 protects photosynthesis from salt stress.** *Mol Cells* 2003, **16:**74-77.
212. Huflejt ME, Tremolieres A, Pineau B, Lang JK, Hatheway J, Packer L: **Changes in membrane lipid composition during saline growth of the fresh water cyanobacterium *Synechococcus* 6311.** *Plant Physiol* 1990, **94:**1512-1521.
213. Ritter D, Yopp JH: **Plasma membrane lipid composition of the halophilic cyanobacterium *Aphanothece halophytica.*** *Arch Microbiol* 1993, **159:**435-439.
214. Riviere ME, Arrio B, Steffan I, Molitor V, Kuntner O, Peschek GA: **Changes of some physical properties of isolated and purified plasma and thylakoid membrane vesicles from the freshwater cyanobacterium *Synechococcus* sp. 6301 (*Anacystis nidulans*) during adaptation to salinity.** *Arch Biochem Biophys* 1990, **280:**159-166.
215. Allakhverdiev SI, Kinoshita M, Inaba M, Suzuki I, Murata N: **Unsaturated fatty acids in membrane lipids protect the photosynthetic machinery against salt-induced damage in *Synechococcus.*** *Plant Physiol* 2001, **125:**1842-1853.
216. Huckauf J, Nomura C, Forchhammer K, Hagemann M: **Stress responses of *Synechocystis* sp. strain PCC 6803 mutants impaired in genes encoding putative alternative sigma factors.** *Microbiology* 2000, **146 (Pt 11):**2877-2889.
217. Shoumskaya MA, Paithoonrangsarid K, Kanesaki Y, Los DA, Zinchenko VV, Tanticharoen M, Suzuki I, Murata N: **Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in synechocystis.** *J Biol Chem* 2005, **280:**21531-21538.
218. Torrecilla I, Leganes F, Bonilla I, Fernandez-pinaz F: **Calcium transients in response to salinity and osmotic stress in the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC7120, expressing cytosolic apoaequorin.** *Plant, Cell and Environment* 2001, **24:**641-648.

219. Apte SK, Haselkorn R: **Cloning of salinity stress-induced genes from the salt-tolerant nitrogen-fixing cyanobacterium *Anabaena torulosa*.** *Plant Mol Biol* 1990, **15:**723-733.
220. Aro EM, McCaffery S, Anderson JM: **Photoinhibition and D1 protein degradation in peas acclimated to different growth irradiances.** *Plant Physiol* 1993, **103:**835-843.
221. Imamura S, Yoshihara S, Nakano S, Shiozaki N, Yamada A, Tanaka K, Takahashi H, Asayama M, Shirai M: **Purification, characterization, and gene expression of all sigma factors of RNA polymerase in a cyanobacterium.** *J Mol Biol* 2003, **325:**857-872.
222. Chen Y, Drysdale J: **Detection of iron binding proteins by a blotting technique.** *Anal Biochem* 1993, **212:**47-49.
223. Wirth P, Romano A: **Staining methods in gel electrophoresis, including the use of multiple detection methods.** *J Chromatogr A* 1995, **698:**123-143.
224. Huang F, Fulda S, Hagemann M, Norling B: **Proteomic screening of salt-stress-induced changes in plasma membranes of *Synechocystis* sp. strain PCC 6803.** *Proteomics* 2006, **6:**910-920.
225. Höper D, Bernhardt J, Hecker M: **Salt stress adaptation of *Bacillus subtilis*: A physiological proteomics approach.** *Proteomics* 2006, **6:**1550-1562.
226. Nakamura Y, Kaneko T, Tabata S: **CyanoBase, the genome database for *Synechocystis* sp. strain PCC6803: status for the year 2000.** *Nucl Acids Res* 2000, **28:**72-73.
227. Norling B, Zak E, Andersson B, Pakrasi H: **2D-isolation of pure plasma and thylakoid membranes from the cyanobacterium *Synechocystis* sp. PCC 6803.** *FEBS Lett* 1998, **436:**189-192.
228. Huang F, Parmryd I, Nilsson F, Persson AL, Pakrasi HB, Andersson B, Norling B: **Proteomics of *Synechocystis* sp. Strain PCC 6803: identification of plasma membrane proteins.** *Mol Cell Proteomics* 2002, **1:**956-966.
229. Pisareva T, Shumskaya M, Maddalo G, Ilag L, Norling B: **Proteomics of *Synechocystis* sp. PCC 6803. Identification of novel integral plasma membrane proteins.** *FEBS Journal* 2007, **274:**791-804.
230. Kashino Y, Harayama T, Pakrasi HB, Satoh K: **Preparation of membrane proteins for analysis by two-dimensional gel electrophoresis.** *J Chromatogr B Biomed Appl* 2007, **849:**282-292.
231. Fulda S, Mikkat S, Schroder W, Hagemann M: **Isolation of salt-induced periplasmic proteins from *Synechocystis* sp. strain PCC 6803.** *Arch Microbiol* 1999, **171:**214-217.
232. Herranen M, Battchikova N, Zhang P, Graf A, Sirpio S, Paakkarinen V, Aro EM: **Towards functional proteomics of membrane protein complexes in *Synechocystis* sp. PCC 6803.** *Plant Physiol* 2004, **134:**470-481.
233. Fulda S, Huang F, Nilsson F, Hagemann M, Norling B: **Proteomics of *Synechocystis* sp. strain PCC 6803: Identification of periplasmic proteins cells grown at low and high salt concentrations.** *Eur J Biochem* 2000, **267:**5900-5907.
234. Fulda S, Mikkat S, Huang F, Huckauf J, Marin K, Norling B, Hagemann M: **Proteome analysis of salt stress response in the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *Proteomics* 2006, **6:**2733-2745.
235. Sudhir PR, Pogoryelov D, Kovacs L, Garab G, Murthy SD: **The effects of salt stress on photosynthetic electron transport and thylakoid membrane**

proteins in the cyanobacterium *Spirulina platensis*. *J Biochem Mol Biol* 2005, **38**:481-485.

236. Neu HC, LA H: **The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts.** *J Biol Chem* 1965, **240**:3685–3692.
237. Wittig I, Braun H-P, Schagger H: **Blue native PAGE.** *Nat Protocols* 2006, **1**:418-428.
238. Camacho-Carvajal MM, Wollscheid B, Aebersold R, Steimle V, Schamel WWA: **Two-dimensional blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates - a proteomics approach.** *Mol Cell Proteomics* 2003:193-206.
239. Norling B, Zak E, Andersson B, Pakrasi H: **2D-isolation of pure plasma and thylakoid membranes from the cyanobacterium *Synechocystis* sp. PCC 6803.** *FEBS Letters* 1998, **436**:189.
240. Wessel D, Flugge UI: **A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids.** *Anal Biochem* 1984, **138**:141-143.
241. Hagemann M, Richter S, Mikkat S: **The ggtA gene encodes a subunit of the transport system for the osmoprotective compound glucosylglycerol in *Synechocystis* sp. strain PCC 6803.** *J Bacteriol* 1997, **179**:714-720.
242. Tolle J, Michel KP, Kruip J, Kahmann U, Preisfeld A, Pistorius EK: **Localization and function of the IdiA homologue Slr1295 in the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *Microbiology* 2002, **148**:3293-3305.
243. Wilkinson MJ, Northcote DH: **Plasma membrane ultrastructure during plant protoplast plasmolysis, isolation and wall regeneration: a freeze-fracture study.** *J Cell Sci* 1980, **42**:401-415.
244. Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S, Vothknecht UC, Soll J, Westhoff P: **VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation.** *Proc Natl Acad Sci U S A* 2001, **98**:4238-4242.
245. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
246. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, et al: **Pfam: clans, web tools and services.** *Nucl Acids Res* 2006, **34**:D247-251.
247. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P: **PROSITE: a documented database using patterns and profiles as motif descriptors.** *Brief Bioinform* 2002, **3**:265-274.
248. Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS: **PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis.** *Bioinformatics* 2005, **21**:617-623.
249. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A: **Prediction of lipoprotein signal peptides in Gram-negative bacteria.** *Protein Sci* 2003, **12**:1652-1662.
250. Liska AJ, Shevchenko A, Pick U, Katz A: **Enhanced photosynthesis and redox energy production contribute to salinity tolerance in *Dunaliella* as revealed by homology-based proteomics.** *Plant Physiol* 2004, **136**:2806-2817.

251. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A: **The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data.** *Mol Cell Proteomics* 2004, **3**:531-533.
252. Hihara Y: **The molecular mechanism for acclimation to high light in cyanobacteria.** *Curr Opin Plant Biol* 1999, **1**.
253. Kok B: **On the inhibition of photosynthesis by intense light.** *Biochim Biophys Acta* 1956, **21**:234-244.
254. Asada K: *Production and action of active oxygen species in photosynthetic tissues.* Boca Raton, FL: CRC Press; 1994.
255. Barber J, Andersson B: **Too much of a good thing: light can be bad for photosynthesis.** *Trends Biochem Sci* 1992, **17**:61-66.
256. Campbell D, Hurry V, Clarke AK, Gustafsson P, Oquist G: **Chlorophyll fluorescence analysis of cyanobacterial photosynthesis and acclimation.** *Microbiol Mol Biol Rev* 1998, **62**:667-683.
257. Niyogi KK: **Photoprotection revisited: genetic and molecular approaches.** *Ann Rev Plant Physiol Mol Biol* 1999, **50**:333-359.
258. Pfannschmidt T, Nilsson A, Tullberg A, Link G, Allen JF: **Direct transcriptional control of the chloroplast genes *psbA* and *psaAB* adjusts photosynthesis to light energy distribution in plants.** *IUBMB Life* 1999, **48**:271-276.
259. Partensky F, Garczarek L: *The photosynthetic apparatus of chlorophyll b- and d-containing oxyphotobacteria.* Dordrecht, The Netherlands: Kluwer Academic Publishers; 2003.
260. de Lorimier RM, Smith RL, Stevens SE: **Regulation of phycobilisome structure and gene expression by light intensity.** *Plant Physiol* 1992, **98**:1003-1010.
261. Belknap WR, Haselkorn R: **Cloning and light regulation of expression of the phycocyanin operon of the cyanobacterium *Anabaena*.** *EMBO J* 1987, **6**:871-884.
262. Schwarz R, Grossman AR: **A response regulator of cyanobacteria integrates diverse environmental signals and is critical for survival under extreme conditions.** *Proc Natl Acad Sci U S A* 1998, **95**:11008-11013.
263. Hagemann M, Fulda S, Schubert H: **DNA, RNA, and protein synthesis in the cyanobacterium *Synechocystis* sp. PCC 6803 adapted to different salt concentrations.** *Curr Microbiol* 1994, **28**:201-207.
264. Hihara Y, Sonoike K, Ikeuchi M: **A novel gene, pmgA, specifically regulates photosystem stoichiometry in the cyanobacterium *Synechocystis* species PCC 6803 in response to high light.** *Plant Physiol* 1998, **117**:1205-1216.
265. Kawamura M, Mimuro M, Fujita Y: **Quantitative relationship between two reaction centers in the photosynthetic system of blue-green algae.** *Plant Cell Physiol* 1979, **20**:697-705.
266. Mohamed A, Jansson C: **Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium *Synechocystis* sp. PCC6803.** *Plant Mol Biol* 1989, **13**:693-700.
267. Mohamed A, Jansson C: **Photosynthetic electron transport controls degradation but not production of *psbA* transcripts in the cyanobacterium *Synechocystis* sp. PCC6803.** *Plant Mol Biol* 1991, **16**:891-897.
268. Tyystjarvi T, Tyystjarvi E, Ohad I, Aro EM: **Exposure of *Synechocystis* 6803 cells to series of single turnover flashes increases the *psbA* transcript level**

by activating transcription and down-regulating *psbA* mRNA degradation. *FEBS Lett* 1998, **436**:483-487.

269. Kulkarni RD, Golden SS: **Adaptation to high light intensity in *Synechococcus* sp. strain PCC 7942: regulation of three *psbA* genes and two forms of the D1 protein.** *J Bacteriol* 1994, **176**:959-965.
270. Brahamsha B, Haselkorn R: **Isolation and characterization of the gene encoding the principal sigma factor of the vegetative cell RNA polymerase from the cyanobacterium *Anabaena* sp. strain PCC 7120.** *J Bacteriol* 1991, **173**:2442-2450.
271. Clarke AK, Schelin J, Porankiewicz J: **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* 1998, **37**:791-801.
272. Mary I, Tu C-J, Grossman A, Vaulot D: **Effects of high light on transcripts of stress-associated genes for the cyanobacteria *Synechocystis* sp. PCC 6803 and *Prochlorococcus* MED4 and MIT9313.** *Microbiology* 2004, **150**:1271-1281.
273. Förster B, Mathesius U, Pogson BJ: **Comparative proteomics of high light stress in the model alga *Chlamydomonas reinhardtii.*** *Proteomics* 2006, **6**:4309-4320.
274. Giacomelli L, Rudella A, van Wijk KJ: **High light response of the thylakoid proteome in Arabidopsis wild type and the ascorbate-deficient mutant vtc2-2. A comparative proteomics study.** *Plant Physiol* 2006, **141**:685-701.
275. Zolla L, Rinalducci S, Timperio AM, Huber CG: **Proteomics of light-harvesting proteins in different plant species. Analysis and comparison by liquid chromatography-electrospray ionization mass spectrometry. Photosystem I.** *Plant Physiol* 2002, **130**:1938-1950.
276. Liu HB, Nolla HA, Campbell L: ***Prochlorococcus* growth rate and contribution to primary production in the Equatorial and Subtropical North Pacific ocean.** *Aquat Microb Ecol* 1997, **12**:39-47.
277. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: The unseen majority.** *Proc Natl Acad Sci U S A* 1998, **95**:6578-6583.
278. Hess WR, Rocap G, Ting CS, Larimer F, Stilwagen S, Lamerdin J, Chisholm SW: **The photosynthetic apparatus of *Prochloroccocus*: Insights through comparitive genomics.** *Photosynthetic Research* 2001, **70**:53-71.
279. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, et al: **Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation.** *Nature* 2003, **424**:1042-1047.
280. West NJ, Schonhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF, Scanlan DJ: **Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides.** *Microbiology* 2001, **147**:1731-1744.
281. Roche JL, van der Staay GWM, Partensky F, Ducret A, Aebersold R, Li R, Golden SS, Hiller RG, Wrench PM, Larkum AWD, Green BR: **Independent evolution of the prochlorophyte and green plant chlorophyll a/b light-harvesting proteins.** *Proc Natl Acad Sci U S A* 1996, **93**:15244-15248.

282. Garczarek L, Hess WR, Holtzendorff J, van der Staay GW, Partensky F: **Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote.** *Proc Natl Acad Sci U S A* 2000, **97:**4098-4101.
283. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW: **Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients.** *Science* 2006, **311:**1737-1740.
284. Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ, Chisholm SW: ***Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method.** *Appl Environ Microbiol* 2006, **72:**723-732.
285. Scanlan DJ, West NJ: **Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*.** *FEMS Microbiology Ecology* 2002, **40:**1-12.
286. West NJ, Scanlan DJ: **Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean.** *Appl Environ Microbiol* 1999, **65:**2585-2591.
287. Duche O, Tremoulet F, Namane A, Labadie J: **A proteomic analysis of the salt stress response of *Listeria monocytogenes*.** *FEMS Microbiol Lett* 2002, **215:**183-188.
288. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al: **Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents.** *Mol Cell Proteomics* 2004, **3:**1154-1169.
289. Fey SJ, Larsen PM: **2D or not 2D.** *Curr Opin Chem Biol* 2001, **5:**26.
290. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA: **The Paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra.** *Mol Cell Proteomics* 2007, **6:**1638-1655.
291. Varvasovszki V, Glatz A, Shigapova N, Josvay K, Vigh L, Horvath I: **Only one dnaK homolog, dnaK2, is active transcriptionally and is essential in *Synechocystis*.** *Biochem Biophys Res Commun* 2003, **305:**641-648.
292. Winer J, Jung CK, Shackel I, Williams PM: **Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro.** *Anal Biochem* 1999, **270:**41-49.
293. Cargile BJ, Bundy JL, Stephenson JL, Jr.: **Potential for false positive identifications from large databases through tandem mass spectrometry.** *J Proteome Res* 2004, **3:**1082-1085.
294. Jacob U, Buchner J: **Assisting spontaneity: the role of Hsp90 and small Hsps as molecular chaperones.**. *Trends Biochem Sci* 1994, **19:**205-211.
295. Suzuki M, Hashiokab A, Mimurac T, Ashihara H: **Salt stress and glycolytic regulation in suspension-cultured cells of the mangrove tree, *Bruguiera sexangula*.** *Physiol Plant* 2005, **123:**246-253.
296. **KEGG** [http://www.genome.jp/kegg/pathway.html] Last accessed 03/08.
297. Hagemann M, Golldack D, Biggins J, Erdmann N: **Salt-dependent protein phosphorylation in the cyanobacterium *Synechocystis* PCC 6803.** *FEMS Microbiol Lett* 1993, **113:**205-209.

298. Nishiyama Y, Allakhverdiev SI, Murata N: **Inhibition of the repair of photosystem II by oxidative stress in cyanobacteria.** *Photosynth Res* 2005, **84:**1-7.
299. Wilkinson MJ, Northcote DH: **Plasma membrane ultrastructure during plant protoplast plasmolysis, isolation and wall regeneration: a freeze-fracture study.** *J Cell Sci* 1980, **42:**401-415.
300. Reddy BR, Apte SK, Thomas J: **Enhancement of cyanobacterial salt tolerance by combined nitrogen.** *Plant Physiol* 1989, **89:**204-210.
301. Fischer G, Schmid FX: **The mechanism of protein folding. Implications of *in vitro* refolding models for de novo protein folding and translocation in the cell.** *Biochemistry* 1990, **29:**2205-2212.
302. Janosi L, Shimizu I, Kaji A: **Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth.** *Proc Natl Acad Sci U S A* 1994, **91:**4249-4253.
303. Smith CA: **Structure, function and dynamics in the mur family of bacterial cell wall ligases.** *J Mol Biol* 2006, **362:**640-655.
304. Singh AK, Summerfield TC, Li H, Sherman LA: **The heat shock response in the cyanobacterium *Synechocystis* sp. Strain PCC 6803 and regulation of gene expression by HrcA and SigB.** *Arch Microbiol* 2006, **186:**273-286.
305. Benson AK, Haldenwang WG: ***Bacillus subtilis* sigma B is regulated by a binding protein (RsbW) that blocks its association with core RNA polymerase.** *Proc Natl Acad Sci U S A* 1993, **90:**2330-2334.
306. Hlaváček O, Váchová L: **ATP-Dependent proteinases in bacteria.** *Folia Microbiologica* 2002, **47:**203-212.
307. Mata J, Marguerat S, Bahler J: **Post-transcriptional control of gene expression: a genome-wide perspective.** *Trends Biochem Sci* 2005, **30:**506-514.
308. Ashcroft S, Periera C: *Practical statistics for the biological sciences.* Basingstoke: Palgrave macmillan; 2003.
309. Waters KM, Pounds JG, Thrall BD: **Data merging for integrated microarray and proteomic analysis.** *Brief Funct Genomic Proteomic* 2006, **5:**261-272.
310. Taylor NL, Heazlewood JL, Day DA, Millar AH: **Differential impact of environmental stresses on the pea mitochondrial proteome.** *Mol Cell Proteomics* 2005, **4:**1122-1133.
311. Ventosa A, Nieto JJ, Oren A: **Biology of moderately halophilic aerobic bacteria.** *Microbiol Mol Biol Rev* 1998, **62:**504-544.
312. Kushner DJ: **Life in high salt and solute concentrations: halophilic bacteria. .** In *Microbial life in extreme environments.* Edited by J. KD. London: Academic Press, Ltd; 1978: 317–368
313. Snijders AP, de Vos MG, de Koning B, Wright PC: **A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and ^{15}N-metabolic labelling.** *Electrophoresis* 2005, **26:**3191-3199.
314. Snijders AP, de Koning B, Wright PC: **Perturbation and interpretation of nitrogen isotope distribution patterns in proteomics.** *J Proteome Res* 2005, **4:**2185-2191.
315. Mackinney G: **Absorption of light by chlorophyll solutions.** *J Biol Chem* 1941, **140:**315-322.

316. Joubert-Caron R, Feuillard J, Kohanna S, Poirier F, Le Caer JP, Schuhmacher M, Bornkamm GW, Polack A, Caron M, Bladier D, Raphael M: **A computer-assisted two-dimensional gel electrophoresis approach for studying the variations in protein expression related to an induced functional repression of NFkappaB in lymphoblastoid cell lines.** *Electrophoresis* 1999, **20:**1017-1026.
317. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4:**406-425.
318. Snijders AP, de Vos MG, Wright PC: **Novel approach for peptide quantitation and sequencing based on ^{15}N and ^{13}C metabolic labeling.** *J Proteome Res* 2005, **4:**578-585.
319. Berkelman T, Stenstedt T: *2D Electrophoresis: Principles and methods.* Uppsala, Sweden Amersham Pharmacia Biotech; 2001.
320. Yuan X, Desiderio DM: **Proteomics analysis of human cerebrospinal fluid.** *J Chromatogr B Biomed Appl* 2005, **815:**179-189.
321. Westermeier R, naven T: *Proteomics in Practice.* 3rd edn. Europe: Amersham Biosciences; 2002.
322. Pandhal J, Biggs C, Wright P: **Proteomics with a pinch of salt: A cyanobacterial perspective.** *Saline Systems* 2008, **4:**1.
323. Cagney G, Amiri S, Premawaradena T, Lindo M, Emili A: ***In silico* proteome analysis to facilitate proteomics experiments using mass spectrometry.** *Proteome Science* 2003, **1:**5.
324. Vinnemeier J, Kunert A, Hagemann M: **Transcriptional analysis of the isiAB operon in salt-stressed cells of the cyanobacterium *Synechocystis* sp. PCC6803.** *FEMS Microbiol Lett* 1998, **169:**323-330.
325. Jakob U, J. B: **Assisting spontaneity: the role of Hsp90 and small Hsps as molecular chaperones.** *Trends Biochem Sci* 1994, **19:**205-211.
326. Espinosa-Ruiz A, Belles JM, Serrano R, Culianez-Macia FA: **Arabidopsis thaliana AtHAL3: a flavoprotein related to salt and osmotic tolerance and plant growth.** *The Plant Journal* 1999, **20:**529-539.
327. Walker GC: *The SOS response of Escherichia coli.* 2nd edn. Washington D.C. : ASM Press; 1996.
328. Domain F, Houot L, Chauvat F, Cassier-Chauvat C: **Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation.** *Molecular Microbiology* 2004, **53:**65-80.
329. Gutekunst K, Phunpruch S, Schwarz C, Schuchardt S, Schulz-Friedrich R, Appel J: **LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator.** *Molecular Microbiology* 2005, **58:**810-823.
330. Oliveira P, Lindblad P: **LexA, a transcription regulator binding in the promoter region of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803.** *FEMS Microbiol Lett* 2005, **251:**59.
331. Schutz K, Happe T, Troshina O, Lindblad P, LeitÃ£o E, Oliveira P, Tamagnini P: **Cyanobacterial H_2 production: a comparative analysis.** *Planta* 2004, **218:**350.
332. Apte SK, Fernandes T, Badran H, Ballal A: **Expression and possible role of stress-responsive proteins in *Anabaena*.** *J Biosci* 1998, **23:**399-406.
333. Kramer G, Rutkowska A, Wegrzyn RD, Patzelt H, Kurz TA, Merz F, Rauch T, Vorderwülbecke S, Deuerling E, Bukau B: **Functional dissection of**

Escherichia coli **trigger factor: unraveling the function of individual domains.** *J Bacteriol* 2004, **186:**3777-3784.

334. Liu J, Zhu J-K: **An Arabidopsis mutant that requires increased calcium for potassium nutrition and salt tolerance.** *Proceedings of the National Academy of Sciences* 1997, **94:**14960-14964.
335. Knight H, Trewavas AJ, Knight MR: **Calcium signalling in Arabidopsis thaliana responding to drought and salinity.** *Plant J* 1997, **12:**1067-1078.
336. John Scott FRJF: **Folic acid and folates: the feasibility for nutritional enhancement in plant foods.** *Journal of the Science of Food and Agriculture* 2000, **80:**795-824.
337. Miller D, Jones J, Yopp J, Tindall D, Schmid W: **Ion metabolism in a halophilic blue-green alga, *Aphanothece halophytica*.** *Archives in microbiology* 1976, **111:**145-149.
338. Takabe T, Incharoensakdi A, Arakawa K, Yokota S: **CO_2 fixation rate and RuBisCO content increase in the halotolerant cyanobacterium, *Aphanothece halophytica*, grown in high salinities.** *Plant Physiol* 1988, **88:**1120-1124.
339. Frolow F, Harel M, Sussman JL, Mevarech M, Shoham M: **Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin.** *Nat Struct Mol Biol* 1996, **3:**452.
340. Dym O, Mevarech M, Sussman JL: **Structural features that stabilize halophilic malate dehydrogenase from an Archaebacterium.** *Science* 1995, **267:**1344-1346.
341. Oren A, Mana L: **Amino acid composition of bulk protein and salt relationships of selected enzymes of *Salinibacter ruber*, an extremely halophilic bacterium.** *Extremophiles* 2002, **6:**217-223.
342. Tokunaga M, Matsuoka K, Tokunaga H: **Identification and NH2-terminal amino acid sequences of DnaK and groEL homologues in moderate eubacterial halophiles.** *Biosci Biotechnol Biochem* 1997, **61:**1388-1390.
343. Pogoryelov D, Sudhir PR, Kovacs L, Gombos Z, Brown I, Garab G: **Sodium dependency of the photosynthetic electron transport in the alkaliphilic cyanobacterium *Arthrospira platensis*.** *J Bioenerg Biomembr* 2003, **35:**427-437.
344. Rao JK, Argos P: **Structural stability of halophilic proteins.** *Biochemistry* 1981, **20:**6536-6543.
345. Samartzidou H, Widger WR: **Transcriptional and posttranscriptional control of mRNA from lrtA, a light-repressed transcript in *Synechococcus* sp. PCC 7002.** *Plant Physiol* 1998, **117:**225-234.
346. Zhu JK: **Plant salt tolerance.** *Trends Plant Sci* 2001, **6:**66-71.
347. Snijders AP, de Koning B, Wright PC: **Relative quantification of proteins across the species boundary through the use of shared peptides.** *J Proteome Res* 2007, **6:**97-104.
348. Zhang R, Sioma CS, Wang S, Regnier FE: **Fractionation of isotopically labeled peptides in quantitative proteomics.** *Anal Chem* 2001, **73:**5142-5149.
349. Oh J-E, Hong S-W, Lee Y, Koh E-J, Kim K, Seo YW, Chung N, Jeong M, Jang CS, Lee B, et al: **Modulation of gene expressions and enzyme activities of methionine sulfoxide reductases by cold, ABA or high salt treatments in Arabidopsis.** *Plant Science* 2005, **169:**1030-1036.

350. Nesvizhskii AI, Aebersold R: **Interpretation of shotgun proteomic data: the protein inference problem.** *Mol Cell Proteomics* 2005, **4:**1419-1440.
351. Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry.** *Anal Chem* 2003, **75:**4646-4658.
352. Gibson JL, Chen JH, Tower PA, Tabita FR: **The form II fructose 1,6-bisphosphatase and phosphoribulokinase genes form part of a large operon in *Rhodobacter sphaeroides*: primary structure and insertional mutagenesis analysis.** *Biochemistry (Mosc)* 1990, **29:**8085-8093.
353. Kneidinger B, Graninger M, Adam G, Puchberger M, Kosma P, Zayni S, Messner P: **Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in *Aneurinibacillus thermoaerophilus* L420-91T.** *J Biol Chem* 2001, **276:**5577-5583.
354. DeSouza L, Diehl G, Rodrigues MJ, Guo J, Romaschin AD, Colgan TJ, Siu KW: **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* 2005, **4:**377-386.
355. Lespinet O, Wolf YI, Koonin EV, Aravind L: **The role of lineage-specific gene family expansion in the evolution of eukaryotes.** *Genome Res* 2002, **12:**1048-1059.
356. Madlung A, Comai L: **The effect of stress on genome regulation and structure.** *Ann Bot (Lond)* 2004, **94:**481-495.
357. Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, Westjohnsrud L, Zettler ER: ***Prochlorococcus marinus* nov. gen. nov. sp.: An oxyphototrophic marine prokaryote containing divinyl chlorophyll a and chlorophyll b.** *Arch Microbiol* 1992, **157:**297-300.
358. Turekian KK: *Oceans.* 1st edn. Englewood Cliffs, NJ: Prentice-Hall; 1968.
359. Kyte J, Doolittle RF: **A simple method for displaying the hydropathic character of a protein.** *J Mol Biol* 1982, **157:**105.
360. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteomics* 2005, **4:**1265-1272.
361. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278:**631-637.
362. **JVIRGEL** [http://www.jvirgel.de/] Last accessed 05/07
363. Garcia-Fernandez JM, Hess WR, Houmard J, Partensky F: **Expression of the psbA gene in the marine oxyphotobacteria *Prochlorococcus* spp.** *Arch Biochem Biophys* 1998, **359:**17-23.
364. Hess WR, Weihe A, Loiseaux-de Goer S, Partensky F, Vaulot D: **Characterization of the single psbA gene of *Prochlorococcus marinus* CCMP 1375 (Prochlorophyta).** *Plant Mol Biol* 1995, **27:**1189-1196.
365. Campbell D, Eriksson M-J, Oquist G, Gustafsson P, Clarke AK: **The cyanobacterium *Synechococcus* resists UV-B by exchanging photosystem II reaction-center D1 proteins.** *PNAS* 1998, **95:**364-369.
366. Sommaruga R, Hofer JS, Alonso-Sáez L, Gasol JM: **Differential sunlight sensitivity of picophytoplankton from surface Mediterranean coastal waters.** *Appl Environ Microbiol* 2005, **71:**2154–2157.

367. Glover JR, Lindquist S: **Hsp104, Hsp70, and Hsp40: A novel chaperone system that rescues previously aggregated proteins.** *Cell* 1998, **94:**73.
368. Tanaka N, Nakamoto H: **HtpG is essential for the thermal stress management in cyanobacteria.** *FEBS Lett* 1999, **458:**117.
369. Hossain M, Nakamoto H: **Role for the cyanobacterial HtpG in protection from oxidative stress** *Curr Microbiol* 2003, **46:**70-76.
370. Clarke AK, Eriksson MJ: **The cyanobacterium *Synechococcus* sp. PCC 7942 possesses a close homologue to the chloroplast ClpC protein of higher plants.** *Plant Mol Biol* 1996, **31:**721-730.
371. Lutkenhaus J: **FtsZ ring in bacterial cytokinesis.** *Mol Microbiol* 1993, **9:**403-409.
372. Cannon GC, bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM: **Microcompartments in prokaryotes: carboxysomes and related polyhedra.** *Appl Environ Microbiol* 2001, **67:**5351-5361.
373. Kaplan A, Reinhold L: **CO_2 concentrating mechanisms in photosynthetic microorganisms.** *Annu Rev Plant Physiol Plant Mol Biol* 1999, **50:**539-570.
374. Scanlan DJ, Silman NJ, Donald KM, Wilson WH, Carr NG, Joint I, Mann NH: **An immunological approach to detect phosphate stress in populations and single cells of photosynthetic picoplankton.** *Appl Environ Microbiol* 1997, **63:**2411-2420.
375. Goericke R, Repeta DJ: **The pigments of *Prochlorococcus marinus*: the presence of divinyl chlorophyll *a* and *b* in a marine procaryote** *Limnol Oceanogr* 1992, **37:**425-433.
376. Hess WR, Partensky F, van der Staay GW, Garcia-Fernandez JM, Borner T, Vaulot D: **Coexistence of phycoerythrin and a chlorophyll a/b antenna in a marine prokaryote.** *Proc Natl Acad Sci U S A* 1996, **93:**11126-11130.
377. Chen J, Balgley BM, DeVoe DL, Lee CS: **Capillary isoelectric focusing-based multidimensional concentration/separation platform for proteome analysis.** *Anal Chem* 2003, **75:**3145-3152.

Appendices

Appendices are available on the accompanying CD (attached to back of thesis) rather than as a hard copy due to the large amount of information contained.

Appendix	Chapter	File name	File type	File description
A	3	Appendix A	Excel	iTRAQ: A list of all proteins and peptides identified in workflow I and II, with quantitations
B	4 (Part A)	Appendix B	Excel	A list of all proteins and peptides identified, with quantitations
C	4 (Part B)	Appendix C	Excel	A list of all proteins identified, with quantitations
D	4 (Part B)	Appendix D	Excel	*De novo* sequencing comparison
E	5	Appendix E	Excel	A list of all proteins identified, with quantitations and proteins shared between both methods (n=47)
F	5	Appendix F	Excel	iTRAQ: A list of all proteins and peptides identified, with quantitations
G	6	Appendix G	Excel	ICP-MS analysis
H	6	Appendix H	Excel	iTRAQ: A list of all proteins and peptides identified in iTRAQ 1 and 2, with quantitations