Proteomics in Microalgae: A Postgenomic Approach for Improved Biofuel Production

By: Joseph Longworth (MSci, MSc)

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

Department of Chemical and Biological Engineering

University of Sheffield

Submitted: September 2013

Thesis Summary

Biofuels from microalgae have become established as a key contender for a renewable transportation fuel. With their relatively high growth rate and lipid content, compared to terrestrial crops, microalgae represent an ideal feedstock for biofuels.

Systems biology and 'omic' techniques have provided potent tools to interrogate an organism under various conditions. One of the most powerful of these tools is proteomics. By quantitatively assessing proteins, the fundamental functional component of biology, greater understanding of the investigated organism can be gained.

This thesis reviews the process of biofuel production from microalgae and the current techniques in proteomics before providing a comprehensive list of previous proteomic investigations in microalgae.

The proteomics data pipeline is then reviewed and adapted to combine multiple proteomic search engine results. This merging process increased the information extractable from the mass spectrometry data by 17% and was subsequently utilised throughout the three iTRAQ based proteomic investigations which form the remainder of the thesis.

The first proteomic investigation was on the model alga *Chlamydomonas reinhardtii* under nitrogen stress. Within which 58 and 280 proteins were confidently identified as being significantly changed during nitrogen stress induced carbohydrate and lipid production respectively. Key trends observed were an increase in energy metabolism, a decrease in translation machinery, an increase in cell wall production and a change of balance between photosystems I and II.

The second, again conducted on *C. reinhardtii*, assessed the effect of carbon source and availability. Utilising a turbidostatic bioreactor, algae were cultured photoautotrophically and mixotrophically with and without CO_2

supplementation. From the 995 confidently identified proteins the major observation was the distinction between carbon un-supplemented versus supplemented growth. Otherwise, the changes associated with the type of supplementation, CO_2 or acetate were minimal.

The third investigation assessed nitrogen stress in the diatom *Phaeodactylum tricornutum*. This facilitated a comparison of nitrogen stress in two microalgae from two distinct Divisions of life. Some 645 proteins were confidently identified as being significantly changed 24 h post nitrogen stress induction. Key trends observed were decreased translation, photosynthesis and fatty acid catabolism whilst observing increased central metabolism, fatty acid biosynthesis, antioxidants, endocytosis and phagocytosis.

The descriptive understanding gained from each of these postgenomic investigations provides greater insight into the effect of culture perturbations on algal culture. These insights can be the basis of new hypotheses and help in identifying possible targets for future manipulation and control of algal culture. The work contained in this thesis has increased the descriptive knowledge of 1028 proteins and described the culture conditions under which they are perturbed.

Contents

Acknowledgements

I am immensely grateful for being given the opportunity to study for a PhD and thus would initially like to thank my sponsor EPSRC and co-supervisors Dr. Raman Vaidyanathan and Prof. Phillip Wright for giving me this opportunity. I would also like to thank them both for considerable guidance and support throughout my studies.

I would also like to thank Dr. Jagroop Pandhal and Dr. Josselin Noirel whose support and guidance as well as friendship have been irreplaceable. For considerable guidance with mass spectrometers I must thank all the ChELSI mass spectrometry researchers past and present including Dr. Caroline Evans, Dr. Ow Saw Yen and Dr. Narciso Couto. The number of algal researchers has expanded during my studies at Sheffield and I would like to thank all members of the Algal Biotechnology Sheffield Network for their help in researching our photosynthetic friends. A special thanks goes to fellow algal PhD students Emily Hounslow, Danying Wu and Rahul Kapoore. In addition I would like to thank all my colleagues too numerous to list here who have helped and guided me in my research.

In addition to my studies I want to thank those who have supported me in life. My first thanks are to my parents Jim and Siân Longworth. You instilled the virtue of academic achievement and always encouraged my curiosity.

My PhD was aided in no small part by the friendships I gained through my studies including: Members of the Sheffield Sabres with whom I was encouraged to take out frustrations. Members of PROGRESS, especially Dr. Hugh Dannatt, Samuel Touchard and Rosanna Milner. Members of my Department with whom I have become good friends over the past few years, especially Stephen Jaffé, Nathan West, and Leon Pybus, thank you for all the good times.

Most of all I would like to thank Joanna Kremer. Her love and support throughout this endeavour has been amazing. Thank you honey!

List of Publications

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

(J – Journal Paper, P – Poster Presentation)

(J1) Longworth, J.; Noirel, J.; Pandhal, J.; Wright, P. C.; Vaidyanathan, S. HILIC- and SCX-based Quantitative Proteomics of *Chlamydomonas reinhardtii* During Nitrogen Starvation Induced Lipid and Carbohydrate Accumulation. Journal of Proteome Research 2012, 11, 5959–5971. (This work forms the basis of Chapter 5)

(P1) Longworth, J.; Kapoore, R.; Wright, P. C.; Vaidyanathan, S. Carbon source effects on *Chlamydomonas reinhardtii*: A proteomic & metabolic investigation.; ChELSI Conference: Sheffield, U.K., 2012.

(P2) Longworth, J.; Pandhal, J.; Noirel, J.; Wright, P. C.; Vaidyanathan, S. Nitrogen stress induced lipid production in the model alga *Chlamydomonas reinhardtii* – A proteomic investigation; EUPA: Glasgow, U.K., 2012.

(P3) Longworth, J.; Pandhal, J.; Wright, P. C.; Vaidyanathan, S. Proteomic analysis of *Chlamydomonas reinhardtii* under nitrogen limited growth; Biofuels C3: Singapore, Singapore, 2011.

Abbreviations

µE	MicroEinstein
µg	Microgram
µL	MicoLitre
µm	Micometer
1D	One Dimensional
2D	Two Dimensional
AC	Accession Number
ACN	Acetonitrile
CCAP	Culture Collection of Algae and Protozoa
ChELSI	Chemical Engineering at the Life Sience Interface
CO_2	Carbon Dioxide
DAVID	Database for Annotation, Visualization and Integrated Discovery
DCW	Dry Cell Weight
dH2O	Distilled Water
DIGE	DIfference Gel Electrophoresis
DMSO	Dimethyl Sulfoxide
DNA	DeoxyriboNucleic Acid
EDTA	Ethylenediaminetetraacetic
EGGNOG	Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups
ELISA	Enzyme-Linked Immunosorbent Assay
Err	Error
ESI	ElectroSpray Ionization
et al.	et alii (and others)
EUPA	EUropean Proteomics Association
FA	Formic Acid
FDR	False Discovery Rate
g	Gravity
GM	Genetically Modified
GO	Gene Ontology
h	Hour(s)
HILIC	Hydrophilic Interaction Liquid Chromatography
HPLC	High-Performance Liquid Chromatography
IC	Isotope Correction
ICAT	Isotope-Coded Affinity Tags
iTRAQ	Isobaric Tags for Absolute and Relative Quantitation
KEGG	Kyoto Encyclopaedia of Genes and Genomes
L	Litre
LC	Liquid Chromatography
m	Meter
m	Mass
m/z	Mass divided by Charge
MALDI	Matrix-Assisted Laser Desorption/Ionization

MC	Median Corrected
mgf	Mascot Generic Format
min	Minute
mL	Millilitre
mM	Milimolar
mm	Millimetre
MMTS	Methyl Methanethiosulfonate
MS	Mass Spectrometry
MudPIT	Multidimensional Protein Identification Technology
NADP	Nicotinamide Adenine Dinucleotide Phosphate
nm	Nanometer
OMSSA	Open Mass Spectrometry Search Algorithm
PC	Principal Component
PCA	Principal Component Analysis
PCC	Pasteur Culture Collection
PH	Phenyx
PK	Peaks
PP	ProteinPilot
ppmv	Parts Per Million by Volume
PS	Peptide Shaker
PSM	Peptide Spectral Match
QTOF	Quadrupole Time-Of-Flight
QUICK	Quantitative Immunoprecipi-tation Combined with Knockdown
RNA	RiboNucleic Acid
RNA-SEQ	RNA Sequencing
ROS	Reactive Oxygen Species
RPLC	Reversed-Phase Liquid Chromatography
RuBisCO.	Ribulose-1,5-Bisphosphate Carboxylase Oxygenase
s	Second
SCX	Strong Cation Exchange (Chromatograph)
SDS	Sodium Dodecyl Sulphate
SILAC	Stable Isotope Labelling with Amino Acid in Culture
Sp.	Species
TAG	Triglyceride
TAP	Tris Acetate Phosphate
TEAB	TriEthylAmmonium Bicarbonate
TMT	Tandem Mass Tags
TOF	Time Of Flight
USA	United States of America
UV	UltraViolet
V	Volts
v/v	Volume for Volume
vvm	Volume for Volume per minute
VSN	Variance Stabilization Normalisation

List of Figures and Tables

Chapter 2

Figures

Tables

Chapter 3

Figures

Chapter 4

Figures

Tables

Chapter 5

Figures

Tables

Chapter 6

Figures

Tables

Chapter 7

Figures

Tables

Chapter 8

Figures

Tables

Chapter 1: Introduction

Chapter 1 – Introduction

1.1: Research Impact

"We are like tenant farmers chopping down the fence around our house for fuel when we should be using Nature's inexhaustible sources of energy — sun, wind and tide. I'd put my money on the sun and solar energy. What a source of power! I hope we don't have to wait until oil and coal run out before we tackle that."

Thomas Edison (1847-1931)[1]

As noted by Thomas Edison in conversation with Henry Ford, our current energy usage is unsustainable. The use of fossil fuels which are effectively derived from ancient sunlight harnessed millions of years ago is inherently unsustainable. Further, the disruption of the carbon balance caused by releasing stored carbon to the atmosphere as CO_2, a greenhouse gas, is believed to be the predominant cause of man-made climate change. For these reasons we should strive to, as Edison states, tackle solar energy, preferably without waiting for oil and coal to run out.

The derivation of energy from current solar emissions can take a variety of forms. It can be indirectly harvested, such as by wind and wave, or directly absorbed by photovoltaic. One promising harvest strategy is to utilise naturally occurring photosynthesis. Photosynthesis converts solar radiation to energy storage molecules. By optimising the cultivation and retrieval of these energy storage molecules a renewable energy source can be derived.

The use of algae as the photosynthesis organisms is supported by their relatively fast growth compared to arable plants and high lipid contents, even when grown photoautotrophically. They, unlike other biofuel crops, should not compete with food and textile production.

Stress induction has become a common tool for the increase in cellular carbon storage, such as carbohydrate and lipid. Though not universal many algae will hyper-accumulate lipid under stress condition. Observed stresses include: nitrogen limitation, iron limitation, sulphur limitation, phosphorus limitation, silicon limitation, salt increase and UV exposure. Whilst these stresses may increase carbon storage they typically have a negative effect on cell growth. Whether this combines to an overall increase in culture carbon source accumulation is dependent on the species and stress involved.

The optimisation of the carbon source accumulation is desired by the algal biofuel community, in particular, lipid accumulation. Such optimisation would be benefitted by greater understanding of the internal mechanism within the cell. It could also be advantageous to identify an artificial induction mechanism with reduced negative effect on growth.

1.2: Systems biology

To optimise the use of algae, more is needed to be understood of the various organisms which comprise this wide ranging term. Taking a top down approach, such as an 'omics' based investigations, can provide an overview of an organism. Such systems biology investigations can be conducted at many 'omic' levels including: proteomics, metabolomics, transcriptomics, and posttranslational modification investigations, such as those on the acetylome and phosphoproteome. However it is the proteome that represents the best starting point of a systems biology investigation. This is best accomplished at the proteomic level, given that proteins are the active machinery of the majority of life, the mode by which the cell attempts to influence its internal and external environment. Unlike transcriptomics, proteomics gives insight into active cell protein content directly rather than by inference.

In an ideal situation a proteomic investigation would provide a list showing the absolute quantity of each protein. However, due to the complexity of the

proteome and limitations of the techniques and time this is rarely achievable. As a compromise to the ideal each protein is reported with three key figures the significance of identity, the significance of change and the degree of change (absolute or relative).

- Such a list of proteins is in essence the results content of any proteomic investigation. From this there are three main ways in which the results can be utilised: Direct observation: Identifying proteins which are of interest due to the particular nature of the investigation based on the background literature and data of the protein.
- Trend analysis: Using classifications systems such as Gene Ontology or Pathway mapping such as KEGG the overall trend of multiple observed proteins can be commented on.
- Data release: By supplying the observed proteomic changes in the correct interrogable manner to the appropriate repositories third party investigations can be supported. This can be through additional descriptive data relating to the protein held of various repositories or investigations which source data from multiple previously reported investigations.

Given the lack of a predefined hypothesis, proteomic investigations can often be defined as a "fishing trip". Whilst it can be argued this can lead to an untargeted investigation it also allows for an unrestricted investigation. Such lack of restriction improves their usefulness as hypothesis deriving investigations. It is this ability to derive unpredicted hypothesis which is the strength of proteomic and indeed omic investigations in general. Due to the specific resources and skills required for proteomics investigations it is not uncommon for targeted investigations hypothesised as a result of a proteomic investigation to be conducted within different research group to that of the proteomic investigation itself.

1.3: This Investigation

Within proteomic investigations there are a variety of tools which can be used. As will be discussed in the literature review these include intrinsic isotopic labelling of peptides themselves, extrinsic labelling with isotopic tags and label free methodologies. Isobaric tags for relative and absolute quantitation (iTRAQ) is an extrinsic isotopic tag methodology. It has become popular in proteomics due to its, lack of need of special culture regimes, ability to be processed on a reasonable quality mass spectrometer without the need of run to run reproducibility that is essential for label free proteomic methods and with the 8-plex variant eight channels of investigations can be run simultaneously. As such multiple conditions or increased numbers of replicates can be assessed simultaneously. To attain the most information out of this technique an initial investigation into the data processing methodologies available will be conducted.

Within current investigation of algae one species and culture perturbation stand out as being of extremely high interest to the community. This is the Chlorophyta *Chlamydomonas reinhardtii* and nitrogen stress. *C. reinhardtii* as the predominant model alga has been sequenced and is the most characterised alga. Nitrogen limitation is the most noted of lipid induction stress and usually derives the greatest carbon source accumulation. For this reason this organism and this stress were selected for the initial investigation.

A noted query with many investigations into *C. reinhardtii* is the carbon source. Able to grow on acetate, mixotrophic culture of this alga is common laboratory practice. It is expected however any culture grown for energy harvest will be grown photoautotrophically. Hence the effect of carbon source is selected as the second investigation.

The term algae covers a wide range of species. The assumption that trends from one will also be found in others is often made, erroneously. To further

the understanding of the proteomic effect of nitrogen stress, further studies were desired in other species of algae. The diatom species *Phaeodactylum tricornutum* was chosen as a divergent species from *C. reinhardtii* with the necessary genomic sequence required for proteomic investigation.

Aims and Objectives

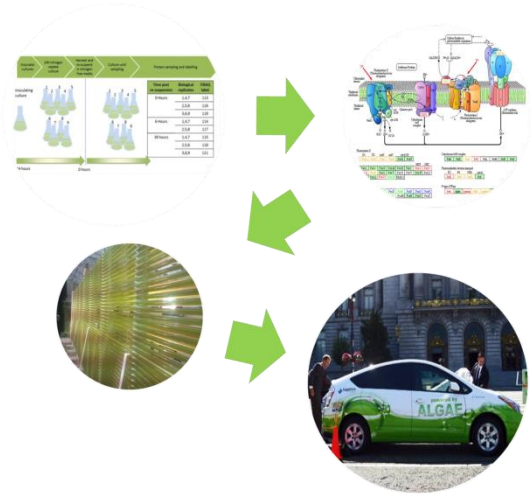
Aim

To increase understanding of microalgae and their potential use as a biofuel feedstock, through the use of quantitative proteomics.

Objectives

1. To optimise the processing of mass spectrometry data to maximise proteomic information.
2. To assess the proteomic changes in *C. reinhardtii* when grown photoautotrophically and mixotrophically.
3. To assess the proteomic changes in two divergent microalgal species under nitrogen induced lipid accumulation.

Chapter 2: Literature Review



Chapter 2 – Literature Review

2.1: Microalgal based biofuels

2.1.1: Problem with fossil fuel

The need to migrate to an energy economy based on renewable energy production rather than non-renewable fossil fuels has become established. This has been described for a variety of reasons including social economic and environmental.

Regarding social economic concerns, the need for energy security has been shown with fossil fuels producing countries able to project financial and political pressure on fossil fuels dependent countries.[2–6] Despite the recent extensions and denials,[7,8] 'Peak oil' is still an inevitable reality. It may be better understood as 'Peak cheap oil'[9,10] which incorporates the understanding that though identification of potential resources increases with improved extraction technology, the cost to retrieve such resources also increases.

Environmental issues revolve around the effect of burning fossil fuels on climate change.[11] Climate change as a phenomenon has been overwhelmingly supported in the literature as a man-made phenomenon with 97% of peer reviewed papers supporting such a link.[12,13] The major contributing factor, the burning of fossil fuels. Such combustion releases CO_2 into the atmosphere. CO_2 is a known greenhouse gas,[14] and its increase in the atmosphere is therefore linked to perturbations of climate. The increase in atmospheric CO_2 has been monitored over the past 50 years with the Keeling data[15] (Figure 2.1B_{p31}) and using geological observations can be observed over the past 1000 years (Figure 2.1A_{p31}). In addition to CO_2, burning of fossil fuels releases methane, nitrous oxide, ozone-depleting substances, hydrofluorocarbons, sulphur hexafluoride and perfluorocarbons, all of which have been linked to climate change.[16]

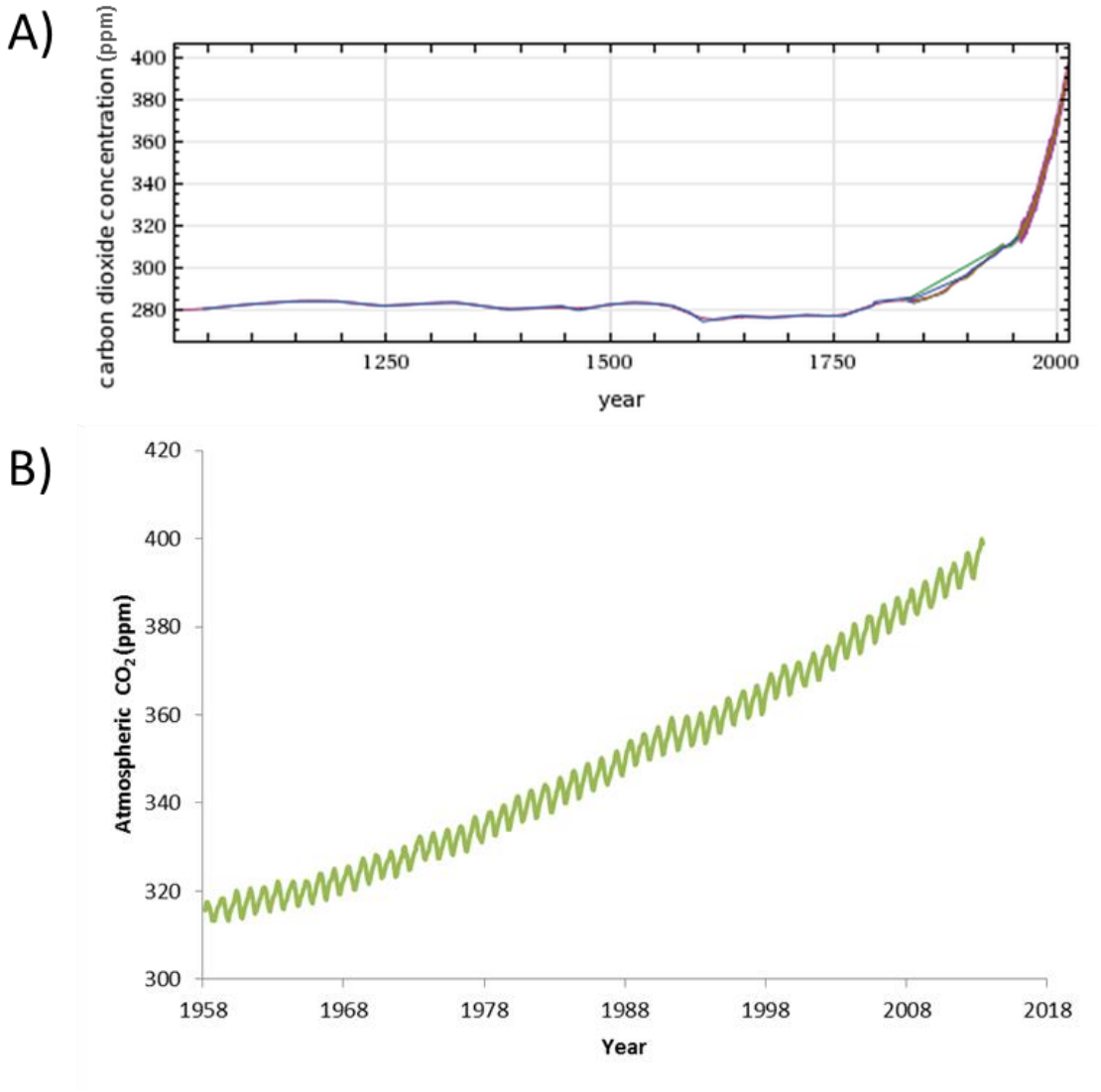


Figure 2.1: Graph of atmospheric CO_2 levels over time given for an extended 1000 year span and a recent 50 year span. (A) was created using Wolphram Alpha B) using Keeling data[15])

2.1.2: Alternative energy sources

There are several alternatives to fossil fuels. This includes nuclear, solar, tidal/wave, geothermal, wind and biofuels. Apart from biofuels all these methods are, however, designed for electricity as the medium of energy. Given that 66%[17] of world energy demand is in the form of fuels (i.e. gaseous natural gas, liquid petroleum and solid coal) rather than electricity a suitable replacement for fuels is also desirable. Possible candidates for this fuel are

biodiesel and bioethanol. Biodiesel and bioethanol being produced from lipid and carbohydrate content of photosynthetic organisms.

2.1.3: 3rd generation biofuels

Biofuels can be classified as first, second or third generation.[18] First generation biofuels refer to those derived from high energy components of biomass. These generally fall into two types: corn and sugar cane, which are typically processed to provide ethanol as the transportable fuel or oleaginous crops such as palm and *Jatropha* processed into biodiesel. These sources frequently require the displacement of food production, thus causing the food/fuel dilemma,[19] an increasing source of humanitarian concern. Second generation biofuels have focused on waste and low energy biomass material, typically lignocellulose material such as that obtained from corn stalks and *Miscanthus.*[20] These second generation biofuel feedstocks are superior to first generation sources with reduced impact on food and textile cultivation.[21] However, they require the conversion of complicated energy stores such as lignocellulose and starch. Though considerable research is being conducted to increase the efficiency[22–25] of such conversions, it is currently still energy intensive and both energetically and financially uneconomical.[26–29] Third generation biofuels have mainly focused on production from algae and in particular microalgae. Like second generation biofuels, these feedstocks are envisaged not to contribute to the food/fuel dilemma taking advantage of the microalgae's ability to grow on non-arable land. Energy from microalgae is in the form of high energy storage compounds naturally found in algae, namely lipids and carbohydrates,[30] thus do not have the conversion problem associated with second generation fuels.

2.1.4: From microalgae to biofuel

The processing of microalgae into biofuels has been repeatedly reviewed.[17,31–45] The major cycle is shown in Figure 2.2$_{p33}$ and can be envisaged as four steps:

1. Cultivation: The microalgal culture is grown, typically in photobioreactors[46,47] or open ponds.[48,49] This growth should be photoautotrophic, converting CO_2 to biomass to be a truly cyclical fuel.
2. Harvest: The microalgal culture being >90% water must be de-watered and refined before being utilised as a biofuel.[44,50,51] This involves de-watering[52] and lipid extraction.[53] Though typically conducted in this order there are proposals which involve lipid extraction from the cells prior to de-watering.[54,55]
3. Conversion: Upon extraction of the lipid content, processing is still required to convert the lipid to a fuel such as biodiesel suitable for combustion in standard or modified engines. This can be achieved by transesterification with methanol, though other processes are being developed.[56]
4. Combustion: Using either a standard or modified internal combustion engine the derived biodiesel is used, returning CO_2 to the atmosphere and completing the cycle.

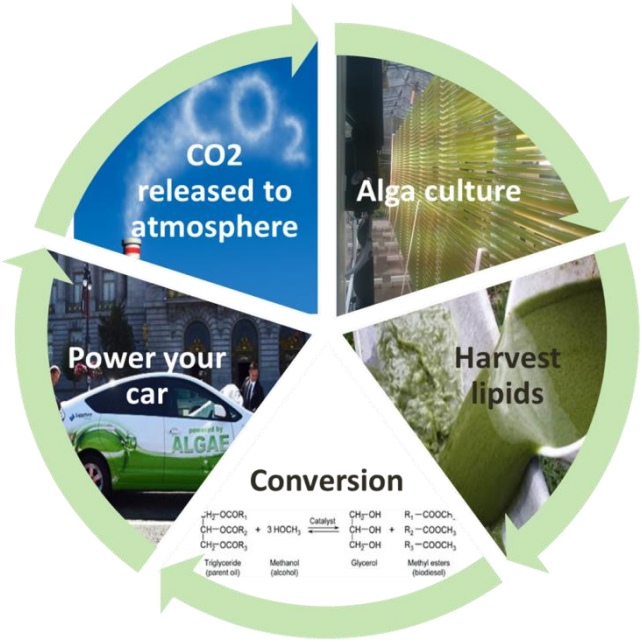


Figure 2.2: Diagram showing the cyclical nature of microalgae derived biofuels (image was self-modified [31,57–59]).

Lipid production rates of these single-celled photosynthetic organisms exceed that of traditional terrestrial crops,[60] and can be grown photoautotrophically without the need for organic sources of energy or carbon.[61] Microalgae are typically cultivated either in photobioreactors or open raceway ponds.[46] The major cost of production is in the dewatering of the culture and extraction of lipid from the microalgae.[51] Yields of lipid from algae can be as high as 66%,[62] but these vary vastly between different species and different growth conditions.[63]

Estimates of land requirement and cost vary dramatically. For land requirement, Borowitzka and Moheimani[64] estimate between 245 and 653 ha for the production of 100,000 bbl per year (depending on productivity and lipid content). With global oil demand in 2008 at 84.3 x 10^6 per day a growing area between 750 thousand and 2 million square kilometres would be required to meet 2008 levels of consumption. Typically estimates of cost are between \$0.42[65] and 2.80[66,67] per litre of biodiesel. This is suggestive that the price may become competitive with the \$0.84 per litre for wholesale diesel[68], and if given non punitive tax would likely be able to compete with the post-tax at pump price of \$2.17.[68]

2.1.5: Species selection

The single term microalgae is often used as a collective classification of microscopic photosynthetic organisms. It is important however not to misinterpret this collective as a similar group of organisms. Microalgae can refer to a wide variety of different single celled photosynthetic species including prokaryotic cyanobacteria and eukaryotic Chlorophyta, Rhodophyta and Bacillariophyceae (also known as diatoms). The eukaryotic algae are generally classified based on their photosynthetic pigments, but represent considerable evolutionary separation. Various terms are used to name sub-groups of algae including; blue-green algae referring to cyanobacteria, Chlorophyta (green algae) referring to those belonging to the

Plantae kingdom and diatoms referring to those algae containing silica cell walls.[69]

As noted by Klein-Marcuschamber *et al.,*[70] it is important not to rely on observations from one algal species being applicable to all algal species. The diversity between even relatively similar Chlorophyta such as *Chlamydomonas reinhardtii* and *Chlorella vulgaris* is considerable.[71,72] To compare, humans, mice and zebra fish share a 52.8% orthologarity, whilst *C. reinhardtii, Thalassiosira pseudonana* and *Cyanidoschyzon merolae* share only a 16.3% orthologarity.[73]

Given the wide variety of algae to choose from, an important decision is which species to use for production of biofuels. There are many factors which affect the suitability of a particular species for production. These can include whether the species is a fresh water species or marine, the degree light penetrates through culture, factors affecting its harvesting and the lipid profile. Perhaps one of the most important factors however is the lipid production rate, a product of the microalgal growth rate and the lipid content. This question of species selection was one of the goals for the Aquatic Species Programme,[36] a large research effort that operated for almost 20 years. Conducted by the USA Department of Energy, the programme investigated the feasibility of biodiesel production from microalgae and CO_2 sequestration. Despite extensive research effort, no clear candidate species was identified and species selection remains a key research goal in many current algal research projects.[74–76] Despite the uncertainty, there are some general trends, such as the preference for eukaryotic algae rather than cyanobacteria which typically have lipid contents below 10% DCW (dry cell weight).[39] Amongst the Eukaryotic species, there have emerged some common contenders, either for production or as a model to understand the use and manipulation of microalgae for lipid production (Table 2.1$_{p36}$).

Table 2.1: Table of various microalgae species and the genomic availability, typical growth rate, oil content and nitrogen effect. (Phylum: He-Heterokontophyta; Ha-Haptophyta; C –Chlorophyta; R-Rhodophyta) (Class): Ch –Chlorophyceae; Pr-Prymnesiophycea E-Eustigmatophyceae; Ba-Bacillariophyceae (Diatoms); tr-Trebouxiophyceae; Pra- Prasinophyceae; Cya-Cyanidiophyceae)

Phylum	Class	Genus	Degree of genomic data[77]	~Doubling time (h)	Fresh/ Marine	Lipid % DCW	Nitrogen stressed Lipid % DCW
C	Ch	*Botryococcus braunii*	Partial	72.[36]	Fresh	23[78]11[79]38.6[80]25[31]	55[36]21[79]61.4[80]75[31]
C	Ch	*Chlamydomonas reinhardtii*	Complete	8[63]	Fresh	21[63]	Non-increasing[36]
C	Ch	*Dunaliella salina*	Soon to be published[81]	72.6[82]	Marine	19[63]28.1[78]33[82]	10[63]
C	Ch	*Dunaliella tertiolecta*	None		Marine	15[63]	18[63]
C	Ch	*Neochloris oleoabundans*	None		Fresh	17[83]	37[83]
C	Ch	*Haematococcus pluvialis*	none	21[84]	Fresh	15.61[84]	
C	Ch	*Scenedesmus obliquus*	Partial		Fresh	21[63]	42[63]
C	Ch	*Tetraselmis suecica*	None	36[63]	Marine	17[63]	26[63]
C	Ch	*Volvo carteri*	Complete[85]		Fresh		
C	Pra	*Ostreococcus* Sp.	Partial		Marine	27[63] 27[86]	50[63] 49.5[86]
C	Pra	*Ostreococcus lucimarinus*	Complete		Marine		
C	Pra	*Ostreococcus tauri*	Complete	~34[87]	Marine		
C	Pra	*Micromonas pusilla*	Complete		Marine	13.3[88]	
C	Tr	*Chlorella* Sp.	Partial	17-28[62]	Fresh	18.7[89]32[62]	66[62]
C	Tr	*Chlorella vulgaris*	Partial	20[63]	Fresh	25[63]	42[63]
Ha	E	*Pleurochrysis carterae*	None	45[82]	Marine	32[82]	
Ha	Pr	*Emiliania huxleyi*	Partial		Marine	23[78]	
Ha	Pr	*Isochrysis galbana*	Partial	48[90]22[63]	Marine	25[63]	29[63]
Ha	Pr	*Pavlova lutheri*	Partial	13.3[91]	Marine	35[63] 12,8.8[78]	
He	Ba	*Phaeodactylum tricornutum*	Partial	27[63]	Marine	21[63]	26[63]
He	Ba	*Thalassiosira*	Complete	12[63]	Marine	24.3[86]16[63]	26.6[86]26[63]

		pseudonana					
He	E	*Nannochloropsis* Sp.	None	28[63]	Marine	31[63]	41[63]
He	E	*Nannochloropsis oculata*	In progress[81]		Marine	13.2[78]	
He	E	*Nannochloropsis salina*	None	18.5[91]	Marine	16.9[78],27[63]	46[63]
R	Cya	*Cyanidioschyzon merolae*	Complete[92]		Fresh		

2.1.6: Lipid profile

When discussing lipid production from microalgae, it is not only important to consider the quantity, but also the type of lipids. Not all lipids are suitable for biodiesel production. Various chain lengths and degrees of unsaturation are required to provide the correct ignition quality, cold flow and oxidative stability.[39] To regulate these properties, both the USA and EU have limits on aspects such as amount of unsaturated fatty acids allowable in biodiesel measured by its iodine value (e.g. EN 14214 and EN 14213).[61]

The lipid profile of different species of microalgae varies significantly between species and modification such as nitrogen stress will also induce dramatic changes in the lipid profile. Typically microalgae lipids have a chain length of between C16 and C18 with high levels of saturated and mono-unsaturated C16 and C18.[39] If these profiles could be manipulated to contain shorter chain fatty acids between C12:0 and C14:0 it would improve the quality of biodiesel produced.[39,61] This could theoretically be achieved by manipulation of the acyl carrier proteins and acyl carrier protein thioesterases.[93]

2.1.7: Biochemical manipulation of microalgae

For over twenty five years, it has been recognised that a simple method to improve lipid content of microalgae and thus biodiesel production was to limit the nitrogen source within the culture.[94] This nitrogen limitation effect has since been demonstrated to be a general but not universal trend within microalgae, with it being observed more in Chlorophyta than diatoms.[86] The lipid content of several algal species, both under nitrogen replete and deplete

conditions, is included in Table 2.1$_{p36}$. As well as nitrogen starvation, similar perturbations of the culture medium such as phosphorous limitation,[95] iron limitation,[96] silicon limitation,[63,97] salt stress[98] and UV radiation[99] have also been shown to increase lipid content.

These culture based modifications to the algal product can be classified as biochemical engineering.[100] Whilst being relatively simple to implement and potentially less regulated compared to genetic engineering, there are numerous negative side effects. Stressed cells may produce greater lipid content but will often have reduced cell growth rates.[101] As lipid productivity is a product of lipid content and growth rate both influence total culture productivity. So whilst nitrogen stress may increase lipid content its side effect of decreased growth rate is greater commonly leading to an overall reduction in culture productivity. It is envisaged that through better understanding of how biochemical manipulations work an optimal balance between lipid content and growth rate could be developed.

Increases in lipid content under stress could be a possible survival mechanism. Given an environment limiting propagation, cells may have evolved to increase energy stores. These would then give the cell a boost upon return to conditions favourable for propagation. Another hypothesis for the storage of lipids beyond that of simple carbon storage is that the accumulation of long chain polyunsaturated fatty acids aids the rapid construction of polyunsaturated fatty acid rich chloroplast membranes required upon return to normal reproductive levels.[102] Further still, the increase in lipid can affect the buoyancy of the organism and thus their location within the water column[103]. Such a movement may potentially increase nutrient supply of the deprived component.

2.1.8: Genetic manipulation of microalgae

Beyond crude biochemical manipulation, it is often suggested that genetic engineering could increase lipid production rates.[19,32,93,100] This, it is

envisaged, will provide the increased levels of lipid with reduced negative growth aspects experienced in biochemical engineering. Maximal productivity therefore is a balance between lipid content and growth rate.

Genetic engineering has been performed in a various microalgae species.[104] Though it is more advanced in Chlorophyta,[105–108] manipulations have also been achieved in diatoms[109,110] and red algae.[111,112]

There are multiple possibilities for the modification of microalgae to improve the production of the desired biodiesel product. These can be broken down into those which increase:

- the rate of solar energy conversion and thus energy available for storage.
- the rate of carbon fixation
- the lipid concentration of the algae.
- quality of the produced lipid.
- secondary factors affecting the cost of production such as ease of harvest.

Increases in energy conversion can be attempted by increasing the productivity of the photosynthetic apparatus and optimisation of metabolic pathways such as increasing the levels of ribulose-1,5-bisphosphate carboxylase.[46] Whilst increased productivity for a given alga may be achieved by increasing the number of photosystems, for the collective culture decreasing the number of photosystems may be more productive. This is due to the effect photosystem number and type has on light penetration. The light penetration within a dense culture, which is preferable for harvesting purposes, can be a limiting factor for growth.[113] This would be inconsequential if the culture depth is shallow enough in raceway ponds or culture chambers thin enough in photobioreactors. However, culturing apparatus with shallow depth may have other related problems such as culture flow and increased footprint requirements. Reduction of the algae's

antenna size and pigment production has been demonstrated to increase such light penetration.[114,115]

As well as providing sufficient solar energy throughout the culture, tolerance to excess irradiation can increase productivity by preventing photoinhibition. Photoinhibition,[116] an effect caused by oxidative damage from the oxygen produced during photosynthesis, can also be manipulated to increase the energy available for conversion to lipid. Due to photoinhibition, even at a relatively low illumination levels, photosynthetic productivity decreases.[31] If this effect could be reduced, the energy conversion of sunlight could be increased.

Limited investigations have been conducted to manipulate the lipid content of microalgae as the understanding of lipid synthesis in microalgae is incomplete. Some investigations[117–120] have been conducted using inferences based on understanding of lipid synthesis in other organisms such as higher plants and yeast.[93] Three strategies, which could be used to increase lipid content, are detailed in the following paragraphs.

Firstly, the synthesis of lipids can be increased metabolically as was done by Dunahay *et al.*[120] Acetyl-CoA carboxylase, hypothesised to be a key enzyme in the fatty acid biosynthesis pathway, was over-expressed. Despite successful transformation of the diatom *Cyclotella cryptica*, where a 2-3 fold increase in acetyl-CoA carboxylase expression was achieved, no significant increases in lipid levels were identified.[120] Other attempts that modified enzyme expression in organisms other than microalgae, such as increasing glycerol-3-phosphate dehydrogenase in *Brassica napus* (oil seed rape) has been shown to increase lipid content by 40%.[121]

The second method is to decrease the use of alternative energy stores, such as starch. This has been demonstrated with significant success in *C. reinhardtii* using the starch synthesis mutant *sta6*, which lacks glucose-1-phosphate adenylyltransferase an enzyme involved in starch synthesis.

Wang *et al.* [119] demonstrated a 30 fold increase in lipid bodies 48 hours after nitrogen stress opposed to a 15 fold increase in non-*sta6* mutants after nitrogen stress.

The third method to increase lipid content is to decrease lipid catabolism. Though not demonstrated in microalgae, this has been achieved in the higher plant, *Arabidopsis thaliana*. Fluda *et al.* [122] demonstrated an increase in lipid content on inactivation of the peroxisomal long-chain acyl-CoA synthetase. Though this has potential to increase lipid content, manipulation of lipid catabolism genes may have deleterious side effects, such as the improper seedling development seen in Fluda *et al.*'s study.[93]

There are also various manipulations that might be desired to decrease the cost of microalgae production. These include manipulating properties such as flocculation which can dramatically reduce de-watering costs, a considerable portion of production costs.[51] In the future, it might be desirable to develop the microalgae to actively secrete lipids themselves such has been demonstrated in the cyanobacterium S*ynechocystis* sp. PCC6803.[123]

When considering genetic manipulations, it should be noted that lipid productivity is a product of the lipid content and growth rate under any given condition. Thus, when considering the genetic manipulation of microalgae for the reasons discussed thus far the effect on growth rate must also be considered. Total culture yield may not, however, be the optimum as considerable cost is involved in de-watering and extraction. Thus, a lower total lipid yield per cell in a higher concentration may be more profitable.

It is still unknown to what effect governmental regulation may have on the use of genetically modified (GM) algae for biodiesel production. This is a complex legal area, especially when considered on the global scale, though perhaps not as controversial as GM food crops. Without the concern of introduction into the food chain GM fuel crops may be better received.

However, GM fuel crops will still come under criticism due to the difficulty of containment.[124,125] This regulatory hurdle is one reason why identification of natural species and utilisation of biochemical manipulation could be favoured. However, greater understanding of the biochemical manipulation may lead to optimised implementation without the need for genetically modified organisms and the resultant regulation.

2.2: Proteomics

2.2.1: Case for increased systems level understanding of microalgae

From the literature, it is clear that biochemical engineering approaches can be used to increase lipid contents. However, this is commonly linked to decreases in growth rates. The degree to which the increase in lipid is a simple re-direction of consequential excess energy production is unknown. To aid advances in both biochemical and genetic engineering methodologies, further understanding of the microalgae on a molecular biology level is desired.

Transcriptomic[126] (study of all mRNA transcripts within the cell typically to make inference on proteomic changes), proteomic[127] (study of all proteins within the culture) and metabolomic[128] (study of all the low molecular weight (<1500 Da) metabolites within the cell) investigations, the bases of systems biology, allow for increased understanding of the changes occurring to the functional entities within the cell itself rather than observing changes on a culture and bulk biomass level. Systems investigations provide a broad level of understanding that can then facilitate numerous hypothesis driven investigations.

Transcriptomics is often used to quantitatively predict the relative abundance of the corresponding proteins, inferring that protein abundance is predominantly a product of transcriptional changes. Protein abundances however, are also affected by variations on the translation level, splice

variants as well as degradation rates. This can lead to incorrect assumptions of protein abundances based on transcription abundance data.[129,130] Further still, the protein presence does not directly demonstrate protein activity as many post-translational modifications are used to regulate protein function.

Modern quantitative proteomics techniques largely based on mass spectrometry (MS), are allowing direct quantitative assessment of protein abundances rather than relying on inferences from the transcriptional level.[131] It should be noted though, that the decreased coverage[132] of these proteomic techniques means that transcriptomics can sometimes provide insight not observed in limited coverage proteomic investigations. The choice of which method to employ is often a matter of equipment and particulars of the condition under investigation.

Metabolomics investigations attempt to identify all low molecular weight components (< 1500 Da) of a cell.[133] This can give insight into the processes within the cell[134]. Transcriptomics and proteomics use the genomic sequence to derive theoretical transcript and protein sequences. Further, though extremely varied the base unit of all transcripts is a chain of only 4 possible bases and for proteins a chain of only 20 possible amino acids. Metabolomics however, can be an extremely heterogeneous[135] mix of thousands of compounds. Metabolome data can be used to derive flux balance analysis, such as Boyle *et al.*[136] Metabolomics technologies are developing,[137] but are best used in conjunction with targeted studies or in transcriptomic and proteomic investigations which have better predisposition to pathway mapping.

By understanding the changes within the proteome and metabolome that regularly occur, under lipid inducing stress conditions, several hypotheses should present themselves. For example, proteins whose abundance under a particular stress condition is changed are presumably involved with the stress response itself or involved in the increased lipid accumulation. Whilst

proteins with hypothesised function may be identified as relating to lipid accumulation or general stress undescribed or ambiguously described proteins may be more challenging. If however, these proteins are seen in multiple perturbations which induce higher lipid contents then this will suggest the proteins are related to lipid synthesis rather than the individual stress response.

Understanding which proteins are required for increased lipid accumulation, which are likely to be homologous in multiple strains of algae, could facilitate in their manipulation to induce over expression in the lipid production pathways. This may be possible without or with reduced negative effects observed with crude biochemical manipulation.

Furthermore, a reverse engineering approach to genetic engineering based on proteomic understanding can often identify which proteins are bottlenecks.[138,139] Increasing a specific protein will often not result in an increase in the pathway product. Instead, a subsequent or previous metabolite or protein becomes the rate limiter to the pathway.[140] Such a situation is a possible cause of failure in the investigation by Dunahay *et al.*,[120] which was based on modification of acetyl-CoA carboxylase to increase lipid levels.

With a systems biology approach utilising comparative proteomics, theoretically all the protein changes that occur to increase lipid levels can be identified. This can identify multiple changes that may be required to perform the desired manipulation without any preconception of their involvement. Proteomics offers the ability to better understand the complex dynamic nature of cellular metabolism, allowing identification of the multiple, rather than single modifications, which may be needed to achieve the desired manipulation.

The use of proteomics for the work described in this thesis was selected over transcriptomics due to the poor accuracy of the transcriptome to reflect true

protein conent.[132,141] This was predicted to be especially problematic during nitrogen stress with re-allocation of nitrogen from existing protein content a potential mechanism of regulation. Further, proteomics was selected over metabolomics as proteins can be described as drivers of cellular changes and as such have a greater potential for understanding the processes in the cell whilst under active modification. This is not to suggest that transcriptomics and metabolomics are not able to provide significant insights in such investigation but rather the proteomic level is often the best omic level to process first. Further, from review of the field at the start of this project it was evident that modern proteomics within algae had only had limited use and as such investigations of this sort would provide important information for the algal community.

2.2.3: Proteomic technologies

With the advent of the post genomic era, the investigation of proteins could shift from a reductionist approach to a systems biology approach based on 'omic' level investigations. Reductionist biology had previously investigated the protein content of pre-selected proteins under various conditions. This was often achieved with the use of antibody labelling techniques such as western blots[142,143] and ELISA assays.[144] In the post-genome era, theoretically the entire proteome, or a sub proteome, can be investigated without any preconceptions of which proteins to investigate. Though in theory any given protein can be identified, the realistic coverage varies greatly depending on: the level of expression in the organism under any given condition, the amenability of any protein to the extraction methods used and the dynamic range of proteins within the extracted proteome.[145]

2.2.4: Gel based proteomics

Initial proteomic studies largely derive from the two dimensional gel electrophoresis (2D-gel) investigations first described 35 years ago.[146] 2D gel studies involve the use of separation on an acrylamide gel in two dimensions, typically by *p*I and molecular weight. This separates a collection

of proteins so that upon staining single or a small number of proteins can be identified as a distinct spot. By comparison of the spots, intensities from one condition to another, the changes in protein abundance can be deduced with spots being identified by various sequencing techniques. In these, low throughput, 2D gel investigations only a limited number of proteins are taken forward for identification each being run sequentially such that 2D gel proteomics is labour-intensive and has a relatively low throughput.[147]

2.2.5: Modern proteomics

Modern proteomics refers to techniques that utilise the mass spectrometer both for identification and for quantitation, either absolute or relative.[148,149] This combination can increase the accuracy of quantitation, and sensitivity of detection for proteins with relatively lower abundance proteins. With modern proteomics, multiple proteins can be analysed in a mixed manner (i.e. multiple samples run together with different labels) reducing mass spectrometer run time. Modern proteomic techniques can be broken down into three types;[150] isotope labelling,[151] chemical tag[152] and label free.[150]

Isotope labelling relies on the identification of heavy and light isotopes of various elements (e.g. $^{12}C/^{13}C$, $^{14}N/^{15}N$, $^{1}H/^{2}H$ for each incorporation in the reported *m/z*). Isotope labelling can be further divided into sub-groups: metabolic labelling[151] and stable isotope labelling with amino acid in culture (SILAC).[153] Metabolic labelling, aims to replace all instances of a given element within an organism and thus all amino acids. This can be achieved by growth of the culture on labelled feedstocks, such as [^{13}C]glucose or [^{15}N]H_4Cl.[151] SILAC[153] however, is targeted to labelling only proteins. By supplementation of the culture medium with isotopic amino acids each protein can be identified with a variation of 1 mass unit for each of the relevant amino acids in the peptide *m/z*. When used with cultures auxotrophic to the given amino acid, suitably complete incorporation can be achieved. Both these methods allow for a two sample comparison in a given mass spectrometry run. They provide relative assessments of proteins but

can also be combined with a spiked standard to give absolute quantifications.

Chemical tags typically also use isotopes but rather than labelling being conducted *in-vivo* as with isotopic labelling the labelling is completed subsequent to protein harvest *in-vitro.* There are multiple labelling varieties available, but four commonly used are di-methyl labelling,[154] isotope-coded affinity tags (ICAT),[155] tandem mass tags (TMT(s)),[156] and isobaric tags for absolute and relative quantitation (iTRAQ).[157] These four can then be further separated into fragmenting and non-fragmenting.

Dimethyl labelling and ICAT are non-fragmenting chemical tags. They consist of a linker, which binds to the peptide and the isotopically varied reporter. In the case of dimethyl labelling (Figure 2.3A_{p48}), there are three versions; standard isotopes, two ^{2}H and three ^{2}H two ^{13}C. In the case of ICAT (Figure 2.3B_{p48}), there are two versions heavy 'd_8' and light 'd_0'. 'd_8' having 8 deuterium with 'd_0' having 8 regular hydrogen atoms. In addition, ICAT has a biotin label allowing labelled peptides to be purified and enriched.

Being non-fragmenting when analysed by mass spectrometry, the precursors are isotopically different altering the ion's *m/z* depending on the isotopic version of the tag applied. Thus by comparing the intensity (e.g. area under peak or centroid) of the mass shifted peak to the non-shifted peak a relative difference between two samples can be determined.

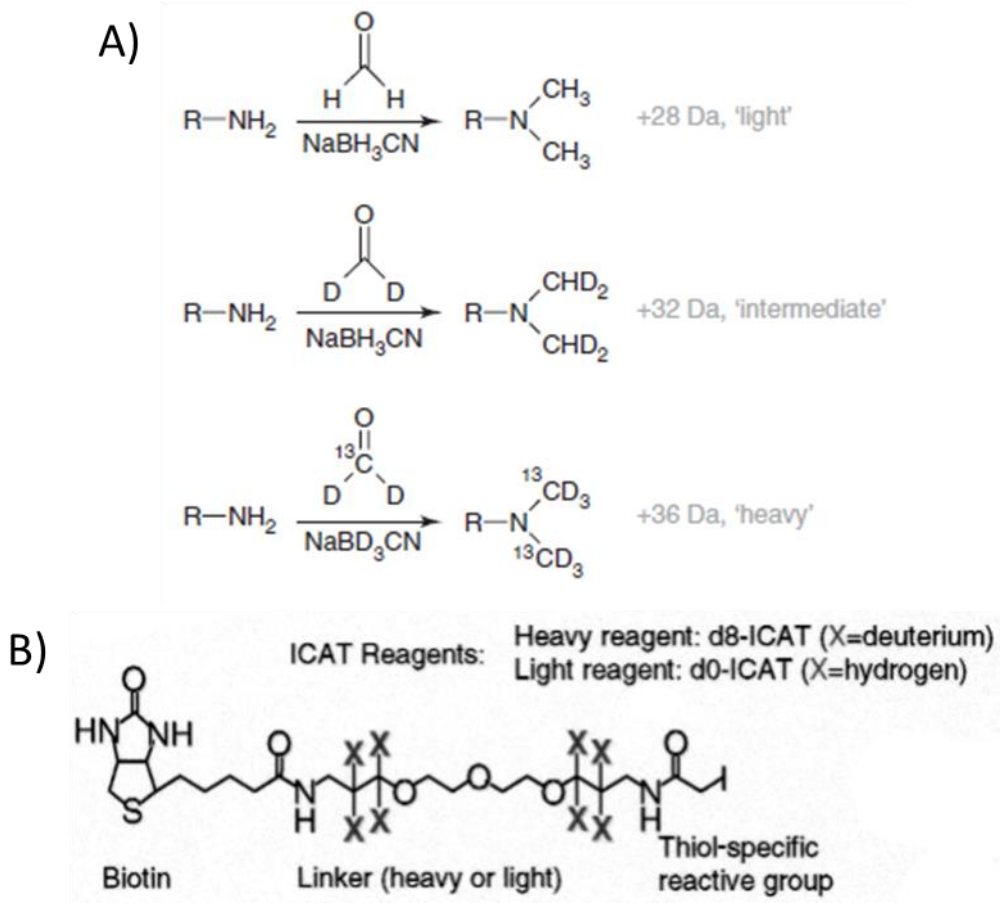


Figure 2.3: Structure of A) dimethyl labelling reproduced from Boersema *et al.*[154] and B) ICAT labels reproduced from Gygi *et al.*[155]

TMT and iTRAQ are fragmenting chemical tags. They consist of a linker, which binds to the peptide, a balancer and the isotopically varied reporter. Fragmented tags are identified at the post fragmentation MS/MS level. Prior to fragmentation the tags are linked to a balancer. This balancer has an isotopic difference to the various reports such that pre fragmentation the tags are isotopically indistinguishable. Post fragmentation, the reporter ion is released from the balancer and peptide such that each variant can be identified as isotopically distinct. In the case of TMT (Figure 2.4A_{p49}), a Thermo Scientific (Thermo scientific, Northumberland, UK) product, there are three versions; a TMTduplex*, TMTsixplex* and TMT10plex*.[158] In the case of iTRAQ (Figure 2.4B_{p49}) a AB Sciex (ABSciex , Framingham, MA, USA) product there are four varieties; mTRAQ duplex and triplex as well as the iTRAQ 4plex and 8plex.[159] TMT reporters have a mass between 126 and 131 *m*

whilst iTRAQ reports are between 113 and 121 *m*. The need for resolution within this range, whilst maintaining resolution at the higher *m/z* typical of most peptide fragments, limits the mass spectrometers usable for these mythologies. Further with the new TMT, 10-plex differentiation is achieved with the subtle difference between an isotopically heavy nitrogen and an isotopically heavy carbon. Use of these tags thus requires instrumentation capable of resolving to 0.006 *m/z*.

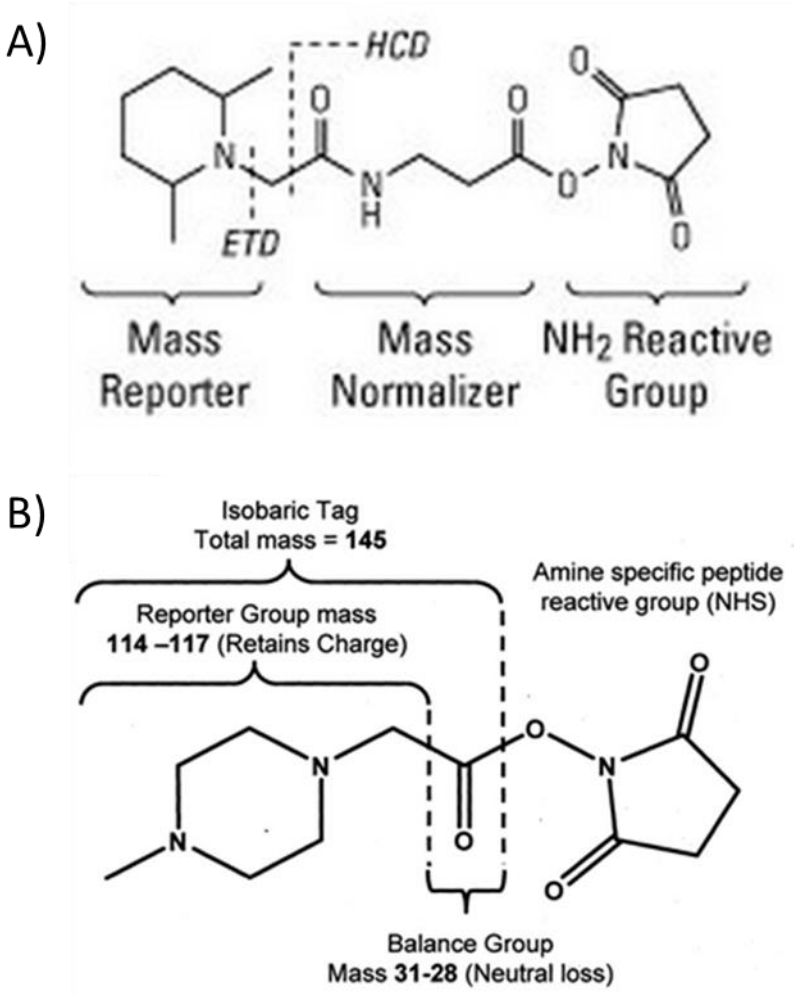


Figure 2.4: Structure of A) TMT reproduced from Thompson *et al.* [156] and B) iTRAQ reproduced from Ross *et al.* [157]

With the advancements in mass spectrometer precision and resolution, the label-free workflow has been developed. There are two broad types of label free methods available; spectral count and peak intensity. [150]

In spectral counting, the number of observations matching to a given protein are counted and compared between conditions.[150] Within a LC-MS/MS dataset, multiple proteins, or rather peptides arising from given proteins, will be repeatedly observed. This number of repeated observations will be correlated with the relative abundance of the protein in the given sample. Spectral counting is however only reliable for proteins with significant number of observations as these observation are the basis of the quantification. Old *et al*[160]. demonstrates that a minimum of 10 counts is required to provide a statistically significant ($p \leq 0.05$) fold changes of 2. Thus proteins identified by only a few spectra cannot be quantified to a high precision. In peak intensity, the strength of signal observed for precursor peptides is taken as an indicator of the intensity of the peptide and thus protein in the sample mix. For peak intensity to be reliable a balance is required such that the peptide is significantly observed in both samples but does not saturate the mass spectrometer.

Label free quantitative proteomics is gaining increased interest in the field of proteomics,[161] but its effective use is still largely based on the availability and quantity of mass spectrometer used. Given that each sample must be run separately, including separate runs to derive statistical repeats, the mass spectrometer run time for label-free techniques is therefore longer than a multiplexed investigation. Furthermore, the run-to-run consistency required means label free techniques are best utilised when control of an instrument can be dedicated to a single investigation at a time.[150]

A significant advantage of the chemical labelling techniques is the increased number of investigatory channels able to be assessed simultaneously. These channels can thus be used to facilitate replicates analysis, as well as assessment of multiple conditions with limited mass spectrometer run time.

For this work, iTRAQ 8-plex was chosen as the analysis methodology. iTRAQ 8-plex was selected for several reasons including:

- Labelling is post digest such that, no modifications to the media composition is required. Preferable given the media manipulation aspect of most proposed investigations.
- Considerable previous experience within the research group with this technique.
- Availability of a QStar XL. Suitable for chemical tag proteomics techniques.
- The availability of maximum number of channels to allow replicate and multiple condition experiments, (TMT 10-plex was not available at the start of this work).

2.2.6: iTRAQ

The typical workflow for iTRAQ is outlined in Figure 2.5A$_{p52}$ and is provided by the manufacturers protocol.[159] In a typical iTRAQ study, proteins are extracted from the samples under investigation. This can be achieved by a variety of techniques such as chemical extraction, french press, bead beating, liquid nitrogen grinding, etc. The protein must then be quantified and may be purified from contaminants through methods such as acetone precipitation. Proteins are then reduced and alkylated breaking sulphur bonds to open up compact proteins. These are then digested to peptides often using trypsin. The peptides can then be labelled with the iTRAQ reagent. Once labelled, the sample will typically be dried, allowing re-suspension in appropriate solution for fractionation. Typically, two orthologous fractionations are used. The first fractionation is typically conducted off-line and the second on-line with the mass spectrometer. MS/MS is them performed with the first round conducting a broad assessment spanning a wide *m/z* window of interest (termed a survey scan). From this, a limited number of ions will be selected for isolation and fragmentation (Figure 2.5B$_{p52}$). The mass spectrum profile of this second scan is used to determine the peptide. In addition, during the fragmentation, the labels will typically fragment, separating the reporter ions from the balancer.

These are observed between 113 and 121*m/z* (Figure 2.5C$_{p52}$). 120 *m/z* is omitted as phenylalanine immonium ion can commonly be found at 120 *m/z* and would otherwise disrupt the results.

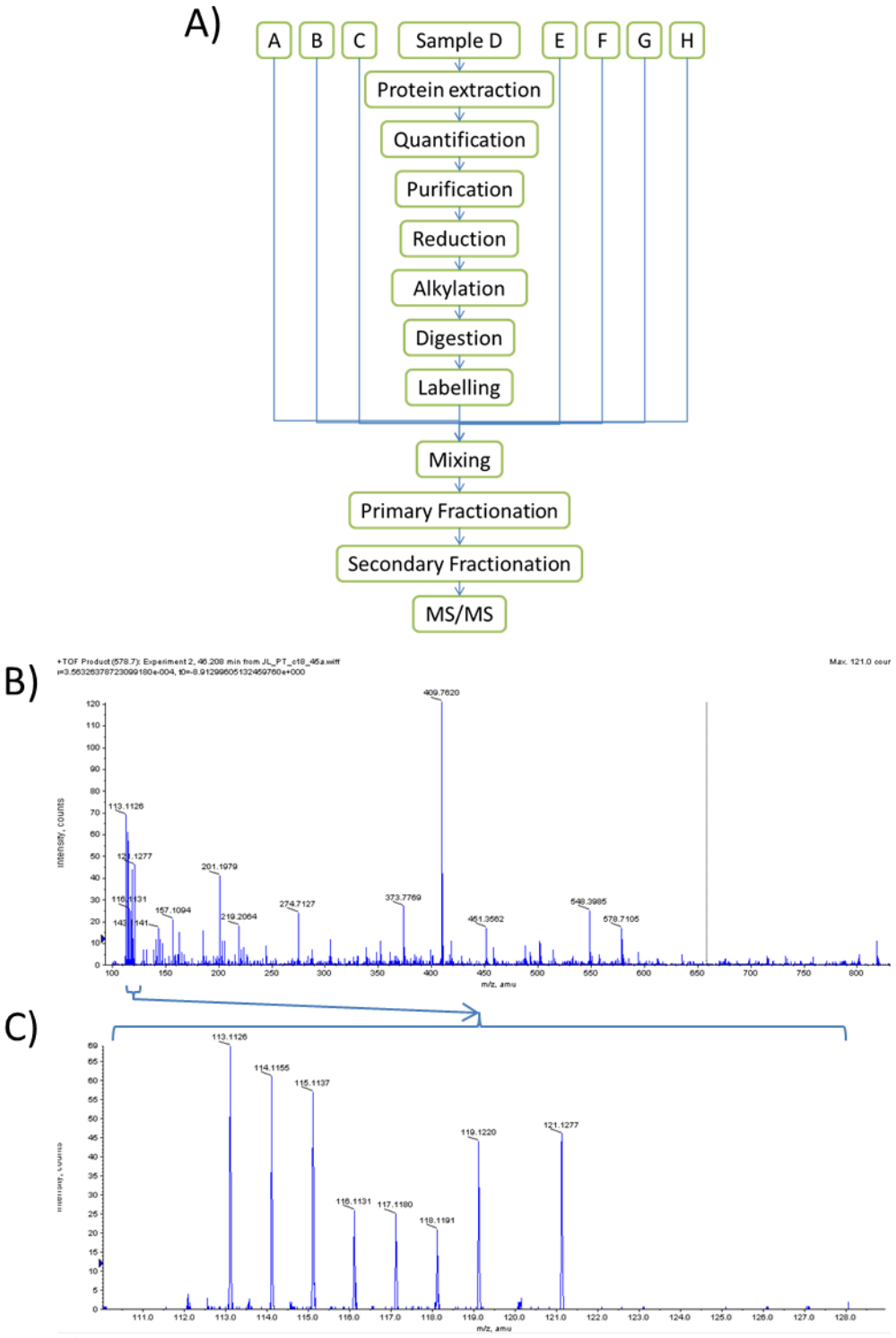


Figure 2.5: A) Flow diagram showing the steps involved in an iTRAQ investigation (here an 8-plex investigation is shown). B) Mass spectrum of product ions post

fragmentation of a selected precursor ion. C) Zoom in of 113-121 *m/z* region of the mass spectrum shown in part B.

2.2.7: Data processing options

Once an iTRAQ labelled peptide mix has been run on a mass spectrometer, the data must be processed. This processing can be referred to as the proteomics pipeline. It typically entails four stages;

- Conversion of data to an interchangeable format
- Peptide spectral matching
- Identification analysis
- Quantitation analysis

Further details of these steps and the options available are given in Chapter 4.

2.2.8: Examples of species engineering based on proteomics studies

From the literature, there are examples where the use of high throughput comparative investigations, have led to improved manipulation of a species, aiding in increased productivity. One such example is a study conducted in a series of papers by Pham *et al.*, on the production of ethanol in yeast.

An iTRAQ study conducted on the effect of high glucose levels on the growth and ethanol production of *Saccharomyces cerevisiae*[162] revealed insight into the metabolism of *S. cerevisiae* under such conditions. One key insight was the down regulation of amino acid production genes. From this insight, investigations were conducted into the effect of amino acid supplementation on the growth and ethanol production of *S. cerevisiae*.[163,164] This supplementation led to an increase in production rates under high glucose conditions achieving 40 g of ethanol per gram of glucose compared to 30 g of ethanol per gram of glucose for non-amino acid supplemented cultures.

Another example of successful use of proteomics based forward engineering is Jagroop *et al.*'s[165] study of glycosylation in recombinant *E. coli*. Within this study an iTRAQ investigation was conducted on induced cell. From the

resultant changes observed target for modification to improve *N*-Glycosylation were identified. These modifications gave a 300% increase in glycosylated proteins. Improved glycosylation, being the driving force of the proteomic investigation.

2.2.9: Previous proteomic investigations in microalgae

Modern research techniques take full advantage of techniques available in sequenced organisms including; transcriptomics, proteomics and metabolomics. Research into algae and their potential as a biofuel feedstock is no different. Jamers *et al.* [166] provides a good review of how microalgae have been investigated on various *omics* levels. This review focuses in much greater detail on the proteomic investigations in microalgae. To understand the current state of proteomic investigation, a table of publications investigating proteomics in microalgae was compiled (Table 2.2$_{p55}$).

To derive this table, the genus names of several species were searched using Scopus. These findings were then reduced to those matching 'proteo*'. These resultant matches were then further refined to identify those conducting a proteomic investigation. Each investigation is grouped by the genus. For each, a brief description of the purpose of the study is given along with the method of quantification (if used) and identification (if used). Each was also given a category type 1-5 as described below:

1. High throughput quantitative proteomics. Investigations that attempt to quantify proteins within a sample without pre selection. These include isotope labelling, chemical tag and label free investigations.
2. Low throughput proteomics. Investigations which use visual quantitative assessment, typically a 2D SDS-page gel, to select specific proteins to be identified by mass spectrometry.
3. Identification investigations where mass spectrometry techniques are used to identify but not quantify proteins within the sample which may represent the entire proteome or some selected sub-proteome.
4. Reviews and *in-silico* analysis of proteomic data.

5. Method development studies.

Table 2.2: Reported proteomic investigations in eukaryotic microalgae. Organised by species the objective of the investigation, method of separation and identification and the type of investigation as defined above are reported for each.

Genera	Objective of investigation	Method of quantita-tion	Method of identifi c-ation	Type	Reference
Chlamydomonas	Investigation of cold stress adapted culture	Intensity based label free	LC-MS/MS	1	167
	Analysis of copper, iron, zinc, and manganese micronutrient deficiency	Intensity based label free (MS^E)	LC-MS/MS	1	168
	Comparison of wild type and two random mutagenesis derived lipid overproducers	2D gel	MALDI-MS/MS	2	169
	Identification of glutathionylated proteins		LC-MS/MS	3	170
	Investigation of salt stress, 300 mM NaCl for 2 h	2D gel	LC-MS/MS	2	171
	Identification investigation focussing on the nuclear proteome		LC-MS/MS	3	172
	Comparison of iTRAQ and intensity based label free using late-log *cw15* vs. *star6* as the model comparison.	iTRAQ & Intensity based label free	LC-MS/MS	1	173
	Comparison of the effect of using different drafts of the *C. reinhardtii* proteome on the effect of proteomics investigations			4	174
	Profiling the proteome of oil bodies		LC-MS/MS	1	175
	Assessment of the proteomic response to herbicides paraquat, diuron, and norflurazon	Spectral count	LC-MS/MS	1	176
	Investigation into protein turnover and salt stress	SILAC	LC-MS/MS	1	177
	Comparison of changes following high light exposure for 1.5, 3 & 6 h	2D gel	LC-MS/MS	2	178
	Investigation into Nitrogen stress cells during carbohydrate and lipid induction	iTRAQ	LC-MS/MS	1	179
	Investigation into ammonium stress	Spectral count	LC-MS/MS	1	180
	Identification of Casein Kinase 1 targets in flagellum		LC-MS/MS	3	181
	Proteomic Analysis of Thylakoid Membranes	2D DIGE	MALDI-MS/MS	3	182
	Review of the chloroplast proteome			4	183

	Report of automated annotation of the *C. reinhardtii* proteome using genomic peptide finder & AUGUSTUS			4	184
	Comparison of duration of a heat stress treatments (42°C for 0, 60 & 180 min) with a ^{15}N ladled mix as reference	Metabolic labelling	LC-MS/MS	1	185
	Investigation into protein turnover within the stroma	Metabolic labelling & 2D gel	LC-MS/MS	2	186
	Evaluation of the extracellular proteins induced under high-CO_2 conditions	2D gel	LC-MS/MS	2	187
	Investigation into the absolute protein abundances when grown mixotrophically vs. autotrophically.	Metabolic labelling	LC-MS/MS	1	188
	Investigation of the anaerobic response focusing on the chloroplast and mitochondrion	Spectral count & SILAC	LC-MS/MS	1	189
	Comparison of the 'mitochondrial alternative oxidase mutant' with wild type	2D DIGE	MALDI-MS/MS	2	190
	Comparison of mitochondrion proteome under different nitrogen source NH_4CL & KNO_3	2D DIGE	MALDI-MS/MS	2	191
	Investigation of proteome changes of an acidophilic strain under BG-11 and metal rich acidic water	2D DIGE	MALDI-MS/MS	2	192
	Analysis of the proteomic changes under sulphur deprivation (0, 24, 32 h), known to increase hydrogen production.	2D gel	MALDI-MS/MS	2	193
	Analysis of the basic proteins at various points after establishing a 12hr day/night cycle	2D gel	LC-MS/MS	2	194
	Review of sub-proteome experiments (eyespot, flagella, mitochondrion, etc…)			4	195
	Investigation into the stoichiometric relationship between various Light harvesting complexes an and photosystems	SILAC	LC-MS/MS	1	196
	General review of proteomics in *Chlamydomonas*			4	197
	Investigation into the use of 1+ charge state ions in mass spectrometry based proteomics investigations		LC-MS/MS	5	198
	Review of tools available on ChlamyC web resource			4	199
	Review of the eye spot proteomics			4	200
	Investigation into the interacting	SILAC	LC-	3	201

	partners of the protein Vesicle-inducing protein in plastids 1 using QUICK (Quantitative Immunoprecipi-tation Combined with Knockdown)		MS/MS		
	Review of proteomic bases insight into iron and cadmium homeostasis			4	202
	Identification assessment of the mitochondrial proteome	1D gel	LC-MS/MS	3	203
	Mini-review of the use of LC-MS for analysis of phosphorylation in the eye spot apparatus			4	204
	Identification of phosphoproteins with in the eyespot		LC-MS/MS	4	205
	Analysis of s-thiolated proteins	Radiolabeling & 2D gel	MALDI-MS/MS	4	206
	Annotation of the proteome creating an annotated MapMan		LC-MS/MS	4	207
	Investigation of the proteins involved in the centriole			3	208
	Review of the phosphorylation in the thylakoid membrane			4	209
	Evaluation under iron stress	Intensity based label free	LC-MS/MS	1	210
	Analysis of the level of phosphorylation in *C. reinhardtii* compared to *A. thaliana* with assessment of phosphopetide enrichment	1D gel	LC-MS/MS	3/5	211
	Identification of the phosphoproteome		LC-MS/MS	3	212
	Identification of phosphorylation on exposed photosynthetic membrane proteins under various light and aerobic conditions	1D gel	LC-MS/MS	3	213
	Identification of phosphorylation in thylakoid membrane under varied CO_2	1D gel	LC-MS/MS	3	214
	Identification of eye spot apparatus proteins	1D & 2D gel	LC-MS/MS	3	215
	Method development of a MALDI proteomic method	ionic copolymer array	MALDI-MS/MS	5	216
	Investigation of the Antarctic Ice Microalga *Chlamydomonas* sp. proteome under low temperature stress	2D gel	MALDI-MS/MS	2	217
	Review of the circadian rhythm proteins			4	218
	Comparison of the proteome under cadmium stress (150 µM)	2D gel	MALDI-MS/MS	2	219

	Comparison of wild type and two high light resistant mutants under normal and high light conditions.	2D gel	MS/MS	2	220
	Investigation of pyruvate formatlyase proteins in the mitochondrion	1D gel	LC-MS/MS	3	221
	Investigation aimed at increasing identification coverage. Utilised a SILAC labelled thylakoid dataset	1D gel & SILAC	LC-MS/MS	3	222
	Review of the circadian rhythm proteome			4	223
	Identification of flagella proteins	1D gel	LC-MS/MS	3	224
	Comparison under iron stress, focusing of light harvesting proteins.	SILAC	LC-MS/MS	1	225
	Analysis of the 80 S cytosolic ribosome proteins	1D & 2D gel	LC-MS/MS & Edman	3	226
	Analysis and identification of centriole proteins associated with the basal body. Separated by centrifugation.	Sucrose gradient	MS with MudPIT analysis	3	227
	In-silico deduction of the Mitochondrial Oxidative Phosphorylation Proteome				228
	Analysis of functional proteins involved in the circadian rhythm	2D gel	LC-MS/MS	4	229
	Review of proteomics in Chlamydomonas reinhardtii			4	230
	Identification of Thioredoxin reducible proteins	2D gel	MALDI-MS/MS	3	231
	Establishment of a two-dimensional protein map of light-harvesting protein	2D gel	LC-MS/MS	3	232
	Comparison within the Thylakoid membrane of cultures grown autotrophically with salt or mixotrophically without	2D gel	MALDI-MS/MS	2	233
	Identification of proteins in the chloroplast ribosome	1D gel	LC-MS/MS	3	234
	Abundant mitochondria proteins were separated and identified	2D gel		3	235
	Characterization of the Small Subunit of Chlamydomonas reinhardtii Chloroplast Ribosome	1D gel	LC-MS/MS	3	236
	Identification of thylakoid membrane proteins and copper stress	2D gel	MS/MS	2/3	237
Chlorella	Comparison of inoculum concentration effect on the proteome ($1*10^4$, $1*10^5$, $1*10^6$ and	iTRAQ	LC-MS/MS	1	

	$1*10^7$ cells mL^{-1})				
	Investigation of cultures under nitrogen starvation	2D DIGE	MALDI-TOF–TOF	2	238
	Investigation of cultures under nitrogen starvation 24, 72 & 144h deprivation.	Spectral count	LC-MS/MS	1	239
	Investigation of cultures under nitrogen starvation using transcriptomic as a proxy quantification	1D gel	LC-MS/MS	2	240
Dunaliella	Comparison of proteome under salt stress (3.5 vs. 1.5 M)	2D gel	LC-MS/MS	2	241
	Identification of the proteins within discarded flagella		LC-MS/MS	3	242
	Comparison of membrane proteins under high vs. low salt conditions (3 vs. 0.5 M)	2D gel	LC-MS/MS	2	243
	Investigation into the ability to identify MS results from a non-sequenced organism through homology.		LC-MS/MS	3	244
	Comparison of high vs. low salt conditions (3 vs. 0.5 M)	2D gel	LC-MS/MS	2	245
Haematococcus	Proteome comparison under oxidative stress induced by sodium orthovanadate	2D gel	MALDI-MS/MS	2	246
	Comparison of proteome under nitrogen stress and high irradiance	2D gel	MALDI-MS/MS	2	247
	Comparison of proteome under astaxanthin inducing high light stress	2D gel	MALDI-MS/MS	2	248
	Investigation of the proteomic differences between a wild type and two cell wall mutants	2D gel	MALDI-MS/MS	2	249
	Analysis of the proteome under oxidative stress induced by addition of acetate, Fe^{2+} and excess light.	2D gel	MALDI-MS/MS	2	250
	Comparison of the cell wall proteins at different stages of the cell cycle	1D gel	MALDI-MS/MS	3	251
	Method development for separation of proteins on 2D gels	2D gel	N/A	5	252
Ostreococcus	Investigation of the synthesis and degradation rate of proteins	Metabolic labelling	LC-MS/MS	1	253
	Shotgun proteomic analysis of dial cycle and various nitrogen source supply levels	Intensity and metabolic labelling	LC-MS/MS	1	254
	Investigation of the effect of	Reductiv	LC-	1	255

	glufosinate	e Isotopic Di-Ethylatio n	MS/MS		
	Database based comparison of the proteome of two *Ostreococcus* species			4	256
	Review of the selenoproteomics in *Ostreococcus* and *Thalassiosira*			4	257
Thalassiosira	Comparison of Phosphorus replete and deplete cultures	Spectral count	LC-MS/MS	1	258
	Analysis upon exposure to benzo(a)pyrene	iTRAQ	LC-MS/MS	1	259
	Identification study of the thylakoid membrane proteome of *T. pseudonana and P. tricornutum*	2D gel	LC-MS/MS	3	260
	Analysis of the *T .pseudonana* proteome upon bloom degradation	Spectral count (limited)	LC-MS/MS	3	261
	Review of the selenoproteomics in *Ostreococcus* and *Thalassiosira*			4	257 (*)
	Identification of Proteins from a Cell Wall Fraction	1D gel	LC-MS/MS	3	262
Nannochloropsis	Investigation of Long-Term Nitrogen Starvation and Recovery	2D gel	MALDI TOF-TOF	2	263
	Comparison of the proteome under cadmium stress	2D gel	ESI-MS/MS	2	264
Phaeodactylum	Analysis of 48h nitrogen stress	2D gel	LC-MS/MS	2	265
	Identification study of the thylakoid membrane proteome of *T. pseudonana and P. tricornutum*	2D gel	LC-MS/MS	3	260 (*)
Emiliania	Comparison of growth under sub-optimal and supraoptimal light	Spectral count	LC-MS/MS	1	266
	Comparison of culture under current and increased CO_2 (395ppmv vs. 1340 ppmv)	iTRAQ	LC-MS/MS	1	267
	Initial identification and extraction report facilitating further proteomics studies	1D gel	LC-MS/MS	3/4	268
Cyanidioschyzon	Comparison between interphase and metaphase of isolated chloroplasts	2D gel	MALDI TOF-MS	2	269
	Analysis of plastid-targeted proteins shared between *Cyanidioschyzon* and A*rabidopsis*			4	270

Alexandrium	Used proteomics comparisons in an attempt to classify different samplings	2D DIGE	MALDI-TOF–TOF	2	271
	Comparison of a loss of toxicity mutant with wild type	2D DIGE	MALDI-TOF–TOF	2	272
	Investigation of extraction techniques using light stress as a comparative example	2D gel	MALDI TOF–TOF	5	273

Figure 2.6$_{p61}$ shows a pie chart representing the predominance of various genera within Table 2.2$_{p55}$. From this, it is clear to see *Chlamydomonas* is the dominant organism representing 65% of proteomic investigations. Further, Chlorophyta represent 83% of investigations. This predominance of *Chlamydomonas* supports its common description as a model eukaryotic alga. It has been used to investigate several research areas including eukaryotic locomotion, eyespot development and phototrophic growth.[274] For this reason, also two in depth reviews of the proteomic studies in *Chlamydomonas* are available.[197,230]

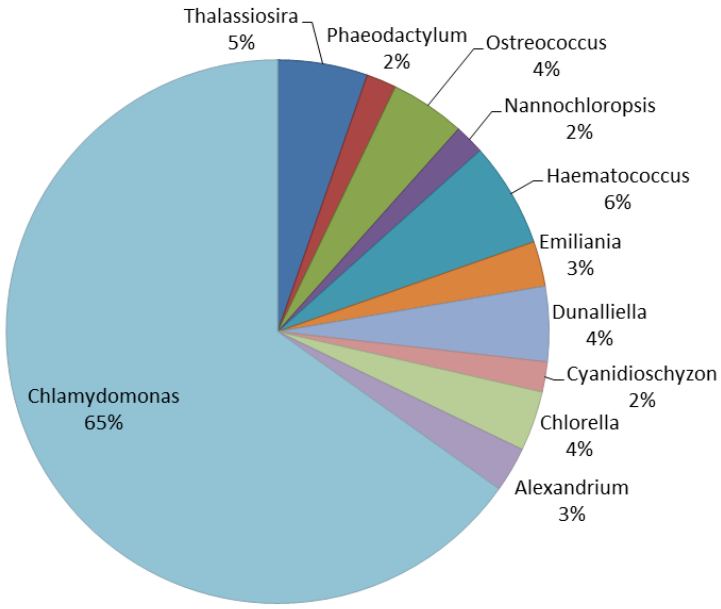


Figure 2.6: Pie chart showing the relative representation of different genera within the proteomic studies on the selected algae.

Figure 2.7$_{p62}$ shows the distribution of papers included in Table 2.2$_{p55}$ by publication year. Though a general increase in investigations can be observed, this isn't consistent with periodic spikes and dips in publications.

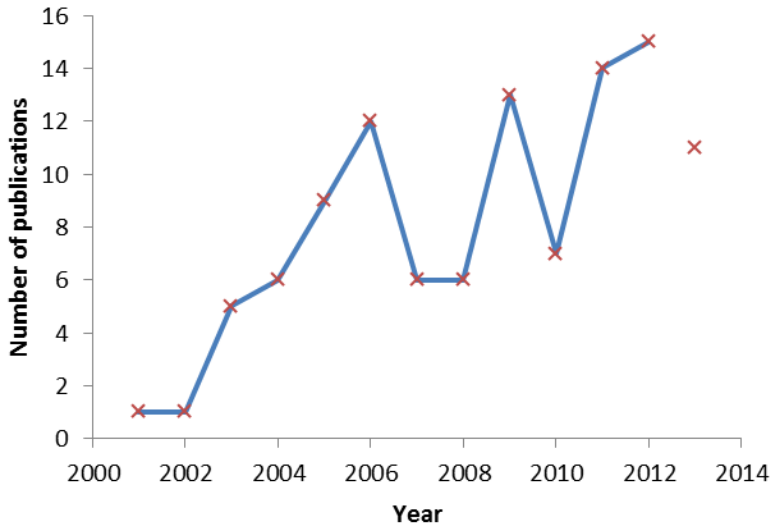


Figure 2.7: Graph of proteomics investigation in algae over time. Current year of 2013 shown only as a point.

2.3: Conclusions

The ability to produce alternative fuels from algae is an intriguing prospect, but requires considerable optimisation before being competitive. There are many areas in which better biological understanding of these organisms could be used to optimise production. This could be achieved through development of genetically modified strains or better application of biochemical manipulations. To identify targets for genetic manipulation or achieve this increased understanding for biochemical manipulation, the use of '-omic' research techniques, particularly comparative proteomics, could be invaluable.

Prior to 2011, comparative proteomics in eukaryotic microalgae was largely based on traditional low throughput techniques and was not been conducted with the aim to optimise microalgae for biofuel use. Increasingly, as with this

thesis, both the use of high-throughput technologies and the aim of biofuel understanding has become a primary focus.

This literature review has covered background of algae for biofuel and use of proteomics as a study tool including a broad overview of the literature as it stands on proteomics in microalgae.

Investigation of microalgae of multiple types under multiple conditions has merit to facilitate improvements in biodiesel production. Furthermore, the use of post genomic and in particular proteomic investigations techniques to conduct such an investigation has been established. However, there are various combinations of organism and condition which could provide insight. A cumulative effect of multiple such investigations is desirable.

In this work we chose to investigate three combinations of organism and perturbation: nitrogen stress in *C. reinhardtii*, Carbon source in *C. reinhardtii* and nitrogen stress in *Phaeodactylum tricornutum*.

- *C. reinhardtii* was selected as it has been demonstrated to be the predominant model organism with the greatest level of genomic description.
- Nitrogen stress as it has been repeatedly demonstrated to induce lipid overproduction.
- Carbon source was selected to better understand the effect of using mixotrophically grown *C. reinhardtii* as a model for photoautotrophic growth.
- *P. tricornutum* was selected as a divergent algal model from the Chlorophyta *C. reinhardtii*. This facilitates comparison between the divergent species. Further, the comparisons could be facilitated by comparison of silicon stress lipid induction in this silicon facultative species.

Whilst these are by no means the only investigations that could be conducted in algae, those selected present some of the more intriguing and currently relevant investigations which can be combined for comparative analysis.

Chapter 3: Materials and Methods

Chapter 3 – Common Materials and Methods

3.1: TAP medium

Tris Acetate Phosphate (TAP) medium was prepared as described elsewhere[275] For 1 L medium 2.42 g Trizma base, 25 mL TAP salts solution (1 L dH_2O with 5 g $MgSO_4.7H_2O$ and 2.5 g $CaCl_2.2H_2O$), 0.375 mL Phosphate solution (100 mL dH_2O with 28.2 g K_2HPO_4 and 14.4 g KH_2PO_4), 1 mL Hutners Elements (detailed below), 2 mL Nitrogen solution (100 mL with 18.75 g NH4Cl) and 1 mL glacial acetic acid. Medium was then made to 1 L with dH_2O. ($note_1$: for N free media Ammonia solution was omitted.)($note_2$: For autotrophic culture glacial acetic acid was omitted with pH of the media adjusted to 7 with HCl.)

Hutners Elements

Hutners Elements was prepared as described elsewhere.[276] For 1 L each of the following was dissolve in the volume of water indicated:

(EDTA was dissolved in boiling water, and the $FeSO_4$ was prepared last to avoid oxidation)

- 50g EDTA disodium salt in 250 mL
- 22g $ZnSO_4.7\ H_2O$ in 100 mL
- 11.4g H_3BO_3 in 200 mL
- 5.06g $MnCl_2.4\ H_2O$ in 50 mL
- 1.61g $CoCl_2.\ 6\ H_2O$ in 50 mL
- 1.57g $CuSO_4\ .\ 5\ H_2O$ in 50 mL
- 1.1g $(NH_4)_6Mo_7O_{24}.\ 4\ H_2O$ in 50 mL
- 4.99g $FeSO_4.\ 7\ H_2O$ in 50 mL

All solutions were mixed together except EDTA. The mix was heated to boiling before addition of EDTA solution. The mix was then cooled to 70 ^{0}C and 85 mL 20% KOH solution (Also 70 ^{0}C) was added. This was then made to 1L using dH_2O. The mixture was then bubbled for 24h with air before returning volume to 1L and filtering through GFC filter paper.

3.2: F/2+Si media

To make F/2+Si the following stocks were prepared:

- Trace elements - 4.16 g of Na_2EDTA, 3.15 g of $FeCl_3.6H_2O$, 0.01 g of $CuSO_4.5H_2O$, 0.022 g of $ZnSO_4.7H_2O$, 0.01 g of $CoCl_2.6H_2O$, 0.18 g of $MnCl_2.4H_2O$ and 0.006 g of $Na_2Mo_4.2H_2O$ were dissolved in 1 L dH_2O.
- Vitamin mix – 0.0005 g of Cyanocobalamin (Vitamin B12) 0.1 g of Thiamine HCl (Vitamin B1) and 0.0005 g of Biotin were dissolved in 1 Litre.
- Sodium Metasilicate – 30 g of $Na_2SiO_3.9H_2O$ was dissolved in 1 L.

To make the final media, 0.075 g $NaNO_3$, 0.00565 g $NaH_2PO_4.2H_2O$ and 33.6 g Ultramarine synthetic salts (Waterlife Research Industries Ltd. Middlesex, U.K.) were dissolved in 997 mL of dH_2O with 1 mL of each of the three stocks added.

This was pH adjusted to 8.0 using NaOH and HCL before sterilising by autoclaving.

3.3: Microscopy

Samples for microscopy were prepared by centrifugation of 1 mL sample at 3000g for 5 min before removing 950 µL and resuspending pellet in remaining 50 µL. 10_µL was then place on a glass slide with a cover slip placed on top. This was then visualised on an Olympus BX51 (Olympys, Southend-on-Sea, U.K.) and image capture conducted using ProgRes CapturePro 2.6 (PandA, Berkshire, U.K.).

3.4: Freeze drying

Samples used for biochemical analysis were centrifuged to form a pellet in a pre-weighed 1.5 mL eppendorf tube. These were then frozen to –20 ^{0}C before placing in a Modulyo freeze drier (Edwards, Crawley, U.K.) where they were kept for >12h. The dried biomass and eppendorf were then weighed and compared to the pre-weight to give the dried cell weight. These samples were then stored at -20^{0}C.

3.5: Pigment assay

Pigments were extracted by agitation of 0.5 mg biomass resuspended in 200 µL phosphate buffer (37.5 mM H3PO4 (pH 7.4)) with glass beads (425-600 μm (Sigma, Dorset, U.K.) for 10 min. Samples were diluted with 800 µL acetone before incubation in dark for 10 min and centrifugation at 10,000g for 1 min. Optical density was then read to determine the pigment concentration as described by Wellburn.[277]

3.6: Carbohydrate assay

Carbohydrates in the microalga samples were estimated using the anthrone method.[278] Briefly, 200 μL of 0.5 μg μL^{-1} algal sample resuspended in distilled water (dH_2O) was transferred to a clean glass test tube. To this, 400 μL of 75% H_2SO_4 and 800 μL anthrone solution (25 mg anthrone, 500 μL EtOH, 12 mL 75% H_2SO_4) were added. Samples were mixed by vortexing and heated at 100 ˚C for 15 min. Samples were then allowed to cool, re-mixed and the OD read at 620 nm. A standard curve (Figure 3.1$_{p69}$) generated using glucose was used to estimate the carbohydrate concentration.

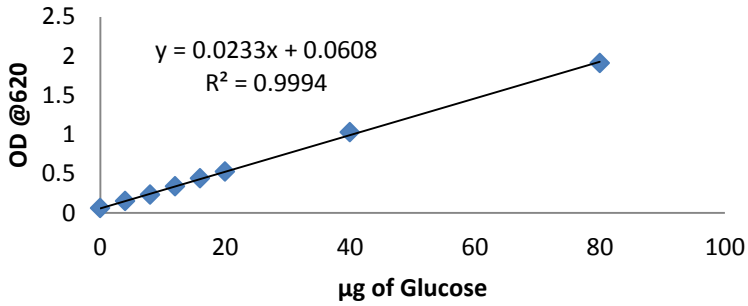


Figure 3.1: Calibration curve showing the colorimetric reaction of glucose to the carbohydrate assay. The equation shown was used to compute a quantity of carbohydrate and thus percentage dry cell weight for carbohydrate assessment of algal samples.

3.7: Lipid assay

The Nile red lipid assay was adapted from Chen *et al.*[279] Dried algal samples were resuspended to 10 μg μL^{-1} in dH_2O. Samples were sonicated for 5 min., samples were then diluted 1:20 to 0.5 μg μL^{-1}. To 6 wells of a 96 black walled clear bottom plate (Greiner Bio-one, Stonehouse, UK), 125 μL dH_2O was added followed by 100 μL of diluted algal sample and 50 μL DMSO. The plate was then shaken for 5 min at 37 °C in a Genios Tecan plate reader (Tecan, Reading, UK). Fluorescence intensity (excitation 530 nm, emission 580 nm) was then read 10 times for each well with 1 min spacing. To all wells 25 μL of Nile red (9 μg mL^{-1}) diluted in DMSO was added. The plate was then shaken for 10 min at 37 °C in the Genios Tecan plate reader. Fluorescence intensity (excitation 530 nm, emission 580 nm) was then read 30 times for each well with 10 min spacing to ensure complete absorption. From this, a trace of fluorescence intensity over time could be drawn as shown in Figure 3.2_{p70}. The increase observed is due to increase in the diffusion of the Nile red dye into the cell to such a point where the signal saturates or the lipid droplets disperse. To ensure reproducibility from run to run the maximal intensity was used for data processing. To process the data, the average fluorescence intensity for each sample without Nile red was subtracted from the average intensity reading for each sample with Nile red staining.

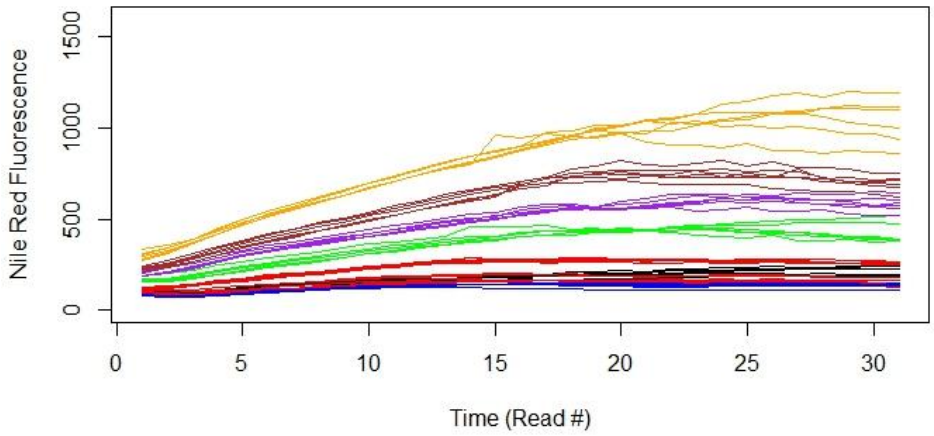


Figure 3.2: Output from eight samples with varying lipid content assessed by the Nile red method. Fluorescence is shown to increase over sample number and hence time. Technical replicates are shown in matching colour. The maximal reading for each sample was averaged to give a Nile red value for the given sample.

3.8: Protein extraction

Protein extraction was achieved by liquid nitrogen grinding. Stored cell samples were resuspended with 500 μL 500mM TEAB (pH 8.5). Samples were immersed in a cooled sonication water bath for 5 min and subsequently ground using a mortar and pestle cooled by liquid nitrogen. Post-grind samples were collected into a fresh protein low bind tube. Samples were then immersed in a cooled sonication water bath for a further 5 min and sonicated for two cycles with a Micro tip Branson sonifier (Enerson, Danbury, CT, USA). Samples were then centrifuged at 18,000*g* for 30 min at 4 ˚C to separate the soluble and insoluble fractions.

3.9: RCDC quantification

Soluble protein samples were acetone precipitated by incubation for 6h with acetone at –20 ˚C at a 6:1 ratio acetone to sample. Sample was then quantified by RCDC (BioRad, Hercules, CA, USA) following the manufacturer's stated protocol. This was assessed against a standard derived using BSA (Sigma, Dorset, UK).

3.10: Protein reduced, alkylated and digestion

An aliquot (100 μg) of acetone-precipitated protein sample was resuspended in 30 μL 500 mM TEAB (pH 8.5) with 0.1% sodium dodecyl sulphate. This was then reduced with 2 μL 50mM tris-(2-carboxyethyl)-phosphine (or for Chapter 5, 0.5 μL 10 mM DL-Dithiothreitol) at room temperature for 30 min, followed by alkylation with 1 μL 200mM methyl methanethiosulfonate in isopropanol (or for Chapter 5, 2 μL of 10 mM Iodoacetamide) for 20 min in dark[185]. Samples were digested with 1:40 trypsin (Promega, Southampton, UK) resuspended in 1 mM HCl at 1 μg μL^{-1}. To the digests, an aliquot (5 μL) acetonitrile (ACN) was added before incubating at 37 ˚C for 16 h.

3.11: 1D SDS-page gel

1D SDS-page gel was prepared using the Mini-PRETEAN® Tetra Cell (BioRad, Hercules, CA, USA) gel casing system. A 10% gel was prepared before loading with marker dye EZ Run (Fisher, Loughborough, U.K.) and 10 μL pre and post digested sample. The gel was placed in the frame and tank with running buffer. A current of 80V was applied for 10 min followed by 180V till the loading dye reached the end of the gel. Upon extraction the gel was rinsed in dH_2O and stained with InstantBlue™ (Expedeon, Cambridgeshire, U.K.). The gel was assessed to ensure equal loading of the samples and correct digestion as demonstrated in Figure 3. 3_{p72}.

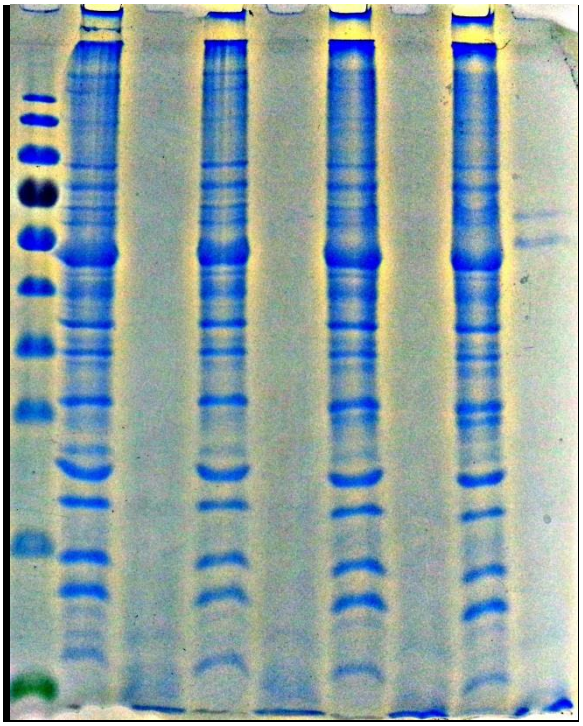


Figure 3. 3: Example of 1D SDS-Page gel showing undigested sample followed by digested sample. Such a gel was run for each sample and digestate used for iTRAQ.

3.12: iTRAQ labelling

Peptides were then labelled with 8-plex iTRAQ reagents (AB Sciex, Framingham, MA, USA), resuspending reagents with 50 μL isopropanol and using a further 20 μL to clean label container. The pH of the sample was tested using indictor paper to ensure a pH between 7.5 and 8.5, using 500mM TEAB to adjust if necessary. Sample and label mix was incubated for 4 h. After incubation all labelled samples were transferred to a fresh lo-bind eppendorf. This mix was centrifuged at 18,000g for 10 min to remove any precipitated protein. Labelled peptides were split in two fractions and vacuum centrifuged for storage at -20 ˚C before fractionation.

3.13: SCX fractionation

Strong cation exchange (SCX) fractionation was carried out using a BioLC HPLC system (Dionex, Sunnyvale, CA, USA). One iTRAQ labelled sample was resuspended in 60 μL buffer A (25% ACN, 0.1% formic acid (FA)). The resuspended sample was loaded onto PolySulfoethyl A column, 5 μm particle size, 20 cm length, 2.1 mm diameter, 200 Å pore size (PolyLC, Columbia, MD,

USA). With a flow of 0.2 mL min^{-1} buffer A was exchanged with buffer B (500 mM KCl, 25% ACN, 0.1% FA) to form a linear gradient as follows. 0% B (0-5 min), 0-40% B (5-45 min), 40-100% B (45-50 min), 100%B (55-60 min), 0%B (60-65 min). Fractions were collected every minute from 16 min through to 48 min. The SCX fractions were cleaned through UltraMicroSpin Columns (Nest, Southborough, MA, USA) according to the manufacturer's guidelines prior to vacuum centrifugation (Eppendorf, Stevenage, UK) in preparation for LC-MS analysis.

3.14: HILIC fractionation

High-resolution hydrophilic interaction chromatography (HILIC) was carried out using an Agilent 100-series HPLC (Agilent, Wokingham, UK). One iTRAQ labelled sample was resuspended in 100 μL buffer A (10 mM ammonium formate, 90% ACN, pH 3 (adjusted with (FA)). The resuspended sample was loaded onto PolyHydroxyethyl A column, 5 μm particle size, 20 cm length, 2.1 mm diameter, 200 Å pore size (PolyLC, Columbia, MD, USA). With a flow of 0.5 mL min^{-1} buffer A was exchanged with buffer B (10 mM ammonium formate, 10% ACN, pH 4 (adjusted with (FA)) to form a linear gradient as follows: 0% B (0-5 min), 0-15% B (5-7 min), 15% B (7-10 min), 15-60% B (10-50 min), 60-100%B (50-55 min), 100%B (55-65 min), 0%B (65-75 min). Fractions were collected every minute from 18 min through to 41 min followed by three, 3 min fractions to 50 min. The fractions were vacuum centrifuged, ready for LC-MS analysis. Desalting clean-up is not required for HILIC samples.

3.15: C18 clean-up

The SCX fractions were cleaned through UltraMicroSpin Columns (Nest, Southborough, MA, USA) according to the manufacturer's guidelines prior to vacuum centrifugation (Eppendorf, Stevenage, UK) in preparation for LC-MS analysis

3.16: RPLC-MS analysis

RPLC-MS was conducted using an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA, USA) coupled to a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer (Applied Biosystems (now ABSciex), Framingham, MA, USA). Samples were resuspended in 20 μL buffer A (3% ACN, 0.1% FA) before loading 9 μL onto a Acclaim PepMap 100 C18 column, 3 μm particle size, 15 cm length, 75 μm diameter, 100 Å pore size (Dionex, Sunnyvale, CA, USA). With a flow of 300 μL min^{-1} buffer A was exchanged with buffer B (97% ACN, 0.1% FA) to form a linear gradient as follows. 3% B (0-5 min), 3-35% B (5-95 min), 35-90% B (95-97 min), 90% B (97-102 min), 3% B (102-130 min). The mass detector range was set to 350-1800 *m/z* and operated in positive ion mode. Peptides with +2, +3, and +4 were selected for fragmentation. The remaining sample was subsequently injected in the same manner to acquire two RPLC-MS runs for each submitted fraction.

3.17: Proteomics processing

The *.wiff files outputted by RPLC-MS were processed, using the method developed and described in Chapter 4 'Proteomics Processing', to produce a list of peptide spectra matches (PSM(s)).

Briefly, each .wiff file was converted into a .mgf file using Mascot Daemon before being run against the species specific database: For *P. tricornutum* (Taxon id: 2850) downloaded from TrEMBL database (www.uniprot.org)[280] on the 1st February 2013. For *C. reinhardtii* (Taxon id: 3055) downloaded from TrEMBL database (www.uniprot.org)[280] on the 1st February 2013 Six search engines were used for the search, Phenyx (Genebio, Geneva, Switzerland), ProteinPilot (AB Sciex UK, Warrington, U.K.), Peaks (BSI, Waterloo, ON, Canada), OMSSA[281], X!Tandem[282]) and Mascot (Matrix Science Ltd, London, U.K.). SearchGUI[283] and Peptide shaker [284] were utilised to pre-process the OMSSA, X!Tandem and Mascot before merging the results as using the R script created in Chapter 4 and reported in Appendix 4.1_{p267}.

Given the PSM output from this script the label intensities were processed using the method outlined in Chapter 4 and an R script reported in Appendix 4.2$_{p273}$. This gave a fold change and p values for significance of the change.

3.18: EGGNOG analysis

EGGNOG analysis was achieved by batch converting Uniprot ID's using the ID Mapping function of Uniprot [280] to identify relevant EGGNOG id(s)[285]. These were then matched to the classification system[286] using a Perl script and processed into graphs using R.

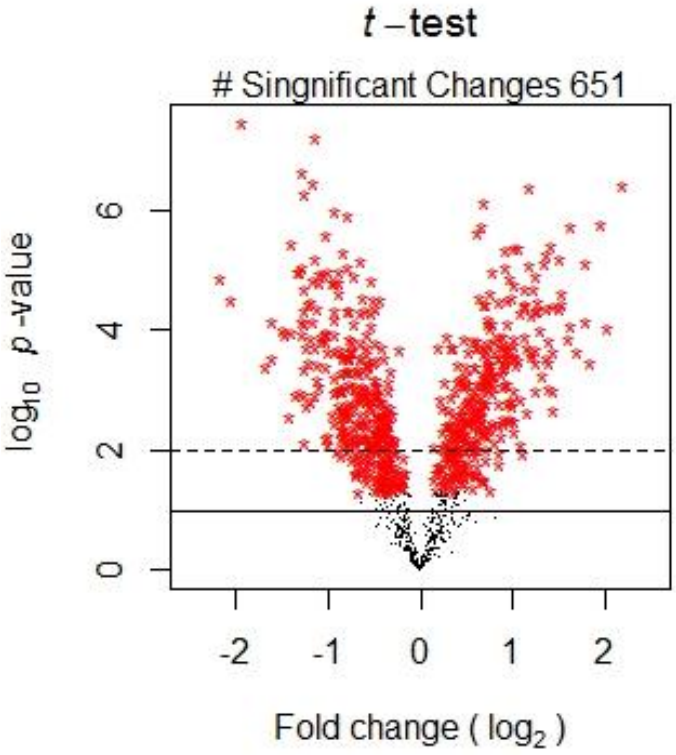
3.19: Gene ontology analysis

Gene ontology was identified using the functional annotation tool DAVID.[287,288] These were exported and processed into graphs using R.

3.20: KEGG analysis

To observe the pathway changes the relevant p-value and fold change were used processed. A term was derived for each entry of the form "UniprotID ,colour". Proteins with a positive fold change were labelled with "dark green" if not significant ($p>0.05$) and "blue" if significant ($p < 0.05$). Proteins with a negative fold change were labelled with "purple" if not significant ($p > 0.05$) and "red" if significant ($p<0.05$). This list of entries was then run through the KEGG "Search&Color Pathway" tool [289] selecting Uniprot id's, appropriate species, clear background and uncoloured diagram.

Chapter 4: Proteomics Processing



Chapter 4: Proteomics Processing

4.1: Abstract

With any proteomics investigation, the proteomics data processing pipeline is an integral part of the investigation. In this chapter, several possible iterations of the pipeline for an iTRAQ based investigation have been assessed. This has included assessment of six commonly used search engines: ProteinPilot, Phenyx, OMSSA, X! Tandem, Mascot and Peaks. It was determined that given a *.wiff input, ProteinPilot gave the greatest number of identifications, but this advantage was reduced given an *.mgf input.

Within the chapter, an R based script was developed to merge together the results from the six search engines. The merger was conducted in such a way so as to remove Peptide Spectral Matches (PSMs) which conflicted between the six search engines. Using this combined dataset, compared to the highest single search engine method, a 17.2% increase in protein (≥2 peptide) identifications was achieved. In addition to increased counts, this merged dataset had increased reliability with removal of conflicting PSM(s).

The various methods available to identify and assess the significance of changes within an iTRAQ investigation were then evaluated. An R based script is reported here that performs 'Variance Stabilization and Normalisation' (VSN), isotopic and loading correction to the merged dataset before performing a *t*-test to assess the significance of biological changes observed.

This analysis provides an appraisal of multiple search engines. Using the merger tool developed, an improved analysis of iTRAQ databases can be achieved. It also demonstrates the principle behind development of a universal merging tool for protein search engines.

4.2: Introduction

Protein investigations and in particular proteomic approaches have been revolutionised by incorporation of mass spectrometry in the analysis stream. Originally mass spectrometers were utilised for identification,

perhaps with quantitation offline via gel electrophoresis. Increasingly though, mass spectrometers have been utilised directly for simultaneous identification and quantitation. Quantitative proteomic mass spectrometry can be achieved by three major means isotopic labelling, chemical tag(s) or label-free. Further details of each can be found in Section 2.2.5$_{p46}$.

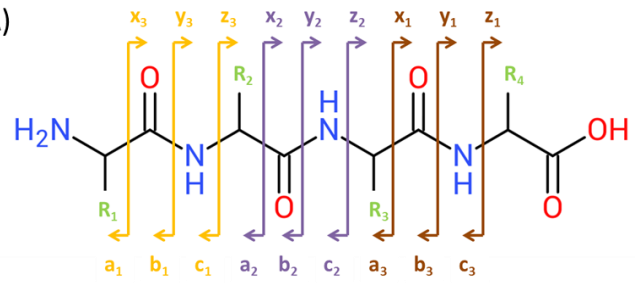
To conduct either identification or quantitation through mass spectrometry it is necessary to have the ability to predict the possible mass over charge (*m/*z) of the subcomponents of protein as it is fragmented. This can be done either by peptide mass fingerprinting or Tandem-MS/MS ion search,[290] the latter being amenable to more complex samples,[291] with multiple proteins being analysed in an ordered but not isolated fashion.

'Peptide mass fingerprinting' relies on the reproducible and predictable way in which proteases like trypsin will digest a protein down into peptides. This process can then be calculated to give a theoretical series of peptide masses, which in turn give a set of possible mass over charge values. By matching this theoretical distribution with that of the digested protein, a match can be made.

'MS/MS ions search' is based on the identification of peptides rather than proteins, though the originating protein is typically deduced subsequently. Initial MS is conducted to identify the series of ions being presented to the mass spectrometer at a given time. In a proteomics experiment, these would be expected to be predominately peptide ions. Using a set of ranking criteria, such as intensity and time previously selected, the instrument software selects one or more ions for further analysis. In tandem mass spectrometry this analysis consists of isolation and fragmentation (typically by collision induced dissociation).[127,290,291] Such fragmentation produces a series of product ions covering the spectrum of possible fragments from the selected peptide (Figure 4.1B$_{p80}$). Given the unique mass of most amino acids, the loss in mass observed from one position in the peptide to the next can allow the determination of the amino acid lost and thus the sequence of the peptide. This can be done either without any pre-expectation of the sequence (*de novo*

sequencing)[292]or by comparison to a set of existing peptides (database matching)[291].

A)



B)

B ion series masses	Peptide KREMER		Y ion series masses
129.10	K	REMER	720.34
285.20	KR	EMER	564.24
414.24	KRE	MER	435.20
545.28	KREM	ER	304.16
674.32	KREME	R	175.11

Figure 4.1: A) Diagram of the Roepstorff-Fohlmann-Biemann nomenclature for the series of ions produced upon peptide fragmentation[291,293]. B) Depiction of the possible resultant fragments and masses from a given peptide fragmented along the B & Y series.

Whilst it is theoretically possible to match the results of both protein methods manually, it is complicated by several factors including:

- The various charge states the fragments can have.
- The presence of both forward and reverse fragmentation series (i.e. the 'C terminal' and 'N terminal' fragment produced from any break)
- Different positions between peptides that the fragmentation can occur (forming the 'A', B','C' or 'Z','Y',X series (Figure 4.1A$_{p80}$)[293]
- Uneven distribution of the possible product ions
- Not all fragments will be observed with some being less amenable to mass spectrometric analysis than others.
- Noise produces by 'contaminants'

For these reasons, the identification manually of even a few MS/MS scans is a difficult task and a typical experiment would contain tens of mass spectrometry runs, each with hundreds of MS and MS/MS scans. Instead, computational search engines have been developed to automatically match peptide sequences to MS/MS scans and report the match and degree of

certainty of such match. Collectively these software form a significant part of the proteomics data pipeline.

4.2.1: Data pipeline

Once the data has been collected by a mass spectrometer, it must then be processed into a form that can be utilised to derive hypothesis of the biological changes observed. This processing can be broken down into several stages, as depicted in Figure 4.2$_{p81}$. Though these set of steps broadly describes data processing, it is common for the software involved to combine some or all of the steps, into one process. Furthermore, although capable of conducting multiple steps many of the multi-step software will allow partially processed data being imported from other software given an interchangeable format.

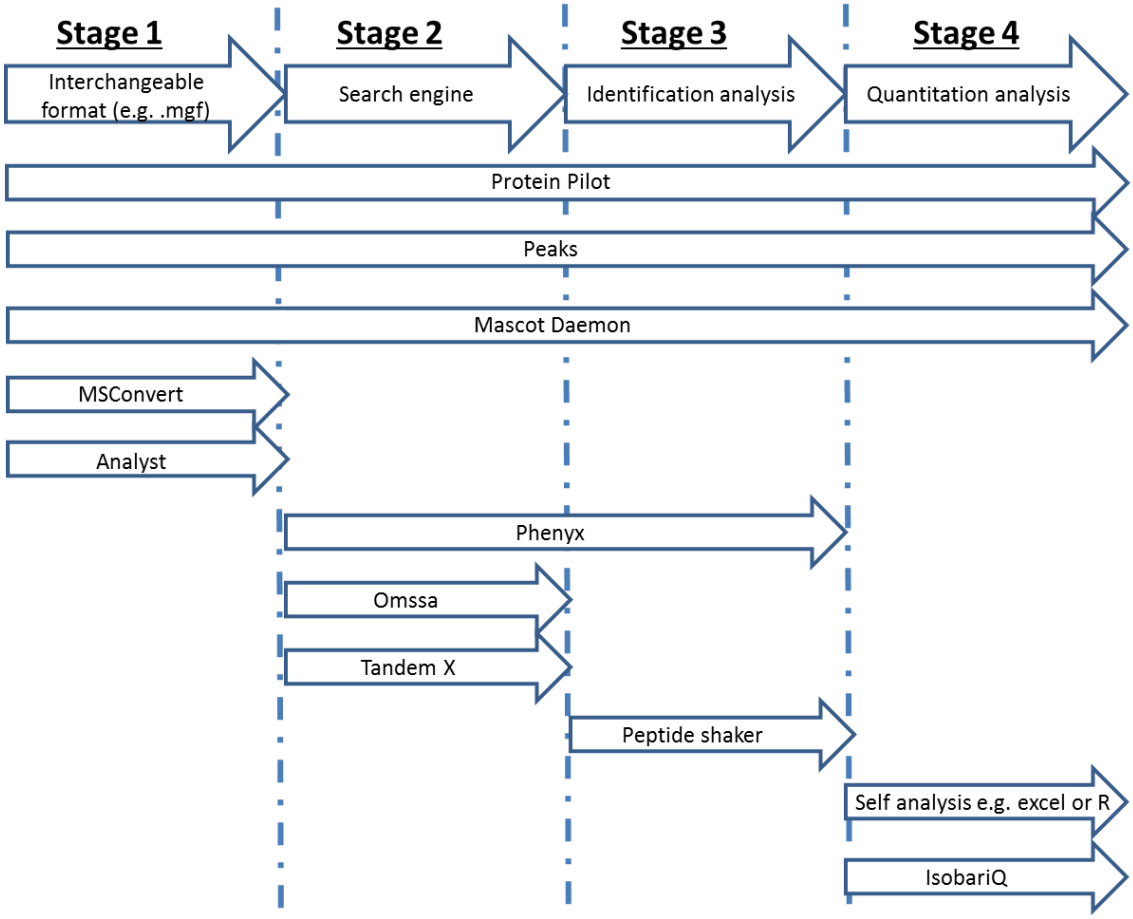


Figure 4.2: Schematic diagram of the proteomic analysis pipeline for iTRAQ and some commonly used software discussed in this analysis. The stage(s) within the proteomic pipeline conducted by each of the software is indicated. (Though able to handle raw data ProteinPilot and Mascot can accept data post conversion to *.mgf and report raw data identification data so processing can be continued in other programs)

The pipeline typically begins with the conversion of the mass spectrometer native file output into an interchangeable format, which can be utilised by multiple software brands (stage one). This can be the mzXML format, designed to represent native mass spectral data.[294,295] Often however, a less informative format is used though still capable for use in MS.MS peptide assignment. Example include, SEQUESTS *.dta and Micromass peak list, but Mascot's *.mgf (Mascot generic file) has become commonly supported by most mass spectrometer vendors. The *.mgf format is a description of MS/MS scans in a standardised text format.[296] Each scan consists of a brief header detailing various information such as precursor mass, elution time, file name and scan number. This is followed by a list of *m/z* values and intensities such that the series of distinct lines represent the profile of the product ions. This need for an interchangeable format is not always required, depending on the maker of the mass spectrometer and the selected processing software, with companies such as ABSciex, which supplies both, facilitating direct input of the raw data format into their own proprietary software.

Stage two in the pipeline can be considered the search itself. This can broadly be broken down into two types, database matching and *de novo* sequencing.

Database matching uses algorithms sometimes proprietary (e.g. Paragon™,[297] Andromeda,[298] Sequest,[299] OLAV[300,301]) in nature to match the spectra of the mass spectrum to theorised spectra conducted from sequence database typically in FASTA format. Though not fully universal with different parsing rules for the header the FASTA format consists of a description line started by a ">" which gives details of the protein ID, description, species etc. This is followed by the protein sequence in single letter amino acid code. FASTA formatted proteome databases such as that for a sequenced species can typically be obtained from several sources, including the organisation who conducted the sequencing, or collections such as Uniprot's SWISS-PROT/TrEMBL (http://www.uniprot.org/[280]).

Alternatively, if sequence information is not available a *de novo* search engine can be used.[292] This uses an algorithm which derives possible peptide sequences without any preconception of the sequence.

Many programs combine the search engine and identification analysis. This is not universal particularly for open source search algorithms such as OMSSA[281] and X!Tandem.[282] From the search engine, a series of peptide spectrum matches (PSM(s)) is created each with an assessment of confidence. Stage three can be considered the compiling of the PSM list to give a list of peptides and proteins with confidence assessments that are a result of the PSM list. For identification, the pipeline could then be considered complete. However for quantitation, further processing is generally required and can be considered stage four.

Depending on the quantitation method used not all identification pipelines will have been compatible and hence the type of quantitation planned will influence the pipeline used. Different methods are available depending on the type (isotope, chemical tag or label free) of quantitation used and the specific (e.g. iTRAQ, TMT, di-methyl) method. For this investigation, the methods available for iTRAQ will be discussed (iTRAQ being selected for this work as discussed in Section 2.2.5$_{p46}$).

4.3: Interchangeable formatting (e.g. *.mgf)

The initial step of the data processing pipeline is typically conversion of the data into a format usable by all software. The various mass spectrometer companies typically supply their mass spectrometers with their own version of software to interface with the instrument and store raw data. Thus, the raw data produced by the instrument can be in various formats (e.g. *.d/*.wiff/*.dta/*.raw). Though there may be a link between the mass spectrometry software and the subsequent search engine, this is not universal. However, most programs now accept the peak list format *.mgf (Mascot generic format) as an input format. Likewise, most mass spectrometry software now facilitate output of an *.mgf with varying degrees of curation for peak inclusion and fragment ion scan merger.

Within the subsequent analysis of various software programs, four different inputs will be utilised:

- Native *.wiff files covering a MS analysis of the effect of carbon source utilised by C*hlamydomonas reinhardtii*, comprising 77 individual mass spectrometer files from 77 LC-MS/MS runs.
- Processed *.mgf formatted using the mass spectrometer manufacturers' software program Analyst (version 2.0). (both as 77 individual *.mgf(s) and as 1 concatenated *.mgf).
- Processed *.mgf formatted using Mascot daemon (version 2.3.2) (both as 77 individual *.mgf(s) and as 1 concatenated *.mgf).
- Processed *.mgf formatted using the open source *.mgf curator MSConvert from Proteowizzard[302] (version 3.0.4268) (as 77 individual *.mgf's only)

4.4: Search engine & identification analysis

Though sometimes distinct steps, the search engine and identification analysis are more commonly merged into a single software operation. Several search engines are available,[283,297–299,303,304] some proprietary in nature and others open source. Figure 4.3$_{p85}$ shows the frequency that terms referring to some commonly used search engines occur as indexable terms within literature also containing 'proteom*' as an indexed term (Conducted 1st May 2013 using Scopus[305]).

Whilst not detailing all instances of each search engine this gives a reasonable approximation of the frequency each is used in the literature. For this investigation, six search engines were used (coloured green in Figure 4.3$_{p85}$). These give an overview of different types of search engines available. Including:

- Open source,
- Commonly used,
- *de novo* capable
- iTRAQ tailored

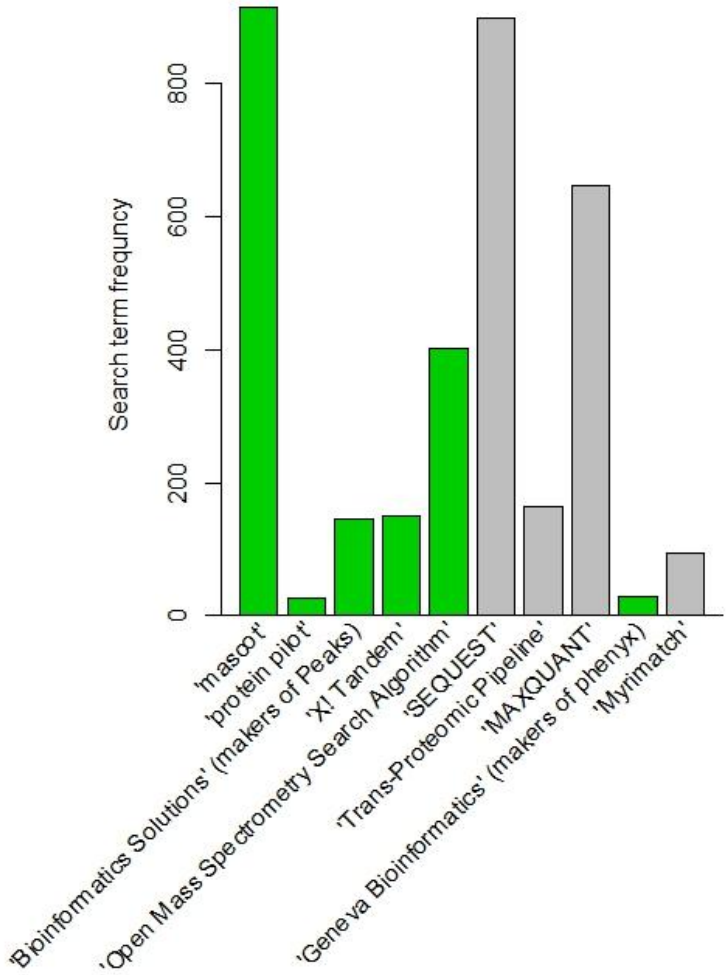


Figure 4.3: Barchart of the frequency each search term is observed in the literature restricted to literature also containing 'proteom*' as an indexed term.

In the subsequent sections of this Chapter each of the selected search engines will be detailed with the parameters selected for search indicated. The effect of some of these parameters will be reported as well as the search results. The format of the results output is then also described. A comparison of the identification made by each will then be made before describing a method derived to merge the results of these six search engines.

4.4.1: ProteinPilot

ProteinPilot (version 4.0.8085) is an identification and analysis software available under licence from ABSciex. It accepts both raw *.wiff files (the ABSciex MS output) and processed *.mgf files. The search uses the Paragon ™ algorithm[297] to obtain PSM(s). ProteinPilot has limited flexibility when it relates to the search setting favouring a more '@click and go' approach. It is also streamlined for the running of iTRAQ which is a licenced product of ABSciex.

This compatibility is reflected in its predominance in use for iTRAQ investigations being incorporated in >50% of investigations[306] compared to <1% as determined from Figure 4.3$_{p85}$.

Search settings

The four inputs described earlier (Section 4.3$_{p83}$) were also run for identification and analysis on ProteinPilot. Each search was performed with setting depicted in Figure 4.4$_{p86}$. iTRAQ 8-plex was selected as the labelling with MMTS and trypsin selected for the cysteine alkylation and digestion. The QSTAR was selected as the instrument used. Quantitate, Bias correction and Background correction were all active and the search effort set to thorough with a protein threshold of 95% confidence.

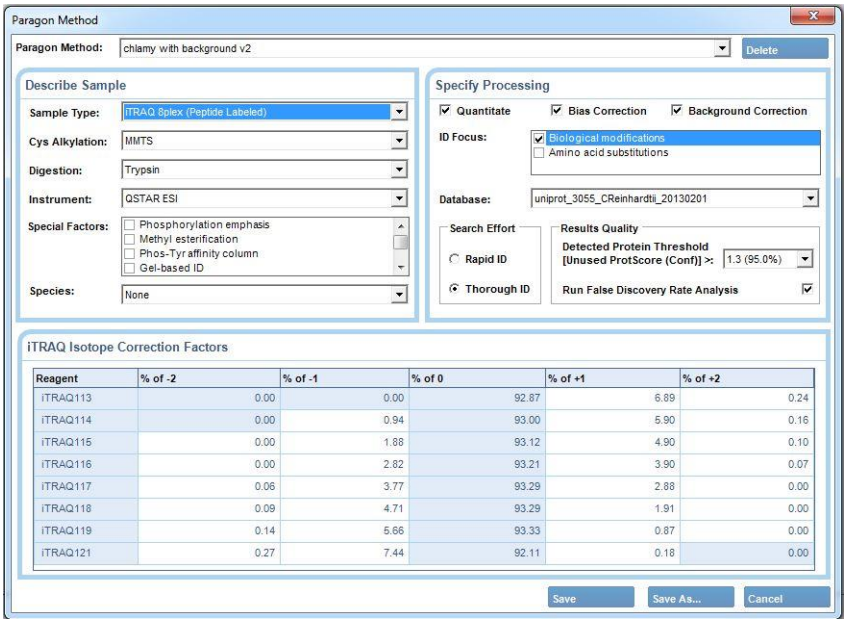


Figure 4.4: Screen print of the search settings for ProteinPilot.

Results handling for identification

Upon completion of the search, ProteinPilot produces a visual overview of the results saved as a group file, as well as an excel file describing information on the false discovery rate if conducted. Within the ProteinPilot itself this data can be visualised in a list with colour coding dictating statistical significance of the change of 114-121, as compared to the base of label 113. This data can then be exported as a text file (*.txt) comprising the peptide or the protein results. For further analysis of this search method, the peptide output was utilised. This output gives several details of each PSM listed in Table 4.1$_{p87}$

Table 4.1: Reported data within the peptide output obtained from ProteinPilot and utilised within the merger script reported in Appendix 4.1$_{p267}$.

Column Title	Description
N	Sequential count of corresponding protein.
Unused	Measurement of all the peptide evidence for a protein which has not been previously assigned
Total	Measurement of all the peptide evidence for a protein
%Cov	Cumulative coverage for matched protein
%Cov(50)	As above excluding those below 50% confidence
%Cov(95)	As above excluding those below 95% confidence
Accessions	Matched protein Accession
Names	Matched protein Accession
Used	Binary of whether peptide was used in protein quantitation calculation
Annotation	Further description of why protein was or wasn't used for quantitation
Contrib	Value describing to what extent the given peptide was used to support the protein identification
Conf	Degree of confidence of match
Sequence	Peptide sequence
Modifications	Modifications identified
Cleavages	Details of missed cleavage point
dMass	The difference in mass between the precursor MW and the theoretical MW of the matching peptide sequence
Prec MW	Observed weight of peptide
Prec m/z	Observed m/z of peptide
Theor MW	Theoretical weight of peptide
Theor m/z	Theoretical m/z of peptide
Theor z	Theoretical charge
Sc	The score for the peptide; this is based on matching ions of various charge states
Spectrum	Code to identify spectrum
Time	Time of precursor elution
PrecursorSignal	Intensity of precursor
PrecursorElution	Time point of elution in LC-MS/MS run
114:113	Ratios label X compared to 113 for each peptide
%Err 114:113	Assessment of the error in the each ration
Continues for each of the 8 label isotopes	
Area 113	Area of the label for each peptide
Err 113	Assessment of the error in each area
Continues for each of the 8 label isotopes	
Background	Quantifier of the background intensity of the label region for each spectra

To process the data, from the peptide output file first the low confidence peptides (below 95%) were removed. From this list, those with a

contribution value of 0 were then removed. With a maximal value of 2.0, the contribution value is determined from the confidence of a protein/peptide/spectral match and the confidence of other protein/peptide matches and peptide/spectra matches for the given protein, peptide or spectra.

The contribution value will therefore be set to 0 if:

- The peptide sequence identified has previously been matched to the protein.
- The peptide sequence has previously been matched to a higher ranking protein identification.
- A different peptide has already been identified from the spectra with a higher confidence.

By removing all matches with a contribution of 0, the resulting list will therefore be devoid of non-unique peptides and where multiple matches can be made the match will be assigned to the most confidently identified peptide and protein.

(Note: When formatting the "protein" output, ProteinPilot only requires that a peptide match is made with a contribution > 0, irrespective of the confidence of the match. Whilst this report also provides the number of matches with a confidence greater than 95%, these matches may not have been those used for the protein identification itself.)

Comparison of identification based on input type

The versions of the data (*.wiff-native/*.mgf-Analyst/*.mgf-Mascot deamon/*.mgf-MS Convert) were run through ProteinPilot and processed as described above. In Table 4.2$_{p89}$, the number of identifications and ≥2 unique peptide identifications are reported for each input method. In parentheses is the number of false positive within each result. The best result was achieved with the *.wiff-native input. This suggests that the internal processing of the native *.wiff file is optimised for the subsequent ProteinPilot search. The two *.mgf(s) created with Analyst and Mascot Daemon gave the next highest identifications, with limited difference between the two. MSConvert gave

significantly poorer results and suggest that this unstructured processing of the *.wiff file did not form appropriate *.mgf spectra for use in ProteinPilot.

The merged versions of all the *.mgf files produced significantly reduced matches. This occurs both for the merged *.mgf file created by Mascot Daemon and for merged *.mgf(s) created by concatenation of the 77 *.mgf(s) made through Analyst and Daemon. This is perhaps due to the way in which ProteinPilot creates its spectral identifiers. These are a code outlined in Section 4.6.1$_{p104}$ comprising the input file (in numerical order uploaded) the scan number and the experiment number. Given a merged input the file input number will be 1 but within that file multiple repeats of scan and experiment number will probably occur such that multiple spectra will have identical identifiers in ProteinPilot. It is this which is likely the cause of the dramatic reduction for merged input.

Table 4.2: Number of identified proteins using ProteinPilot with various inputs. Bracketed numbers indicate the number of decoy hits within.

	# of identified proteins	# of ≥2 unique peptide identified proteins	Percentage of top score at ≥2 unique peptide level
*.wiff direct	1271 (9)	863 (0)	100%
Analyst derived *.mgf	1020 (5)	626 (1)	72.5%
Analyst derived *.mgf merged externally	178 (0)	57 (0)	6.6%
Mascot Daemon derived *.mgf	998 (10)	615 (1)	71.3%
Mascot Daemon derived *.mgf merged by Daemon	246 (0)	84 (0)	9.7%
Mascot Daemon derived *.mgf merged externally	251 (2)	82 (1)	9.5%
Ms Convert derived *.mgf	459 (4)	160 (2)	18.5%

4.4.2: Phenyx

Phenyx (version 2.6) is a licenced program from GENEBIO[307]. Though not currently operating as a commercial entity, it is still an operable program. Phenyx allows input in the format of *.mgf and conducts both a search using its own algorithm and statistical analysis of the results with various outputs in interchangeable forms such as *.csv.

Search settings

Unlike ProteinPilot, Phenyx allows for more control on parameter selection. This can be extremely useful to an advanced user with an understanding of the expected results. However, for a novice user this level of flexibility can increase the use of incorrect or sub-optimal settings.

Some of the parameters required are clear selections such as the digestion enzyme used and modifications expected. However, other parameters can be seen as more dynamic may be set at any point within a scale. Such parameters as tolerance, P-value, AC score (ProteinPilot's accession score) and the number of missed cleavages can be considered dynamic and do not have a clear optimum.

During subsequent investigations, unless stated otherwise, the following settings were used, displayed in the parameters report for Phenyx in Figure 4.5$_{p91}$. The search was simultaneously run against a normal and decoy database for the species under investigation. The instrument set to ESI-QTOF (QStar) and precursor charge of 2+ and 3+ selected. MMTS and iTRAQ_8plex_Nterminal were set as fixed modifications, whilst iTRAQ_8plex_K and Oxidation were set as variable modifications. Trypsin was selected as the digestion enzyme. AC, peptide length and p-values were all set low to facilitate downstream processing with in-house software FDRslice.

IDs	11140
Title	Cr_ CSource_mgf_analyst_0.3_0.1_v1.4
File(s)	part 4.zip (mgf 11253 Kb)
Databank(s)	C_reinharditii_uni_3055 (20130204) C_reinhardtii_uni_decou_3055 (20130204)
AC	
Taxonomy	root
Scoring Model	ESI - QTOF (QStar)
Parent Charge	2,3 (trust=medium)
Round #	1
Modifications	MMTS[fixed, all] iTRAQ_8plex_Nterm[fixed, all] iTRAQ_K[variable, <=4] iTRAQ_8plex_K[variable, <=4] Oxidation_M[variable, <=4]
Enzyme	Trypsin_(KR_noP) miss. cleav. 1 cleav. mode. normal
Parent tol.	0.3Da
Pept thresholds	length>=4 score>=4.0 p-value<=1.0E-4
AC Score	4.0
Conflict resolution	yes
Turbo scoring	tolerance=0.1Da coverage >=0.2 series=b;b++;y;y++

Figure 4.5: Screen print of the search settings for Phenyx (version 2.6).

Using FDRslice the output of any Phenyx run can be conformed to a fixed FDR at the peptide level. As such, during the run, statistical parameters such as the score, p-value and AC score are set at low levels giving the fullest set of peptides to then restrict.

This dynamic threshold selection based on FDR reduces a source of researcher variation where some researchers would choose to sacrifice identifications for an improved FDR or vice versa.

Selection of ion tolerance

A dynamic parameter that must be selected is the precursor and product ion tolerance. This should be set based on instrument type and the resultant accuracy expected. Thus an ion trap with less *m/z* accuracy would need a higher tolerance that that of a time-of-flight instrument. However, the tolerance should not be set too high as this will lead to increased matches to decoy spectra and restrict the number of validated peptides.

To investigate the effect of varying the precursor and parent ion tolerance on our data, repeated runs were conducted changing only these parameters (reported in Table 4.3$_{p92}$). From this assessment, 0.3 was selected as the optimum setting for precursor tolerance for this dataset. It is also notable that the variation between all the tolerance settings is negligible, varying only 1.3% of the midpoint. Thus for a relatively high precision instrument such as that used here (QTOF) it is not a significant factor once data has been curated to significantly identified proteins with two unique peptides.

Table 4.3: Number of identified proteins using Phenyx with various 'Product ion' and 'Precursor ion' tolerances. Each combination of tolerances has the number of identified proteins and number of ≥2 unique peptide identified proteins reported. All lists were restricted to a peptide FDR of 0.03 using FDR slice.

	Precursor	Product	# of identified proteins	# of 2 unique peptide identified proteins
Fixed Product Tolerance	0.1	0.1	1483	791
	0.2	0.1	1490	789
	0.3	0.1	1494	804
	0.4	0.1	1558	796
	0.5	0.1	1524	787
Fixed Precursors Tolerance	0.3	0.001	1509	805
	0.3	0.01	1516	804
	0.3	0.025	1506	798
	0.3	0.05	1512	805
	0.3	0.15	1503	803
	0.3	0.2	1501	801
	0.3	0.3	1492	797

Results handling for identification

Phenyx allows for multiple output settings. For this investigation, the iTRAQ output was used providing the details shown Table 4.4$_{p93}$ for each PSM. This was then processed by in-house software FDRslice which after ranking all PSM(s) restricts the list of PSMs such that the FDR is as desired. For this and subsequent uses, an FDR of 3% at the peptide level was used. This equates to a 0.09% FDR for protein identifications when requiring the ≥2 unique peptide for protein validation.

Table 4.4: Reported data within the 'Global Export' output obtained from Phenyx and utilised within the merger script as reported in Appendix 4.1$_{p267}$

Column Title	Description
jobid	The search number
AC	Accession match
sequence	Sequence of PSM
charge	Precursor ion charge
zscore	Statistical factor
pvalue	Statistical factor
is selected	
moz	Mass over charge of precursor
miss. cleav.	Number if miss cleavages
modif	Any modification's found
113.1	Sum of all intensities at 113.1 ±0.2
Continues likewise through to 121.1	
spectra descr.	Title line of spectra from the .mgf
pm key	Description of the peptide sequence including modifications

4.4.3: Search GUI/PeptideShaker (OMSSA & X! Tandem & Mascot)

SearchGUI[283] (version 1.12.2) and PeptideShaker[284] (version 0.20.1) combine to form an open source alternative to the search engine and identification analysis steps within the proteomics data pipeline. This combination best demonstrates how the search engine and identification analysis, though commonly combined, are two separate tasks. In addition to the output of SearchGUI, PeptideShaker also accepts properly formatted Mascot searches as an input.

SearchGUI

SearchGUI[283] provides a graphical user interface to run two open source command line database search engines, Open Mass Spectrometry Search Algorithm (OMSSA)[281] and X! tandem.[282] It accepts input in the form of *.mgf and FASTA files which it processes into a concatenated Target-Decoy database to perform a false discovery rate search against.

Settings were set as was the case for Phenyx using an optimised precursor and product ion tolerance of 0.3 and 0.1 Da respectively. A screen print of the parameters used is displayed in Figure 4.6$_{p94}$.

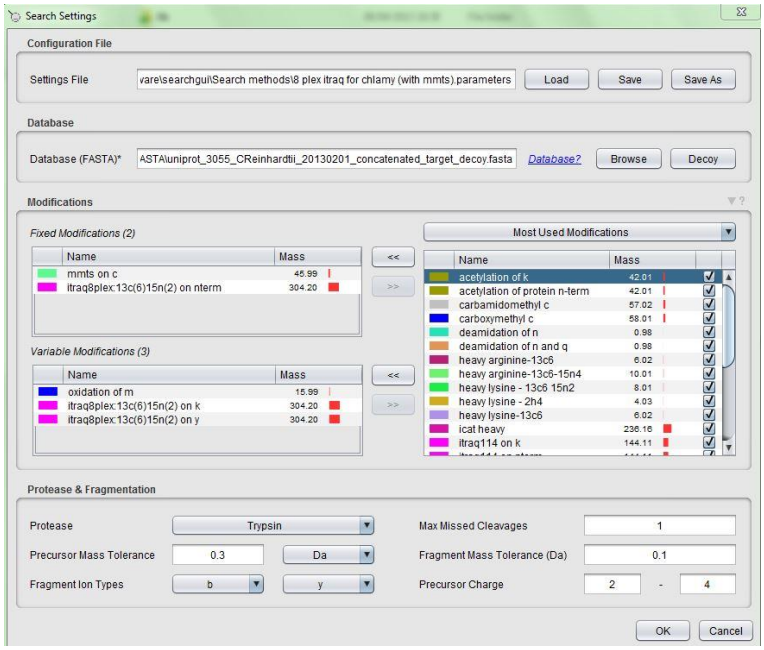


Figure 4.6: Screen print of the parameters setting pane of SearchGUI, with parameters set as used for current analysis.

Mascot

Mascot is a proprietary search engine from Matrix Science.[304] It is also the most commonly used search engine in general (Figure 4.3$_{p85}$) and the second most common for iTRAQ investigations[306]. It typically runs on a server with job submission remotely through Mascot Daemon (version 2.3.2). In addition to acting as a search engine portal, Mascot Daemon can process raw file into *.mgf formatted files. The search parameter page is shown in Figure 4.7$_{p95}$. These were set as was outlined for Phenyx. To make the result compatible with PeptideShaker, the FASTA database used was the target decoy database used by SearchGUI.

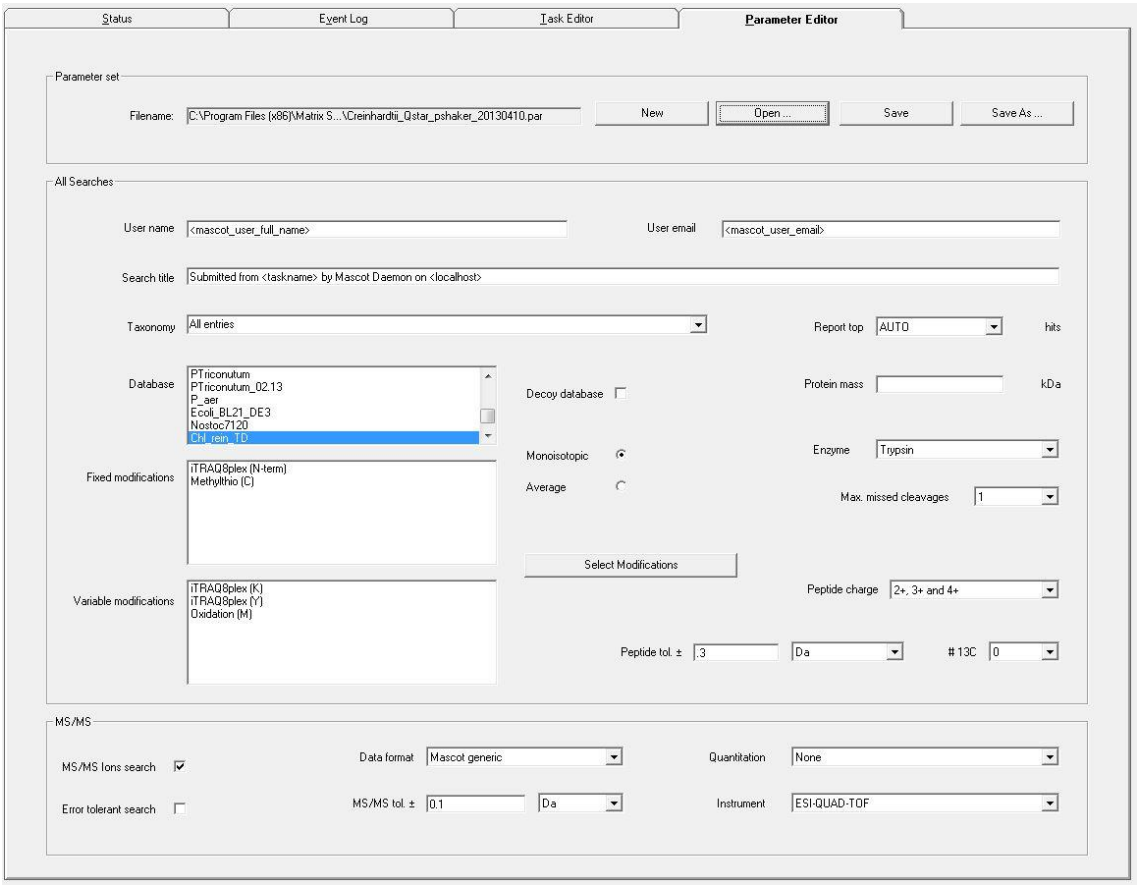


Figure 4.7: Screen print of the parameter setting page of Mascot Daemon (version 2.3.2).

PeptideShaker

PeptideShaker is a data processing program under active construction from Compomics.[284]. It accepts search engine results from OMSSA, X! Tandem and Mascot, combines these results and preforms analysis on the identifications to determine the FDR for each subsequent PSM according to the number of decoy identification scoring higher than the current PSM. Given this the assignment identity threshold can be set based on a FDR threshold as with FDRslice.

As well as reporting the merged and statistically analysed list of identifications from OMSSA, X! Tandem and Mascot. PeptideShaker also provides several assessment tools such as fraction location, coverage mapping on 3D structures and links for each protein to appropriate databases such as UniProt,[280] String[308] and David.[288]

The PSM output of PeptideShaker (Table 4.5$_{p96}$) was used for the merger process reported subsequently.

Table 4.5: Reported data within the PSM output obtained from PeptideShaker and utilised within the merger script as reported in Appendix 4.1$_{p267}$

Column Title	Description
Protein(s)	Accession ID(s) of matched proteins (if multiple matches each is separated by ",")
Protein(s) Descriptions	Description for each match based on FASTA input
Sequence	Sequence of PSM
Spectrum File	Source *.mgf file
Spectrum Title	Title of spectra from *.mgf
Score	Statistical factor
Confidence	Statistical factor

4.4.4: Peaks

Peaks (version 6.0) is a proprietary search engine available from Bioinformatics Solutions.[303] It is noted as a *de novo* search engine deriving sequence identifications without a matching database. In addition to the *de novo* functionality, Peaks can also conduct database searches. Peaks was run using the same parameters as for the other previously utilised software. A screenshot of the search parameters can be seen in Figure 4.8$_{p97}$.

I Peaks has several other functionalities its inChorus feature which allows for the running of several open source software for comparison to Peaks including Mascot, OMSSA, Sequest and X!Tandem. Peaks also provides several analysis tools unavailable in ProteinPilot, including a continual FDR allowing an FDR based threshold at the peptide level and graphic illustrations of quantitations, such as heatmaps, superior to that of ProteinPilot.

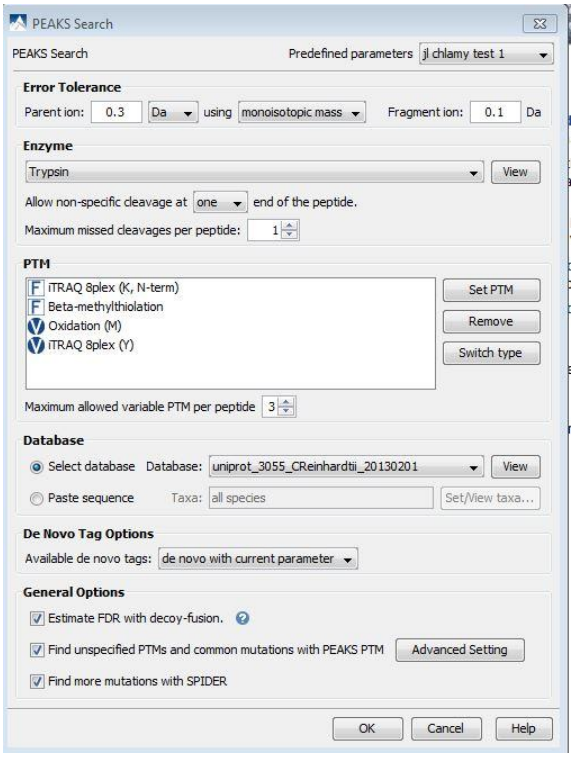


Figure 4.8: Screen print of the search settings for Peaks (version 6.0).

As with PeptideShaker, Peaks ranks identifications based on a statistical factor termed '-10lgP'. The continuum of FDR scores for each PSM is then calculated such that a given FDR score can be selected and set as a threshold for inclusion in validated results. For the merger procedure conducted subsequently a threshold of 3% was selected. Data was exported as a *.csv file with information for each PSM as reported in Table 4.6$_{p97}$.

Table 4.6: Reported data within the peptide output obtained from Peaks and utilised within the merger script as reported in Appendix 4.1$_{p267}$

Column Title	Description
Peptide	Sequence (including modifications)
-10lgP	Statistical factor
Mass	Deconvoluted mass of precursor ion
ppm	Precursor ion charge
m/z	Mass over charge of precursor ion
RT	Retention time
Scan	Spectra count (unique identifier)
Accession	Accession ID
PTM	Y/N for presence of PTM(s)

4.5: Comparison of identification

Due to the complex nature of the identification programs there are multiple ways each can be run. This therefore allows for considerable bias to be placed for a particular program to outperform others for a given comparison. By this means, a developer could run the search in such a way as to benefit their software program such that it outperformed the competition.[300]

Furthermore, due to variance in how each program deduces statistically significant identifications, a strong argument can be made that the greatest number of identification may not be the best output to judge performance and that a reduced but more reliable list is preferable.

Here, we have tried to utilise each program in an unbiased fashion to reveal the best data from each. We used the peptide FDR where possible as a cross platform confidence measure using 3% as the cut-off. With this 3% peptide FDR we should therefore have a theoretical 0.1% FDR on the protein level given the ≥2 unique peptide requirement.

From Table 4.7$_{p99}$ it can be seen that given the stated conditions, set to be unbiased, ProteinPilot achieved the best identification rate followed by Phenyx and Peaks. PeptideShaker which is an amalgamation of Mascot, Omssa and X! Tandem was the poorest performer. This was less notable when allowing multiple accession assignments to any given spectra with assignment being made to the most frequently matched accession (if equal first alphabetical order). This is perhaps indicative of the way in which PeptideShaker handles shared peptides. In addition, given the use of three search engines, PeptideShaker identification is already curtailed to PSM, which have maintained reproducibility between the three search engines.

Table 4.7: Reports the various identifications found with each software program. Currently it is apparent that ProteinPilot gave the greatest number of identifications for followed by Phenyx and Peaks. In brackets are the number identified with multiple accession matches.

	# of 2 unique peptide identified proteins
ProteinPilot	859
Phenyx	786
PeptideShaker	372 (482)
Peaks	516

Figure 4.9$_{p100}$ gives a Venn diagram of the identified proteins (≥2 peptide) from the four search engines. Between the four search engines, 1102 proteins were confidently identified with 247 being identified reproducibly by all 4 engines. Thus a combined approach using multiple search engines could increase identification a further 28.3 % on top of that achieved by using the most abundant single search engine alone.

An alternative to the cumulative approach would be a restrictive approach accepting only those reproducibly identified by all search engines. Whilst it can be argued that the 247 reproducibly identified proteins now have increased confidence, the reverse that those not identified by all search

engines had reduced confidence would be incorrect, rather their confidence level remains the same.

In a similar fashion, that those identified by all search engines can be said to have increased confidence. Those identifications made by three search engine (459 proteins) and those made by two search engines (725 proteins), can also be said to have a level of increased confidence.

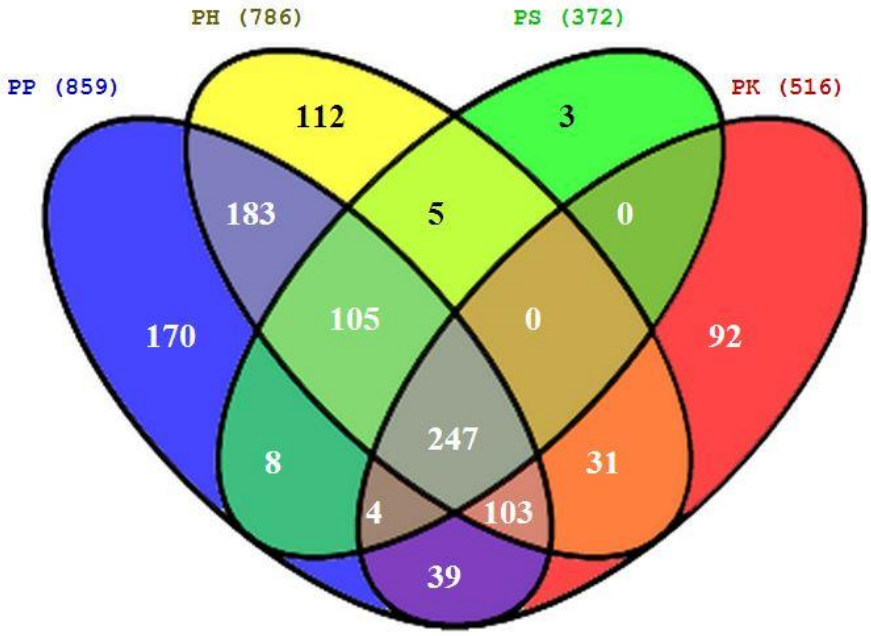


Figure 4.9: Venn diagram of the identified proteins (≥2 peptide) from each search engine. PP (ProteinPilot), PH (Phenyx), PS (PeptideShaker), PK (Peaks).

4.6: Search engine merger

4.6.1: Existing merger tools

Given the comparison between multiple search engines, the next logical step would be to merge multiple search engines together. Software capable of such merging includes:

- Peaks[303]: Included within the proprietary Peaks software is inChorus. This is designed for the merger of Peaks DB, Mascot, OMSSA, SEQUEST and X! Tandem search results. Reporting the consensus matches between search engines (ignoring non matches) InChorus provides a convenient merger tool if already working within the Peaks

environment, however apart from Mascot it doesn't allow for input from other proprietary search engine such as ProtienPilot.

- Scaffold[309]: Propriety software from Proteome software (Proteome software Inc., Portland, Oregon, U.S.). This software accepts inputs from multiple search engines (including: SEQUEST Mascot X!Tandem Phenyx OMSSA Spectrum Mill Waters IdentityE and Proteome Discoverer). Using each search engine's ranking system is used to create PSM probabilities, which in turn can be used to combine matches from multiple search engines.[310] PSM matches are therefore retained as valid based on agreement between software and the scores associated with each individual software utilised. Thus promoting multiple observations of low significance matches and demoting single observations or contradicting observations of previously high scoring PSM matches. Scaffold then combines peptides into proteins using a novel algorithm to define groups and the ProteinProphet algorithm to define probablilities.[309]
- MassSieve[311]: Open source Java-based software is designed for visualizing and parsimony analysis of database search engine results. It can intake common formats of Mascot*.dat files, OMSSA and X!Tandem *.xml files and the standardised pepXML output if available. The software can be used to merge these search engine outputs either removing any spectra peptide discrepancies or retaining and labelling appropriately. The software them makes its own assessment to organise peptides into protein based on parsimony.
- iProphet[312]: Part of the Trans Proteome Pipeline This software combines evidence from multiple identifications of the same peptide sequence across multiple spectra, experiments, precursor ions as well as multiple search engines. It is currently able to combine results from Sequest, Mascot, ProbID, X!Tandem and Phenyx.

- PepArML[313]: An meta-search engine PepArML searches of Mascot, Tandem, OMSSA, KScore (Tandem with the K-Score scoring plugin), SScore, Myrimatch, and InsPecT. Interacting with the associated groups servers and local resources several open source search engines can be combined with Mascot to give a merged list of results. Apart from Mascot it doesn't allow for input from other proprietary search engine such as ProtienPilot.
- MSblender[314]: This software package (for Linux or Mac) intakes several search engine PSM and match statistics and converts them to a standard probability score allowing for cross comparison and merger of search engine data sets.

4.6.2: Developmetn of an R based merging method

Many of the previously reported merging software are capable of merging multiple search engines effectively. However without a standard format for search engines results none of the merging software is capable of merging all search engines. In fact none are capable of incorporating ProteinPilot into the merger, which was identified as the most successful search engine in Table 4.7$_{p99}$. For this reason the derivation of self-derived merging method was selected for the work described in this thesis. To accomplish this in an open accessible clear text format the programming language R was used.

At a simplistic level this could be achieved in two fashions. Either a cumulative total of all the search engine results to provide a higher identification rate, or a restriction to accept only those identifications made by all search engines providing an increased level of certainty to the identifications. I.E. either 1102 or 247 respectively when using the results of the 4 software comparison presented in Figure 4.9$_{p100}$.

The latter option could be seen as restrictive and reducing processing to that of the poorest of the search engines, whilst the former can result in inaccuracies and support multiple matches to single MSMS spectra. Whilst

such a two-match scenario is possible, given two different peptides with similar *m/z* co-eluting, it would be a rare event and still should be excluded from any quantitative analysis.

Some more complicated methods of merging multiple search engines have been published including:

- Vote counting system[315] – In this system, the identifications are ranked first in an order of the number of search engines identifying a particular PSM and subsequently by a combined *p* value based on the significance value of the individual search engines and the FDR of that search. A threshold N_{SELECT} is then set and the top N ranked proteins carried forward.
- FDR score [316] – In this method a combined FDR score is made for all spectra matches. This is achieved by ranking all target decoy matches and calculating the FDR (using Kall *et al.*'s method[317]) at each PSM subsequently. The target-decoy matches are then ordered in reverse order and the minimum FDR score observed thus far is reported for each PSM. This creates the New FDR score for each PSM. This process is repeated for each search engine involved. To combine FDR scores from multiple search engines the χ root of the multiple of the FDR scores, where χ is the number of search engines being combined is calculated. All data can then be merged and ordered by this new FDR selecting a threshold cut-off.
-

Whilst valid, these three methods of combination do have significant drawbacks. The vote system major effect is to select those PSM more reproducibly identified. The degree to which it is akin to either the simplistic cumulative or simplistic restrictive approach is based on where the arbitrary threshold is set. A parameter which cannot be deduced based on the existing data presented. The FDR score and MSblender methods do make a significant

improvement; however, their implementation is complex and software and search engine restrictive.

An alternative method attempted here is to take a simpler accumulative approach removing or resolving instances of conflict (Table 4.8$_{p104}$). Each search engine was used to rank its own PSM(s) with a cut-off set to achieve a specified FDR. With each PSM thus passed an accepted level of accuracy, all PSM (unless disproven) be considered accepted. From the combined search engines, evidence contradicting PSM(s) can be considered multiple sequence matches to any given MS/MS scan identity (msmsid) or multiple accession matches. The remaining PSMs can then be passed on for further analysis.

This merger procedure would not retain the FDR score however, given all the input search engine results were limited to a FDR bellow 0.03 at the peptide level the overall FDR could not be higher than this. Further as both known false positive and unknown false positive are more predispositions to conflicts between the search engines they will be predisposition to selection for removal reducing both in the final PSM list.

Table 4.8: Table showing the possible combinations when combining two search engines. Result when requiring only sequence and both sequence and accession match are shown (green retained, red rejected).

MSMSid	Serach engine A		Serach engine A		Sequnce	Sequence & Accession
	Sequence	Acession	Sequence	Acession		
1	HPES	id 1				
2			SMT	id 2		
3	HTRWGNL	id 3	HTRWGNL	id 3		
4	NAISMI	id 4	NAISMI	id 5		
5	NAHTAN	id 6	WEHTAM	id 7		

4.6.3: Method of merger

To combine the search engine results and process the data the programming and statistical package R[318] was utilised. Figure 4.10$_{p106}$ shows a flow diagram describing the operations conducted in the script, with the script being reported in Appendix 4.1$_{p267}$.

To combine the data an msmsid was created representing each MS/MS scan as reported from within the mfg. files used. The format for the id is "A.B.C ". "A" is the numerical count of the mass spectrometer run (in alpha numerical order). "B" is the scan number reported in the spectra title. "C" is the experiment number reported in the spectrum scan title. Where multiple scan or experiments were combined within the *.mfg, the first scan or experiment number was used.

For ProteinPilot the spectrum (aka scan number) reported is a similar numerical code to that derived for the msmsid consisting of "A.X.Y.B.C". The "X" and "Y" were removed using R such that the code matched that of the msmsid. For Phenyx and PeptideShaker, which report the scan title with the PSM match the msmsid was processed in R. For Peaks, scans are reported as "F5.6835" where "F5 indicates the *.mfg. and "6835" is a numerical count of all the input scans from all *.mfg(s) in sequential order. The Peaks count was obtained from the reporter ion intensities, which likewise was ordered as a sequential list of all the combined MS/MS scans.

It should be noted that although accurate for >99% of the ProteinPilot identifications, the ProteinPilot spectrum code does not consistently use the first scan and experiment number where multiple scan and experiment are enclosed in the MS/MS title. These MS/MS spectra could not be correctly assigned and thus were removed from further analysis.

Once merged, any scans that matched different sequences were removed as questionable identifications. The remaining scans were then either re-formatted for import into in house 'uTRAQ' or curated further to remove any scans which matched to more than one protein accession before re-formatted for import into uTRAQ. uTRAQ is a program run using Perl and Long Hanbourgh, UK). It takes in a PSM list with iTRAQ labels and report the median corrected (MC) and Isotopic corrected (IC) average label intensities for each identified protein.

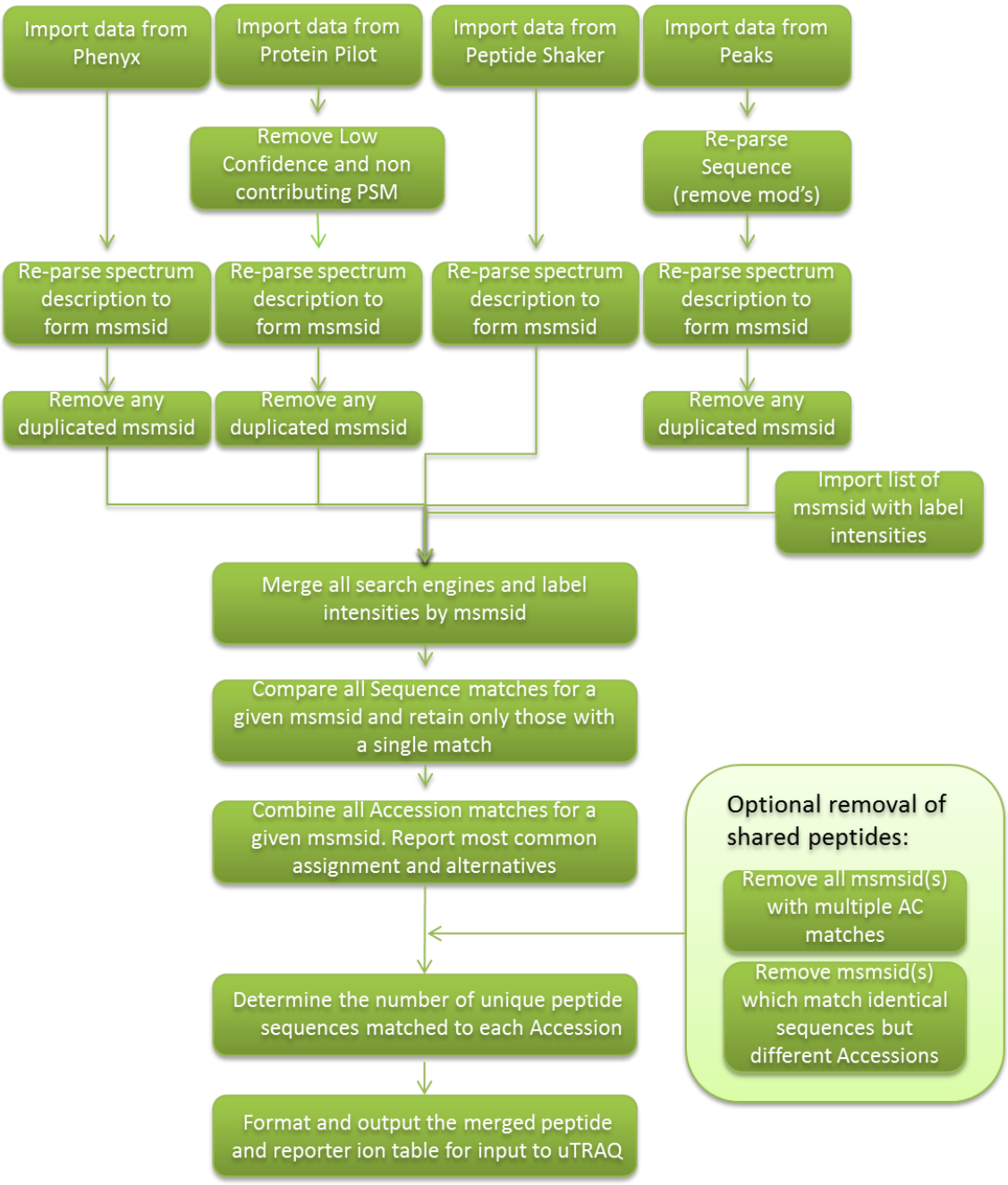


Figure 4.10: Schematic diagram of the processing conducted in R to merge the search engine outputs together. Describes the process accomplished by the R script in Appendix 4.1$_{p267}$

4.6.4: Comparison of merged search engine results

After combining the four search engines with the script, a total of 28944 PSMs were retained, which represented 7969 unique sequence identifications and 1960 protein identifications with 1007 identified by two or more peptide matches. This represents a 17.2% increase compared to ProteinPilot alone, the best single search engine. Compared to the simplistic restrictive

approach, it is a 307.7% increase and an 8.6% decrease from the simplistic cumulative approach. The number of identified proteins (≥2 peptides) achieved for each combination is shown in Figure 4.11[p107].

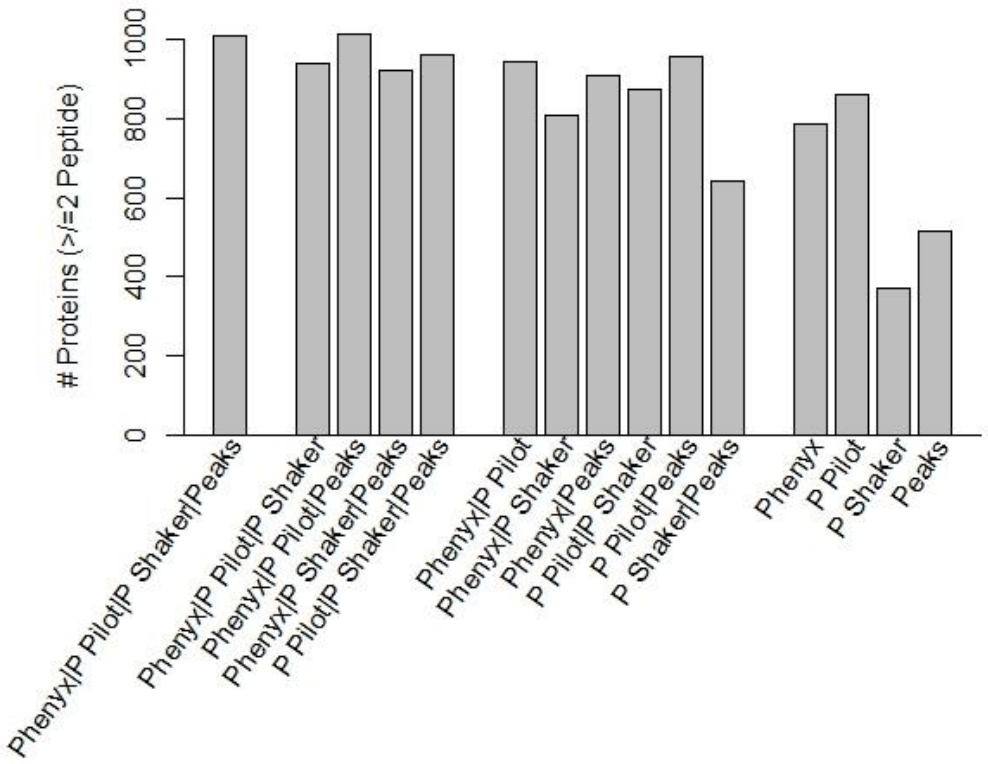


Figure 4.11: Bar chart showing the number of ≥2 peptide identifications made for various combinations of search engines merged together.

To understand what has been lost from each search engine when run separately as compared to the merged dataset, the ≥2 peptide proteins from each engine were compared to the merged dataset (Figure 4.12[p108]). The protein, rather than PSM level, was used to demonstrate the effect on the results that would be continued on to further analysis based on the processing used.

The comparison was conducted both when removing scans without a single sequence and accession matches (Figure 4.12A[p108]) and for scans without sequence match irrespective of accession match (Figure 4.12B[p108]). For ProteinPilot, Phenyx, and Peaks run individually the requirement for single AC match had no effect as these search engines either discount PSM with multiple AC matches or report only a single AC for each PSM.

When considering protein loss due to sequence miss-match (Figure 4.12B$_{p108}$), it can be seen that Peaks drops 5.2% of its individual matches compared to ProteinPilot, which drops 3.0%. This is suggestive that compared to each other, Peaks makes the less reliable sequence identifications with its' PSM(s) more often being lost from the merged pool.

When comparing Figure 4.12A$_{p108}$ vs. Figure 4.12B$_{p108}$ the increase in removal from the merged pool for ProteinPilot, Phenyx and Peaks demonstrates that the level to which given correct sequence identification each makes a non-consensus accession assignment. Here, Phenyx appears to perform poorly, increasing by 2.4% compared to <1.4% for either ProteinPilot or Peaks. The decrease in loss from PeptideShaker suggests that if restricted to PSM made to single AC then the identifications made by PeptideShaker are unlikely to be questioned. This is expected, as PeptideShaker is already an amalgamation of three search engines and as such had already reduced its questionable spectra.

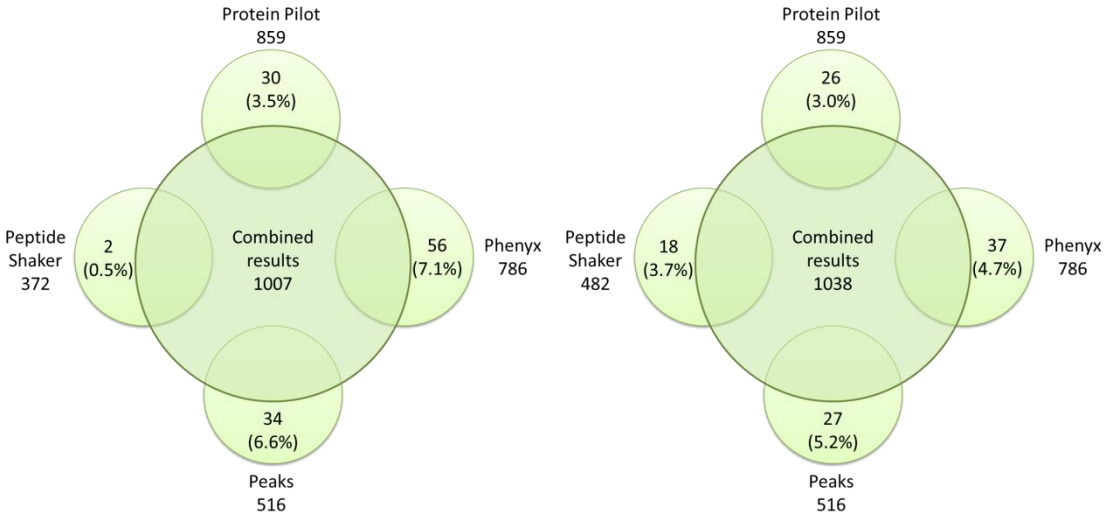


Figure 4.12: How each individual search engine compares to the combined search results (≥2 unique peptide level). A) Results removing PSM's with multiple AC and Sequence matches. B) Results removing PSM's with multiple Sequence matches but retaining those with multiple AC matches.

4.6.5: Investigation of a case of miss match

To look further at the reason for identification loss in the merged pool, a protein (A8HQG9) was selected that was lost from ProteinPilot compared to

the merged data. The msmsid(s) linked to the protein were selected and the identification that each search engine gave to those msmsid(s) reported (Table 4.9$_{p109}$).

Table 4.9: Scans used to identify protein A8HQG9 in ProteinPilot and the corresponding identifications made by other search engines to these Scans. A8HQG9 was selected as a sample as a protein lost from the merged dataset due to sequence mismatches.

msmsid	ph AC	ph SEQ	pp AC	pp SEQ	ps AC	ps SEQ	pk AC	pk SEQ
36.1574.3	A8HQG9	VAEVTEGAEAR	A8HQG9	VAEVTEGAEAR	A8HQG9	VAEVTEGAEAR	A8HQG9	VAEVTEGAEAR
37.1574.3	A8HQG9	VAEVTEGAEAR	A8HQG9	VAEVTEGAEAR	NA	NA	A8HQG9	VAEVTEGAEAR
10.1947.3	NA	NA	A8HQG9	LSGDADVGCLFFSER	NA	NA	A8I6T5	ELLYLDLALEDQAR

Of the three scans used to match identify A8HQG9, two (36.1574.3 & 37.1574.3) matched the same sequence and one (10.1947.3) identified a second unique sequence. This second sequence was lost from the merged data, as it identified a different sequence in Peaks.

The ProteinPilot and Peaks spectra which matched (36.1574.3) and those which did not match (10.1574.3) are displayed in Figure 4.13$_{p110}$ (A & B and C & D respectively). Scrutinising these spectra, it is observable that the identification which corroborated Figure 4.13$_{p110}$ (A & B) are good matches to the spectra with the major ions being used to match many of the predicted product ions. Compared with this, the spectra with different matches can be described as poor, with predicted ions covering less of all the ions as well as less intense ions.

This loss demonstrated the benefit of multiple search engines at removing 'poor' PSM(s). Without manual curation, the spectra would have been used by both search engines for different proteins. In the case of ProteinPilot, it would have supported evidence of a protein to the '2 unique peptide' validation level with weak evidence. Furthermore, for Peaks which matched the spectra to A8I6T5 it would have included the quantitation values of this spectra to that of the A8I6T5 proteins perhaps erroneously.

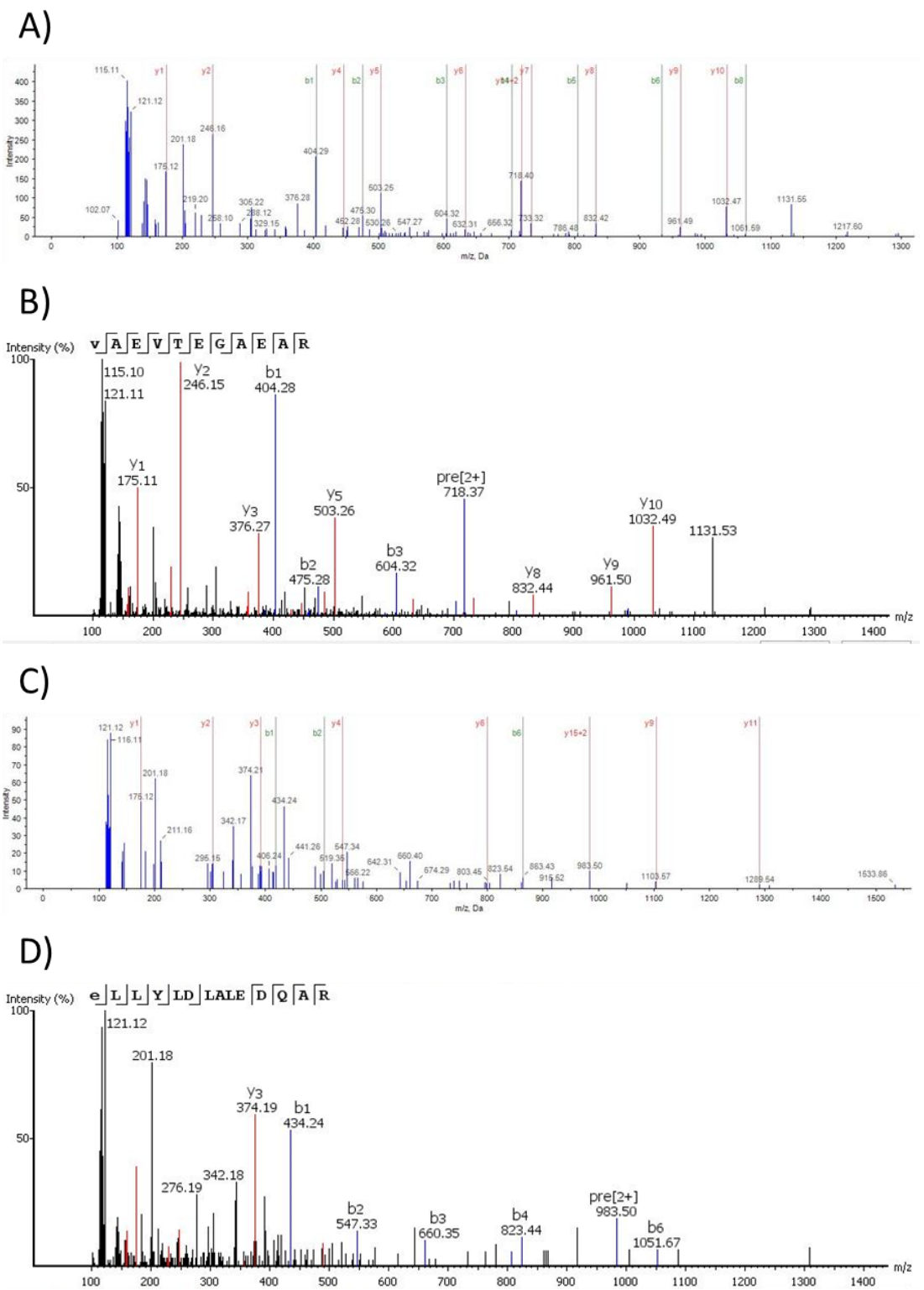


Figure 4.13: Spectra showing PSM matches of ProteinPilot and peaks (A,C and B,D respectively) for spectra 37.1574.3 and 10.1947.3 (A,B and C,D respectively) which represent validated matching PSM(s) and bad miss matching PSM(s) between the two search engines.

4.6.6: Conclusion of merging

Merging the data in the manner outlined achieved a 17.2% increase in protein (≥2 peptides) identifications as compared to the highest ranking individual search engine. This was achieved without inclusion of 'non-rational PSM(s)' with multiple matches to a single spectrum, which would have been the result of a straight merger of the data. Furthermore, as poor quality spectral matches are more likely to be differently identified as opposed to high quality spectral matches, the merger of the data has increased the confidence in the quality of the PSM as a whole. The merging technique applied here was utilised throughout out the rest of this thesis to generate the list of identifications reported.

4.7: Comparison of quantitation

For an identification based investigation, this list of identified proteins would be the conclusion of the proteomics data pipeline. However, for a quantitative investigation an analysis of the degree and significance of any observed changes should be made.

4.7.1: Integrated methods

Given the establishment of iTRAQ reagents as a proteomics technique, many search engine suppliers have to some extent integrated iTRAQ into the software workflow. The degree of processing also varies from software to software. Phenyx, for example, will report the raw label intensities matched to the relevant PSM. Other software programs such as ProteinPilot will additionally apply correction methods such as background and bias correction.

Although some of the analysis may identify significant changes for one label to another, none provides an assessment of significance based on the biological variation observed.

4.7.2: Standalone methods

With the development of iTRAQ, several standalone software programs have been developed for use in such investigations. These have included IsobariQ[319] and iQuantitator.[320] These programs have varying degrees of success.

The program IsobariQ processes the output of the Mascot search engine, preforming various label adjustments such as VSN. The results are then reported with a level of significance change observed. IsobariQ, however, requires the use of Mascot alone, a search engine, previously identified, in combination with OMSSA and X! Tandem, to provide approximately half of the identifications of other search engines (Table 4.7$_{p99}$).

iQuantitator, unlike IsobariQ, accepts as input a table format containing iTRAQ label peak area, peptide and protein matched. This is a report obtainable from most search engines. iQuantitator however has not been maintained through successive generations of R and as such currently requires considerable de-bugging to function.

4.7.3: Methods utilised in this investigation

To be able to utilise the combined dataset formed for the merger of the six search engines derived above, a new method of label analysis was required. The analysis preformed can be broken down into several steps, the outline of which can be seen in Figure 4.14$_{p113}$.

To process the data, iTRAQ label intensities were first matched to the PSM list. This could have been achieved through merger of the reported label intensities as they are outputted by the various search engines being combined. Such a merger would be complicated by the varying ways each output reports the label intensity; Peak area vs. centroid peak and the degree of processing such as isotopic correction performed by each search engine. The alternative utilised here was to retrieve the label intensities from the raw *.mgf files. This was achieved with a Perl script that progressively

extracted the spectra title from each *.mgf and formatted a spectral identity as described earlier. This was then coupled to the sum of intensities associated with each label 113.1:121.1 with a tolerance of 0.2 *m/z*. These were merged to the PSM list in the script shown in Appendix 4.1$_{p267}$. One fault with this method is the loss of some labels from some spectra matched using the ProteinPilot *.wiff based import.

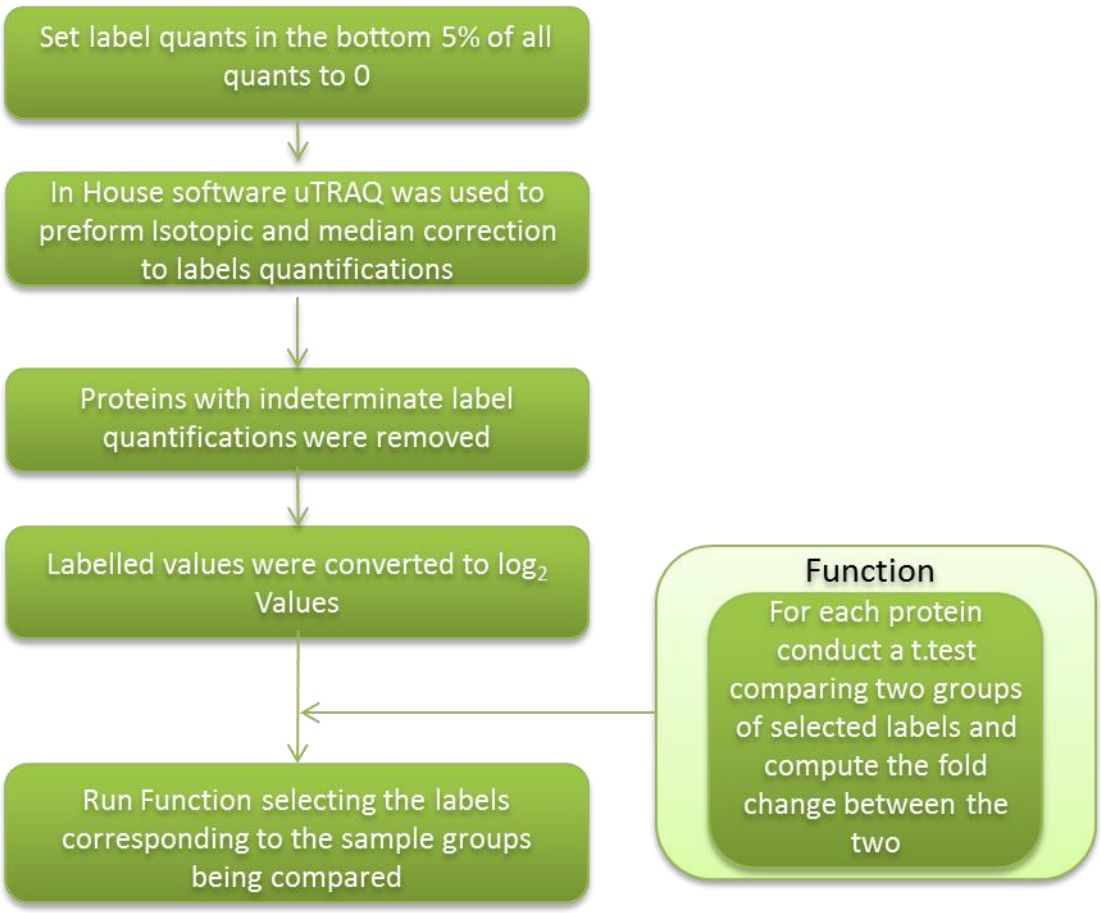


Figure 4.14: Schematic diagram of the processing conducted to compare biological conditions under investigation. Describes the process accomplished by the R script in Appendix 4.2$_{p273}$

4.8: Label correction

Though the raw iTRAQ intensities can be uses directly, several processing techniques can be applied to improve statistical analysis. This includes variance stabilization normalisation (VSN), isotopic correction and loading correction (aka median correction).

VSN has commonly been used in microarray datasets. It attempts to correct for the varying variance depending on intensity that is inherent to both microarrays and iTRAQ datasets [321]. To preform VSN, an R package, released as part of the bioconductor suite, was utilised. This manipulation was performed on the data as part of the script reported in Appendix 4.1$_{p267}$. Figure 4.15$_{p114}$ shows the effect of VSN on the dataset. From the untreated but log_2 transformed data in Figure 4.15A$_{p114}$ to the treated data in Figure 4.15B$_{p114}$ the variation can be observed to be more uniform. This is particularly true for the low intensity ions which are reduced. This treatment is extremely beneficial in reducing erroneous large differentials (e.g. 113/114) which are a result of large degree of noise at the low intensities.

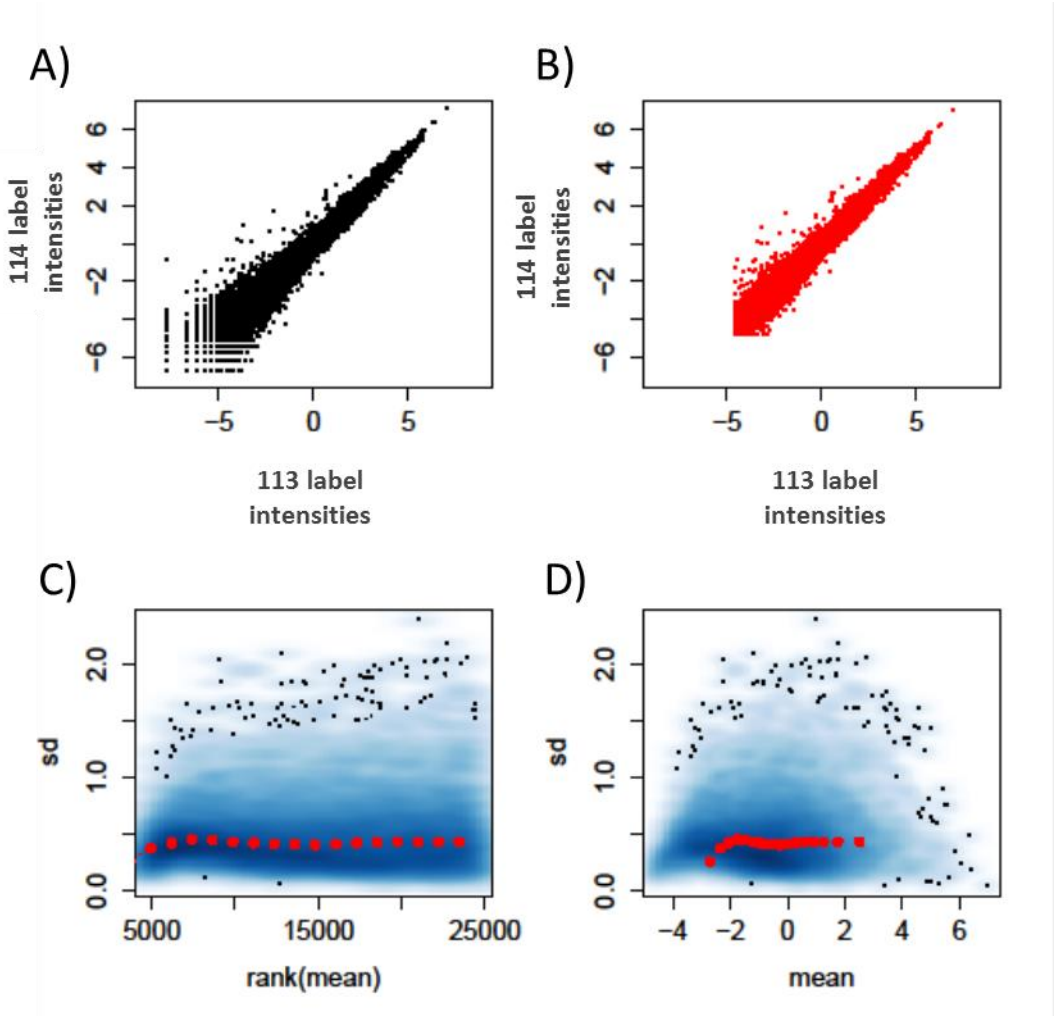


Figure 4.15: Graphs showing the VSN manipulation on the data. A) Untreated comparison of label 113 vs. 114 transformed by log2. B) VSN treated comparison of 113 vs. 114. C&D) Ranked and unranked means of the resultant distribution.

To conduct isotopic and loading correction, an in-house Mathematica script called "uTRAQ" was utilised. Effectively, loading correction brings the mean label intensity for each isotope 113:121 into alignment. This corrects for differences in loading when labelling the various samples under investigation. It is based on the assumption that rather than absolute quantitation it is the change in proportion of the entire identified protein that are of relevance. Isotopic correction corrects for the intensity of adjacent labels. This corrects in such a way that an intense reading for 113 &115 does not artificially increase 114.

4.9: Determination of significance

Once corrected, the next vital step is to determine how to identify significant changes between the biological replicates. How to accomplish this has been a matter of much debate and inconsistencies in the literature.[306,322–325] This can be considered a problem of channels. With many mass spectrometric investigations there is a limit to the availability of the mass spectrometer. As such a limited number of runs could be used for replication. This creates a challenge between the desire to derive information from the replicates available and the desire to increase the statistical rigour of the investigation.

To minimise this fault it was often assumed that each peptide and observation of a peptide could be considered a replicate of a protein observation. These however are pseudo-replicates and should, if avoidable, not be used as true replicates.

With reduction in cost and increased availability of mass spectrometers, allowing multiple replicate runs and quantitative methods (e.g. TMT[156], ITRAQ[157]) allowing multiple channels to be simultaneously run, this problem has been reduced. However to maximize throughput it is still common to run experiments with the minimum number of replicates possible allowing for maximum number of biological comparisons. This is perhaps most evident in 8-plex iTRAQ investigations where one condition can have a minimum of

a 3 vs. 3 comparison the remaining channels are sometimes used for a secondary comparison rather than to increase the significance of the primary comparison.

Perhaps the best method to deal with this conflict is the incorporation of biological knowledge. As mentioned above each peptide cannot be considered a true replicate, but rather is a pseudo-replicate. The same is not true however of protein with the observation of each being independent. Therefore, given a set of biologically related proteins demonstrating a consistent pattern the statistical trust in the pattern ban is bolstered. In practice this can be used as two thresholds. A more stringent to suggest individual protein observations and a less stringent for patterned observations such as gene ontology or pathway based observations.

Figure 4.16$_{p117}$ shows three different attempts to assess the significance of the protein changes. Figure 4.16A$_{p117}$ is an in-house tool termed "SignifiQuant" which uses pseudo-replicates in the form of multiple spectra, matching a given protein to determine significance. Figure 4.16B$_{p117}$ uses an R script shown in Appendix 4.2$_{p273}$ and performs a *t*-test comparing the two conditions. Figure 4.16C$_{p117}$ which uses the same R script shown in Appendix 4.2$_{p273}$, performs a non-parametric Wilcox.Test instead of a *t*-test. Figure 4.16D$_{p117}$ shows the three way comparisons of all three statistical assessments.

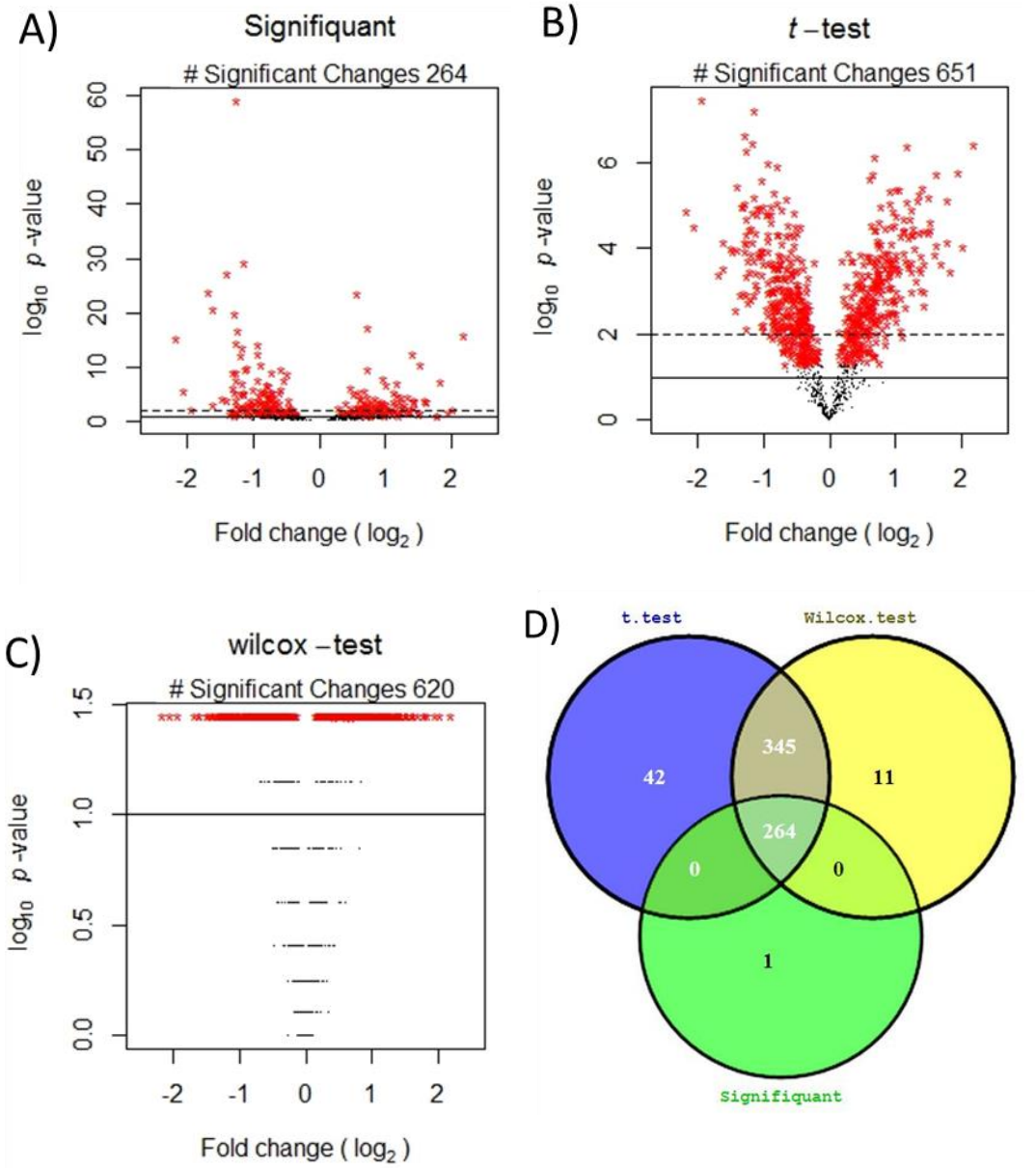


Figure 4.16: Volcano plots derived from three methods of statistical comparison: A) SignifiQuant (utilising pseudo-replicates), B) Wilcox-test, and C) t-test. D) Venn diagram of the significant proteins identified to a *p*-value of 0.05.

From Figure 4.16$_{p117}$ SignifiQuant, of the methods tested appears to provide the most significant changes with significance identified to several orders of magnitude greater than that of the *t*-test or Wilcox-test. This is derived from its use of pseudo-replicates. With the SignifiQuant method, the statistical comparison of each biological replicate is made to each replicate of the comparison condition with the least significant comparison being used to set the proteins' significance. As such, with increasing numbers of replicates the

number of significant proteins below a given threshold will decrease. In circumstances where only a limited number or even one biological replicate(s) has been used this is perhaps a reasonable compromise and as seen in the Venn-diagram predominantly identifies a subset of changes found as significant by both a *t*-test and a Wilcox-test. However, its use of pseudo-replicates means the significance reported is an assessment of the investigated sample differences and not that of the biological variable investigated.

Given the ratio compression issues which have been identified in iTRAQ investigations,[17,18] a non-parametric test such as the Wilcox test is the most appropriate of the three shown. However, only a comparison of at least 5 vs. 3 replicates can derive significance lower than 0.05. The *t*-test method doesn't rely on pseudo- replicates, nor is it restricted to comparisons of > 5 vs. 3 replicates. As such, it should be used for any comparison with less than 5 vs. 3 replicates with Wilcox being used for those with greater than 5 vs. 3 replicates.

4.10: Conclusions

Given the comparison of search engines made in this investigation, it is clear that given a *.wiff output ProteinPilot is the superior search engine of the engines tested. This observation is far less certain however given another input. For *.mgf input at the ≥2 peptide level, Phenyx identified more proteins (786) than ProteinPilot (626). Peaks identified 516, whilst PeptideShaker an amalgamation of OMSSA, X! Tandem and Mascot identified 372 proteins. Hence Phenyx would be described as the highest identifying search engine though with the optimised peak picking of ProteinPilot it can supersede Phenyx.

Upon merging, it is also noteworthy that subsequently ProteinPilot identifications reduced only 3.5%, compared to 6.6% for Peaks and 7.1% for Phenyx. This supports the concept that in addition to providing increased

numbers of identifications, ProteinPilot has increased reliability in the identifications it makes.

Whilst given for the dataset utilised here would suggest ProteinPilot as a favoured selection it should be noted that compared to the other commercial search engines it has the least versatility being unsuitable for de-Novo analysis such as Peaks or for SILAC such as Mascot.

Using a relatively simple merger script an increase in identifications (≥2 peptide) of 17.2% has been achieved with more importantly a corresponding increase in identification confidence. With known and unknown false positives being removed at during the merger, though not directly computable, the overall FDR of this list is less than 0.03 at the peptide level. This script has been compiled in such a way as to allow direct input from standard outputs of ProteinPilot, Phenyx, PeptideShaker and Peaks. Using these search engines and outputs the script can now be utilised in subsequent investigations within this thesis and beyond.

A recent publication by Sheteynberg et al.[326]published in September's issue of Molecular and Cellular Proteomics conducted an assessment of different techniques available for combining search engine results. There investigation sis not investigate the incorporation of such a process in the iTRAQ workflow. However, their observation that the combining of search engine output provided a dramatic improvement agreed with the observations of this study. Finally, they comment on the need for an easily installed, run and trubleshootable methodology. The concept of which, has been investigated here and will be discussed further in future work section.

The derived analysis pipeline has also included an assessment of the methods available to identify significant differences. These have been incorporated into an R based script that can be manipulated to conduct a comparison between any given set of iTRAQ labels. This script can now be utilised in the iTRAQ assessments conducted in this thesis and beyond.

4.11: Future work

Given the assessment conducted here, several lines of future investigation present themselves. These are discussed below.

The assessment carried out was conducted on a QStar based instrument outputting a *.wiff file. Though the assessment based only on the interchangeable input *.mgf was reported, the effect of the mass spectrometer and original format was not. Hence, a comparison to a dataset of similar size and nature but using another instrument and source format would be of interest.

Within the script as it is at present, any contradicting evidence of a PSM and protein peptide match results in its removal from the merged dataset. A future development of the process could be the retrieval of such contradicting matches where there is agreement between the majority (or some other threshold such as 66%) of search engines discarding as it were minority match.

As called for by Shteynberg et al.[326] the field is in need of an easily installed, run and trubleshootable merging method. In addition it would be advantageous for the method to be open source, facilitating wider implementation in the field. Thus a list of design criteria may be:

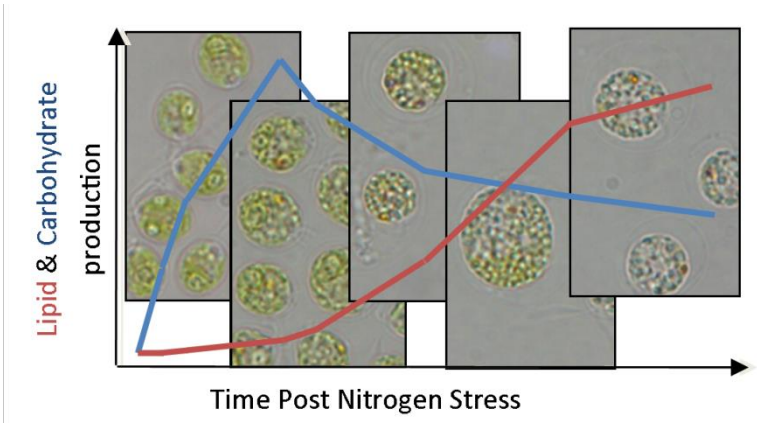
- Simplistic processing, sacrificing advance handling for robustness and easy troubleshooting.
- Simplistic installation and implementation as well as simple installation of required operating software (e.g. apache, R, Mathematica, Perl).
- Open source.
- Maximum search engine coverage.
- Future proof (often with academic software releases the release of an update to the language (i.e. R, Perl etc.) causes problems beyond the

troubleshooting ability of novice users. A solution for this could be to provide with the merging script/software fixed release versions of the languages required tested for compatibility at time of release.)

The process described for the merger here, whilst functional with limited R understanding, could be converted for use by a novice R user. This perhaps would best be achieved by development of several scripts each designed to convert a single search engine output into a plain text common format. This common format could then in effect be an interchangeable format for proteomics search engines such as the *.mgf is a plain text interchangeable format for spectra description. Any given number of these converted outputs could then be processed into a merged and conflict removed dataset as achieved in this investigation.

The major obstacle to such a process is the identification of spectra which is considerably varied from search engine to search engine and perhaps would best be solved by a Perl based script which should be run on the collections of *.mgf files to both retrieve the raw iTRAQ labels and formulate several identifiers based on spectra order as used with peaks or Tile description as used with PeptideShaker. Such tools could easily be placed online allowing increased utilisation of multiple search engines throughout the proteomics community.

Chapter 5: *Chlamydomonas reinhardtii* under nitrogen stress: A proteomic investigation



Chapter 5: *Chlamydomonas reinhardtii* Under Nitrogen Stress: A Proteomic Investigation

5.1: Abstract

Nitrogen starvation induced changes in carbohydrate and lipid content has been described in several algal species. Although these phenotypic changes are desirable, such manipulations also significantly deteriorate culture health, ultimately halting growth. To optimize biofuel production from algae, it is desirable to induce lipid accumulation without compromising cell growth and survival. In this study, we utilized an 8-plex iTRAQ-based proteomic approach to assess the model alga *Chlamydomonas reinhardtii* CCAP 11/32CW15+ under nitrogen starvation. First dimension fractionation was conducted using HILIC and SCX. 591 proteins were identified (≥2 unique peptides) of which 58 and 280 were differentially expressed at significant levels ($p < 0.05$), during nitrogen stress induced carbohydrate and lipid production, respectively. Forty seven percent more changes with significance were observed with HILIC compared to SCX. Several trends were observed including increase in energy metabolism, decrease in translation machinery, increase in cell wall production and a change of balance between photosystems I and II. These findings point to a severely compromised system, where lipid is accumulated at the expense of normal functioning of the organism, suggesting that a more informed and controlled method of lipid induction than gross nutrient manipulation would be needed for development of sustainable processes.

5.2: Introduction

The case for algae as a 3rd generation biofuel producer has been extensively outlined in the literature.[31,32,36,50] However, it has been repeatedly noted that cultivation of microalgae for high volume biofuel production is not yet

economically viable.[327] Through research and greater understanding of the organisms used, it is hoped that better classical manipulation of the cultivated organism or metabolic engineering will improve the prospects of biofuel production from such algae and make it economically viable.[93]

Manipulation of the culture medium, commonly achieved by nitrogen starvation, is used to increase the carbohydrate and lipid content in numerous algal species.[63,86] Stress induction via nitrogen starvation reduces the efficiency by which the cell converts sunlight to chemical energy. Ideally, we require a method to induce lipid accumulation without the corresponding side effects inherent in such stress induction. This would also facilitate production methodologies aimed at 'milking' algae for their lipid content, rather than collecting the whole biomass for harvest.[54,55]

For post-genomic investigations, the model alga *Chlamydomonas reinhardtii* has been the most extensively utilized species primarily due to its genome sequence becoming available in 2007.[328] Research on *C. reinhardtii* initially focused on its use for hydrogen production,[329,330] but it has recently been identified as a species that is capable of accumulating carbohydrate and lipids under nitrogen starvation.[119,331] Given that the *C. reinhardtii* genome is sequenced and responds to nitrogen starvation, it is an ideal model organism to study nitrogen stress at the proteomic level.

Although insights have been gained from investigations at the transcriptomic level,[332] the reliability of transcriptomics in accurately reflecting proteomic changes has been questioned.[129] This non-correspondence may be of more relevance during nitrogen starvation, where it can be envisioned that the degradation of proteins may be a key regulatory mechanism employed by the organism to change its proteome, which would not be revealed in a transcriptomics based approach.

Various algal proteomics reviews are available, both general,[166,333] and *C. reinhardtii* specific.[197,230] Most of the investigations conducted thus far have

been either non quantitative [201,203,208] and therefore unable to infer metabolic changes, or employed 2D gel based approaches.[190,192–194] Such studies are labour intensive[334] and typically show only a small fraction of the proteome.

Modern proteomic methodologies which label the peptides either intrinsically, such as in stable isotope labelling by amino acids (SILAC)[153] or extrinsically, such as in isobaric tag for relative and absolute quantitation (iTRAQ)[157] have greatly increased the number of proteins accessible in any comparison.[335] Though proteome coverage has considerably increased in recent years, modern proteomics studies still identify only a fraction of the total proteome with any certainty and must restrict discussions to those changes that have been observed with a significant level of confidence, and this includes any interpretations made with respect to the protein changes and their relation to pathways they may be involved in.

With the development of these high-throughput methodologies, further proteomic investigations have been conducted in *C. reinhardtii*. SILAC-based investigations have been conducted on heat stress,[185] anaerobic stress,[189] iron deficiency,[210] heterotrophic vs. autotrophic growth.[188] Other investigations include an iTRAQ study to compare a starch mutant with cell wall mutants (sta6 and cw15)[173]

A recent report by Lee *et al.* [180] utilised a label-free proteomic approach combined with metabolomics to investigate nitrogen stress. Their investigation focussed on the effect of reduced nitrogen on the proteome of *C. reinhardtii* CC125 and did not consider the nitrogen free condition with respect to energy storage.

Although often used, nitrogen induced carbohydrate and lipid accumulation is not an ubiquitous phenomenon in algae.[63,86] It is not even common in all strains of *C. reinhardtii*,[336] which as a group contains several isolates as well as laboratory generated deletion mutants. One strain that has been the subject of numerous investigations [173,337–341] is the cell wall deficient mutant

CCAP 11/32CW15+. This mutant was derived by Davies and Plaskitt,[342] and classified as a type c cell wall mutant (the mutant has a reduced cell wall without attachment to the plasma membrane). It should also be noted that the strain is derived from a nitrogen mutant with mutations in Nia 1 and Nia 2 genes, rendering it unable to utilize nitrate as a nitrogen source.[274] Due to its common use, we chose this strain for our investigation.

The strain-to-strain variations observed and the difficulty in exactly replicating growth environments for photosynthetic cultures has led to considerable lab to lab variations in the assessment of microalgae. Thus, in this study, we first conducted a detailed time course assessment of the changes in the composition of carbohydrates, lipids and pigments (hereinafter referred to as biochemical composition) of cultures under the conditions employed for proteomic analysis, followed by assessment of cultures harvested for proteomic investigation themselves.

In this investigation, nitrogen starvation was induced at a fixed time point by re-suspension of the culture at a known cell density in fresh Tris Acetate Phosphate (TAP)[343] medium that did not contain a nitrogen source (referred to as nitrogen starvation hereon). This was subsequent to 24 h culture in nitrogen replete TAP medium containing 7 mM ammonium chloride, such that the cultures were past the lag phase but prior to the stationary phase (termed 'mid-log' from hereon). During initial investigations, samples were taken for biochemical analysis to assess the profiles of carbohydrate and lipid production. The observations were used in sampling subsequent cultures both for proteomics and biochemical analysis to validate the culture condition. Protein samples were then analysed using the iTRAQ methodology to understand the changes occurring within the culture proteome.

The goal of the research reported here was to understand the relationship between nitrogen stress, and carbohydrate and lipid accumulations in *C.*

reinhardtii. Such understanding would allow superior control over induction of lipid accumulation in microalga and as such increase their productivity as a biofuel feedstock. To understand this relationship the question of how the proteome was affected during nitrogen induced accumulation was proposed.

5.3: Materials and methods

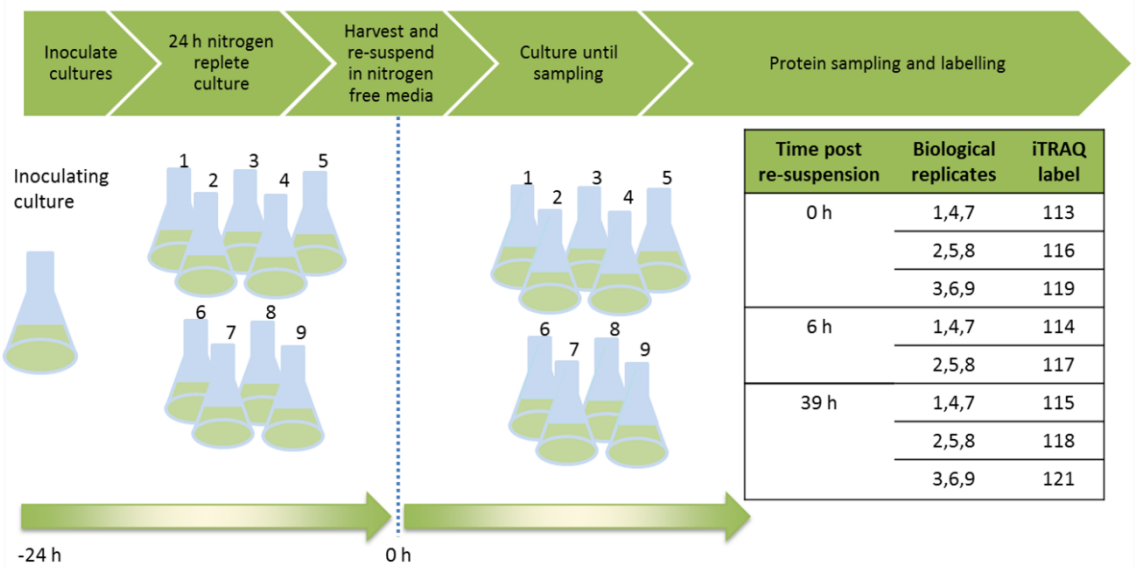
5.3.1: Organism and medium

C. reinhardtii (CCAP 11/32CW15+) was obtained from the Culture Collection of Algae and Protozoa (CCAP, Oban, UK). Cultures were grown in ether TAP medium (as described in Section 3.1$_{p66}$) or TAP medium omitting ammonium chloride for nitrogen free medium.

5.3.2: Culture

For biochemical analysis, mid log phase cultures were harvested and resuspended in nitrogen-free TAP medium to give an optical density (OD) of 0.4 at 750 nm. Cultures were maintained under illumination (350±25 µE m^{-2} s^{-1}), at 25 °C on an orbital shaker at 250 rpm. An aliquot (10 mL) of culture was sampled at intervals and freeze dried in a Modulgo freeze dryer (Edwards, Crawley, UK) for 12 h before obtaining dry cells that were estimated for dry cell weight (DCW) and subsequently used for biochemical analysis.

Cultures for proteome analysis were prepared as outlined in Figure 5.1$_{p129}$. Briefly, nine biological replicates of *C. reinhardtii* were grown in 100 mL TAP medium on an orbital shaker as detailed above. Cultures were grown for 24 h in nitrogen-replete TAP medium before being harvested by centrifugation at 3000*g* for 5 min and resuspended in nitrogen-free TAP medium to give an OD of 0.4 at 750 nm. An aliquot (5 mL) was taken for proteomics and another (10 mL) for biochemical analysis, at 0, 6 and 39 h after transfer to nitrogen free media.



Time post re-suspension	Biological replicates	iTRAQ label
0 h	1,4,7	113
	2,5,8	116
	3,6,9	119
6 h	1,4,7	114
	2,5,8	117
39 h	1,4,7	115
	2,5,8	118
	3,6,9	121

Figure 5.1: A schematic representation of the experimental design. Stress induction methodology is shown along with arrangements of biological replicates and iTRAQ labels applied.

5.3.3: Biochemical analysis

For biochemical analysis:

- Pigment assessment was conducted using a method modified from Wellburn, as described in Section 3.5_{p70}.
- Carbohydrate assessment was conducted using the Anthrone method, as described in Section 3.6_{p68}.
- Lipid assessment was conducted using Nile Red assay, as described in Section 3.7_{p69}.

5.3.4: Proteomic sampling

For protein extraction, 5 mL of algal samples were centrifuged at 3000*g* for 10 min at 4 ˚C, then resuspended in 1 mL 500 mM Triethylammonium bicarbonate buffer (TEAB) (pH 8.5) and transferred to protein low bind tube and further centrifuged at 3000*g* for 5 min finally resuspending the pellet in 200 µL 500 mM TEAB (pH 8.5). Samples were then stored at -20 °C till all harvests were completed.

5.3.5: Proteomic sample processing

Protein samples were extracted by liquid nitrogen grinding (Section 3.8$_{p70}$), before quantification by RCDC (Section 3.9$_{p70}$)). These samples were reduced, alkylated and digested (Section 3.10$_{p74}$) before assessing digestion by 1D SDS page gel as shown in Section 3.11$_{p71}$,and then labelled as described in Section 3.12$_{p72}$ with label assignment as shown in Figure 5.1$_{p129}$.

5.3.6: Off line HPLC fractionation and clean-up

Labelled proteins were fractionated using Strong cation exchange (SCX) as described in Section 3.13$_{p72}$ and Hydrophilic interaction liquid chromatography (HILIC) as described in Section 3.14$_{p73}$. Post processing SCX samples were cleaned prior to LC-MS/MS using C18 micro spin columns as described in Section 3.15$_{p73}$.

5.3.7: LC-MS/MS

Reverse phase LC-MS/MS was conducted using an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA, USA) and a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer (Applied Biosystems (now ABSciex), Framingham, MA, USA) as described in Section 3.16$_{p74}$.

5.3.8: Data analysis

For the SCX vs. HILIC assessment ProteinPilot was used to provide a list of peptide spectra match(es) (PSM). For the biological analysis, the LC-MS/MS data was processed using the optimised method, as developed and described in Chapter 4 (summarised in Section 3.17$_{p74}$). This gave an output of PSM(s) from the combine and conflict resolved outputs of 6 proteomic search engines. These were then processed using an R script developed and described in Chapter 4, to give a list of fold changes between relevant conditions each with a t-test based significance factor. EGGNOG and KEGG data processing was conducted as described in Sections 3.18$_{p75}$ and 3.20$_{p75}$.

5.4: Results and discussion

5.4.1: Biochemical characterization

Given the previously discussed variation from lab to lab in carbohydrate and lipid temporal profiles post nitrogen starvation, a biochemical assessment of the cultures in the environment proposed for protein sampling was desired. To eliminate any culture growth phase and density effect, cultures were grown for 24-48 h in normal TAP media bringing them to growth phase before harvesting and resuspending to achieve an OD of 0.475 at 750 nm. Figure $5.2A_{p133}$ and Figure $5.2B_{p133}$ shows the changes in DCW, chlorophyll, percentage carbohydrate and lipid post fixed point nitrogen starvation created by harvest and re-suspension of the culture in nitrogen-free media.

Through initially biomass (DCW) continues to increase under nitrogen stress, this increase plateaus over 48-72 h (Figure $5.2A_{p133}$). It should be noted that increase in biomass could be a result of either an increase in individual cell mass or on cell numbers. The photosynthetic pigment content can also be seen to decrease as the cells shut down growth.

It is also evident that over the first 24 h, it is mainly carbohydrates that increase as a percentage of DCW, with minimal change in lipid content. After 24 h, lipids increase and the proportion of carbohydrate in the DCW begins to fall.

The profile in Figure $5.2B_{p133}$ shows that within the given growth environment, a statistically significant ($p < 0.01$) increase in carbohydrate begins between 4 h and 8 h. Significant ($p < 0.01$) increases in lipid levels are observed between 32 h and 40 h. These points of initial increase were used to select the sampling points for proteomics.

Separate cultures were grown in identical condition to the biochemical analysis detailed above for proteomic sampling (Figure 5.1_{p129}). To ensure that the proteomic samples were representative of carbohydrate and lipid

production, samples from the proposed sampling points were also concurrently analysed for carbohydrates (Figure 5.2C_{p133}) and lipids (Figure 5.2D_{p133}). As was observed in Figure 5.2A_{p133}, a significant ($p < 0.01$) increase in carbohydrate can be seen before and after the sampling point taken at 6 h (Figure 5.2B_{p133}). Likewise, a significant increase in lipids can be seen before and after the sampling point at 39 h. These two respective sampling points were therefore considered to be representative of *C. reinhardtii* during active accumulation of carbohydrate and lipid.

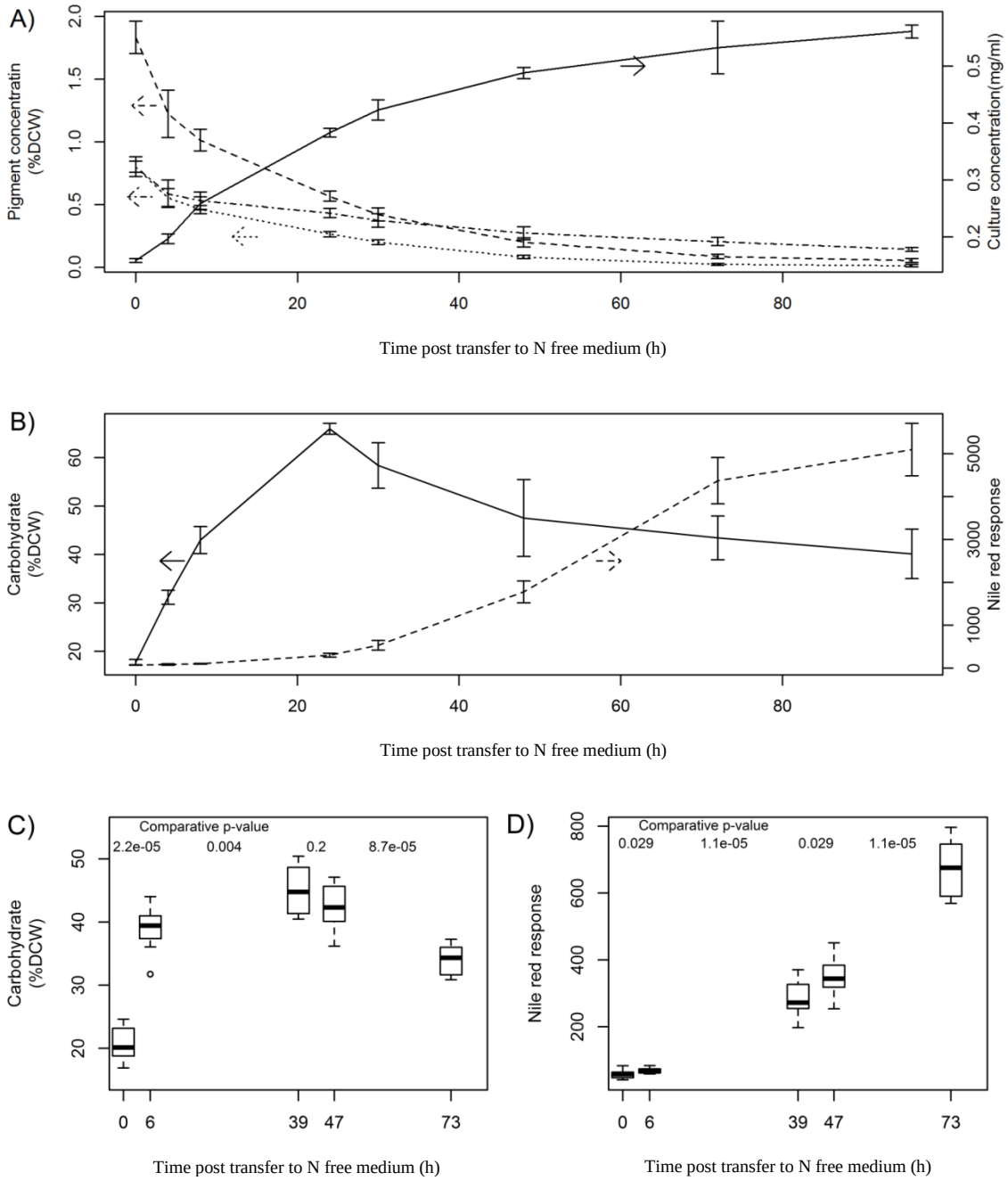


Figure 5.2: General time course of culture composition post nitrogen stress: (A) DCW and pigment changes over nitrogen starvation (DCW depicted by solid line and chlorophyll a, chlorophyll b and carotenoids shown by dashed, dotted and dotdash respectively); (B) carbohydrate and lipid changes over nitrogen stress (carbohydrate changes depicted by solid line and lipid by dashed). (C&D) Composition of cultures sampled specifically for proteomic analysis (box-whisker plots of the carbohydrate (C) and lipid (D) time courses are shown); subsequent samplings are also shown to demonstrate the continuation of the trend.

5.4.2: SCX vs. HILIC comparison

As samples from individual replicates did not have sufficient proteins for reliable proteomic analysis, pooling of samples was required to increase the quantity of protein at each sample point, while maintaining separation of the biological replicates as described in Figure 5.1_{p129}. Furthermore, samples were taken from the same cultures at different points in their progressive nitrogen starvation such that the biological replicates at each time point could be considered paired rather than independent. This pooling also aided in reducing biological variation and thus increased the statistical certainty derivable when comparing the labelled replicates for statistically significant changes.

Collected protein samples were pooled and labelled with iTRAQ reagents as shown in Figure 5.1_{p129} before being fractionated by either HILIC or SCX (25 and 31 fractions respectively). Samples were then analysed by RP LC-MS/MS. For SCX vs. HILIC comparison only identifications made with ProteinPilot were used. Mass spectrometry data was processed using the proprietary software ProteinPilot and matched against the UniProt proteome database of *C. reinhardtii* (taxon id: 3055). In total 108,910 (FDR = 0.51%) MS/MS were assigned with 19,038 (FDR = 0.02%) above confidence threshold of 95 being carried on to further analysis.

Using the peptide report from ProteinPilot (Appendix 5.1_{p275}) the matches could be split based on fractionation technique (SCX or HILIC) and thus be compiled as separate protein reports. A comparison between HILIC and SCX derived data (Appendix 5.4_{p276}) shows that HILIC achieved more identifications compared to SCX (512 vs. 401 proteins, 10,462 vs. 8,575 assigned MS/MS scans). A t-test assessing the protein changes between 0 h and 39 h revealed 280 significantly ($p < 0.05$) changed proteins using the HILIC derived data compared to 190 for SCX derived data and 311 when taking the data combined.

The use of HILIC in iTRAQ based proteomics is therefore supported not only by a 22% increase in assigned MS/MS scans and a resultant 27% increase in protein identification but also a 47% increase in significantly changed proteins as compared to SCX pre-fractionation. As already mentioned, SCX involves a desalting step that is inherent to the technique and not required for HILIC separations. The differences observed between the techniques can be attributed in part to sample loss during de-salting. However, it does not entirely explain the 47% increase in significantly changed proteins using HILIC. This we believe is due to the higher resolution of HILIC compared to SCX.

Interestingly, however, there were 21 proteins identified as significant from SCX or HILIC analysed separately, but not from the combined data set (Appendix 5.4_{p276}) suggesting a loss of p value for some changes, presumably due to increased noise.

A scatter graph of the changes observed according to each fractionation technique was plotted (Figure 5.3_{p136}). A linear model was fitted to the data that showed a significant deviation from the "y = x" line (p = 4.59 x 10-15) supporting Ow *et al.*.'s [344] earlier report that the high resolution HILIC technique reduced iTRAQ compression compared to SCX. It is thus probable that the 47% increase in significant changes observed by HILIC is not simply a result of increased identification (22% for assigned MS/MS scans and 27% for proteins) but is influenced by the reduced compression found in HILIC-iTRAQ.

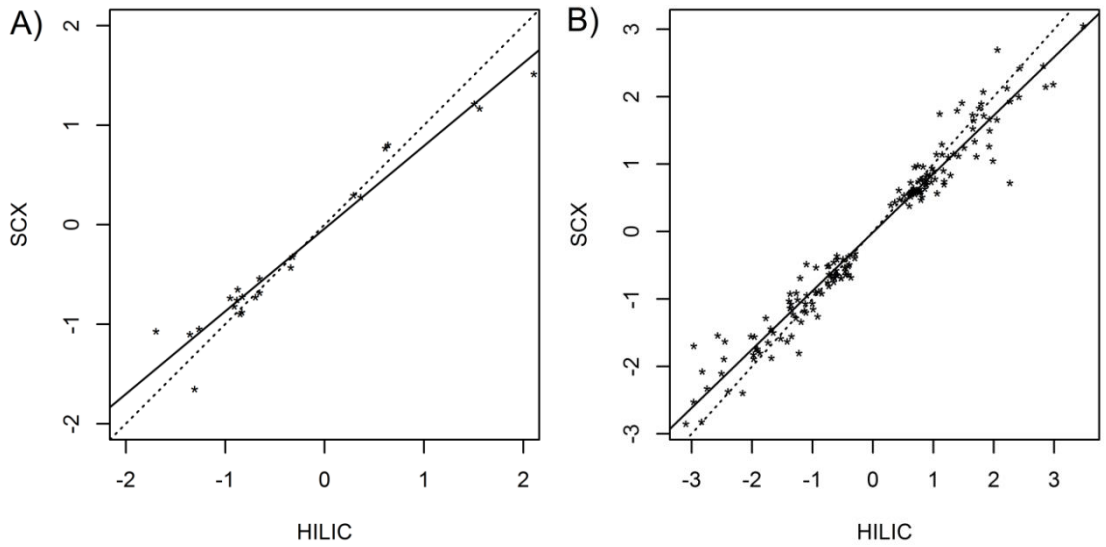


Figure 5.3: Scatter plot comparing protein changes from SCX vs. HILIC derived data for carbohydrate production (A) and lipid production (B). Proteins observed with significant change by both SCX and HILIC are plotted based on the fold change (log_2). A line of best fit is shown with significant deviation ($p = 3.91*10^{-4}$ for carbohydrate comparison and $p = 4.59*10^{-15}$ for lipid comparison) from the dotted y = x observed for both comparisons.

Combining the datasets led to minimal increase in significant changes obtained: from 70 to 71 for the carbohydrate comparison and 280 to 311 for the lipid comparison. This can be explained by the fact that gains from the use of complementary separation techniques were lost due to variation of compression. The number of significant changes may be increased if a new method for combining heterogeneous datasets were developed in the future. Although an exhaustive comparison of the two techniques was not the focus of this investigation, differences noted here suggest how the techniques fare when applied within an iTRAQ proteomics workflow.

5.4.3: Biological interpretations

For the combined analysis the merger method developed in Chapter 4 was utilised. (This analysis is divergent from the previously published paper [179] of this investigation. In that investigation only ProteinPilot was used for PSM identification rather than the combined method subsequently developed and reported in Chapter 4. Despite the change in proteomics search engine(s) utilised, many of the observations remain the same as in the published paper).

Using this combined method 22,248 PSM(s) were made without conflict amongst the search engines (Appendix 5.2$_{p275}$). This represented 1095 protein identifications 591 of which were identified by 2 or more unique peptides.

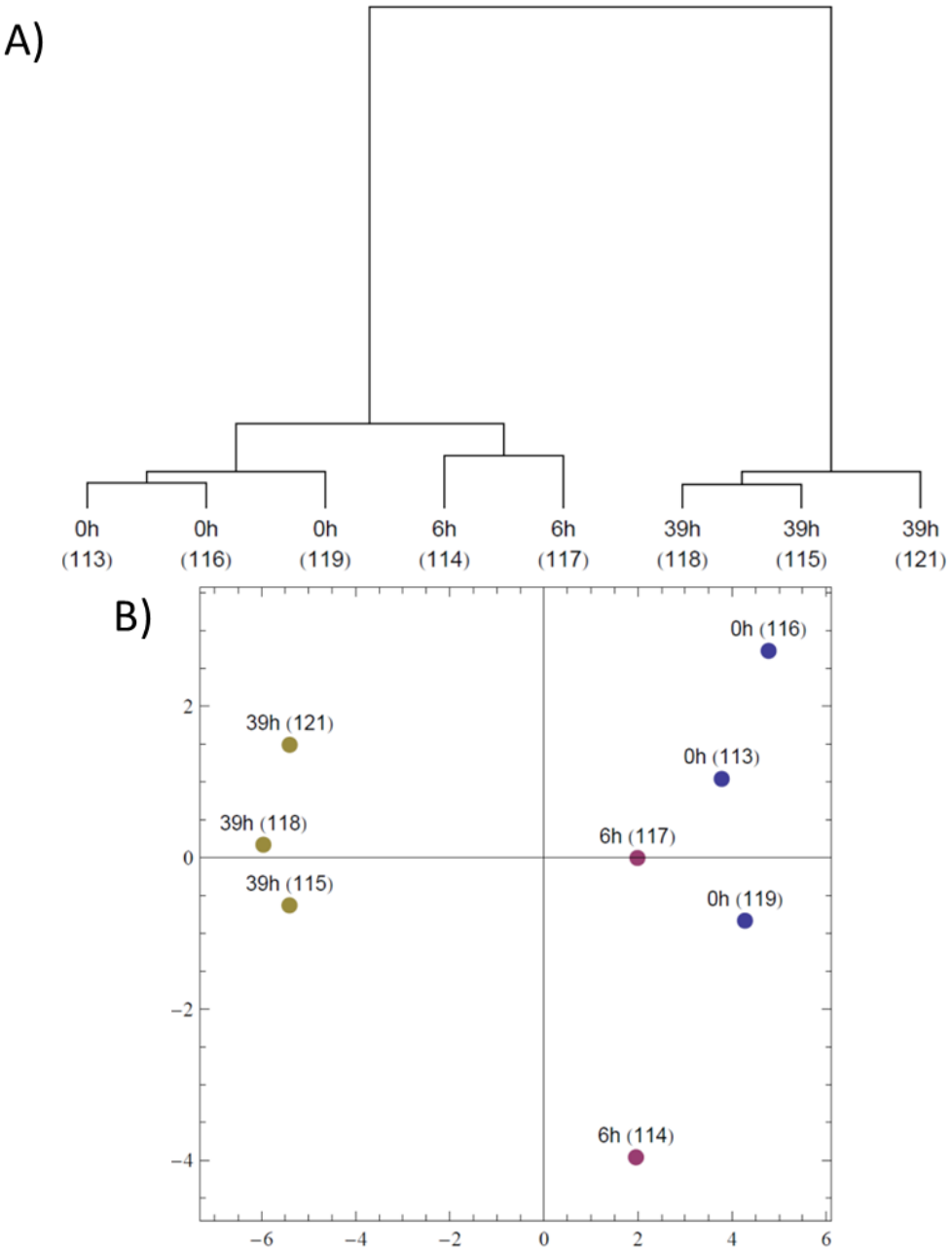


Figure 5.4: A) Hierarchal clustering using ward's linkage method of the 8 iTRAQ labels identifying correct grouping of biological replicates (113,116 and 119 derived from 0 h samplings; 114 and 117 derived from 6h samplings; 115,118 and 121 derived from 39h sampling). B) Principal component analysis showing clustering of the 8 iTRAQ labels.

Hierarchal clustering along with a principal component analysis (Figure 5.4_{p137}) shows correct clustering of the replicates with the majority of the variance between the high lipid producing cells at 39 h and the other samples.

Analysis was conducted based on direction and significance of change rather than intensity of change, since the latter was shown to be unreliable in iTRAQ-based methodologies due to the compression of quantifications.[345] For each protein and comparison, the fold change was calculated and a two sample t-test performed (Appendix 5.3_{p275}). In total, 58 proteins were identified to be significantly ($p < 0.05$) changed during carbohydrate production and 280 during lipid production.

To understand the spread of the data, volcano plots of the full set of identified proteins for both the 0 h to 6 h and 0 h to 39 h comparisons are shown in Figure 5.5_{p139}. Each point represents a protein identified by three or more peptides. The x-axis indicates the fold change observed between 0 h and either 6 h or 39 h (Figure $5.5A_{p139}$ and Figure $5.5B_{p139}$, respectively). The y-axis indicates the statistical certainty of the change observed for that given protein.

As would be expected, the volcano plots show that there is an increase in the magnitude of change observed for the 39 h comparison as against the 6 h comparison. This also leads to an increase in the statistical certainty of any change. Furthermore, while the iTRAQ methodology allows for multiple assessments at once, it is limited to eight simultaneous channels. As such in this investigation, three replicates were analysed for the 0 h (non-stressed) and 39 h (lipid producing) conditions while two replicates were used for the 6 h (carbohydrate induced) condition.

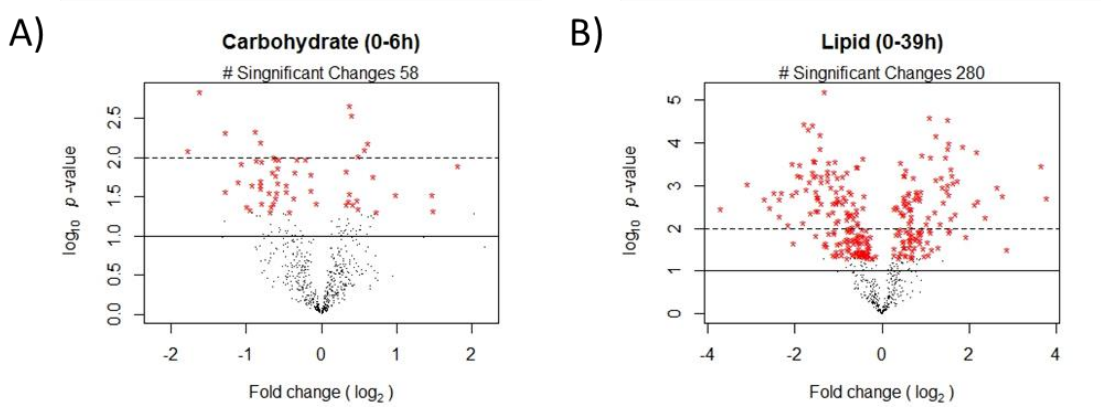


Figure 5.5: Volcano plots of overall proteomic changes between (A) non stressed and nitrogen-stressed carbohydrate producing and (B) nitrogen-stressed lipid producing conditions. Proteins with significant changes ($p < 0.05$) are shown by red star. The more relaxed cut-off of $p < 0.1$ is indicated by a solid line and the more stringent cut-off of $p < 0.01$ shown by a dotted line.

5.4.4: Analysis of cultures under carbohydrate accumulation

To derive an understanding of the changes in the proteome under carbohydrate production conditions at 6 h, those proteins identified as being significantly changed ($p < 0.05$) were collated into functional categories using the assigned name and descriptive information for the relevant Uniprot entry. Table 5.1_{p140} shows these significantly changed proteins omitting proteins that did not show any functional prediction.

Due to the nitrogen stress exerted on the cell, it is unsurprising that there were significant changes in amino acid production and decreased abundance in translation processes. The regulation of the amino acid production does follow the hydrophilic status with increased abundance of proteins producing hydrophilic amino acids and a relative decrease of hydrophobic amino acid production (Table 5.1_{p140}), though further investigation is needed to support this observation.

Within energy metabolism, there is no clear consensus, with a relative decrease of multiple proteins involved in photosynthesis or photosynthetic apparatus production, while there was increased abundance of proteins involved in glycolysis and the TCA cycle. There is also some evidence of

increased abundance of carbohydrate production and a relative decrease of fatty acid biosynthesis, which is in line with biochemical data.

Unanticipated changes were also observed with the increased abundance of proteins previously reported to be induced under low CO_2 conditions and a relative decrease of proteins involved in thiamine biosynthesis.

Table 5.1: Proteins observed to change between 0 h and 6 h nitrogen starvation at point of carbohydrate production. Changes observed with $p < 0.05$ are displayed and grouped by functional category. Each protein is reported by its UniProt ID, name, number of unique peptides, significance of change and fold change. Positive changes over time are indicated in bold.

Uniprot ID	Protein Name	# Unique Peptides	*p*.value	Fold Change
	Hydrophilic Amino Acid Synthesis			
A8J933	Acetylornithine aminotransferase	5	2.89E-02	**1.99**
A8J1S7	Bifunctional aspartate kinase/homoserine dehydrogenase	2	1.60E-02	-1.11
	Hydrophobic Amino Acid Synthesis			
A8J6Q7	3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase	3	1.92E-02	-1.75
A8JH37	Cobalamin-independent methionine synthase	28	2.43E-02	-1.75
	Other Amino Acid Synthesis			
A8IXE0	Adenosylhomocysteinase	10	3.83E-02	-1.74
	Photosynthesis			
Q5NKW4	Photosystem I reaction center subunit II, 20 kDa	12	4.45E-02	**1.40**
A8J1M9	Thylakoid lumenal 17.4 kDa protein	2	3.89E-02	**1.34**
B7U1J0	ATP synthase subunit alpha, chloroplastic	42	1.03E-02	-1.16
A8IMZ5	Magnesium chelatase subunit I	7	4.72E-03	-2.42
A8HTK7	PsbA translation factor	2	9.39E-03	**1.40**
Q9S7V1	Coproporphyrinogen III oxidase	3	2.23E-02	-1.89
A8HNE8	Geranylgeranyl reductase	7	2.74E-02	-1.62
A8J7H3	Glutamate-1-semialdehyde aminotransferase	4	4.56E-03	-1.84
A8JFB1	Porphobilinogen deaminase	18	6.25E-03	-1.75
A8JC21	Uroporphyrinogen-iii decarboxylase	5	2.67E-02	-2.42
A8J4K4	Uroporphyrinogen-iii synthase	3	2.75E-02	-1.51
	CO2 Limitation			
Q0Z9B8	LciD (Limiting CO2 inducible protein)	3	6.43E-03	**1.53**
	Energy Metabolism			
A8J5F7	6-phosphogluconate dehydrogenase, decarboxylating	2	3.95E-02	**1.26**
A8J1T4	Dihydrolipoyl dehydrogenase	9	3.51E-02	**1.29**
A8J5T0	Mitochondrial cytochrome c	6	7.80E-03	**1.49**
A8IWJ8	Succinyl-CoA ligase	9	1.48E-02	**1.26**
A8ITS8	Dihydrolipoamide succinyltransferase, oxoglutarate dehydrogenase E2 component	6	2.18E-03	**1.30**
A8HXT4	Pyruvate carboxylase	8	3.83E-02	-1.05
A8HQW1	6-phosphogluconolactonase-like protein	9	2.60E-02	-1.10
	Fatty Acid Biosynthesis			
A8JGF4	Biotin carboxylase, acetyl-CoA carboxylase	9	3.27E-02	-1.28

	component			
Q6UKY5	Acyl carrier protein	6	4.20E-02	-1.99
	Flagella			
A8I748	Flagellar associated protein	2	4.72E-02	**2.80**
A8I7W0	Flagellar associated protein	6	4.50E-02	-1.93
	Thiamine Biosynthesis			
A8J841	Hydroxymethylpyrimidine phosphate synthase	2	1.45E-03	-3.08
A8J9T5	Thiamine thiazole synthase, chloroplastic	10	2.03E-02	-2.14
	Translation			
A8J0R4	Acidic ribosomal protein P2	4	3.78E-02	-1.56
A8IJQ4	Ribosomal protein S23	3	4.92E-02	-1.35
A8HWZ8	30S ribosomal protein S15	5	1.03E-02	-1.25
E3SC50	40S ribosomal protein S9	11	2.23E-02	-1.76
A8J5Z0	Acidic ribosomal protein P0	4	9.79E-03	-1.55
A8ISZ1	Elongation factor EF-3	5	2.06E-02	-1.49
P17746	Elongation factor Tu, chloroplastic	23	3.23E-02	-1.53
A8JHM2	Eukaryotic initiation factor	2	1.51E-02	-1.57
A8IW44	Plastid ribosomal protein L19	3	2.69E-02	-1.38
A8HTY0	Plastid ribosomal protein L7/L12	7	4.84E-02	-1.60
A8JDP6	Plastid ribosomal protein S13	2	1.33E-02	-1.50
A8J8M5	Plastid ribosomal protein S5	6	4.19E-02	-1.59
A8I8A3	Plastid-specific ribosomal protein 3	8	1.67E-02	-1.53
A8I403	Ribosomal protein S19	2	1.09E-02	-1.73
A8J4I2	SR protein factor	3	2.20E-02	-1.38
	Protein Processing			
A8IDN1	Chaperonin 10	4	3.44E-02	**1.39**
	Miscellaneous			
A8JAV1	Actin	8	2.86E-02	**1.30**
A8JD06	Agmatine iminohydrolase	2	1.25E-02	**3.51**
A8I7T8	Binding protein 1	17	1.54E-02	-1.27
A2BCY1	Chloroplast nucleosome assembly protein-like	10	2.82E-03	**1.33**
A8IZS5	Glycine-rich RNA-binding protein	8	1.08E-02	-1.82
Q65Z22	NSG1 protein	3	8.15E-03	-3.42
Q6V9B1	NADH:ubiquinone oxidoreductase subunit 8	2	1.04E-02	-1.52
A8JFK4	Serine hydroxymethyltransferase 2	2	1.19E-02	-2.09

5.4.5: Analysis of cultures under lipid production

To analyse proteomic changes occurring as the alga converted from carbohydrate to lipid production, the comparison between lipid production at 39 h and carbohydrate production at 6 h was used (Appendix 5.3_{p275}). The significant protein changes ($p < 0.05$) were categorized in the same manner as the carbohydrate analysis and are reported in Table 5.2_{p142}, omitting proteins without functional classification in the UniProt database.

Continuing the trend of reduced translation observed at 6 h (carbohydrate production), there was a further reduction in multiple proteins involved in translation, at 39 h. Furthermore, during the 39 h (lipid production) as well

as reduced protein production, several enzymes involved in proteolysis such as translational inhibitor protein and cysteine endopeptidase were increased. In addition, the trends of increase in hydrophilic and decrease in hydrophobic amino acid synthesis continued. The increase in arginine synthesis may be related to an attempt to maintain levels of arginine, which has the highest content of nitrogen of all amino acids, during nitrogen stress.

Within energy metabolism, certain trends can be identified in this comparison between 39 h and 6 h, with continued decrease of photosynthetic activity and carbon fixation pathways. Yet again, there is increased abundance of many of the energy metabolism proteins involved in glycolysis and the TCA cycle. Considering the scenario with the carbohydrate production, as discussed above, there is a lack of consensus but a trend for a relative decrease of carbohydrate synthesis enzymes and increased abundance of proteins involved in carbohydrate catabolism. As before, there is only limited evidence specific to fatty acid production with up regulation of acetyl-CoA carboxylase and long chain acyl-CoA synthase.

Table 5.2: Protein changes observed in nitrogen starved cultures between 6 h and 39 h at point of lipid production. Changes observed with a $p < 0.05$ are displayed and grouped by functional category. . Each protein is reported by its UniProt ID, name, number of unique peptides, significance of change and fold change. Positive changes over time are indicated in bold.

Uniprot ID	Protein Name	# Unique Peptides	*p*.value	Fold Change
	Hydrophilic Amino Acid Synthesis			
A8J933	Acetylornithine aminotransferase	5	3.44E-02	**1.46**
P22675	Argininosuccinate lyase	15	3.23E-02	**1.61**
A8J2W0	Glutamic-gamma-semialdehyde dehydrogenase	3	1.53E-03	**2.09**
A8ILJ9	Ornithine carbamoyltransferase	5	3.06E-02	**1.67**
	Hydrophobic Amino Acid Synthesis			
A8J129	Aspartate aminotransferase	8	3.04E-02	-1.38
A8J355	Cystathionine gamma-synthase	4	4.28E-02	-1.38
	Other Amino Acid Synthesis			
A8JH72	Alanine-glyoxylate transaminase	2	6.37E-03	-1.32
A8IXE0	Adenosylhomocysteinase	10	2.85E-03	**2.20**
A8JHB4	Ferredoxin-dependent glutamate synthase	27	2.42E-02	**1.31**
A8IVM9	Glycine cleavage system, P protein	15	1.95E-02	-1.44
Q8W4V3	Serine hydroxymethyltransferase	13	1.75E-02	-1.70
	Photosynthesis			
A8IMZ5	Magnesium chelatase subunit I	7	2.27E-02	-3.52
Q93WD2	Chlorophyll a-b binding protein CP29	2	3.53E-02	-1.92

A8JGW2	Cytochrome b6f complex subunit V	6	1.59E-02	-4.70
A8HNE8	Geranylgeranyl reductase	7	5.37E-03	-2.27
A8ICF4	Heme oxygenase	2	1.39E-02	-1.56
A8HPJ2	Light-dependent protochlorophyllide reductase	2	2.27E-02	-3.58
A8JEV1	Oxygen evolving enhancer protein 3	31	7.15E-04	-2.41
A8JH68	Plastocyanin, chloroplast	6	2.17E-03	-1.89
A8JFB1	Porphobilinogen deaminase	18	1.47E-04	-1.98
A8I5N5	Tetrapyrrole-binding protein	2	6.59E-03	-1.92
A8HY43	Thylakoid lumenal protein	10	3.62E-02	-1.52
A8HNG8	Uncharacterized lumenal polypeptide	5	3.35E-02	-2.65
A8J4K4	Uroporphyrinogen-iii synthase	3	1.05E-02	-2.10
	Carbon Fixation			
P00877	Ribulose bisphosphate carboxylase large chain	46	3.80E-03	-2.24
P08475	Ribulose bisphosphate carboxylase small chain 2, chloroplastic	12	2.56E-02	-2.59
P23489	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic	4	1.43E-03	-2.03
A8JGP5	Ribulose bisphosphate carboxylase small chain	5	4.02E-02	-2.98
	CO_2 Limitation			
Q39590	Carbonic anhydrase	2	3.15E-02	-1.64
Q0ZAI6	LciB (Low-CO2-inducible protein)	13	6.17E-04	-3.22
Q0Z9B8	LciD (Limiting CO2 inducible protein)	3	6.60E-03	-2.70
Q75NZ1	Low-CO2 inducible protein (LCIC)	26	2.98E-04	-3.81
Q39589	Carbonic anhydrase	24	2.28E-02	-2.03
A8IGD9	Low-CO2-inducible protein	9	6.53E-03	-3.49
	Energy Metabolism			
A8IVG0	2-oxoglutarate dehydrogenase, E1 subunit	11	2.48E-02	**1.14**
A8J5F7	6-phosphogluconate dehydrogenase, decarboxylating	2	9.82E-03	**2.41**
A8JFR9	Acetyl CoA synthetase	12	1.24E-05	-1.54
A8J4D3	Alpha-amylase	5	3.80E-02	**1.71**
Q96550	ATP synthase subunit alpha	33	3.24E-03	**1.28**
A8J7F6	Dihydrolipoamide acetyltransferase	11	1.83E-02	**1.31**
A8JH98	Enolase	11	1.12E-02	**1.64**
A8IKQ0	Fructose-1,6-bisphosphatase	4	9.95E-06	-1.85
A8HQP1	Glucokinase	2	4.51E-02	**1.48**
A8HS14	Glucose-1-phosphate adenylyltransferase	5	5.43E-03	-1.66
A8JHR9	Glyceraldehyde 3-phosphate dehydrogenase, dominant splicing variant	31	7.22E-03	**3.13**
A8HP84	Glyceraldehyde-3-phosphate dehydrogenase	14	3.08E-02	-7.93
Q0ZAZ1	Glycolate dehydrogenase	6	1.79E-02	-1.61
A8JBU2	Isoamylase, starch debranching enzyme	2	4.73E-02	**1.73**
Q6X898	Malate synthase	23	3.46E-02	-1.62
A8J3U9	Mitochondrial ATP synthase subunit 5, OSCP subunit	11	9.13E-04	**1.34**
A8J8Z2	Phosphoglucomutase	2	3.85E-02	**1.97**
Q2VA40	Phosphorylase	15	2.33E-03	**2.39**
A8HMX2	Pyruvate-formate lyase	27	3.20E-02	**1.68**
A8IUU3	Transaldolase	3	2.90E-03	-1.30
A8I164	Vacuolar ATP synthase, subunit A	14	2.19E-02	**1.45**
A8JH65	5,10-methylenetetrahydrofolate dehydrogenase	3	3.21E-02	-1.50
P12113	ATP synthase gamma chain, chloroplastic (F-ATPase gamma subunit)	3	1.90E-02	-1.51

A8J0R7	Isocitrate dehydrogenase, NAD-dependent	9	3.29E-02	**1.23**
A8J6V1	Isocitrate dehydrogenase, NAD-dependent	3	4.75E-02	**1.70**
Q6QAY3	Mitochondrial F1F0 ATP synthase associated 19.5 kDa protein	6	2.84E-02	**1.10**
A8J4Z4	Mitochondrial F1F0 ATP synthase associated 45.5 kDa protein	30	2.55E-02	**1.30**
Q2VA41	Phosphorylase	2	4.77E-02	**1.23**
	Fatty Acid Biosynthesis			
A8J646	Acetyl-CoA carboxylase	4	3.46E-02	**1.45**
A8JH58	long-chain acyl-CoA synthetase	2	8.46E-03	**4.21**
	Transcription			
A8J4I2	SR protein factor	3	1.13E-02	**2.16**
A8J9F6	Mitochondrial transcription termination factor	3	1.41E-02	-1.23
	Translation			
A8HXM1	Mitochondrial ribosomal protein L29	4	6.43E-03	-2.27
A8JDP4	Ribosomal protein L9	3	3.76E-02	-1.60
A8J1G8	40S ribosomal protein S6	7	3.89E-02	-1.44
A8HS59	Ribosomal protein L17	7	3.14E-02	-1.37
A8J1A3	Ribosomal protein L24	2	3.44E-02	-1.59
A8J2G4	Ribosomal protein L32	3	4.27E-02	-1.26
A8J1Q3	Ribosomal protein L39	2	3.35E-02	**1.21**
A8HWZ8	30S ribosomal protein S15	5	1.12E-02	-1.62
O20032	30S ribosomal protein S18, chloroplastic	2	2.56E-02	-3.63
P11094	50S ribosomal protein L14, chloroplastic	2	1.31E-02	-1.83
A8J3Z3	50S ribosomal protein L31	2	2.18E-02	-2.92
Q8HTL1	50S ribosomal protein L5, chloroplastic	4	7.04E-03	-1.85
A8I8Z4	Plastid ribosomal protein L1	7	8.10E-03	-1.87
A8ICE4	Plastid ribosomal protein L11	5	4.36E-03	-2.00
A8I3M4	Plastid ribosomal protein L17	4	6.38E-03	-2.42
A8IW44	Plastid ribosomal protein L19	3	5.01E-03	-1.55
A8J9D9	Plastid ribosomal protein L24	3	3.86E-03	-2.86
A8J503	Plastid ribosomal protein L6	5	3.05E-03	-2.30
A8HTY0	Plastid ribosomal protein L7/L12	7	1.67E-02	-3.08
A8JDP6	Plastid ribosomal protein S13	2	2.05E-02	-1.75
A8JDN8	Plastid ribosomal protein S16	5	2.36E-02	-2.08
A8JGS2	Plastid ribosomal protein S17	5	6.52E-03	-1.88
A8J8M5	Plastid ribosomal protein S5	6	2.03E-02	-2.60
A8J8B2	Plastid-specific ribosomal protein 1	3	4.66E-02	-1.58
A8I8A3	Plastid-specific ribosomal protein 3	8	1.25E-04	-1.76
A8JF05	Ribosomal protein L28	2	6.50E-03	-1.38
A8HP55	Ribosomal protein L5	4	2.32E-02	-1.50
A8JGI9	Ribosomal protein S7	8	1.66E-03	-2.17
A8JB48	Translational inhibitor protein	2	1.67E-02	**2.34**
A8HX38	Eukaryotic translation elongation factor 1 alpha 1(Eukaryotic translation elongation factor 1 alpha 2)	23	4.92E-02	**1.54**
Q5QEB2	Elongation factor Ts	5	5.72E-03	-2.03
P17746	Elongation factor Tu, chloroplastic	23	1.58E-02	-1.56
	Protein Processing			
A8JHQ7	Small rab-related GTPase	2	2.26E-02	**2.56**
Q945T2	GrpE protein homolog	8	1.40E-02	-1.19
A8J9Q1	Importin subunit alpha	2	2.72E-02	**1.51**
A8ITH8	Chaperonin 60B2	15	1.67E-03	-1.18
A8J014	T-complex protein, zeta subunit	3	3.08E-02	**1.21**

A8IKI9	Mitochondrial processing peptidase alpha subunit	13	4.67E-02	**1.35**
A8JD56	Chloroplast trigger factor	2	7.24E-03	-2.13
A8J7T7	Cysteine endopeptidase	3	6.62E-04	**2.83**
	Prtoeolysis			
A8HSS0	Oligopeptidase A	2	2.14E-02	-1.48
	Cell Wall Structures			
A1KR25	Cell wall glycoprotein GP2	14	4.81E-03	**3.34**
A8J1N3	Cell wall protein pherophorin-C10	2	5.05E-03	**3.22**
Q3HTK4	Cell wall protein pherophorin-C3	5	9.25E-03	**2.30**
A8J094	Chitinase-related protein	2	2.09E-02	**2.76**
A8JBI9	Hydroxyproline-rich cell wall protein	2	8.33E-03	**3.02**
Q3HTK5	Pherophorin-C2 protein	6	9.69E-03	**2.55**
Q9FPQ6	Vegetative cell wall protein gp1 (Hydroxyproline-rich glycoprotein 1)	4	9.23E-04	**2.76**
	Flagella			
Q19VH4	AGG4 (Flagellar flavodoxin)	4	1.76E-02	**2.74**
Q84X82	Flagella membrane glycoprotein	10	4.44E-03	**1.59**
A8I748	Flagellar associated protein	2	5.37E-03	**2.41**
A8IAA8	Flagellar associated protein	2	8.32E-03	**5.41**
A8HW56	Flagellar associated protein	2	5.40E-03	**1.47**
A8IZV9	Flagellar associated protein	17	2.09E-03	**1.70**
	Heat Shock Protein			
A8JEU4	Heat shock protein 70A	12	3.49E-04	**1.45**
A8JES1	Mitochondrial grpE-type co-chaperone of the HSP70 system	3	4.54E-02	-1.28
	Thioredoxin Biosynthesis			
A8HNQ7	Thioredoxin reductase	2	1.29E-02	-1.49
	Miscellaneous			
A8JEV9	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	6	6.68E-03	-1.62
A8I268	ABC transporter	11	2.63E-02	-3.45
A8JAV1	Actin	8	1.37E-03	**2.22**
A8JDN2	Adenylate kinase 3	6	8.55E-03	-1.37
A8J7X9	Cytochrome c peroxidase	5	3.39E-02	**1.88**
A8JFP5	DnaJ-like zinc-finger protein	2	6.20E-03	-1.77
A8JBI5	Ferredoxin-sulfite reductase	2	2.38E-02	**1.69**
A8IZS5	Glycine-rich RNA-binding protein	8	7.70E-03	**1.33**
A8HVA3	Histone H4	13	2.88E-02	**1.80**
Q65Z22	NSG1 protein	3	2.77E-03	**8.66**
A8HPG8	Peroxiredoxin, type II	2	2.28E-02	**1.91**
Q4VKB7	Putative chloroplast 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase	4	1.82E-02	-1.75
A2PZC0	Zygote-specific Zys3 like protein	4	2.41E-02	-1.34
A8J276	Rubredoxin	2	3.42E-02	-1.25

Unlike the observations seen with the carbohydrate samples (6 h vs. 0 h comparison), there is increased abundance of cell wall-associated proteins and flagella-associated proteins in the lipid production samples (39 h vs. 6 h comparison). Further, in contrast to the changes observed between 0 and 6 h,

there is a relative decrease of low CO2 induced proteins, suggesting again that carbon fixation is decreased.

5.4.6: Non-consistent observations.

Given that the two comparisons in this investigation are of a progression post nitrogen starvation, it could be expected that the identified changes would follow a consistent trend (i.e. increased for 0 to 6 h and 6 to 39 h or down regulated for 0 to 6 h and 6 to 39 h). Within this dataset, 5 observations (Appendix 5.5_{p277}) were made with a significance of $p < 0.05$ where this did not hold true with the protein abundance switching from either increase to decreased in abundance or decreased to increased in abundance.

The increased abundance followed by a relative decrease of LciD previously linked to low CO2 stress[346,347] suggest a decrease in photosynthetic carbon dioxide fixation.

Among the proteins identified as decreased at 6 h but increased at 39 h is NSG1, which was previously identified in a transcriptomics study investigating nitrogen stress induced sexual reproduction.[348]

5.4.7: EggNOG analysis

To identify some of the broad patterns of change in the culture under nitrogen starvation induced carbohydrate and lipid production, an analysis of the functional category changes observed was conducted. The EggNOG functional classification was retrieved for each protein identified to have a significant change along with the direction of change. The comparative changes in protein assignments compared to functional class is shown for both carbohydrates and lipids in Figure 5.6_{p147}.

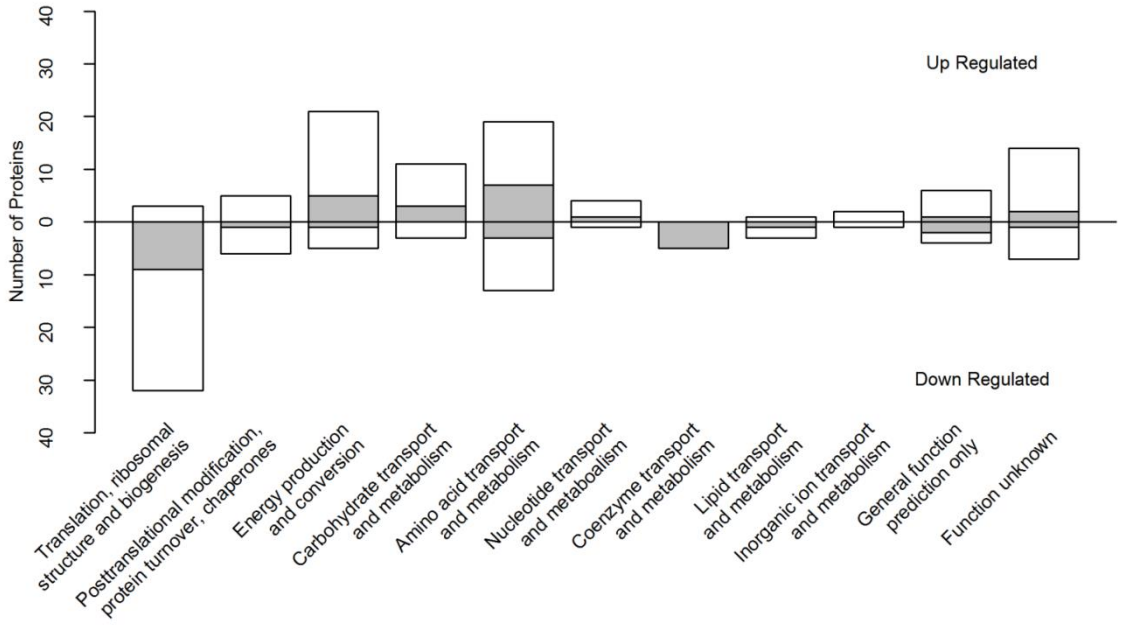


Figure 5.6: Comparison of changes observed in the EggNOG functional categories restricted to changes observed with a significance of $p < 0.05$. Grey bars indicate changes observed during carbohydrate production (6 h vs. 0 h comparison) overlaid on white bars indicating lipid production (39 h vs. 0 h).

Within the carbohydrate classification, a decrease in abundance of proteins involved in translation is shown, predominantly due to reduction in ribosomal proteins. Similarly, decreases in proteins involved in coenzyme and lipid transport/metabolism were also observed. Increases in energy production, carbohydrate transport/metabolism and amino acid transport/metabolism were observed.

Observations of cultures during lipid production were similar to that during carbohydrate, but to a greater degree multiple proteins involved in translation are decreased. The observed increase in energy production and carbohydrate synthesis was also seen to be continuing to the lipid producing cultures.

5.4.8: Pathway analysis

In order to further understand the proteomic changes, proteins observed to be significantly ($p < 0.05$) changed were mapped onto the KEGG database,[349,350] showing either up- or a relative decrease. These mappings revealed several pathways with trends corroborating the observations

discussed earlier, with much of the glycolysis and the TCA cycle (Appendix 5.7_{p279}) being identified as increased and multiple decreased proteins being mapped to the ribosomal machinery (Appendix 5.6_{p278}). In addition some suggestion of increased phagocytosis was seen with up regulation in the Phagosome (Appendix 5.8_{p281}).

Highlighting the changes within the photosynthetic apparatus between 0 h and 39 h (Figure 5.7_{p148}), reveals a relative decrease of photosystem II coupled with increased abundance of photosystem I. This preference of photosystem I over photosystem II is suggestive of cyclic photophosphorylation.

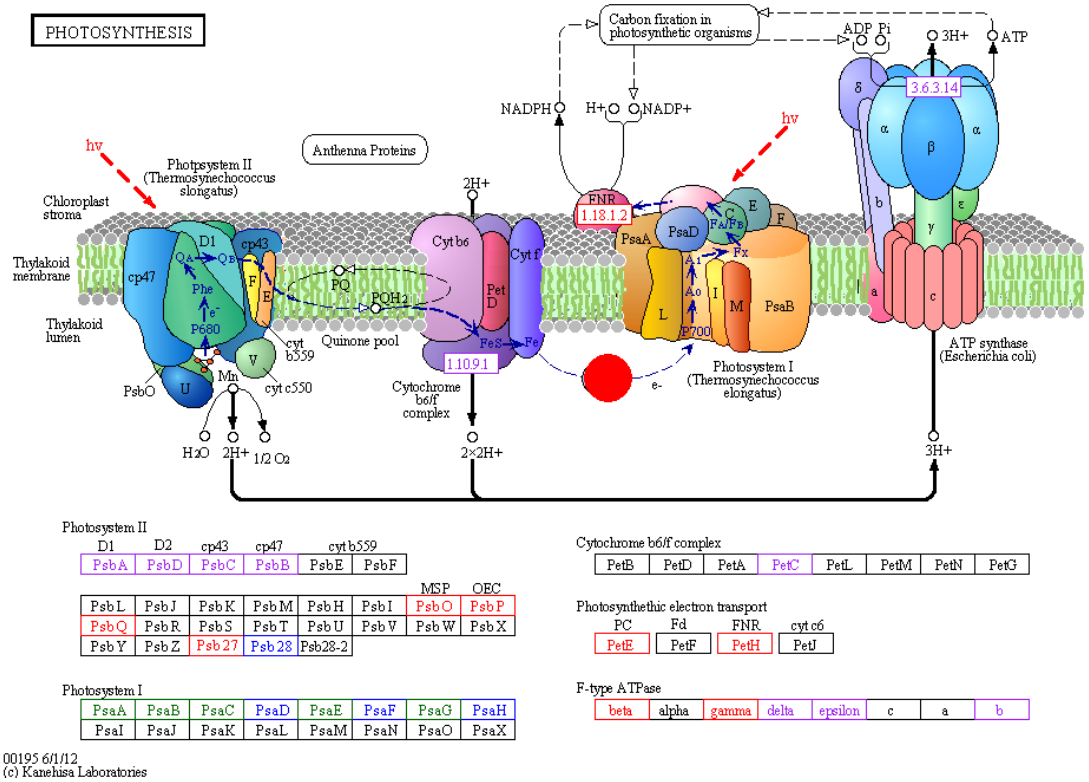


Figure 5.7: KEGG map of the photosynthetic apparatus with identified protein changes (0 h vs. 39 h) coloured based on observed regulation. Increased abundance in dark green and blue outline ($p > 0.05$ and $p < 0.05$ respectively). Decrease abundance in purple and red outline ($p > 0.05$ and $p < 0.05$ respectively).

When observing lipid biosynthesis (Appendix 5.9_{p282}) there is little evidence of increased production with some up and some down regulation observed. However, in lipid metabolism considerable down regulation was observed.

This could suggest that the increase in lipid observed is a result of decreased breakdown rather than increased production.

Although *C. reinhardtii* was sequenced in 2007 and is the most extensively characterized algal species, much of *C. reinhardtii*'s proteome is only annotated by automated homology algorithms, leaving numerous gaps and undescribed predicted proteins. There are various significantly changed proteins identified in this study which may warrant further investigation to deduce their role, including proteins which showed non-continuous changes when comparing the carbohydrate producing against the lipid producing proteome.

5.4.9: Energy pathways

This study provides evidence that, despite a 66% decrease in total pigment content and 63% decrease in growth, the cells increased activity in their energy metabolism pathways with significant coverage of the proteins involved in glycolysis and the TCA cycle being increased. This coupled with the biochemical analysis showing an increase in energy storage compound (carbohydrates or lipids) production, suggesting the cells actively increased energy processing and storage mechanisms rather than shutting down completely.

The relative increase and decrease in abundance of "low CO_2 inducible proteins" between 6 h and 39 h could be suggestive of a switch from photosynthetic growth to heterotrophic growth both on acetate available in the media and previously generated carbohydrate compounds, which appear to be degraded at the latter stage of nitrogen starvation.

The favouring of heterotrophic growth over photoautotrophic growth, post-nitrogen starvation, may explain why other groups have found difficulty in replicating the nitrogen effect in autotrophic conditions. When researching the nitrogen effect, it should be considered that while *C. reinhardtii* is photosynthetically active in the presence of acetate,[274] cultures will

increasingly rely on heterotrophic growth during nitrogen starvation. This leads to a suggestion that, as nitrogen starvation continues, the carbon availability to the algae is fixed by the acetate concentration per algal cell at the time of stress induction. This is unlike non-stressed cultures that can fix carbon dioxide. With such fixed carbon availability, the amount that can be converted to first carbohydrates and subsequently lipids is fixed, based on the acetate availability per cell. Thus, culture density will affect the carbohydrate and lipid accumulation achieved.

5.4.10: Cyclic photophosphorylation

One unexpected observation of this study was the decrease of photosystem II co-current with increased abundance of photosystem I. This reliance on photosystem I suggests that the cells shifted from regular photosynthesis to cyclic photophosphorylation thereby allowing the continuation of light harvesting in the absence of NADH. This shift in photosynthetic activity is supported by the reduction in chlorophyll a and b observed under nitrogen starvation (Figure 5.2A$_{p133}$). While maintaining limited photosynthetic activity, this method of photosynthesis is energetically less favourable, as it might eventually limit the production of energy storage compounds, such as carbohydrates and lipids. Cyclic phosphorylation is unlikely to be a prerequisite for accumulating carbohydrates and lipids and as such could be an undesirable side effect of nitrogen starvation that it would be advantageous to remove.

5.4.11: Dormancy

While the culture was actively producing lipid beyond 39 h, from this study it can be postulated that the cultures were entering into a type of dormancy. This is evidenced by the relative decrease of proteins involved in translation (Table 5.1$_{p140}$ and Table 5.2$_{p142}$) and a general decrease of the enzymes involved in the various amino acid synthesis pathways. However, these may also be indicative of rationing limited nitrogen sources. The concept of dormancy is further supported by changes in the NSG1 associated protein, which is

involved in sexual reproduction, and by the increased abundance of eight proteins involved in production of the cell wall formation. This build-up of the cell wall is possibly related to the zygospore created by *C. reinhardtii* during sexual reproduction that can be triggered within heterosexual cultures by nitrogen starvation.

5.5: Conclusions

This investigation reveals insights into nitrogen starvation of *C. reinhardtii*, including changes in photosystem stoichiometry, modifications in cell wall and possible shift in balance between autotrophic and heterotrophic growth in this mixotrophic organism. Further investigation can now establish whether such proteomic changes as reported here are unique to *C. reinhardtii* or unique to nitrogen starvation.

The comparison of SCX and HILIC conducted during this investigation supports previous findings that HILIC provides increased identification with less compression. This study demonstrates that such improvement translates into increased identification of significant changes, on which biological interpretations can be made. As such, utilisation of HILIC provides superior data compared to SCX without significantly increased workload.

This study highlights the global effect that the media manipulation has on *C. reinhardtii.* Such gross manipulations may have the desired effect of increased lipid accumulations but also have several undesirable side effects such as reduced chlorophyll content, altered photosynthesis and reduced energy production. A greater understanding of how the organism is affected as a result of gross manipulations will be needed before such methods are employed for increased lipids from microalgae. As such, this study provides further support that a more informed and focussed method of lipid induction, would be preferable to the gross manipulation by nutrient

limitation observed with nitrogen depletion. Methods involving metabolic engineering will provide the opportunity for such focussed manipulations.

5.6: Further work

Given the stated goal of this thesis to investigate microalgae and their potential for biofuel production, it is sensible that *C. reinhardtii* as the predominant model alga be selected for initial investigations. Given its dramatic effect of increasing lipid content it is also sensible that nitrogen stress be the initial comparative condition. Within this investigation several changes were observed. Initial further work would likely involve detailed analysis of these changes. This could be done with more targeted methods of analysis such as SRM or immune blotting. This would allow both a confirmation by a second method of the change and a more detailed understanding of the change temporally.

Depending on this further assessment these changes would be intriguing candidates for knock out or down investigations, perhaps using RNA silencing. Further the approach of reducing lipid catabolism as a means to induce lipid storage is of considerable interest.

Within this investigation a substantial re-organisation of the proteome was observed. This complicates identification of which changes are those involved with lipid accumulation and which involved with the other aspects of nitrogen stress. Through further analysis of multiple proteomic investigations commonalities and differences may aid in such identification. Regarding further proteomic based investigations three tacks are evident.

First there is scope for further investigations looking into the combination of *C. reinhardtii* and nitrogen stress. This investigation utilised the mixotrophic nature of *C. reinhardtii* to increase the speed of the lipid accumulation. It is however more likely that any productive strain of microalgae for energy with be grown photoautotrophically. Thus a similar

investigation omitting acetate from the culture media could be envisaged. In addition our study focused on a general approach probing the soluble faction. This could be complimented with other general investigations such as the insoluble proteome or cellular sub fractionation experiments such as those targeted to the chloroplast, nucleus, or endoplasmic reticulum. This would increase the coverage of the active proteome from that obtained in this investigation.

Second, in addition to nitrogen, several other culture perturbations have been noted to act as lipid triggers. Through comparison of the changes from other triggers, the changes which are specific to nitrogen and those more general to stress and lipid induction may be identified.

Third, whilst *C. reinhardtii* has been the predominant model alga, microalgae cover a wide selection of organisms including members of the multiple phyla and kingdoms. An investigation of nitrogen stress in a different alga species would allow comparisons to this investigation revealing common traits of nitrogen stressed microalgae.

Chapter 6: Comparison of mixotrophic and photoautotrophic growth of *Chlamydomonas reinhardtii* grown under turbidostatic conditions: A Proteomic Investigation

Chapter 6: Comparison of mixotrophic and photoautotrophic growth of *Chlamydomonas reinhardtii* grown under turbidostatic conditions: A Proteomic Investigation

6.1: Abstract

In laboratory based investigations of *Chlamydomonas reinhardtii*, growth has generally been conducted in mixotrophic conditions. This is generally due to a desire to increase specific growth rates. However, in an ideal biofuel production situation, photoautotrophic growth is desired. In this investigation, we aimed to try and understand the changes occurring within the organism when grown under different trophic conditions.

Using turbidostatic cultures in stirred tank photobioreactors, *C. reinhardtii* was grown photoautotrophically and mixotrophically, each with and without CO_2 supplementation. Carbohydrate, lipid and pigment content were assessed for each culture condition and an 8-plex iTRAQ-based proteomic assessment was carried out. 995 Proteins were confidently (≥2 unique peptides) identified with the significantly changed proteins reported for each comparison.

From the biochemical observation we observed that the source of the carbon has a greater impact on the biochemical makeup of the cell rather than the degree of carbon availability. Conversely however it was the degree of carbon availability which influenced the proteome to the greatest extent.

It is therefore suggested that whilst proteomic findings in one carbon source can be used to predict those in another. The same is not true for biochemical composition which varies dependent on source of carbon. This investigation suggests that the increased carbon availability of mixotrophic growth as

compared to un-supplemented photoautotrophic growth, significantly impacts on the cell at the proteomic level. The carbon availability must therefore be taken into consideration when relating mixotrophic investigations to photoautotrophic cultivations.

6.2: Introduction

The use of carbon based compounds for the storage and interchange of energy is common to almost all living organisms. These carbon molecules in the form of carbohydrates and lipids form the majority of dry biomass. These compounds are also the precursors of most biofuels. As such a key factor of any biofuel feedstock organism is the method by which it derives its' carbon compounds.

In any organism, the derivation of any carbon compounds requires both a source of carbon molecules themselves and energy. The means by which an organism derives these two components, carbon and energy, is used to describe its trophism.

The carbon source can be from the fixation of inorganic carbon, such as CO_2, termed autotrophy or from the consumption of organic carbon, termed heterotrophy. The energy source can come from photosynthesis of solar radiation called phototrophy or from the breakdown of organic or inorganic compounds termed chemotrophy[351].

Given these two levels, the trophism of an organism's growth can be defined as photoautotrophic, chemoautotrophic, photoheterotrophic or chemoheterotrophic. In the literature, however, the more common combinations of chemoheterotrophic and photoautotrophic are shortened to heterotrophic and to autotrophic, with the energy source aspect taken as assumed. In addition, the term mixotrophic is used to refer to organisms which can conduct multiple of these four growth options.[352] Most commonly, mixotrophic refers to the ability to grow both photoautotrophically and

chemoheterotrophically. Mixotrophic organisms can either be considered facultative, preferring one growth condition over others or may use both conditions simultaneously.

The model alga *Chlamydomonas reinhardtii* has been demonstrated as growing chemoheterotrophically (heterotrophically) by growth in the dark with acetate, photoautotrophically (autotrophically) by growth without acetate in the light and mixotrophically by growth with acetate and light[136]. In addition, it remains one of the rare examples of a model organism identified as synthesising a full photosynthetic apparatus under mixotrophic growth with acetate[188].

Given growth in mixotrophic conditions, the degree to which any organism is facultative photoautotrophic, photoheterotrophic, chemoautotrophic or chemoheterotrophic is often not reported. The study by Heifetz *et al.*,[353] reporting ^{13}C incorporation, identified that *C. reinhardtii* grew as a true mixotroph with >50% of incorporated carbon from photosynthetic processes whilst gown in acetate containing TAP media. This percentage was dependent on light intensity. Whilst this describes the carbon sourcing, the balance for energy between phototrophic and chemotrophic is not described.

Due to its use a model algae the need for a greater understanding of *C. reinhardtii* is needed. A core understanding that is currently lacking in the field is the true effect of carbon source and availability.

Proteomics being the functional omic level can be argued to provide the greatest insight into the cells active response to perturbations. It would therefore be beneficial for the field to have an overview of the proteomic effect of carbon source and availability in the model alga *C. reinhardtii.*

To our knowledge, two previous publications have reported proteomic changes in *C. reinhardtii* under variations of trophism. The first by, Baba *et al.,*[187] investigated the effect of high versus low CO_2 in photoautotrophic

cultures. Within this investigation they assessed both intracellular and extracellular proteins. For the extracellular component, 22 proteins were identified as significantly regulated by means of 2D gel based proteomics. For the intracellular they observed no detectable changes in the proteome. The second by, Weinkoop *et al.,*[188] investigated autotrophic vs. mixotrophic cultivation in batch cultivations through a targeted proteomics focusing on low CO_2 induced proteins.

A serious limiter of these investigation was however the coupling of carbon availability and carbon source. These are two different factors and the comparison of four possible combinations in a single investigation is needed.

Given the desired comparison of carbon source and availability it is evident that the growth rate between the four conditions will vary dramatically. Given this the use of a chemostat, where supply of nutrients is fixed would lead to varying culture densities. Such changes in culture density would likely have a severe impact on the culture. Hence, to investigate the carbon source and availability effect in isolation to culture density effects, a turbidostatic culture was used. In such a culture the turbidity (a proxy for culture density) is maintained as fixed rather than the feed rate as with a traditional chemostat. In effect, rather than growing each culture chemostatically with a fixed feed rate, a feed rate specific to the growth rate of each culture is selected.In this investigation *C. reinhardtii* was assessed under carbon starved photoautotrophic culture against non-starved cultures whose carbon source was either supplemented CO_2 , acetate or both. We utilised turbidostatic culture to control the carbon availability at a per cell basis (OD being a proxy for cell number). The changes were assessed both biochemically and proteomically by means of an 8-plex iTRAQ investigation.

The goal of this investigation was understand the way carbon source affected *C. reinhardtii* culture. In particular how laboratory scale investigations can be related to productive systems given a change in carbon source between

the two. This relation is of critical importance when relating laboratory investigations conducted using acetate to accelerate growth to that of production systems which are envisaged to rely on atmospheric or CO_2 supplemented growth. To advance this goal the proteomic changes under four carbon regimes were investigated.

6.3: Materials and methods

6.3.1: Organism and medium

C. reinhardtii (CCAP 11/32CW15+) was obtained from the Culture Collection of Algae and Protozoa (CCAP, Oban, UK). Cultures were grown in either TAP media (as described in Section 3.1$_{p66}$ or TAP media omitting Acetate and pH adjusted to 7.0 using HCl.

6.3.2: Bioreactor culture

The bioreactor was set up as described in Figure 6.1$_{p161}$. A Sartorius Biostat B plus was utilised for culturing. The reactor was mixed by two 6 bladed Ruston impellers connected to an axel rotating at 100 rpm. The reactor had 'on-line' pH measurements controlling flow of 1 N HCL and 1N NaOH into the reactor to maintain a constant pH of 7. A thermometer with 'on-line' measurement was used to control the temperature of an external water jacket for the reactor maintaining a constant culture temperature of 25^0C. Aeration was provided by sparging at the bottom of the impeller at 1 L min^{-1} with either atmospheric air containing 0.04% CO_2 or with atmospheric air supplemented with a percentage of CO_2 (v/v). The reactor was surrounded with a light shroud containing six 7 W fluorescent bulbs. These provided 550 ± 50 µE $m^{-2}s^{-1}$ of irradiance measured at the centre of the reactor when filled with TAP media.

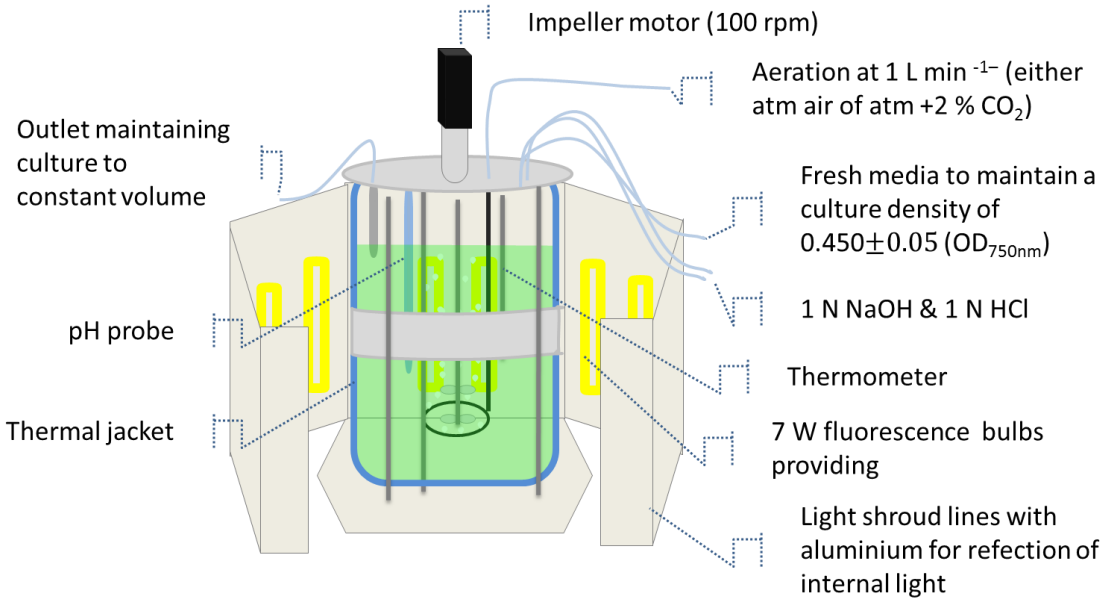


Figure 6.1: Graphical depiction of the bioreactor growth environment detailing parameter setting.

6.3.3: Turbidostatic culture

For turbidostatic culture, the bioreactor conditions as described for batch culture were replicated with the addition of fresh appropriate media (TAP or Acetate free TAP) at such a rate that the dilution rate equalled the specific growth rate. This was achieved with 'at-line' optical density measurements (750 nm) and manual control of the fresh media flow rate. Media removal rate was matched to addition by maintaining a constant culture volume as shown in Figure 6.1A$_{p161}$.

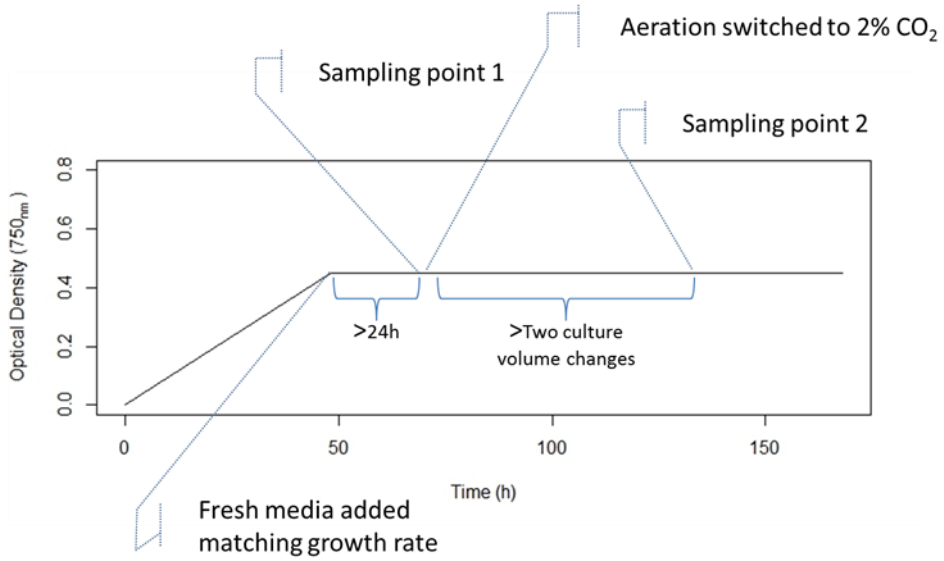
Given the 'at-line' nature of the turbidostatic control, a degree of variation was observed in the culture density. Cultures were however maintained at OD of 0.450 ±0.05 at 750nm for at least 12 h prior to any sampling event.

The profile and sample arrangement for the turbidostatic sampling are shown in Figure 6.2$_{p162}$. The reactor was prepared and seeded and allowed to grow to an OD$_{750nm}$ of 0.450 with the relevant medium for the condition and 0.04% CO_2. Fresh medium was then added to match the growth rate giving a turbidostatic environment. This was maintained for at least 24 h prior to sampling both for biochemical and proteomic analysis.

Once samples for the 0.04% CO_2 condition the sparging gas was increased to 2% CO_2 increasing the flow of fresh media to match the growth rate. At least 2 culture volumes were then exchanged under the new condition before taking a second sampling for biochemical and proteomic analysis under the CO_2 supplemented condition.

Each biological replicate and media condition was derived from a fresh seeding of the bioreactor as seen in Figure 6.2$_{p162}$.

A)



B)

Culture Media		Sampling point 1 0.0387% CO_2	Sampling point 2 2% CO_2
		Condition A	Condition B
TAP media With Acetate	Culture seeding 1	Biological replicate 1	Biological replicate 1
	Culture seeding 2	Biological replicate 2	Biological replicate 2
	Culture seeding 3	Biological replicate 3	Biological replicate 3
		Condition C	Condition D
TAP media Without Acetate	Culture seeding 4	Biological replicate 1	Biological replicate 1
	Culture seeding 5	Biological replicate 2	Biological replicate 2
	Culture seeding 6	Biological replicate 3	Biological replicate 3

Figure 6.2: A) Graphical representation of sampling procedure for each culture seeding as listed in part B. B) Table showing traits of each culture seeding (1-6) and the appropriate condition (A through D) used to refer to each.

6.3.4: Biochemical sampling

For biochemical sampling, three 15 mL algal samples were centrifuged at 3000*g* for 10 min at 4 ˚C, then resuspended in 1 mL dH_2O in a 1.5 mL Eppendorf tube. This was centrifuged at 3000*g* for 5 min before storing at -80 °C prior to freeze drying (Section 3.4_{p68}) to obtain a dry cell weight for biochemical analysis.

6.3.5: Proteomic sampling

For protein extraction, two 50 mL algal samples were centrifuged at 3000*g* for 10 min at 4 ˚C, then resuspended in 1 mL 500 mM Triethylammonium bicarbonate buffer (TEAB) (pH 8.5) and transferred to protein low bind tube and further centrifuged at 3000*g* for 5 min finally resuspending the pellet in 200 *µ*L 500 mM TEAB (pH 8.5). Samples were then stored at -80 °C until all harvests were completed.

6.3.6: Biochemical analysis

For biochemical analysis, the following were carried out, each condition being composed of three biological replicates each of which consisted of three sampling replicates:

- Pigment assessment was conducted using a method modified from Wellburn, as described in Section 3.5_{p68}.
- Carbohydrate assessment was conducted using the Anthrone method, as described in Section $3.6_{p68.}$
- Lipid assessment was conducted using Nile Red assay, as described in Section 3.7_{p69}.

6.3.7: Proteomic sample processing

Protein samples were extracted by liquid nitrogen grinding (Section 3.8_{p70}), before quantification by RCDC (Section 3.9_{p70}). These were reduced, alkylated and digested (Section 3.10_{p71}) before assessing digestion by 1D SDS page gel as shown in Section 3.11_{p71}. These were then labelled as described in Section 3.12_{p72} with label assignment as shown in Table 6.1_{p170}.

6.3.8: Off line HPLC fractionation and clean-up

Hydrophilic interaction liquid chromatography (HILIC) was conducted as described in Section 3.14$_{p73}$. Post processing HILIC samples were cleaned prior to LC-MS/MS using C18 micro spin columns as described in Section 3.15$_{p73}$.

6.3.9: RPLC-MS analysis

Reverse phase LC-MS/MS was conducted using an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA, USA) and a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer (Applied Biosystems (now ABSciex), Framingham, MA, USA) as described in Section 3.16$_{p74}$.

6.3.10: Data analysis

The LC-MS/MS data was processed using the optimised method, as developed and described in Chapter 4. This gave an output of PSM(s) from the combined and conflict-resolved outputs of 6 proteomic search engines. These were then processed using the script developed and described in Chapter 4 (summarised in Section 3.17$_{p74}$) to give a list of fold changes between relevant conditions each with a *t*-test based significance factor. Gene ontology and KEGG data processing was conducted as described in Section 3.19$_{p75}$ and 3.20$_{p75}$.

6.4: Results and discussion

The use of a bioreactor environment was chosen to investigate the effect of carbon source on *C. reinhardtii* due to its increased level of control over the culture environment. Whilst limiting the ability to conduct a large number of simultaneous cultures, bioreactors can provide a superior level of control to the culture conditions such as mixing, pH regulation and gas regime. This was particularly important given the carbon dioxide supplementation aspect of this investigation.

6.4.1: Batch growth curves

To provide initial assessment of the growth conditions, several batch growth curves of *C. reinhardtii* were conducted (Figure 6.3$_{p166}$). These conditions included growth without acetate with either 0.04% CO_2, 2% CO_2 or 5% CO_2 and growth with acetate under 0.04% CO_2 and 2% CO_2.

As may be expected, photoautotrophic growth using atmospheric CO_2 without acetate was the slowest (Figure 6.3$_{p166}$) and showed a more linear profile than in the other conditions. Growth rate under 2% and 5% CO_2 was higher than that under air (0.04%). Further, given the flow rate of 1 vvm the profile of acetate-free culture supplemented with 2% CO_2 was comparable to that of 5% CO_2 supplementation. This comparability suggests that supplementation beyond 2% CO_2 is non-beneficial either due to reaching a maximal of dissolution or maximal organismal uptake. As such, further supplementation was not beneficial (2% supplementation was the minimal operable range of Sartorius system given the mixture of pure CO_2 with atmosphere at 1L min^{-1} s^{-1}).

Mixotrophic growth using acetate described a more typical exponential growth curve, as may be expected from a heterotrophic growth profile. The combination of acetate and 2% CO_2 shows a growth profile similar to that of the acetate only profile suggesting that the level of acetate in the media was not the limiting factor in growth rate. The degree to which the mixotrophic condition (whether CO_2 supplemented or not) were autotrophic or heterotrophic was not monitored. However, based on Heifetz *et al.*[353]an assumption of true mixotrophic culture rather that facultative autotrophic or heterotrophic culture is assumed.

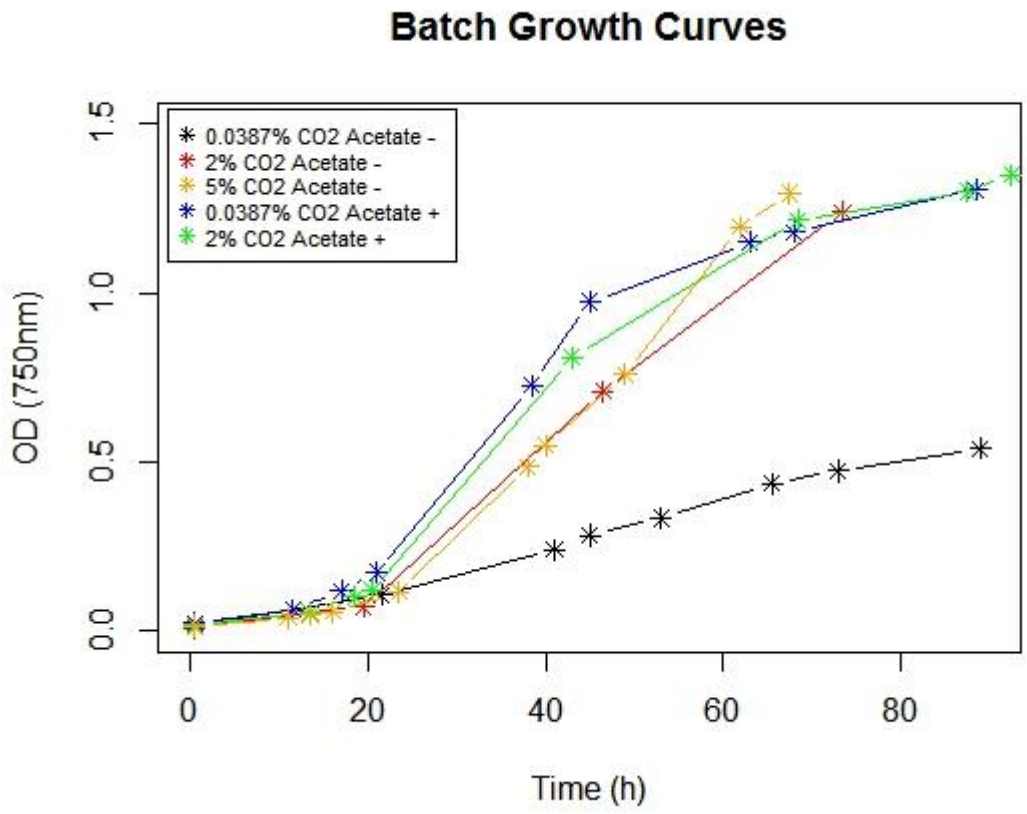


Figure 6.3: Batch growth curves of *C. reinhardtii* in a bioreactor environment. Optical density of the culture over time for each batch run is shown.

6.4.2: Turbidostatic culture

Given the initial batch culture investigation above, four conditions were chosen for further investigation. These are:

- Un-supplemented photoautotrophic, involving growth without acetate and atmospheric air termed "Photoautotrophic 0.04% CO_2".
- Supplemented photoautotrophic, involving growth without acetate and atmospheric air supplemented with 2% CO_2 termed "Photoautotrophic 2% CO_2."
- Un-supplemented mixotrophic, involving growth with acetate and atmospheric air termed "Mixotrophic 0.04% CO_2."
- Supplemented mixotrophic, involving growth with acetate and atmospheric air supplemented with 2% CO_2 termed "Mixotrophic 2% CO_2."

These conditions were also termed condition A, B, C and D (Table 6.1_{p170}) for identification by shorthand notation.

With any batch growth investigation, the point of sampling can have a considerable impact on the culture, due to current culture density and nutrient depletion profile. This is particularly evident in an investigation involving carbon sources where factors such as culture density and time since carbon compound provisions will affect the culture greatly. Further as the cell density would presumable significantly impact the organism this should ideally also be controlled. For these reasons, this investigation was conducted in a turbidostatic fashion.

6.4.3: Carbohydrate, lipid and chlorophyll assessments

The carbohydrate assessment Figure $6.4A_{p170}$ showed a significant (p = 0.007) difference between the conditions based on the presence of CO_2 supplementation. With samples grown under 2% CO_2 containing an average of 14% carbohydrate (% DCW) compared to 18% when grown under normal atmospheric concentrations of CO_2. Though the degree of noise in samples varied, no significant changes in carbohydrate concentration were observed comparing the presence or absence of acetate.

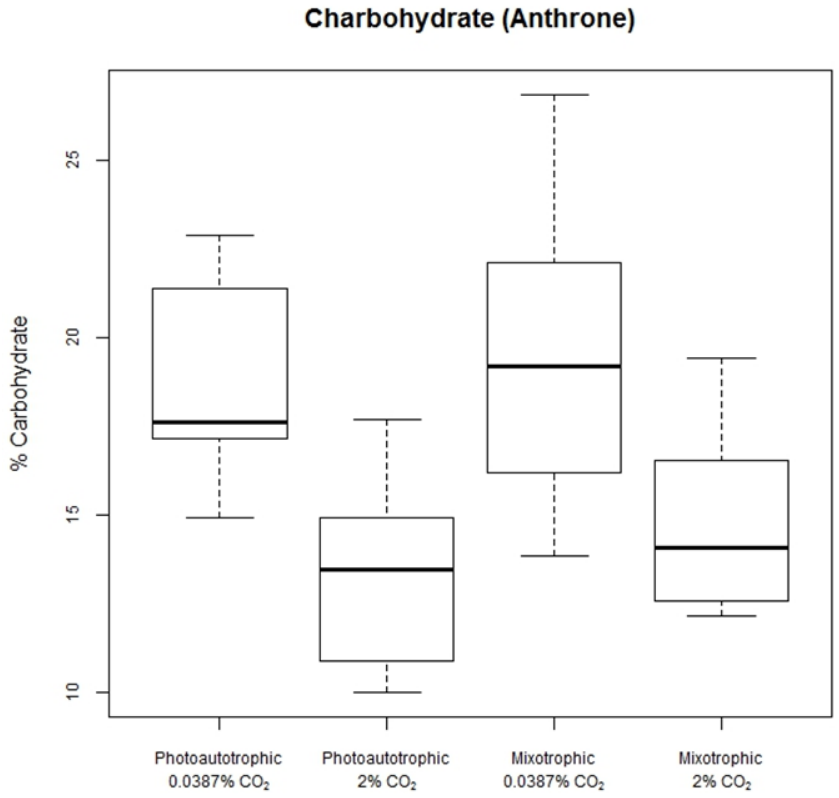
The trend in lipid (Figure $6.4B_{p170}$) variation for conditions A, B & C corresponds with carbohydrate observations, reducing with CO_2 supplemented photoautotrophic growth. However, unlike with carbohydrates lipid content is significantly (p = $9.38x10^{-5}$) accumulated under mixotrophic CO_2 supplementation.

The level of pigments (chlorophyll a, b and total carotenoids) is shown in Figure $6.4C_{p170}$. As seen with carbohydrate and lipid, pigment levels varied depending on the carbon source, though observations were marginal and many more biological replicates would be needed to provide statistical significance. Within the acetate free conditions, total pigment increased from 2.4% (%DCW) to 3.4% (p = 0.069) under CO_2 supplemented growth. This

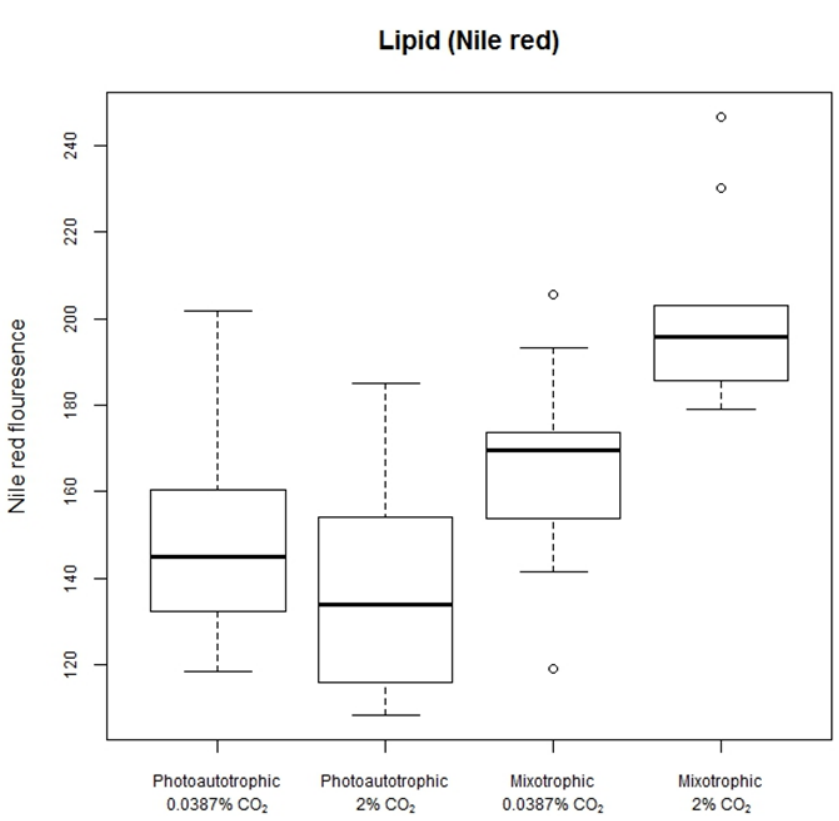
increase is reversed within the acetate growth conditions. Decreasing from 3.0% to 2.6% (p = 0.073).

The increase in pigment when comparing condition A to B could suggest an increase in photosynthesis under increased carbon availability. Evolutionarily, it would have been expected that both photoautotrophic cultures would have more pigmentation than the mixotrophic culture. The observation of condition A as the lowest pigmentation level therefore suggests that the cultures are limited in their pigment production due to carbon availability limitations.

A)


Charbohydrate (Anthrone)
% Carbohydrate
10
15
20
25
Photoautotrophic 0.0387% CO2
Photoautotrophic 2% CO2
Mixotrophic 0.0387% CO2
Mixotrophic 2% CO2

B)


Lipid (Nile red)
Nile red flouresence
120
140
160
180
200
220
240
Photoautotrophic 0.0387% CO2
Photoautotrophic 2% CO2
Mixotrophic 0.0387% CO2
Mixotrophic 2% CO2

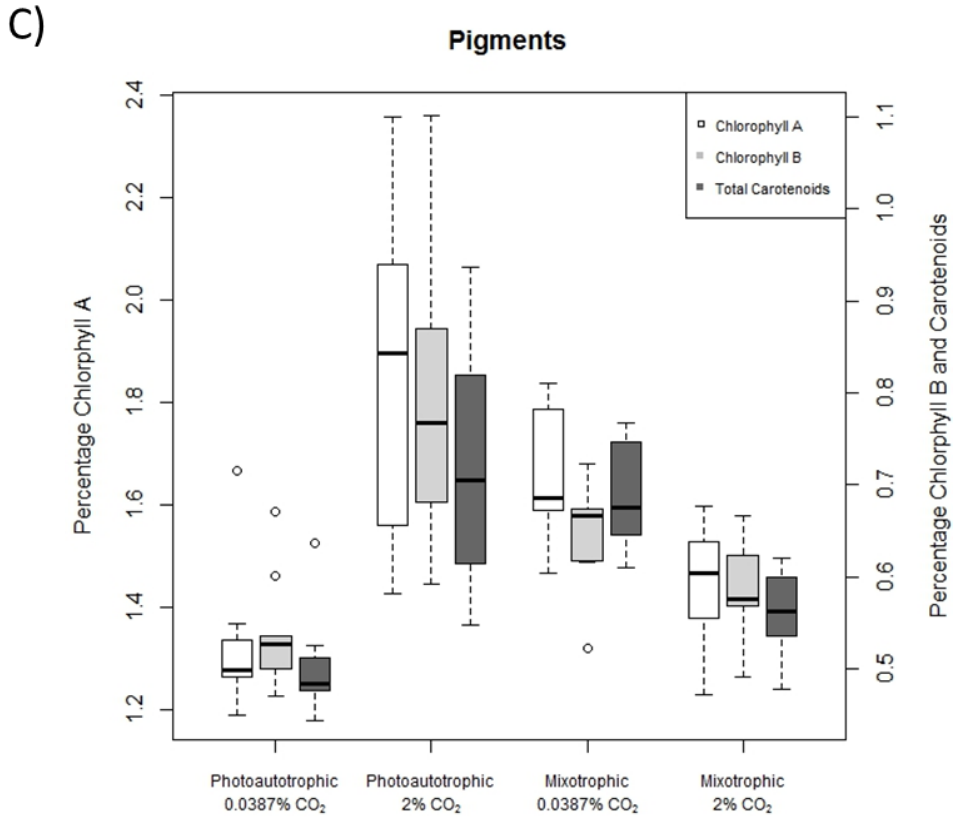


Figure 6.4: Boxplot of carbohydrate, lipid and pigment content observed for *C. reinhardtii* grown both photoautotrophically and mixotrophically each either under 0.04% CO_2 and 2%CO_2. Each boxplot was derived from 9 observations comprised of three biological replicates and three sampling replicates.

6.4.4: Overview of proteomic data

To investigate the changes occurring under the different trophic conditions an 8-plex iTRAQ experiment was performed. Two biological replicates of each condition were labelled as described in Table 6.1_{p170} facilitating investigation of multiple comparisons

Table 6.1: Table showing iTRAQ label arrangement for the four conditions under investigation. Also shown is the labelling 'Condition A-D' used as shorthand to describe the four conditions.

	0.0387% CO_2 at 1L min^{-1}		2% CO_2 at 1L min^{-1}	
Photoautotrophic	Condition A	113	Condition B	114
		115		116
Mixotrophic	Condition C	117	Condition D	118
		119		121

Within the dataset, 28,218 peptide spectral match(es) (PSM(s)) were made each with an iTRAQ quantification. Being derived from the merging script outlined in Chapter 4 (Appendix 6.1_{p284}), this is a combination of six search engines with conflicting matches removed. These matches represented 7,892 unique peptide sequences and 1,862 proteins of which 995 were identified by two or more unique peptides.

To assess sample arrangement, principal component analysis (Figure 6.5_{p172}) was performed on the merged PSM list. The PC1 can be seen to account for 74% of the variation compared to 10% for the PC2. Hence changes observed along the 'X' axis are 7.4 fold more significant than those on the 'y'. From the PCA, It is observable that condition A separates from the other conditions with condition D being the most extreme opposite from condition A. This is suggestive that the primary effect of differentiation is the carbon availability. As such, condition A 'un-supplemented photoautotrophic' with limited carbon availability and condition D 'supplemented mixotrophic' with high carbon availability, form the two extremes of the primary principal component space.

With the exception of replicate 'D|118', the PCA also suggest a distinction between autotrophic (A and B) with mixotrophic growth (C & D) along the second principal component. This however is only suggestive given the limited number of observations under comparison and the peculiarity of D|118.

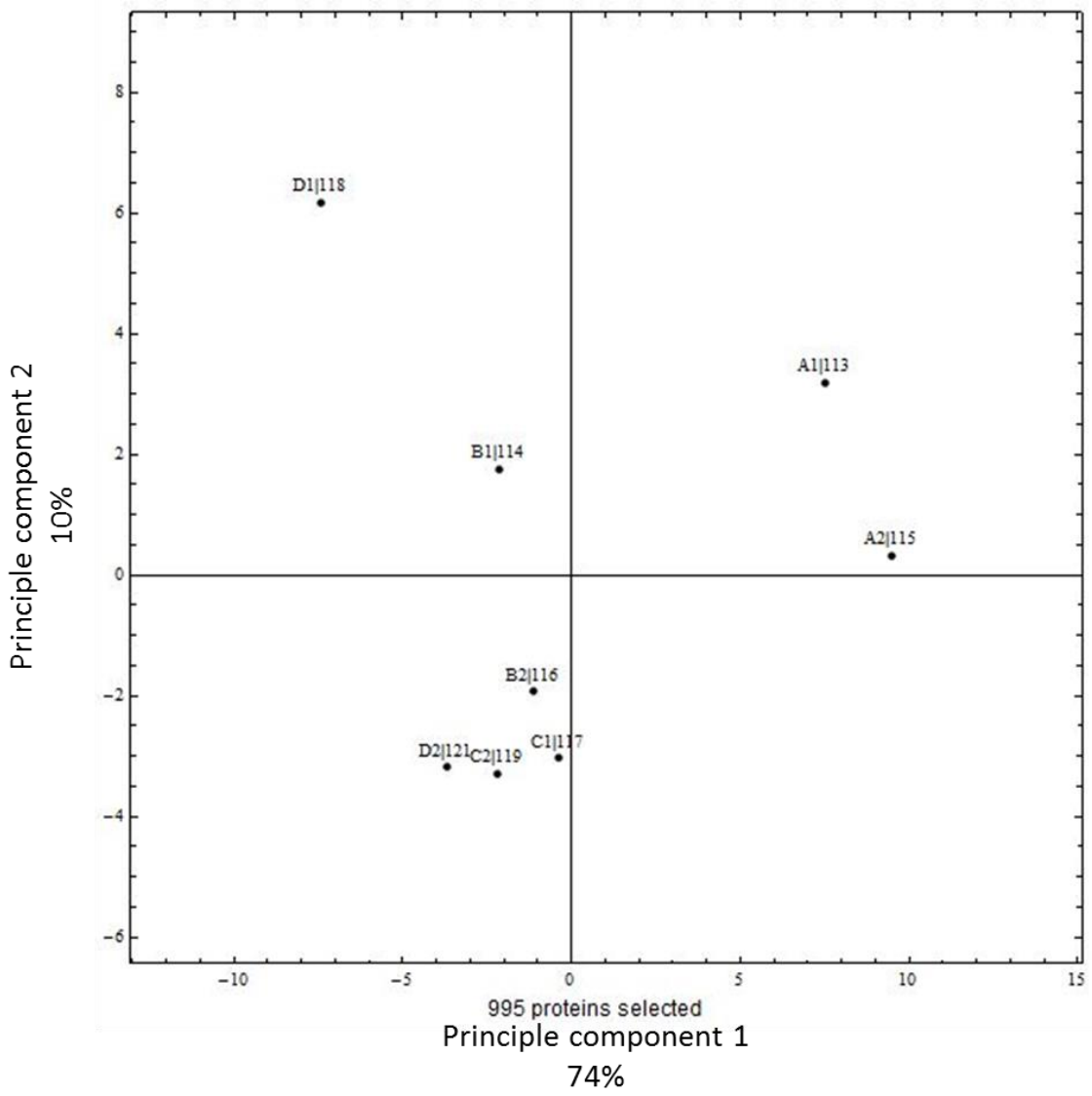


Figure 6.5: Principal component analysis showing the clustering for samples based on VSN, IC, MC corrected reporter ion intensities, limited to ≥2 peptide proteins.

6.4.5: Proteomic changes

Within this dataset, there are multiple comparisons that can be made including:

- Six two way comparisons of each condition against each other.
- Two single factor comparisons (e.g. acetate +/- irrespective of CO_2 and 0.04%/2% CO_2 irrespective of acetate.)
- Condition A vs. condition B,C and D combined (based on PCA arrangement)

Some of these comparisons will be used in subsequent analysis with all comparisons being included in the fold change and significance table (Appendix 6.2$_{p284}$).

With any iTRAQ based experiment (unless combining multiple iTRAQ reactions), there is a limit of 8 channels of investigation available. In this investigation, we selected to scrutinize closely related conditions each with two replicates. Thus, when considering statistical comparisons, extremely high significance should not be expected. To investigate the changes occurring two thresholds were used.

A *p*-value less than 0.05 was required for inclusion in grouped analysis such as KEGG pathways and Gene Ontology observations. This lower significance level is accepted for these evaluations, as discussion derived from such is based on multiple identifications providing a cumulative effect. For tables of proteins the threshold of 0.01 was used allowing comment on single proteins observations rather than on observations of multiple protein trends.

Carbon low vs. high

From the PCA (Figure 6.5$_{p172}$), it is clear that the majority of the variance between the samples is between condition A and each of the other three conditions. This would presumably be a result of the availability of carbon, with condition A being more carbon-limited compared to conditions B, C and D. Based on this assumption the comparison of A vs. D was conducted (volcano plot shown in Figure 6.6$_{p174}$) and termed 'Carbon Low vs. High'.

Within the low vs. high comparison, 165 proteins were significantly identified at as different (p <0.05) with 45 highly significant proteins (p < 0.01).

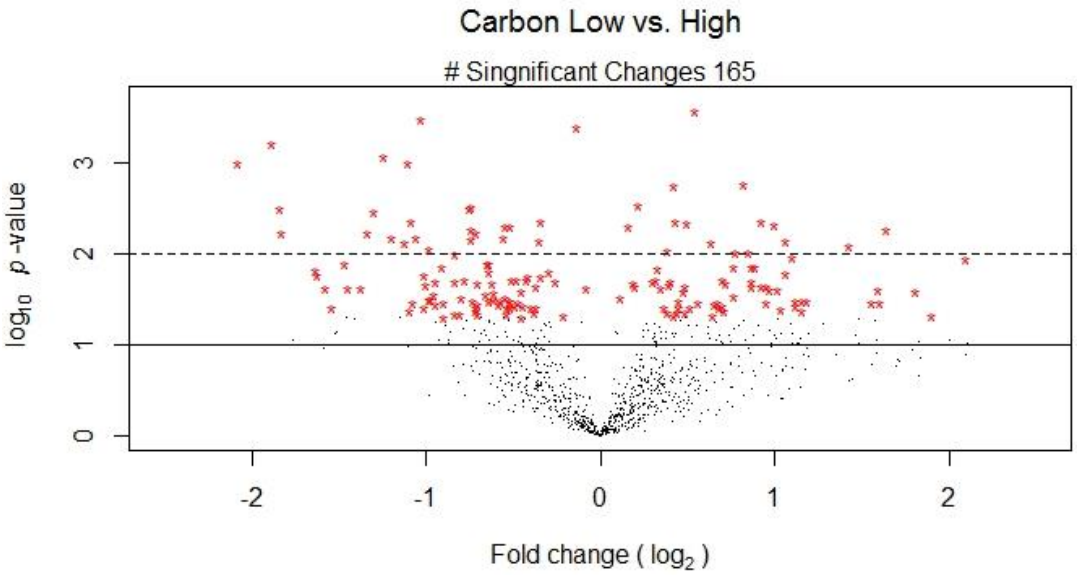


Figure 6.6: Volcano plot showing the changes observed when comparing the low carbon condition Photoautotrophic 0.04% CO_2 vs. high carbon source Mixotrophic 2% CO_2.

To assess the differences identified, the significant changes were grouped by gene ontology (GO). The number of representatives of each grouping were identified is shown in Figure 6.7$_{p175}$.

When considering statistically significant (p-value < 0.05) protein changes, the ratio of increase vs. decrease was 69 vs. 96. However, when considering assignments to GO groups the up vs. down is 112 vs. 15. This suggests that the database assignment and description is more complete for the increased proteins compared to the decreased proteins, perhaps indicating the increased abundance is in core proteins whilst less common proteins comprise the decreased proteins.

Groups identified as decreased included homeostasis and electron transport chain. These are indicative of an improved cellular environment with reduced pressures for homeostatic control.

Within the increased GO(s), multiple categories correspond to biosynthetic processes (Figure 6.7$_{p175}$). This is indicative of an increased cellular

production in what can be considered the more favourable growth environment of increased carbon availability.

To assess the proteomic changes, with respect to known pathways, the significant ($p < 0.05$) proteins were used to colour the KEGG pathway map. Despite a slight increase in lipid content demonstrated in Figure 6.4B$_{p170}$, the limited matches made to the lipid biosynthesis pathway were inconclusive in nature (Appendix 6. 3$_{p284}$).

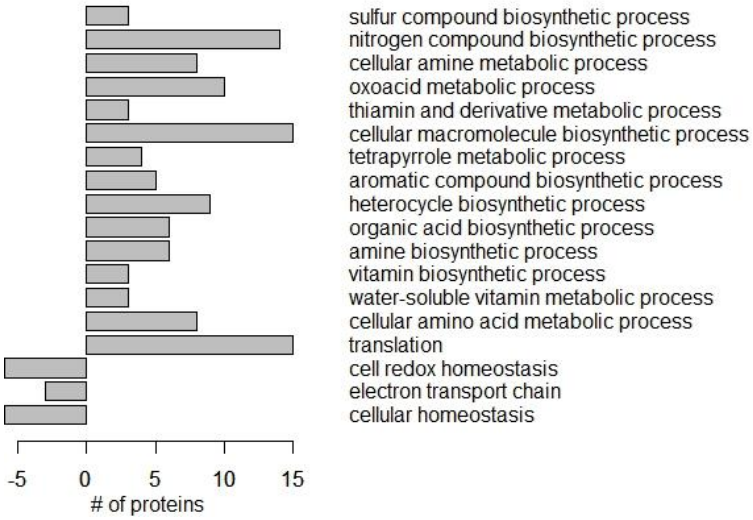


Figure 6.7: Bar chart of assigned GO groups. Based on significantly ($p < 0.05$) identified protein changes comparing condition A vs. D (photoautotrophic 0.04% CO_2 vs. Mixotrophic 2% CO_2).

Of the 45 highly significant ($p < 0.01$) changes, 7 matched predicted proteins. The remaining 38 proteins were grouped by function and are reported in Table 6.2$_{p176}$.

From these significant protein changes, several trends can be discerned. As observed in the GO analysis, translation is shown to be increased. In addition, amino acid biosynthesis is also increased.

As would be expected, given the increase in carbon and CO_2 availability, multiple proteins involved in carbon dioxide absorption such as carbonic anhydrase and LCI A,B & C [347] were increased.

Despite the marginal increase in photosynthetic pigments observed in Figure 6.4C_{p170}, from the proteomic changes it appears that photosynthetic energy production was decreased under CO_2 and acetate. This decreased reliance on phototrophism is also reflected in the decrease in chloroplast ATP synthase (B7U1J0 & A8HXL8) and corresponding increase in mitochondrial ATP synthase (A8HTX7). This increase in respiration is however not consistent with reduction in two subunits of 'NADH:ubiquinone oxidoreductase', an initiating enzyme in oxidative phosphorylation.

Contrary to the lipid assessment of Figure 6.4B_{p170}, the fatty acid initiator Acetyl-CoA biotin carboxyl carrier is observed to be decreased. This suggests that despite the increased lipid level, the level of active lipid production was not increased. There are multiple possible explanations for this observation. These could include the use of an alternative pathway under the changed carbon source availability. A further explanation is that under the carbon restricted condition the consumption of lipid is larger than that of the non-'carbon limited' condition. Thus, despite a reduction in lipid production, an increase in lipid content could be observed. Both these and other hypothesised explanation would require further investigations to understand in detail.

Table 6.2: Protein observed to be significantly (p <0.01) changes when comparing condition A to that of condition D.

Uniprot ID	Protein Name	# Unique Peptides	p.value	Fold Change
	Amino Acid Biosynthesis			
A8IMY5	Anthranilate synthase, alpha subunit	4	4.34E-03	**1.35**
A8I0V1	Aspartokinase	5	4.37E-03	**1.90**
A8JH48	3-phosphoshikimate 1-carboxyvinyltransferase	2	4.74E-03	**2.00**
	Photosynthesis			
A8JHU0	Malate dehydrogenase	39	3.26E-04	-2.04
B7U1J0	ATP synthase subunit alpha, chloroplastic	29	7.05E-03	-1.27
A8HTK7	PsbA translation factor	4	7.50E-03	-2.18
A8HXL8	Chloroplast ATP synthase gamma chain	13	4.04E-04	-1.10
A8I835	Photosystem I reaction center subunit N	4	4.42E-03	-2.12
A8IMZ5	Magnesium chelatase subunit I	17	4.89E-03	**1.12**
P18068	Plastocyanin, chloroplastic	2	5.80E-03	-1.64
A8JEV1	Oxygen evolving enhancer protein 3	28	4.36E-03	-1.27
	Respiration			
A8JI60	Type-II calcium-dependent NADH dehydrogenase	3	9.38E-03	**1.80**

Q6UKY6	NADH:ubiquinone oxidoreductase 78 kDa subunit	10	5.24E-03	-1.67
Q6V505	NADH:ubiquinone oxidoreductase 17.8 kDa subunit	2	4.93E-03	-1.46
A8HTX7	Mitochondrial F1F0 ATP synthase associated 31.2 kDa protein	17	9.10E-03	**1.30**
	Fatty Acid Biosynthesis			
A8JDA7	Acetyl-coa biotin carboxyl carrier	7	6.53E-03	-1.47
	CO2 limitation			
A8J4Z8	Carbonic anhydrase 3	2	3.03E-03	-1.67
A8JJ91	Mitochondrial carbonic anhydrase, beta type	3	2.40E-03	-11.27
O24451	Envelope protein (Low-CO2-inducible chloroplast envelope protein)	2	2.59E-04	-8.96
Q0Z9B8	LciD (Limiting CO2 inducible protein)	6	6.02E-04	-3.70
Q0ZAI6	LciB (Low-CO2-inducible protein)	30	3.08E-03	-3.58
Q39590	Carbonic anhydrase	16	2.03E-04	-18.31
Q75NZ1	LciC (Low-CO2 inducible protein)	30	9.83E-04	-4.23
	Thiamine Biosynthesis			
A8J841	Hydroxymethylpyrimidine phosphate synthase	13	1.71E-03	**1.77**
	Protien Folding			
A8IVN2	Peptidyl-prolyl cis-trans isomerase	10	6.52E-03	-2.29
A8J3L3	Peptidyl-prolyl cis-trans isomerase	5	8.36E-04	-2.37
	Translation			
A8HXM1	Mitochondrial ribosomal protein L29	9	2.84E-03	**1.16**
A8IZ36	Ribosomal protein S25	11	7.11E-03	**2.08**
A8J888	Tryptophanyl-trna synthetase	3	2.67E-04	**1.46**
A8JDP4	Ribosomal protein L9	7	1.79E-03	**1.34**
A8JGK1	Ribosomal protein S17	3	7.97E-03	**2.68**
	Miscellaneouse			
A8HRZ0	Histone H1	7	4.97E-03	-1.43
A8IJ19	Aldehyde dehydrogenase	11	1.01E-03	-2.15
A8IS90	Aldo/keto reductase	3	5.87E-03	-2.53
A8HN52	Glutaredoxin, CGFS type	5	7.30E-03	**1.55**
A8JBZ2	Pyridine nucleotide binding protein	3	8.88E-03	-1.98
A8J2E9	Glycolate dehydrogenase	14	3.38E-03	-2.46
O64985	Poly(A) binding protein RB47	3	5.24E-03	**3.12**

Single carbon – CO_2 vs. acetate

An aim of this study was to investigate the validity of mixotrophic culture as a model of photoautotrophic culture. As is seen above, considerable change is seen due to the increase in carbon availability associated with acetate or CO_2 culture when compared to atmospheric photoautotrophic culture.

When comparing the CO_2 un-supplemented photoautotrophic to the CO_2 un-supplemented mixotrophic conditions (Figure $6.8A_{p178}$), 191 significant protein changes were observed. This suggests a considerable reorganisation of the proteome between these two culture environments, which must be considered when relating findings of one to the other. By comparing supplemented photoautotrophic cultures to mixotrophic cultures (Figure

6.8B$_{p178}$) there is a reduction in significant differences to 92, suggesting that mixotrophic growth is more akin to supplemented photoautotrophic than un-supplemented photoautotrophic culture.

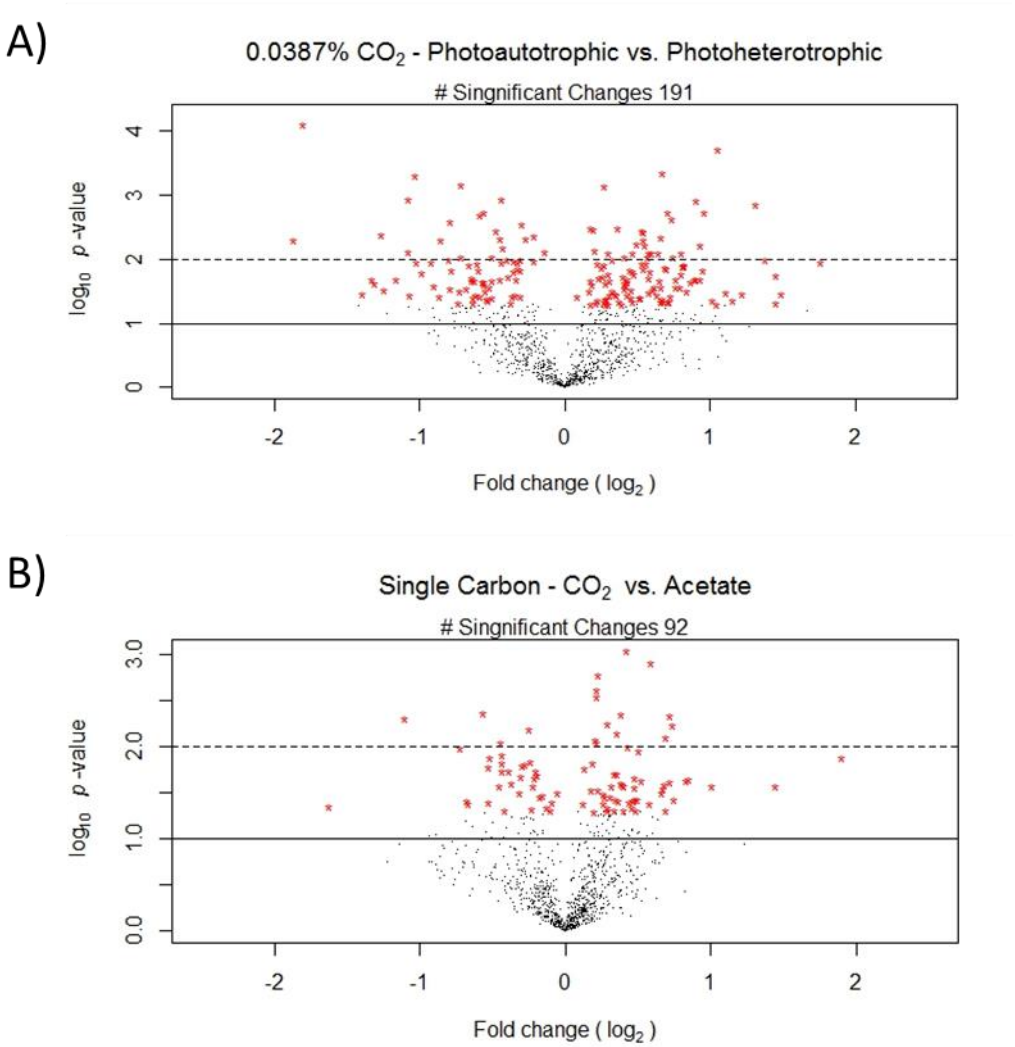


Figure 6.8: Volcano plots showing the changes observed when comparing A) Photoautotrophic vs. Mixotrophic both without CO_2 supplementation. B) Photoautotrophic growth on 2% CO_2 with Mixotrophic growth on 0.04% CO_2.

As with the carbon low vs. high analysis these changes observed between condition B and C were matched to GO classifications (Figure 6.9$_{p179}$). Several trends are apparent for the changes in these GO groups. Reflecting the increase in carbohydrate content (Figure 6.4A$_{p170}$) with acetate rather than CO_2, several GO groups involved in carbohydrate metabolic and catabolic processes are increased. In addition, 'tetrapyrrole', 'heterocycle' and

'terpenoid biosynthesis' are increased. Among the decreased GO groups are 'protein-DNA complexes' and 'nucleosomes'. This suggests increased dormancy within CO_2 cultured samples compared to acetate cultures. This increase in nucleosomes was unexpected and as such further investigation into this observation would be warranted.

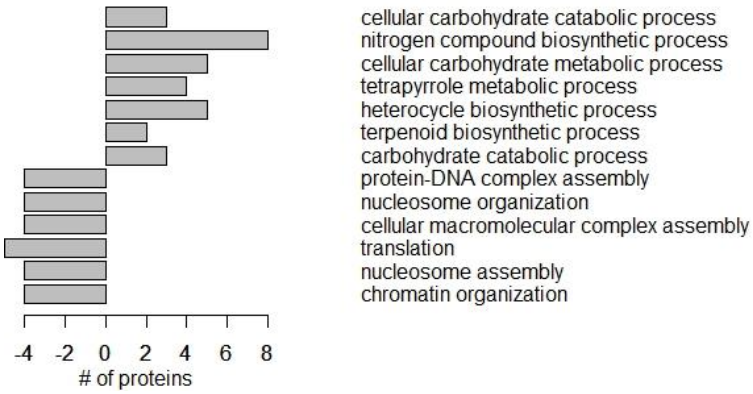


Figure 6.9: Bar chart of assigned GO groups. Based on significantly (p < 0.05) identified protein changes comparing condition B vs. C (photoautotrophic 2% CO2 vs. Mixotrophic 0.04% CO2.

The highly significant changes observed ($p < 0.01$) are reported in Table 6.3_{p180}. Among these was an increase in phosphorylase, responsible for the breakdown of glucagon to glucose and increase in respiration proteins. This is perhaps indicative of a switch from photosynthesis to respiration, as the energy source. Within the mitochondria, there is also evidence of perturbations in the contents of the inter-membrane space with increase in A8HQS2 and decrease in A8IVN6, responsible for transport across the inner and outer mitochondrial membrane respectively.

Within the mixotrophic acetate culture it could have been assumed that heterotrophic growth would be prioritised with consequential reduction in photosynthetic apparatus. Such a reduction is not apparent suggesting either that the culture maintains a true mixotrophic growth or that the photosynthetic apparatus is maintained despite its reduced requirement. This could perhaps improve the cell's fitness given a return to reliance on photoautotrophic growth.

Table 6.3: Protein observed to be significantly ($p < 0.01$) changes when comparing condition B to that of condition C.

Uniprot ID	Protein Name	# Unique Peptides	p.value	Fold Change
	amino acid biosynthesis			
A8IWA6	Glutamate synthase, NADH-dependent (EC 1.4.1.14)	4	4.49E-03	**1.64**
A8JG03	Isopropylmalate dehydratase, large subunit	14	5.57E-03	**1.22**
E9N731	Glutamyl-tRNA reductase binding protein	5	9.10E-04	**1.34**
	Photosynthesis			
A8J4K4	Uroporphyrinogen-iii synthase	8	1.21E-03	**1.50**
Q9S7V1	Coproporphyrinogen III oxidase	9	7.11E-03	**1.28**
	Respiration			
A8HP06	Succinate dehydrogenase subunit A	23	2.40E-03	**1.16**
A8JCF6	Mitochondrial pyruvate dehydrogenase complex, E1 component, alpha subunit	7	9.06E-03	**1.16**
A8HQS2	Mitochondrial inner membrane translocase	5	1.67E-03	**1.17**
A8IVN6	Translocase of outer mitochondrial membrane	2	4.89E-03	-2.15
	Carbohydrate breakdown			
Q2VA41	Phosphorylase (EC 2.4.1.1)	5	7.83E-03	**1.61**
	protien production / processing			
A8HMC0	Calreticulin 2, calcium-binding protein	4	2.86E-03	**1.16**
A8J2N4	Dual specificity protein phosphatase 8 (Fragment)	2	6.46E-03	-1.19
A8HQT1	Protein disulfide isomerase	22	9.97E-03	**1.35**
A8I6B6	Peptidyl-prolyl cis-trans isomerase	4	4.26E-03	-1.48
A8INR7	Plastid ribosomal protein L27	4	9.09E-03	-1.36
	Miscellaneous			
A8IZU0	Heat shock protein 70C	27	5.74E-03	**1.66**
A8JIN8	N-carbamyl-L-amino acid amidohydrolase	2	8.40E-03	**1.15**

The above comparison is a direct approach to investigating the difference between carbon dioxide and acetate as a carbon source. An alternative to this direct comparison is to investigate contradictory differences when each are compared to condition A. To identify these contradictions, all proteins with a $p > 0.1$ for both comparisons were removed. This gave a pool of 166 proteins, of these 166 only 2 were identified as contradictory. These were:

- Glyceraldehyde 3-phosphate dehydrogenase, dominant splicing(A8JHR9) – Involved in glycolysis this was identified as being decreased by 1.94 fold upon increasing CO_2 but increased 1.93 fold upon addition of acetate.
- Predicted protein (A8IVP1) which matched with 46% similarity to I0YP46, SMAD/FHA domain-containing protein –A possible

transcription factor was identified as increased 1.17 fold upon increasing CO_2 but decreased 1.23 fold upon addition of acetate.

These two anomalies warrant further investigation as peculiarities, but the major observation of this comparison is the similarity of the two comparisons. Given 166 proteins significantly identified by two or more unique peptides and identified as significantly changing in both condition B and condition C as compared to condition A, only 2 were contradictory in regulation. This further suggests that carbon availability is the predominant factor influencing the culture rather than the source of carbon.

6.5: Conclusions

From compositional analysis within photoautotrophic cultures, addition of CO_2 caused a reduction in carbohydrate and lipid content. This trend was also observed for carbohydrate upon CO_2 supplementation of mixotrophic cultures, but for lipid accumulation, CO_2 supplementation increased rather than decreased content.

CO_2 supplementation, but not acetate, was shown to influence the culture's biochemical composition. Therefore, investigations observing compositional changes should consider whether CO_2 concentrations are being correctly controlled. This is particularly important for experimental set-ups with limited gas exchange, where CO_2 levels are affected by culture density.

When conducting the proteomic analysis, we chose to analyse all four conditions each with two biological replicates. As may have been expected, this caused a decrease in the level of statistical significance derived from any single comparison. The inclusion of all four conditions did however facilitate a greater understanding of the factors involved. Given a solely acetate +/- investigation or a solely 0.04% / 2% CO_2 investigation the effect of carbon availability would have overshadowed that of the effect of the source itself.

The lack of observe proteomic changes between acetate and CO_2 growth may suggest that whilst capable of taking advantage of heterotrophic growth opportunities *C. reinhardtii* does not considerably re-model its proteome to favour such conditions when available, but rather maintains an ability to utilise both.

This investigation identified significant changes based on the trophism. It identifies the need to ensure consistent carbon availability between compared conditions during investigations. However, the degree of changes observed was limited as compared to those found in nitrogen stress (Chapter 5) and as such the observations made from one trophism should be largely comparable to the others.

6.6: Further work

Through this investigation we noted minimal changes, at the proteomic level, between the three cultures without carbon limitation (i.e. un-supplemented photoautotrophic). As such the mixotrophic nature of the two acetate based cultures could be questioned. A repeat of the Heifetz *et al*[353]. isotopic carbon study utilising the turbidostatic culture demonstrated within this study would allow clarification to the truly mixotrophic nature of the culture.

Within this investigation some intriguing observations were made about the effect of carbon source, both type and amount, on carbohydrate, lipid and pigment concentrations. These observations were based on only three biological replicates and thus statistical significances were mostly marginal. The subtlety of the changes observed suggest a follow-up investigation with more replicates is desired. Given the effect culture density has on carbon availability on a per cell basis it would be preferential for such further replicates to be conducted in turbidostatic conditions. The size of the culture however would not be required to be 1 L as in this investigation and smaller

but appropriately controlled reactors would be preferential to facilitate larger number of replicates.

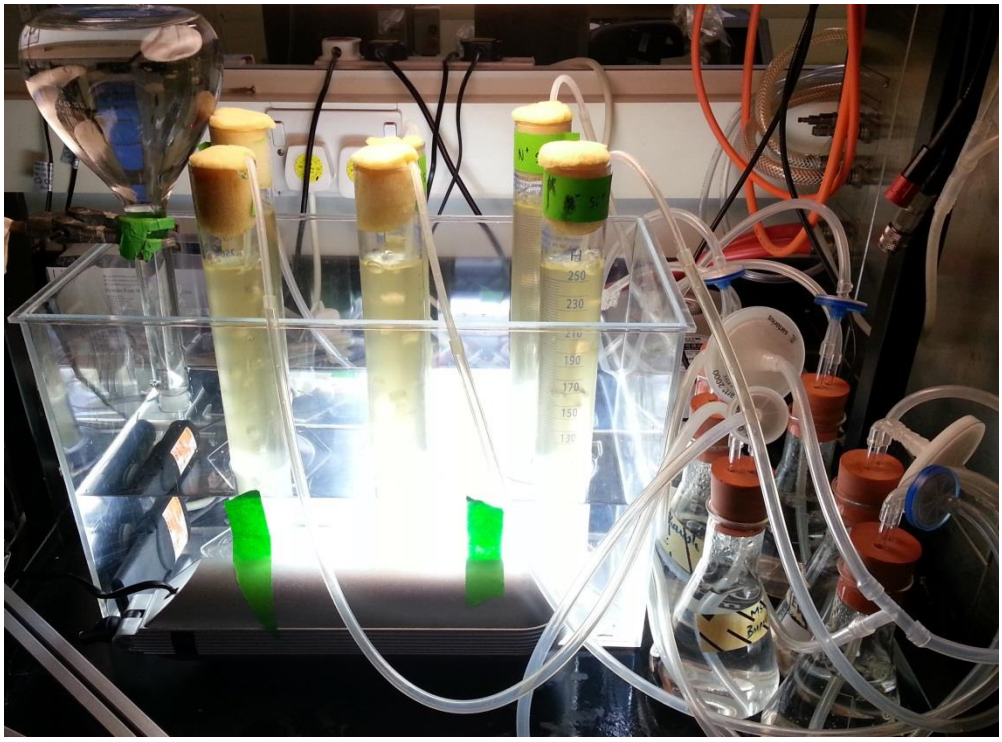
The possible increased dormancy observed during CO_2 supplementation was an unexpected observation and as such should be investigated secondarily by targeted proteomics approaches such as MRM or non-targeted proteomics investigation but focused on the nucleosome, perhaps through organelle fractionation.

The observation of CO_2 influences on carbohydrate and lipid content were unexpected prior to the investigation. Though this had a limited effect as compared to other triggers such as nitrogen depletion, this warrants further investigation. The biochemical pathways underpinning this were not clear from this investigation. A more targeted approach such as MRM may help with this. This effect of carbon source should also be investigated in combination with other triggers such as nitrogen stress to see if a combined effect can be observed.

A significant hindrance of this investigation was the limited number of replicates within the proteomics analysis. From this investigation, the comparison of condition A vs. C for general carbon availability and B vs. C for source type, suggest themselves as the important comparisons. Such investigation could be conducted as two 8-plex iTRAQ investigations, or through SILAC or Label free methods which depending on experimental set up can be less restricted in the number of channels of investigation.

Given this study suggests that the proteome does not dramatically change due to type of carbon source but rather the availability of carbon it would be advantageous to assess the cell at other omic levels. In particular metabolomics, given that whilst it is proteins which allow the cell to influence its environment it is the proteins along with available metabolites which dictate the essence of the cell under any given condition

Chapter 7: *Phaeodactylum tricornutum* under nitrogen stress: A proteomic investigation



Chapter 7: *Phaeodactylum tricornutum* Under Nitrogen Stress: A Proteomic Investigation

7.1: Abstract

The use of nitrogen stress as a trigger mechanism for increased lipid content of microalgae has become well established. Investigations have predominantly focused on Chlorophyta species, but increasing interest has developed in other lineages of microalgae such as the diatoms. The sequenced diatom *Phaeodactylum tricornutum* is a good candidate for investigation of this nitrogen trigger beyond the Chlorophyta lineage.

In this investigation, 24 h nitrogen stressed cultures were investigated at the proteomic level through an LC-MS/MS based iTRAQ investigations. 1043 proteins were confidently identified (≥2 unique peptides) with 645 significant (p<0.05) changes observed.

Through gene ontology classifications, KEGG pathways and detailed analysis of highly significant changes several trends were observed, which included decreased translation, photosynthesis and fatty acid catabolism whilst observing increased central metabolism, fatty acid biosynthesis, antioxidants, endocytosis and phagocytosis.

This investigation increased the number of identified proteomic changes within *P. tricornutum* under nitrogen stress by 17-fold. This allows for a deeper understanding of nitrogen induced lipid induction, as well as providing further descriptive information of 645 proteins. Further, our investigation supports the concept that a reduction in lipid catabolism perhaps more than lipid synthesis is a key factor in lipid accumulation under nitrogen stress.

7.2: Introduction

With the increased interest in microalgae for a biofuel feedstock, there has been increasing interest in the commonly observed nitrogen trigger. The nitrogen trigger is a common feature, with multiple algal species where reduced nitrogen stress causes increase in storage of energy reserves. This has been demonstrated in multiple species.[39,63] Recent studies (including that presented earlier in Chapter 5) have attempted to further understand this stress using the model algal species *Chlamydomonas reinhardtii.*[173,179,180] However, extrapolation of findings in this one species to all microalgae would be inappropriate, particularly given the diverse lineage of organisms referred to as microalgae.[354]

Microalgae encompass a broad spectrum of species and include members of Chlorophyta, Rhodophyta and Bacillariophyta (diatoms). As a collective term they contain a wide diversity, sharing only 16.3% orthologarity.[73]

Diatoms play a significant role in the world ecology, accounting for ~20% of global photosynthesis[355] and as such warrant investigation from an ecological standpoint alone. The marine diatom *Phaeodactylum tricornutum* has been utilised for aquaculture[356] and as a model for cell morphological investigations.[357] Its' marine nature is of particular interest for its use as a biofuel crop due to the reduced requirement for fresh water, a dwindling resource.[358] *P. tricornutum* has also been noted as unique among diatoms, having only a facultative requirement for silicon.[357] Further, Youngmanitchai and Ward showed an increase in lipid content under silicon deprived conditions.[359]

With the increasing interest in microalgal based biofuels *P. tricornutum* has also begun to be used, along with *Thalassiosira pseudonana,* as a model for diatom based biofuels.[360] Song *et al.*[361] recently selected *P. tricornutum* as a favourable strain for biodiesel production, citing its high lipid content of 61%

and lipid productivity of 26.75 mg L^{-1} d^{-1} as well as, suitable lipid profiles for octane, iodine number and cloud point.

P. tricornutum has been shown to respond to the nitrogen trigger increasing from 21% to 26% lipid (% dry cell weight (DCW)) upon nitrogen stress.[63] The genome of *P. tricornutum* has been sequenced[362] with descriptive information available in UniProt, and KEGG. As such, it is a good candidate to investigate the nitrogen trigger in a diatom species and allows for comparison to Chlorophyta species.

Transcriptomics can be used to make inferences of proteomic changes. This however provides only assessment of transcriptional control and not of translational control and degradation control both of which affect to the amount of protein present.[141] This is of particular relevance in a nitrogen stress environment where protein degradation for nitrogen recovery may play a significant role.[179] Hence, transcriptomic studies cannot be relied upon to represent the true protein levels.[129] They should either be supported with targeted protein analysis such as western blots or MRM or a proteomic investigation itself.

Previous transcriptomic studies have utilised both microarray[363] and RNA-SEQ[364,365] to investigate the effect of nitrogen stress. Here we attempt to support these initial findings with proteomic based data. For proteomic analysis, we utilised isobaric tags for relative and absolute quantitation (iTRAQ).[157] iTRAQ utilises amine linking isobaric tags to allow quantitative comparison of numerous proteins in an unbiased way and has become a popular tool for proteomics.[161]

Whilst several investigations have been published in the Chlorophyta,[166,197,230] only limited proteomics investigations has been published in other phyla. Within diatoms, *T. pseuodonana* has been investigated sporadically[257–260,262,366] with even more limited investigations on *P. tricornutum.*[260,265] These include an identification study of the

thylakoid membrane proteome and a recent study by Yang *et al.,*[265] where they reported proteomic changes under 48 h nitrogen stress in *P. tricornutum.* This investigation utilised a 2D SDS Page gel methodology and identified 37 proteins as being significantly differentially regulated. Through the use of mass spectrometry based quantification, we aim to increase their observations.

In this study, *P. tricornutum* was cultured in an air lift reactor in a batch manner harvesting within exponential phase. The cultures were then pooled before resuspending in nutrient replete or deplete conditions giving a fixed time point of stress. Both nitrogen and silicon depletion were assessed at 24h post stress induction. Samples were assessed biochemically for pigment, carbohydrate and lipid content. In conjunction, samples were assessed for proteomic changes using the iTRAQ methodology to understand the changes occurring within the culture proteome.

The goal of this investigation was to further the understand of the relationship between nitrogen stress and lipid accumulation within microalgae to non-chlorophyta. The sequenced diatom P. tricornutum was selected for investigation. Such understanding from multiple species would help further understanding of this common but not universal relationship. Thus potentially allowing superior control over induction of lipid accumulation in microalga during biofuel production. To understand this relationship the question of how the proteome was affected during nitrogen induced accumulation was proposed. To this end an iTRAQ was conducted comparing the *P. tricornutum* proteome before nitrogen starvation to 24 hrs post nitrogen stress induction.

7.3: Materials and method

7.3.1: Organism and medium

P. tricornutum (CCAP 1055/1) was obtained from the Culture Collection of Algae and Protozoa (CCAP, Oban, U.K.). Cultures were grown in ether F/2+Si media (as described in Section 3.2_{p67}) or F/2+Si media omitting respectively sodium nitrate or sodium metasilicate for nitrogen and silicon free media.

7.3.2: Culture

P. tricornutum was cultured (Figure $7.1A_{p191}$) in 250 mL measuring cylinders sparged with air at 2.4 L min^{-1}. Filtered air was first passed through sterile water to humidify it, before being introduced by silicone tubing to the bottom of the cylinder such that it's addition provided mixing as well as gas transfer. The tops of culture vessels were sealed using a foam bung. Cylinders were placed in a water bath maintained at 25 ^{0}C and under 24 h illumination with two side facing halogen lamps providing 200 ± 50 μE m^{-2} s^{-1} to the centre of a dH_2O filled cylinder. Cultures were grown using F/2+Si media for 72h reaching an optical density (750nm) > 0.4. Culture from four culture cylinders was pooled and the combined optical density used to calculate the culture volume required giving an optical density of 0.2 upon re-suspension to 250 mL. The calculated culture volume was harvested from the mixed culture by centrifugation at 3000 g for 5 min. This was then resuspended in F/2+Si medium with or without nitrate and silicon to give four culture conditions:

I. Normal replete conditions (N^+ Si^+)
II. Nitrogen stressed (N^- Si^+)
III. Silicon stressed (N^+ Si^-)
IV. Nitrogen & Silicon stressed (N^- Si^-)

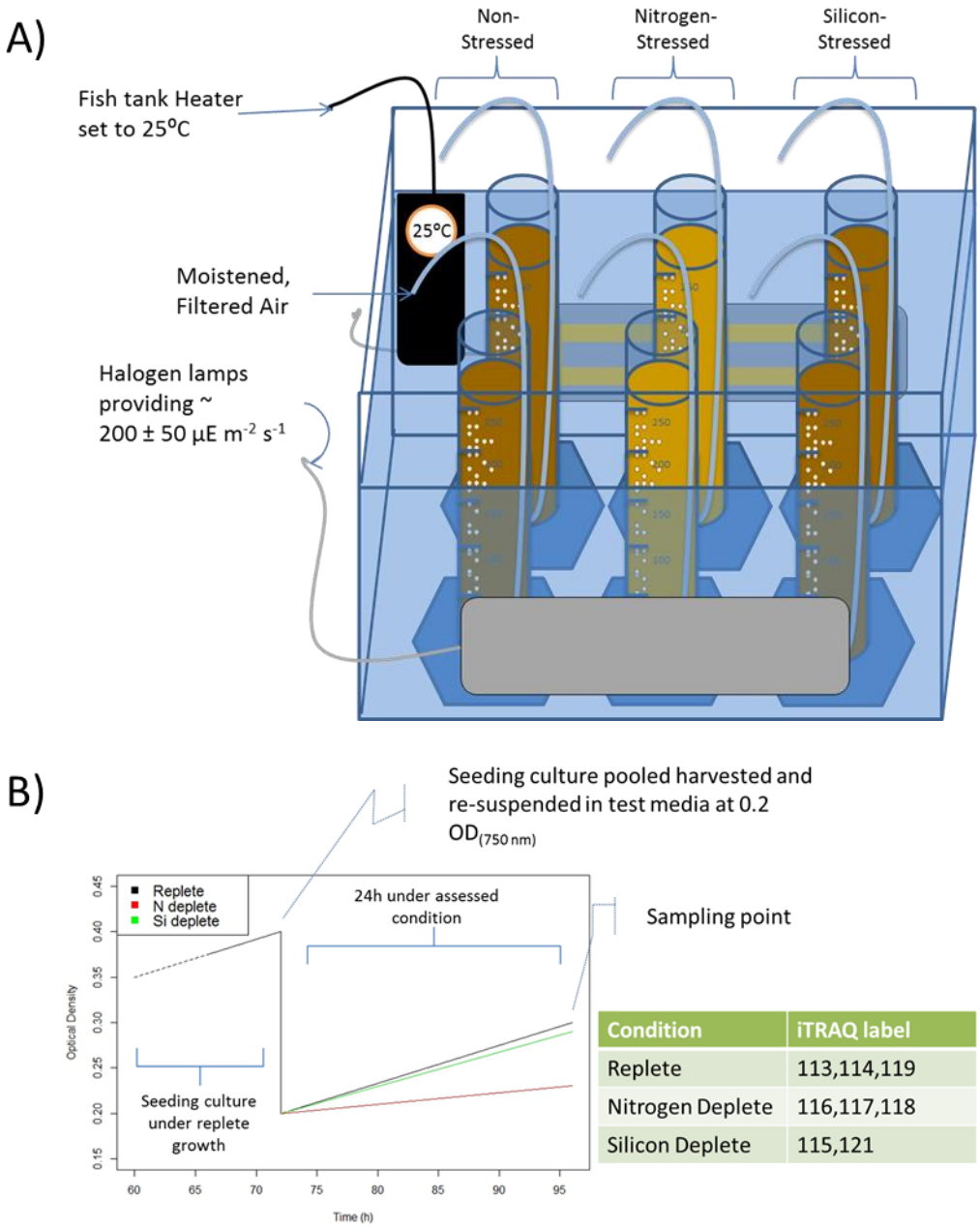


Figure 7.1: A) Diagram of culture set-up. Cultures were grown in 250 mL measuring cylinders submerged a water bath maintained at 25°C, illuminated with two halogen lamps creating 200± 50 µE m-2 s-1. Cultures were mixed and aerated with moistened, filtered air. B) Graphical depiction of the culture progression from pre-stress through stress induction and harvest. Conditions under assessment are shown along with appropriate iTRAQ labels.

7.3.3: Sampling for biochemical analysis

For the initial assessment, 15 mL samples were taken from each at 0, 25, 47, 70 & 143 hours measuring the optical density (750nm). Cultures were then

centrifuged at 3000 *g* before transferring to a pre weighed eppendorf tube and freeze dried according to Section 3.4$_{p68}$.

For the biochemical assessment of proteomics a 50 mL sample was processed in the same way.

7.3.4: Microscopy

Microscopy was conducted as described in Section 3.3$_{p67}$.

7.3.5: Biochemical analysis

For biochemical analysis:

- Pigment assessment was conducted using a method modified from Wellburn, as described in Section 3.5$_{p68}$. Chlorophyll b, which is not present in *P. tricornutum*, was as expected not detectable.
- Carbohydrate assessment was conducted using the Anthrone method, as described in Section 3.6$_{p68}$.
- Lipid assessment was conducted using a Nile Red assay, as described in Section 3.7$_{p69}$.

7.3.6: Proteomic sampling

At 24 h, two 50 mL aliquots were taken by siphoning from the aeration tube with a syringe. The collected sample was centrifuged at 3000 *g* for 10 min at 4 ˚C, then resuspended in 1 mL 500 mM triethylammonium bicarbonate buffer (TEAB) (pH 8.5) and transferred to protein low bind tubes. Samples were then stored at -20 °C till all harvests were completed.

7.3.7: Proteomic sample processing

Protein samples were extracted by liquid nitrogen grinding (Section 3.8$_{p70}$), before quantification by RCDC (Section 3.9$_{p70}$). These were reduced, alkylated and digested (Section 3.10$_{p71}$) before assessing digestion by 1D SDS page gel as shown in Section 3.11$_{p71}$. These were then labelled as described in Section 3.12$_{p72}$ with label assignment as shown in Figure 7.1B$_{p191}$.

7.3.8: Off line HPLC fractionation and clean-up

Hydrophilic interaction liquid chromatography (HILIC) was conducted as described in Section 3.14$_{p73}$. Post processing HILIC samples were cleaned prior to LC-MS/MS using C18 micro spin columns as described in Section 3.15-$_{p73}$.

7.3.9: LC-MS/MS

Reverse phase LC-MS/MS was conducted using an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA, USA) and a QStar XL Hybrid ESI Quadrupole time-of-flight tandem mass spectrometer (Applied Biosystems (now ABSciex), Framingham, MA, USA) as described in Section 3.16$_{p74}$.

7.3.10: Data analysis

The LC-MS/MS data was processed using the optimised method, as developed and described in Chapter 4. This gave an output of PSM(s) from the combined and conflict-resolved outputs of 6 proteomic search engines. These were then processed using a script developed and described in Chapter 4 (summarised in Section 3.17$_{p74}$) to give a list of fold changes between relevant conditions each with a *t*-test based significance factor. Gene ontology and KEGG data processing was conducted as described in Section 3.19$_{p75}$ and 3.20$_{p75}$.

7.4: Results and Discussion

7.4.1: Initial assessment of profiles

To develop an understanding of the time frame within which stress induction affected *P. tricornutum* under the specified growth conditions, an initial trial assessment was conducted. This was done as a single replicate to provide a rapid overview of the gross trend facilitating time point selection for later biological replicated analysis. These trends were not used for understanding of the temporal effect of the stress. For this assessment, four cultures were grown and resuspended in N+ Si+, N- Si+, N+ Si- & N- Si-. Optical density (Figure 7.2A$_{p195}$) shows a distinction between the nitrogen stress

conditions and non-nitrogen stressed with differences developing rapidly at 25 h. Nitrogen replete cultures approximately doubled in density as opposed to a 50% increase for nitrogen deplete cultures. Beyond 25 h, the nitrogen deplete cultures only marginally increased whilst the nitrogen replete cultures continued to grow, though at a reduced rate, beyond 47 h. No distinction between replete and deplete silicon was observed.

For energy storage, whilst lipid concentration (Figure 7.2C_{p195}) showed an expected rapid increase in content for the nitrogen deplete cultures, carbohydrate concentration (Figure 7.2B_{p195}) suggested that silicon deplete culture gave an increased content around 47 to 70 h. However, in addition to the stressed conditions, the replete condition also resulted in an increase in lipid content beyond 47 h. In addition the growth rate of the replete conditions was reduced beyond 47h. These observations suggest that beyond 47h replete is no longer replete and some factor(s) (nutrient, light, self-shadowing etc…) is perturbing the cultures. The chlorophyll a concentration (Figure 7.2D_{p195}), whilst not distinctive, suggests a decrease in nitrogen deplete condition as compared to replete conditions. Based on this initial assessment, 24h post stress was selected as the appropriate time point to conduct the proteomics investigation on nitrogen and silicon stress to study induction of lipid accumulation.

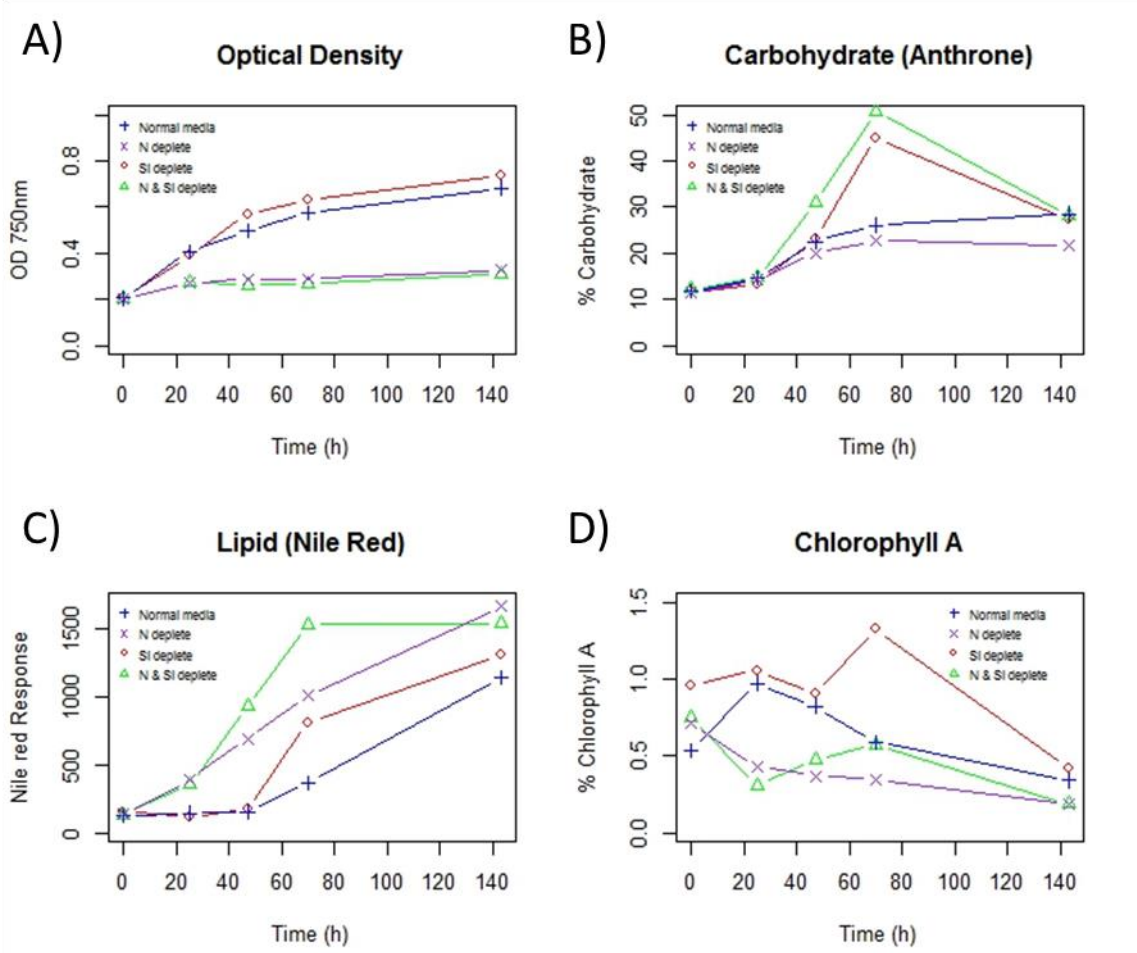


Figure 7.2: Initial culture assessment of P. tricornutum. Assessment shows culture response when transferred to four conditions: normal F/2+Si media , F/2+Si without nitrogen, F/2 without silicon and F/2 without nitrogen or silicon. An assessment of the A) optical density, B) carbohydrate, C) lipid and D) chlorophyll A are shown. The carbohydrate, lipid and chlorophyll are all based from dry cell weight reconstituted to a fixed biomass concentration.

7.4.2: Assessment of proteomic samples

To assess the proteomic changes occurring within *P. tricornutum* under nitrogen and silicon stress an 8-plex iTRAQ experiment was designed. Given the restriction of an 8-plex to 8 investigatory channels, the effect of nitrogen stress was selected as the primary investigation, as preliminary investigations (detailed above) suggested this to be the dominant of the two effects studied. A single time point analysis was chosen using 24 h based on the initial assessment described above. Given this selection, three biological replicates were used for each of the normal, nitrogen stressed and silicon

stressed conditions. Each of these was then used for biochemical and microscopic analysis. For the iTRAQ restricted to 8 channels, one silicon stress condition was dropped from the analysis allowing for a 3 vs. 3 comparison of the primary investigation - nitrogen stress, and a 3 vs. 2 analysis of the secondary investigation - silicon stress.

Cultures of *P. tricornutum* were grown for simultaneous biochemical and proteomic analysis sampling at 24h post stress. Carbohydrate, lipid, and pigment analysis were conducted prior to proteomic assessment.

A *t*-test showed a statistically significant ($p < 0.05$) increase in carbohydrate (Figure 7.3A$_{p197}$) and lipid (Figure 7.3B$_{p197}$) when cultures were under nitrogen stress for 24 h. Silicon stress at 24h did not show a significant effect. Conversely, pigmentation (Figure 7.3C$_{p197}$ and Figure 7.3D$_{p197}$) showed a significant reduction under nitrogen stress with no significant effect being identified in silicon stress.

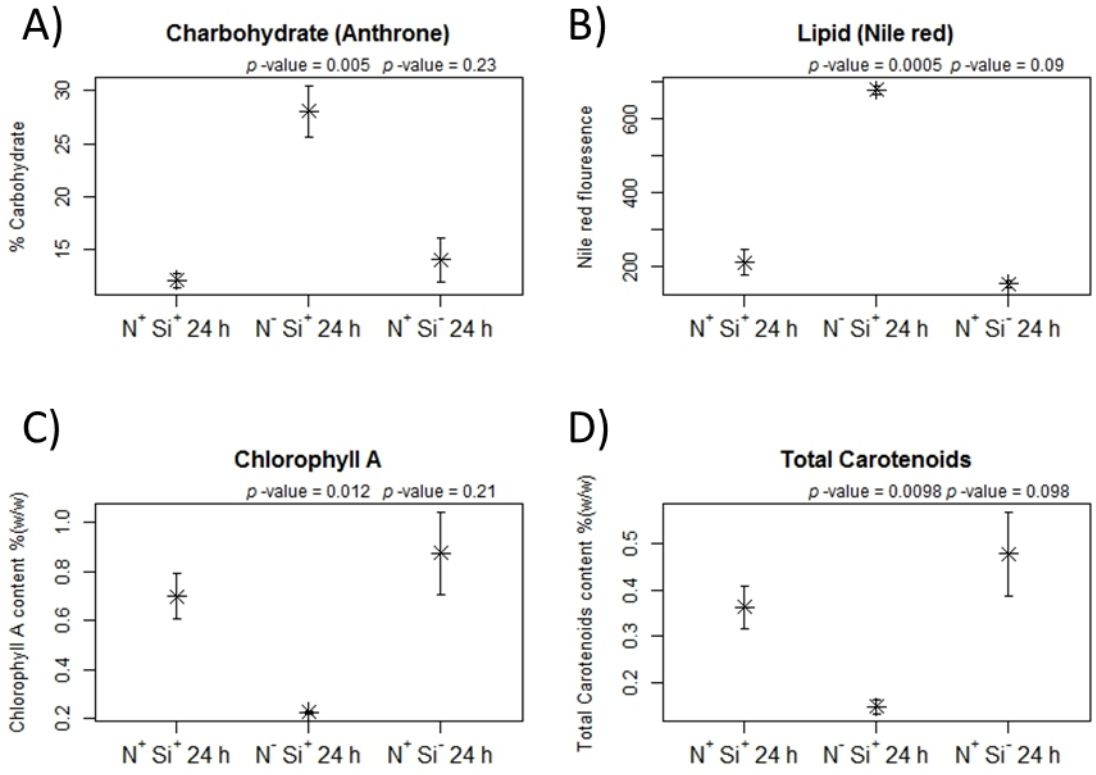


Figure 7.3: Plots showing the biochemical assessment of samples used for proteomic analysis. Samples taken 24h post change of media to test condition. Assessment of A) carbohydrate, B) lipid, C) chlorophyll A and D) total carotenoids are shown. All assessments were based on dry cell weight reconstituted to a fixed biomass concentration. *p*-values are reported indicating the significant difference of the two stress conditions to that of the normal replete condition.

Concurrent with proteomics and biochemical analysis, 1 mL of culture was also prepared for microscopy. Given the saline nature of the medium and the propensity for undissolved compounds in artificial sea water, considerable debris was present in the microscopy imagery. A representative image for each condition is shown in Figure 7.4_{p197}. The nitrogen stressed cells were observed to have reduced pigmentation with reduced chloroplast content, which is in accordance with the observations made for the pigment concentration (Figure $7.3C_{p197}$ and Figure $7.3D_{p197}$) It is suggested that the nitrogen stressed cells had an increase in the number of intracellular vesicles, but without staining their type and intensity is difficult to clarify.

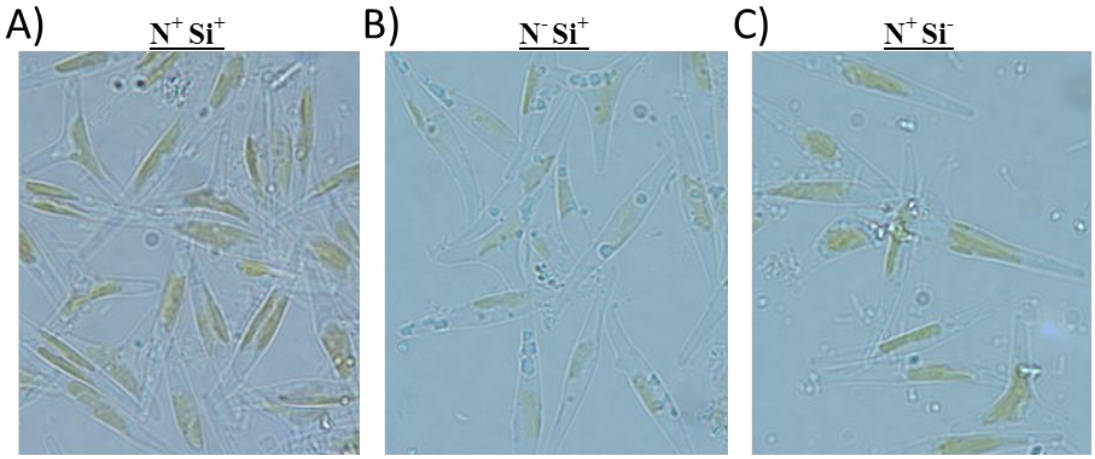


Figure 7.4: Microscope images of *P. tricornutum* 24h after transfer to test conditions. A) Replete culture, B) nitrogen deplete and C) silicon deplete. Taken at 100x magnification.

7.4.3: Overview of proteomic data

Within the proteome dataset 23,544 spectra were matched to peptide and protein without disagreement amongst the six search engines. Each of which were limited to a false discovery rate <0.03 at the peptide level. The derived

PSM list represented 7777 unique sequences matched to 1761 proteins of which 1043 had two or more unique peptides.

To assess sample arrangement, hierarchal clustering (Figure 7.5A$_{p199}$) and principal component analysis (Figure 7.5B$_{p199}$) was performed on the merged PSM list. From this analysis, it can be seen that the nitrogen stress replicates cluster well. The silicon stress samples however do not cluster. Furthermore, the differential between nitrogen stress and non-nitrogen stress is responsible for >80% of the variation between the samples.

The list of PSM(s) was then processed as described to determine significantly identified and changed proteins using the method and script developed and described in Chapter 4 'Proteomic processing'. The full table of changes and significance is available in Appendix 7.2$_{p285}$.

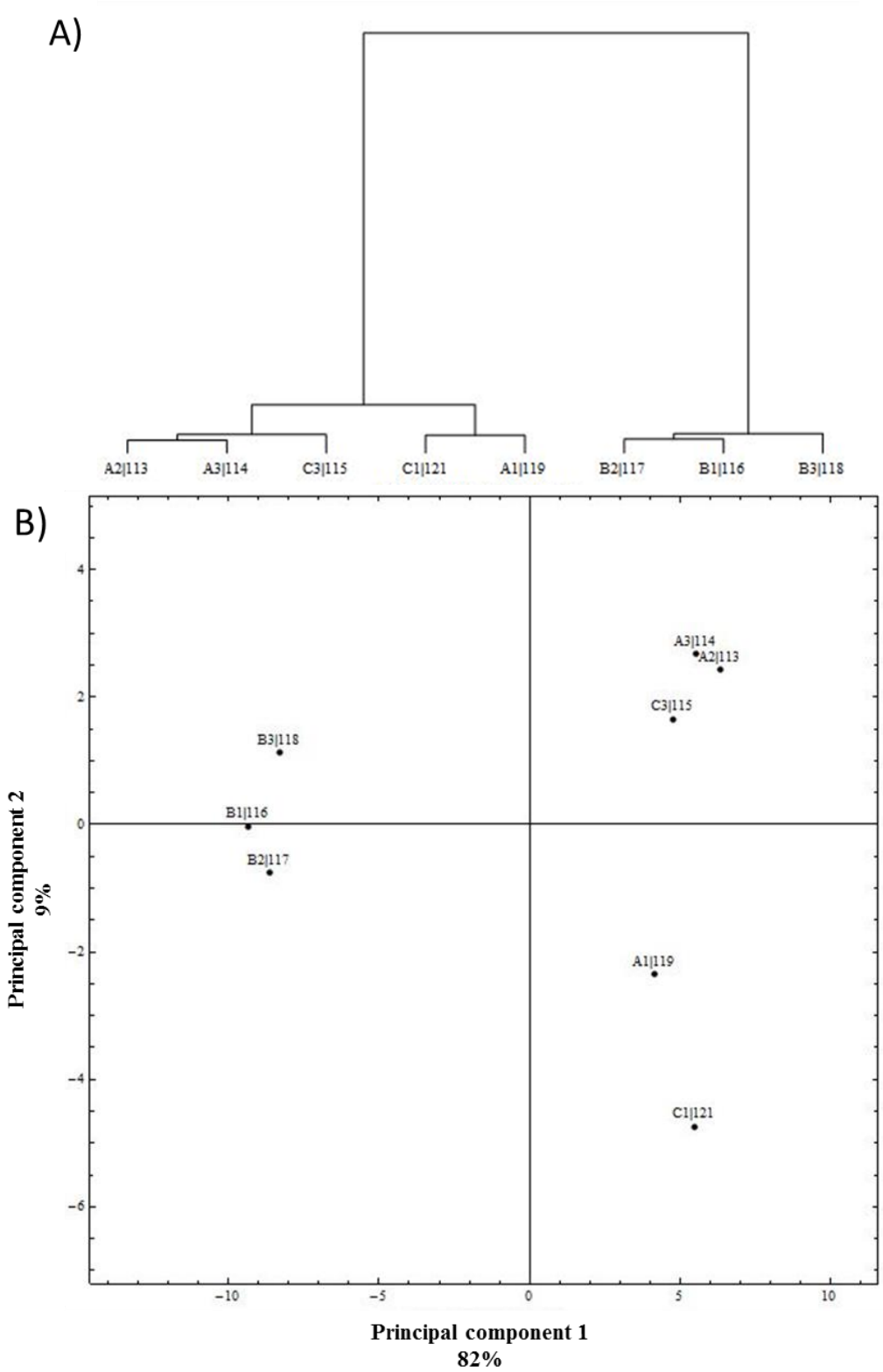


Figure 7.5: A) Dendrogram of the samples association with distance on y axis signifying degree of variation. A, B and C indicate normal, nitrogen stressed and silicon stressed respectively. B) Principal component analysis showing the clustering for samples based on VSN, IC, MC corrected reporter ion intensities. Limited to ≥2 peptide proteins.1, 2, 3 indicate biological replicates. Samples also indicate there label assignment 113:121.

For the nitrogen stress investigation, 605 significant changes were observed (Figure 7.6A$_{p201}$) at a significance level of 0.05. For the silicon stress investigation, 28 significant changes were observed (Figure 7.6B$_{p201}$) at a significance level of 0.05. Given that 1043 proteins were assessed, it would be expected that ~52 would be identified at the selected significance level by random chance.

Given the incorrect clustering of the silicon stressed replicates (Figure 7.6B$_{p201}$) and the insignificant number of changes observed, the silicon stress investigation was not pursued. Furthermore, from the clustering of silicon stressed cultures with the replete cultures, the variation of the proteomic changes under this stress can be described as within the biological variation observed. As such, the assessment of the effect of nitrogen stress can be strengthened by pooling both replete samples and silicon stressed samples into one condition. This pooling resulted in an increase to 645 significant changes (Figure 7.7$_{p202}$).

For biological description two sets of statistically significant proteins were used. 645 identified with a significant difference greater than 0.05 and 498 identified with a significance greater than 0.01. The more stringent list was then utilised for direct hypothesis derivation in Table 7.1$_{p209}$, whilst the less restrictive set was utilised for pathway and gene ontology analysis which requires deduction of hypothesis based on protein clusters rather than individual observations.

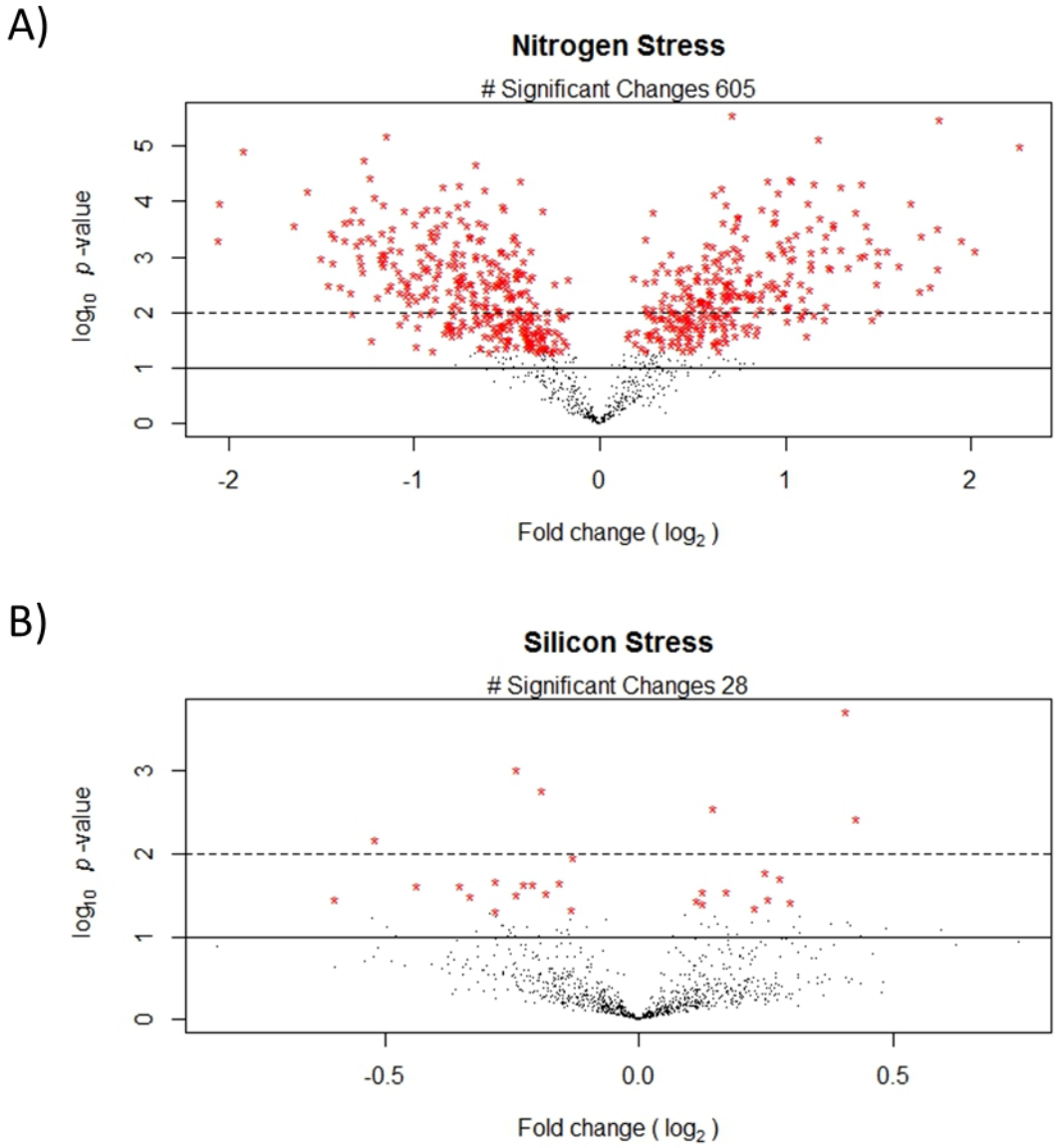


Figure 7.6: Volcano plots showing the changes observed when comparing nitrogen stressed samples to normal (A) and silicon stressed samples to normal (B). Significant changes to a *p*-value of 0.05 are indicated by red *. The *p*-value cut-off of 0.01 and 0.1 are indicated by a dotted and solid line respectively.

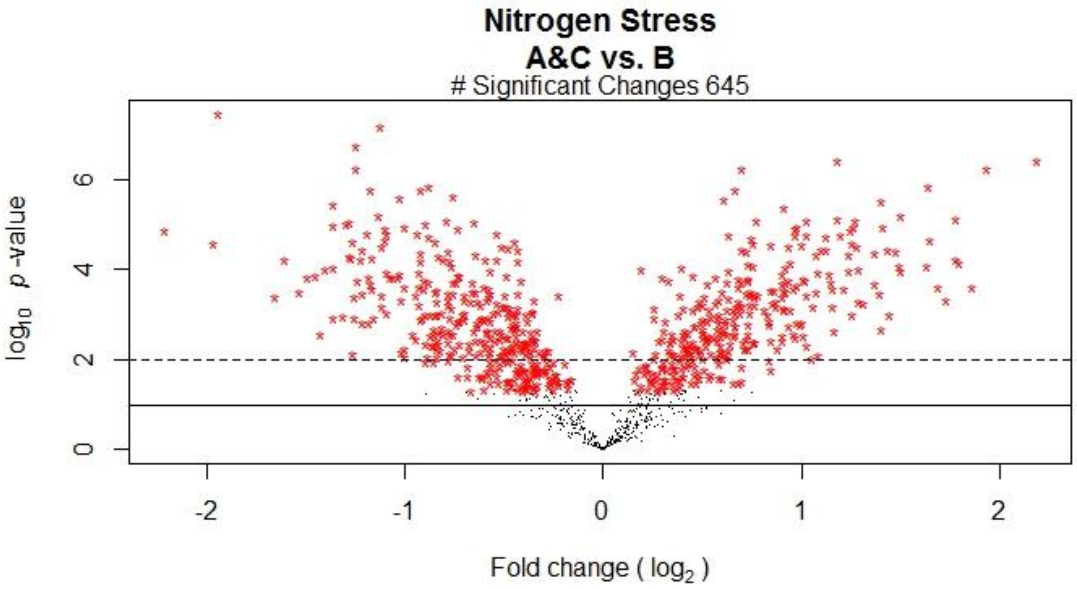


Figure 7.7: Volcano plot of the nitrogen stress comparison given the pooling of normal and silicon stressed samples. Significant changes to a *p*-value of 0.05 are indicated by red *. The *p*-value cut-off of 0.01 and 0.1 are indicated by a dotted and solid line respectively.

7.4.4: Gene ontology assessment

The significant changes ($p < 0.05$) were separated into two pools, those increased and those decreased. Each was then matched to a gene ontology (GO) category. These GO categories were then merged by category and separated to form three charts reporting those identified exclusively as increased (Figure 7.8A_{p204}), decreased (Figure 7.8B_{p204}) and those identified in both pools(Figure 7.8C_{p204}).

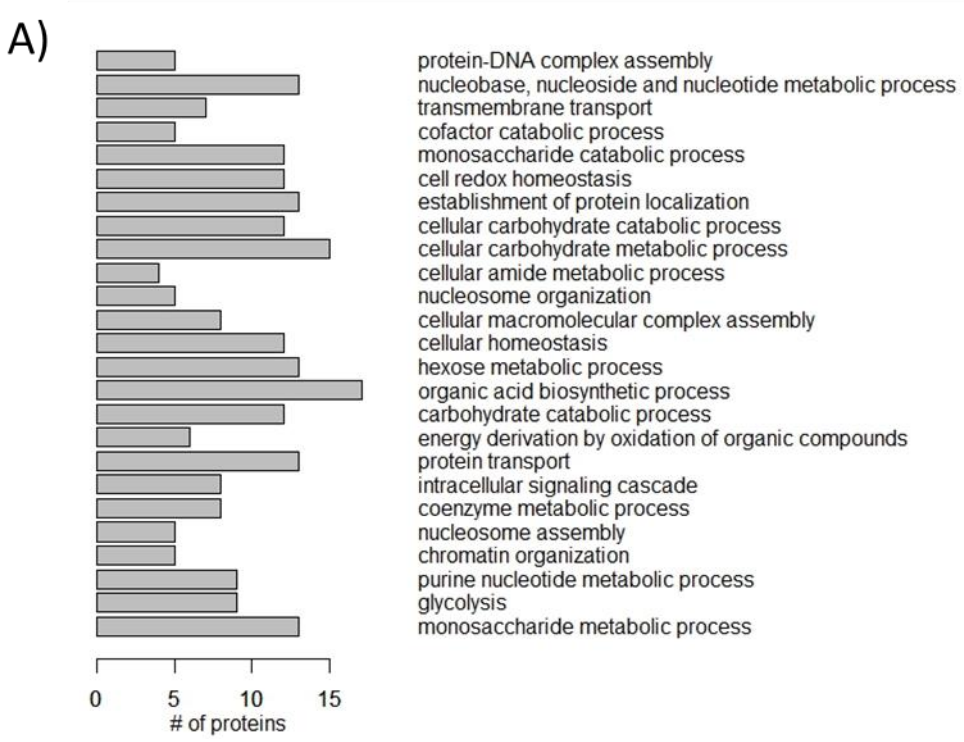
Within these, several groupings and concepts can be identified.

Included in the increased categories are several categories involved in central energy metabolism such as glycolysis. Carbohydrate metabolic and catabolic processes are also identified as increased, suggesting increased activity in carbohydrate processing. Whether this is synthesis, degradation or a combination is not determinable at this point. Protein transport also appears to be increased, suggesting a reorganisation of the cell's proteome complement. Also identified as increased are several categories involved in protein-DNA interactions, possibly suggesting activation of dormancy in the

culture. The increase observed in homeostasis is also indicative of the stressed nature of the culture under nitrogen deprivation.

Within the decreased categories, the decreased abundance of translation and protein metabolic processes is apparent. This is expected given the essential nature of nitrogen for protein production. Also identified as decreased are photosynthesis and processes involved in the production of photosynthetic machinery. This corroborates the reduced pigment observation of Figure $7.3C_{pl97}$ and Figure $7.3D_{pl97}$. Notably, the electron transport chain was identified as being decreased despite the increased glycolysis observed earlier, suggesting that whilst some parts of respiration are increased, others are reduced.

Within those categories identified as both increased and decreased are categories involved in amino acid production, nitrogen compound production and proton transport. These mixed categories suggest a re-arrangement rather than reduction or increase in the particular process the category describes.



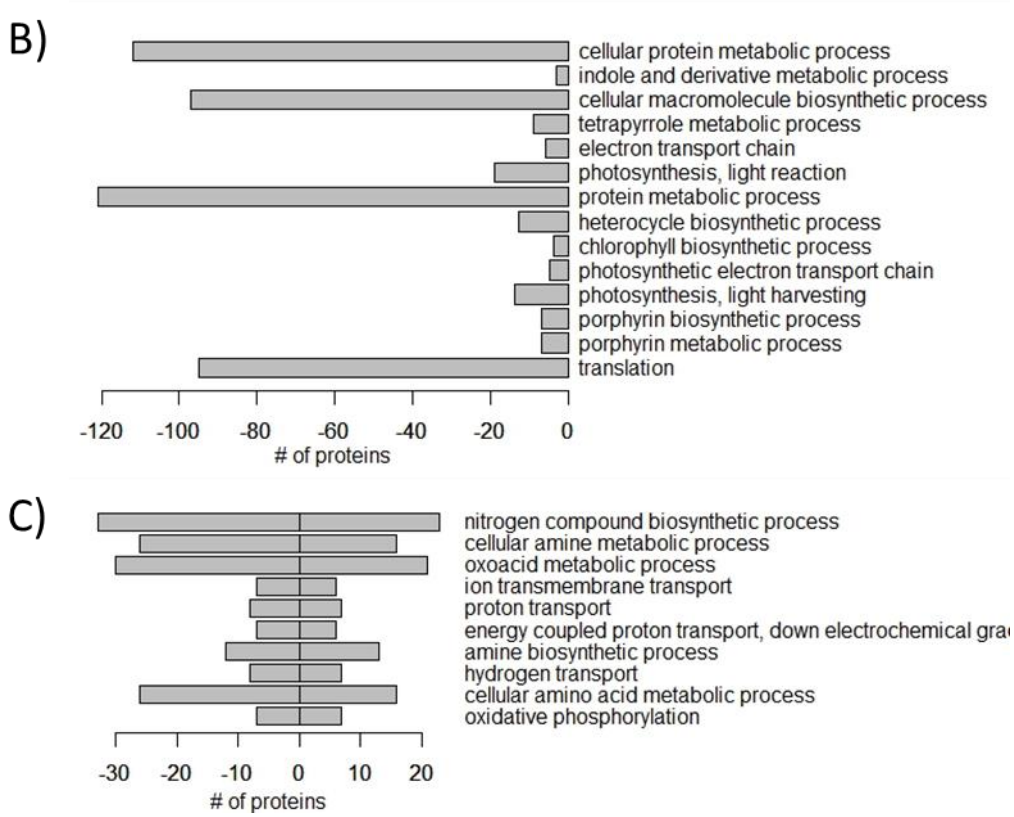


Figure 7.8: Gene Ontology terms assignments indicating the number of protein assignments identified as significantly A) increased, B) decreased and C) both.

7.4.5: KEGG pathway assessment

The significant (p<0.05) changes were then used to colour KEGG maps using clear diagrams with changes identified in blue being those identified as increased and those in red as decreased. The overall map of the metabolism is shown in Figure 7.9$_{p205}$ with several pathway maps being shown grouped by concept in Appendix 7.3$_{p285}$ to Appendix 7.8$_{p297}$.

These pathways show several trends:

Decreased abundance of ribosomal proteins can be identified with considerable coverage (Appendix 7.8$_{p297}$). This agrees with the observation of the gene ontology and is again suggestive of a decrease in protein production associated with nitrogen stress.

Agreeing with the decrease in translation, pathways involved in amino acid production also contain predominantly decreased proteins. Proteins involved in 'phenylalanine, tyrosine and tryptophan biosynthesis', 'Valine,

Leucine and Isoleucine biosynthesis' and 'Aminoacyl-tRNA Biosynthesis' were reduced (Appendix 7.4_{p286}). There was, however, observation of an increase in serine tRNA suggesting that whilst decreased in general some amino acids were increased.

Considering energy production, pathways involved in respiration ('Pentose Phosphate Pathway', 'Citrate Cycle', 'Fructose and Mannose Metabolism' and 'Glycolysis/Gluconeogenesis' (Appendix 7.6_{p292})) were increased. Conversely, those involved in photosynthesis ('Photosynthesis' and 'Porphyrin and Chlorophyll metabolism' (Appendix 7.7_{p295})) decreased in abundance. This increase in central energy metabolism, despite decreases in photosynthetic activity has been demonstrated in other algae[179] and through transcription analysis within *P. tricornutum.*[364]

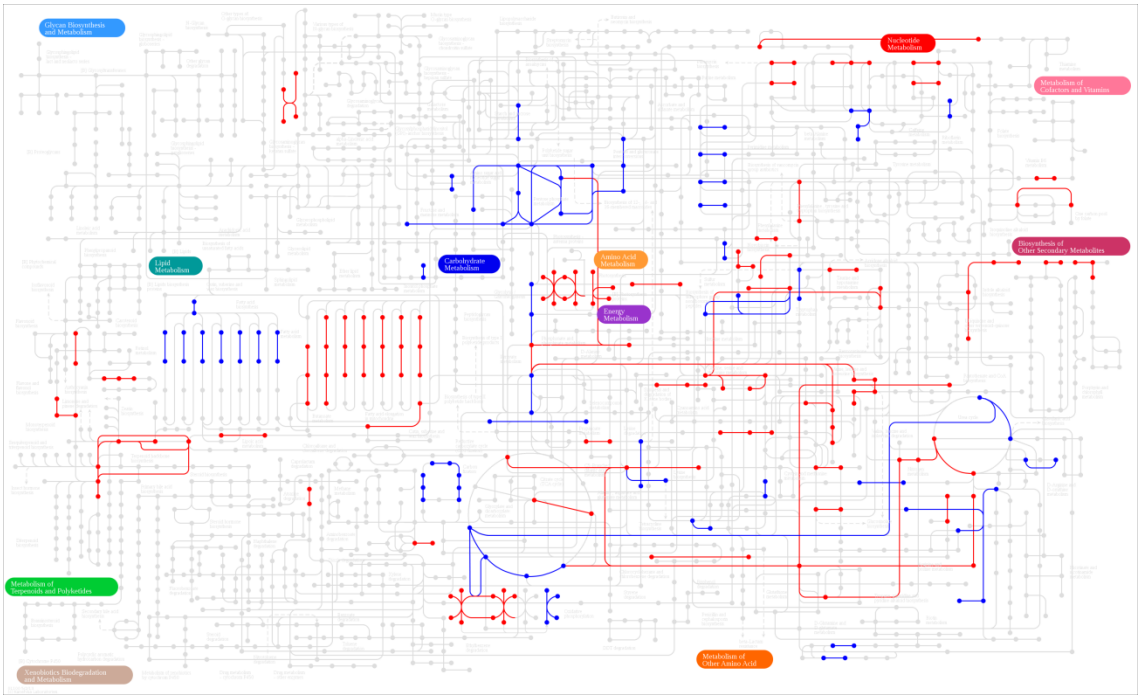


Figure 7.9: Metabolic pathway diagram from KEGG painted with significantly ($p < 0.05$) increased and decreased proteins, shown in blue and red respectively. (Large image is available with appended CD.)

Fatty acid metabolism pathways are reported in Appendix 7.5_{p289}. Increased KEGG pathways include 'biosynthesis of unsaturated fatty acids' and 'fatty acid biosynthesis'. A relative decrease was observed in 'Fatty acid elongation' and 'Fatty acid metabolism'. These agree with the observation of Figure

7.3B$_{p197}$ identifying an increase in lipid content under nitrogen stress. Further, the observations suggest an increase in shorter chain fatty acids.

Not previously realised from GO analysis, pathway analysis suggests an increase in 'Endocytosis' and 'Phagosome' activity under nitrogen stress (Figure 7.10$_{p207}$). Such increases in phagosomal activity has previously been reported in other alga under nitrogen stress. For example in Bihan *et al.*'s proteomic study on *Ostreococcus tauri.*[254] This is suggestive of a scavenging response. With the reduced nitrogen availability, the cells could have increased the intake and processing of extracellular debris and perhaps attempted to consume other organisms such as bacteria to obtain additional nitrogen supplies. Thus, nitrogen stress could be suggested to induce phagotrophy [367,368] describing the engulfing of other smaller organisms for nutrients.[369]

In addition to external nutrient retrieval many of the proteins associated with endocytosis and phagocytosis are reported to be similarly involved in autophagy.[370,371] Transcriptional evidence of a link between nitrogen stress and autophagy induction was previously shown in the chlorophyta *Neochloris.*[372]

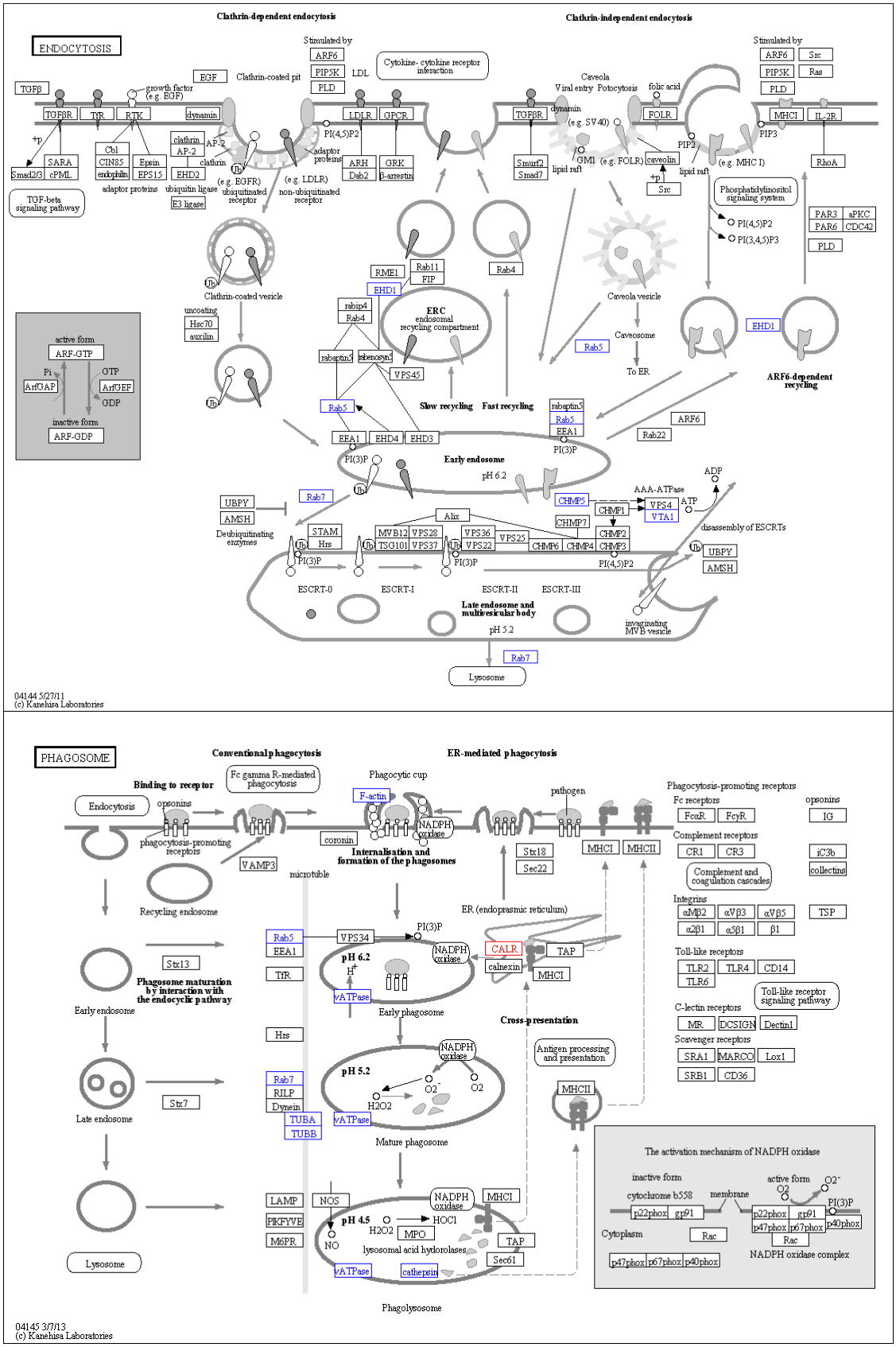


Figure 7.10: KEGG diagrams of A) endocytosis and B) phagocytosis, painted with significantly (p < 0.05) increased and decreased proteins, shown in blue and red respectively.

7.4.6: Highly significant changes ($p < 0.01$)

Given limited annotation available with *P. tricornutum*, it is prudent to investigate the most significant changes individually. Within the dataset, 500 confidently identified proteins (≥2 unique peptides) were significantly ($p < 0.01$) altered. These were matched to protein names using Uniprot. Of the 498, 305 are named as 'Predicted Protein' or 'Predicted protein (Fragment)' with 193 described to some degree.

The *P. tricornutum* proteome is 88.5% described as 'Predicted Protein' or 'Predicted protein (fragment)'. As such, the identification of 61% of the significant changes as unclassified is coherent. The reduction from 88.5 to 61% is likely due to an increase propensity for abundant or commonly identified proteins to be described. Despite the non-characterisation of 9,267 proteins in the Uniprot database, when these proteins are searched in the KEGG database 938 match to pathways and when matched to GO terms in DAVID[288] 2,632 assignments are made. This highlights the need to utilise multiple methods to probe the discovered protein changes as conducted within this investigation.

The 193 identifications with descriptive names were grouped using the protein name and information provided on the Uniprot entry page. These are reported in Table 7.1_{p209}. From these groupings several observations made from the GO and KEGG analysis are further supported. These include decreases in translation, photosynthesis and increase in central energy metabolism. Again, as previously observed, amino acid production appears to be re-organised with some increased and some decreased proteins. Grouping the amino acid production based on their type (e.g. aromatic and hydrophobicity) didn't reveal any consensus.

Table 7.1: Table of all significant (p < 0.01) changes observed omitting Predicted Proteins. Each protein is reported with it Uniprot ID, Descriptive name, Number of unique peptides, the p-value and fold change observed under nitrogen stress. Positive fold changes are shown in bold.

Uniprot ID	Protein Name	# Unique Peptides	*p*-value	Fold Change
	Hydrophilic Amino Acid Synthesis			
B7GEJ6	Acetylornithine aminotransferase	7	3.51E-03	**1.28**
B7G5H9	Aspartokinase	2	4.67E-04	-1.50
B7GBH2	Delta l-pyrroline-5-carboxylate synthetase	12	1.53E-03	-1.35
B7G3A2	Diaminopimelate decarboxylase	6	9.06E-05	**1.31**
	Hydrophobic Amino Acid Synthesis			
B7FUP6	2-isopropylmalate synthase	6	5.28E-04	-1.55
B7FRJ9	3-deoxy-7-phosphoheptulonate synthase	15	1.82E-04	-1.69
B7FT14	Adenosylhomocysteinasee	18	1.77E-04	-2.36
B7G2T9	Carboxy-lyase	3	3.39E-03	-1.27
B7FS76	Chorismate synthasee	2	2.57E-03	**1.27**
B7G117	O-acetylhomoserine	8	6.13E-03	-1.22
	Other Amino Acid Synthesis			
B7FT50	Asparagine synthetasee	4	2.56E-03	**1.53**
B7G5Z8	Glycine decarboxylase p-protein	16	4.83E-04	-1.63
B7FZB0	Synthase of glutamate synthase	18	5.15E-03	**1.33**
	Photosynthesis			
A0T0C9	Apocytochrome f	17	6.20E-04	-1.91
A0T0D1	ATP synthase epsilon chain, chloroplastic	7	1.80E-05	-1.84
A0T0F1	ATP synthase subunit alpha, chloroplastic	40	6.52E-08	-2.18
A0T0E9	ATP synthase subunit b, chloroplastic	5	2.99E-04	-2.89
A0T0E8	ATP synthase subunit b', chloroplastic	2	1.51E-04	-2.80
A0T0D2	ATP synthase subunit beta, chloroplastic	49	5.75E-07	-2.37
A0T0F0	ATP synthase subunit delta, chloroplastic	3	1.49E-04	-2.26
A0T0A3	Cytochrome b559 subunit alpha (PSII reaction center subunit V)	6	1.30E-04	-2.73
B7FSN7	Delta-aminolevulinic acid dehydratase	8	4.56E-04	-1.27
B7G3I6	Fucoxanthin chlorophyll a/c protein, deviant	9	7.46E-03	-1.71
Q41093	Fucoxanthin-chlorophyll a-c binding protein E, chloroplastic	11	1.38E-04	-1.65
A0T0B5	Magnesium-chelatase subunit I	16	4.86E-05	-2.42
B7FZ96	Oxygen-evolving enhancer protein 1	14	1.83E-03	-1.35
A0T0B9	Photosystem I ferredoxin-binding protein	28	3.34E-04	-1.48
A0T0M1	Photosystem I protein F	9	6.56E-03	-1.54
A0T0M6	Photosystem I reaction center subunit XI	4	4.56E-03	-1.78
A0T096	Photosystem II CP43 chlorophyll apoprotein	15	1.09E-03	-2.49
A0T0B2	Photosystem II CP47 chlorophyll apoprotein	17	4.01E-04	-3.15
A0T097	Photosystem II D2 protein	3	2.71E-03	-2.69
A0T0H5	Photosystem II reaction center psb28 protein	9	1.42E-03	**1.78**
A0T0G9	Photosystem Q(B) protein	3	8.84E-03	-1.85
B7FZL9	Phytoene dehydrogenase	2	1.94E-04	-1.67
B7FRW2	Protein fucoxanthin chlorophyl a/c protein	24	5.93E-04	-1.53
B7FQE0	Protein fucoxanthin chlorophyll a/c protein	4	9.20E-03	-1.40
B7FQE1	Protein fucoxanthin chlorophyll a/c protein	7	8.36E-03	-1.63
B7FR60	Protein fucoxanthin chlorophyll a/c protein	4	3.86E-04	**1.87**
B7FRW4	Protein fucoxanthin chlorophyll a/c protein	3	8.07E-03	-1.49
B7FV42	Protein fucoxanthin chlorophyll a/c protein	4	2.37E-03	-1.77
B7G6Y1	Protein fucoxanthin chlorophyll a/c protein	3	4.25E-03	-1.85

B7G955	Protein fucoxanthin chlorophyll a/c protein	3	4.17E-03	-1.72
B7GCV9	Protein fucoxanthin chlorophyll a/c protein	3	2.26E-05	**3.13**
B5Y3F4	Protoporphyrin IX magnesium chelatase, subunit H	6	2.10E-05	-2.14
B7GDU9	Protoporphyrinogen oxidase	2	4.28E-03	-1.65
B7FUT6	Uroporphyrinogen decarboxylase	10	6.46E-06	-2.19
B7FUR6	Violaxanthin deepoxidase	4	1.09E-04	**1.56**
	Carbon Fixation			
Q9TK52	Ribulose bisphosphate carboxylase large chain	38	5.64E-05	-2.33
A0T0E2	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	5	1.19E-05	-1.92
	Energy Metabolism			
B7FXB6	6-phosphogluconate dehydrogenase, decarboxylating	6	1.73E-04	**1.66**
B5Y3C9	Cytochrome b6-f complex iron-sulfur subunit	11	2.19E-03	-1.72
Q8GTB5	Cytochrome c6 (Precursor cytochrome c6)	4	3.69E-03	**1.79**
B5Y578	Cytochrome c6, cytochrome c553	10	3.54E-03	**1.43**
B7FRC1	Cytosolic aldolase	8	1.86E-04	**1.90**
Q9M7R3	Cytosolic glyceraldehyde-3-phosphate dehydrogenase	18	9.11E-05	**1.89**
Q84XB5	Fructose-1,6-bisphosphate aldolase	16	1.30E-03	**1.42**
B7GDK9	Glucose-6-phosphate isomerase	7	4.73E-04	**1.50**
B7G6T5	Glutamine-fructose-6-phosphate transaminase	4	9.12E-04	**1.61**
B7G518	Isocitrate lyase	3	1.16E-04	-1.90
B7FYD8	Kinase adenylate kinase	3	1.26E-03	-1.53
B7G0K7	Ligase succinate-coa ligase	3	2.64E-03	**1.58**
B7G9G3	Lipoamide dehydrogenase	15	1.13E-03	**1.19**
B7FYT9	Malate synthase	6	3.62E-04	-1.41
B7GCG9	PFP pyrophosphate dependent phosphofructokinase	11	2.98E-03	**1.58**
B7GEI2	Phosphoglycerate mutase	4	2.04E-04	**1.87**
B7G492	Phosphomannose mutase	5	2.75E-03	**1.66**
B7GEF2	Plastidic enolase	17	9.30E-03	**1.20**
B7G0M9	Precursor of ATPase ATPase gamma subunit	11	1.82E-04	-1.58
B7FZE1	Precursor of dehydrogenase pyruvate dehydrogenase E1, alpha and beta subunits	24	3.68E-05	**1.62**
F1SXA3	Putative phosphoenolpyruvate carboxykinase	5	7.35E-03	**1.54**
B7FZG7	Pyruvate kinase	6	7.22E-03	**1.40**
Q2TSW8	Pyruvate kinase	3	6.77E-03	-1.25
Q2TSW9	Pyruvate kinase	7	4.01E-03	**1.37**
Q2TSX0	Pyruvate kinase	5	5.05E-03	**1.34**
B5Y5N6	Succinate dehydrogenase flavoprotein	19	7.27E-05	**1.66**
B7GA40	Succinate dehydrogenase iron sulfur protein	5	2.04E-04	**2.59**
B7FUU0	Transketolase	36	9.69E-04	-1.37
B7G5R3	Transketolase	10	1.84E-03	**1.58**
B7FT67	Triosephosphate isomerase	6	2.01E-03	**1.42**
B7G3C1	Triosephosphate isomerase	4	1.12E-03	**1.46**
	Fatty Acid Biosynthesis			
B7G1R8	3-oxoacyl-[acyl-carrier protein	11	8.27E-06	**1.71**
B7GCM0	3-oxoacyl-[acyl-carrier-protein] synthase	10	4.83E-03	**1.51**
B7G7H8	3R-hydroxyacyl-[acyl carrier protein] dehydrase	4	5.98E-03	**1.41**
A0T0F8	Acyl carrier protein	2	3.23E-03	**1.96**
B7FRX6	Acyl carrier protein	3	2.76E-05	**2.00**
B7G3D4	Malonyl-CoA:ACP transacylase	2	1.50E-03	**1.41**

Q2TSX2	Mitochondrial glyceraldehyde-3-phosphate dehydrogenase	2	1.55E-03	**1.53**
E6Y9B3	Stearoyl-ACP desaturase	2	3.10E-03	**1.48**
	Fatty Acid Catabolism			
B5Y4D9	Long chain acyl-CoA synthetase	2	5.83E-03	-1.78
B7FXX6	Long chain acyl-coa synthetase	7	4.09E-03	-1.31
B7FW77	Peroxisomal 2,4-dienoyl-CoA reductase	2	6.61E-03	-2.03
B5Y5R5	Short chain acyl-coenzyme A dehydrogenase	5	2.38E-03	-1.28
	Nucleotide Biosynthesis			
B7FP55	Inosine-5'-monophosphate dehydrogenase	2	8.23E-04	-1.44
B7FPE8	Nucleoside diphosphate kinase 1	6	1.35E-03	**1.25**
B7FR80	Nucleoside diphosphate kinase 3	6	1.75E-05	**1.55**
	Translation			
A0T0J8	30S ribosomal protein S13, chloroplastic	5	5.40E-04	-1.45
A0T0B3	30S ribosomal protein S14, chloroplastic	2	1.22E-04	-2.04
B7FU91	30S ribosomal protein S15	3	4.74E-03	-1.82
Q5D704	30S ribosomal protein S16, chloroplastic	2	8.56E-06	-1.57
A0T0E0	30S ribosomal protein S2, chloroplastic	3	2.78E-03	-1.71
A0T0I5	30S ribosomal protein S3, chloroplastic	4	1.33E-04	-1.66
A0T0J5	30S ribosomal protein S5, chloroplastic	9	2.36E-04	-1.51
A0T0K5	30S ribosomal protein S7, chloroplastic	3	3.41E-03	-1.43
A0T0K2	30S ribosomal protein S9, chloroplastic	2	1.05E-05	-2.56
B7FPA1	40S ribosomal protein S12	7	1.67E-03	-1.66
B7FPM3	40S ribosomal protein S3a	8	2.61E-04	-1.87
B5Y4X4	40S ribosomal protein S6	15	1.58E-05	-2.28
B7FP80	40S ribosomal protein S8	3	8.03E-03	-1.56
A0T0C1	50S ribosomal protein L1, chloroplastic	7	1.34E-03	-1.66
A0T0C2	50S ribosomal protein L11, chloroplastic	6	1.55E-05	-1.45
A0T0C0	50S ribosomal protein L12, chloroplastic	17	3.28E-04	-1.69
A0T0K1	50S ribosomal protein L13	2	1.70E-04	-1.87
A0T0I9	50S ribosomal protein L14, chloroplastic	3	6.48E-04	-1.42
A0T0I6	50S ribosomal protein L16, chloroplastic	4	3.67E-04	-2.17
A0T0C7	50S ribosomal protein L19, chloroplastic	2	2.73E-03	-1.81
A0T0I1	50S ribosomal protein L2, chloroplastic	6	2.55E-06	-2.03
A0T0G3	50S ribosomal protein L21, chloroplastic	2	6.31E-03	-1.60
A0T0I4	50S ribosomal protein L22, chloroplastic	2	2.46E-05	-1.36
A0T0H8	50S ribosomal protein L3, chloroplastic	4	3.54E-04	-1.92
A0T0J3	50S ribosomal protein L6, chloroplastic	2	1.55E-05	-1.91
B7G9G2	60S ribosomal protein L13	7	6.50E-05	-1.34
B7G0R5	60S ribosomal protein L18a	10	4.90E-05	-1.77
B7FTL3	60S ribosomal protein L36	9	1.53E-04	-2.08
B7FUV3	60S ribosomal protein L6	7	1.36E-03	-1.48
E9PAI7	Elongation factor Ts, mitochondrial	6	9.82E-04	-1.36
B7GA11	Elongation factor Tu	10	1.24E-03	-1.62
B7G0T8	Eukaryotic translation initiation factor 3 subunit A	9	9.60E-04	-1.30
B7GCT6	Glutamyl-trna synthase	4	1.21E-03	-1.52
B5Y502	Ribosomal protein L15	6	8.86E-04	-1.62
B7GAA5	Ribosomal protein L19	7	4.02E-04	-1.49
	Protein Processing			
A0T0H6	60 kDa chaperonin, chloroplastic	11	1.73E-04	-1.33
B7FUB7	ER luminal binding protein	33	6.09E-05	-1.45
B7G5I4	Importin subunit alpha	10	2.64E-03	**1.54**
B7GE38	Oligosaccharyl transferase	3	2.31E-03	-1.26
B5Y4H4	Peptidyl-prolyl cis-trans isomerase	9	5.87E-04	**1.61**

B7FQT3	Peptidyl-prolyl cis-trans isomerase	17	4.08E-03	**1.63**
B7FSV6	Peptidyl-prolyl cis-trans isomerase	4	1.67E-06	**1.59**
B7FPA6	Peptidyl-prolyl cis-trans isomerase	2	2.74E-04	**1.88**
B7FZL3	Peptidyl-prolyl cis-trans isomerase	7	1.02E-03	**1.50**
B7G5J3	Peptidyl-prolyl cis-trans isomerase	2	1.88E-04	**1.55**
B7GB02	T-complex protein 1 subunit delta	3	8.59E-04	-1.28
	Proteolysis			
B7FU90	Proteasome subunit alpha type	2	8.35E-03	-1.29
B7G2F7	Regulatory proteasome non-atpase subunit 1	2	3.36E-05	-1.39
B7FY02	Ubiquitin extension protein 3	18	2.98E-04	**1.51**
	Nitrogen Metabolism			
B7G8X8	Aliphatic amidase	2	2.20E-03	**2.24**
B7GEG8	CPS III, carbamoyl-phosphate synthase mitochindrial	39	3.53E-06	-2.57
B7FYS6	Formidase	5	7.27E-06	**2.27**
B7G997	Nitrate reductase	22	5.85E-05	-2.33
B7FZW5	Urea transporter	3	9.77E-04	**2.38**
	Cytoskeleton / Cellular Transport			
B7G5C0	Actin/actin like protein	9	3.94E-03	**1.70**
B7G878	Actin/actin like protein	23	2.41E-03	**1.42**
B7FY56	Coronin	5	3.56E-04	**1.31**
B7FTS7	Det3-like protein	7	7.55E-04	**1.54**
B7FUJ2	Gelosin/severin like protein	6	7.93E-06	**2.42**
	Histone			
B7FR39	Histone H3	8	8.70E-03	**1.29**
B7FX68	Histone H4	12	8.90E-04	-2.13
B7FX66	Histone linker H1	6	6.80E-03	**1.51**
B7FTP2	N-terminal histone linker H1	5	1.58E-03	**1.66**
	Antioxidant			
B7G384	Ascorbate peroxidase	5	7.59E-03	**2.11**
B7GDY5	Glutaredoxin	5	2.14E-04	**1.96**
B7GDI2	Glyoxalase	2	2.85E-04	**1.80**
B7GLJ9	L-ascorbate peroxidase	6	6.76E-03	**1.50**
B7G0L6	Superoxide dismutase	4	1.88E-04	**2.22**
B7FP57	Thioredoxin	2	1.52E-03	**2.00**
B7G0C9	Thioredoxin	5	3.15E-05	**2.43**
B7G0P5	Thioredoxin f	3	9.94E-03	**1.31**
B7G7L6	Thioredoxin h	3	9.83E-05	**2.44**
	Heat Shock Protein			
Q41074	BiP	6	1.44E-06	-1.84
A0T0H7	Chaperone protein dnaK	33	1.37E-04	-1.40
B7FXQ8	Heat shock protein Hsp20	2	1.48E-03	**1.97**
B7GEF7	Heat shock protein Hsp90	11	5.48E-03	-1.42
B7GCE9	Protein heat shock protein	10	2.98E-05	-1.42
	Miscellaneous			
B7G5Y2	14-3-3-like protein	11	5.46E-03	**1.64**
B7FV10	1-hydroxy-2-methyl-2-	7	8.09E-05	-1.70
B7S4B2	Alcohol dehydrogenase	2	4.79E-04	**3.30**
A0T0F2	ATP-dependent zinc metalloprotease FtsH	11	8.81E-04	-1.75
B7FQH4	Calcyclin-binding protein	2	4.43E-03	**1.56**
B7FNY6	Early light induced protein	3	2.58E-03	-1.95
B7FU89	Farnesyltranstransferase	5	1.54E-05	-2.13
B7GB73	FeS assembly protein suf	5	8.88E-04	-1.36
B7FUG8	Glycolate oxidase	10	5.07E-03	-1.50
B7FWY2	Hydroxymethylbilane synthase	22	1.92E-04	-1.84

B7FYL2	Iron starvation induced protein	6	2.49E-05	-3.90
A0T0E5	Iron-sulfur cluster formation ABC transporter ATP-binding subunit	5	6.69E-03	-1.36
B7G6D3	Metacaspase	5	2.67E-05	**1.69**
B7S4C8	Methionine aminopeptidase	2	1.10E-03	-1.39
Q8LKV0	Microsomal cytochrome b5	3	4.64E-03	-1.81
B7FQ72	Mitochondria-targeted chaperonin	58	7.43E-03	-1.30
B7FU88	P2B, P type ATPase	3	8.96E-04	-1.28
B5Y5C8	Short-chain alcohol dehydrogenase with NAD or NADP as acceptor	7	3.40E-04	**1.65**
B5Y3S6	Transaldolase	5	1.27E-05	**1.96**
B7GEF3	Translocator of the inner chloroplast envelope membrane 110k	13	3.64E-04	-1.63

Further, from supporting GO and KEGG pathway observations some new trends can be commented on using the proteins in Table 7.1$_{p209}$.

Given the nature of the stress condition, it is expected that nitrogen scavenging would be strongly promoted within the cell. When observing nitrogen metabolism, increased proteins included aliphatic amidase and formidase, both of which free ammonia from other compounds. Nitrate reductase, responsible for converting nitrates to nitrite, an initial step for nitrate assimilation was decreased. It is this nitrate which is removed to induce stress. This increase in nitrogen scavenging simultaneous to reduction in the nitrogen uptake enzyme nitrate reductase suggests a more active rather than passive response to the stress. This need for scavenging for nitrogen resources is perhaps also demonstrated by the increases in proteasome proteins and earlier observations of endocytosis and phagosome.

Within fatty acid metabolism as suggested by the biochemical analysis (Figure 7.3B$_{p197}$) the production of fatty acids is increased. This is supported by both Table 7.1$_{p209}$ and to some extent by KEGG analysis (Appendix 7.5$_{p289}$). This perhaps contradicts the transcriptomic analysis conducted by Valenzuela *et al.*[365] Such disagreement highlights the inappropriateness of using transcriptomic data to infer proteomic changes. The discord between these findings could suggest a translational control for fatty acid synthesis proteins.

In addition to the increased abundance of fatty acid biosynthesis proteins, there was a decrease in fatty acid catabolic related proteins suggesting a decrease in degradation of fatty acids maybe a driving factor in their accumulation. This is supported by Burrows *et al.*[373] observation of preservation of existing TAGs post nitrogen stress.

Nine proteins with antioxidant properties are observed as increased under nitrogen stress. These suggest a change in reactive oxygen species (ROS) concentration within the cellular environment. Candan and Tarhan[374] suggest that the increase in ROS is a major source of damage to cells under abiotic and biotic stresses in plants. It is therefore suggestive that the increase in antioxidant proteins is a mechanism used by the alga to limit this damage.

Also noted in Table 7.1$_{p209}$ is the ~ 2 fold decrease of both the long and short chain of RuBisCO. RuBisCO conducts the first major step in carbon fixation suggesting carbon fixation is decreased under nitrogen stress. This is logically sound given the reduced photosynthesis observed, but contradicts findings of Yang *et al.*[364] transcriptomic investigation. From[364] the Burrows *et al.*[373] study of de-novo lipid production, it is known that at 72 h post N depletion 60% of TAG's are from post starvation fixation. Hence, it is evident that though there is a reduction in this highly abundant RuBisCO, considerable carbon fixation must still be occurring.

7.4.7: Predicted proteins analysis

Within the 498 significantly identified proteins, 305 were termed predicted proteins. To investigate the most significantly identified (top 10 significant) of these predicted proteins, the sequences for each was searched using the BlastP tool of Uniprot. The top match along with the percentage identity is reported in Table 7.2$_{p215}$.

From these top ten predicted proteins several observations made previously by GO, KEGG and Table 7.1$_{p209}$ are supported. A notable addition is B7GOY4,

which matched the ammonium transporter AMT2a (Q5EXJ7) which further demonstrates the cell's attempt to source nitrogen.

This demonstrates that within *P. tricornutum*, effort is still required to increase the level of descriptive information linked to proteins within the database. However, within this investigation which aims to identify the trend changes such matching would derive only limited further understanding. Further, due to the reduced confidence of the true function of the protein when based on such an inference this may lead to increased erroneous conclusions.

Table 7.2: Top ten most significantly identified protein described as 'Predicted Protein'. For each the highest non-predicted blastp search result is reported.

P. tricornutum UniProt AC	*p*-value	Fold Change	Identity	Best Match UniProt AC	Description of matched Protein	Matched Species
B7G4W8	3.51E-08	-3.8378	82%	B8C8G5	Thiamin biosynthesis protein	*Thalassiosira pseudonana*
B7FP19	1.79E-07	-2.3737	70%	D7FRX5	Geranylgeranyl reductase; geranylgeranyl diph...	*Ectocarpus siliculosus*
B7FTY3	3.63E-07	2.2649				
B7G0Y4	3.72E-07	4.5389	71%	Q5EXJ7	Ammonium transporter AMT2a	*Cylindrotheca fusiformis*
B7S3I6	5.49E-07	1.62032	76%	D8LP22	CDGSH iron sulfur domain-containing protein	*Ectocarpus siliculosus*
B7G350	5.61E-07	3.80656				
B7GDW4	1.45E-06	3.11038				
B7FTR6	1.69E-06	-2.2486	81%	B5YP04	Acyl-CoA dehydrogenase-like protein	*Thalassiosira pseudonana*
B7G5I7	1.69E-06	-1.8885	76%	B8BY08	RL17, ribosomal protein 17-like 60S large rib...	*Thalassiosira pseudonana*
B7GAG0	2.28E-06	-1.6858	83%	B5YMU6	RL18, ribosomal protein 18, 60S large ribosom...	*Thalassiosira pseudonana*

7.5: Conclusions

From the biochemical analysis, it is evident that nitrogen stress increases energy storage molecules in *P. tricornutum*. These increases are coupled with a decrease in photosynthetic pigments. Through use of an iTRAQ methodology, 1,043 proteins were confidently identified, of which 645 were shown to be significantly regulated under nitrogen stress. This represents a 17 fold increase of previous nitrogen stress assessments of *P. tricornutum*,

and as such provides greater understanding of the effects of nitrogen stress in this model diatom species.

The extent to which the proteome changes in response to nitrogen stress has been demonstrated at >60%, with over 60% of the confidently identified proteins being significantly changed ($p < 0.05$) in abundance. Rather than a consistent reduction across all proteins, a controlled up and down regulation was observed.

Several patterns of response have been identified within the proteome. These include increases in central energy pathways, whilst reduced photosynthesis, reduction in translation and evidence of nitrogen reassignment both for internal stores and exogenous sources by phagocytosis and endocytosis.

Whilst the biochemical analysis of Figure $7.3A_{p197}$ and Figure $7.3B_{p197}$ demonstrates increases in both carbohydrate and lipid content, the proteomics results show a distinction between the pathways for these two storage resources. The proteomic investigation suggests active production and consumption of the carbohydrate compounds. Conversely for lipids, anabolism was increased whilst catabolism was decreased.

7.6: Future work

Within this investigation the silicon stress comparison was abandoned. Further investigation into the literature suggests that one possible cause of this was the glass ware used. Glass contains ~75% silica and the low requirement for silica or *P. tricornutum* could perhaps leach or erode from the glass. Though silicon stress studies have been conducted in glass[359] they are also conducted in polycarbonate[375,376] to eliminate such issues.

The high degree of predicted proteins identified during this investigation suggests investigations in *P. tricornutum* would be aided by an increase in proteome annotation. As demonstrated by the BlastP search summarised in

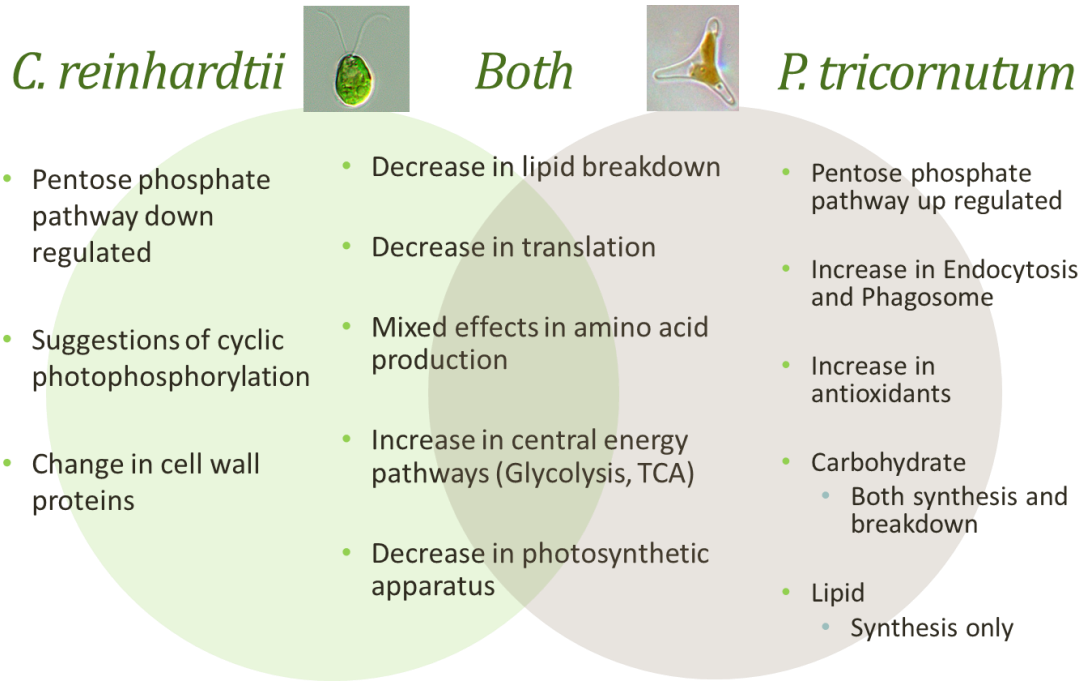
Table 7.2$_{p215}$, this could be achieved in a large part by automated annotation algorithms based on homology. Alternatively annotation could be gained based on multiple investigations describing expression changes of culture under pertubations.[355]

This investigation reports an example of the nitrogen trigger from the Bacillariophyceae group of algae. It adds to previous studies in Chlorophyta. However, these are not the only major branches of microalgae. For a more complete observation of the nitrogen trigger in microalgae, we would require observations from other species as well as other classes such as Rhodophyta and Heterokontophyta.

This investigation was conducted on the effect of nitrogen stress. Further work could therefore focus on other stresses such as phosphorous, temperature and irradiance. These may then be assessed for their effect on the compositional changes. Changes thus seen under nitrogen stress can thus be identified as common to stress conditions or common to stress and compositional change induction.

Within this investigation, 605 proteins were identified as being significantly regulated under nitrogen stress. Many of these are therefore candidates for further investigation of lipid induced nitrogen stress within the species.

Chapter 8: Discussion with comparison of nitrogen stress



Chapter 8 – Thesis Discussion with Comparison of Nitrogen Stress

8.1: Proteomics data pipeline

Within Chapter 4 an assessment of the proteomic data pipeline was conducted. This identified ProteinPilot as the most successful search engine given a limit of 0.03 at the peptide FDR level. During this investigation the advantages of using a combinatorial approach were discussed. Such a combination improves both identification rates and accuracy of matches as questionable matches are more prone to exclusion by disagreement than less questionable matches. A merging method was derived which could be utilised for the subsequent investigations within this thesis. In such a fashion this chapter is directly related to the other three in the way it improved identifications made within each.

8.2: Carbon source effect

The carbon source investigation highlights the importance of relating the laboratory conditions to the 'real world' conditions. Identified in this study was the effect that carbon availability had on the proteome. This study's influence on the others is perhaps best demonstrated by the caution it suggest on relating nitrogen stress in these two species. For the nitrogen stress investigation *C. reinhardtii* was grown mixotrophically whilst *P. tricornutum* was grown photoautotrophically. However, whilst caution is required, the degree of change in the proteome wasn't as dramatic as that of nitrogen stress and such comparative observations should still have a level of significance.

8.3: Nitrogen stress comparison

8.3.1: Introduction

Within this thesis two species were investigated for the effect of nitrogen stress induced lipid accumulation. A comparison noting similarities and differences between these two species is desirable. It is important however to recognise the limitation of such a comparison.

Firstly, the lack of significant observation is not the same as identification of a lack of change. Due to the nature of proteomic analysis there are multiple observations which though identified confidently and found to be significantly different in one investigation will not be confident in all or significant in a repeat of that investigation. The reliability is however improved through observation of trend and pathway changes that can be interpreted from multiple related protein observations. Thus comparison is best conducted on pathway or group observations.

Secondly, these investigations were conducted at different time points of nitrogen stress. The time points of this investigation were selected due to their alignment with active lipid accumulation. As such it is more appropriate to compare these different time points than identical time points, which would likely be in a different state of stress progression. This is not to say that the duration of stress may still play a significant factor and as such further temporal investigation would be needed to assess the consistency over stress duration.

To compare the two investigations two techniques were employed using the GO categories (Table 8.1$_{p224}$) and KEGG pathways (Figure 8.1$_{p228}$)

8.3.2: Similarities

To compare the observations from each investigation in the two species, under nitrogen stress the GO categories observed from each were compared. The number of assigned proteins within each category was obtained for each

species and direction of regulation. These four category lists were then merged by category term and reported in Table 8.1$_{p224}$. Colour coding indicates exclusively up, down or mixed identification for each species. Blanks indicate non-identification for that category within that species.

Taking the GO comparison as a whole it can be seen that the similarity between the two species was greater than any differences. This is demonstrated by the lack of any category being contradictory between the two, though some categories which were exclusively up or down in one species did demonstrate a mixed result in the other species.

The observed decreases in photosynthetic pigments and repeated observation of reduction in photosynthetic proteins it could be suggested that reduced photosynthesis is a common response of both species. This has been previously noted in other investigations[332,377] but progression of this loss may be more controlled rather than a consistent reduction given the increased abundance of some components in *C. reinhardtii* suggesting cyclic photophosphorylation. This observation based on decreased abundance of photosystem II with increased abundance of photosystem I, has since been further supported in *Synechocystis* by Huang *et al.*[378] Further, from the GO analysis (Table 8.1$_{p224}$), the dark reactions can be identified as decreased in *C. reinhardtii* whilst the light reactions associated GO terms were decreased within *P. tricornutum*.

This controlled regulation was further noted within amino acid metabolism. Within both species a mixed response was observed. The arrangement of this regulation is difficult to conclude based on proteomic analysis alone, due to the limited coverage of such pathways in these general investigations. However, the suggestion of consistency within the aromatic and non-aromatic amino acids has since been supported by Courant *et al.*[379] metabolic study.

Given the reduction of nitrogen it is logical that the processes which synthesis nitrogen compounds would be down regulated. This is most evident in the decreased abundance of translational machinery with 47 and 95 proteins identified to this GO category during nitrogen stress of C. *reinhardtii* and *P. tricornutum* respectively. Within both species, 'cellular protein metabolic processes' and 'heterocycle biosynthetic processes' are consistently decreased. Within both species there is a mixed response, as within 'Nitrogen compound biosynthetic processes' mixed results were observed, suggesting the cells were promoting the synthesis of certain compounds to the detriment of others. This is evidence of an evolved controlled response to nitrogen stress.

A theme commonly identified in both species is the increased interaction between DNA and proteins. This is manifested by increases in 'Chromatin' 'Nucleosomes assembly', 'Nucleosome organization' and 'Protein DNA complexes' GO terms (Table 8.1$_{p224}$). This observation could be a side effect of general decrease in protein content. In such a theory, the retention of DNA interacting proteins as a per-cell content could remain the same but given a relative to total protein content (as within these investigation) appear to be increasing. Beyond this however, the induction of cell gametogenesis by nitrogen[348] stress in *C. reinhardtii* could also explain such observations. This induction of cell gametogenesis has not been previously noted in *P. tricornutum*. An alternative to both these is an induction of dormancy. Such nitrogen induced dormancy has been observed in yeast.[380] Within algal culture the imitation of dormancy could be a survival mechanism, perhaps linked to gametogenesis in *C. reinhardtii* that in effect suspends cell activity until return to favourable conditions amenable to proliferation. This would have an evolutionary benefit to the organism allowing survival during times of prolonged nitrogen deprivation.

Table 8.1: Table showing the GO terms assigned to significant changes identified in each investigation. The Category as well as the number of up or down regulated proteins observed in each investigation is indicated. Colour coding is used to signal if observed changes were decreased abundance (red) up regulation (blue) or a mix of both (yellow).

GO category	*C. reinhardtii*	*P. tricornutum*
amine biosynthetic process	-6/8	-12/13
aromatic amino acid family metabolic process	-3	
aromatic compound biosynthetic process	-5	
carbohydrate biosynthetic process	-4	
carbohydrate catabolic process	8	12
cation transport	8	
cell redox homeostasis		12
cellular amide metabolic process		4
cellular amine metabolic process	-13/10	-26/16
cellular amino acid metabolic process	-13/9	-26/16
cellular carbohydrate catabolic process	8	12
cellular carbohydrate metabolic process	11	15
cellular homeostasis		12
cellular macromolecular complex assembly	34	8
cellular macromolecule biosynthetic process	-48	-97
cellular protein metabolic process	-55	-112
chlorophyll biosynthetic process	-4	-4
chromatin organization	34	5
coenzyme metabolic process	7	8
cofactor biosynthetic process	-8	
cofactor catabolic process	6	5
electron transport chain		-6
energy coupled proton transport, down electrochemical gradient	7	-7/6
energy derivation by oxidation of organic compounds	8	6
establishment of protein localization		13
glycolysis	6	9
heterocycle biosynthetic process	-12	-13
hexose metabolic process	9	13
hydrogen transport	8	-8/7
indole and derivative metabolic process		-3
intracellular signaling cascade		8
ion transmembrane transport	7	-7/6
monosaccharide catabolic process	7	12
monosaccharide metabolic process	9	13
nitrogen compound biosynthetic process	-22/17	-33/23
nucleobase, nucleoside and nucleotide metabolic process		13
nucleosome assembly	34	5
nucleosome organization	34	5
organic acid biosynthetic process	7	17
oxidative phosphorylation	8	-7/7
oxoacid metabolic process	-18	-30/21
photosynthesis, dark reaction	-3	
photosynthesis, light harvesting		-14
photosynthesis, light reaction		-19
photosynthetic electron transport chain		-5
porphyrin biosynthetic process	-8	-7
porphyrin metabolic process	-8	-7
protein metabolic process	-57	-121
protein transport		13
protein-DNA complex assembly	34	5
proton transport	8	-8/7
purine nucleotide metabolic process	7	9
reductive pentose-phosphate cycle	-3	
tetrapyrrole metabolic process	-10	-9
translation	-47	-95
transmembrane transport	7	7

Within both species, despite the decrease in photosynthesis an increase in central energy metabolism was noted. Within *C. reinhardtii*, where cells were cultivated mixotrophically, this could be suggested to be a redirection to heterotrophic rather than photo autotrophic growth. However, within the *P. tricornutum* investigation, cultures were grown photoautotrophically. A better interpretation of this increased abundance could be that it facilitates re-direction of energy stores from one source to another. This could be (as suggested by Msanne *et al.*[377]) to re-direct energy produced from the breakdown of proteins. It could also be to facilitate a re-direction of some lipid classes in favour of others, as suggested by Burrows *et al.*[361] Finally, as suggested in this thesis it could be to facilitate the re-direction of energy stored in carbohydrates, hence the observation of both increased carbohydrate metabolism and catabolism (Table 8.1$_{p224}$). It is likely however that all three of these re-directions occur during the nitrogen starvation process.

The understanding of lipid processes within the cell from the various investigations within this thesis cannot be described as extensive. This is perhaps due in part to the fact that predominantly only the soluble proteome has been characterised in the investigations conducted. Given their association with lipids it is not uncommon to find lipid metabolic proteins within the in-soluble proteome fraction, which were not specifically analysed here. Further, the broader nature of our approach precluded targeting lipid metabolic proteins and thus their identification within these studies is sporadic. Despite this limitation, some common trends could be observed and interpreted. The regulation of biosynthetic proteins appears to be mixed with some significant increased abundance increased abundance observed in addition to some decreased abundance . This has previously been noted by Msanne *et al.*[377] in *C. reinhardtii.* The study into *de novo* lipid synthesis by Burrows *et al.*[373] would however contradict this lack of increased synthesis. This could perhaps therefore suggest an increase in the

pathway without corresponding increases in pathway proteins. More notable within these investigations was the down –regulation of lipid catabolic processes. This suggests a significant cause of lipid accumulation may be the decreased degradation rather than the increased production.

8.3.3: Differences

As discussed in the introduction, several observations were made in one species that was not sufficiently captured in the other. However, given the nature of the proteomic investigation the lack of observation does not necessarily imply a contradictory observation, unless multiple corroborating proteomic observations are noticeable by the absence in the other investigation. One such situation might be suggested in Table 8.1_{p224} regarding 'Chromatin' 'Nucleosomes assembly', 'Nucleosome organization' and 'Protein DNA complexes'. Though, as discussed above, these were observed as increased in both species seven times as many proteins were identified as increased in *C. reinhardtii* than in *P. tricornutum.* However, this is possibly due to the increased duration of stress at which *C. reinhardtii* (39h) was assessed compared to *P. tricornutum* (24h).

Another contradictory observation is the increase in antioxidant stress observed in *P. tricornutum*. This is seen in the GO analysis Table 8.1_{p224} as well as by comparison of Table 5.2_{p142} and Table 7.1_{p209} from Chapter 5 and Chapter 7 respectively.

When considering the contrasts between *C. reinhardtii* and *P. tricornutum*, two related KEGG pathways stand out as contrasting. These are the 'Pentose phosphate pathway' and the 'Reductive pentose phosphate cycle' otherwise known as the Calvin cycle. Both of these are highlighted in Figure 8.1_{p228} with the specific KEGG 'Pentose phosphate pathway' reported in Appendix 8.1_{p298}. These pathways can be considered mirror images of each other.[381] Whilst the Calvin cycle uses NADPH and CO_2 to produce glyceraldehyde-3-phosphate which then is used to make glucose, Pentose phosphate pathway converts

glucose to CO_2 to create NADPH and ribulose 5 phosphate. These two pathways combine with the light reactions to drive energy from photosynthesis to fatty acids. Its increased abundance in *P. tricornutum* is therefore coherent with the increases in lipid observed. It's decreased abundance in *C. reinhardtii* suggests lipid increases must be driven by either an alternate energy source such as the heterotrophic conversion of acetate or re-tasking of other energy sources in the cell or a reduction in lipid degradation rather than an increase in lipid synthesis.

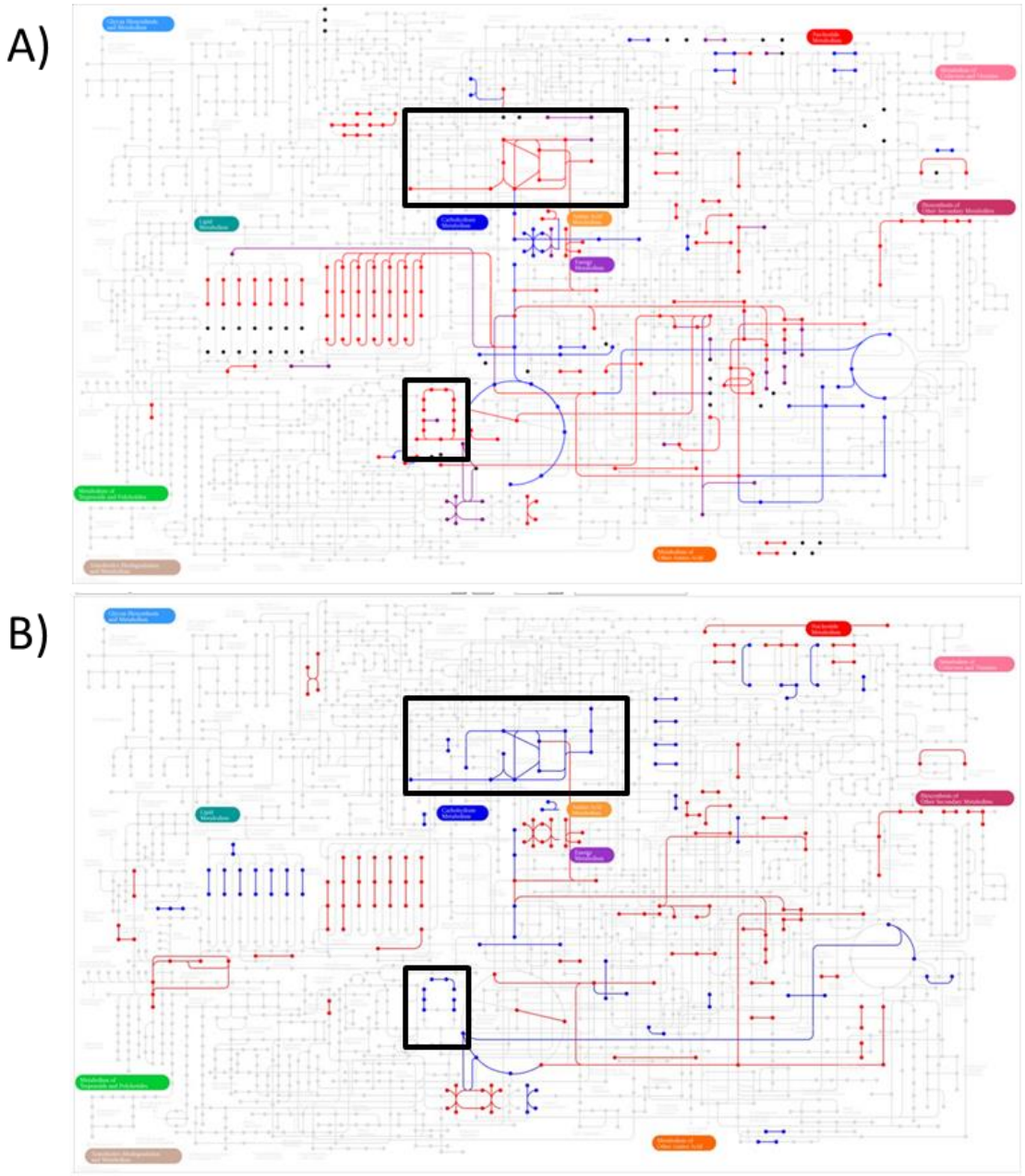


Figure 8.1: KEGG's metabolic map showing identified up and down (blue and red respectively) regulation during nitrogen stress of A) *C. reinhardtii* and B) *P. tricornutum*. To highlight the divergence between the two species, the Calvin cycle and Pentose phosphate pathway are ringed on each with a black outline.

Chapter 9: Conclusions

Chapter 9: Conclusions and Further Work

9.1: Thesis conclusions

The overall aim of this thesis was "To increase understanding of microalgae and their potential use as a biofuel feedstock through the use of quantitative proteomics." This was subdivided into three objectives each being discussed below.

Within Chapter 4, Objective 1 was tackled "To optimise the processing of mass spectrometry data to maximise proteomic information". As a result of this investigation the processing of proteomic datasets was improved by 17.2% through rational combination of 6 available search engines. Further, due to the removal of conflicting PSM matches the accuracy of the PSMs was improved. This provided a reusable R based script which could be utilised through-out this thesis for the subsequent proteomic investigations. It further demonstrated the value of combining multiple search engines in an inclusive but conflict removed fashion. With the insight gained form this processing the format of a standardised PSM and the processing of a merging algorithm were developed and subsequently utilised.

Objective 2 was "To assess the proteomic changes in *C. reinhardtii* when grown photoautotrophically and mixotrophically." This was driven by the use of mixotrophic culture for *C. reinhardtii* in the laboratory to represent what will likely be photoautotrophic growth when at productive scale. In Chapter 6 four comparable conditions regarding carbon source were investigated. These were the photoautotrophic and mixotrophic, as mentioned above, as well as CO_2 supplementation in each to identify the difference due to carbon source and carbon availability. Within this investigation, 995 proteins were identified (≥2 peptides) allowing for several comparisons within the 4 conditions. The results of this investigations showed that carbon availability had significant impact on the proteome.

Whether this source was supplementation by CO_2 or acetate did not have a significant impact.

Objective 3 was "To assess the proteomic changes in two divergent microalgae species under nitrogen induced lipid accumulation." This was conducted as detailed in Chapter 5 in the Chlorophyta *C. reinhardtii* and as elaborated in Chapter 7 in the diatom *P. tricornutum.* Several observations from both investigations were compared in Chapter 8. These investigations contributed significantly to the proteomics based information of nitrogen stress in microalgae.

The regulation of lipid increase through nitrogen stress could have been suggested to have been due to a flow of energy to storage when energy can no longer be utilised for propagation. The proteomic findings, from this thesis, contradict such a suggestion. Many essential components of energy production are deteriorated by the lack of nitrogen availability. As such the energy productiveness of the cell is co-currently damaged by nitrogen stress as lipid is accumulated. It is therefore suggested that a targeted approach that induced lipid accumulation without the negative side effect of a crude biochemical manipulation has potential to provide superior lipid production rates.

Referring back to the overall aim of this thesis, "To increase understanding of microalgae and their potential use as a biofuel feedstock through the use of quantitative proteomics." there are two main paths which this could have been achieved.

One method by which this investigation could have increased the potential of microalgae for biofuels is by identifying targets of manipulation. However, the ability of proteomics, at the scales conducted in this thesis, to provide specific targets for manipulation is limited. Even in a highly studied organism the function and pathway involvement of much of the proteome is limited. In addition, the key proteins whose regulation affects the productive

rate of such pathways is even further limited. This lack of description is even more prominent in even the most studied alga *C. reinhardtii*. Further, current proteomics investigations, given their nature, are only able to provide a limited coverage. Whether the key regulatory proteins are within this coverage is indeterminable. Hence though targets for manipulation can be derived from this investigation they remain subjective and need to be verified with a more targeted approach.

The other method to increase microalgae's potential is to improve understanding. Given the functional nature of proteins, investigations at this 'omic' level provided the greatest cellular insight. Through this thesis proteomic level understanding of two variables were further advanced, the carbon source and nitrogen availability. The latter and more industrially relevant of which was conducted in two divergent algal models. Through these investigations and their respective findings, as well as the comparison of Chapter 8 the aim of the thesis has been accomplished.

9.2: Further work

As discussed within the further work section of Chapter 4, the field of proteomics would benefit by implementation of a combinatorial approach to search engines. Though there is some software available to conduct such mergers at present they typically require a level of computer or computer language knowledge which is beyond that of common biological users. As presented in Chapter 4, a robust user friendly tool for the merger of proteomic search engine results, is desirable for the field of proteomics. Such a tool should be designed for wider uptake by the entire field, something currently lacking for those tools currently available. The requires methodology for such a tool is the essence of Appendices 4.1 and 4.2 but, would require r-working to support a more simplistic user interface.

Within the carbon source investigation, of note was the difficulty in identifying statistically significant changes at the replicate level used. Despite this difficulty some observations were made. This suggests that a larger study is called for with more replicates and more conditions relating to the degree of carbon availability.

As discussed in the literature review 'omic' investigations are best used when they are across multiple ‘omic’ levels. Whilst within this investigation we have demonstrated insights from the proteomic level other investigations at the metabolomics and transcriptomic and post translational modification level would be of benefit to complement the proteomics conducted here. Further, the conditions investigated here are not the only conditions of interest and many more could be useful to increase the knowledge of algae.

As stated in the conclusions with our global approach to proteomics there is difficulty in identifying which proteins could be potential targets of genetic manipulation. This could be improved by either organelle restricted investigations such as the nucleus. Such an investigation would plausibly

identify transcription factor changes which would then be ideal targets for manipulation.

Knock out methods such as RNA interference (RNAi) could be a potential method to assess the significant changes identified here. Using the list of highly significant changes, including those currently without description, a collection of RNAi molecules could be derived. This could then be transfected into culture, perhaps within a 96 well plate. Once allowed to incubate this plate could be replicated, with some replicated culture plates being used to assess growth curves of the knock downs and others being used for Nile red assessment as conducted in a 96 well plate within this investigation.

Within Chapter 8 similarities and differences between the two species under nitrogen stress were discussed. There are numerous further investigations which could be proposed based on these investigations.

From Chapter 8 the suggestion that reduced lipid catabolism may be a key mode of action for increased lipid is of interest. Further, investigations to identify the validity of this suggestion could involve the tracking of carbon flux through the alga. By switching the feed to an isotopically heavy source upon nitrogen starvation the proportion of each carbon compound freshly synthesised rather than pre-existing or re-allocated could be determined.

Within Chapter 8 the increased condensation of chromatin was suggested. This could be tested relatively easily using a commercial dye such as Nuclear-ID (Enzo Life Sciences, Exeter, U.K.). The fluorescence dye responds more intensely when intercalated in condensed chromatin than non-condensed chromatin. This degree of condensation of DNA is a common suggestion of dormancy.

A contradicting observation between the *C. reinhardtii* and *P. tricornutum* discussed in Chapter 8 was the levels of antioxidants within each culture.

With the availability of measurement protocols and kits for reactive oxygen species (ROS) such as the DCFDA-Cellular Reactive oxygen detection kit (abcam, Cambridge, U.K.), the levels of antioxidants within the cell can be tracked. Though not describing the reason for the increase or decrease in ROS it would support proteomic observation of increased antioxidant as resultant of the cellular ROS levels or not.

Also contradictory between the two species was the pentose phosphate pathway. This suggests a further investigation of the proteins involved in this pathway would be informative. Such an investigation could be conducted in a variety of ways including western blot with appropriate antibodies, or multiple reaction monitoring analysis using the mass spectrometer.

An interesting investigation that could be conducted to further understand lipid accumulation in algae would be the induction of nitrogen stress whilst supplementing cells with appropriate amino acids. This could provide evidence as to whether lipid induction was caused by some sensing of the media or was induced as a result of the subsequent consequences of such stress, i.e. the reduction in proteins.

References

References

(1) Newton, J. D. Uncommon Friends: Life with Thomas Edison, Henry Ford, Harvey Firestone, Alexis Carrel, and Charles Lindbergh; Harcourt Brace Jovanovich, 1989.

(2) Measuring Short-Term Energy Security - Moses.pdf.

(3) Ndimba, B. K.; Ndimba, R. J.; Johnson, T. S.; Waditee-Sirisattha, R.; Baba, M.; Sirisattha, S.; Shiraiwa, Y.; Agrawal, G. K.; Rakwal, R. Biofuels as a sustainable energy source: An update of the applications of proteomics in bioenergy crops and algae. J Proteomics **2013**, 93, 234–244.

(4) Mason, R. Middle East unrest puts oil on a slippery slope. Telegraph.co.uk, 2011.

(5) Crane, K.; Goldthau, A.; Toman, M.; Light, T.; Johnson, S. E.; Nader, A.; Rabasa, A.; Dogo, H. Imported Oil and U.S. National Security http://www.rand.org/pubs/monographs/MG838.html (accessed Jul 18, 2013).

(6) Knox-Hayes, J.; Brown, M. A.; Sovacool, B. K.; Wang, Y. Understanding attitudes toward energy security: Results of a cross-national survey. Global Environmental Change **2013**, 23, 609–622.

(7) Chapman, I. The end of Peak Oil? Why this topic is still relevant despite recent denials. Energy Policy.

(8) The receding threat from "peak oil." BBC, 2013.

(9) Jakobsson, K.; Bentley, R.; Söderbergh, B.; Aleklett, K. The end of cheap oil: Bottom-up economic and geologic modeling of aggregate oil production curves. Energy Policy **2012**, 41, 860–870.

(10) Hall, C. A. S.; Ramírez-Pascualli, C. A. What Do We Know About "Peak Oil" Today? In The First Half of the Age of Oil; SpringerBriefs in Energy; Springer New York, 2013; pp. 71–87.

(11) Kjärstad, J.; Johnsson, F. Fossil Fuels: Climate Change and Security of Supply. International J. Sustainable Water and Environmental Systems **2012**, 4, 79–87.

(12) Oreskes, N. The Scientific Consensus on Climate Change. Science **2004**, 306, 1686–1686.

(13) Survey finds 97% of climate science papers agree warming is man-made http://www.theguardian.com/environment/climate-consensus-97-per-cent/2013/may/16/climate-change-scienceofclimatechange (accessed Aug 2, 2013).

(14) Schneider, S. H. The Greenhouse Effect: Science and Policy. Science **1989**, 243, 771–781.

(15) Keeling, C. D.; Piper, S. C.; Bacastow, R. B.; Wahlen, M.; Whorf, T. P.; Heimann, M.; Meijer, H. A. Atmospheric CO2 and 13CO2 Exchange with the Terrestrial Biosphere and Oceans from 1978 to 2000:

Observations and Carbon Cycle Implications. In A History of Atmospheric CO2 and Its Effects on Plants, Animals, and Ecosystems; Baldwin, I. T.; Caldwell, M. M.; Heldmaier, G.; Jackson, R. B.; Lange, O. L.; Mooney, H. A.; Schulze, E.-D.; Sommer, U.; Ehleringer, J. R.; Dearing, M. D.; et al., Eds.; Ecological Studies; Springer New York, 2005; pp. 83–113.

(16) Montzka, S. A.; Dlugokencky, E. J.; Butler, J. H. Non-CO2 greenhouse gases and climate change. Nature **2011**, 476, 43–50.

(17) Schenk, P.; Thomas-Hall, S.; Stephens, E.; Marx, U.; Mussgnug, J.; Posten, C.; Kruse, O.; Hankamer, B. Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. BioEnergy Research **2008**, 1, 20–43.

(18) Costa, J. A. V.; de Morais, M. G. The role of biochemical engineering in the production of biofuels from microalgae. Bioresource Technology **2011**, 102, 2–9.

(19) Wijffels, R. H.; Barbosa, M. J. An Outlook on Microalgal Biofuels. Science **2010**, 329, 796–799.

(20) Mittelbach, M. Biodiesel - Quo vadis? Eur. J. Lipid Sci. Technol. **2009**, 111, 745–746.

(21) James, T. No sugar rush [Power Biofuels]. Engineering & Technology **2010**, 5, 46–47.

(22) Boudet, A. M.; Kajita, S.; Grima-Pettenati, J.; Goffner, D. Lignins and lignocellulosics: a better control of synthesis for new and improved uses. Trends in Plant Science **2003**, 8, 576–581.

(23) Halpin, C. Re-designing lignin for industry and agriculture. Biotechnol. Genet. Eng. Rev **2004**, 21, 229–245.

(24) Torney, F.; Moeller, L.; Scarpa, A.; Wang, K. Genetic engineering approaches to improve bioethanol production from maize. Current Opinion in Biotechnology **2007**, 18, 193–199.

(25) Chen, F.; Dixon, R. A. Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotech **2007**, 25, 759–761.

(26) Festel, G. W. Biofuels – Economic Aspects. Chemical Engineering & Technology **2008**, 31, 715–720.

(27) Lange, J.-P. Lignocellulose conversion: an introduction to chemistry, process and economics. Biofuels, Bioproducts and Biorefining **2007**, 1, 39–48.

(28) Searchinger, T.; Heimlich, R.; Houghton, R. A.; Dong, F.; Elobeid, A.; Fabiosa, J.; Tokgoz, S.; Hayes, D.; Yu, T.-H. Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change. Science **2008**, 319, 1238–1240.

(29) Yuan, J. S.; Tiller, K. H.; Al-Ahmad, H.; Stewart, N. R.; Stewart Jr, C. N. Plants to power: bioenergy to fuel the future. Trends in Plant Science **2008**, 13, 421–429.

(30) Singh, A.; Nigam, P. S.; Murphy, J. D. Renewable fuels from algae: An answer to debatable land based fuels. Bioresource Technology **2011**, 102, 10–16.

(31) Chisti, Y. Biodiesel from microalgae. Biotechnol. Adv. **2007**, 25, 294–306.
(32) Chisti, Y. Biodiesel from microalgae beats bioethanol. Trends Biotechnol. **2008**, 26, 126–131.
(33) Spolaore, P.; Joannis-Cassan, C.; Duran, E.; Isambert, A. Commercial applications of microalgae. Journal of bioscience and bioengineering **2006**, 101, 87–96.
(34) Rosenberg, J. N.; Oyler, G. A.; Wilkinson, L.; Betenbaugh, M. J. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Current Opinion in Biotechnology **2008**, 19, 430–436.
(35) Walker, T. L.; Purton, S.; Becker, D. K.; Collet, C. Microalgae as bioreactors. Plant cell reports **2005**, 24, 629–641.
(36) Sheehan, J.; Dunahay, T.; Benemann, J.; Roessler, P. A look back at the US Department of Energy's Aquatic Species Program: Biodiesel from Algae; NERL report TP-580-24190 NERL report TP-580-24190; NERL report TP-580-24190; National Renewable Energy Laboratory, 1998.
(37) Huntley, M. E.; Redalje, D. G. CO_2 mitigation and renewable oil from photosynthetic microbes: A new appraisal. Mitigation and Adaptation Strategies for Global Change **2007**, 12, 573–608.
(38) Fukuda, H.; Kondo, A.; Noda, H. Biodiesel fuel production by transesterification of oils. Journal of Bioscience and Bioengineering **2001**, 92, 405–416.
(39) Hu, Q.; Sommerfeld, M.; Jarvis, E.; Ghirardi, M.; Posewitz, M.; Seibert, M.; Darzins, A. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. The Plant Journal **2008**, 54, 621–639.
(40) Durrett, T. P.; Benning, C.; Ohlrogge, J. Plant triacylglycerols as feedstocks for the production of biofuels. The Plant Journal **2008**, 54, 593–607.
(41) Demirbas, A. Biofuels securing the planet's future energy needs. Energy Conversion and Management **2009**, 50, 2239–2249.
(42) Mata, T. M.; Martins, A. A.; Caetano, N. S. Microalgae for biodiesel production and other applications: A review. Renewable and Sustainable Energy Reviews **2010**, 14, 217–232.
(43) Dillschneider, R.; Steinweg, C.; Rosello-Sastre, R.; Posten, C. Biofuels from microalgae: Photoconversion efficiency during lipid accumulation. Bioresource Technology **2013**, 142, 647–654.
(44) Sing, S. F.; Isdepsky, A.; Borowitzka, M. A.; Moheimani, N. R. Production of biofuels from microalgae. Mitig Adapt Strateg Glob Change **2013**, 18, 47–72.
(45) Bahadar, A.; Bilal Khan, M. Progress in energy from microalgae: A review. Renewable and Sustainable Energy Reviews **2013**, 27, 128–148.
(46) Pulz, O. Photobioreactors: production systems for phototrophic microorganisms. Applied microbiology and biotechnology **2001**, 57, 287–293.

(47) Fernández, F. G. A.; Sevilla, J. M. F.; Grima, E. M. Photobioreactors for the production of microalgae. Rev Environ Sci Biotechnol **2013**, 12, 131–151.
(48) Jorquera, O.; Kiperstok, A.; Sales, E. A.; Embiruçu, M.; Ghirardi, M. L. Comparative energy life-cycle analyses of microalgal biomass production in open ponds and photobioreactors.
(49) Borowitzka, M. A.; Moheimani, N. R. Open Pond Culture Systems. In Algae for Biofuels and Energy; Borowitzka, M. A.; Moheimani, N. R., Eds.; Developments in Applied Phycology; Springer Netherlands, 2013; pp. 133–152.
(50) Brennan, L.; Owende, P. Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. Renewable Sustainable Energy Rev. **2010**, 14, 557–577.
(51) Molina Grima, E.; Belarbi, E. H.; Acién Fernández, F. G.; Robles Medina, A.; Chisti, Y. Recovery of microalgal biomass and metabolites: process options and economics. Biotechnology Advances **2003**, 20, 491–515.
(52) Uduman, N.; Qi, Y.; Danquah, M. K.; Forde, G. M.; Hoadley, A. Dewatering of microalgal cultures: A major bottleneck to algae-based fuels. Journal of Renewable and Sustainable Energy **2010**, 2, 012701–012701–15.
(53) Ríos, S. D.; Castañeda, J.; Torras, C.; Farriol, X.; Salvadó, J. Lipid extraction methods from microalgal biomass harvested by two different paths: Screening studies toward biodiesel production. Bioresource Technology **2013**, 133, 378–388.
(54) Hejazi, M. A.; Wijffels, R. H. Milking of microalgae. Trends Biotechnol. **2004**, 22, 189–194.
(55) Ramachandra, T. V.; Mahapatra, D. M.; B, K.; Gordon, R. Milking Diatoms for Sustainable Energy: Biochemical Engineering versus Gasoline-Secreting Diatom Solar Panels. Ind. Eng. Chem. Res. **2009**, 48, 8769–8788.
(56) Pragya, N.; Pandey, K. K.; Sahoo, P. K. A review on harvesting, oil extraction and biofuels production technologies from microalgae. Renewable and Sustainable Energy Reviews **2013**, 24, 159–171.
(57) Harvesting Algae http://www.et.byu.edu/ ~wanderto/ homealgaeproject/ Harvesting%20Algae.html (accessed Aug 4, 2013).
(58) GOVERNMENT MISSES 2ND CO2 TARGET: WE DEMAND DEBATE | Green MSPs http://www.greenmsps.org/government-misses-2nd-co2-target-we-demand-debate/ (accessed Aug 4, 2013).
(59) First Algae-Powered Car Attempts to Cross US on 25 Gallons | Inhabitat - Sustainable Design Innovation, Eco Architecture, Green Building http://inhabitat.com/first-algae-powered-car-attempts-to-cross-us-on-25-gallons/ (accessed Aug 4, 2013).
(60) Weyer, K. M.; Bush, D. R.; Darzins, A.; Willson, B. D. Theoretical Maximum Algal Oil Production. Bioenerg. Res. **2010**, 3, 204–213.

(61) Feofilova, E.; Sergeeva, Y.; Ivashechkin, A. Biodiesel-fuel: Content, production, producers, contemporary biotechnology (Review). Applied Biochemistry and Microbiology **2010**, 46, 369–378.
(62) Hsieh, C.-H.; Wu, W.-T. Cultivation of microalgae for oil production with a cultivation strategy of urea limitation. Bioresource Technology **2009**, 100, 3921–3926.
(63) Griffiths, M. J.; Harrison, S. T. L. Lipid productivity as a key characteristic for choosing algal species for biodiesel production. J Appl Phycol **2009**, 21, 493–507.
(64) Borowitzka, M. A.; Moheimani, N. R. Sustainable biofuels from algae. Mitig Adapt Strateg Glob Change **2013**, 18, 13–25.
(65) Nagarajan, S.; Chou, S. K.; Cao, S.; Wu, C.; Zhou, Z. An updated comprehensive techno-economic analysis of algae biodiesel. Bioresource Technology.
(66) Algae to solve the Pentagon's jet fuel problem http://www.theguardian.com/environment/2010/feb/13/algae-solve-pentagon-fuel-problem (accessed Aug 7, 2013).
(67) Delrue, F.; Setier, P.-A.; Sahut, C.; Cournac, L.; Roubaud, A.; Peltier, G.; Froment, A.-K. An economic, sustainability, and energetic model of biodiesel production from microalgae. Bioresource Technology **2012**, 111, 191–200.
(68) July fuel price update | AA https://www.theaa.com/newsroom/news-2013/aa-fuel-price-update-july-2013.html (accessed Aug 7, 2013).
(69) Wilhelm, C.; Büchel, C.; Fisahn, J.; Goss, R.; Jakob, T.; LaRoche, J.; Lavaud, J.; Lohr, M.; Riebesell, U.; Stehfest, K.; et al. The Regulation of Carbon and Nutrient Assimilation in Diatoms is Significantly Different from Green Algae. Protist **2006**, 157, 91–124.
(70) Klein-Marcuschamer, D.; Chisti, Y.; Benemann, J. R.; Lewis, D. A matter of detail: Assessing the true potential of microalgal biofuels. Biotechnology and Bioengineering **2013**, 110, 2317–2322.
(71) Lewis, L. A.; McCourt, R. M. Green algae and the origin of land plants. Am. J. Bot. **2004**, 91, 1535–1556.
(72) Radmer, R. J. Algal diversity and commercial algal products. Bioscience **1996**, 46, 263–270.
(73) Gimpel, J. A.; Specht, E. A.; Georgianna, D. R.; Mayfield, S. P. Advances in microalgae engineering and synthetic biology applications for biofuel production. Current Opinion in Chemical Biology **2013**, 17, 489–495.
(74) Mutanda, T.; Ramesh, D.; Karthikeyan, S.; Kumari, S.; Anandraj, A.; Bux, F. Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production. Bioresource Technology **2011**, 102, 57–70.
(75) Nascimento, I. A.; Marques, S. S. I.; Cabanelas, I. T. D.; Pereira, S. A.; Druzian, J. I.; de Souza, C. O.; Vich, D. V.; de Carvalho, G. C.; Nascimento, M. A. Screening Microalgae Strains for Biodiesel Production: Lipid Productivity and Estimation of Fuel Quality Based

on Fatty Acids Profiles as Selective Criteria. Bioenergy Research **2013**, 6, 1–13.
(76) Ratha, S. K.; Prasanna, R. Bioprospecting microalgae as potential sources of "green energy"--challenges and perspectives (review). Prikl. Biokhim. Mikrobiol. **2012**, 48, 133–149.
(77) genomesonline.org http://www.genomesonline.org/gold.cgi?want=Eukaryotic+Ongoing+Genomes (accessed Sep 15, 2009).
(78) Basova, M. M. Fatty acid composition of lipids in microalgae. Inter J Algae **2005**, 7, 33–57.
(79) Zhila, N. O.; Kalacheva, G. S.; Volova, T. G. Influence of nitrogen deficiency on biochemical composition of the green alga Botryococcus. Journal of Applied Phycology **2005**, 17, 309–315.
(80) Singh, Y.; Kumar, H. D. Lipid and hydrocarbon production by Botryococcus spp. under nitrogen limitation and anaerobiosis. World Journal of Microbiology & Biotechnology **1992**, 8, 121–124.
(81) Genome Project Result http://www.ncbi.nlm.nih.gov/sites/entrez (accessed Feb 4, 2010).
(82) Moheimani, N.; Borowitzka, M. The long-term culture of the coccolithophore Pleurochrysis carterae (Haptophyta) in outdoor raceway ponds. Journal of Applied Phycology **2006**, 18, 703–712.
(83) Pruvost, J.; Van Vooren, G.; Cogne, G.; Legrand, J. Investigation of biomass and lipids production with Neochloris oleoabundans in photobioreactor. Bioresource Technology **2009**, 100, 5988–5995.
(84) Damiani, M. C.; Popovich, C. A.; Constenla, D.; Leonardi, P. I. Lipid analysis in Haematococcus pluvialis to assess its potential use as a biodiesel feedstock. Bioresource Technology **2010**, 101, 3801–3807.
(85) Prochnik, S. E.; Umen, J.; Nedelcu, A. M.; Hallmann, A.; Miller, S. M.; Nishii, I.; Ferris, P.; Kuo, A.; Mitros, T.; Fritz-Laylin, L. K.; et al. Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science **2010**, 329, 223–226.
(86) Shifrin, N. S.; Chisholm, S. W. Phytoplankton Lipids: Interspecific Differences and Effects Of Nitrate, Silicate and Light-Dark Cycle. J. Phycol. **1981**, 17, 374–384.
(87) Worden, A. Z.; Nolan, J. K.; Palenik, B. Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. Limnology and Oceanography **2004**, 49, 168–179.
(88) Brown, M. R.; Jeffrey, S. W. Biochemical composition of microalgae from the green algal classes Chlorophyceae and Prasinophyceae. 1. Amino acids, sugars and pigments. Journal of Experimental Marine Biology and Ecology **1992**, 161, 91–113.
(89) Rodolfi, L.; Chini Zittelli, G.; Bassi, N.; Padovani, G.; Biondi, N.; Bonini, G.; Tredici, M. R. Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. Biotechnology and Bioengineering **2008**, 102, 100–112.

(90) Flynn, K. J.; Zapata, M.; Garrido, J. L.; Opik, H.; Hipkin, C. R. Changes in carbon and nitrogen physiology during ammonium and nitrate nutrition and nitrogen starvation in Isochrysis galbana. European Journal of Phycology **1993**, 28, 47–52.
(91) Emdadi, D.; Berland, B. Variation in lipid class composition during batch growth of Nannochloropsis salina and Pavlova lutheri. Marine Chemistry **1989**, 26, 215–225.
(92) Nozaki, H.; Takano, H.; Misumi, O.; Terasawa, K.; Matsuzaki, M.; Maruyama, S.; Nishida, K.; Yagisawa, F.; Yoshida, Y.; Fujiwara, T.; et al. A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. BMC Biology **2007**, 5, 28.
(93) Radakovits, R.; Jinkerson, R. E.; Darzins, A.; Posewitz, M. C. Genetic Engineering of Algae for Enhanced Biofuel Production. Eukaryotic Cell **2010**, 9, 486–501.
(94) Piorreck, M.; Baasch, K. H.; Pohl, P. Biomass production, total protein, chlorophylls, lipids and fatty acids of freshwater green and blue-green algae under different nitrogen regimes. Phytochemistry **1984**, 23, 207–216.
(95) Xin, L.; Hong-ying, H.; Ke, G.; Ying-xue, S. Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid accumulation of a freshwater microalga Scenedesmus sp. Bioresource Technology **2010**, 101, 5494–5500.
(96) Chen, M.; Tang, H.; Ma, H.; Holland, T. C.; Ng, K. Y. S.; Salley, S. O. Effect of nutrients on growth and lipid accumulation in the green algae Dunaliella tertiolecta. Bioresource Technology In Press, Corrected Proof.
(97) Jeffryes, C.; Rosenberger, J.; Rorrer, G. L. Fed-batch cultivation and bioprocess modeling of Cyclotella sp. for enhanced fatty acid production by controlled silicon limitation. Algal Research **2013**, 2, 16–27.
(98) Takagi, M.; Karseno; Yoshida, T. Effect of salt concentration on intracellular accumulation of lipids and triacylglyceride in marine microalgae Dunaliella cells. Journal of Bioscience and Bioengineering **2006**, 101, 223–226.
(99) Skerratt, J. Effect of uv-b on lipid content of three antarctic marine phytoplankton. Phytochemistry **1998**, 49, 999–1007.
(100) Courchesne, N. M. .; Parisien, A.; Wang, B.; Lan, C. Q. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. Journal of biotechnology **2009**, 141, 31–41.
(101) Converti, A.; Casazza, A. A.; Ortiz, E. Y.; Perego, P.; Del Borghi, M. Effect of temperature and nitrogen concentration on the growth and lipid content of Nannochloropsis oculata and Chlorella vulgaris for biodiesel production. Chemical Engineering and Processing: Process Intensification **2009**, 48, 1146–1151.

(102) Khozin-Goldberg, I.; Cohen, Z. Unraveling algal lipid metabolism: Recent advances in gene identification. Biochimie **2011**, 93, 91–100.
(103) Smith, C. B.; Manoylov, K. M. The effects of variable light and lipids on the water column distribution and interactions of phytoplankton. Plant **2013**, 1, 16–24.
(104) Qin, S.; Lin, H.; Jiang, P. Advances in genetic engineering of marine algae. Biotechnology Advances **2012**, 30, 1602–1613.
(105) Ioki, M.; Baba, M.; Nakajima, N.; Shiraiwa, Y.; Watanabe, M. M. Codon usage of Botryococcus braunii (Trebouxiophyceae, Chlorophyta): implications for genetic engineering applications. Phycologia **2013**, 52, 352–356.
(106) Talebi, A. F.; Tohidfar, M.; Tabatabaei, M.; Bagheri, A.; Mohsenpor, M.; Mohtashami, S. K. Genetic manipulation, a feasible tool to enhance unique characteristic of Chlorella vulgaris as a feedstock for biodiesel production. Mol Biol Rep **2013**, 40, 4421–4428.
(107) Chen, H.-C.; Melis, A. Marker-free genetic engineering of the chloroplast in the green microalga Chlamydomonas reinhardtii. Plant Biotechnol. J. **2013**, 11, 818–828.
(108) Rupprecht, J. From systems biology to fuel—Chlamydomonas reinhardtii as a model for a systems biology approach to improve biohydrogen production. Journal of Biotechnology **2009**, 142, 10–20.
(109) Radakovits, R.; Eduafo, P. M.; Posewitz, M. C. Genetic engineering of fatty acid chain length in Phaeodactylum tricornutum. Metab. Eng. **2011**, 13, 89–95.
(110) Poulsen, N.; Chesley, P. M.; Kröger, N. Molecular Genetic Manipulation of the Diatom Thalassiosira Pseudonana (bacillariophyceae)1. Journal of Phycology **2006**, 42, 1059–1065.
(111) Lapidot, M.; Raveh, D.; Sivan, A.; Arad, S. M.; Shapira, M. Stable chloroplast transformation of the unicellular red alga Porphyridium species. Plant Physiol. **2002**, 129, 7–12.
(112) Minoda, A.; Sakagami, R.; Yagisawa, F.; Kuroiwa, T.; Tanaka, K. Improvement of Culture Conditions and Evidence for Nuclear Transformation by Homologous Recombination in a Red Alga, Cyanidioschyzon merolae 10D. Plant Cell Physiol **2004**, 45, 667–671.
(113) Mussgnug, J. H.; Thomas-Hall, S.; Rupprecht, J.; Foo, A.; Klassen, V.; McDowall, A.; Schenk, P. M.; Kruse, O.; Hankamer, B. Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. Plant Biotechnology Journal **2007**, 5, 802–814.
(114) Kirst, H.; Garcia-Cerdan, J. G.; Zurbriggen, A.; Ruehle, T.; Melis, A. Truncated Photosystem Chlorophyll Antenna Size in the Green Microalga Chlamydomonas reinhardtii upon Deletion of the TLA3-CpSRP43 Gene1[C][W][OA]. Plant Physiol **2012**, 160, 2251–2260.
(115) Mussgnug, J. H.; Thomas-Hall, S.; Rupprecht, J.; Foo, A.; Klassen, V.; McDowall, A.; Schenk, P. M.; Kruse, O.; Hankamer, B. Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. Plant Biotechnol. J. **2007**, 5, 802–814.

(116) Long, S. P.; Humphries, S.; Falkowski, P. G. Photoinhibition of Photosynthesis in Nature. Annual Review of Plant Physiology and Plant Molecular Biology **1994**, 45, 633–662.
(117) Roessler, P. G.; Bleibaum, J. L.; Thompson, G. A.; Ohlrogge, J. B. Characteristics of the Gene That Encodes Acetyl-CoA Carboxylase in the Diatom Cyclotella crypticaa. Annals of the New York Academy of Sciences **1994**, 721, 250–256.
(118) Dunahay, T. G.; Jarvis, E. E.; Dais, S. S.; Roessler, P. G. Manipulation of microalgal lipid production using genetic engineering. Appl Biochem Biotechnol **1996**, 57-58, 223–231.
(119) Wang, Z. T.; Ullrich, N.; Joo, S.; Waffenschmidt, S.; Goodenough, U. Algal Lipid Bodies: Stress Induction, Purification, and Biochemical Characterization in Wild-Type and Starchless Chlamydomonas reinhardtii. Eukaryotic Cell **2009**, 8, 1856–1868.
(120) Dunahay, T. G.; Jarvis, E. E.; Roessler, P. G. Genetic transformation of the diatoms Cyclotella cryptica and Navicula saprophila. Journal of Phycology **1995**, 31, 1004–1012.
(121) Vigeolas, H.; Waldeck, P.; Zank, T.; Geigenberger, P. Increasing seed oil content in oil-seed rape (Brassica napus L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnology Journal **2007**, 5, 431–441.
(122) Fulda, M.; Schnurr, J.; Abbadi, A.; Heinz, E.; Browse, J. Peroxisomal Acyl-CoA Synthetase Activity Is Essential for Seedling Development in Arabidopsis thaliana. Plant Cell **2004**, 16, 394–405.
(123) Liu, X.; Sheng, J.; Curtiss III, R. Fatty acid production in genetically modified cyanobacteria. Proc Natl Acad Sci U S A **2011**, 108, 6899–6904.
(124) Jacobsen, S.-E.; Sørensen, M.; Pedersen, S. M.; Weiner, J. Feeding the world: genetically modified crops versus agricultural biodiversity. Agron. Sustain. Dev. 1–12.
(125) Conner, A. J.; Glare, T. R.; Nap, J.-P. The release of genetically modified crops into the environment. The Plant Journal **2003**, 33, 19–46.
(126) Sánchez-Pla, A.; Reverter, F.; Ruíz de Villa, M. C.; Comabella, M. Transcriptomics: mRNA and alternative splicing. Journal of Neuroimmunology **2012**, 248, 23–31.
(127) Graves, P. R.; Haystead, T. A. J. Molecular Biologist's Guide to Proteomics. Microbiol. Mol. Biol. Rev. **2002**, 66, 39–63.
(128) Dettmer, K.; Aronov, P. A.; Hammock, B. D. Mass spectrometry-based metabolomics. Mass Spectrom Rev **2007**, 26, 51–78.
(129) Rogers, S.; Girolami, M.; Kolch, W.; Waters, K. M.; Liu, T.; Thrall, B.; Wiley, S. H. Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. Bioinformatics **2008**, 24, 2894–2900.
(130) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Corrigendum: Global quantification of mammalian gene expression control. Nature **2013**, 495, 126–127.

(131) Ghazalpour, A.; Bennett, B.; Petyuk, V. A.; Orozco, L.; Hagopian, R.; Mungrue, I. N.; Farber, C. R.; Sinsheimer, J.; Kang, H. M.; Furlotte, N.; et al. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. PLoS Genet **2011**, 7, e1001393.
(132) Hegde, P. S.; White, I. R.; Debouck, C. Interplay of transcriptomics and proteomics. Current Opinion in Biotechnology **2003**, 14, 647–651.
(133) Garcia, D. E.; Baidoo, E. E.; Benke, P. I.; Pingitore, F.; Tang, Y. J.; Villa, S.; Keasling, J. D. Separation and mass spectrometry in microbial metabolomics. Current Opinion in Microbiology **2008**, 11, 233–239.
(134) Fernie, A. R.; Obata, T.; Allen, A. E.; Araújo, W. L.; Bowler, C. Leveraging metabolomics for functional investigations in sequenced marine diatoms. Trends in Plant Science **2012**, 17, 395–403.
(135) Patti, G. J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol **2012**, 13, 263–269.
(136) Boyle, N. R.; Morgan, J. A. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. BMC Systems Biology **2009**, 3, 4.
(137) Fukusaki, E.; Kobayashi, A. Plant metabolomics: potential for practical operation. Journal of Bioscience and Bioengineering **2005**, 100, 347–354.
(138) Hatzimanikatis, V.; Choe, L. H.; Lee, K. H. Proteomics: Theoretical and Experimental Considerations. Biotechnology Progress **1999**, 15, 312–318.
(139) Woo, H. M.; Murray, G. W.; Batth, T. S.; Prasad, N.; Adams, P. D.; Keasling, J. D.; Petzold, C. J.; Lee, T. S. Application of targeted proteomics and biological parts assembly in E. coli to optimize the biosynthesis of an anti-malarial drug precursor, amorpha-4,11-diene. Chemical Engineering Science **2013**, 103, 21–28.
(140) Fell, D. A.; Thomas, S. Physiological control of metabolic flux: The requirement for multisite modulation. Biochemical Journal **1995**, 311, 35–39.
(141) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. Nature **2011**, 473, 337–342.
(142) Towbin, H.; Staehelin, T.; Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A **1979**, 76, 4350–4354.
(143) Renart, J.; Reiser, J.; Stark, G. R. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. Proc Natl Acad Sci U S A **1979**, 76, 3116–3120.
(144) Engvall, E.; Perlmann, P. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. Immunochemistry **1971**, 8, 871–874.

(145) Nilsson, T.; Mann, M.; Aebersold, R.; Iii, J. R. Y.; Bairoch, A.; Bergeron, J. J. M. Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Meth **2010**, 7, 681–685.
(146) O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. Journal of Biological Chemistry **1975**, 250, 4007 –4021.
(147) Beranova-Giorgianni, S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. TrAC Trends in Analytical Chemistry **2003**, 22, 273–281.
(148) Lamond, A. I.; Uhlen, M.; Horning, S.; Makarov, A.; Robinson, C. V.; Serrano, L.; Hartl, F. U.; Baumeister, W.; Werenskiold, A. K.; Andersen, J. S.; et al. Advancing Cell Biology Through Proteomics in Space and Time (PROSPECTS). Mol Cell Proteomics **2012**, 11.
(149) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. Nature **2003**, 422, 198–207.
(150) Zhu, W.; Smith, J. W.; Huang, C.-M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. BioMed Research International **2009**, 2010.
(151) Beynon, R. J.; Pratt, J. M. Metabolic Labeling of Proteins for Proteomics. Mol Cell Proteomics **2005**, 4, 857–872.
(152) Borràs, E.; Espadas, G.; Mancuso, F. M.; Maier, T.; Chiva, C.; Sabidó, E. Integrative quantitation enables a comprehensive proteome comparison of two Mycoplasma pneumoniae genetic perturbations. Molecular BioSystems **2013**, 9, 1249.
(153) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; 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.
(154) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc **2009**, 4, 484–494.
(155) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. **1999**, 17, 994–999.
(156) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K. A.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. **2003**, 75, 1895–1904.
(157) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; 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.
(158) Amine-reactive, 10-plex Tandem Mass Tag Reagents http://www.piercenet.com/browse.cfm?fldID=8F2F91B7-A573-C989-C39E-FA0FA962411A (accessed Aug 10, 2013).

(159) iTRAQ® Reagents | Multiplexed, Amine-Specific, Stable Isotope | SCIEX http://www.sciex.com/products/standards-and-reagents/iTRAQ-Reagents.xml (accessed Aug 10, 2013).
(160) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. Mol Cell Proteomics **2005**, 4, 1487–1502.
(161) Evans, C.; Noirel, J.; Ow, S. Y.; Salim, M.; Pereira-Medrano, A. G.; Couto, N.; Pandhal, J.; Smith, D.; Pham, T. K.; Karunakaran, E.; et al. An insight into iTRAQ: where do we stand now? Analytical and Bioanalytical Chemistry **2012**, 404, 1011–1027.
(162) Pham, T. K.; Chong, P. K.; Gan, C. S.; Wright, P. C. Proteomic Analysis of Saccharomyces cerevisiae under High Gravity Fermentation Conditions. Journal of Proteome Research **2006**, 5, 3411–3419.
(163) Pham, T. K.; Wright, P. C. Proteomic Analysis of Calcium Alginate-Immobilized Saccharomyces cerevisiae under High-Gravity Fermentation Conditions. Journal of Proteome Research **2008**, 7, 515–525.
(164) Pham, T. K.; Wright, P. C. The Proteomic Response of Saccharomyces cerevisiae in Very High Glucose Conditions with Amino Acid Supplementation. Journal of Proteome Research **2008**, 7, 4766–4774.
(165) Pandhal, J.; Ow, S. Y.; Noirel, J.; Wright, P. C. Improving N-glycosylation efficiency in Escherichia coli using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. Biotechnology and Bioengineering **2011**, 108, 902–912.
(166) Jamers, A.; Blust, R.; De Coen, W. Omics in algae: Paving the way for a systems biological understanding of algal stress phenomena? Aquatic Toxicology **2009**, 92, 114–121.
(167) Valledor, L.; Furuhashi, T.; Hanak, A.-M.; Weckwerth, W. Systemic cold stress adaptation of chlamydomonas reinhardtIII. Molecular and Cellular Proteomics **2013**, 12, 2032–2047.
(168) Hsieh, S. I.; Castruita, M.; Malasarn, D.; Urzica, E.; Erde, J.; Page, M. D.; Yamasaki, H.; Casero, D.; Pellegrini, M.; Merchant, S. S.; et al. The proteome of copper, iron, zinc, and manganese micronutrient deficiency in Chlamydomonas reinhardtii. Molecular and Cellular Proteomics **2013**, 12, 65–86.
(169) Choi, Y.-E.; Hwang, H.; Kim, H.-S.; Ahn, J.-W.; Jeong, W.-J.; Yang, J.-W. Comparative proteomics using lipid over-producing or less-producing mutants unravels lipid metabolisms in Chlamydomonas reinhardtii. Bioresource Technology **2013**, 145, 108–115.
(170) Zaffagnini, M.; Bedhomme, M.; Groni, H.; Marchand, C. H.; Puppo, C.; Gontero, B.; Cassier-Chauvat, C.; Decottignies, P.; Lemaire, S. D. Glutathionylation in the photosynthetic model organism Chlamydomonas reinhardtii: A proteomic survey. Molecular and Cellular Proteomics **2012**, 11.

(171) Yokthongwattana, C.; Mahong, B.; Roytrakul, S.; Phaonaklop, N.; Narangajavana, J.; Yokthongwattana, K. Proteomic analysis of salinity-stressed Chlamydomonas reinhardtii revealed differential suppression and induction of a large number of important housekeeping proteins. Planta **2012**, 235, 649–659.
(172) Winck, F. V.; Riaño-Pachón, D. M.; Sommer, F.; Rupprecht, J.; Mueller-Roeber, B. The nuclear proteome of the green alga Chlamydomonas reinhardtii. Proteomics **2012**, 12, 95–100.
(173) Wang, H.; Alvarez, S.; Hicks, L. M. Comprehensive Comparison of iTRAQ and Label-free LC-Based Quantitative Proteomics Approaches Using Two Chlamydomonas reinhardtii Strains of Interest for Biofuels Engineering. J. Proteome Res. **2011**, 11, 487–501.
(174) Valledor, L.; Recuenco-Munoz, L.; Egelhofer, V.; Wienkoop, S.; Weckwerth, W. The different proteomes of Chlamydomonas reinhardtii. Journal of Proteomics **2012**, 75, 5883–5887.
(175) Nguyen, H. M.; Baudet, M.; Cuiné, S.; Adriano, J.-M.; Barthe, D.; Billon, E.; Bruley, C.; Beisson, F.; Peltier, G.; Ferro, M.; et al. Proteomic profiling of oil bodies isolated from the unicellular green microalga Chlamydomonas reinhardtii: with focus on proteins involved in lipid metabolism. Proteomics **2011**, 11, 4266–4273.
(176) Nestler, H.; Groh, K. J.; Schönenberger, R.; Eggen, R. I. L.; Suter, M. J.-F. Linking proteome responses with physiological and biochemical effects in herbicide-exposed Chlamydomonas reinhardtii. Journal of Proteomics **2012**, 75, 5370–5385.
(177) Mastrobuoni, G.; Irgang, S.; Pietzke, M.; Aßmus, H. E.; Wenzel, M.; Schulze, W. X.; Kempa, S. Proteome dynamics and early salt stress response of the photosynthetic organism Chlamydomonas reinhardtii. BMC Genomics **2012**, 13.
(178) Mahong, B.; Roytrakul, S.; Phaonaklop, N.; Wongratana, J.; Yokthongwattana, K. Proteomic analysis of a model unicellular green alga, Chlamydomonas reinhardtii, during short-term exposure to irradiance stress reveals significant down regulation of several heat-shock proteins. Planta **2012**, 235, 499–511.
(179) Longworth, J.; Noirel, J.; Pandhal, J.; Wright, P. C.; Vaidyanathan, S. HILIC- and SCX-based quantitative proteomics of chlamydomonas reinhardtii during nitrogen starvation induced lipid and carbohydrate accumulation. Journal of Proteome Research **2012**, 11, 5959–5971.
(180) Lee, D. Y.; Park, J.-J.; Barupal, D. K.; Fiehn, O. System Response of Metabolic Networks in Chlamydomonas reinhardtii to Total Available Ammonium. Mol. Cell. Proteomics **2012**, 11, 973–988.
(181) Boesger, J.; Wagner, V.; Weisheit, W.; Mittag, M. Application of phosphoproteomics to find targets of casein kinase 1 in the flagellum of chlamydomonas. International Journal of Plant Genomics **2012**, 2012.
(182) Yadavalli, V.; Nellaepalli, S.; Subramanyam, R. Proteomic analysis of thylakoid membranes. Methods Mol. Biol. **2011**, 684, 159–170.

(183) Terashima, M.; Specht, M.; Hippler, M. The chloroplast proteome: a survey from the Chlamydomonas reinhardtii perspective with a focus on distinctive features. Curr. Genet **2011**, 57, 151–168.
(184) Specht, M.; Stanke, M.; Terashima, M.; Naumann-Busch, B.; Janßen, I.; Höhner, R.; Hom, E. F. Y.; Liang, C.; Hippler, M. Concerted action of the new Genomic Peptide Finder and AUGUSTUS allows for automated proteogenomic annotation of the Chlamydomonas reinhardtii genome. Proteomics **2011**, 11, 1814–1823.
(185) Mühlhaus, T.; Weiss, J.; Hemme, D.; Sommer, F.; Schroda, M. Quantitative Shotgun Proteomics Using a Uniform ^{15}N-Labeled Standard to Monitor Proteome Dynamics in Time Course Experiments Reveals New Insights into the Heat Stress Response of Chlamydomonas reinhardtii. Mol. Cell. Proteomics **2011**, 10.
(186) Bienvenut, W. V.; Espagne, C.; Martinez, A.; Majeran, W.; Valot, B.; Zivy, M.; Vallon, O.; Adam, Z.; Meinnel, T.; Giglione, C. Dynamics of post-translational modifications and protein stability in the stroma of Chlamydomonas reinhardtii chloroplasts. Proteomics **2011**, 11, 1734–1750.
(187) Baba, M.; Suzuki, I.; Shiraiwa, Y. Proteomic analysis of high-CO 2-inducible extracellular proteins in the unicellular Green Alga, Chlamydomonas reinhardtii. Plant and Cell Physiology **2011**, 52, 1302–1314.
(188) Wienkoop, S.; Weiss, J.; May, P.; Kempa, S.; Irgang, S.; Recuenco-Munoz, L.; Pietzke, M.; Schwemmer, T.; Rupprecht, J.; Egelhofer, V.; et al. Targeted proteomics for Chlamydomonas reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. Mol. BioSyst. **2010**, 6, 1018–1031.
(189) Terashima, M.; Specht, M.; Naumann, B.; Hippler, M. Characterizing the anaerobic response of Chlamydomonas reinhardtii by quantitative proteomics. Mol. Cell. Proteomics **2010**, 9, 1514–1532.
(190) Mathy, G.; Cardol, P.; Dinant, M.; Blomme, A.; Gérin, S.; Cloes, M.; Ghysels, B.; DePauw, E.; Leprince, P.; Remacle, C.; et al. Proteomic and Functional Characterization of a Chlamydomonas reinhardtii Mutant Lacking the Mitochondrial Alternative Oxidase 1. J. Proteome Res. **2010**, 9, 2825–2838.
(191) Gérin, S.; Mathy, G.; Blomme, A.; Franck, F.; Sluse, F. E. Plasticity of the mitoproteome to nitrogen sources (nitrate and ammonium) in Chlamydomonas reinhardtii: The logic of Aox1 gene localization. Biochimica et Biophysica Acta - Bioenergetics **2010**, 1797, 994–1003.
(192) Cid, C.; Garcia-Descalzo, L.; Casado-Lafuente, V.; Amils, R.; Aguilera, A. Proteomic analysis of the response of an acidophilic strain of Chlamydomonas sp. (Chlorophyta) to natural metal-rich water. Proteomics **2010**, 10, 2026–2036.
(193) Chen, M.; Zhao, L.; Sun, Y.-L.; Cui, S.-X.; Zhang, L.-F.; Yang, B.; Wang, J.; Kuang, T.-Y.; Huang, F. Proteomic Analysis of Hydrogen

Photoproduction in Sulfur-Deprived Chlamydomonas Cells. J. Proteome Res. **2010**, 9, 3854–3866.
(194) Wagner, V.; Mittag, M. Probing circadian rhythms in Chlamydomonas rheinhardtii by functional proteomics. Methods Mol. Biol. **2009**, 479, 173–188.
(195) Wagner, V.; Boesger, J.; Mittag, M. Sub-proteome analysis in the green flagellate alga Chlamydomonas reinhardtii. J. Basic Microbiol. **2009**, 49, 32–41.
(196) Stauber, E. J.; Busch, A.; Naumann, B.; Svatoscaron, A.; Hippler, M. Proteotypic profiling of LHCI from Chlamydomonas reinhardtii provides new insights into structure and function of the complex. Proteomics **2009**, 9, 398–408.
(197) Rolland, N.; Atteia, A.; Decottignies, P.; Garin, J.; Hippler, M.; Kreimer, G.; Lemaire, S. D.; Mittag, M.; Wagner, V. Chlamydomonas proteomics. Curr. Opin. Microbiol. **2009**, 12, 285–291.
(198) Nesatyy, V. J.; Groh, K.; Nestler, H.; Suter, M. J.-F. On the acquisition of +1 charge states during high-throughput proteomics: Implications on reproducibility, number and confidence of protein identifications. Journal of Proteomics **2009**, 72, 761–770.
(199) May, P.; Christian, J.-O.; Kempa, S.; Walther, D. ChlamyCyc: an integrative systems biology database and web-portal for Chlamydomonas reinhardtii . BMC Genomics **2009**, 10, 209.
(200) Kreimer, G. The green algal eyespot apparatus: a primordial visual system and more? Curr. Genet **2009**, 55, 19–43.
(201) Heide, H.; Nordhues, A.; Drepper, F.; Nick, S.; Schulz-Raffelt, M.; Haehnel, W.; Schroda, M. Application of quantitative immunoprecipitation combined with knockdown and cross-linking to Chlamydomonas reveals the presence of vesicle-inducing protein in plastids 1 in a common complex with chloroplast HSP90C. Proteomics **2009**, 9, 3079–3089.
(202) Clemens, S.; Naumann, B.; Hippler, M. Proteomics of metal mediated protein dynamics in plants - Iron and cadmium in the focus. Frontiers in Bioscience **2009**, 14, 1955–1969.
(203) Atteia, A.; Adrait, A.; Brugière, S.; Tardif, M.; van Lis, R.; Deusch, O.; Dagan, T.; Kuhn, L.; Gontero, B.; Martin, W.; et al. A proteomic survey of Chlamydomonas reinhardtii mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. Mol. Biol. Evol. **2009**, 26, 1533–1548.
(204) Wagner, V.; Kreimer, G.; Mittag, M. The power of functional proteomics: Components of the green algal eyespot and its light signaling pathway(s). Plant Signal Behav **2008**, 3, 433–435.
(205) Wagner, V.; Ullmann, K.; Mollwo, A.; Kaminski, M.; Mittag, M.; Kreimer, G. The phosphoproteome of a Chlamydomonas reinhardtii eyespot fraction includes key proteins of the light signaling pathway. Plant Physiology **2008**, 146, 772–788.

(206) Michelet, L.; Zaffagnini, M.; Vanacker, H.; Le Maréchal, P.; Marchand, C.; Schroda, M.; Lemaire, S. D.; Decottignies, P. In Vivo Targets of S-Thiolation in Chlamydomonas reinhardtii . Journal of Biological Chemistry **2008**, 283, 21571–21578.
(207) May, P.; Wienkoop, S.; Kempa, S.; Usadel, B.; Christian, N.; Rupprecht, J.; Weiss, J.; Recuenco-Munoz, L.; Ebenhoh, O.; Weckwerth, W.; et al. Metabolomics- and Proteomics-Assisted Genome Annotation and Analysis of the Draft Metabolic Network of Chlamydomonas reinhardtii . Genetics **2008**, 179, 157–166.
(208) Keller, L. C.; Marshall, W. F. Isolation and proteomic analysis of Chlamydomonas centrioles. Methods Mol. Biol. **2008**, 432, 289–300.
(209) Vener, A. V. Environmentally modulated phosphorylation and dynamics of proteins in photosynthetic membranes. Biochim. Biophys. Acta **2007**, 1767, 449–457.
(210) Naumann, B.; Busch, A.; Allmer, J.; Ostendorf, E.; Zeller, M.; Kirchhoff, H.; Hippler, M. Comparative quantitative proteomics to investigate the remodeling of bioenergetic pathways under iron deficiency in Chlamydomonas reinhardtii. Proteomics **2007**, 7, 3964–3979.
(211) Krüger, R.; Wolschin, F.; Weckwerth, W.; Bettmer, J.; Lehmann, W. D. Plant protein phosphorylation monitored by capillary liquid chromatography-element mass spectrometry. Biochemical and Biophysical Research Communications **2007**, 355, 89–96.
(212) Wagner, V.; Gessner, G.; Heiland, I.; Kaminski, M.; Hawat, S.; Scheffler, K.; Mittag, M. Analysis of the Phosphoproteome of Chlamydomonas reinhardtii Provides New Insights into Various Cellular Pathways. Eukaryotic Cell **2006**, 5, 457–468.
(213) Turkina, M. V.; Kargul, J.; Blanco-Rivero, A.; Villarejo, A.; Barber, J.; Vener, A. V. Environmentally Modulated Phosphoproteome of Photosynthetic Membranes in the Green Alga Chlamydomonas reinhardtii. Molecular & Cellular Proteomics **2006**, 5, 1412–1425.
(214) Turkina, M. V.; Blanco-Rivero, A.; Vainonen, J. P.; Vener, A. V.; Villarejo, A. CO_2 limitation induces specific redox-dependent protein phosphorylation in Chlamydomonas reinhardtii. Proteomics **2006**, 6, 2693–2704.
(215) Schmidt, M.; Gessner, G.; Luff, M.; Heiland, I.; Wagner, V.; Kaminski, M.; Geimer, S.; Eitzinger, N.; Reissenweber, T.; Voytsekh, O.; et al. Proteomic analysis of the eyespot of Chlamydomonas reinhardtii provides novel insights into its components and tactic movements. Plant Cell **2006**, 18, 1908–1930.
(216) Muck, A.; Ibáñez, A. J.; Stauber, E. J.; Mansourova, M.; Svatoscaron, A. Atmospheric molding of ionic copolymer MALDI-TOF/MS arrays: A new tool for protein identification/profiling. Elecetrophoresis **2006**, 27, 4952–4959.
(217) Kan, G.-F.; Miao, J.-L.; Shi, C.-J.; Li, G.-Y. Proteomic Alterations of Antarctic Ice Microalga Chlamydomonas sp. Under Low-Temperature Stress. J Integrative Plant Biology **2006**, 48, 965–970.

(218) Iliev, D.; Voystekh, O.; Mittag, M. The circadian system of Chlamydomonas reinhardtii . Biological Rhythm Research **2006**, 37, 323.
(219) Gillet, S.; Decottignies, P.; Chardonnet, S.; Le Maréchal, P. Cadmium response and redoxin targets in Chlamydomonas reinhardtii: a proteomic approach. Photosynthesis Research **2006**, 89, 201–211.
(220) Förster, B.; Mathesius, U.; Pogson, B. J. Comparative proteomics of high light stress in the model alga Chlamydomonas reinhardtii. Proteomics **2006**, 6, 4309–4320.
(221) Atteia, A.; van Lis, R.; Gelius-Dietrich, G.; Adrait, A.; Garin, J.; Joyard, J.; Rolland, N.; Martin, W. Pyruvate Formate-lyase and a Novel Route of Eukaryotic ATP Synthesis in Chlamydomonas Mitochondria. Journal of Biological Chemistry **2006**, 281, 9909–9918.
(222) Allmer, J.; Naumann, B.; Markert, C.; Zhang, M.; Hippler, M. Mass spectrometric genomic data mining: Novel insights into bioenergetic pathways in Chlamydomonas reinhardtii. Proteomics **2006**, 6, 6207–6220.
(223) Wagner, V.; Gessner, G.; Mittag, M. Functional proteomics: A promising approach to find novel components of the circadian system. Chronobiology International **2005**, 22, 403–415.
(224) Pazour, G. J.; Agrin, N.; Leszyk, J.; Witman, G. B. Proteomic analysis of a eukaryotic cilium. Journal of Cell Biology **2005**, 170, 103–113.
(225) Naumann, B.; Stauber, E. J.; Busch, A.; Sommer, F.; Hippler, M. N-terminal Processing of Lhca3 Is a Key Step in Remodeling of the Photosystem I-Light-harvesting Complex Under Iron Deficiency in Chlamydomonas reinhardtii. Journal of Biological Chemistry **2005**, 280, 20431–20441.
(226) Manuell, A. L.; Yamaguchi, K.; Haynes, P. A.; Milligan, R. A.; Mayfield, S. P. Composition and Structure of the 80 S Ribosome from the Green Alga Chlamydomonas reinhardtii : 80 S Ribosomes are Conserved in Plants and Animals. Journal of Molecular Biology **2005**, 351, 266–279.
(227) Keller, L. C.; Romijn, E. P.; Zamora, I.; Yates, J. R.; Marshall, W. F. Proteomic analysis of isolated Chlamydomonas centrioles reveals orthologs of ciliary-disease genes. Curr. Biol **2005**, 15, 1090–1098.
(228) Cardol, P.; González-Halphen, D.; Reyes-Prieto, A.; Baurain, D.; Matagne, R. F.; Remacle, C. The mitochondrial oxidative phosphorylation proteome of Chlamydomonas reinhardtii deduced from the genome sequencing project. Plant Physiology **2005**, 137, 447–459.
(229) Wagner, V.; Fiedler, M.; Markert, C.; Hippler, M.; Mittag, M. Functional proteomics of circadian expressed proteins from Chlamydomonas reinhardtii. FEBS Letters **2004**, 559, 129–135.
(230) Stauber, E. J.; Hippler, M. Chlamydomonas reinhardtii proteomics. Plant Physiol. Biochem. **2004**, 42, 989–1001.
(231) Lemaire, S. D.; Guillon, B.; Le Maréchal, P.; Keryer, E.; Miginiac-Maslow, M.; Decottignies, P. New thioredoxin targets in the

unicellular photosynthetic eukaryote Chlamydomonas reinhardtii . Proceedings of the National Academy of Sciences of the United States of America **2004**, 101, 7475–7480.
(232) Stauber, E. J.; Fink, A.; Markert, C.; Kruse, O.; Johanningmeier, U.; Hippler, M. Proteomics of Chlamydomonas reinhardtii Light-Harvesting Proteins. Eukaryotic Cell **2003**, 2, 978–994.
(233) Rexroth, S.; Meyer zu Tittingdorf, J. M. W.; Krause, F.; Dencher, N. A.; Seelert, H. Thylakoid membrane at altered metabolic state: Challenging the forgotten realms of the proteome. Electrophoresis **2003**, 24, 2814–2823.
(234) Yamaguchi, K.; Beligni, M. V.; Prieto, S.; Haynes, P. A.; McDonald, W. H.; Yates, J. R.; Mayfield, S. P. Proteomic characterization of the Chlamydomonas reinhardtii chloroplast ribosome. Identification of proteins unique to th e70 S ribosome. J. Biol. Chem **2003**, 278, 33774–33785.
(235) Van Lis, R.; Atteia, A.; Mendoza-Hernandez, G.; Gonzalez-Halphen, D. Identification of Novel Mitochondrial Protein Components of Chlamydomonas reinhardtii . A Proteomic Approach. Plant Physiol. **2003**, 132, 318–330.
(236) Yamaguchi, K.; Prieto, S.; Beligni, M. V.; Haynes, P. A.; Hayes McDonald, W.; Yates III, J. R.; Mayfield, S. P. Proteomic characterization of the small subunit of Chlamydomonas reinhardtii chloroplast ribosome: Identification of a novel S1 domain-containing protein and unusually large orthologs of bacterial S2, S3, and S5. Plant Cell **2002**, 14, 2957–2974.
(237) Hippler, M.; Klein, J.; Fink, A.; Allinger, T.; Hoerth, P. Towards functional proteomics of membrane protein complexes: analysis of thylakoid membranes from Chlamydomonas reinhardtii . Plant J **2001**, 28, 595–606.
(238) Li, Y.; Yuan, Z.; Mu, J.; Chen, D.; Feng, B. Proteomic analysis of lipid accumulation in chlorella protothecoides cells by heterotrophic N deprivation coupling cultivation. Energy and Fuels **2013**, 27, 4031–4040.
(239) Guarnieri, M. T.; Nag, A.; Yang, S.; Pienkos, P. T. Proteomic analysis of Chlorella vulgaris: Potential targets for enhanced lipid accumulation. J Proteomics **2013**, 93, 245–253.
(240) Guarnieri, M. T.; Nag, A.; Smolinski, S. L.; Darzins, A.; Seibert, M.; Pienkos, P. T. Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalga. PLoS ONE **2011**, 6.
(241) Cui, L.; Chai, Y.; Li, J.; Liu, H.; Zhang, L.; Xue, L. Identification of a glucose-6-phosphate isomerase involved in adaptation to salt stress of Dunaliella salina. J Appl Phycol **2010**, 22, 563–568.
(242) Jia, Y.; Xue, L.; Li, J.; Liu, H. Isolation and proteomic analysis of the halotolerant alga Dunaliella salina flagella using shotgun strategy. Molecular Biology Reports **2010**, 37, 711–716.

(243) Katz, A.; Waridel, P.; Shevchenko, A.; Pick, U. Salt-induced Changes in the Plasma Membrane Proteome of the Halotolerant Alga Dunaliella salina as Revealed by Blue Native Gel Electrophoresis and Nano-LC-MS/MS Analysis. Molecular & Cellular Proteomics **2007**, 6, 1459–1472.
(244) Waridel, P.; Frank, A.; Thomas, H.; Surendranath, V.; Sunyaev, S.; Pevzner, P.; Shevchenko, A. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. Proteomics **2007**, 7, 2318–2329.
(245) Liska, A. J.; 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.
(246) Tran, N.-P.; Park, J.-K.; Hong, S.-J.; Lee, C.-G. Proteomics of proteins associated with astaxanthin accumulation in the green algae Haematococcus lacustris under the influence of sodium orthovanadate. Biotechnol. Lett **2009**, 31, 1917–1922.
(247) Tran, N.-P.; Park, J.-K.; Lee, C.-G. Proteomics analysis of proteins in green alga Haematococcus lacustris (Chlorophyceae) expressed under combined stress of nitrogen starvation and high irradiance. Enzyme and Microbial Technology **2009**, 45, 241–246.
(248) Kim, J.-D.; Lee, W.-S.; Kim, B.; Lee, C.-G. Proteomic analysis of protein expression patterns associated with astaxanthin accumulation by green alga Haematococcus pluvialis (chlorophyceae) under high light stress. Journal of Microbiology and Biotechnology **2006**, 16, 1222–1228.
(249) Wang, S.-B.; Chen, F.; Sommerfeld, M.; Hu, Q. Isolation and proteomic analysis [corrected] of cell wall-deficient Haematococcus pluvialis mutants. Proteomics **2005**, 5, 4839–4851.
(250) Wang, S.-B.; Chen, F.; Sommerfeld, M.; Hu, Q. Proteomic analysis of molecular response to oxidative stress by the green alga Haematococcus pluvialis(Chlorophyceae). Planta **2004**, 220, 17–29.
(251) Wang, S.-B.; Hu, Q.; Sommerfeld, M.; Chen, F. Cell wall proteomics of the green alga Haematococcus pluvialis (Chlorophyceae). Proteomics **2004**, 4, 692–708.
(252) Wang, S.-B.; Hu, Q.; Sommerfeld, M.; Chen, F. An optimized protocol for isolation of soluble proteins from microalgae for two-dimensional gel electrophoresis analysis. Journal of Applied Phycology **2003**, 15, 485–496.
(253) Martin, S. F.; Munagapati, V. S.; Salvo-Chirnside, E.; Kerr, L. E.; Le Bihan, T. Proteome turnover in the green alga Ostreococcus tauri by time course 15N metabolic labeling mass spectrometry. Journal of Proteome Research **2012**, 11, 476–486.
(254) Le Bihan, T.; Martin, S. F.; Chirnside, E. S.; van Ooijen, G.; Barrios-Llerena, M. E.; O'Neill, J. S.; Shliaha, P. V.; Kerr, L. E.; Millar, A. J. Shotgun proteomic analysis of the unicellular alga Ostreococcus tauri. J Proteomics **2011**, 74, 2060–2070.

(255) Barrios-Llerena, M. E.; Pritchard, J. C.; Kerr, L. E.; Le Bihan, T. The use of a novel quantitation strategy based on Reductive Isotopic Di-Ethylation (RIDE) to evaluate the effect of glufosinate on the unicellular algae Ostreococcus tauri. Journal of Proteomics **2011**, 74, 2798–2809.
(256) Jancek, S.; Gourbiere, S.; Moreau, H.; Piganeau, G. Clues about the Genetic Basis of Adaptation Emerge from Comparing the Proteomes of Two Ostreococcus Ecotypes (Chlorophyta, Prasinophyceae). Mol Biol Evol **2008**, 25, 2293–2300.
(257) Lobanov, A. V.; Fomenko, D. E.; Zhang, Y.; Sengupta, A.; Hatfield, D. L.; Gladyshev, V. N. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. Genome Biol **2007**, 8, R198–R198.
(258) Dyhrman, S. T.; Jenkins, B. D.; Rynearson, T. A.; Saito, M. A.; Mercier, M. L.; Alexander, H.; Whitney, L. P.; Drzewianowski, A.; Bulygin, V. V.; Bertrand, E. M.; et al. The Transcriptome and Proteome of the Diatom Thalassiosira pseudonana Reveal a Diverse Phosphorus Stress Response. PLoS ONE **2012**, 7, e33768.
(259) Carvalho, R.; Lettieri, T. Proteomic analysis of the marine diatom Thalassiosira pseudonana upon exposure to benzo(a)pyrene. BMC Genomics **2011**, 12, 159.
(260) Grouneva, I.; Rokka, A.; Aro, E.-M. The thylakoid membrane proteome of two marine diatoms outlines both diatom-specific and species-specific features of the photosynthetic machinery. J. Proteome Res. **2011**, 10, 5338–5353.
(261) Nunn, B. L.; Aker, J. R.; Shaffer, S. A.; Tsai, S.; Strzepek, R. F.; Boyd, P. W.; Freeman, T. L.; Brittnacher, M.; Malmström, L.; Goodlett, D. R. Deciphering diatom biochemical pathways via whole-cell proteomics. Aquat Microb Ecol **2009**, 55, 241–253.
(262) Frigeri, L. G.; Radabaugh, T. R.; Haynes, P. A.; Hildebrand, M. Identification of proteins from a cell wall fraction of the diatom Thalassiosira pseudonana: insights into silica structure formation. Mol. Cell Proteomics **2006**, 5, 182–193.
(263) Dong, H.-P.; Williams, E.; Wang, D.; Xie, Z.-X.; Hsia, R.; Jenck, A.; Halden, R.; Li, J.; Chen, F.; Place, A. R. Responses of Nannochloropsis oceanica IMET1 to Long-Term Nitrogen Starvation and Recovery. Plant Physiol. **2013**, 162, 1110–1126.
(264) Kim, Y. K.; Yoo, W. I.; Lee, S. H.; Lee, M. Y. Proteomic Analysis of Cadmium-Induced Protein Profile Alterations from Marine Alga Nannochloropsis oculata. Ecotoxicology **2005**, 14, 589–596.
(265) Yang, Z.-K.; Ma, Y.-H.; Zheng, J.-W.; Yang, W.-D.; Liu, J.-S.; Li, H.-Y. Proteomics to reveal metabolic network shifts towards lipid accumulation following nitrogen deprivation in the diatom Phaeodactylum tricornutum . J Appl Phycol **2013**, 1–10.
(266) McKew, B. A.; Lefebvre, S. C.; Achterberg, E. P.; Metodieva, G.; Raines, C. A.; Metodiev, M. V.; Geider, R. J. Plasticity in the proteome of

Emiliania huxleyi CCMP 1516 to extremes of light is highly targeted. New Phytologist **2013**, 200, 61–73.
(267) Jones, B. M.; Iglesias-Rodriguez, M. D.; Skipp, P. J.; Edwards, R. J.; Greaves, M. J.; Young, J. R.; Elderfield, H.; O'Connor, C. D. Responses of the Emiliania huxleyi proteome to ocean acidification. PLoS ONE **2013**, 8, e61868.
(268) Jones, B. M.; Edwards, R. J.; Skipp, P. J.; O'Connor, C. D.; Iglesias-Rodriguez, M. D. Shotgun Proteomic Analysis of Emiliania huxleyi, a Marine Phytoplankton Species of Major Biogeochemical Importance. Marine Biotechnology **2011**, 13, 496–504.
(269) Yoshida, M.; Yoshida, Y.; Fujiwara, T.; Misumi, O.; Kuroiwa, H.; Kuroiwa, T. Proteomic comparison between interphase and metaphase of isolated chloroplasts of Cyanidioschyzon merolae (Cyanidiophyceae, Rhodophyta). Phycological Research **2011**, 59, 1–15.
(270) Suzuki, K.; Miyagishima, S.-Y. Eukaryotic and Eubacterial Contributions to the Establishment of Plastid Proteome Estimated by Large-Scale Phylogenetic Analyses. Molecular Biology and Evolution **2010**, 27, 581–590.
(271) Li, C.; Zhang, Y.; Xie, Z.-X.; He, Z.-P.; Lin, L.; Wang, D.-Z. Quantitative proteomic analysis reveals evolutionary divergence and species-specific peptides in the Alexandrium tamarense complex (Dinophyceae). Journal of Proteomics **2013**, 86, 85–96.
(272) Wang, D.-Z.; Li, C.; Zhang, Y.; Wang, Y.-Y.; He, Z.-P.; Lin, L.; Hong, H.-S. Quantitative proteomic analysis of differentially expressed proteins in the toxicity-lost mutant of Alexandrium catenella (Dinophyceae) in the exponential phase. Journal of Proteomics **2012**, 75, 5564–5577.
(273) Wang, D.-Z.; Lin, L.; Chan, L. L.; Hong, H.-S. Comparative studies of four protein preparation methods for proteomic study of the dinoflagellate Alexandrium sp. using two-dimensional electrophoresis. Harmful Algae **2009**, 8, 685–691.
(274) Harris, E. H. Chlamydomonas as a Model Organism. Annu. Rev. Plant Physiol. Plant Mol. Biol. **2001**, 52, 363–406.
(275) Gorman, D. S.; Levine, R. P. Cytochrome F and Plastocyanin: Their Sequence in the Photosynthetic Electron Transport Chain of Chlamydomonas reinhardi. PNAS **1965**, 54, 1665–1669.
(276) Hutner, S. H.; Provasoli, L.; Schatz, A.; Haskins, C. P. Some Approaches to the Study of the Role of Metals in the Metabolism of Microorganisms. Proceedings of the American Philosophical Society **1950**, 94, 152–170.
(277) Wellburn, R. W. The spectral determination of chlorophylls a and b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. J. Plant Physiol. **1994**, 144, 307–313.
(278) Gerhardt, P.; Murray, R. G. E.; Wood, W. A.; Krieg, N. R. Methods for general and molecular bacteriology; American Society for Microbiology, 1994; Vol. 40.

(279) Chen, W.; Zhang, C.; Song, L.; Sommerfeld, M.; Hu, Q. A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae. J. Microbiol. Methods **2009**, 77, 41–47.
(280) UniProt http://www.uniprot.org/ (accessed Jul 17, 2013).
(281) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. J. Proteome Res. **2004**, 3, 958–964.
(282) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. Bioinformatics **2004**, 20, 1466–1467.
(283) Vaudel, M.; Barsnes, H.; Berven, F. S.; Sickmann, A.; Martens, L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Proteomics **2011**, 11, 996–999.
(284) Barsnes, H.; Vaudel, M.; Colaert, N.; Helsens, K.; Sickmann, A.; Berven, F. S.; Martens, L. compomics-utilities: an open-source Java library for computational proteomics. BMC Bioinformatics **2011**, 12, 70.
(285) Muller, J.; Szklarczyk, D.; Julien, P.; Letunic, I.; Roth, A.; Kuhn, M.; Powell, S.; von Mering, C.; Doerks, T.; Jensen, L. J.; et al. eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. Nucleic Acids Res. **2010**, 38, D190–D195.
(286) Tatusov, R. L.; Koonin, E. V.; Lipman, D. J. A genomic perspective on protein families. Science **1997**, 278, 631–637.
(287) DAVID: Functional Annotation Result Summary http://david.abcc.ncifcrf.gov/summary.jsp (accessed Jul 17, 2013).
(288) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protocols **2008**, 4, 44–57.
(289) KEGG Mapper – Search&Color Pathway http://www.genome.jp/kegg/tool/map_pathway2.html (accessed Jul 16, 2013).
(290) Westermeier, R.; Naven, T.; Höpker, H.-R. Proteomics in Practice: A Guide to Successful Experimental Design; John Wiley & Sons, 2008.
(291) Steen, H.; Mann, M. The abc's (and xyz's) of peptide sequencing. Nat Rev Mol Cell Biol **2004**, 5, 699–711.
(292) Dančík, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De Novo Peptide Sequencing via Tandem Mass Spectrometry. Journal of Computational Biology **1999**, 6, 327–342.
(293) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. **1984**, 11, 601.
(294) Keller, A.; Eng, J.; Zhang, N.; Li, X.; Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol **2005**, 1.
(295) Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpp, A.; Neumann, S.; Pizarro, A. D.; et al.

mzML—a Community Standard for Mass Spectrometry Data. Mol Cell Proteomics **2011**, 10.

(296) Mascot database search | Data file format for mass spectrometry peak lists http://www.matrixscience.com/help/data_file_help.html (accessed Sep 8, 2013).

(297) Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A. 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.

(298) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. **2011**, 10, 1794–1805.

(299) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom **1994**, 5, 976–989.

(300) Colinge, J.; Masselot, A.; Giron, M.; Dessingy, T.; Magnin, J. OLAV: towards high-throughput tandem mass spectrometry data identification. Proteomics **2003**, 3, 1454–1463.

(301) Colinge, J.; Masselot, A.; Cusin, I.; Mahé, E.; Niknejad, A.; Argoud-Puy, G.; Reffas, S.; Bederr, N.; Gleizes, A.; Rey, P.-A.; et al. High-performance peptide identification by tandem mass spectrometry allows reliable automatic data processing in proteomics. Proteomics **2004**, 4, 1977–1984.

(302) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotech **2012**, 30, 918–920.

(303) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. Mol Cell Proteomics **2012**, 11.

(304) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **1999**, 20, 3551–3567.

(305) Scopus http://www.scopus.com/home.url;jsessionid=7D8BD6BA49A8CCB84654DB180577FFD7.FZg2ODcJC9ArCe8WOZPvA (accessed Sep 7, 2013).

(306) Noirel, J.; Evans, C.; Salim, M.; Mukherjee, J.; Yen Ow, S.; Pandhal, J.; Khoa Pham, T.; A. Biggs, C.; C. Wright, P. Methods in Quantitative Proteomics: Setting iTRAQ on the Right Track. Current Proteomics **2011**, 8, 17–30.

(307) GeneBio : Phenyx http://www.genebio.com/products/phenyx/ (accessed Sep 7, 2013).

(308) STRING: functional protein association networks http://string-db.org/newstring_cgi/show_input_page.pl?UserId=eOHsWyLhfEER&sessionId=NNjO5FESuvNG&info_box_type_input_page=general (accessed Sep 7, 2013).
(309) Searle, B. C. Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. PROTEOMICS **2010**, 10, 1265–1269.
(310) Searle, B. C.; Turner, M.; Nesvizhskii, A. I. Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. J. Proteome Res. **2008**, 7, 245–253.
(311) Slotta, D. J.; McFarland, M. A.; Markey, S. P. MassSieve: panning MS/MS peptide data for proteins. Proteomics **2010**, 10, 3035–3039.
(312) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol. Cell Proteomics **2011**, 10, M111.007690.
(313) Edwards, N. J. PepArML: A Meta-Search Peptide Identification Platform for Tandem Mass Spectra: PepArML: A Meta-Search Peptide Identification Platform. In Current Protocols in Bioinformatics; Bateman, A.; Pearson, W. R.; Stein, L. D.; Stormo, G. D.; Yates, J. R., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; pp. 13.23.1–13.23.23.
(314) Kwon, T.; Choi, H.; Vogel, C.; Nesvizhskii, A. I.; Marcotte, E. M. MSblender: A probabilistic approach for integrating peptide identifications from multiple database search engines. J. Proteome Res. **2011**, 10, 2949–2958.
(315) Chen, Y.-Y.; Dasari, S.; Ma, Z.-Q.; Vega-Montoto, L. J.; Li, M.; Tabb, D. L. Refining comparative proteomics by spectral counting to account for shared peptides and multiple search engines. Anal Bioanal Chem **2012**, 404, 1115–1125.
(316) Jones, A. R.; Siepen, J. A.; Hubbard, S. J.; Paton, N. W. Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines. PROTEOMICS **2009**, 9, 1220–1229.
(317) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J. Proteome Res. **2008**, 7, 29–34.
(318) R Development Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2011.
(319) Arntzen, M. Ø.; Koehler, C. J.; Barsnes, H.; Berven, F. S.; Treumann, A.; Thiede, B. IsobariQ: software for isobaric quantitative proteomics using IPTL, iTRAQ, and TMT. J. Proteome Res. **2011**, 10, 913–920.
(320) Schwacke, J. H.; Hill, E. G.; Krug, E. L.; Comte-Walters, S.; Schey, K. L. iQuantitator: a tool for protein expression inference using iTRAQ. BMC Bioinformatics **2009**, 10, 342.

(321) Huber, W. Introduction to robust calibration and variance stabilisation with VSN. **2012**.
(322) Lilley, K. S.; Deery, M. J.; Gatto, L. Challenges for proteomics core facilities. PROTEOMICS **2011**, 11, 1017–1025.
(323) Karp, N. A.; Lilley, K. S. Design and Analysis Issues in Quantitative Proteomics Studies. PROTEOMICS **2007**, 7, 42–50.
(324) Molloy, M. P.; Brzezinski, E. E.; Hang, J.; McDowell, M. T.; VanBogelen, R. A. Overcoming technical variation and biological variation in quantitative proteomics. PROTEOMICS **2003**, 3, 1912–1919.
(325) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem **2007**, 389, 1017–1031.
(326) Shteynberg, D.; Nesvizhskii, A. I.; Moritz, R. L.; Deutsch, E. W. Combining results of multiple search engines in proteomics. Mol. Cell Proteomics **2013**, 12, 2383–2393.
(327) Gallagher, B. J. The economics of producing biodiesel from algae. Renewable Energy **2011**, 36, 158–162.
(328) Merchant, S. S.; Prochnik, S. E.; Vallon, O.; Harris, E. H.; Karpowicz, S. J.; Witman, G. B.; Terry, A.; Salamov, A.; Fritz-Laylin, L. K.; Maréchal-Drouard, L.; et al. The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. Science **2007**, 318, 245–250.
(329) Esquível, M. G.; Amaro, H. M.; Pinto, T. S.; Fevereiro, P. S.; Malcata, F. X. Efficient H2 production via Chlamydomonas reinhardtii. Trends Biotechnol. **2011**, 29, 595–600.
(330) Vijayaragh, K.; Karthik, R.; Kamala Nal, S. P. Hydrogen Generation from Algae: A Review. J. Plant Sci. **2010**, 5, 1–19.
(331) Moellering, E. R.; Benning, C. RNA Interference Silencing of a Major Lipid Droplet Protein Affects Lipid Droplet Size in Chlamydomonas reinhardtii. Eukaryotic Cell **2010**, 9, 97–106.
(332) Miller, R.; Wu, G.; Deshpande, R. R.; Vieler, A.; Gärtner, K.; Li, X.; Moellering, E. R.; Zäuner, S.; Cornish, A. J.; Liu, B.; et al. Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. Plant Physiol. **2010**, 154, 1737–1752.
(333) Jinkerson, R. E.; Subramanian, V.; Posewitz, M. C. Improving biofuel production in phototrophic microorganisms with systems biology. Biofuels **2011**, 2, 125–144.
(334) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of Two-Dimensional Gel Electrophoresis-Based Proteome Analysis Technology. PNAS **2000**, 97, 9390–9395.
(335) Wu, W. W.; Wang, G.; Baek, S. J.; Shen, R.-F. Comparative Study of Three Proteomic Quantitative Methods, DIGE, cICAT, and iTRAQ, Using 2D Gel- or LC−MALDI TOF/TOF. J. Proteome Res. **2006**, 5, 651–658.

(336) Siaut, M.; Cuiné, S.; Cagnon, C.; Fessler, B.; Nguyen, M.; Carrier, P.; Beyly, A.; Beisson, F.; Triantaphylidès, C.; Li-Beisson, Y.; et al. Oil accumulation in the model green alga Chlamydomonas reinhardtii: characterization, variability between common laboratory strains and relationship with starch reserves. BMC Biotechnol. **2011**, 11, 7.
(337) Neupert, J.; Shao, N.; Lu, Y.; Bock, R. Genetic transformation of the model green alga Chlamydomonas reinhardtii. Methods Mol. Biol. **2012**, 847, 35–47.
(338) Scholz, M.; Hoshino, T.; Johnson, D.; Riley, M. R.; Cuello, J. Flocculation of wall-deficient cells of Chlamydomonas reinhardtii mutant cw15 by calcium and methanol. Biomass Bioenergy **2011**, 35, 4835–4840.
(339) Ladygin, V. G.; Boutanaev, A. M. Transformation of Chlamydomonas reinhardtii CW-15 with the Hygromycin Phosphotransferase Gene as a Selectable Marker. Russ. J. Genet. **2002**, 38, 1009–1014.
(340) Bonente, G.; Pippa, S.; Castellano, S.; Bassi, R.; Ballottari, M. Acclimation of Chlamydomonas reinhardtii to different growth irradiances. J. Biol. Chem. **2012**, 287, 5833–5847.
(341) Chankova, S. G.; Yurina, N. Micro-algae as a Model System for Studying of Genotype Resistance to Oxidative Stress and Adaptive Response. In Radiobiology and Environmental Security; Mothersill, C. E.; Korogodina, V.; Seymour, C. B., Eds.; NATO Science for Peace and Security Series C: Environmental Security; Springer Netherlands, 2012; pp. 19–30.
(342) Davies, D. R.; Plaskitt, A. Genetical and structural analyses of cell-wall formation in Chlamydomonas reinhardtii. Genet. Res. **1971**, 17, 33–43.
(343) Harris, E. H.; Stern, D. B.; Witman, G. The Chlamydomonas Sourcebook: Introduction to Chlamydomonas and its laboratory use; Academic Press: San Diego, CA, 2009.
(344) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Wright, P. C. Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. Proteomics **2011**, 11, 2341–2346.
(345) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly."J. Proteome Res. **2009**, 8, 5347–5355.
(346) Wang, Y.; Spalding, M. H. An inorganic carbon transport system responsible for acclimation specific to air levels of CO2 in Chlamydomonas reinhardtii. PNAS **2006**, 103, 10110–10115.
(347) Miura, K.; Yamano, T.; Yoshioka, S.; Kohinata, T.; Inoue, Y.; Taniguchi, F.; Asamizu, E.; Nakamura, Y.; Tabata, S.; Yamato, K. T.; et al. Expression Profiling-Based Identification of CO_2-Responsive Genes Regulated by CCM1 Controlling a Carbon-Concentrating Mechanism in Chlamydomonas reinhardtii. Plant Physiol. **2004**, 135, 1595–1607.
(348) Abe, J.; Kubo, T.; Takagi, Y.; Saito, T.; Miura, K.; Fukuzawa, H.; Matsuda, Y. The transcriptional program of synchronous

gametogenesis in Chlamydomonas reinhardtii. Curr. Genet. **2004**, 46, 304–315.
(349) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **2000**, 28, 27–30.
(350) Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. **2012**, 40, D109–D114.
(351) Lawrence, E. Henderson's dictionary of biology; Prentice Hall, 2005.
(352) Tittel, J.; Bissinger, V.; Zippel, B.; Gaedke, U.; Bell, E.; Lorke, A.; Kamjunke, N. Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs. PNAS **2003**, 100, 12776–12781.
(353) Heifetz, P. B.; Förster, B.; Osmond, C. B.; Giles, L. J.; Boynton, J. E. Effects of Acetate on Facultative Autotrophy inChlamydomonas reinhardtii Assessed by Photosynthetic Measurements and Stable Isotope Analyses. Plant Physiol. **2000**, 122, 1439–1446.
(354) Tirichine, L.; Bowler, C. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. The Plant Journal **2011**, 66, 45–57.
(355) Maheswari, U.; Jabbari, K.; Petit, J.-L.; Porcel, B. M.; Allen, A. E.; Cadoret, J.-P.; De Martino, A.; Heijde, M.; Kaas, R.; La Roche, J.; et al. Digital expression profiling of novel diatom transcripts provides insight into their biological functions. Genome Biology **2010**, 11.
(356) Wilson, J. H. The food value of Phaeodactylum tricornutum Bohlin to the larvae of Ostrea edulis L. and Crassostrea gigas Thunberg. Aquaculture **1978**, 13, 313–323.
(357) Bowler, C.; De Martino, A.; Falciatore, A. Diatom cell division in an environmental context. Current Opinion in Plant Biology **2010**, 13, 623–630.
(358) Gleick, P. H.; Palaniappan, M. Peak water limits to freshwater withdrawal and use. PNAS **2010**, 107, 11155–11162.
(359) Yongmanitchai, W.; Ward, O. P. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. Appl. Environ. Microbiol. **1991**, 57, 419–425.
(360) Hildebrand, M.; Davis, A. K.; Smith, S. R.; Traller, J. C.; Abbriano, R. The place of diatoms in the biofuels industry. Biofuels **2012**, 3, 221–240.
(361) Song, M.; Pei, H.; Hu, W.; Ma, G. Evaluation of the potential of 10 microalgal strains for biodiesel production. Bioresource Technology **2013**, 141, 245–251.
(362) Armbrust, E. V.; Berges, J. A.; Bowler, C.; Green, B. R.; Martinez, D.; Putnam, N. H.; Zhou, S.; Allen, A. E.; Apt, K. E.; Bechner, M.; et al. The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. Science **2004**, 306, 79–86.

(363) Osborn, H. L.; Hook, S. E. Using transcriptomic profiles in the diatom Phaeodactylum tricornutum to identify and prioritize stressors. Aquatic Toxicology **2013**, 138-139, 12–25.
(364) Yang, Z.-K.; Niu, Y.-F.; Ma, Y.-H.; Xue, J.; Zhang, M.-H.; Yang, W.-D.; Liu, J.-S.; Lu, S.-H.; Guan, Y.; Li, H.-Y. Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. Biotechnology for Biofuels **2013**, 6.
(365) Valenzuela, J.; Mazurie, A.; Carlson, R. P.; Gerlach, R.; Cooksey, K. E.; Peyton, B. M.; Fields, M. W. Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. Biotechnol Biofuels **2012**, 5, 40.
(366) Nunn, B. L.; Ting, Y. S.; Malmström, L.; Tsai, Y. S.; Squier, A.; Goodlett, D. R.; Harvey, H. R. The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation. Limnology and Oceanography **2010**, 55, 1790–1804.
(367) Caron, D. A.; Sanders, R. W.; Lim, E. L.; Marrasé, C.; Amaral, L. A.; Whitney, S.; Aoki, R. B.; Porters, K. G. Light-dependent phagotrophy in the freshwater mixotrophic chrysophyte Dinobryon cylindricum. Microb Ecol **1993**, 25, 93–111.
(368) Carvalho, W. F.; Granéli, E. Contribution of phagotrophy versus autotrophy to Prymnesium parvum growth under nitrogen and phosphorus sufficiency and deficiency. Harmful Algae **2010**, 9, 105–115.
(369) Schmidt, S.; Raven, J. A.; Paungfoo-Lonhienne, C. The mixotrophic nature of photosynthetic plants. Functional Plant Biology **2013**, 40, 425–438.
(370) Lamb, C. A.; Dooley, H. C.; Tooze, S. A. Endocytosis and autophagy: Shared machinery for degradation. Bioessays **2013**, 35, 34–45.
(371) Shui, W.; Sheu, L.; Liu, J.; Smart, B.; Petzold, C. J.; Hsieh, T.; Pitcher, A.; Keasling, J. D.; Bertozzi, C. R. Membrane proteomics of phagosomes suggests a connection to autophagy. PNAS **2008**, 105, 16952–16957.
(372) Rismani-Yazdi, H.; Haznedaroglu, B. Z.; Hsin, C.; Peccia, J. Transcriptomic analysis of the oleaginous microalga Neochloris oleoabundans reveals metabolic insights into triacylglyceride accumulation. Biotechnology for Biofuels **2012**, 5, 74.
(373) Burrows, E. H.; Bennette, N. B.; Carrieri, D.; Dixon, J. L.; Brinker, A.; Frada, M.; Baldassano, S. N.; Falkowski, P. G.; Dismukes, G. C. Dynamics of Lipid Biosynthesis and Redistribution in the Marine Diatom Phaeodactylum tricornutum Under Nitrate Deprivation. Bioenerg. Res. **2012**, 5, 876–885.
(374) Candan, N.; Tarhan, L. The correlation between antioxidant enzyme activities and lipid peroxidation levels in Mentha pulegium organs grown in Ca2+, Mg2+, Cu2+, Zn2+ and Mn2+ stress conditions. Plant Science **2003**, 165, 769–776.
(375) Brzezinski, M. A.; Olson, R. J.; Chisholm, S. W. Silicon availability and cell-cycle progression in marine diatoms. Marine ecology progress series. Oldendorf **1990**, 67, 83–96.

(376) Sapriel, G.; Quinet, M.; Heijde, M.; Jourdren, L.; Tanty, V.; Luo, G.; Le Crom, S.; Lopez, P. J. Genome-Wide Transcriptome Analyses of Silicon Metabolism in Phaeodactylum tricornutum Reveal the Multilevel Regulation of Silicic Acid Transporters. PLoS ONE **2009**, 4, e7458.
(377) Msanne, J.; Xu, D.; Konda, A. R.; Casas-Mollano, J. A.; Awada, T.; Cahoon, E. B.; Cerutti, H. Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae Chlamydomonas reinhardtii and Coccomyxa sp. C-169. Phytochemistry **2012**, 75, 50–59.
(378) Huang, S.; Chen, L.; Te, R.; Qiao, J.; Wang, J.; Zhang, W. Complementary iTRAQ proteomics and RNA-seq transcriptomics reveal multiple levels of regulation in response to nitrogen starvation in Synechocystis sp. PCC 6803. Mol. BioSyst. **2013**, 9, 2565–2574.
(379) Courant, F.; Martzolff, A.; Rabin, G.; Antignac, J.-P.; Bizec, B. L.; Giraudeau, P.; Tea, I.; Akoka, S.; Couzinet, A.; Cogne, G.; et al. How metabolomics can contribute to bio-processes: a proof of concept study for biomarkers discovery in the context of nitrogen-starved microalgae grown in photobioreactors. Metabolomics **2013**, 9, 1286–1300.
(380) Su, S. S.; Tanaka, Y.; Samejima, I.; Tanaka, K.; Yanagida, M. A nitrogen starvation-induced dormant G0 state in fission yeast: the establishment from uncommitted G1 state and its delay for return to proliferation. Journal of Cell Science **1996**, 109, 1347–1357.
(381) Stryer, L.; Berg, J. M.; Tymoczko, J. L. Biochemistry, 5th Ed.; 5th Revised edition.; W.H.Freeman & Co Ltd, 2002.

Appendices

Chapter 4 Appendices

Appendix 4.1: R script of for the merger of output of ProteinPilot, Phenyx, Peptide shaker and Peaks. Script merges the search engine outputs by spectra before removing spectra with disagreement between the search engines. The raw iTRAQ labels derived from the .mgf is also merged with the spectra before preforming VSN on the labels and reformatting the output for uTRAQ.

########standard starting lines###########

```
rm (list=ls())
library(calibrate)
library(XLConnect)
library(gdata)
par(mfrow=c(1,1))
stringsAsFactors=FALSE
#####################################data
input#################################
setwd("C:/Users/Joseph Longworth/Dropbox/Chapter 4 search optimisation/Merger of
pp,ph,ps/")
phenyxids <- readWorksheet(loadWorkbook("Cr_
CSource_mgf_analyst_0,3_0,1_FDR0,03.xls"),sheet=2)#,
startRow=1,endRow=200)#,startCol=1,endCol=8)
# phenyxids<-phenyxids[1,]
ppilotids<-read.csv("JL_Cr_CSource_77_frac_wiff_PeptideSummary_truncated_v3.csv")
# ppilotids<-ppilotids[1,]
pshakerids <- readWorksheet(loadWorkbook("psm export of m,o,x.xlsx"),sheet=1)#,
startRow=1,endRow=200)#,startCol=1,endCol=2)
# pshakerids<-pshakerids[1,]
peaksids <- read.csv("DB search psm.csv")
# peaksids<-peaksids[1,]
fileorder<-  readWorksheet(loadWorkbook("filename.xls"),sheet=1,header=FALSE,startCol=1)
 fileorder<- na.omit(fileorder)
  Reporterions <-read.delim("MSMSid_24-04-2013_204642_v2.csv",sep=",",header=F)
colnames(Reporterions)=c("msmsid","i113","i114","i115","i116","i117","i118","i119","i121")
setwd("C:/Users/Joseph Longworth/Desktop/R/")

###############pshakerids <- read.table("search engine export of
m,o,x.txt",sep="\t",header=T)
#############################################################
############################
###########Remove low cofidence and non contrib
psm#########################

ppilotids<-ppilotids[ppilotids$Conf>94.99,]
ppilotids<-ppilotids[order(ppilotids$Contrib,decreasing=T),]
ppilotids<-ppilotids[!duplicated(ppilotids$Spectrum),]

ACcol<-as.numeric(which( colnames(ppilotids)=="Accessions" ))
SEQcol<-as.numeric(which( colnames(ppilotids)=="Sequence" ))
SEQAC<-c()
for(i in seq(nrow(ppilotids))){
 SEQAC1<-
paste(c(as.character(ppilotids[i,SEQcol]),as.character(ppilotids[i,ACcol])),collapse="|")
```

```
  SEQAC<-c(SEQAC,SEQAC1)
}
ppilotids<-cbind(ppilotids,SEQAC,c(1:nrow(ppilotids)))
ppilotids2<-ppilotids[!duplicated(ppilotids$SEQAC),]
ppilotids3<-ppilotids2[ppilotids2$Cont ==0,]
faulty.spectra<-c()
SEQACcol<-which( colnames(ppilotids)=="SEQAC")
for(i in seq(nrow(ppilotids))){
  if(ppilotids[i,SEQACcol]%in%ppilotids3$SEQAC) {a<-1}else{a<-0}
  faulty.spectra<-c(faulty.spectra,a)
}
ppilotids<-cbind(ppilotids,faulty.spectra)
ppilotids<-ppilotids[ppilotids$faulty.spectra==0,]
ppilotids<-ppilotids[,1:57]
############################################################
###############
####################parse sequence for
peaks#########################
SEQpk <- c()
seqcol<-which( colnames(peaksids)=="Peptide" )
for(i in seq(nrow(peaksids))){
  seqpk1<-strsplit(as.character(peaksids[i,seqcol]), "[()]")
  seqpk2<-unlist(seqpk1)
  seqpk3<-as.list(seqpk2)
seqpk5<-c()
for(n in seq(length(seqpk3))){
seqpk4<-pmatch(x="+",seqpk3[n],0)
seqpk5<-c(seqpk5,seqpk4)}
seqpk6<-seqpk3[!seqpk5==1]
seqpk7<-paste(seqpk6,collapse="")
SEQpk<-c(SEQpk,seqpk7)}
peaksids<-cbind(peaksids,SEQpk)
##################re-name sequnce
columns#####################################
colnames(phenyxids) [which(colnames(phenyxids)  == "sequence")] = "SEQph"
colnames(ppilotids) [which(colnames(ppilotids)  == "Sequence")] = "SEQpp"
colnames(pshakerids)[which(colnames(pshakerids) == "Sequence")] = "SEQps"
############################################################
####################
#############re-parse the msmsid for protein pilot#############
msmsidpp <- c()
specppcol<-which( colnames(ppilotids)=="Spectrum" )
for(i in seq(nrow(ppilotids))){
  msmsidpp1<-strsplit(as.character(ppilotids[i,specppcol]), "[.]")
  msmsidpp2<-unlist(msmsidpp1)
  msmsidpp3<-paste(msmsidpp2[1],msmsidpp2[4],msmsidpp2[5],sep=".")
  msmsidpp <- c(msmsidpp, msmsidpp3)
}
ppilotids<-cbind(ppilotids,msmsidpp)
############################################################
####################
#################re-parse the msmsid for phenyx######
specphcol<-which( colnames(phenyxids)=="spectra.descr." )
msmsidph <- c()
for(i in seq(nrow(phenyxids))){
```

```
msmsidph1<-unlist(strsplit(as.character(phenyxids[i,specphcol]), "[,]"))
msmsidph2<-unlist(strsplit(msmsidph1[5], "[\ ]"))
msmsidph3<-unlist(strsplit(msmsidph2[5], "[)]"))
msmsidph4<-fileorder[which(fileorder==msmsidph1[1]),2]
msmsidph5<-unlist(strsplit(msmsidph2[3],"[-]"))
msmsidph6<-paste(msmsidph4,msmsidph5[1],msmsidph3,sep=".")
msmsidph<- c(msmsidph, msmsidph6)
}
phenyxids<-cbind(phenyxids,msmsidph)#,stringsAsFactors=FALSE)
##########################################################
##############
#################re-parse the msmsid for peptide shaker######
specpscol<-which( colnames(pshakerids)=="Spectrum.Title" )
msmsidps <- c()
for(i in seq(nrow(pshakerids))){
  msmsidps1<-unlist(strsplit(as.character(pshakerids[i,specpscol]), "[,]"))
  msmsidps2<-unlist(strsplit(msmsidps1[5], "[\ ]"))
  msmsidps3<-unlist(strsplit(msmsidps2[5], "[)]"))
  msmsidps4<-fileorder[which(fileorder==msmsidps1[1]),2]
  msmsidps5<-unlist(strsplit(msmsidps2[3],"[-]"))
  msmsidps6<-paste(msmsidps4,msmsidps5[1],msmsidps3,sep=".")
  msmsidps<- c(msmsidps, msmsidps6)
}
pshakerids<-cbind(pshakerids,msmsidps)#,stringsAsFactors=FALSE)
##########################################################
##############
#################re-parse the msmsid for peaks shaker######
ms.scan.number <-c(1:nrow(Reporterions))
   Reporterions1<-cbind(Reporterions,ms.scan.number)
Reporterions1<-Reporterions1[,c(1,10)]
msmsidpk1<-as.data.frame(strsplit(as.character(peaksids$Scan),"[:]"))
msmsidpk2<-t(msmsidpk1)
ms.scan.number<-msmsidpk2[,2]
peaksids<-cbind(peaksids,ms.scan.number)
peaksids <-peaksids
[!duplicated(peaksids$ms.scan.number)|duplicated(peaksids$ms.scan.number,fromLast=T),]
peaksids<-
merge(peaksids,Reporterions1,by.x="ms.scan.number",by.y="ms.scan.number",all.x=TRUE,all
.y=F)
colnames(peaksids) [which(colnames(peaksids)  == "msmsid")] = "msmsidpk"
##########################################################
#######
##########remove duplicate msmsid assignments###########
ppilotids <-ppilotids
[!duplicated(ppilotids$msmsidpp)|duplicated(ppilotids$msmsidpp,fromLast=T),]
pshakerids<-
pshakerids[!duplicated(pshakerids$msmsidps)|duplicated(pshakerids$msmsidps,fromLast=
T),]
phenyxids <-phenyxids
[!duplicated(phenyxids$msmsidph)|duplicated(phenyxids$msmsidph,fromLast=T),]
peaksids  <-peaksids
[!duplicated(peaksids$msmsidpk)|duplicated(peaksids$msmsidpk,fromLast=T),]
###########merge together by msmsid#########################
mergeddata<-
merge(phenyxids,ppilotids,by.x="msmsidph",by.y="msmsidpp",all.x=TRUE,all.y=TRUE)
```

```
colnames(mergeddata)[which(colnames(mergeddata) == "Accessions")] = "ACpp"
mergeddata<-
merge(mergeddata,pshakerids,by.x="msmsidph",by.y="msmsidps",all.x=T,all.y=T)
mergeddata<- merge(mergeddata,peaksids,by.x="msmsidph",by.y="msmsidpk",all.x=T,all.y=T)
mergeddata<-
merge(mergeddata,Reporterions,by.x="msmsidph",by.y="msmsid",all.x=T,all.y=F)
colnames(mergeddata)[which(colnames(mergeddata) == "AC")] = "PHACfull"
##################################################################
###
#################remove matches with different sequence
assignment###############
SEQphcol<-as.numeric(which( colnames(mergeddata)=="SEQph" ))
SEQppcol<-as.numeric(which( colnames(mergeddata)=="SEQpp" ))
SEQpscol<-as.numeric(which( colnames(mergeddata)=="SEQps" ))
SEQpkcol<-as.numeric(which( colnames(mergeddata)=="SEQpk" ))

#FUN1 Returns a binary comparison that two sets of sequenes if present match
#inputs are FUN1(col#1,col#2)
FUN1<-function(a,b){
 i<-1
 passed <- c()
 for(i in seq(nrow(mergeddata))){
  if(is.na(mergeddata[i,a])) {passed <- c(passed,1)
  }   else if(is.na(mergeddata[i,b])){passed <- c(passed,1)
  }       else if(as.character(mergeddata[i,a])==as.character(mergeddata[i,b])){passed <-
c(passed,1)
  }              else{  passed<-c(passed,0)}}
 return(passed)}

phvspp<-FUN1(SEQphcol,SEQppcol)
phvsps<-FUN1(SEQphcol,SEQpscol)
phvspk<-FUN1(SEQphcol,SEQpkcol)

ppvsps<-FUN1(SEQppcol,SEQpscol)
ppvspk<-FUN1(SEQppcol,SEQpkcol)

psvspk<-FUN1(SEQpscol,SEQpkcol)

match.seq.pass<- phvspp*phvsps*phvspk*ppvsps*ppvspk* psvspk

mergeddata<-cbind(mergeddata,match.seq.pass)
mergeddata<-mergeddata[!(mergeddata$match.seq.pass==0),]
#################merge AC differently#################
ACphcol<-which( colnames(mergeddata)=="PHACfull" )
ACppcol<-which( colnames(mergeddata)=="ACpp" )
ACpscol<-which( colnames(mergeddata)=="Protein.s." )
ACpkcol<-which( colnames(mergeddata)=="Accession" )

ACunique<-c()
AC<-c()
AC.others<-c()
for(i in seq(nrow(mergeddata))){
 #i<-2172
 ACph1<-strsplit(mergeddata[i,ACphcol], "[|]")
 ACph2<-unlist(ACph1)
```

```
 ACph3<-paste(ACph2[2])
 ACpp1<-strsplit(as.character(mergeddata[i,ACppcol]), "[|]")
 ACpp2<-unlist(ACpp1)
 ACpp3<-paste(ACpp2[2])
 ACps1<-strsplit(as.character(mergeddata[i,ACpscol]), "[,]")
 ACps2<-unlist(ACps1)
 ACpk1<-strsplit(as.character(mergeddata[i,ACpkcol]), "[|]")
 ACpk2<-unlist(ACpk1)
 if(length(ACpk2)==3){ACpk3<-paste(ACpk2[2])}else{ACpk3<-paste(ACpk2[1])}
 AC1<-trim(c(ACps2,ACph3,ACpp3,ACpk3))
 AC2<-na.omit(AC1)
 AC3<- AC2[AC2!="NA"]
if(all(AC3[1]==AC3)){ACunique1<-1
}else{ACunique1<-0}
  ACunique<-c(ACunique,ACunique1)
 AC4<-names(which.max(table(AC3)))
  AC<-c(AC,AC4)
  if(ACunique1==1){AC.others.1<-"NA"}else{AC.others.1<-
paste(trim(AC3[AC3!=AC4]),collapse="|")}
   AC.others<-c(AC.others,AC.others.1)
 }
AC<-trim(AC)

mergeddata<-cbind(mergeddata,AC,AC.others,ACunique)
#########################################################
########################Find # unique peptides per AC##########
##########create a combined list of SEQ###################
SEQ<-c()
i<-1
for(i in seq(nrow(mergeddata))){
  if(!is.na(mergeddata[i,SEQphcol])) {SEQ <- c(SEQ,as.character(mergeddata[i,SEQphcol]))
  }   else if(!is.na(mergeddata[i,SEQppcol])){SEQ <-
c(SEQ,as.character(mergeddata[i,SEQppcol]))
  }       else if(!is.na(mergeddata[i,SEQpscol])){SEQ <-
c(SEQ,as.character(mergeddata[i,SEQpscol]))
  }           else{ SEQ <- c(SEQ,as.character(mergeddata[i,SEQpkcol]))}}
mergeddata<-cbind(mergeddata,SEQ)
#########################################################################
###################
##############remove items with non unique AC############
mergeddata<-mergeddata[!(mergeddata$ACunique==0),]
#########################################################################
####
#################remove duplicated SEQ with miss matched
AC####################
ACcol<-as.numeric(which( colnames(mergeddata)=="AC" ))
SEQcol<-as.numeric(which( colnames(mergeddata)=="SEQ" ))
SEQAC<-c()
i<-1
for(i in seq(nrow(mergeddata))){
  SEQAC1<-
paste(c(as.character(mergeddata[i,SEQcol]),as.character(mergeddata[i,ACcol])),collapse="|")
  SEQAC<-c(SEQAC,SEQAC1)
}
mergeddata<-cbind(mergeddata,SEQAC)
```

```
mergeddata4<-mergeddata[!duplicated(mergeddata$SEQAC),]
mergeddata5<-
mergeddata4[duplicated(mergeddata4$SEQ)|duplicated(mergeddata4$SEQ,fromLast=TRUE),
]
if(nrow(mergeddata5)>0){
missmatch1<-as.character(mergeddata5$msmsidph)
missmatch3<- cbind(missmatch1,1)
mergeddata6<-
merge(mergeddata,missmatch3,by.x="msmsidph",by.y="missmatch1",all.x=T,all.y=T)
mergeddata<-mergeddata6[is.na(mergeddata6$V2),]}
##################################################################
####
#################remove duplicated SEQ's###################
mergeddata2<-mergeddata[!duplicated(mergeddata$SEQ),]
#################table of # of instances of each (primary)AC##########
uPEPtable<-as.data.frame(table(mergeddata2$AC))
uPEPtable2<-subset(uPEPtable,uPEPtable$Freq>1)
############link unique pep count to table#############
ACcol<-as.numeric(which( colnames(mergeddata)=="AC" ))
uPEP<-c()
i<-1
for(i in seq(nrow(mergeddata))){
  uPEP<-
c(uPEP,as.numeric(uPEPtable[which(uPEPtable==as.character(mergeddata[i,ACcol])),2]))
}
mergeddata<-cbind(mergeddata,uPEP)
##################################################################
########
############fomat for utraq#####################################
utraqcol<-
c("msmsidph","AC","SEQ","uPEP","AC.others","SEQAC","V2","ACunique","SEQps","SEQpp"
        ,"il13","il14","il15","il16","il17","il18","il19","il21","spectra.descr.","pm.key")

utraqcolnumber<-c()
for(i in seq(utraqcol)){
utraqcolnumber1<-as.numeric(which( colnames(mergeddata)==utraqcol[i] ))
utraqcolnumber<-c(utraqcolnumber,utraqcolnumber1)}
utraq.input<-mergeddata[,utraqcolnumber]

#write.csv(uPEPtable2,file="2pep.proteins_nonUAC_ph,pp,ps,pk.csv",row.names=F)
#write.csv(utraq.input,file="utraqinput_nonUAC_ph,pp,ps,pk.csv",row.names=F)
############################temp boot from mergeddata#########
# write.csv(mergeddata,file="mergeddata.csv")
# rm (list=ls())
# mergeddata<-read.csv("mergeddata.csv")
```

Appendix 4.2: R script used to conduct a t-test comparing any given set of iTRAQ labels associated to on condition to another given set of iTRAQ labels associated with another condition.

```
##########################################################

rm (list=ls())
library(calibrate)
library(XLConnect)
########################################data
input################################
setwd("C:/Users/Joseph Longworth/Desktop/R/")
utraq <- readWorksheet(loadWorkbook("utraqinput-quants.xls"),sheet=2,
startRow=2)#,startCol=1,endCol=8)
phenyx <- read.csv("mergeddata.csv")
############### remove "indeteminate" values##############
ind<-c()
for(i in seq(nrow(utraq))){
if(utraq[i,43]=="Indeterminate"|
  utraq[i,44]=="Indeterminate"|
  utraq[i,45]=="Indeterminate"|
  utraq[i,46]=="Indeterminate"|
  utraq[i,47]=="Indeterminate"|
  utraq[i,48]=="Indeterminate"|
  utraq[i,49]=="Indeterminate"|
  utraq[i,50]=="Indeterminate"){ind1<-1}else{ind1<-0}
ind<-c(ind,ind1)}
  utraq<-cbind(utraq,ind)
utraq<- utraq[!(utraq$ind==1),]
##########################################################
#####################
################select lable data#############
log113 <- log(as.numeric(utraq$il13.5),2)
log114 <- log(as.numeric(utraq$il14.5),2)
log115 <- log(as.numeric(utraq$il15.5),2)
log116 <- log(as.numeric(utraq$il16.5),2)
log117 <- log(as.numeric(utraq$il17.5),2)
log118 <- log(as.numeric(utraq$il18.5),2)
log119 <- log(as.numeric(utraq$il19.5),2)
log121 <- log(as.numeric(utraq$il21.5),2)

allcomparisons<-utraq[,1:2]
colnames(allcomparisons)[] = c("AC","Unique.Peptides")
##########################################################
######
######################data processing to detemine p-value and fold
change############
comp<-function(comparison.name,base.condition,compared.condition){

i<-1
biopval<- c()
for(i in seq(nrow(base.condition))){
  p <- t.test(base.condition[i,],compared.condition[i,],var=TRUE,Paired=FALSE)$p.value
biopval <- c(biopval,p)
}
```

```
allcomparisons<- cbind(allcomparisons,biopval)
pval.comparison<-paste(c("p-value",as.character(comparison.name)),sep="",collapse="|")
colnames(allcomparisons)[which(colnames(allcomparisons) == "biopval")] = pval.comparison

basecondition <- rowMeans (base.condition)
comparedcondition <- rowMeans (compared.condition)
logfoldchange <- comparedcondition-basecondition
foldchange <- 2**(abs(logfoldchange))*sign(logfoldchange)
allcomparisons<- cbind(allcomparisons,foldchange)
fold.comparison<-paste(c("fold.change",as.character(comparison.name)),sep="",collapse="|")
colnames(allcomparisons)[which(colnames(allcomparisons) == "foldchange")] =
fold.comparison
assign('allcomparisons',allcomparisons,envir=.GlobalEnv)
}
##########################################################
##############described desired comparisons and run#################
name<- "0.04% vs 2% CO2"
base<- cbind(log113,log115,log117,log119)
compared<- cbind(log114,log116,log118,log121)
comp(name,base,compared)
```

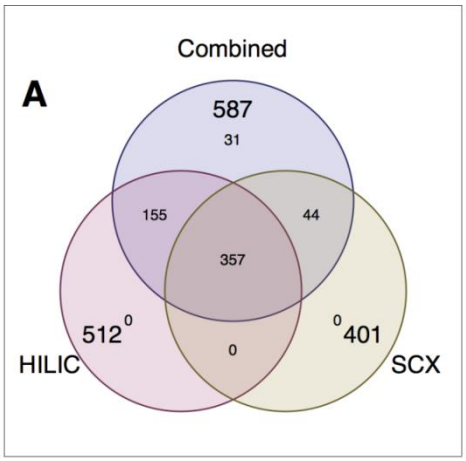
Chapter 5 Appendices

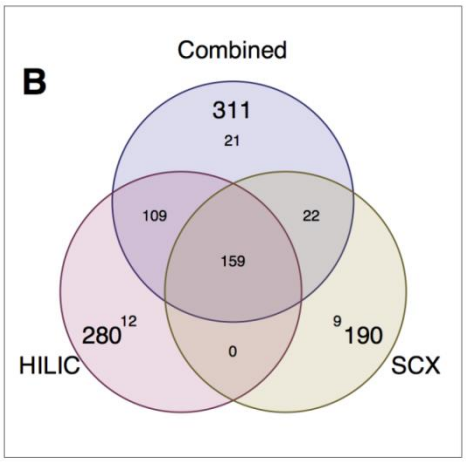
Appendix 5.1: Raw output from ProtienPilot allowing for separation of SCX and HILIC derived PSMs.

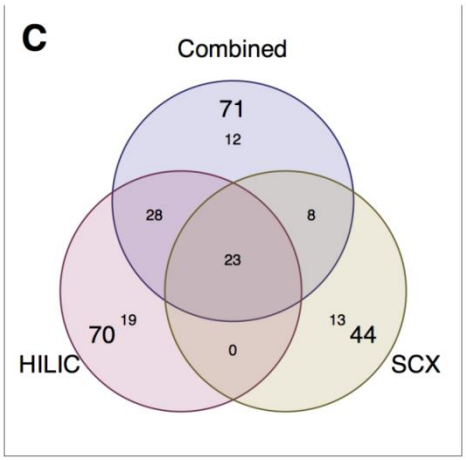
Appendix 5.2: Table of merged PSMs as outputted for the outputted by the script developed in Chapter 4 and reported in Appendix 4.1$_{p267}$. This Appendix is supplied on the enclosed CD only due to space.

Appendix 5.3: Table of protein changes identified as outputted by the script developed in Chapter 4 and reported in Appendix 4.2$_{p273}$. For each identified protein the number of unique peptide is reported with several comparisons, each of which reports the fold change and *p*-value for significance of change. This Appendix is supplied on the enclosed CD only due to space.

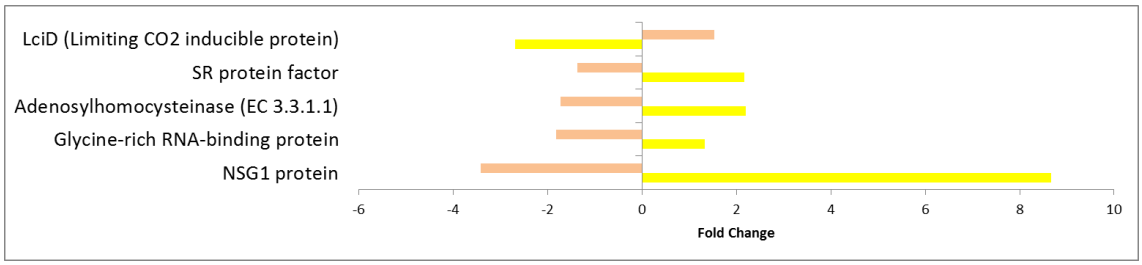
Appendix 5.4: Venn diagrams comparing data derived from HILIC fractionation, SCX fractionation and data from both sources combined. Comparisons of: A) Identified proteins, B) Significant changes (p < 0.05) identified in the lipid comparison, C) Significant changes (p < 0.05) identified in the carbohydrate comparison.



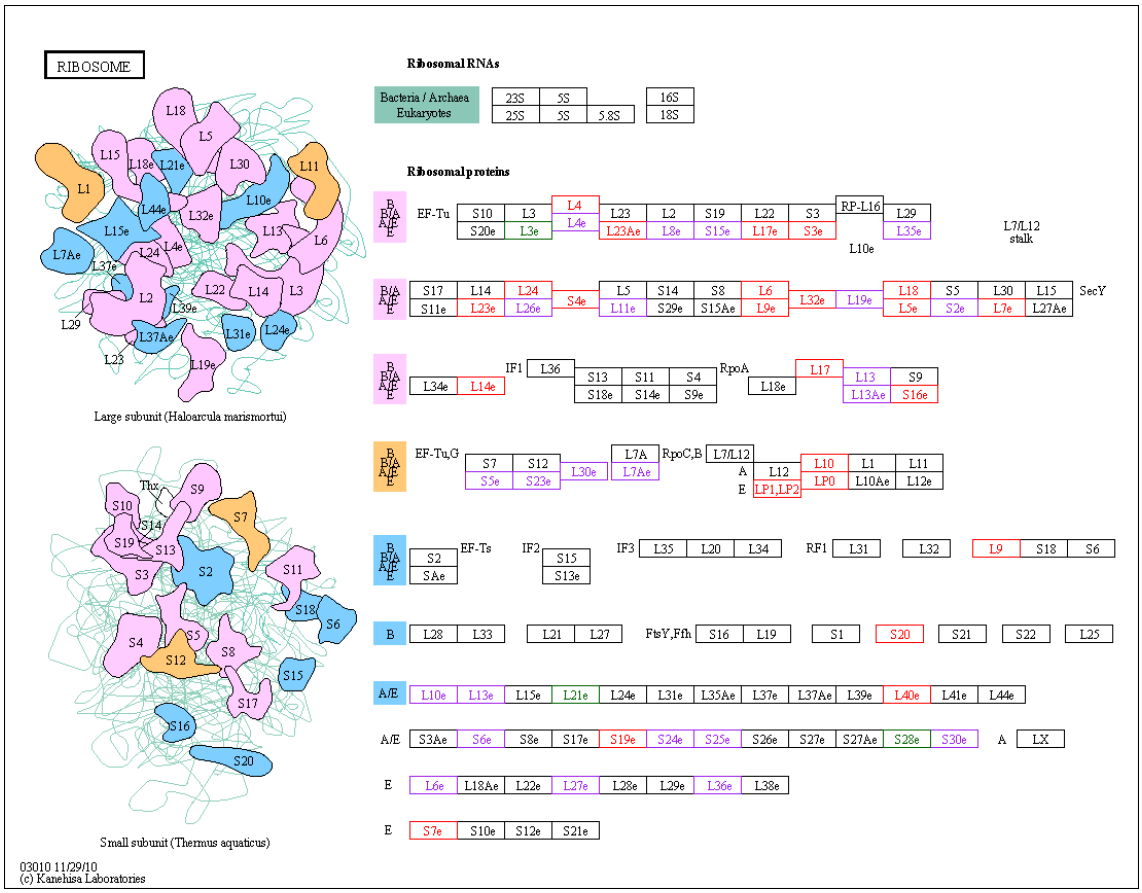




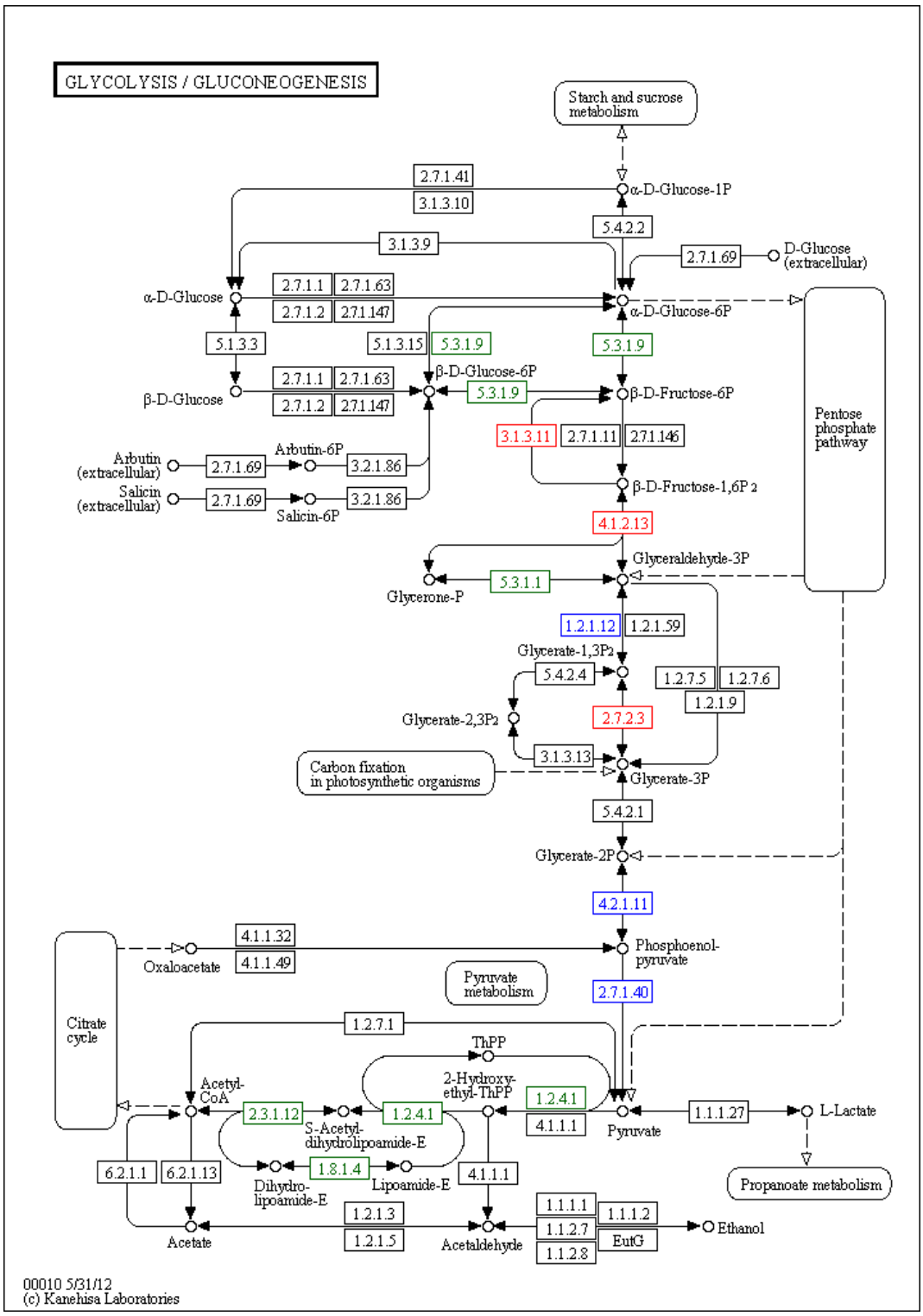
Appendix 5.5: Significantly observed changes that did not follow a continual trend under nitrogen starvation, being increased or decreased at 6 h post-nitrogen starvation and showing the opposite trend at 39 h post-nitrogen starvation. Changes between 0 h and 6 h are shown in pink and between 6 h and 39 h are shown in yellow.

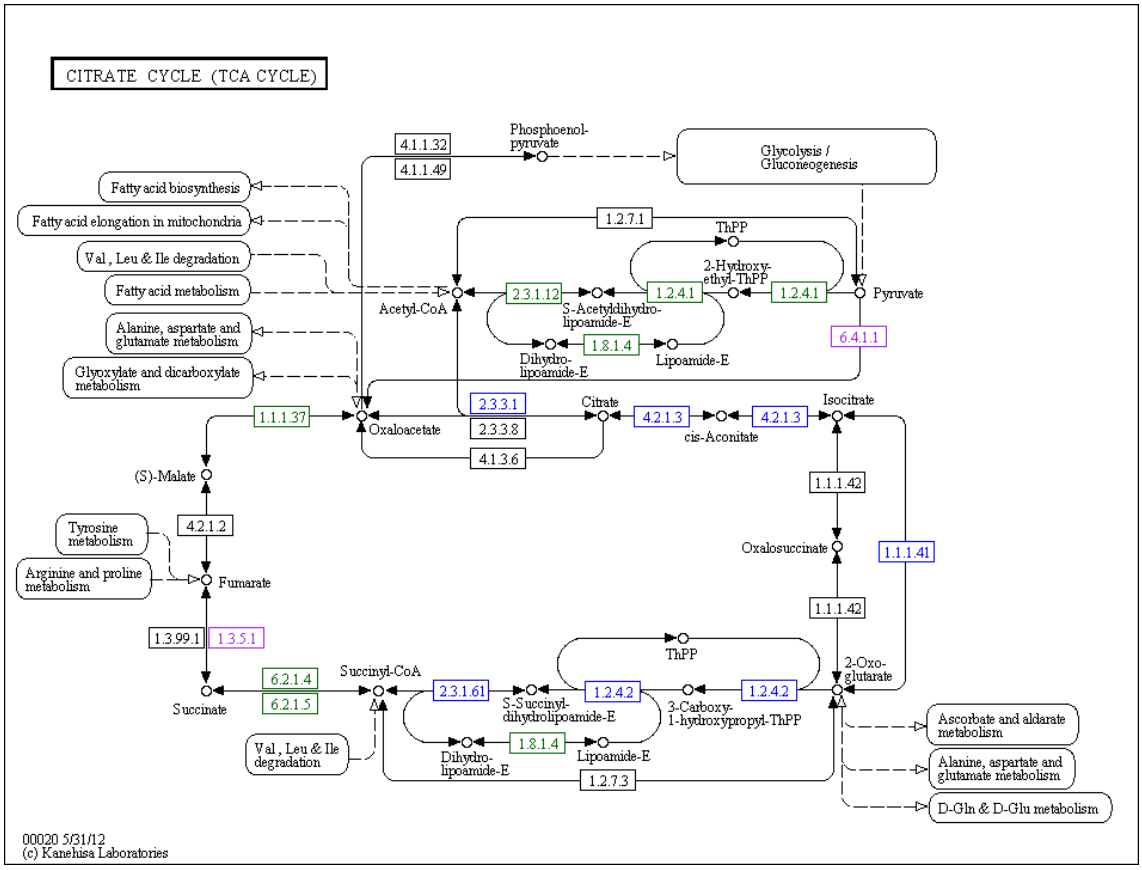


Appendix 5.6: KEGG map of the ribosome with identified proteins changes (0 h vs. 39 h) coloured based on observed regulation. Increased abundance in dark green outline with blue outline indicating significant changes ($p < 0.05$). Decreased abundance in purple outline with red out indicating significant changes ($p < 0.05$).

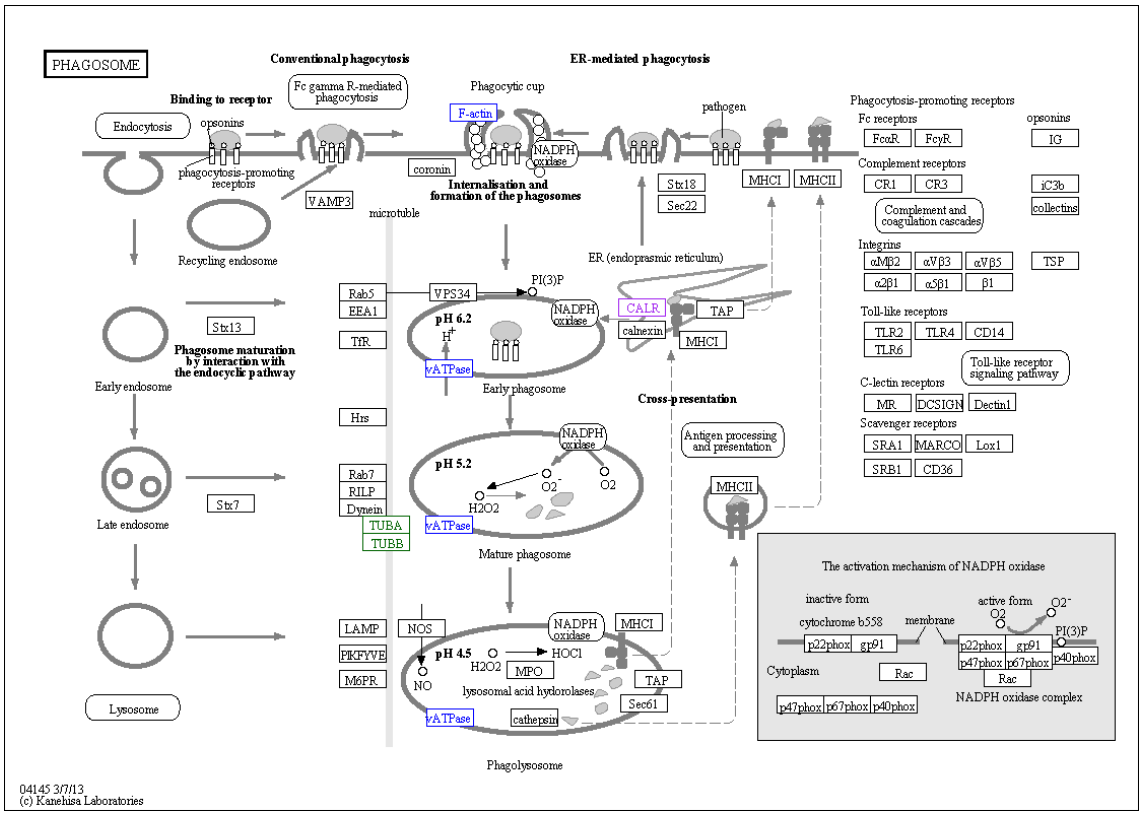


Appendix 5.7: KEGG map of the central energy metabolism showing Glycolysis and the TCA cycle with identified proteins changes (0 h vs. 39 h) coloured based on observed regulation. Increased abundance in dark green outline with blue outline indicating significant changes ($p < 0.05$). Decreased abundance in purple outline with red outline indicating significant changes ($p < 0.05$).

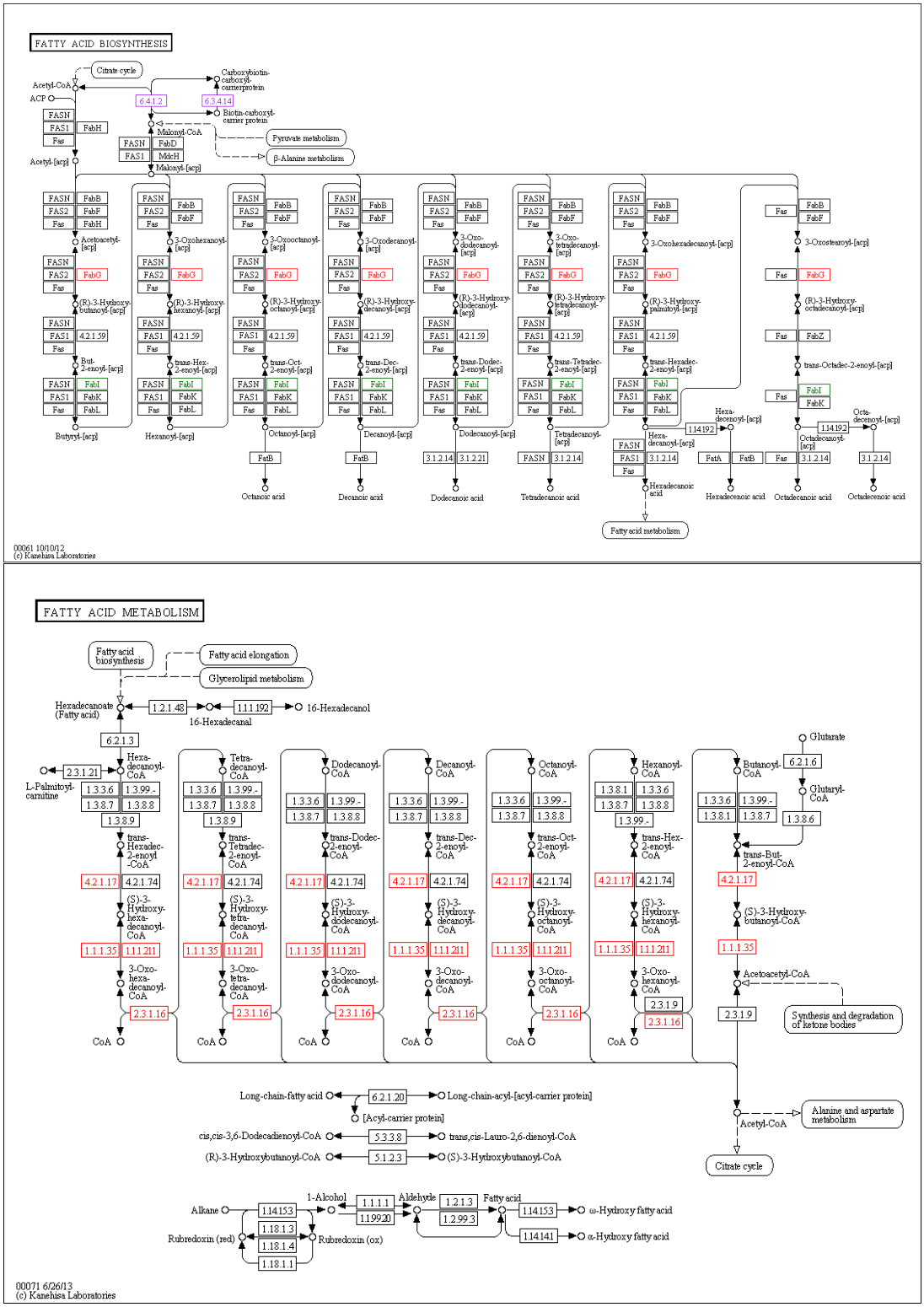



CITRATE CYCLE (TCA CYCLE)
Phosphoenol-pyruvate
4.1.1.32
4.1.1.49
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val, Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
1.2.7.1
ThPP
2-Hydroxy-ethyl-ThPP
2.3.1.12
1.2.4.1
1.2.4.1
Pyruvate
Acetyl-CoA
S-Acetyldihydro-lipoamide-E
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
6.4.1.1
2.3.3.1
Citrate
4.2.1.3
4.2.1.3
Isocitrate
1.1.1.37
2.3.3.8
Oxaloacetate
cis-Aconitate
4.1.3.6
(S)-Malate
1.1.1.42
Tyrosine metabolism
4.2.1.2
Oxalosuccinate
1.1.1.41
Arginine and proline metabolism
Fumarate
1.1.1.42
1.3.99.1
1.3.5.1
ThPP
2-Oxo-glutarate
6.2.1.4
6.2.1.5
Succinyl-CoA
2.3.1.61
1.2.4.2
1.2.4.2
Succinate
S-Succinyl-dihydrolipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
Ascorbate and aldarate metabolism
Val, Leu & Ile degradation
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
Alanine, aspartate and glutamate metabolism
1.2.7.3
D-Gln & D-Glu metabolism
00020 5/31/12
(c) Kanehisa Laboratories

Appendix 5.8: KEGG map of the Phagosome with identified proteins changes (0 h vs. 39 h) coloured based on observed regulation. Increased abundance in dark green outline with blue outline indicating significant changes ($p < 0.05$). Down-regulation in purple outline with red outline indicating significant changes ($p < 0.05$).



Appendix 5.9: KEGG map of proteins changes within fatty acid biosynthesis and metabolism when comparing 0 h with 39 h nitrogen stress. Based on observed regulation, increased abundance is in dark green outline with blue outline indicating significant changes ($p < 0.05$). Decreased abundance in purple outline with red outline indicating significant changes ($p < 0.05$).

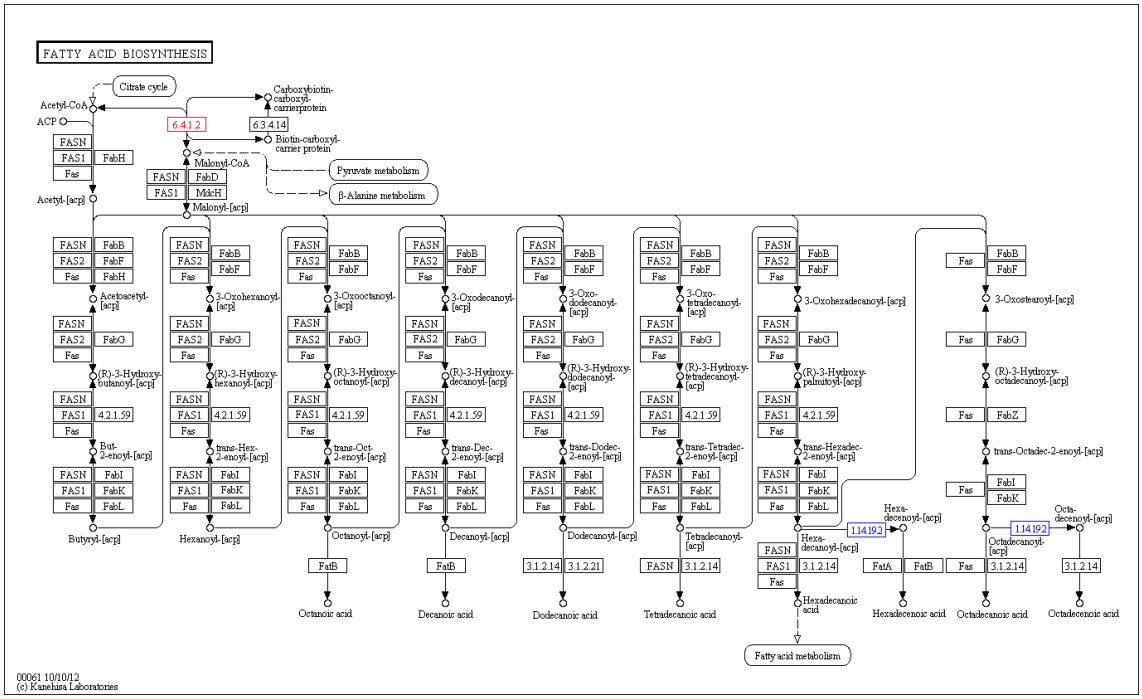


Chapter 6 Appendices

Appendix 6.1: Table of merged PSMs as outputted for the outputted by the script developed in Chapter 4 and reported in Appendix 4.1$_{p267}$. This Appendix is supplied on the enclosed CD only due to space.

Appendix 6.2: Table of protein changes identified as outputted by the script developed in Chapter 4 and reported in Appendix 4.2$_{p273}$. For each identified protein the number of unique peptide is reported with several comparisons, each of which reports the fold change and p-value for significance of change. This Appendix is supplied on the enclosed CD only due to space.

Appendix 6. 3: Fatty acid biosynthesis pathway diagrams from KEGG painted with significantly ($p < 0.05$) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.

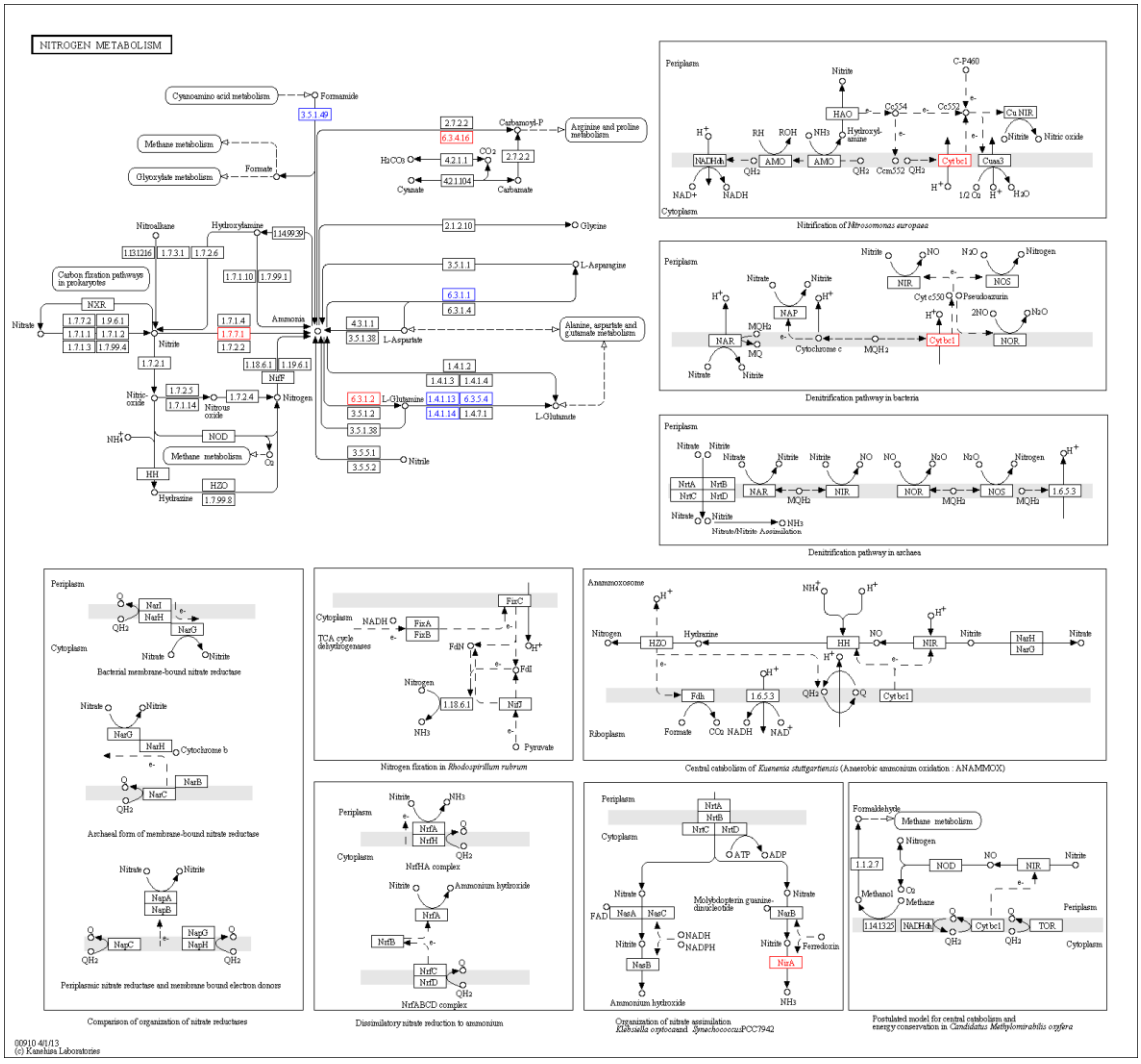


Chapter 7 Appendices

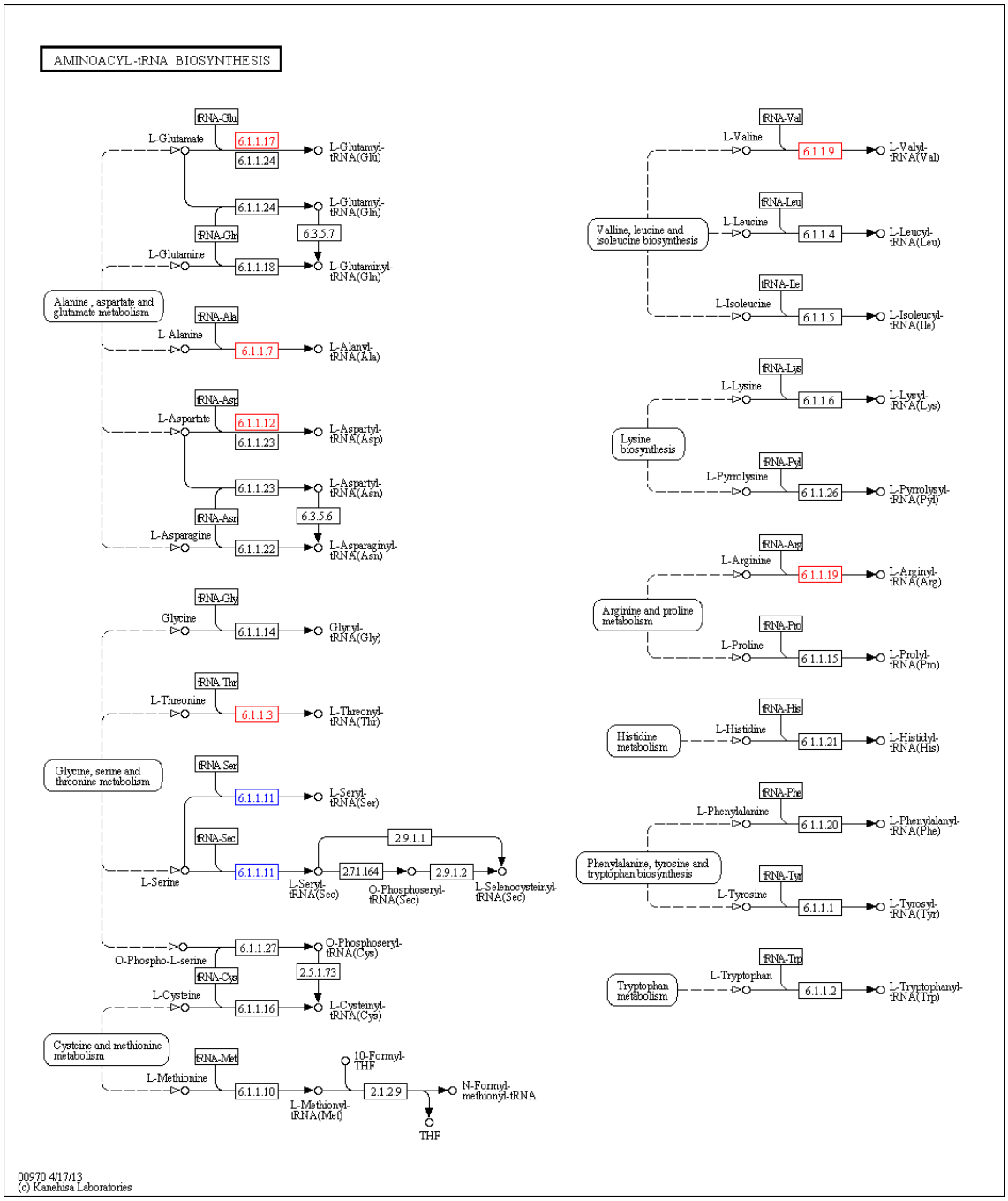
Appendix 7.1: Table of merged PSMs as outputted for the outputted by the script developed in Chapter 4 and reported in Appendix 4.1$_{p267}$. This Appendix is supplied on the enclosed CD only due to space.

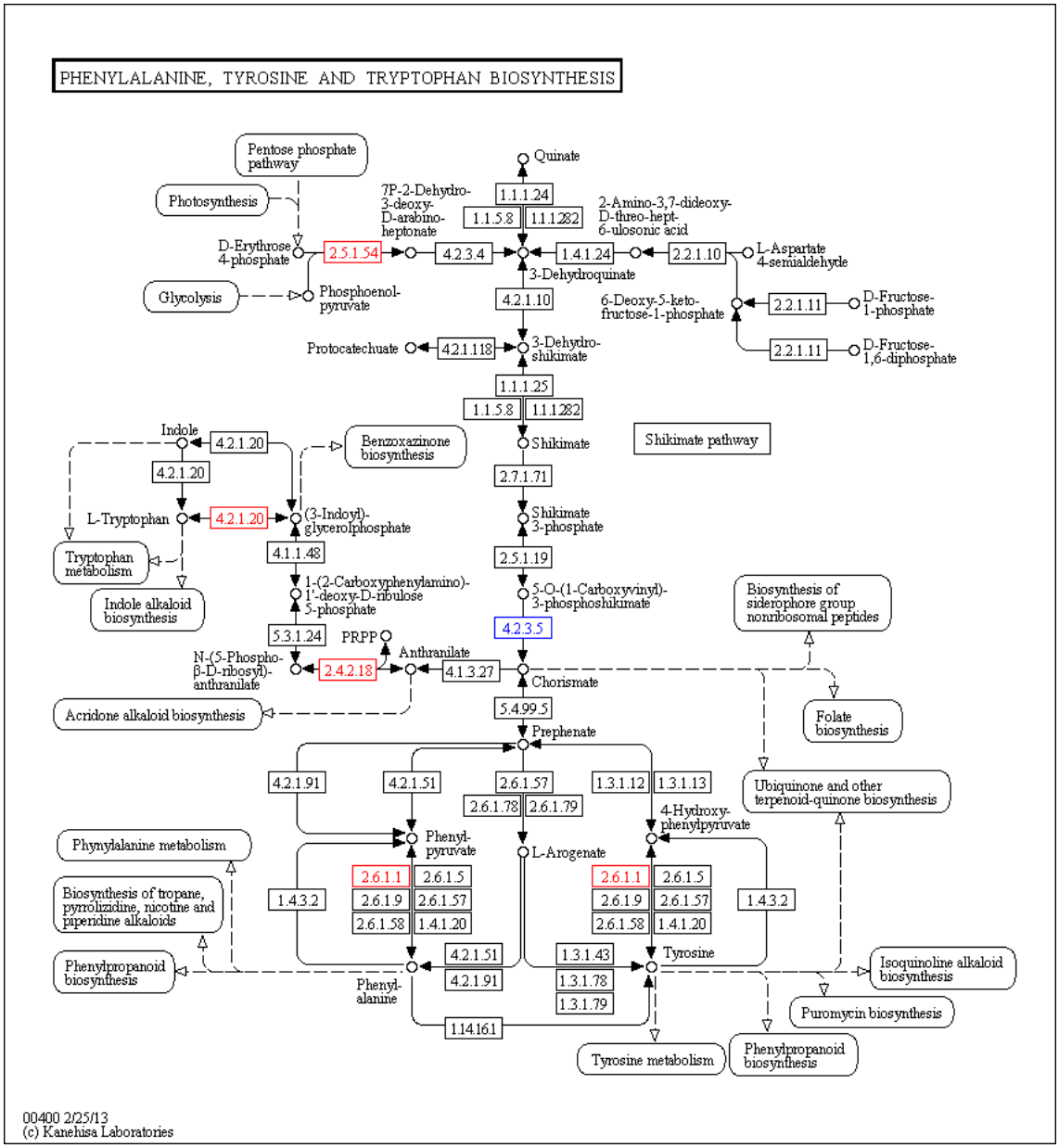
Appendix 7.2: Table of protein changes identified as outputted by the script developed in Chapter 4 and reported in Appendix 4.2$_{p273}$. For each identified protein the number of unique peptide is reported with several comparisons, each of which reports the fold change and *p*-value for significance of change. This Appendix is supplied on the enclosed CD only due to space.

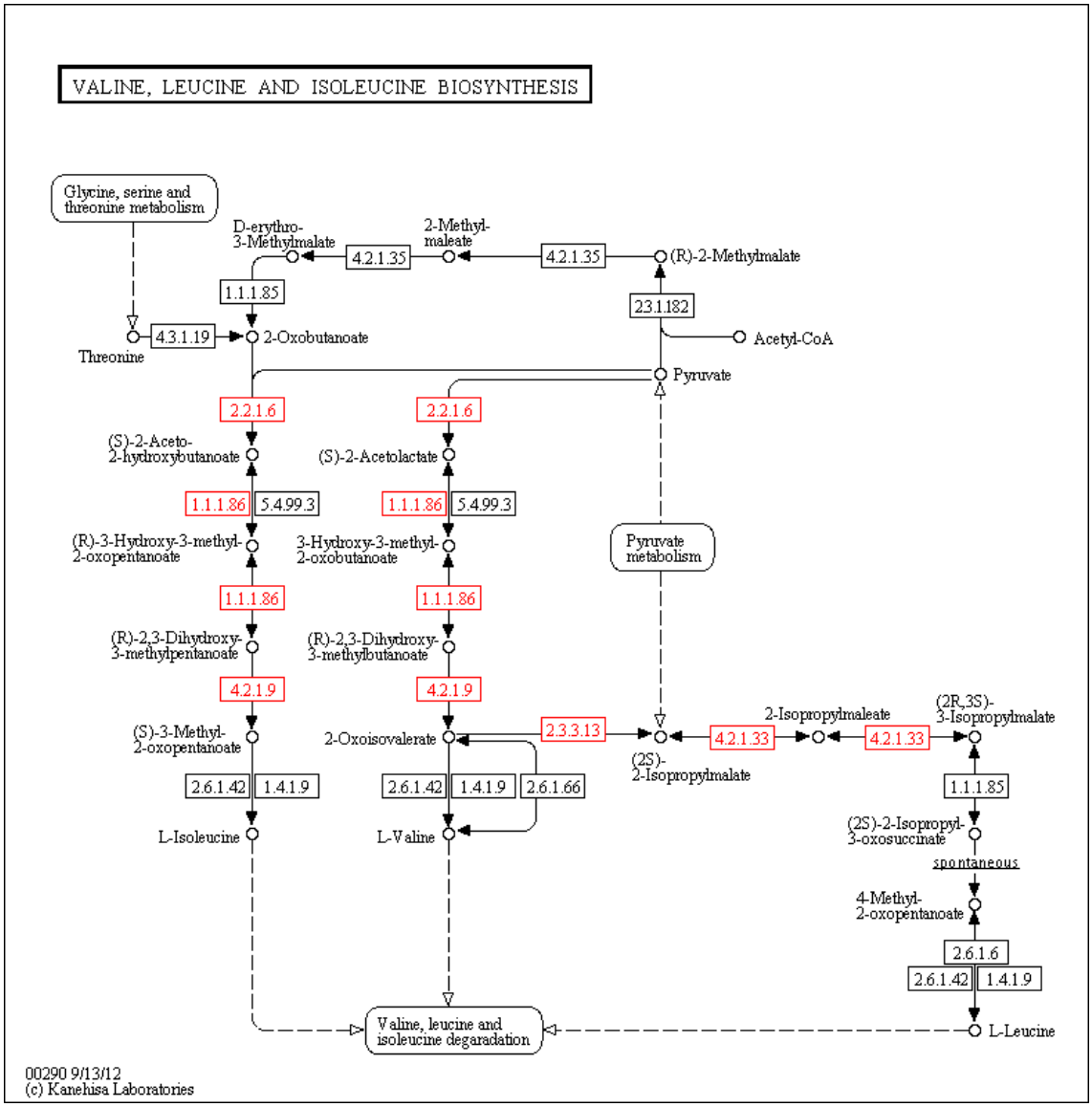
Appendix 7.3: Nitrogen metabolism pathway diagrams from KEGG painted with significantly ($p < 0.05$) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.



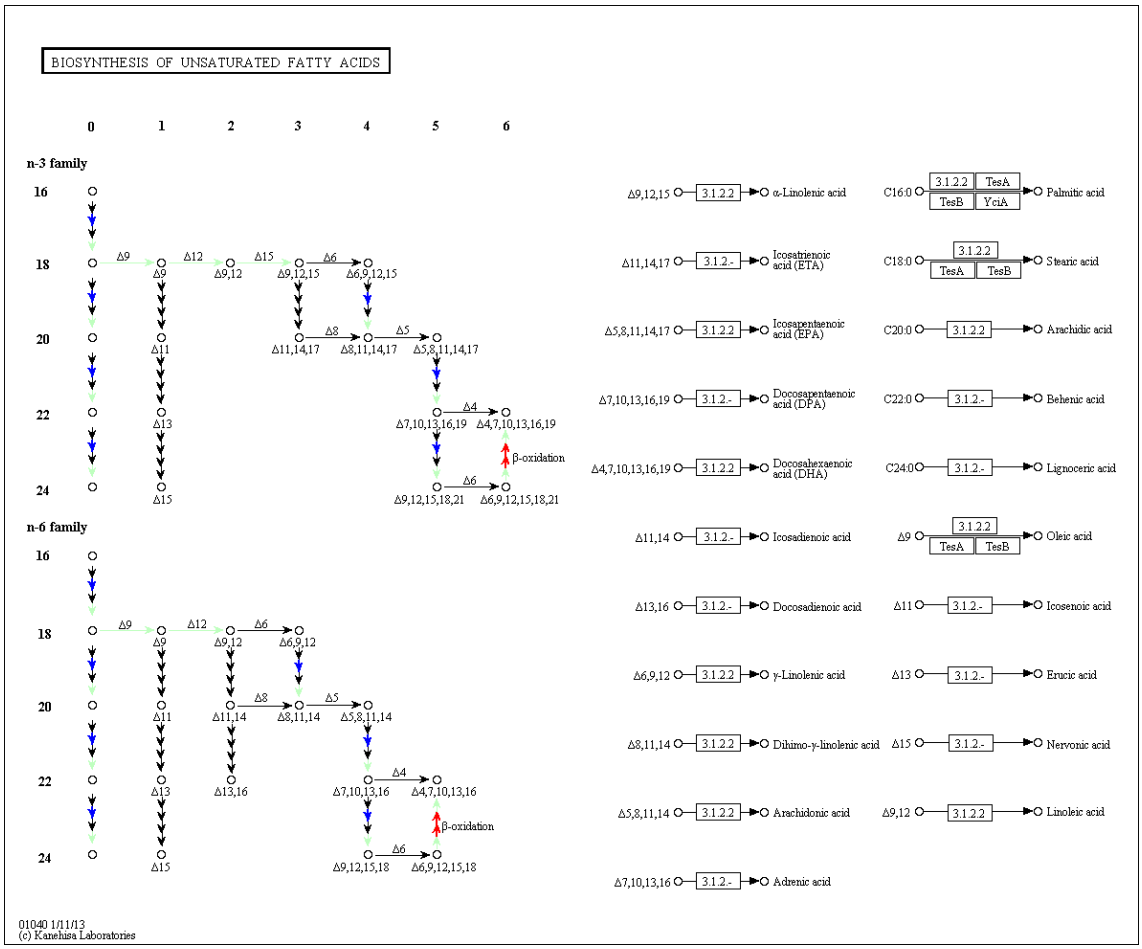
Appendix 7.4: Amino acid production pathway diagrams from KEGG painted with significantly ($p < 0.05$) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.

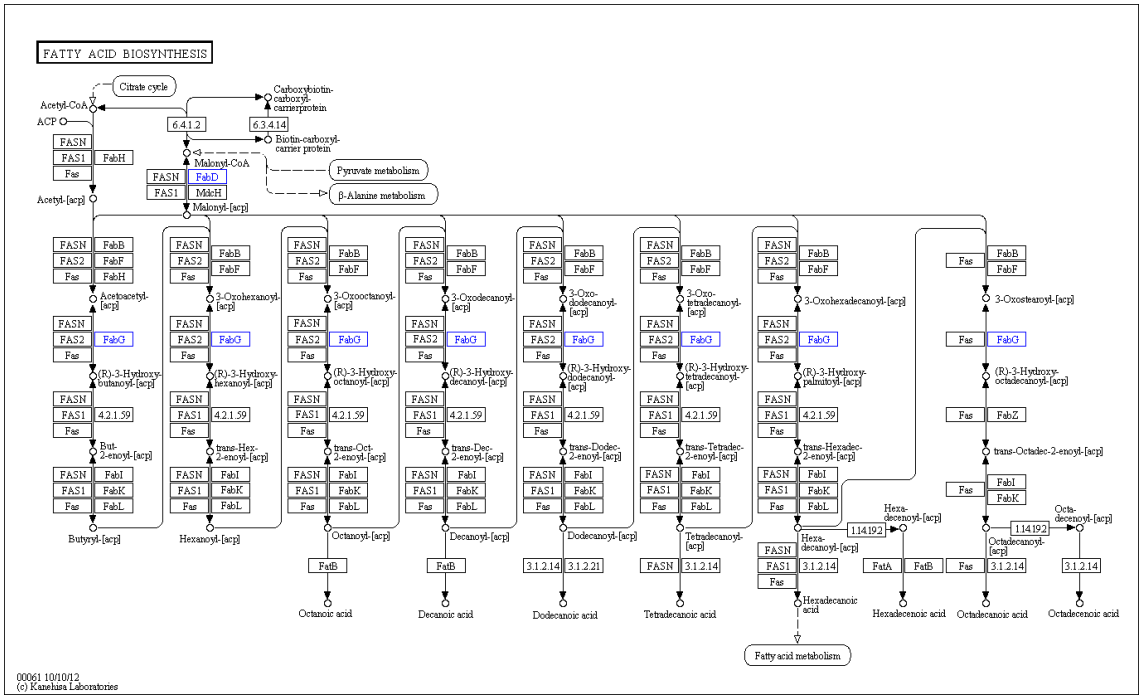


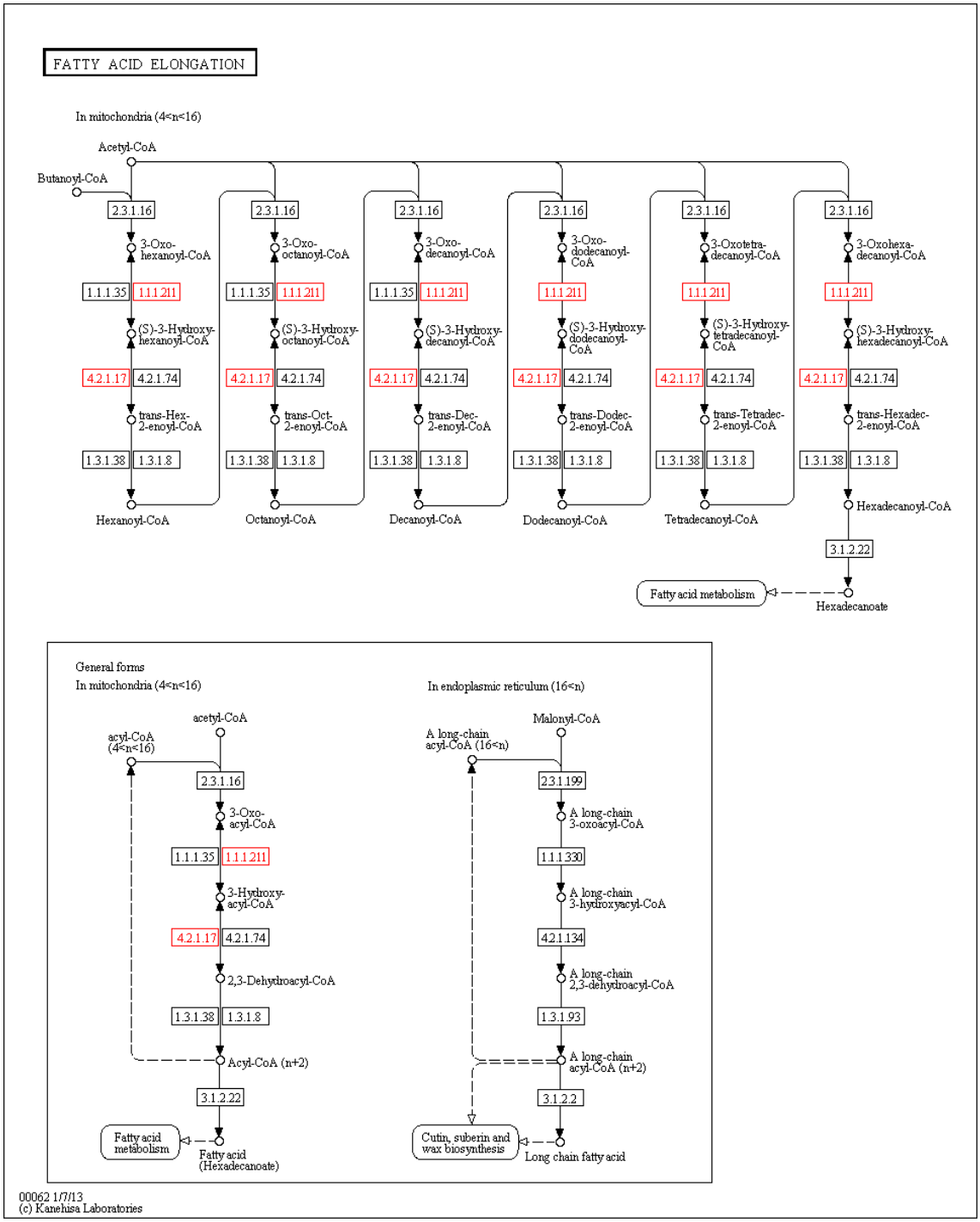

PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS
Pentose phosphate pathway
Photosynthesis
Glycolysis
D-Erythrose 4-phosphate
Phosphoenol-pyruvate
7P-2-Dehydro-3-deoxy-D-arabino-heptonate
Quinate
2-Amino-3,7-dideoxy-D-threo-hept-6-ulosonic acid
L-Aspartate 4-semialdehyde
3-Dehydroquinate
6-Deoxy-5-keto-fructose-1-phosphate
D-Fructose-1-phosphate
D-Fructose-1,6-diphosphate
Protocatechuate
3-Dehydro-shikimate
Shikimate
Shikimate pathway
Indole
Benzoxazinone biosynthesis
L-Tryptophan
(3-Indoyl)-glycerolphosphate
Shikimate 3-phosphate
Tryptophan metabolism
Indole alkaloid biosynthesis
1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5-phosphate
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
Biosynthesis of siderophore group nonribosomal peptides
PRPP
N-(5-Phospho-β-D-ribosyl)-anthranilate
Anthranilate
Chorismate
Acridone alkaloid biosynthesis
Folate biosynthesis
Prephenate
Ubiquinone and other terpenoid-quinone biosynthesis
4-Hydroxy-phenylpyruvate
Phenyl-pyruvate
L-Arogenate
Phynylalanine metabolism
Biosynthesis of tropane, pyrrolizidine, nicotine and piperidine alkaloids
Phenylpropanoid biosynthesis
Phenyl-alanine
Tyrosine
Isoquinoline alkaloid biosynthesis
Puromycin biosynthesis
Tyrosine metabolism
Phenylpropanoid biosynthesis
00400 2/25/13
(c) Kanehisa Laboratories

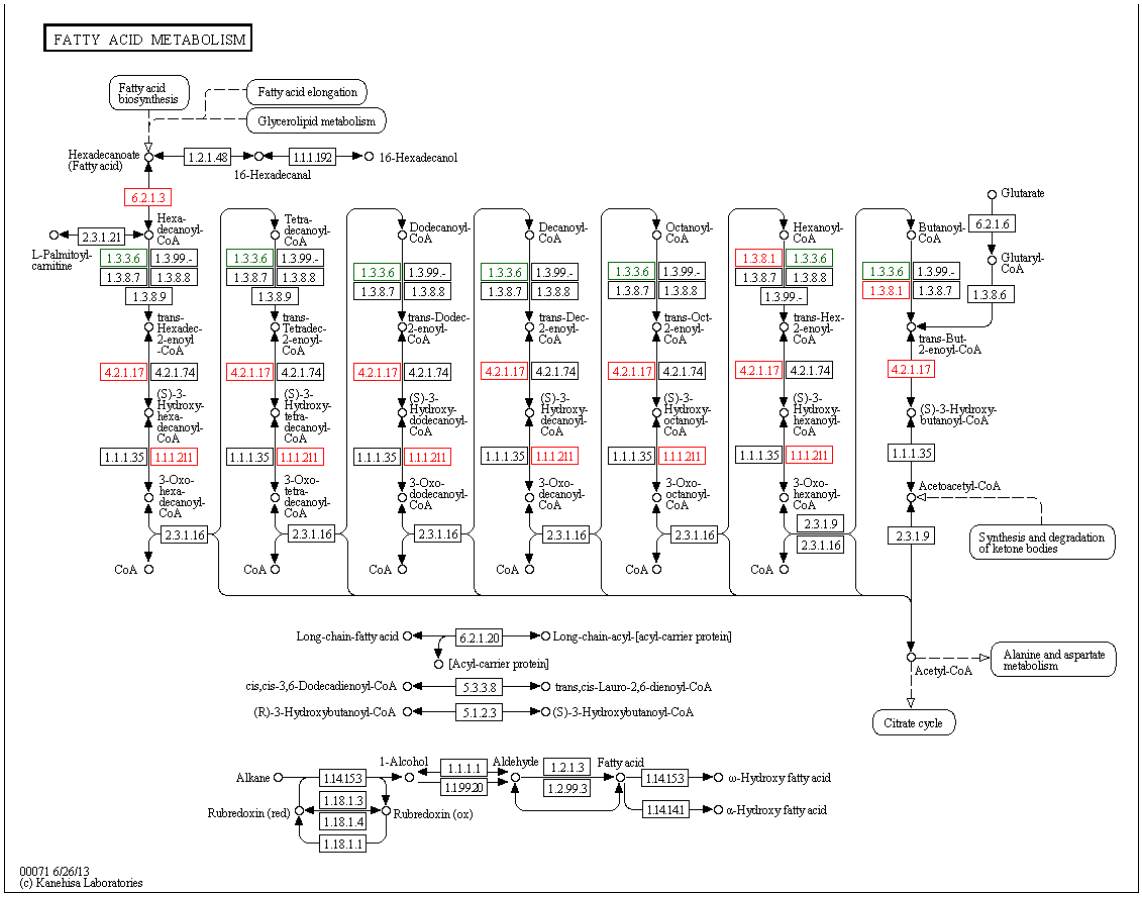

VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS
Glycine, serine and threonine metabolism
D-erythro-3-Methylmalate
4.2.1.35
2-Methyl-maleate
4.2.1.35
(R)-2-Methylmalate
1.1.1.85
2.3.1.182
4.3.1.19
2-Oxobutanoate
Acetyl-CoA
Threonine
Pyruvate
2.2.1.6
2.2.1.6
(S)-2-Aceto-2-hydroxybutanoate
(S)-2-Acetolactate
1.1.1.86
5.4.99.3
1.1.1.86
5.4.99.3
(R)-3-Hydroxy-3-methyl-2-oxopentanoate
3-Hydroxy-3-methyl-2-oxobutanoate
Pyruvate metabolism
1.1.1.86
1.1.1.86
(R)-2,3-Dihydroxy-3-methylpentanoate
(R)-2,3-Dihydroxy-3-methylbutanoate
4.2.1.9
4.2.1.9
2-Isopropylmaleate
(2R,3S)-3-Isopropylmalate
(S)-3-Methyl-2-oxopentanoate
2-Oxoisovalerate
2.3.3.13
4.2.1.33
4.2.1.33
(2S)-2-Isopropylmalate
2.6.1.42
1.4.1.9
2.6.1.42
1.4.1.9
2.6.1.66
1.1.1.85
L-Isoleucine
L-Valine
(2S)-2-Isopropyl-3-oxosuccinate
spontaneous
4-Methyl-2-oxopentanoate
2.6.1.6
2.6.1.42
1.4.1.9
Valine, leucine and isoleucine degaradation
L-Leucine
00290 9/13/12
(c) Kanehisa Laboratories

Appendix 7.5: Fatty acid pathways diagrams from KEGG painted with significantly ($p < 0.05$) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.

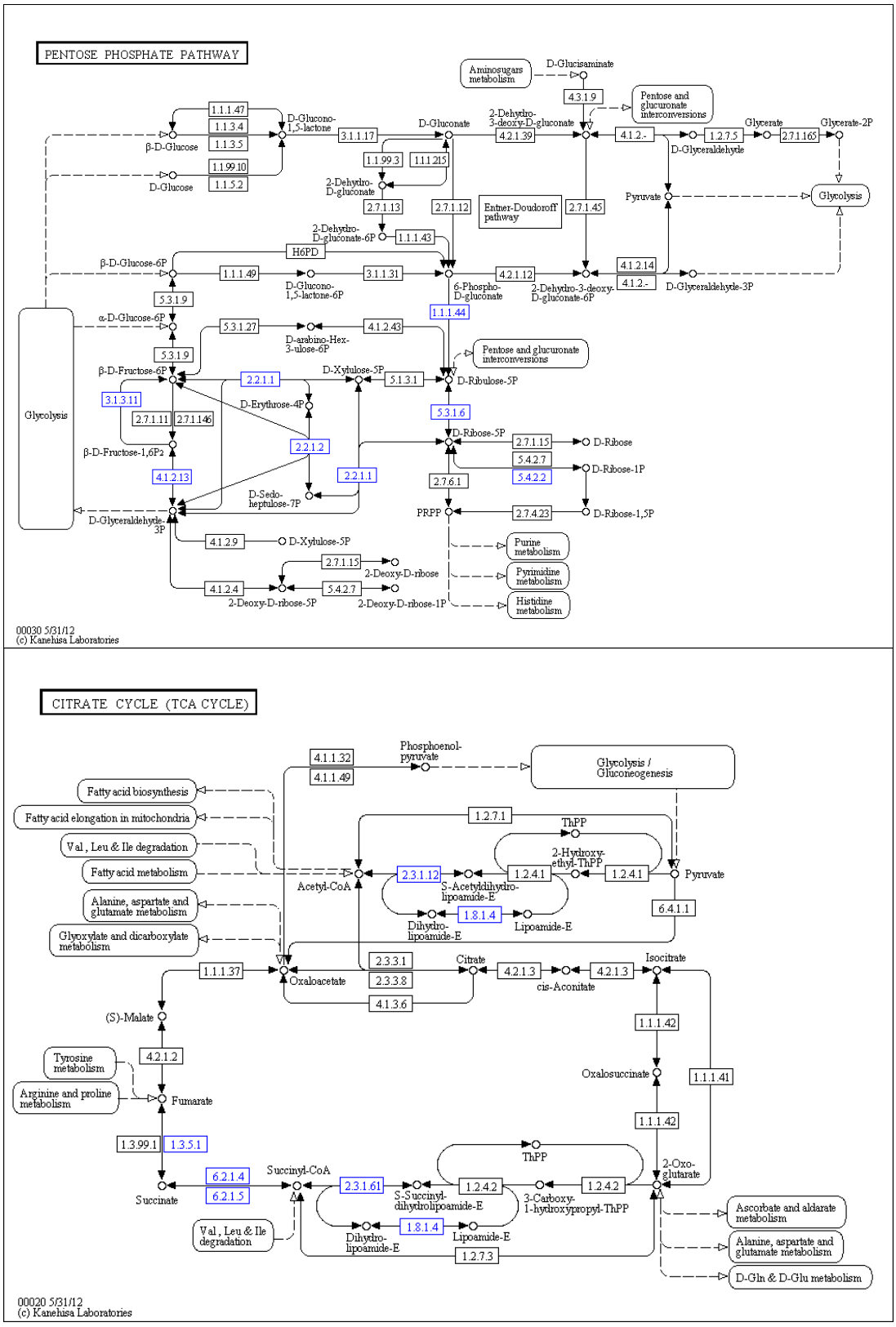


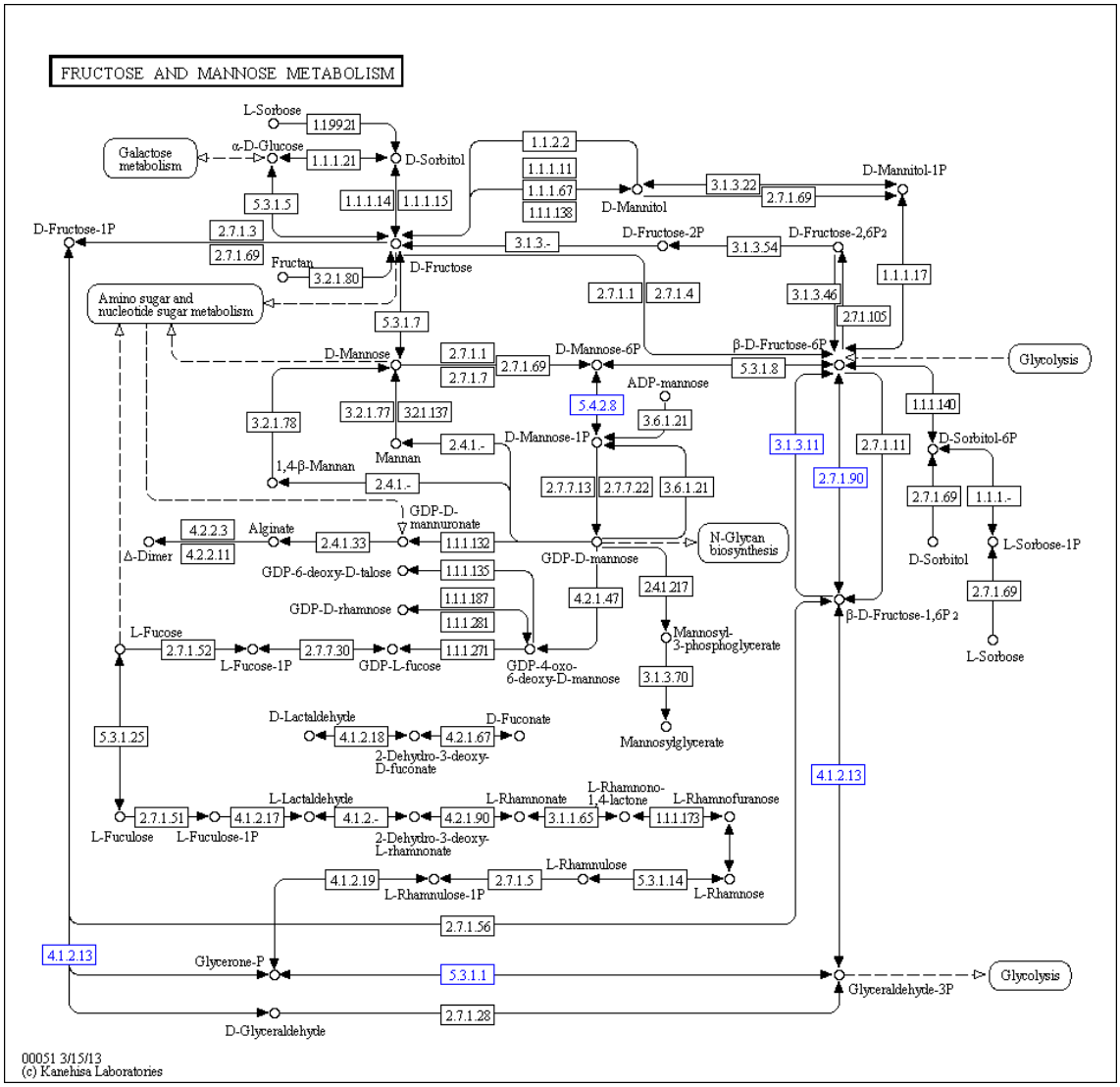


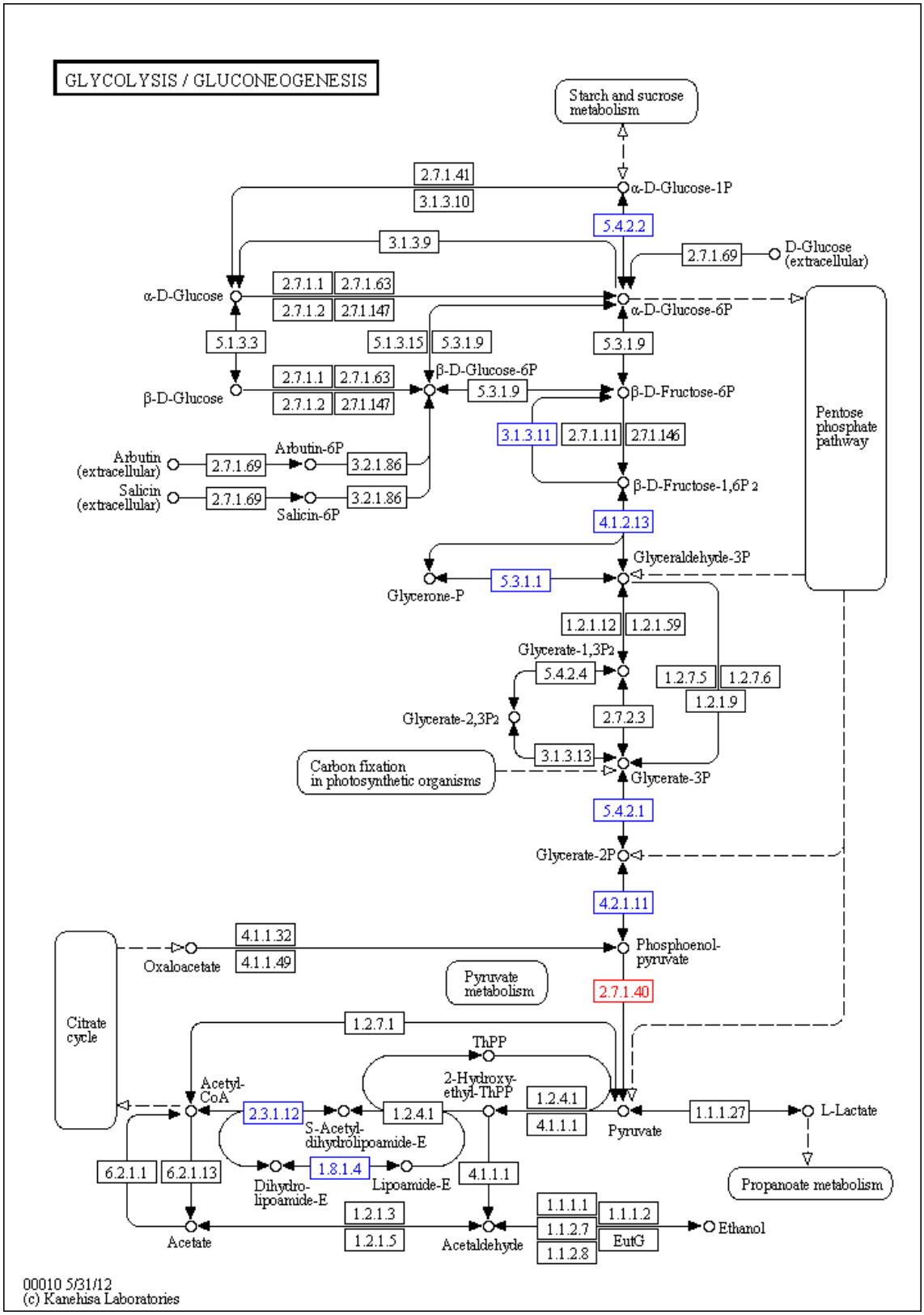

FATTY ACID ELONGATION
In mitochondria (4<n<16)
Acetyl-CoA
Butanoyl-CoA
2.3.1.16
3-Oxo-hexanoyl-CoA
1.1.1.35
1.1.1.211
(S)-3-Hydroxy-hexanoyl-CoA
4.2.1.17
4.2.1.74
trans-Hex-2-enoyl-CoA
1.3.1.38
1.3.1.8
Hexanoyl-CoA
3-Oxo-octanoyl-CoA
(S)-3-Hydroxy-octanoyl-CoA
trans-Oct-2-enoyl-CoA
Octanoyl-CoA
3-Oxo-decanoyl-CoA
(S)-3-Hydroxy-decanoyl-CoA
trans-Dec-2-enoyl-CoA
Decanoyl-CoA
3-Oxo-dodecanoyl-CoA
(S)-3-Hydroxy-dodecanoyl-CoA
trans-Dodec-2-enoyl-CoA
Dodecanoyl-CoA
3-Oxotetra-decanoyl-CoA
(S)-3-Hydroxy-tetradecanoyl-CoA
trans-Tetradec-2-enoyl-CoA
Tetradecanoyl-CoA
3-Oxohexa-decanoyl-CoA
(S)-3-Hydroxy-hexadecanoyl-CoA
trans-Hexadec-2-enoyl-CoA
Hexadecanoyl-CoA
3.1.2.22
Fatty acid metabolism
Hexadecanoate
General forms
In mitochondria (4<n<16)
acetyl-CoA
acyl-CoA (4<n<16)
2.3.1.16
3-Oxo-acyl-CoA
1.1.1.35
1.1.1.211
3-Hydroxy-acyl-CoA
4.2.1.17
4.2.1.74
2,3-Dehydroacyl-CoA
1.3.1.38
1.3.1.8
Acyl-CoA (n+2)
3.1.2.22
Fatty acid metabolism
Fatty acid (Hexadecanoate)
In endoplasmic reticulum (16<n)
Malonyl-CoA
A long-chain acyl-CoA (16<n)
2.3.1.199
A long-chain 3-oxoacyl-CoA
1.1.1.330
A long-chain 3-hydroxyacyl-CoA
4.2.1.134
A long-chain 2,3-dehydroacyl-CoA
1.3.1.93
A long-chain acyl-CoA (n+2)
3.1.2.2
Cutin, suberin and wax biosynthesis
Long chain fatty acid
00062 1/7/13
(c) Kanehisa Laboratories


FATTY ACID METABOLISM
Fatty acid biosynthesis
Fatty acid elongation
Glycerolipid metabolism
Hexadecanoate (Fatty acid)
1.2.1.48
16-Hexadecanal
1.1.1.192
16-Hexadecanol
6.2.1.3
Hexadecanoyl-CoA
Tetradecanoyl-CoA
Dodecanoyl-CoA
Decanoyl-CoA
Octanoyl-CoA
Hexanoyl-CoA
Butanoyl-CoA
2.3.1.21
L-Palmitoyl-carnitine
1.3.3.6
1.3.99.-
1.3.8.7
1.3.8.8
1.3.8.9
1.3.8.1
trans-Hexadec-2-enoyl-CoA
trans-Tetradec-2-enoyl-CoA
trans-Dodec-2-enoyl-CoA
trans-Dec-2-enoyl-CoA
trans-Oct-2-enoyl-CoA
trans-Hex-2-enoyl-CoA
trans-But-2-enoyl-CoA
4.2.1.17
4.2.1.74
(S)-3-Hydroxyhexadecanoyl-CoA
(S)-3-Hydroxytetradecanoyl-CoA
(S)-3-Hydroxydodecanoyl-CoA
(S)-3-Hydroxydecanoyl-CoA
(S)-3-Hydroxyoctanoyl-CoA
(S)-3-Hydroxyhexanoyl-CoA
(S)-3-Hydroxybutanoyl-CoA
1.1.1.35
1.1.1.211
3-Oxohexadecanoyl-CoA
3-Oxotetradecanoyl-CoA
3-Oxododecanoyl-CoA
3-Oxodecanoyl-CoA
3-Oxooctanoyl-CoA
3-Oxohexanoyl-CoA
2.3.1.16
2.3.1.9
CoA
Glutarate
6.2.1.6
Glutaryl-CoA
Acetoacetyl-CoA
Synthesis and degradation of ketone bodies
Alanine and aspartate metabolism
Acetyl-CoA
Citrate cycle
Long-chain-fatty acid
6.2.1.20
Long-chain-acyl-[acyl-carrier protein]
[Acyl-carrier protein]
cis,cis-3,6-Dodecadienoyl-CoA
5.3.3.8
trans,cis-Lauro-2,6-dienoyl-CoA
(R)-3-Hydroxybutanoyl-CoA
5.1.2.3
(S)-3-Hydroxybutanoyl-CoA
Alkane
1.14.15.3
1-Alcohol
1.1.1.1
1.1.99.20
Aldehyde
1.2.1.3
1.2.99.3
Fatty acid
1.14.15.3
ω-Hydroxy fatty acid
1.14.14.1
α-Hydroxy fatty acid
Rubredoxin (red)
1.18.1.3
1.18.1.4
1.18.1.1
Rubredoxin (ox)
00071 6/26/13
(c) Kanehisa Laboratories

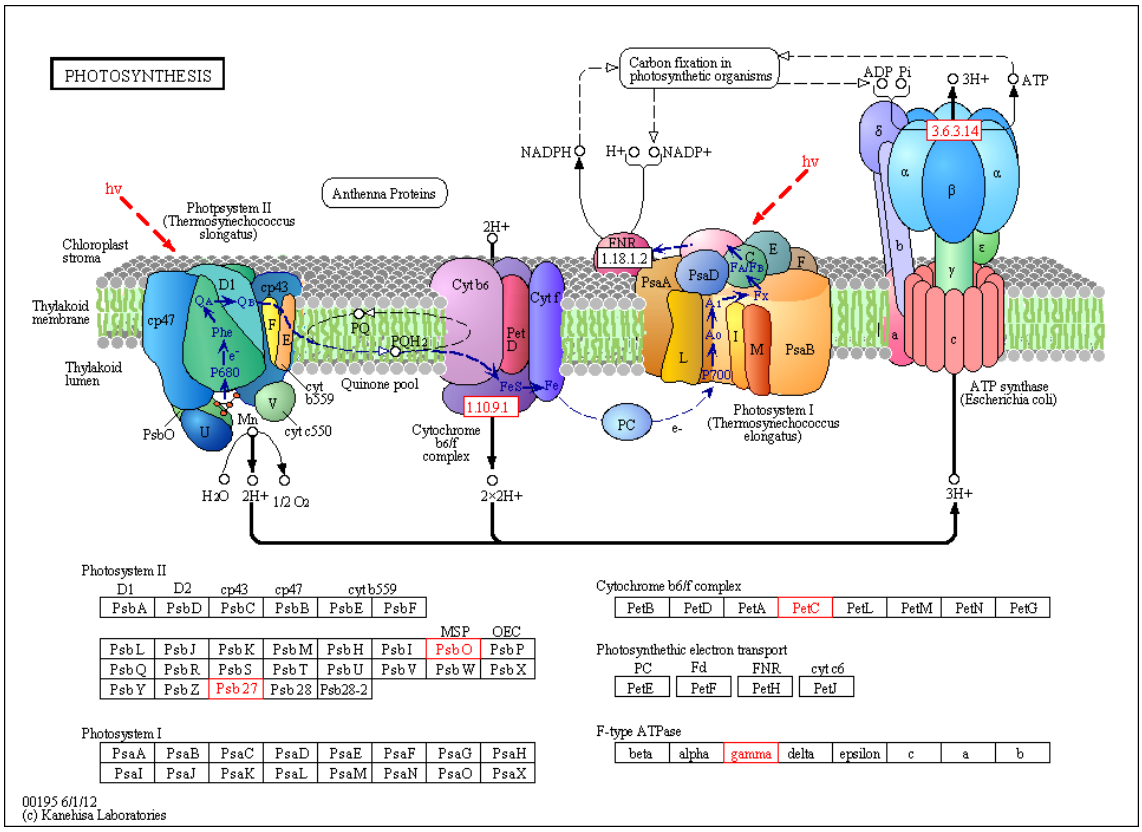
Appendix 7.6: Central energy metabolism pathway diagrams from KEGG painted with significantly (p < 0.05) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.

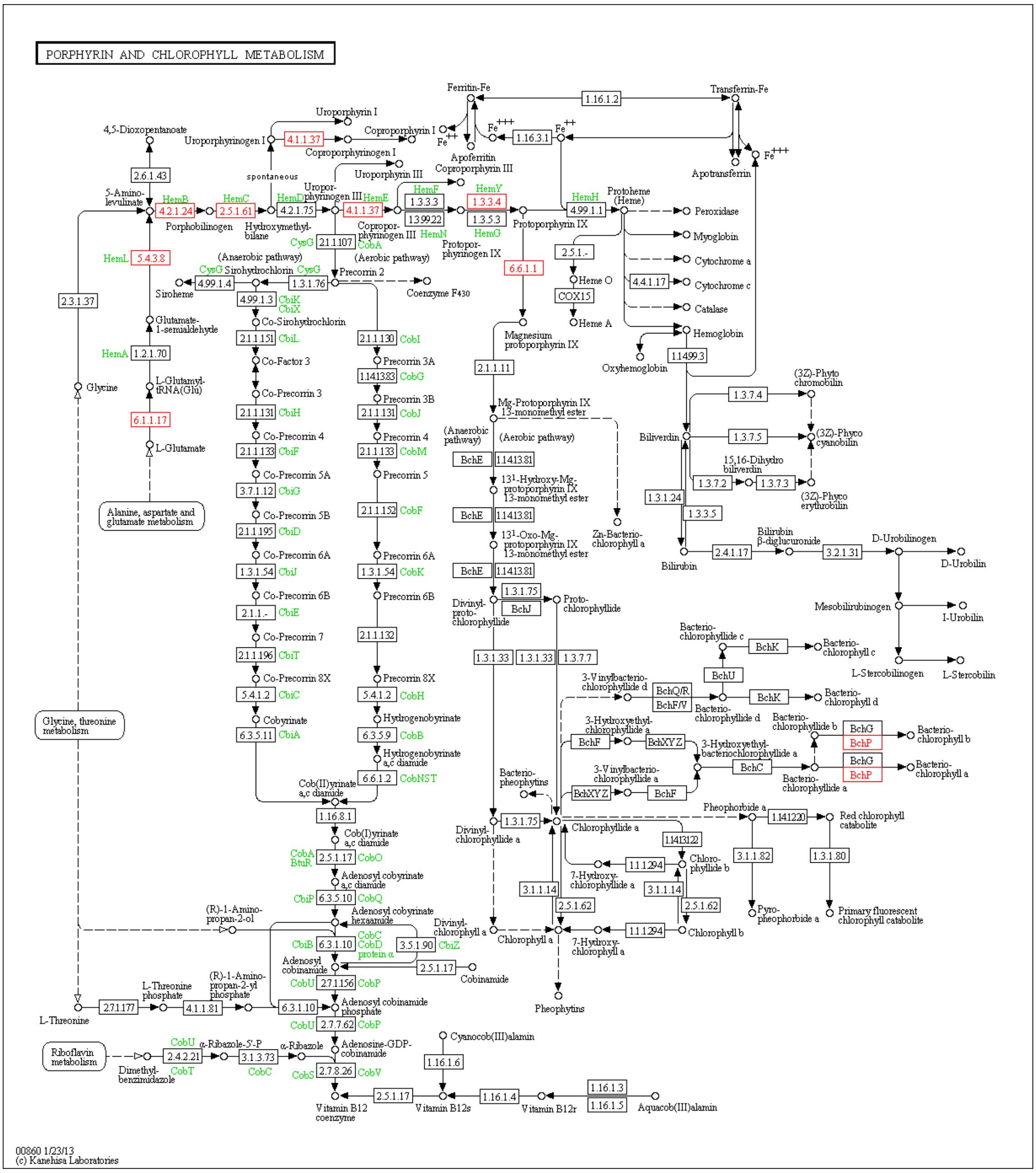



FRUCTOSE AND MANNOSE METABOLISM
L-Sorbose
1.1.99.21
Galactose metabolism
α-D-Glucose
1.1.1.21
D-Sorbitol
1.1.2.2
1.1.1.11
1.1.1.67
1.1.1.138
D-Mannitol
3.1.3.22
2.7.1.69
D-Mannitol-1P
5.3.1.5
1.1.1.14
1.1.1.15
D-Fructose-1P
2.7.1.3
2.7.1.69
3.1.3.-
D-Fructose-2P
3.1.3.54
D-Fructose-2,6P2
Fructan
3.2.1.80
D-Fructose
1.1.1.17
Amino sugar and nucleotide sugar metabolism
2.7.1.1
2.7.1.4
3.1.3.46
2.7.1.105
5.3.1.7
D-Mannose
2.7.1.1
2.7.1.7
2.7.1.69
D-Mannose-6P
β-D-Fructose-6P
5.3.1.8
Glycolysis
ADP-mannose
5.4.2.8
3.6.1.21
1.1.1.140
3.2.1.78
3.2.1.77
3.2.1.137
D-Mannose-1P
2.4.1.-
Mannan
3.1.3.11
2.7.1.11
D-Sorbitol-6P
1,4-β-Mannan
2.4.1.-
2.7.7.13
2.7.7.22
3.6.1.21
2.7.1.90
2.7.1.69
1.1.1.-
GDP-D-mannuronate
4.2.2.3
Alginate
2.4.1.33
1.1.1.132
N-Glycan biosynthesis
Δ-Dimer
4.2.2.11
GDP-D-mannose
D-Sorbitol
L-Sorbose-1P
GDP-6-deoxy-D-talose
1.1.1.135
2.4.1.217
2.7.1.69
1.1.1.187
GDP-D-rhamnose
1.1.1.281
4.2.1.47
β-D-Fructose-1,6P2
L-Fucose
2.7.1.52
2.7.7.30
1.1.1.271
Mannosyl-3-phosphoglycerate
L-Sorbose
L-Fucose-1P
GDP-L-fucose
GDP-4-oxo-6-deoxy-D-mannose
3.1.3.70
D-Lactaldehyde
D-Fuconate
5.3.1.25
4.1.2.18
4.2.1.67
Mannosylglycerate
2-Dehydro-3-deoxy-D-fuconate
4.1.2.13
L-Lactaldehyde
L-Rhamnonate
L-Rhamnono-1,4-lactone
L-Rhamnofuranose
2.7.1.51
4.1.2.17
4.1.2.-
4.2.1.90
3.1.1.65
1.1.1.173
L-Fuculose
L-Fuculose-1P
2-Dehydro-3-deoxy-L-rhamnonate
L-Rhamnulose
4.1.2.19
2.7.1.5
5.3.1.14
L-Rhamnulose-1P
L-Rhamnose
2.7.1.56
4.1.2.13
Glycerone-P
5.3.1.1
Glycolysis
Glyceraldehyde-3P
2.7.1.28
D-Glyceraldehyde
00051 3/15/13
(c) Kanehisa Laboratories

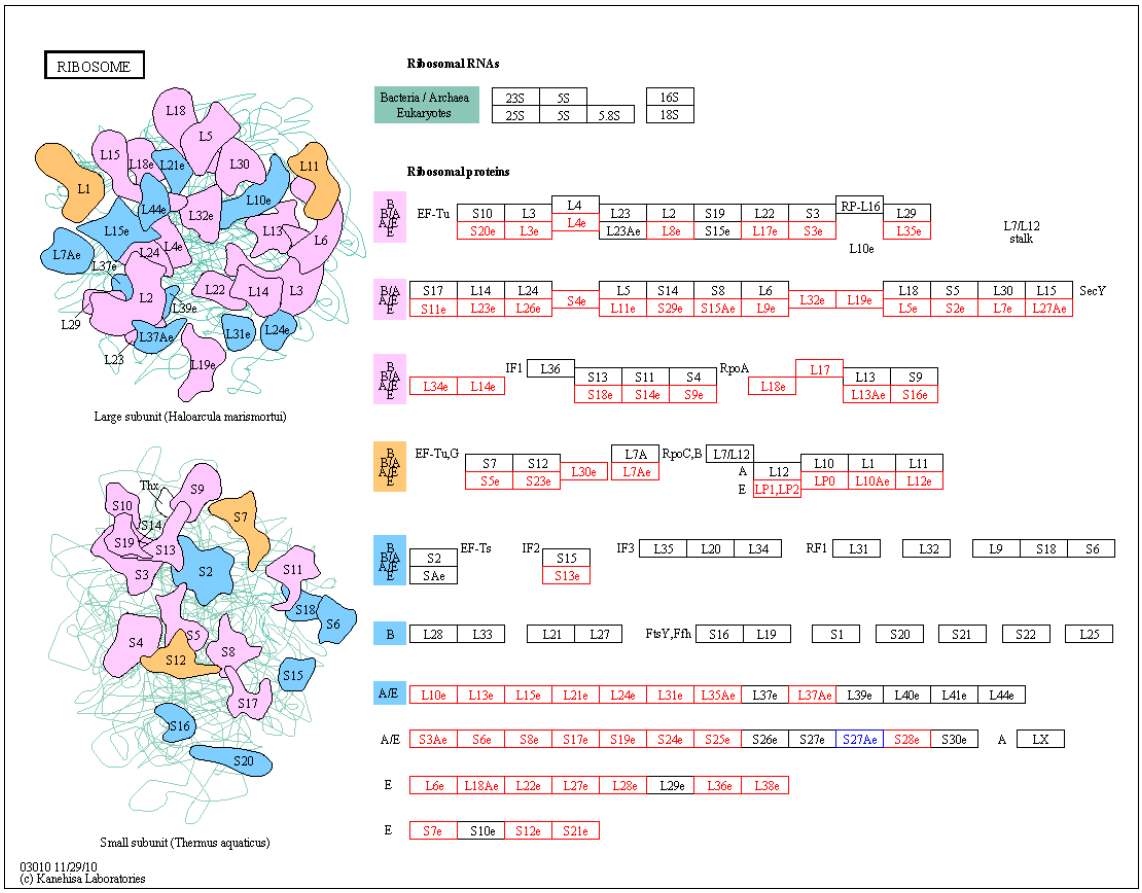

GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
2.7.1.41
3.1.3.10
α-D-Glucose-1P
5.4.2.2
3.1.3.9
2.7.1.69
D-Glucose (extracellular)
α-D-Glucose
2.7.1.1
2.7.1.63
2.7.1.2
2.7.1.147
α-D-Glucose-6P
5.1.3.3
5.1.3.15
5.3.1.9
5.3.1.9
β-D-Glucose-6P
β-D-Glucose
β-D-Fructose-6P
Pentose phosphate pathway
Arbutin (extracellular)
Arbutin-6P
3.2.1.86
Salicin (extracellular)
Salicin-6P
3.1.3.11
2.7.1.11
2.7.1.146
β-D-Fructose-1,6P 2
4.1.2.13
Glyceraldehyde-3P
Glycerone-P
5.3.1.1
1.2.1.12
1.2.1.59
Glycerate-1,3P2
5.4.2.4
1.2.7.5
1.2.7.6
1.2.1.9
Glycerate-2,3P2
2.7.2.3
3.1.3.13
Carbon fixation in photosynthetic organisms
Glycerate-3P
5.4.2.1
Glycerate-2P
4.2.1.11
4.1.1.32
4.1.1.49
Oxaloacetate
Phosphoenol-pyruvate
Pyruvate metabolism
2.7.1.40
Citrate cycle
1.2.7.1
ThPP
Acetyl-CoA
2-Hydroxy-ethyl-ThPP
2.3.1.12
1.2.4.1
1.2.4.1
4.1.1.1
Pyruvate
1.1.1.27
L-Lactate
S-Acetyl-dihydrolipoamide-E
6.2.1.1
6.2.1.13
1.8.1.4
4.1.1.1
Dihydro-lipoamide-E
Lipoamide-E
Propanoate metabolism
1.2.1.3
1.2.1.5
1.1.1.1
1.1.1.2
1.1.2.7
EutG
1.1.2.8
Ethanol
Acetate
Acetaldehyde
00010 5/31/12
(c) Kanehisa Laboratories

Appendix 7.7: Photosynthetic apparatus diagram from KEGG painted with significantly ($p < 0.05$) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.




PORPHYRIN AND CHLOROPHYLL METABOLISM
4,5-Dioxopentanoate
Uroporphyrinogen I
Uroporphyrin I
Coproporphyrin I
Coproporphyrinogen I
Ferritin-Fe
Transferrin-Fe
Apoferritin
Apotransferrin
Coproporphyrin III
Uroporphyrin III
5-Amino-levulinate
Porphobilinogen
Hydroxymethyl-bilane
Uroporphyrinogen III
Coproporphyrinogen III
Protoporphyrinogen IX
Protoporphyrin IX
Protoheme (Heme)
Peroxidase
Myoglobin
Cytochrome a
Cytochrome c
Catalase
Hemoglobin
Oxyhemoglobin
Heme O
Heme A
spontaneous
(Anaerobic pathway)
(Aerobic pathway)
Siroheme
Sirohydrochlorin
Precorrin 2
Coenzyme F430
Glutamate-1-semialdehyde
L-Glutamyl-tRNA(Glu)
L-Glutamate
Glycine
Alanine, aspartate and glutamate metabolism
Glycine, threonine metabolism
Co-Sirohydrochlorin
Co-Factor 3
Co-Precorrin 3
Co-Precorrin 4
Co-Precorrin 5A
Co-Precorrin 5B
Co-Precorrin 6A
Co-Precorrin 6B
Co-Precorrin 7
Co-Precorrin 8X
Cobyrinate
Precorrin 3A
Precorrin 3B
Precorrin 4
Precorrin 5
Precorrin 6A
Precorrin 6B
Precorrin 8X
Hydrogenobyrinate
Hydrogenobyrinate a,c diamide
Cob(II)yrinate a,c diamide
Cob(I)yrinate a,c diamide
Adenosyl cobyrinate a,c diamide
Adenosyl cobyrinate hexaamide
Adenosyl cobinamide
Adenosyl cobinamide phosphate
Adenosine-GDP-cobinamide
Cobinamide
(R)-1-Amino-propan-2-ol
(R)-1-Amino-propan-2-yl phosphate
L-Threonine phosphate
L-Threonine
Riboflavin metabolism
Dimethyl-benzimidazole
α-Ribazole-5'-P
α-Ribazole
Vitamin B12 coenzyme
Vitamin B12s
Vitamin B12r
Cyanocob(III)alamin
Aquacob(III)alamin
Magnesium protoporphyrin IX
Mg-Protoporphyrin IX 13-monomethyl ester
Divinyl-protochlorophyllide
Protochlorophyllide
Zn-Bacteriochlorophyll a
Divinyl-chlorophyllide a
Chlorophyllide a
Divinyl-chlorophyll a
Chlorophyll a
Chlorophyll b
Chlorophyllide b
7-Hydroxy-chlorophyllide a
7-Hydroxy-chlorophyll a
Pheophytins
Pheophorbide a
Pyro-pheophorbide a
Red chlorophyll catabolite
Primary fluorescent chlorophyll catabolite
Bacteriopheophytins
Biliverdin
Bilirubin
Bilirubin β-diglucuronide
D-Urobilinogen
D-Urobilin
Mesobilirubinogen
I-Urobilin
L-Stercobilinogen
L-Stercobilin
(3Z)-Phytochromobilin
(3Z)-Phycocyanobilin
(3Z)-Phycoerythrobilin
15,16-Dihydrobiliverdin
Bacteriochlorophyllide c
Bacteriochlorophyll c
Bacteriochlorophyllide d
Bacteriochlorophyll d
Bacteriochlorophyllide b
Bacteriochlorophyll b
Bacteriochlorophyllide a
Bacteriochlorophyll a
00860 1/23/13
(c) Kanehisa Laboratories

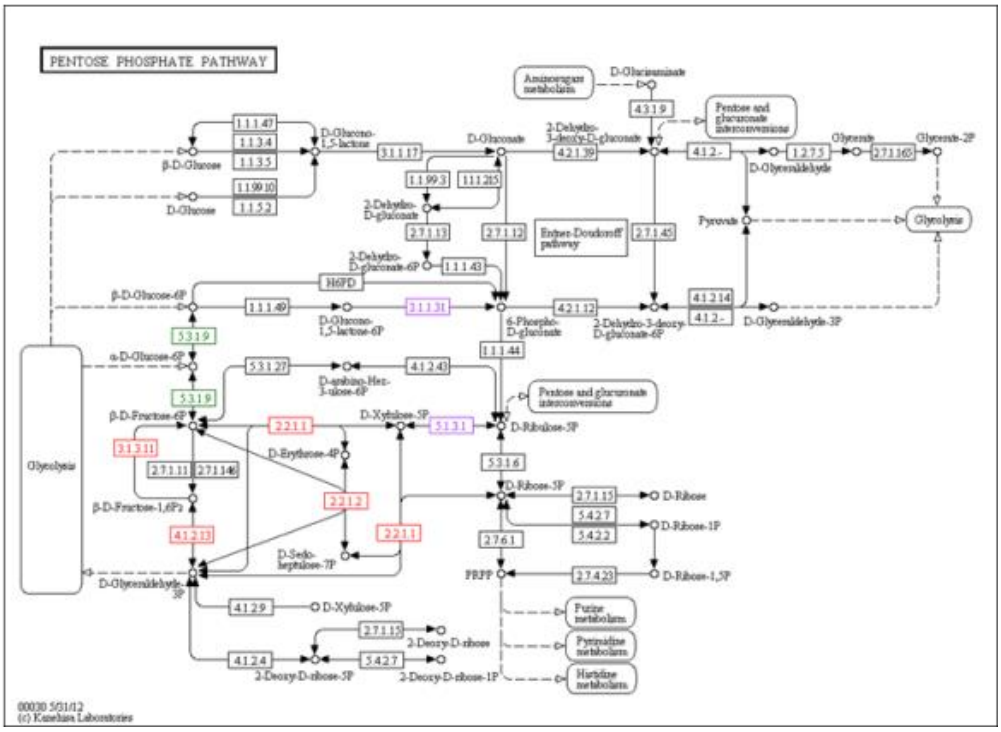
Appendix 7.8: Ribosome diagram from KEGG painted with significantly (p < 0.05) regulated proteins. Increased abundance is shown in blue and decreased abundance shown in red.



Chapter 8 Appendices

Appendix 8.1: KEGG's Pentose Phosphate Pathway of A) C. reinhardtii and B) P. tricornutum. Increased abundance is in dark green outline with blue outline indicating significant changes ($p < 0.05$). Decreased abundance in purple outline with red outline indicating significant changes ($p < 0.05$).

A)



B)