universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Proteomic studies on *Chlamydomonas reinhardtii*“

verfasst von / submitted by

Dipl.-Biochem. Luis Recuenco-Muñoz

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Biologie
Betreut von / Supervisors:	Univ.-Prof. Dr. Wolfram Weckwerth Ass.-Prof. Dipl.-Biol. Dr. Stefanie Wienkoop, Privatdoz.

Declaration of authorship

I, Luis Recuenco-Muñoz, declare that this thesis, titled 'Proteomic studies on *Chlamydomonas reinhardtii*' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Equal goes it loose

(Ernst Goyke)

Aknowledgements

- I wish to thank Prof. Dr. Wolfram Weckwerth and Dr. habil. Stefanie Wienkoop for giving me the chance to work in this utterly interesting field, tutoring and mentoring me throughout my PhD Thesis, and for all the teaching, support, advice and fun I have had both on a working and on a personal level during my whole stint in Vienna.

- Very special thanks go to Dr. Luis Valledor for the extensive and intensive coaching (both *in situ* and *in silico*), the productive work and also his additional readings and corrections on my PhD-Thesis. On this last behalf I would also like to thank Dr. Thomas Nägele.

- I would also like to thank Dr. Wolfgang Hoehenwarter, Dr. Alexander Seidel, Dr. Takeshi Furuhashi, Przemisław Kmiecik, Dr. Maria Angeles Castillejo, Dr. Volker Egelhofer, Dr. Gert Bachmann, Dr. David Lyon, Prof. Dr. Irene Lichtscheidl and Ella Nukarinen, whose help and/or advice have definitely made a difference for my PhD-Thesis. I would also like to thank Dr. Christian Molitor and Dr. Pierre Offre for their fruitful collaboration.

- Since it has been an absolute pleasure to go to work every single day, I would like to thank the whole MoSys Department for creating such a cheerful and productive working atmosphere, especially Lena Fragner, Dr. Mette Hanak, Doris Dell'mour, Dr. Xiaoliang Sun, Yangfan Qiu, Dr. Vlora Mehmeti, Dr. Till Ischebeck, Manuela Leonardelli, Fernando Torralbo and Jana Klimashevskaia. I would also like to thank my good friends Torsten, Netti, Mahtin, Franzi, Natalie, Alois and Jens for their longtime personal support.

- I would finally like to give the most special thanks to my parents, my sister and my brother for their inconditional support at any respect, help, encouragement, and too many other motives to be all named here.

List of Publications

This cumulative PhD-Thesis includes the following publications:

1. Wolfgang Hoehenwarter, Yanmei Chen, Luis Recuenco-Munoz, Stefanie Wienkoop, and Wolfram Weckwerth. Functional analysis of proteins and protein species using shotgun proteomics and linear mathematics. *Amino acids*, 41(2):329–41, jul 2011

2. Luis Valledor, Luis Recuenco-Munoz, Stefanie Wienkoop, and Wolfram Weckwerth. The different proteomes of Chlamydomonas reinhardtii. *Journal of Proteomics*, 2012

3. Luis Valledor, Takeshi Furuhashi, Luis Recuenco-Muñoz, Stefanie Wienkoop, and Wolfram Weckwerth. System-level network analysis of nitrogen starvation and recovery in Chlamydomonas reinhardtii reveals potential new targets for increased lipid accumulation. *Biotechnology for biofuels*, 7:171, jan 2014

4. Luis Recuenco-Muñoz, Pierre Offre, Luis Valledor, David Lyon, Wolfram Weckwerth, and Stefanie Wienkoop. Targeted quantitative analysis of a diurnal RuBisCO subunit expression and translation profile in Chlamydomonas reinhardtii introducing a novel Mass Western approach. *Journal of Proteomics*, 113:143–153, 2015

Any other publications quoted in this PhD-Thesis that are coauthored by me will be cited in bold letters.

Contents

List of Abbreviations

2DE= Two-Dimensional-Electrophoresis
AAE= Acyl-ACP Esterase
ACN= Acetonitrile
ACP= Acyl-Carrier Protein
AFAT= Algal Functional Annotation Tool
AGC= Automatic Gain Control
ALE= Adaptive Laboratory Evolution
APEX= Absolute Protein EXpression (Profiling)
AQUA= Absolute Quantification
ASLs= Annotated Spectrum Libraries
AspN= Endoproteinase: Flavastacin
AUC= Area Under Curve
BAC= Bacterial Artificial Chromosome
BHT= Butylated Hydroxytoluene
BLAST= Basic Local Alignment Search Tool
BN-PAGE= Blue Native PolyAcrylamide Gel Electrophoresis
BRI1= Brassinosteroid insensitive (mutant) 1
BSA= Bovine Serum Albumine
C18= Octadecyl (reversed-phase-chromatography sorbent material)
CAH= Carbonic Anhydrase
CCM= Carbon Concentrating Mechanisms
cDNA= Complementary DesoxyriboNucleic Acid
CI= Chemical Ionisation
CID= Collision-Induced Dissociation
COX= CycloOxygenase
CR= *Chlamydomonas reinhardtii*
DB= Database
DDA= Data Dependent Acquisition
DGAT= DiacylGlycerol AcylTransferase
DIA= Data Independent Acquisition
DIGE= Difference Gel Electrophoresis
DNA= DesoxyriboNucleic Acid
DTT= DiThioThreitol
EBI= European Bioinformatics Institute ECD= Electron Collision Dissociation
EMPAI= Exponentially Modified Protein Abundance Index
ESI= Electron-Spray-Ionization
EST= Expression Sequence Tag
EtOH= Ethanol
FAB= Fast Atom Bombardment
FAME= Fatty Acid Methyl Ester

FDR= False Discovery Rate
FPLC= Fast Performance Liquid Chromatography
FT= Fourier Transformation FTICR= Fourier Transformation Ion Cyclotrone Resonance
FT-MS= Fourier Transformation Mass Spectrometry
GEO= Gene Expression Omnibus
GC-MS= Gas Chromatography (coupled) Mass Spectrometry
GluC= Endoproteinase: Staphylococcus aureus Protease V8
GPMDB= Global Proteome Machine DataBase
HA= Heteroautotrophic
HCD= High-energy Collisional Dissociation
HILEP= Hydroponic Isotope Labelling of Entire Plants
HILIC= Hydrophilic Interaction Chromatography
HPI= Human Protein Index
HPLC= High Performance Liquid Chromatography
HUGO= Human Genome Organization
HUPO= Human Proteome Organization
IC= Independent Component
ICA= Independent Component Analysis
ICAT= Isotope-Coded Affinity Tags
ICL= Instrument Control Language
IEA= International Energy Agency
IFT= IntraFlagellar Transport
IRMPD= Infrared multiple photon dissociation (spectroscopy)
iTRAQ= Isobaric Tag for Relative and Absolute Quantification
JGI= Joint Genome Institute (http://genome.jgi-psf.org/)
JIPID= Japan International Protein Information Database
KAAS= KEGG Automatic Annotation Server
KEGG= Kyoto Encyclopedia of Genes and Genomes (www.genome.jp/kegg/)
LB= Lipidic Bodies
LECA= Last Eukaryotic Common Ancestor
LHC= Light Harvesting Complex
LC= Liquid Chromatography
LC-MS= Liquid Chromatography-coupled Mass Spectrometry
LSU= (RuBisCO) Large Subunit
LTQ= Linear Trap Quadrupole
m/z= Charge Ratio (Fragment Mass / Charge)
MLDP= Major Lipid Droplet Protein
mRNA= Messenger Ribonucleic Acid
MALDI= Matrix Assisted Laser Desorption Ionisation
MAPA= Mass Accuracy Precursor Alignment
MASCP= Multinational Arabidopsis Steering Committee - Proteomics (Subcommittee)
MIPS= Munich Information Center for Protein Sequences
MO= MicroOrganism
MPI= Mean Protein Intensity
MRM= Multiple Reaction Monitoring
MS= Mass Spectrometry
MudPIT= Multidimensional Protein Identification Technology

MW= Molecular Weight
NBRF= National Biomedical Research Foundation
NCBI= National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov)
NIST= National Institute of Standards and Technology
NGS= Next Generation Sequencing
NSAF= Normalized Spectral Abundance Factor
OMSSA= Open Mass Spectrometry Search Algorithm
OmpT= Outer membrane protease T
PA= PhotoAutotrophic
PAI= Protein Abundance Index
PBR= PhotoBioReactor
PC= Partial Component
PCA= Partial Component Analysis
PIR= Protein Information Resource
PIR-PSD= PIR-International Protein Sequence Database
PLS= Partial Least Square (Analysis)
PMSF= Phenylmethanesulphonylfluoride
PN= (unique) Peptide Number
ppm= parts per million
PRIDE= PRoteomics IDEntifications (Database)
PRM= Parallel Reaction Monitoring
ProMEX= Protein Mass Spectral Extraction (http://promex.mpimp-golm.mpg.de/)
PS= PhotoSynthesis
PSAQ= Protein Standard Absolute Quantification
PSI= Proteomics Standards Initiative
PSM= Peptide per Spectrum Matches
PTM= Post-Translational Modification
QFISH= Quantitative Fluorescent In Situ Hybridization
rbcLS= RuBisCO Large Subunit (transcript)
rbcSS= RuBisCO Small Subunit (transcript)
RF= Retention Factor
RNAi= Ribonucleic Acid interference
RP= Reverse Phase
RSc= (Protein) Ratio from Spectral Counting
RSU= RuBisCO SubUnit
RT= Retention Time
RuBisCO= Ribulose Bisphosphat Carboxylase Oxygenase
SAX= Strong Anion Exchange
$scCO_2$= Super Critical Carbon Dioxyde
SCX= Strong Cation Exchange
SDS-PAGE= Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis
SE= Size Exclusion
SIB= Swiss Institute of Bioinformatics
SILAC= Stable Isotope Labeling with Amino acids in Cell culture
SILIP= Stable Isotope Labeling In Planta
SIn= normalized Spectral Index
SISCAPA= Stable Isotope Standards and Capture by Anti-Peptide Antibodies

sPLS= Sparse Partial Least Square (Analysis)
SRM= Single Reaction Monitoring
SSU= (RuBisCO) Small Subunit
TAG= Triacylglyceride
TOF= Time of Flight (Mass Spectrometer)
TMT= Tandem Mass Tag
TPM= Targeted Peptide Monitoring
TPP= Trans-Proteomic Pipeline
Tris= Tris(hydroxymethyl)aminomethane
TSQ= Triple Quadrupole (Mass Spectrometer)
UHPLC= Ultra-High Performance Liquid Chromatography
UV= Ultra Violet
WCX-HILIC= Weak Cation Exchange - Hydrophilic Interaction Chromatography
WT= Wild Type
Xcorr= Correlation factor

1 Introduction

1.1 Proteomics

1.1.1 The Foundations of Proteomics

What is "Proteome"

The term "**Proteome**" was coined by geneticist Marc Wilkins in 1994 in a symposium on "2D Electrophoresis: from protein maps to genomes" held in Siena in Italy. It appeared in print in 1996([340] [341]). If Proteome was sparsely described as "the PROTEin complement expressed by a genOME", the resulting field of Proteomics comprises the large-scale study of protein properties (expression level, post-translational modification, interactions etc.) of a tissue or an organism at a certain time point to obtain a global, integrated view of disease processes, cellular processes and networks at the protein level [30].

Protoproteomics

The origins of Proteomics go back to late 70s, even before the HUGO project was set up: after the development of 2D-Electrophoresis by Klose, O'Farrell and Scheele [151] [225] [262] there was such an optimism with its capabilities that it was considered to establish a Human Protein Index (HPI), a kind of equivalent of the periodic table featuring systematic enumeration of human proteins, even leading the initiation of a Molecular Anatomy Program at the Argonne National Laboratory and the formation in 1980 of the Human Protein Index Task Force (unluckily, the lack of political consensus and the poor image of large-scale science at these days doomed the initiative from the start) . Nevertheless, 2D-Electrophoresis still established itself as the method of choice for visualising and separating complex protein mixtures: due to the

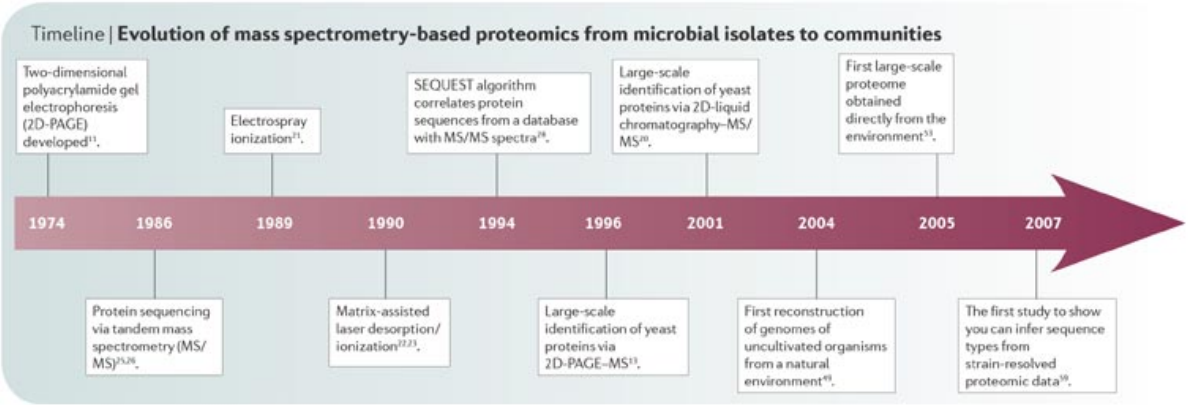

Figure 1.1: Proteomics: from the development of 2D-PAGE to its current application in microbal community studies (VerBerkmoes et al. 2009) [317]

high resolution and reproducibility of immobilized pH gradients that allowed the pI and the MW of the proteins to be predicted , it could display up to 10000 proteins simultaneously in a single gel (with approximately 2000 proteins being routine), and detecting and quantifying protein amounts of less than 1ng per spot [183]. 2D-Electrophoresis is a standard procedure for autoradiographic imaging [100] and through different staining techniques (such as Coomassie blue, silver staining or fluorescent dyes as in Difference Gel Electrophoresis (DIGE)) it allows a quick visualisation of differences between two different protein extracts or alterations on a certain protein in a complex sample (as degradation or PTMs) [210][246].

Peptide Volatisation and Ionisation

The other technical breakthrough that paved the way to Proteomics before the concept came to existence is the ability to volatise and ionise peptidic molecules. Since the first applications of Mass Spectrometry in Protein Analysis there has been a huge evolution: despite some previous experiments on Peptides with Field Desorption Mass Spectrometry, it wasn't until the development of the **FAB (fast atom bombardment)** soft ionisation technique by Morris (1981) that peptides could be brought as intact, singly protonated/deprotonated molecules in the gas phase for MS analysis [215]. The first applications were the determination of peptide molecular mass, and the elucidation of the aminoacid sequence of a certain peptide with higher speed and efficiency than with the classical Edman-Degradation technique, especially after the development of **TANDEM-Mass Spectrometry** by Hunt (1981)[133], incorporating the use of quadrupole mass spectrometers to create fractionation patterns and measure both parent and fragment ions. A specific nomenclature for peptide fragments was originally suggested by Roepstorff (1984) [256] and improved by Johnson (1987) [139]. The next decisive step into the development of Proteomics is, as previously mentioned, the development of more sophisticated soft ionization techniques that could not only allow, especially two: the **Electrospray Ionisation (ESI)** and the **Matrix-Assisted Laser Desorption Ionisation (MALDI)**.

Electrospray Ionisation (ESI) Developed by Fenn and colleagues [84], the Electrospray Ionisation is a procedure in which peptides from a steady flow enter the gas-phase through a voltage-driven desolvatisation process (Coulomb explosions), thus resulting a spray of single-charged peptide ions directed to the nozzle of the Mass Spectrometer for further analysis. This makes possible to couple the Mass Spectrometer with liquid chromatographies, thus enabling continuous measurements. Furthermore, applying different separation strategies at the chromatography - such as reverse phase (RP), size exclusion (SE),or hydrophilic interaction (HILIC) — more comprehensive online studies can be performed on a certain subject. Currently, the most used setting for LC-MS-Proteomics is an RP-column on a UHPLC combined with linear quadrupole Mass Spectrometers. For these achievements, Fenn was awarded the Nobel Prize of Medicine 2002.

Matrix-assisted Laser Desorption Ionisation (MALDI) Developed by Tanaka, Karas and Hillenkamp towards end of the 1980s [290] [125], this soft ionisation technique was an evolution on previous laser desorption techniques (until then only capable of ionising particles up to 1000 Da): biopolymers would be embedded on a stable, slightly acidic polymer matrix (the exact composition can be adjusted depending on the object of study) that would get excited by bombardment with short duration laser pulses, and subsequently transmit that excitation energy to the embedded molecules. The resulting ions could be then measured with the also

pulse-based **time-of-flight (TOF)** MS-technique: ion pulses excited under an electric field get separated according to their trajectory to the detector — direcly proportional to the m/z-ratio. TOF mass spectrometers are neither the quickest instruments nor well suited for continuous measurements, but they feature very high mass accuracy and range). MALDI-TOF also allowed for detection of higher charged molecules. Tanaka received the Nobel Prize of Medicine for MALDI in 2002.

Peptide Fragmentation

Peptide sequence identification by tandem mass spectrometry involves ionization of volatized peptides or even proteins, usually getting protonated and therefore scanned by the mass spectrometer on positive mode. Since ESI brings analytes in the gas phase without rupturing any covalent bonds, it provides the molecular weight via m/z, but further fragmentation of these analytes is needed to confirm the aminouacid sequence and determine the structure of an unknown molecule [351] [14]. Some of the different ion activation and fragmentation mechanisms are featured on Table 1.1.1 [278] [162][351] [119].

Table 1.1: Overview: different ion activation methods

Ion activation	Related MS	Application
IRMP (infrared multiphoton dissociation) Little et al., 1994 [177]	IT, FTCIR	Unimolecular dissociation Structural elucidation top-down proteomics
SID (surface induced dissociation) Mabud, 1985 [184] Cooks, 1990 [52] Laskin, 2000 [161]	IT, Hybrid TOF, FTCIR, TSQ	Specific fragmentation pathways, isomer distinction
SORI (sustained off-resonance irradiation) Gauthier, Trautman, Jaconson, 1991 [90]	FTCIR	Radial separation of mass-selected ions, MSn structure determination
CID (collision induced dissociation) Jennings, 1968 [137]	IT, TOF, FTCIR, LTQ, TSQ, Orbitrap	Peptide dissociation Most extended method
PQD (pulse Q induced dissociation) Schwartz et al, 2005 [269]	Orbitrap (Velos)	Overcomes low-mass cut-off (Peptide tag identification)
HCD (higher energy collisional dissociation) Olsen 2007 [226]	Orbitrap (Velos)	Full mass range with high accuracy e.g. PRM
ECD (electron collision dissociation) Zubarev 1998 [361]	FTCIR	Peptide backbone dissociation conserving PTMs; multiply charged protein/peptide ions
ETD (electron transfer dissociation) McLafferty 2001 [196]	modifield LTQ QTOF, OTOF	Peptide backbone dissociation conserving PTMs; multiply charged protein/peptide ions

Ideally, measured m/z values of these pieces can be assembled to produce the original sequence (direct sequence ions) [117]. However, some peptides, due to their physical-chemical properties, cannot be efficiently ionized or fragmented, therefore producing MS/MS spectra unidentifiable by the current database search tools [221]. It can also happen that many fragment ions undergo complex rearrangements and therefore differ from the direct sequence ion series (also known as nondirect sequence ions) [117]. Many models have been proposed to explain the behaviour of ionized proteins and peptides, such as the "mobile proton model", defined initially in several publications by Wysocki et al. as a qualitative description of peptide fragmentation [351], and further expanded to a more quantitative-oriented models such as the "Pathways in Competition" model of Paizs [231] and the kinetic model of Zhang [357]. Even though it is commonly accepted that peptide cleavage occurs predominantly through charge-directed pathways (i.e. initiated by a charge that is transferred to the vicinity of the cleavage site [351]), questions such as the dependence of a fragmentation pattern on the aminoacid sequence of the precursor peptides or the stabilization of the correspondent peptide fragmentes still remain controversial [14].

Peptide properties affecting ionization and fragmentation Despite the great complications to dilucidate the processes leading to backbone fragmentation, many of the properties of the peptides can be used to predict loosely how the analyte might ionize and fragmentate:

- Peptide length (the longer, the more energy is needed to fragment)
- Peptide hydrophobicity (higher insolubility in the buffers compatible with electrospray difficults peptide ionization — lower peptide abundance)
- Higher charged peptides require less energy to fragment
- Cyclic peptides feature poorer fragmentation efficiency
- Amino acid sequence: different aminoacids can produce different effects depending on the type of fragmentation
- on MALDI: K terminated peptides give more evenly distributed patterns than R terminating peptides
- on CID: R, K or H on either peptide terminus can lead to production of b/y ions
- on CID: Proline enhances fragmentation (as C-side of His, Asp and Blu, and on Asp-Pro), on the other hand Gly frequently suppresses it, particularly Gly-Gly or Gly-Ala.

Besides, the collision gas used at the ion source also plays a role, being peptide fragmentation proportional to molecule size and gas pressure.

Peptide fragment nomenclature A nomenclature for peptide fragmentation was already established in the middle 80s by Roepstorff and Johnson [256] [139]. This system both describes the fragment ion types that are produced by cleavage of different bonds along the peptide backbone and/or fragments from side chains.

Cleavage of the backbone typically occurs at the peptide amide bond, producing **b** ions in case the amino terminal fragment retains the charge, or **y** ions if the carboxy-terminal fragment retains the charge (it's notable that **b** and **y** ions are not always present at the same spectrum

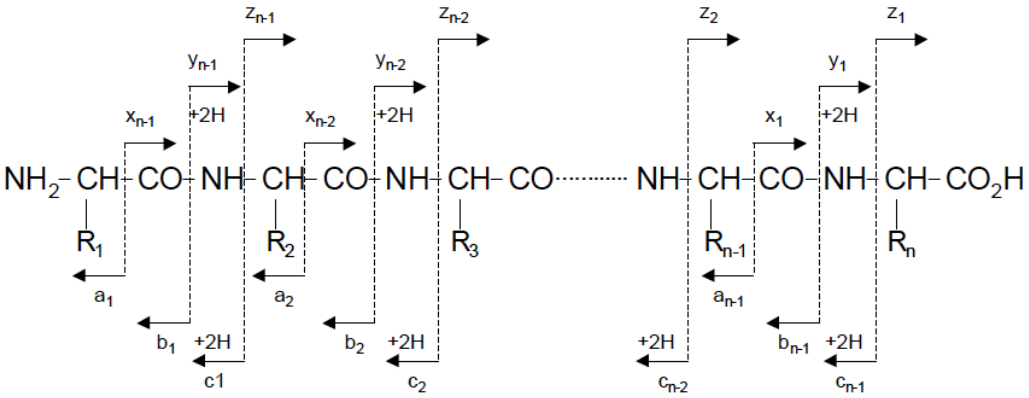

Figure 1.2: Roeppstorf Nomenclature Scheme for Peptide Fragmentation Patterns — Production of **a**-/**x**-Ions, **b**-/**y**-Ions, **c**- Ions, **y**-Ions.

due to different stability). Beside **b** and **y** ions, considered to be the most useful sequence ion types, other ion types can be observed — for instance, **a** ions are the result of **b** ions losing a CO group (thus, a m/z difference of 28 between two peaks could hint at an **a**-**b** ion pair). Alternative fragmentation methods (such as ECD fragmentation - see Table 1.1.1) produce **c**/**z** ion pairs, with the cleavage at the N-$C\alpha$ bond. This comes as a result of the energy pathway of these alternative methods being lower, so that modifications that would usually split with neutral loss through CID fractionation are still conserved — which is a huge advantage for PTM characterization [78]. Other relevant ions are side-chain cleavage ions (also known as **d**, **v** and **w** ions), which result out the cleavage of side-chain bonds, are allow distinction between isobaric or isomeric ions (e.g. Ile from Leu) [351].

Mass Spectrometry

First developments in Protein Mass Spectrometry One of the classical dilemmas in Protein Mass Spectrometry is perfectly reflected in the two classic mass spectrometer types that became most prominent: FTICR and TOF mass spectrometers. **FTICR** (Fourier Transformation Ion Cyclotrone Resonance) spectrometers offered high resolution and accuracy at cost of poor multiplexing and relatively low speed; meanwhile, **TOF** (Time-of-Flight) mass spectrometers had a wide dynamic range and high speed but due to poor transmission and pulsed ionisation they were not very appropiate for continous measurements e.g. coupled to LCs. Also Ion Traps, despite great transmission and sensitivity, lacked the speed and the accuracy of the FTICR. The more primitive **Sector Field** Mass Spectrometers or **Linear Quadrupole** Spectrometers were also stuck to similar limitations, bringing great speed at costs of resolution and sensitivity.

The Orbitrap Around 2000, Dr. Alexander Makarov presented a new mass spectrometer design that brought many advantages from the previous types: improving on an original idea of Kingdon, it became possible to stabilize ions through pulsed injection into an electrostatic field, bringing them to rotate in orbitals around a central electrode inside an ion trap. The LTQ Orbitrap featured two coupled ion traps (linear and orbital) in which a first linear quadrupol that works both as a first detector and as an ion source for a second, orbital-shaped trap performing high accuracy ion measurements. Thus, the LTQ Orbitrap allowed parallel mea-

surement of parent and fragment ions with very high mass resolution. Further developments of the Orbitrap aimed principally at increasing its transmission (and consequently also its speed) through use of S-lenses for better focusing and filtering out of neutrals before entering the first quadrupole, lowering of the ion cut-off values (thus broadening its accuracy for fingerprint ions enabling the measurement of reporter ions and so the application of techniques like iTRAQ) and increasing mass accuracy and PSM (peptide per spectrum matches) through more effective fragmentation methods (see Table 1.1.1). Indeed, many improvements on Mass Spectrometers have made possible to incorporate alternative fragmentation techniques to the standard CID procedure (see Table 1.1.1), both on hardware (e.g. through adding an additional chamber for HCD fragmentation [226]) and software level (introduction of Data-Dependent Decision Trees for combining different fractionation strategies as e.g. Swaney 2008 [288]).

Latest Developments For Protein MS-Analysis there has always been the compromise between high mass accuracy and resolution on one hand, and speed and sensitivity on the other — the recent developments on Orbitrap Mass Spectrometers (Thermo) aim at reaching enough speed and sensitivity to compete with previous triple quadrupole models hence being suitable also for quantitative protein/peptide analysis. New devices incorporated to the original LTQ-Orbitrap construction, such as the S-Lens or straight/bent flatapoles from the Q-Exactive Spectrometers (also from Thermo), aim at better ion focusing and filtering out neutrals. Other alternatives like the triple TOF from ABSciex ditch the DDA, rather focusing on speed and powerful post-acquisition processing; another alternative from another manufacturer is the development of iFunnel technology (aiming to increase the amount of measured ions minimising ion loss on the way from the ion source to the mass analyzer) by Agilent. Still, there are still manufacturers splitting clearly between high accuracy and high speed measurements e.g. Bruker with the maXis II models.

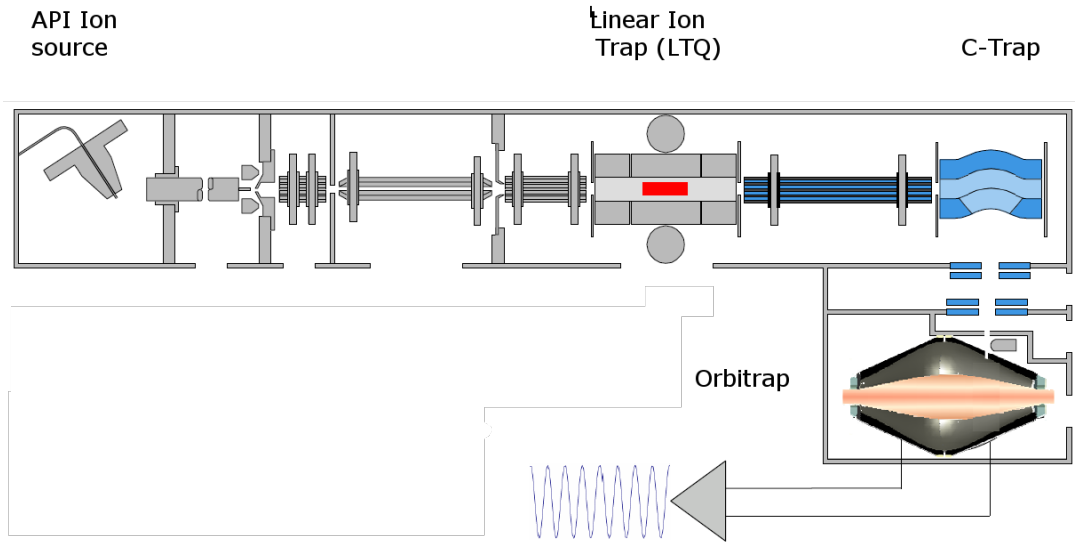

Figure 1.3: Original LTQ-Orbitrap structure

Table 1.2: Overview: current Mass Spectrometer models and their corresponding features

Instrument	Applications	Resolution	Mass Accuracy	Sensitivity	Dynamic range	Scan rate
LIT (LTQ)	Bottom-up protein identification in high-complexity, high -throughput analysis, LC-MS^n capabilities	2000	100 ppm	Femtomole	1e4	Fast
TQ (TSQ)	Bottom-up peptide and protein quantification; medium complexity samples, peptide and protein quantification (SRM, MRM, precursor, product, neutral fragment monitoring)	2000	100 ppm	Attomole	1e6	Moderate
LTQ-Orbitrap	Protein identification Protein quantification PTM identification	100.000	2 ppm	Femtomole	1e4	Moderate
LTQ-FTCIR, Q-FTCIR	Protein identification Protein quantification PTM identification top-down protein identification	500.000	<2 ppm	Femtomole	1e4	Slow, slow
Q-TOF, IT-TOF	Bottom-up protein identification top-down protein identification PTM identification	10.000	2-5 ppm	Attomole	1e6	Moderate, fast
Q-LIT	Bottom-up peptide and protein quantification; medium complexity samples, peptide and protein quantification (SRM, MRM, precursor, product, neutral fragment monitoring)	2000	100 ppm	Attomole	1e6	Moderate, fast
Orbitrap VELOS Pro	Intact Antibody Characterization, Leachables and Extractables Drug Impurities Analysis	> 100.000	< 1ppm RMS		5,000 within a single scan guaranteeing specified mass accuracy	-
Q-Exactive	Untargeted proteomics and metabolomics, targeted screening	140.000	< 1ppm RMS (internal), 3ppm RMS (external)	Full MS: 500 fg Buspirone column S/N 100:1 SRM: 50fg Buspirone column S/N 100:1	5000:1 (intra-scan)	<12 Hz
Orbitrap ELITE	Bottom-up Proteomics targeted quantitation, top-down proteomics intact protein characterization PTMs (glycosylation, phosphorylation) relative quantitation (SILAC, TMT, label-free)	15.000- > 240.000	< 3 ppm RMS (external calib.) < 1 ppm RMS (internal calib.)	2 μ L of a 50 fg/μL reserpine solution (100 fg total) 500μl /min S/N 100:1 (HESI)	5000 (single scan)	-
Q-Fusion	Proteins Bottom-up Proteomics targeted quantitation, top-down proteomics intact protein characterization PTMs (glycosylation, phosphorylation) relative quantitation (SILAC, TMT, label-free)	450.000	< 3 ppm RMS (external calib.) < 1 ppm RMS (internal calib.)	2 μL of a 50 fg/μL reserpine solution (100 fg total) S/N 100:1 (HESI)	5,000 within a single scan	< 20

Computer Tools and Strategies

Another key element for the consolidation of Proteomics was the development of computer tools to both optimize the measurements in the Mass Spectrometers and to correlate the results to peptides and proteins.

From the first Algorithms to Data Acquisition Strategies Even though there had already been worked on algorithms for capital aspects as multiply charged ions or peak deconvolution [189], the first truely disruptive innovation was the introduction of **Instrument Control Language (ICL)** on the Finnigan MAT TSQ-70 (Lammert, S. A. Finnigan-MAT Technical Report Number 603, 1987). ICL could be used to create computer programs to interact with data and to control the instrument operations based on that data in real time [355]. Therefore, it was posible to establish automated methods for investigation of fundamental instrument parameters, or to maximize information per analysis time unit — early applications were e.g. automatic acquisition of an MS/MS scan for every parent ion found in the mass spectrum above a given threshold with the mass spectrum not being known in advance [118]. These methods are commonly known as **Data-dependent-acquisition (DDA)** and have decisively contributed to the development of large-scale MS/MS measurements — in fact, the DDA approach has been the most common data acquisition strategy for peptide analysis since the development of Shotgun Proteomics. For instance, on a current LC-MS/MS analysis a mass spectrometer can sequentially survey all the peptide ions that elute from the LC column at a particular time MS^1 and subsequently isolate and fragment a selection of parent ions from that MS^1 (usually the top few most intense) to generate MS/MS (MS^2) spectra [304]. The increasing speed and resolution of mass spectrometers, and the development of more powerful software has contributed to alternative methods to complement DDA, like the **Data Independent Acquisition (DIA)** [316]. DIA features no data selection for TANDEM analysis, but rather a serial fragmentation of all measured peptides in different isolation windows (mass, time $\cdots$) [170]. Even if DIA measurements need existing libraries and selects mostly already known peptides — therefore not being suitable for discovery-based applications [67] — it overcomes the limitations of dynamic range of DDA, which filters a considerable amount of low abundant peptide signals. Since its introduction with the Waters Expression System™ in 2002, Data Independent Acquisition has already been further implemented in different approaches, such as **FT-ARM** (spanning defined m/z ranges and assigning a score based on comparison with empirical and theoretical peptide fragment spectra [332], **multiplexed DIA** (random selection of five non-overlapping smaller m/z windows that got analysed simultaneously, instead of sequentially — Egertson 2013 [76]) or **MSPLIT-DIA** (mixture-spectrum partitioning using libraries of identified tandem mass spectra, using spectrum projections to match library spectra to each DIA spectrum [324])). A comprehensive review of DIA methods can be found on Bilbao 2015 [26]. Other applications of ICL on mass spectrometry measurements are e.g. switching between ionization types or creating decision algorithms [288].

Peptide Sequence Assignment A critical aspect to make protein analysis quicker and more effective was the establishment of sequence databases to which single peptides could be matched without having to sequence the whole protein. Since Henzel introduced the first algorithms to match mass spectral data to digital protein databases at the Third Symposium of The Protein Society in 1989 [25], many different algorithms have been developed for this task.

The most popular algorithms are SEQUEST (Yates 1994) [79] and MASCOT (Perkins 1999)

[238]. The SEQUEST algorithm converts the character-based representations of amino acid sequences from a protein database to a fragmentation pattern that can subsequently be used to match fragment mass values to an uninterpreted tandem mass spectrum, then picking putative ions for each peptide, performing a cross-correlation analysis between theoretical and experimental mass spectra, giving every peptide match a corresponding X-corr factor, and finally selecting the highest scoring matches. The MASCOT algorithm [238] enhances with MS/MS-Ion search other capabilites of its precursor, the MOWSE algorithm [234], such as peptide mass fingerprinting and sequence query. Furthermore, MASCOT adds a probability-based score for each identified peptide, and a further Protein score gathering all significant peptides of a single protein.

Other newer algorithms, mostly open-source, have been recently developed, e.g. OMSSA [91] (in collaboration with the NCBI institute) or X!Tandem (Craig, Cortens, Beavis 2004 [56]). These algorithms can also be combined with search engines, such as MaxQuant [53] or the more recent MS-Amanda [70].

Validation Strategies While these algorithms are very powerful, there can however be substantial overlap between the scores for correct and incorrect peptide identifications [144] — therefore, additional validation must be made. The standard procedure is the **False Discovery Rate (FDR)**, introduced into proteomics procedures in 2007 and mainly applied in two approaches: (a) *empirical Bayes approach* (from the overlap of two score distributions corresponding to correct and incorrect identifications, being the FDR computed as the fraction of the mixture density attributable to the incorrect distribution above a given threshold — nevertheless both distributions are modelled for every measurement, being highly variable depending on the search engine and the dataset) and (b) *target-decoy searching*, more extended in sequence searching (not spectral searching), in which peptide search with both the original and a decoy database (featuring false peptide sequences e.g. inverted) are concatenated, being the FDR computed out of the probability of finding false hits in the final search results [157]. Many algorithms to generate decoy databases can be found on Hoopman and Moritz 2013 [128]. Nowadays, most search engines have integrated FDR in their toolboxes [155]. It has also been discussed to apply **p-Value correction** in multiple testing since FDR can be considered additive (originally via Bonferroni correction). However, this procedure might be too stringent depending on the type of the experiment, eliminating too many true positives as false negatives. The usual FDR value for confidence intervals in proteomics experiments is 5%.

Other factors affecting the quality of the peptide identifications/score factors can be the quality of the spectra and, depending on the used algorithm or search engine, search time or the size of the used database.

Development of Protein Databases and Repositories

The first comprehensive database of classified and functionally annotated protein sequences was produced by the **Protein Information Resource** (**PIR**, established in 1984 by the National Biomedical Research Foundation, NBRF) in collaboration with MIPS (Munich Information Center for Protein Sequences) and JIPID (Japan International Protein Information Database). This dataset was built from the already collected Atlas of Protein Sequence and Structure (Margaret Dayhoff, 1965–1978) — full releases have been published quarterly since 1984 [17] and finally presented in 2004 as the **PIR-International Protein Sequence-Database** (**PIR-PSD**). Besides, PIR joined EBI (European Bioinformatics Institute) and

SIB (Swiss Institute of Bioinformatics) to form the *UniProt* consortium, integrating all its sequences and annotations in the UniProt Knowledgebase. Bidirectional cross-references between UniProt (UniProt Knowledgebase and/or UniParc) and PIR-PSD are established to allow easy tracking of former PIR-PSD entries. Furthermore, PIR-PSD unique sequences, reference citations, and experimentally-verified data can also be found in the relevant UniProt records. Currently there are 21.176 entries (12.02.2016), 705 of which are reference proteomes (http://www.uniprot.org/taxonomy/complete-proteomes).

There are two further active databases of high relevance (and with daily updates):

- **GPMDB (Global Proteome Machine and Database)**, which is built after reprocessing the information collected in the GPM servers with the X!Tandem algorithm [55] and focuses on holding the minimum amount of information necessary for certain bioinformatics-related tasks, such as sequence assignment validation. Most of the data is held in a set of XML files: the database serves as an index to those files, allowing for very rapid lookups and reduced database storage requirements (http://thegpm.org).
- **PeptideAtlas (Desiere 2006) [65]** Mass spectrometer output files are loaded into the PeptideAtlas data repository, after which they are analyzed through the Trans-Proteomic Pipeline (TPP) using either Sequest or X!Tandem as database-dependent search algorithm. Further analysis with PeptideProphet and ProteinProphet derives a probability of correct identification for all results in a uniform manner. Besides, to ensure a high quality database, false discovery rates at the whole atlas level are also included.

Other repositories were not as successful due to funding problems, like e.g. the **Proteome Commons Tranche repository** — linked to the ProteomeCommons.org community in order to redundantly store and disseminate data sets for the proteomics community. It featured support for large data files, prepublication access controls, licensing options, and it ensured both data provenance and integrity ([280], linked to [80] [optional]). The **NCBI-Peptidome** [138] had a similar fate: both repositories discontinued on 2011 and migrated to massIVE and PRIDE, respectively [236], [59].

Proteogenomics Even though a big majority of protein databases comes from translation of genomic databases (an overview of all the genome projects to this date can be found in https://gold.jgi.doe.gov/ [254]. Proteomics can also play a very important role for the development of protein databases, e.g. in database enhancement through functional annotation or even for constructing metabolic networks (as in **May et al 2008** [195]). In fact, the recently defined field of **Proteogenomics** (studies in which proteomic data are applied to improve genome annotation and characterization of the protein-coding potential) aims at protein-level validation of expression of the novel transcripts nominated by genomic and transcriptomic technologies. Proteogenomics includes such procedures as confirmation of translation, reading-frame determination, identification of gene and exon boundaries, evidence for post-translational processing, identification of splice-forms including alternative splicing, and also, prediction of completely novel genes [45]. Furthermore, it also deals with computational strategies for building and using customized protein sequence databases, e.g. including EST sequences or six-frame translation of cDNAs [220]. Many of these strategies have been applied at the development of the mentioned repositories — the SwissProt section even features manual annotation. Nevertheless, for plant proteomics this procedure has barely been applied on other organisms than *Arabidopsis thaliana* and *Oryza sativa*, remaining still the rest of the proteomes ([266]).

Actual Situation of Protein Databases Good introductions to Proteomics Databases and repositories have been written by Nesvizhskii [219] or Vizcaíno [321] — more actual developments can be read under (Perez-Riverol 2015 [236]) or, for plant proteomics, under Sakata & Komatsu 2014 [141] — including a philogenetic tree of proteomics repositories for plants. A challenge coming out of this variety in repositories is how to avoid data dissemination and unify or coordinate data for collaborating groups — the MASCP recommendation is to submit data to PRIDE because of its compliance with standards developed by the **Proteomics Standards Initiative** (**PSI**) [228] [292] and because it can harbour basically any kind of processed proteome data (http://www.masc-proteomics.org/mascp/index.php/Data). Other initiatives to overcome this pitfall are the **ProteomeXchange** consortium (Ternent 2014 [296]; http://www.proteomexchange.org) and the **Biomart** platform (http://www.biomart.org), which has been recently developed as a way to integrate large datasets and biomedical datasets ([113] [279]). Other key aspects regarding protein database standardization are database size, indexing, clustering vs filtering approaches and correction factors/scores. An important shortcoming is the scarce development of tools to evaluate database performance, especially considering that protein sequence databases show different levels of completeness and sequence redundancy. This issue is adressed on the publication "**The different proteomes of *Chlamydomonas reinhardti***" **[310]**, included in this PhD-Thesis.

Spectral Libraries Spectral library searching has been proposed as a useful complement and in some cases even as a promising alternative to sequence database searching ([353]). In this approach, peptide identification is made by comparing the query MS/MS spectrum to a library of reference spectra for which the identifications are known and selecting candidates using a set of filters (as in Domokos 1984 [68]). This strategy (previously known as **Annotated Spectrum Libraries**, **ASLs**) has been commonly practiced for mass spectrometric analysis of small molecules since the 1950s ([158]), but it was in the 1990s when spectral libraries were brought in connection with cross correlation methods ([230]) and applied on protein analysis as a way to enhance the current manual interpretation of MS/MS spectra ([353]). The development of computational methods was not only peptide identification through MS/MS spectra assignment, but also for validating experimental against experimental spectra, as in SILVER [97] [77]. Nevertheless, both theoretical spectra and generating spectra from synthesized peptides have been too costly — instead, an alternative strategy was to develop quality controls and curation schemes that could help building sets of composite spectra, each the result of averaging together multiple observations of the same peptide [54]. In fact, the exponential growth of data volume owing to rapid experimental advances in the field played an important role in the development of spectral libraries [156].

Spectral libraries have many advantages to sequence assignment database queries: smaller databases can save a lot of time for every search (e.g. the P3 Algorithm by Craig 2005 managed 20fold time reduction compared to standard searches with whole protein sequences [57]), fragment intensity distributions in product ion spectra (nonexistent in sequence search) can be taken into account as an additional parameter to match unknowns and avoid false positives, and the storage of high quality spectra from standard peptides featuring modifications such as phosphorylation or oxidation can enhance the detection of PTMs. Other advantages could be the simple visual discrimination of false positives, and providing information about sequence coverage [132]. Despite some limiting factors , it is expected that spectral databases become more extended, with even software producers including spectral libraries on their programs.

Classic examples of mass spectral libraries are NIST-related **MSPepSearch** (featuring spectra from NIST repositories — validation through SEQUEST search, PeptideProphet, integrated in an SBEAMS module and finally correlated to their respective genome sequences via BLAST [65]), **X!Hunter** (featuring *Homo sapiens*, *Mus musculus*, *Saccharomyces cerevisiae*, it is built from annotated spectra of the Global Proteome Machine Database, GPMDB; peptide sequence identification is based on scoring the similarity of the experimental spectrum with the contents of the library) [54], **Bibliospec** (modular approach for development of spectral libraries out of high quality spectra: BlibBuild + BlibFilter + BlibSearch — dot product metric is used to measure the similarity between reference spectrum and unknown spectrum = dot product similarity score for higher efficience than usual cross correlation methods. Featured examples: *Caenorhabditis elegans* and *Escherichia coli*) [86], **SpectraST** (*Saccharomyces cerevisiae*, integrated on Trans Proteomic Pipeline, also designed for Targeted Proteomics, originally over 30000 spectra [158] or NIST **MSPepSearch** (http://peptide.nist.gov). Very relevant for plant physiologists is **ProMEX**, a mass spectral library developed by Hummel et al. (2007) and extended by Wienkoop et al. (2012) which features different plant species such as *Arabidopsis thaliana*, *Chlamydomonas reinhardtii* or *Medicago truncatula*, and fully integrated in the SwissProt database [132][336].

Mass Spectrometry Measurements Depending On The Fractionation Strategy

On a technical level, there are three different possibilities for protein identification on Proteomics-related mass spectrometry analysis:

Bottom-up Proteomics "Bottom up" strategies to proteome analysis involve cleaving the protein into peptide fragments that are smaller but still sufficiently distinctive to allow protein identification. There are two generally applicable "bottom up" approaches for protein identification: a first one based upon MS fragmentation (MS/MS) of one or more of these peptides mass measurements for a set of peptide digestion products from the parent protein, and a second approach usually referred to as *peptide mass fingerprinting* that consists of the creation of a set of peptide fragments unique to each protein. The m/z value of each one of these fragments is then used as a "fingerprint" to identify the original protein. Bottom-up Proteomics is presently unsurpassed in its ability to identify large numbers of proteins — however, it suffers from the loss of information required to distinguish proteoforms [281]), many of them functionally relevant, and therefore essential to conform a functionally complete proteome that describes the studied biological system accurately [8].

Top-down Proteomics As already mentioned, bottom-up proteomics strategies show some limitations at the time of identifying site-specific mutations and post-translational modifications of invidual proteins which could be relevant in biological function. This is partially due to the fact that bottom-up proteomics rarely reaches complete sequence coverage of proteins in proteome analysis. A way to overcome these limitations and gain otherwise unobtainable insights at the peptide level is the MS-analysis of intact proteins, commonly known as "top-down Proteomics". Here, mostly single proteins or simple protein mixes are separated and fragmented directly in the mass spectrometer to achieve both protein identification and characterization, without the need for prior chemical or enzymatic proteolysis. Nevertheless, "top down" strategies are generally less effective for protein identification than peptide level measurements [33][8].

Middle-down Proteomics Despite this technique was already conceived by Fenselau et al. around 1994 [318], the term "Middle-down proteomics" became coined by García and Kelleher in 2007 to describe a middle term between both previous approaches used basically for histone analysis: analysis of digested peptidic fragments, but with a much higher molecular weight than in the bottom up approach (3 to 9kDa), so that some disadvantages of bottom-up proteomics can be bypassed [88]. This disadvantages are e.g. sensitivity loss due to low ionization efficiency of a given peptide, poor LC retention of short (<6 aa) and hydrophilic peptides, digestion efficiency of peptides carrying PTMs [274], or discrimination between protein isoforms with presence of PTMs [221]. Fragments of this size can be achieved through protein digestion with GluC instead of Trypsin, relaying on the high concentration of Glc in the Histone ends, and blocking any sites for tryptic digestion through comprehensive propionylation of the free lysines. Alternative digestion enzymes for middle down proteomics are AspN [213] and outer membrane protease T (OmpT) [349]. With the recent development of WCX-HILIC and online ECD applications in the middle-down range [145], a major breakthrough on the analytical side is in sight. The current bottleneck is the development of robust bioinformatics pipelines that can automatically interpret the collected data [213].

1.1.2 Shotgun Proteomics / qualitative analysis

Introduction: Shortcomings of 2D-PAGE for Proteomic Studies

Proteomics was initially established as a combination of 2D-PAGE with MS-spectrometry for direct and rapid analysis of the entire protein complement of complex biological systems, aiming at separation, identification and comprehensive coverage of all proteins present in complex biological samples from whole organelles, cells and tissues [150] [334], [208]. Nevertheless, even though 2D-Gels were an essential tool for proteome analysis, many disadvantages of 2-D became evident: lack of reproducibility, failure to resolve most proteins greater than approximately 100 kDa, failure to routinely detect more than 1000 spots that could be identified by mass spectrometry, and the inability to separate most membrane proteins — extremely acidic or basic proteins also required special gels to get resolved [188]. Additionally, its limited dynamic range could limit analyses to less than half of the detectable proteins, as reported by Gygi on a study on yeast [111] [100].

The first alternative to 2D-PAGE-based Proteomics was presented by Washburn et al. (2001) as a complementary technique based upon a method previously developed by Link et al. [176]: the **MudPIT** approach (/textbfmultidimensional protein identification technology) consisted on loading different samples with digested protein extracts from cell lysates on a biphasic microcapillary column packed with SCX and RP materials. MudPIT could not only overcome the previously mentioned issues of 2D-Gels, but could also mantain the separation of two components after they have been resolved in one step. For instance, Washburn et al. detected and identified in their first trials with *Saccharomyces cerevisiae* a total of 1,484 proteins, representing 24% coverage of the predicted open reading frames of this organism [329].

Even though the concept "Shotgun protein analysis" had already been coined by Yates [354], MudPIT is regarded as the first successful application of the Shotgun proteomics approach.

Shotgun Proteomics

Definition A term adapted from molecular biology (*Shotgun Genomics*), **Shotgun proteomics** implies the identification of different proteins from a complex sample through de-

tection of one or more peptide fragments produced by enzymatic -usually tryptic- or chemical digestion (therefore, a bottom-up proteomics approach). This process fully automates the separation and identification of peptides after digestion of complex protein mixtures [329] [100] and allows also quantitative approaches [331] (see Section 1.1.3).

Basic experimental design A Shotgun Proteomics study usually consists of the following steps:

1. Sample preparation, including proteolytic digestion of the proteins
2. Online separation of the generated peptides by LC
3. Mass-spectrometric analysis
4. Bioinformatic interpretation of the data series

Even if the general output of proteins in a proteomics study depends on the sampling (studied conditions, replicates, additional purification or enrichment steps), an expected yield calculated on a *Saccharomyces cerevisiae* study estimated an identification of 50 to 106 molecules per cell [95] [175]. Furthermore, a standard shotgun proteomics study features a dynamic range of 10^4-10^5 molecules per cell [60] [344].

A nice overview of (Shotgun) Proteomics Tools can be found in Mallick and Kuster (2010) [187]. Interesting overviews on different developments on Shotgun Proteomics have been written by Gilmore and Washburn(2010) [99] mainly on different chromatography strategies (e.g. different gradients and column combinations) and sample prefractionation techniques.

Functions and Aspects

Implicit in this methodology is the ability to monitor the system both qualitatively and quantitatively [348]. Considering its features, Shotgun Proteomics can be successfully applied on the following tasks:

1. Identify the entire, dynamic protein complement of an organism, avoiding the modest separation efficiency and poor mass spectra sensitivity associated with intact protein analysis
2. Detect covalent modifications on known proteins
3. Allow for quantitative comparisons between samples [348]
4. Validation of genes that are expressed on the protein level and discovery of novel gene products and variants — for instance, high quality MS/MS spectra that are left unassigned when searched against protein sequence databases can be reanalyzed more comprehensively by searching genomic databases
5. Elucidation of global protein expression patterns that would otherwise be missed in an analysis of a single experiment)[221]
6. Optimally suited for discovering protein functions [198]
7. Peptide identification data can be used to improve the quality of the protein sequence databases by making them more complete and accurate [195], to create proteome maps [1] and even for enhancing Genome Annotation and Metabolic draft networks [195].

Drawbacks and Challenges

Despite the many advantages of Shotgun Proteomics over 2D-PAGE, there are still many challenges to face. These issues are often related to the complexity of the samples, the properties of the detected peptides, and also methodology aspects like the peptide-to-protein inference or optimization of the peptide databases. Regarding sample complexity, the limiting factors at the time of dealing with the enormous dynamic range of a shotgun proteomics measurement can be both associated with the limit of detection in the mass spectrometer and/or the presence of high- and low- abundant proteins as a bias towards high-abundant proteins. As shown already in Liu et al. [179] — among others — the lower the abundance of a protein is, the less frequently it gets identified. Replicate analyses of a complex sample can increase the total number of proteins identified by increasing detection of lower abundance proteins, and different peptide enrichment techniques can be applied as well. Another strategy to avoid bias towards high concentrated proteins is setting dynamic exclusion parameters during DDA measurements. Low abundant peptides can also be the result of protein digestion efficiency/constraint. A further reason for low peptide detection beyond low peptide abundance can be low peptide ionization, since poorly ionizing peptides are less likely to be selected for MS/MS sequencing. Poor peptide ionization can be due to the peptide's physical-chemical properties, but also to ion suppression — for instance, creating methods to reduce ion suppression and obtain a more uniform ionization would improve both qualitative and quantitative analysis [355]. Peptide-to-protein inference relies on the identification of the N- and C-terminal peptides — therefore, since the whole protein is assigned to single peptide sequences, both the different sequence coverage and the size of the correspondent protein do play a role in correct protein identification and quantitation. Likewise, it is posible to obtain false IDs from sequence redundacy, e.g. with distinct proteins having a high degree of sequence homology (as is the case in protein families, alternative splice forms of the same gene, or differently processed proteins) [220]. Unexpected post-translational modifications (PTMs) which are not included in the correspondent peptide database can also impede correct peptide identification. Further flaws in the used databases, such as artificial redundancies coming from truncated sequences, sequence alternatives arising from sequencing errors, or redundant peptide sequences under different gene names, should also be considerer as possible error sources.

Many of these drawbacks can be overcome through different chromatography techniques, the continuous improvement of Mass Spectrometers (including higher speed, sensitivity or development of new ionization strategies), more stringent search machines, and steady updating of the present databases. An alternative approach that has been gaining importance throughout the last years is database-independent protein identification.

Database-independent Shotgun Proteomics

Methods for Data Dependent Acquisition (DDA) Despite the capability of spectral libraries of identifying post-translational modifications (PTMs) that might not be reflected in protein sequence databases, many unknown peptide modifications (not only PTMs, but also e.g. alternative splicing) might still get unnoticed. Considering phosphorylation as a key regulatory mechanism, the identification of phosphopeptides that could serve as potential biomarkers is more than relevant. One possible approach to search for unknown significant peptides using high mass accuracy MS-measurements and spectral counting is the database-independent approach **MAPA** (**Mass Accuracy Precursor Alignment**; **Hoehenwarter 2011**) **[126]**. This strat-

egy is constructed around the **ProtMax** algorithm (Hoehenwarter 2008), which extracts all peptide precursor ions with the same mass to charge ratio (m/z) from high mass accuracy mass spectrometric raw data and groups them in a data matrix without any preliminar database search [127]. This matrix allows statistical treatment of the binned peptides and further analysis e.g. through existent databases or De Novo sequencing (see next paragraph). Initially, only the frequency of observed m/z-fragments according to the concentration of peptides was counted (*spectral count*) or ion intensity and added to the data matrix — in a new revision, retention time became also considered for binning the correct m/z-precursors, introducing local Rt-windows in order to discriminate peptides that share the same error-tolerated m/z values but that do not elute within an expected time window, binning them separately, and adding an absolute intensity-based noise filter [75]. ProtMax can be applied on both DDA- and DIA measurements.

Methods for Data Independent Acquisition (DIA) Other strategies have focused on DIA measurements, frequently based upon scanning cycles of limited m/z isolation windows that should be repeated over the whole MS measurement. Two examples of these approaches are **PAcIFIC** [233]) scanning CID fragments (no precursors) in 15m/z windows on the whole range, or **SWATH** [98], which records a complete, high accuracy fragment ion spectrum of all precursors selected in one isolation window. The same precursor isolation window is fragmented over and over at each cycle during the entire chromatographic separation, thus providing a time-resolved recording of the fragment ions of all the peptide precursors that elute on the chromatography.

***De Novo* Sequencing** Strategy for the identification of new peptide sequences without matching them to a peptide database — the resulting *De novo* sequenced peptides can be further investigated e.g. by homology studies with other sequenced organisms. *De novo* sequencing can be very useful in case of PTMs or neutral loss processes that haven't been included in any protein/peptide databases, alternative splicing or even detecting bioactive peptides that are heavily modified and rarely detected in high throughput experiments involving tryptic digestion [202]. Through predictable fragmentation patterns it is possible to find fragments that can be matched to approximate masses of amino acid residues measured on the MS (a good overview of the influence of different mass spectrometry approaches and fragmentation mechanisms on De Novo sequencing can be found on Seidler 2010 [271]). These fragments can be further aligned following logical ion series (peptide ladder) [2]. Besides, other rules have to be considered e.g. there are 16 rules regarding loss of ammonia and water, spectral intensity, aminoacid composition, isobaric mass (as with Leu/Ileu or Lys/Glu), among others. The main challenge at this process is that the peptide segments are not measured sequencially, so that a huge computing power has to be invested to calculate and match all these measured fragments simultaneously on a complex sample. The most extended program for De Novo sequencing is IonSource's **PEAKS** — many other algorithms and different strategies for De Novo sequencing are commented in Allmer 2014 [2] .

1.1.3 Quantitative Proteomics

Definition

Quantitative proteomics is an essential procedure for comparative studies in regulatory mechanisms or different responses on stress conditions. Its main advantage over e.g. transcript analysis is that, despite not having any amplification step such as PCR, it offers an accurate reflection of post-transcriptional and post-translational modification of the target proteins. Furthermore, the use of LC-MS allows a more reproducible, precise and accurate measurement than more "traditional" methods such as densitometric quantification or immunoassays.

Accuracy versus Comprehensiveness

For quantitative protein analysis there are other key factors to be prioritized than with qualitative analysis: basically, a reliable chromatography is essential in order to maximize peak capacity for MS/MS, and fast scanning instruments that can track the presence of an ion throughout a gradient. The priorities for the MS- measurements are contradictory: on one hand, there should be captured as many time points as possible in the chromatogramm for each ion, but on the other hand also move on swiftly to other m/z values, thus including as many different ions as possible for each ion in a duty cycle without compromising the precision of the quantification. This conflict between brevity and persistence can lead to trade-offs in data quality, since the limit of detection for peptide identification most often exceeds the limit of quantitation [355]. A solution to this problem is to optimize measurements for identification and quantify well enough to observe the trends in changes which can then be measured more accurately and precisely with more focused mass spectrometry methods as done by Dong et al (2007) [69]. In fact, the level of practical experience with quantitative proteins is of major significance as the choice of the quantitation method to reach a high technical reproducibility [306]. Besides, the experimental design should also consider the biological variation of the quantified samples, hence including a corresponding number of replicates [268].

Choice of Mass Spectrometers

In general, quantitative proteomic analyses are performed as a straight-ahead comparison between two organisms or states using already selected peptides as references (e.g. a metabolic pathway). The instruments of choice are usually triple quadrupole spectrometers due to their superior speed and capability of measuring in a continuous flow — nevertheless, the latest evolutions have brought new possibilities that blur these preconceptions, e.g. adjusting Orbitrap spectrometers to achieve competitive results through enhanced resolution, lower background noise or higher efficiency through better ion transmission and elimination of neutrals, or triple TOF spectrometers gaining ion coverage through discarding data-dependent acquisition (DDA).

Normalisation

A key step in proteome analysis, especially when comparing different datasets, is to normalise the different identified peptides — a factor is applied to minimise the different variation sources that might exist between different experiments or even different technical replicates. A simple approach would be referring the measured data to housekeeping proteins, as e.g. on classical techniques such as Western Blots [42]; nevertheless, this strategy is rather limited considering that many proteomics approaches such as targeted proteomics or PTM analysis might exclude

housekeeping proteins, and also that many variation sources relate to the properties of individual proteins even in the same measurement (such as peptide ionisation or chain length). Therefore, different indices have been derived: **NSAF** (large proteins tend to contribute more peptide/spectra than small ones, a **normalized spectral abundance factor** (NSAF) — this factor compensates for that), **RSc** (actually an alternative to Spectral Counting (see *Label-free vs. labeled Analysis*) — log 2 on the ratio of spectral count — better reduction of variation than NSAF), **SIn** (normalized spectral Index, [104] — more powerful reducing variability between samples than NSAF and RSC — fragment ion intensities calculated from peak heights rather than integrating precursor ions [218]- **PN** (**Unique peptide number**, [104]), **MPI** (**Mean Protein Intensity**, [104]), **Q-FISH** [166]; scale standardisation through division of intensity values by its maximum value, noise reduction through creation of mass windows, quantification through clustering of duplicated peptides with similar patterns — reference spectra: average built based on patterns of precursor ions and identification of differentially expressed peptides through beta-binomial tests), **PAI** (protein abundance index: number of identified peptides divided by the number of theoretically observable tryptic peptides for each protein — usually used to estimate protein abundance for absolute quantification), **EMPAI** (exponentially modified PAI), or **APEX** (absolute quantification through a score built from the proportion between the expected number of unique peptides of a protein and the observed number of peptides expected for a protein — with every protein getting a correspondent critical correction factor [360]).

Label-free vs. Labeled Analysis

Label-free Analysis Many of the strategies for protein quantification do not necessarily rely on absolute values, but on ratios between two different states, or on the relative amounts measured in a sample. One of the first quantitative label-free shotgun proteomics studies was developed in combination with a metabolomics protocol (Weckwerth et al. 2004) [330]. The label-free shotgun proteomics method is way more economic since no expensive labeling or internal standard procedures are needed, and it accomodates to most workflows [27]. Besides, the proteome coverage of quantified proteins is high because basically every protein that is identified by one or more peptide spectra can be quantified. Therefore, label-free methods usually have high analytical depth and dynamic range, which is very favourable when large, global protein changes between treatments are expected [268]. Nevertheless, label free analyses should be thouroughly planned and executed to minimize the inherent error sources in quantitative measurements, such as ionization competition between complex samples or non-homogeneous sample introduction. Two standard strategies for label-free comparisons are **spectral ion intensity** and **spectral count**.

Spectral ion intensity Since signal intensity from electrospray ionization (ESI) recorded chromatographs correlates with ion concentration [323], relative quantification can be achieved by extracting and calculating peak areas for a certain mass value, and comparing the elution tags and resulting values from separate LC-MS runs for that same ion. A reproducible chromatographic separation is key, since the confidence of the measurement relies heavily on both reproducible RTs and accurate peptide extraction for a correct peak alignment, and integration of the measured peptides [27]. For more accurate matching and quantification it is also recommendable to average and normalize the chromatographic peak intensities, thus eliminating background noise and reducing experimental variations coming from differences in sample

preparation and LC-MS/MS injection [360]. It is also important to employ a high mass accuracy mass spectrometer so that the influence of interfering signals of similar but distinct ma/z can be minimized [15]. A major drawback is that parent ion survey scans are interrupted by the fragment ion scan events (MS/MS), resulting in discontinuous coverage of the peptide ion peaks: the more interruptions, the more coverage, but also the less accuracy for the detected ions. Therefore, a balance between acquisition of survey and fragment spectra has to be found [268],[15].

Spectral intensity (*spectral count*) Here, relative protein quantification is achieved by comparing the number of identified MS/MS spectra from the same protein in each of the multiple LC- MS/MS or LC/LC-MS/MS datasets. It was demonstrated that among all the factors of identification, only spectral count showed strong linear correlation with relative protein abundance [360], all this having a dynamic range over 2 orders of magnitude [179]. Even though low abundance proteins might get lost, quantification through spectral intensity guarantees a similar linearity of quantification, less variability between peptides of the same protein, it covers a higher linear dynamic range in complex matrices and detects a higher number of identified proteins. Furthermore, it doesn't need an additional extraction algorithm, thus making further processing much more comfortable [335].

Labeled Analysis This kind of measurement has been frequently associated with the **Single Reaction Monitoring Technique** (**SRM**). The SRM technique is based upon the selection of certain ions through a mass window and monitoring them through a complete gradient - these selected ions become subsequently also fragmented, and the resulting ions also get selected according to their m/z and measured. Because of the peptide and ion selection, lower selectivity is required from the machine, shifting the priorities rather to speed, sensitivity and dynamic range - in fact, the SRM can increase the sensitivity / dynamic range of an average full scan measurement by one to two orders of magnitude [160] [47] [347]. For this approach, the most frequently used mass spectrometers are triple quadrupol models, sometimes also LTQ-TOF-TOFs (Q1: precursor detection / mass filter — Q2/TOF1: fragmentation cell — Q3/TOF2: ion detector). The multiplexed application of the SRM is also known as **MRM** (**Multiple Reaction Monitoring**). In the recent years, it has been attempted to adapt this approach to absolute quantification to the slower Orbitrap-based devices, even though the presence of an ion trap is usually at odds with quantification because of its lower speed. To compensate for this minor speed though, the more advantageous aspects of Orbitraps have been brought to the front: lower background noise, better resolution and consequently also higher selectivity. Further strategies have been incorporated in different methods, such as use of mass or time windows throughout the LC-gradient (as in the TPM process [124]), or improvement in the fragmentation and stabilization of ions (e.g. given via HCD-Fragmentation [226], leading to the development of the **Parallel Reaction Monitoring** (**PRM**) approach. In this DIA-based procedure both parent and fragment ions are measured simultaneously with high accuracy in the Orbitrap, instead of serially acquiring selected transitions in a Triple Quadrupole like in MRMs [240][87].

Chemical labeling Chemical labeling is based upon marked tags of a known m/z binding with certain functional groups of the peptide chains. This is usually achieved through isotope-coded affinity tags, with the binding taking place either at protein level (**ICAT** [112]) or

peptide level (isobaric tags, commercially known as **iTRAQ** [258] or **TMT** [298]). Even though both approaches are based upon the same basic principle of a 3-part-tag, there are some clear differences: on ICAT, the first component acts as a reagent, binding specifically to Cys rests, a second component acts as a linker where the isotopes are introduced (up to 8 Deuterium rests) and the third component is a Biotin rest that allows Protein enrichment via Avidin binding. Besides, the actual quantification of the protein is achieved on the precursor ions, before the Biotin fragment gets dissociated by CID. On the other hand, isobaric tags bind to peptides instead of proteins, thus having a wider peptide coverage (the binding specifity of their peptide reactant fraction is on primary amine groups) and quantitation takes place in the MS/MS stage, coming with higher time cost but also higher peptide detection [258]. Isobaric tags, unlike ICAT, always feature the same mass: this is achieved through concatenation of different reporter ions with balancer regions to keep the total tag mass constant (the carbonyl-based mass balancers will later split through neutral loss in MS/MS) . Finally, instead of a Biotin group these tags feature a cyclic N-methylpiperazine-based group in the reporter peptide fraction to simplify the interpretation of MS/MS spectra. Quantitation relies on the intensity ratios of so-called reporter ions in the fragment spectra.

A cheaper alternative for chemical labeling can be **Dimethyl labeling**, which consists on substituting the H-atoms from Amin groups for isotope-labeled methyl groups, causing shifts both in mass and retention time [129]

Metabolic labeling One way to reduce the errors coming from sample loss during protein extraction [27], purification or digestion is to add the internal standard before the sampling takes place — for instance, feeding the live organisms with isotopically labeled nutrients, like essential aminoacids on cell cultures (as in the **SILAC** approach by Ong 2002 [227], usually arginine and lysine) and ^{15}N labeled salts (like HILEP, SILIP), or ^{13}C labeled glucose (*SMIRP*) on high plants [27]. These techniques usually feature a high labeling efficiency [5] — nevertheless, they also depend on great part on the labeling efficiency and are relatively limited for multiplexing.

Labeled standards The most extended strategy for absolute protein quantification is the use of internal standards, emulating on Proteomics the classical mass spectrometry approach of stable isotope dilution. These internal standards aim at concrete peptide sequences (determined either experimentally or after an *in silico* digestion of the analysed proteins) with selected aminoacids being marked by stable isotopes, thus producing a mass shift.

The first successful approach to stable isotope labeled standard dilution techniques on peptides was first attempted as an alternative to immunoassays by Barr 1996 [18] and aiming to quantify phosphorylated Separase on HeLa cells by Stemmann 2001 [284]. Nevertheless, it was with the **AQUA** ("**Absolute Quantification**") approach [93]) that the technique was firmly established on protein analysis: here, spots from 2D-Gels were cut, digested in the presence of a known amount of standard peptide and finally analysed through SRMs: quantification resulted from the ratio of the AUC of the measured peptide to the AUC of its respective standard. Despite allowing accurate peptide measurements and leaving also space for further developments, such as as coupling with purification techniques like *SISCAPA* (*Stable Isotope Standards and Capture by Anti-Peptide Antibodies* [6]), the AQUA approach featured some drawbacks, like differential ionisation properties of the different peptides or keeping equimolarity among the different standards at multiplexing.

An alternative method was also developed for whole proteins, usually expressed on cell free

systems or bacteria [41]: through labeling of complete proteins by **PSAQ** ("**Protein Standard Absolute Quantification**" [73]) different peptides of the same proteins, isoforms, more reliable digestion and larger sequence coverage should be possible. This strategy not only aimed at minimizing issues such as ionization competition or affinity issues in the chromatography through a wider choice of reporter ions, but also left room for other digestion enzymes than Trypsin. Furthermore, it also pointed at investigating isoforms and quantify proteins on top-down proteomic approaches. Unfortunately, unlike AQUA peptides, the PSAQ standards are not feasible for PTM analysis.

A less expensive method that also allows multiplexed quantification (up to 50 tryptic peptides) and guarantees equimolarity between all the different standard peptides is the **QconCAT** approach [23]. Here, chimeric, isotope-labeled proteins consisting of concatenated standard peptide sequences are obtained from heterologous expression of synthetic genes in vivo — after tryptic digestion of a concatamere, a full stoichiometric set of labeled standard peptides is obtained. Further developments on QconCAT incorporating some other isotope labeling strategies can be found in Simpson and Beynon (2012) [275]. QconCAT has the great advantage of multiplexing, thus enabling the measurement of different peptides from the same protein, and ensuring standard equimolarity in case of quantifying protein complexes or subunits. Besides, the production of standard peptides doesn't need to be outsourced: once the vectors with the synthetic genes are produced, standard peptides can be produced at wish. Nevertheless, the calibration of the QconCAT standard and its influence on the accuracy of the final quantification could become a challenge [41]. Another limitation is the lack of PTMs, since QconCAT standards are synthesized mostly by *E.coli*.

An interesting approach which brings together many of the advantages of the previously mentioned techniques is included in this PhD-Thesis: the **Mass Western** (Lehmann 2008). Initially developed in 2008 as an application of SRM with added synthetic stable isotope labeled standard peptides on complex biological samples with no prefractionation whatsoever [167], the latest development of the Mass Western has been enhanced through the use of synthetic peptides that include cross-concatenated peptide sequences from different protein subunits (thus keeping standard equimolarity and regarding digestion efficiency on the final results) and also an additional, differentially marked equalizer peptide to compensate other sources of variation such as peptid e loss during sample desalting **[253]** — to sum up, the Mass Western is a method that succeeds bringing together both the high accuracy and flexibility of the classic AQUA approach, and the robustness and flexibility of PSAQ and QconCAT.

1.1.4 Different Subareas in Proteomics

The field of Proteomics can be divided in different subareas — depending on many aspects of the correspondent study. These are some of the most relevant criteria:

Depending on the Initial Hypothesis

Depending on the initial hypothesis, it can be differenciated between an **unbiased** approach (with a comprehensive study of the entire protein complement of an organism as a goal) and a **targeted** approach (if the study aims at detection and quantification of specific proteins, gene products or groups of proteins from a specific biological phenomenon according to their physicochemical or biological properties [141].

Depending on the Number of Different Biological Samples Involved

Depending on the number of different biological samples involved in the study, we can speak of **Descriptive Proteomics** (aiming at identifying the protein complement found in a particular sample) or **Comparative Proteomics** (Analysis of proteome changes in response to development, disease, or environment changes in different biological samples) [210].

Depending on the Object of the Study

Subcellular proteomics are defined as proteomic studies on subcellular structures, either aiming at the protein complement of the whole organel, or protein assignment to specific subcellular locations). They usually involve the use of subcellular fractionation techniques — a relatively recent review of subcellular proteomics methods has been written by Drissi 2013 [72]. In plant physiology there have also been established Organism-specific subareas, such as **Leaf/seed/stem proteomics** [207].

Depending on How Proteins are Related

Depending on how proteins relate or interact with each other, it can be differenciated between **Expression Proteomics** (comparative proteomics aiming to confirm functional relation between proteins with similar expression) and **Interactomics** (analysis of protein interactions from scales of binary interactions to proteome- or network-wide). Most proteins function via protein-protein interactions, and one goal of interaction proteomics is to identify binary protein interactions, protein complexes, and interactomes). Interactomics also incorporates methods from wide-genomics approaches (such as yeast-two-hybrid), same as **Structural Proteomics**, another subarea that aims at generating protein 3D structures after in silico , crystallographic, or spectroscopic analysis [141], and that also includes the investigation of the contact sites at protein interaction.

Depending on the Investigated Macromolecules

Proteomics does not only include the study of proteins as they are featured on a standard protein database, but some other variations are included: **Posttranslational Proteomics** is the large scale screening and mapping of biological protein samples looking for covalent modifications [343] — these modified peptides (e.g. phosphopeptides) being usually putative markers for specific physiological processes. Another approach dealing with peptides of "a different kin" is **Peptidomics**, aiming at the identification of functionally active single peptide fragments coming from limited proteolysis of precursor proteins, gaining information about sites for signal peptide removal, propeptide removal, peptide hormone processing, and domain shedding [305].

1.2 *Chlamydomonas reinhardtii*

1.2.1 History and Taxonomy

Chlamydomonas reinhardtii is the most intensively studied and well-developed model for investigation of a wide-range of microalgal processes ranging from basic development through understanding triacylglycerol production. This alga is easy to cultivate, grows quickly, and is

tolerant to varying growth conditions. Additionally, its complete genome has been sequenced [204], so that the algae can be easily engineered at the genetic level. This makes *Chlamydomonas reinhardtii* an attractive model system for investigation of a wide range of underlying biology processes, including photosynthesis, cell motility, phototaxis, cell-wall biogenesis and other fundamental cellular processes [140].

The Chlamydomonas genus was first described 1833 by botanist Ehrenberg in his "Dritter Beitrag zur Erkenntnis grosser Organisation in der Richtung des kleinen Raumes". Currently, there are 575 taxonomically accepted entries among 1163 species and infraspecific names in the algal database (http://www.algaebase.org). This genus is considered as a representative of the critically important, early-diverging lineage leading to plants, and as a microbe retaining important features of the last eukaryotic common ancestor (LECA) that have been lost in the highly studied yeast lineages [58]. Relatively rare to find in nature, the Chlamydomonas species has been localized in many different habitats all around the world (such as soil, fresh water, oceans, and even in snow on mountaintops) but only in a handful of places - all in all, very well documented, though (see Table 1.3).

Table 1.3: *Chlamydomonas reinhardtii*: distribution around the world (source: http://www.algaebase.org)

Continent	Country
Europe	Germany [291], Great Britain [333], Netherlands [314], Romania [44], Russia [224], Spain [43]
North America	USA [116]
South America	Argentina [293], Brazil [201] [285]
South West Asia	Iraq [194], India [110]
Asia	China [130], Tajikistan [16], Taiwan (http://http://taibnet.sinica.edu.tw/),
Oceania	New Zealand [40]
Antarctica	Lake Bonney [243]

The *Chlamydomonas reinhardtii* species was originally described by Dangeard (1888), and suggested by Pröschold et al (2001) as the conserved species of this genus. The most widely used species used in laboratories has been long assumed to come out of clones originated from one single zygospore isolated in a potato field in Amherst, Massachussetts in 1945 and designated 137c by G.M. Smith. Nevertheless, this topic has been recently revised by Pröschold 2005 and Kubo 2002 [245] [153], concluding that two further different / nonrelated lines are actually involved in the development of all the actual strangs of *Chlamydomonas reinhardtii*[116]. From these 3 strangs, the third one was inable to use nitrate as a sole nitrogen source, and has also been consistently designated 137c in publications since then [116]. Recently, the strang cc503 has gained considerable relevance, since the genome sequencing of *Chlamydomonas reinhardtii* was performed on it.

All of the different *Chlamydomonas reinhardtii* mutants are available at the Chlamydomonas Center [http://www.chlamy.org].

1.2.2 Morphology

Chlamydomonas are ∼10 μm, typically spherical to subspherical unicellular algae, featuring a cell wall, a central nucleus, usually two contractile vacuoles (depending on the species they might be either absent or more numerous, though), multiple mitochondria, Golgi vesicles, starch granula, and its most distinctive features: two anterior flagella for motility and mating, and a single chloroplast per cell with a large pyrenoid that houses the photosynthetic apparatus and also other critical metabolic pathways [204]. The two isokont, anterior flagella (with which they can swim using a breast-stroke type motion) are rooted in basal bodies, built from a typical eukaryotic (9+2) microtubule pattern, and contain intraflagellar transport (IFT) particle arrays between the axoneme and the flagellar membrane. The basal, cup-shaped chloroplast can be very variable depending on the cell, but many characteristics remain constant: a pyrenoid, and a light-perceiving eyespot (stigma). The pyrenoid is a proteinaceous structure almost absent from the chloroplasts of terrestrial plants but very common on eukaryotic, photosynthetic algae [34] that contains most of the RuBisCO from the chloroplast, separating it from the carbonic anhydrases in the stroma [214]. Pyrenoids also undergo a dramatic morphological change when cells are switched from high-to-low-CO_2 conditions [251]: when the CCM is functional, a starch sheath appears around it, and 90% of the total RuBisCO is accumulated. Further morphological changes can be caused by nitrogen depletion: lipidic bodies can be recognized and cells get deflagelated, developing into two different mating types that can fuse, forming a diploid zygospore with a hard outer wall that protects it from adverse environmental conditions.

1.2.3 Cell cycle

A haploid in vegetative state, under light and N-starvation Chlamydomonas cells develop into mt+ and mt- gametes, as previously mentioned. These two gametes can fuse, forming a diploid zygospore that will produce 4 haploid zoospores through meiosis once nitrogen has been repleted (in presence of enough light and water) or the environmental conditions have improved. Vegetative cells can also divide during light/dark cycles by multiple fission into 4-8 zoospores [116], usually after entering the 3rd hour of the dark cycle. The Chlamydomonas cell cycle has a striking temporal and functional separation between cell growth and rapid cell divisions [58], probably connected to the interplay between diurnal cycles that drive photosynthetic cell growth with the cell division cycle — if a circadian clock is also involved in the regulation of the cell cycle remains still controversial [58] [320]. *Chlamydomonas reinhardtii* also features a highly choreographed interaction between the cell cycle and its centriole/basal body/flagellar cycle.

1.2.4 Genome

Chlamydomonas reinhardtii features a 110Mb nuclear genome [204], which for meiosis get divided into 17 chromosomes (packaged with histones as standard eukarytic chromatin [58]). It also contains a 203 kbp chloroplast genome [193] and a 15.8 kbp mitochondrial genome [206]) both of them multicopy and packaged into nucleoids that are distributed throughout the stroma and the mitochondrial network, respectively. It is also remarkable that the chloroplastic genome replicates independently of the nuclear genome during the growth phase [58]. Through genomic comparisons to photosynthetic and ciliated organisms two different gene groups could be distinguished: **Green Cut** (comprising 349 Chlamydomonas proteins with homologs in

representatives of the green lineage of the Plantae (Chlamydomonas, Physcomitrella, *Ostreococcus tauri* and *O. lucimarinus*), but not in nonphotosynthetic organisms) and **Cilia Cut** (195 Chlamydomonas proteins with homologs in human and species of Phytophthora, but not in nonciliated organisms) [204].

Genome research in *Chlamydomonas reinhardtii* started already in the mid 20th century, with Ralph Lewin and Paul Levine isolating the first flagellar and photosynthesis mutants, respectively [169] [74]. Another breakthrough from these early years was the demonstration of Chloroplast inheritance, first through the isolation of drug-resistant mutants and later through tetrad analysis [260] [261] [311]. Further forerunners were Pete Lefebvre and Carolyn Silflow, who constructed a BAC library [146] and found useful polymorphic strains [105], the Kazusa Institute, which generated the first set of ESTs for Chlamydomonas (http://www.kazusa.or.jp/en/plant/chlamy/EST) [9], or Jin Billy Li, who developed the gene-finding **GreenGenie** algorithm for analyzing the first large-scale sequence of the nuclear genome, generating over 200,000 additional sequences assembled into over 10.000 "unique" cDNAs [171]. The decisive catalysator in the genome research was, though, Arthur Grossman, who launched the **Chlamydomonas Genome Project** with funding from the National Science Foundation and, in collaboration with several Chlamydomonas laboratories, produced a large EST library, the first cDNA microarray, the chloroplast genome sequence, a more complete set of nuclear markers, and set up a team for sequencing the nuclear genome with Dan Rokhsar at the Department of Energy's Joint Genome Institute (JGI). The first assembly of the nuclear genome was finished in 2002 by Grossman et al. [108], and enabled the first proteomic studies of single organelles such as flagella [235], basal body [147]), eyespot [264], mitochondria [312]), or thylakoids [3]. However, it took to 2007 until the complete *Chlamydomonas reinhardtii* genome sequence was released [204]). This definitive version marked the beginning of more ambitious, comprehensive studies featuring proteomics data integrated with e.g. genome annotation (through EST mapping or computational gene prediction combined with homology-based search methods [346] **[195]**) for validation of *in silico* gel models, evaluation of the accuracy of existing gene models, or, through combination of proteomics and metabolomics data, for completing metabolic draft networks, as in **May et al (2008)**. More recent studies have dealt with more specific proteomes (like its Phosphoproteome [324], the Nuclear Proteome [342] or the Chloroplast proteome [294]), mechanisms of metabolic/dynamic regulation [191], development of new analysis techniques, e.g. genome microarray [301], or even the comparison of many proteomes using databases from different genome assemblies and annotations **[310]**.

The results of these decades of investigation are collected in different resources: genome (Phytozome, http://www.phytozome.net) and protein databases in **JGI**, **Chlamy.org**, **ChlamyCyc**, **NCBI**, **KEGG**, **SwissProt** (**UniProt**), also Plasmid, cosmid, bacterial artificial chromosome (BAC),RNAseq transcriptome data in different conditions and stresses (**GEO**, **Gene Expression Omnibus**), spectral libraries (**ProMex**), and many other resources from genetic and transcript studies with algae, like a library of indexed mutant libraries for reverse genetic studies [172], the **AFAT** database integrating gene expression data with metabolic pathway information [182] or the AlgaePath database including gene information, biological pathways and NGS datasets [358]. A recommendable and relatively recent overview of the current *Chlamydomonas reinhardtii* resources can be found in Blaby 2014 [28].

1.2.5 Relevance as a model plant

As previously metioned, *Chlamydomonas reinhardtii* diverted from plants over a billion years ago, thus conserving important features inherited from the LECA that have been lost in the highly studied yeast lineages but shared in land plants and animal cells, such as e.g. Cyclin A, Cyclin D [58] or the structure, assembly and function of eukaryotic flagella (motility, microtubuli assembly, flagellar functions, or establishing links between ciliopathy and the composition and function of flagella [204]). This brought the current understanding of the ancestral eukaryotic cell considerably further, and also allowed researchers to learn more about regulation of gene expression in more complex plants and animals. Another side-effect of this early divergence from high plants is Chlamydomonas not developing some very effective high copy numbers of some genes as higher plants did in the paleopolyploidization [89]. Thus, *Chlamydomonas reinhardtii* has remained a haploid organism with very few gene duplicates, what makes it ideal for loss-of-function genetic studies — e.g. by suppression of specific gene activities with antisense or RNA interference (RNAi) constructs, or developing reporter genes to identify regulatory factors and sequences that are involved in regulating gene expression. Chlamydomonas is also the first organism in which all three genomes could be easily transformed [37] [148] [248] [311] and sequenced [193] [308] [32]) [259]. This ease of transformation in a photosynthetic organism has brought this algae the nicknames of "the green yeast" or "the photosynthetic yeast" [102] [255]. Many applications have been found for *Chlamydomonas reinhardtii* e.g. in induction and isolation of repair-deficient mutants for studying various DNA repair mechanisms, conserved both in Chlamydomonas and many other organisms.

Among all its features, the ability of Chlamydomonas to grow both photoautotrophically under light conditions and heteroautotrophically in dark conditions (being even capable of synthesizing a complete photosynthetic apparatus while mixotrophically growing on acetate [337]), makes this algae an ideal model system for studying chloroplast-based photosynthesis. Many aspects of Chloroplast-based photosynthesis have been studied in Chlamydomonas, like Chloroplast biogenesis [255], abundance and rate of synthesis of individual complexes, assembly of the photosynthetic apparatus, regulatory molecules and distribution of excitation energy (state transitions / non-photochemical quenching) — [299][94][107] . In recent years, Chlamydomonas has been increasingly used to study additional biological processes, including among others lipid biosynthesis [131][326] [211] [178], pigment biosynthesis and regulation [180] [322], enzyme stoichiometry **[253]**, basal body functions [283]), Flux analyses [36] **[338]**, carbon-concentrating mechanisms [11][325] [143] [250], growth during nutrient deprivation [101][46] [29] [122] [106] [265], responses to heat stress [121], photoreception [19], mating [307] [92], cell cycle [320] [58], circadian rhythms [39] [267], bioaccumulation processes [242], biotoxicity [164], or cellular quiescence [303][172]. Especially the studies in TAG accumulation or biomass and hydrogen gas production [96] [217][123][352] are becoming more relevant due to the potential role of *Chlamydomonas reinhardtii* as a bioproducer in the biofuel industry.

1.2.6 Industrial applications

It has already been mentioned that *Chlamydomonas reinhardtii* is relatively easy to cultivate, growing both in light conditions on simple medium made of inorganic salts (both liquid or solid medium) and using photosynthesis to provide energy for growth, or in total darkness if acetate is provided as an alternative carbon source (an alternative to added acetate would be algae cocultivation with other mutants that would act as a C-source [297]). Furthermore, Chlamy-

domonas can be also cryopreserved at a cell density lower than 2.5 x 10^6 cells ml^{-1} [241] and even its asexual and sexual reproduction can be controlled by synchronization through nitrate deprivation [116]. The doubling rate of green algae can be 5-6h under laboratory conditions [282] and 24h under mass culture ambient conditions ([21]). Many different strains, both motile and nonmotile, have been developed for specific research purposes and industrial applications, such as bioremediation [66] and wastewater treatment [350] [252], herbicide biosensors [159], production of recombinant proteins [163] or even medical use (CR-based vaccines proteins [4], [257] and immunotoxin cancer therapeutics [302]). Nevertheless, considering the easy cultivation and conservation, the capabilities of *Chlamydomonas reinhardtii* as a bioproducer and also the possibility of scaling-up cultures in bioreactors (which offer many possibilites to develop economically viable production systems [115] or even its application in Biorefineries), it seems that this algae will still gain most relevance as a biomass, H, EtOH and TAGs source for the Biofuel industry [115] [152] [115], [120][131].

1.3 Biofuel production and *Chlamydomonas reinhardtii*

1.3.1 The Energy Problem

Since the first steps after the oil crisis in the mid 1970s, the necessity to look for alternatives for fossil fuels has increased dramatically: not only is the actual consumption level unsustainable (at the actual levels of energy consumption 90% of the total energy is being generated from fossil fuels — therefore only 10% comes from renewable energy sources [186]), but it is also estimated that at the present rates of consumption the conventional oil reserves that can be commercially exploited might last about 40 years — according to the Institute of Mechanical Engineers http://www.ibtimes.co.uk/world-energy-day-2014-how-much-oil-left-how-long-will-it-last-1471200. Furthermore, other side-effects such as global environmental pollution (98% of carbon emissions result from fossil fuel combustion [63]), ecological degradation and biotic health problems [114], and also the raising awareness about the climate change have made the transition to renewable energy sources a major necessity for the world's future energy supply.

1.3.2 Biofuels: Definition and types

Definition of Biofuel and Advantages over Fossil Fuels

One of the alternatives to fossil fuels are Biofuels, which can be defined as renewable energy sources produced from biobased material. Biofuels do not only have the potential to provide energy services with zero or almost zero emissions of both air pollutants and greenhouse gases, but also are expected to reduce dependence on imported petroleum (and its associated political and economic vulnerability), also reduce greenhouse gas emissions and other pollutants, and revitalize the economy by increasing demand and prices for agricultural products [63]. The most common biofuels are biohydrogen, bioethanol, biodiesel, being the latter one the most relevant one due to its compatibility with actual motor engines, its energy density, and its low emission of SO_x and NO_x after combustion [249].

Classes of Biofuels

The different types of biofuels can be also be classified based on their original feedstocks:

First generation feedstocks include such food and oil crops, such as rape seed, palm, soy, sorghum or sunflower and other oil producing plants. This type of biofuels have raised a lot of controversy, since the need of arable land for they production competes with food and fibre. Besides, regionally constrained market structures, lack of well managed agricultural practices in emerging economies, high water and fertiliser requirements, and a need for conservation of biodiversity (International Energy Agency (IEA) 2007 https://www.iea.org/publications/freepublications/publication/essentials2.pdf [134]) bring a negative impact on impact on global food markets and on food security [38].

Second generation feedstocks can include non-edible sources like animal fats, waste oils [249] or whole plant matter (e.g. jatropha, mahua, jojoba oil salmon oil, sea mango)[212], mostly for biodiesel production. These sources do not suffer as much from the food versus fuel dilemma since only low quality crops and lower amounts of water and fertilizer are needed — neverthelesss, their yield is too low to guarantee a stable supply that fulfills future energy needs.

Third generation feedstocks include microorganisms, but are especially referred to the cultivation of microalgae as a source of different products, such as methane, biohydrogen, biodiesel and bioethanol, and even further products from processing of such residues as glycerol [359]. Microalgae are considered the most promising source for biodiesel since they offer many advantages to other plants, like comfortable cultivation, faster growth, more efficient biomass production, higher CO_2-fixing rate or higher oil content and biofuel extractability. However, there are many aspects about microalgae cultivation that should be discussed with more detail.

Microalgae as an Energy Source

The use of algae as an energy source started in the late 1950s when it was suggested that carbohydrate fractions of algal cells could be used for the production of methane gas via anaerobic digestion [197] [229] [328]. Nowadays, algae may serve as potential sources of many types of biofuels, these including: biogas produced in processes of anaerobic degradation of biomass, biodiesel produced from lipids accumulated algae, alcohol, hydrogen from photobiological transformations or algae biomass that may be used for direct combustion. The resulting sludge after biomass fermentation could be also additionally used as a fertilizer or a complement in growth medium for algal biomass cultivation [61]. Microalgal biofuel systems can theoretically achieve higher yields per hectare than traditional crops due to their short life cycles and high production efficiencies [263]. This together with the fact that they can offer improved water use efficiencies and being potential providers of a wide range of fuels (e.g. biodiesel, methane, ethanol or hydrogen) explains why the microalgal biofuel sector has seen a rapid increase in investment over the past years [152].

1.3.3 Biodiesel Production

The standard methodology to obtain biodiesel from algae consists of the following steps: Strain isolation, Algae cultivation, Harvesting, Drying, Lipid extraction and Biodiesel production.

Strain Isolation

Since the first steps of investigations on algae as an energy source, the continuous search for the most appropriate species narrowed the algae collection down to the 300 most promising strains, mainly Chlorophyceae (green algae) and Bacillariophyceae (diatoms) [24]. Criteria for this selection are varied: Very popular algae for biofuel production contained in these families (considering the oil content per dry weight) are *Botryococcus braunii*, *Crypthecodinium cohnii*, *Cylindrotheca sp.*, *Nitzschia sp.*, *Phaeodactylum tricornutum*, *Schizochytrium sp.*, *Nannochloropsis occulata*, and most of all, *Chlorella sp.* [249] [24]. Three different strategies are used for the screening and selection of strains. The first strategy is selection from local environments where they are to be grown on a large scale. Local species have competitive advantages under the local geographical, climatic and ecological conditions. The second strategy is acclimatization under an environment in which the microalgae do not normally grow well. The third strategy is the use of a genetic approach to understand and modify the regulation of metabolic pathways[149]. The difficulties in efficient biodiesel production from algae are not in extracting the oil, but in finding an algal strain with a high lipid content and fast growth rate that is not too difficult to harvest, and a cost-effective cultivation system (i.e. type of photobioreactor and/or possible optimization of two-stage cultivation strategies) that is best suited to that strain.

Algae Cultivation

For algae cultivation different strategies can be adopted. The classic method is the cultivation on open ponds, with sunlight as a light source and frequently using wastewater as a C-source. Despite being the most economic approach, this strategy has many drawbacks — some of them, such as the heterogeneous conditions inside the pond terms of medium composition, dissolved O_2 or temperature can be partially solved through more intensive mixing e.g. in race ponds (continuously operated, closed-loop recirculation systems built as individual ponds or as groups of ponds arranged in a series connection, typically constructed of concrete or compacted earth [50] [149]). Nevertheless, other drawbacks of open ponds, such as evaporative losses due to the large surface area or poor species control due to external contamination, are ineludible.

A very effective way to overcome these problems is the use of **Photobioreactors**. Photobioreactors are arrays of straight transparent, UV-resistent tubes, usually made of plastic or glass -thus allowing maximum exposure to sunlight for photosynthesis [50]- through which microalgal broth is circulated from a reservoir as in a circular culture. Photobioreactors can also feature an internal light source, allowing continuous lightning, customized light cycles, and selection of the most appropriate wavelength and light intensity. They not only avoid species contamination or evaporative losses through isolation from the atmosphere [168] [272], but allow much closer monitoring of other cultivation parameters (salinity, pH, temperature, mixing, gas exchange, flow regime...), more homogeneouos culture conditions, higher cell densities and more effective biomass recovery. This reflects in a up to 13-fold higher productivity compared to traditional race ponds [24]. Nevertheless, the high energy consumption of Photobioreactors compared to ordinary fermenters or open ponds makes the global energy balance lower [277] and thus too expensive for biodiesel production on a pilot scale [149]. Therefore, until the development of more sustainable energy sources, the most economically viable strategy is the development of two-stage hybrid systems e.g. combining incubation in PBRs and upscaling in open ponds. A Table comparing more in detail the characteristics of open ponds, fermenters

and photobioreactors can be found in Zhu et al. 2014 [359], and a comparison between different types of photobioreactors can be bound in Bahadar et al. 2013 [12].

Another important factor to be considered for algae cultivation is the C-source: depending on the presence of a carbon source and the use of light, it can be differenciated between **photoautotrophic growth** (microalgae grow in the same way as other photosynthetic plants, by utilizing carbon dioxide and sunlight), **photoheterotrophic growth** (microalgae use organic carbon compounds as a source of energy instead of carbon dioxide, and sunlight), **mixotrophic growth** (algae are capable of using both autotrophic (sun- light and carbon dioxide) and **heterotrophic growth** (organic compounds instead of both carbon dioxide and sunlight). In comparison to photoautotrophic growth, heterotrophic growth usually involves higher growth rate and lipid content, reportedly reaching yields 3.4 times higher than in photoautotrophic growth conditions [205]. A few autotrophic microalgae can be converted to heterotrophic by changing the cultivation conditions or using genetic engineering modification [149]. Generally, heterotrophic algae cultivation is reported to provide higher biomass and lipid productivity than photoautotrophic cultivation — to cope with the higher costs of these strategy, alternative C-sources to glucose or acetate have been searched, being crude glycerol a very popular resource [174]. Another strategy is the CO_2-capture / sequestration or utilization of CO_2 from power plant flue gases and other fossil fuel combustion systems [22]. One example of a biodiesel production system including an upstream processing section aimed at sequestering the CO_2 from flue gras produced at an industrial source (e.g. a power plant) can be found in [244]. This publication also offers a techno-economic analysis to determine the optimal design of a flue gas to biodiesel system through the cultivation of algae a production chain.

Harvesting

The choice of a harvesting technique is dependent on the characteristics of the cultivated microalgae, e.g. size, density, and the value of the target products [38].Generally, microalgae harvesting is a two-stage process, in which biomass has to get first separated from the bulk suspension (Bulk harvesting) followed by concentration of the resulting slurry (Thickening). Different techniques can be used: most common for bulk harvesting are **Centrifugation** (based on Stoke's Law and only suitable for large microalgae with a diameter over 70 μm [109] and preferably with cell wall), **Flotation** (trapping of algae through micro-air bubbles, therefore chemical-free — not economically very viable, though) and **Flocculation** (taking away the negative charges of the algae surface through addition of metallic salts as FeCl, thus facilitating aggregation and precipitation), whereas for thickening more intensive techniques favoring aggregation, such as filtration, ultrasonic aggregation and also centrifugation are used [38].

Drying

The harvested biomass slurry (typical 5-15% dry solid content) is perishable and must be processed rapidly after harvest. Protocols for wet extraction have been developed, but with lower yields — however, since drying is a costly procedure, a balance has to be found between efficiency and cost to maximize the energy output of the fuels [173]. The drying method of choice depends on the final product required, e.g. sun drying is the cheapest method but also has the highest material loss, spray drying is widely applied for high quality fats but still might damage some pigments, and freeze drying is just as expensive but reccomendable for oil extraction — both from biomass slurry and wet algal biomass). Drying temperature also plays

an important role, not only on the yield, but also on the lipid composition [38].

Lipid extraction

The development of the method for the extraction and purification of the lipids from dry biomass is critical for diesel production from microalgae [252]. Conventional extractions usually involve a previous dewatering process, usually conducted in an expeller or through pressing. The extraction itself can happen either mechanically (e.g. through mechanical pressing after steaming the algae under high pressure) or chemically (with addition of chemical solvents like hexane or toluene). Usually the lack of solvents brings lower efficiency and yields than chemical extraction — on the other hand, it helps preserving the qualities of the extracted oil since no further distillation or processing is needed to remove the added solvents. Other methods have been more recently developed to overcome these limitations: e.g. incorporating ultrasound and microwave techniques [286] or through supercritical-fluid extraction: the use of SC-CO_2 instead of organic solvents avoids both oil contamination and thermal degradation, thus obtaining high-quality oil [200] — nevertheless, despite possible scale-up strategies have already been discussed as in Taher et al 2014 [289], it is too time-consuming for large-scale production.

After oil extraction from algae, the remaining biomass fraction can be used as a high protein feed for livestock — thus giving further value to the process and reducing waste [63]. It is important to note that allegedly 51% of the environmental impact of the microalgae production is not allocated to the biodiesel, but to the algae cake and the glycerine, in line with the European Directive on Renewable Energy [51]. Therefore, it is key to optimize procedures that obtain energy from the organic matter of the algae cake, such as fermentation to biogas (also in presence of other reactants such as maize [61]) or even use that energy to enhance the algae production (e.g. producing electricity by anaerobic digestion of algal waste residuals following biodiesel processing, or injecting the CO_2 removed during biogas production to the cultivation broth in order to stimulate algal growth [339]).

Biodiesel production

Various microalgal biofuels have been covered, but biodiesel has received intense focus due to its diversified use and superior financial viability [149]. Biodiesel is produced by transesterifying the parent oil or fat to achieve a viscosity close to that of petrodiesel. Transesterification (also called alcoholysis) is the reaction of fat or oil with an alcohol (usually in excess to favor the right side of the reaction — it can be recovered and reused [63]) with the involvement of catalysator to form esters and glycerol (see equation on Fig 1.4. This reaction usually can happens in three steps (breaking all the three fatty acid chains from the glycerol - see Figure 1.4) although two step catalyzed production has also been reported [24]. The resulting crude glycerol can be either purified or, as previously mentioned, be used as C-feedstock for further biodiesel production [286].

There are many methods for transesterification, mostly depending on the fats to be processed and the catalysators used: acidic (cheapest and most suitable method for organic substrates, homogeneous acidic catalists such as H_2SO_4, HCl, BF_3 or H_3PO_4 offer — especially when added in excess- high conversion rates and high yields despite being a relatively slow process), alkali (homogeneous base catalists such as NaOH, KOH or CH_3Na are frequently used for industrial scale-up due to higher conversion rates, usually taking place at 60 ℃ as a three-step process with MetOH and EtOH as alkyl acceptors) or non-catalytic super critical methanol

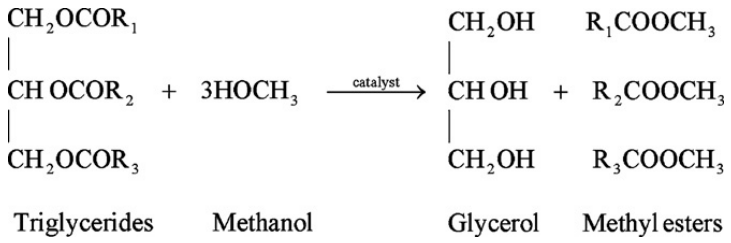

Figure 1.4: The chemical reaction of the transesterification process [286]

(which avoids such problems as moisture-related saponification or the formation of two-phase oil/MetOH mixtures, therefore offering easier biodiesel purification. Nevertheless, this method is too expensive for industrial application). More advanced strategies involve the use of lipases instead of chemical catalysators (thus avoiding the removal of alkaline waste, catalists and glycerol, and also not promoting side reactions), which can be either added or immobilized and also intracellular (having certain bacteria or fungal species immobilized in catalytic beds) or extracellular (more effective but also more expensive due to complicated downstream processes for enzyme extraction from MOs and immobilization on carriers to allow repeated use). For lipase-related transesterification there has also been experimented with alternative acyl-acceptors to EtOH and MetOH, since low chain alcohols in high concentration can inactivate lipases, e.g. Ethylacetate, Methylacetate or Dimethylcarbonate [24].

1.3.4 Growth optimization mechanisms for lipid production

Many algal metabolic pathways have been obtained through homology with higher plants or model organisms [82][203]. *Chlamydomonas reinhardtii* has always been considered rather a model organism than a serious candidate for biofuel production: thanks to its fully sequenced genome, many annotation projects have managed to reconstruct metabolic pathways thus providing important insights for studies on other oleaginous algae with more limited molecular genomic and genetic resources [328][203]. However, recent studies have shown that the lipid production of *Chlamydomonas reinhardtii* can be considerable increased through cultivation under certain conditions, accumulating TAG up to 46-65% their dry weight [203], thus reaching comparable yields to Chlorella and Scenedesmus (50% [339]. Some of the investigated conditions are e.g. pH, salinity, extreme temperatures, light, macronutrient supply or macro- and micronutrient deficiency such as sulfur, phosphorus, zinc, iron, and very especially nitrogen [35][328].

The most extensively investigated variations on the cultivation of *Chlamydomonas reinhardtii* for enhanced lipid storage are provision of a reduced C-source (usually acetate, a standard procedure for algae cultivation on heterotrophic or mixotrophic conditions [116]) and nutrient depletion (making Chlamydomonas cells reprogram their metabolism, thus retarding growth and accumulating starch and neutral lipids — particulary TAGs [328]. Starch and TAGs are the two main storage sinks for reduced carbon in this algae, starch being preferentially synthesized and mobilized, and oil representing rather a long-term storage in case of prolonged shortage or stress [273]. Both pathways are temporally and spatially separated (as shown e.g. in oilseeds by Focks & Benning [85]1998 and Smith 2010 [7]): both take place inside the chloroplast, but whereas amylopectin synthesis relies on soluble enzymes and both amylose synthesis and elongation take place mostly inside insoluble granules and plates surrounding the pyrenoid [116],

TAG biosynthesis begins with de novo fatty acid synthesis in the stroma [82] . The resulting TAGs are subsequently directed to the membrane or stored in lipidic droplets both inside and outside the chloroplast [103] — which acyl-transferases are responsible for cytoplasmic vs plastidic lipid storage is still unanswered. These lipidic droplets / bodies are not only considered a housing site for neutral lipids, but also dynamic compartments that participate actively in lipid metabolism via trafficking and signalling [203]. Further insights into characterization and biosynthesis of algal lipidic bodies can be found in Wang et al 2009 [326].

There are still many questions about the nature of the interaction of these pathways and the factors controlling carbon partitioning between these two storage products — actually, despite the starch biosynthetic pathway has been particularly well characterized at the molecular genetic level in Chlamydomonas [13][273], the TAG metabolism pathway on Chlamydomonas has been mostly adapted by homology e.g. to *Saccharomyces cerevisiae* [203]. However, it could be demonstrated that starch and TAG pathways have a close reciprocal relationship, with the starch pathway being the predominant one. This is evident e.g. in the fact that rapid oil synthesis only happens when the C-supply exceeds the capacity of starch synthesis [82] — in fact, TAG accumulation raises dramatically when the cultivation medium contains added acetate [24], whereas the increase in starch production is much lower. This leads to the consideration of carbon availability as a key metabolic factor controlling oil biosynthesis and carbon partitioning between starch and oil in *Chlamydomonas reinhardtii* [82].

Since N-free cultivation of *Chlamydomonas reinhardtii* reportedly increases their lipidic content, many studies have been led to clarify the mechanisms of C- and N-metabolism under N(-) stress. Proteomics analyses have been performed on N-deprived Chlamydomonas' chloroplasts [295], mitochondria [10] and lipidic bodies [211] [222] [136] — many of these studies have brought especially relevant results about TAG accumulation pathways [209] [35] [181] [203] [216] [28] [273]. For instance, after switching from nitrogen-replete to nitrogen-deprived conditions there was reportedly a notable decrease on ribosomal, photosynthetic and lipid pathway genes that have been shown to remain stable under N+ conditions — both on protein and transcript levels.On the other hand, the increased abundance of many proteins related to for glycolysis and oxydative phosphorylation points at a higher activity of the photosynthetic electron transport to ensure an adequate supply of metabolic cofactors for TAG accumulation, even though the abundance of many photosynthesis-related proteins such as photosystems or LHCs is reduced, thus slowing down cell growth and replication [328]. However, in spite of the upregulation of some proteins related to aminoacid metabolism and nitrogen assimilation, the enzymatic machinery associated with fatty acid synthesis to TAG assembly is already present in excess of, or readily inducible in response to metabolic needs of cells (since most transcripts encoding *de novo* fatty acid and membrane lipid synthetic enzymes were downregulated but the amount of many metabolites related to the synthesis of fatty acids and complex lipids increased) — therefore, oil accumulation can be considered independent from de novo protein synthesis, and mainly conditioned by C-availability. To increase that C-availability, the use of carbon in N-metabolism and cell growth is restrained, and hydrolytic release of carbon skeletons from proteins and other cellular compounds is stimulated as well [216]. The rapid incorporation of added acetate into TAG and membrane lipids, as measured in ^{1}4C-Acetate, N-free medium fed algae by Fan et al. [83], also supports C-availability as a key metabolic factor in oil synthesis. Other intermediates in TAG synthesis could be Acyl-CoA and recycled membrane lipid-derived fatty acids (as indicated by gene expression and lipid profiling studies [209][81] [203], as well as exogenous fatty acids exclusively channeled into TAG but not used for starch synthesis [81]). Other notable variations reported on algae C-metabolism under N-deprivation are the shift

from glucose synthesis to storage as starch, and the downregulation of the gluconeogenic and glyoxylate pathways [209], same as the abundance for other selected ribosomal, photosynthetic and lipidic pathways [216]).

Further insights can be found after N-replenishment: when nitrogen is added back to media, nearly 70% of the starch is degraded within 20h, with TAG breakdown occuring more slowly. According to Siaut et al 2011, oil and starch accumulated during N starvation phase are rapidly mobilised upon switching to nutrient replete conditions [273] — starch degradation was found to occur very rapidly after switching to dark, starting earlier than oil degradation, which was degraded between 20 and 24 hours of N resupply — the same results were obtained by Wase et al. 2014 [328].

There are still many other strategies aiming at higher lipid accumulation, e.g. through genetic engineering, both through overexpression of key enzymes for lipid biosynthesis and through inactivation of the dominant starch metabolism. Targeted overexpression has been met with mixed success: whereas overexpression of type 2 diacylglycerol acyltransferases (DGATs) DGAT2-1 and DGAT2-5 led to increased lipid content, DGAT2-a,b,c overexpression had no effect [64][154] and acyl-ACP (acyl carrier protein) esterase (AAE) overexpression led to an altered lipid profile but not an increase in lipid content [31] [270]. Genetic inactivation of starch synthesis has also brought diverse results: higher lipid accumulation in some species like Chlamydomonas or Chlorella [247], no significant changes in other species like *Brassica napus L* [319] and even a decrease in the final seed content of other species like *Arabidopsis thaliana* [237]. In case of Chlamydomonas, many studies have been performed on mutants with disrupted isoamylase (sta-7) or ADP glucose pyrophosphorylase (sta6) genes. An interesting list featuring frequent mutant strains, their genotype, nomenclature and original oil synthesis is enclosed in a study performed by Siaut et al. 2011 [273].

Another possible strategy to find different strains with enhanced lipid productivity in algae is **Adaptive Laboratory Evolution** (**ALE**, [232], [71]), as performed e.g. by Velmurugan 2014 (with an increase of the C:N ratio at the end of the adaptive period in both WT and sta-mutants [315]), Yu 2013 (higher biomass concentrations and total lipid content in 3 different *Chlamydomonas reinhardtii* strains after performing ALE according to Palsson in 3-day-cycles [356]) and Perrineau 2014 (with increased size and grow rate, significant upregulation of genes involved in protein synthesis, the cell cycle and cellular respiration and improved acetate metabolism after 1880 generations under serial dilution and continous light [239]) .

1.3.5 The Concept of Biorefinery

In general, large-scale production of algal biodiesel still is very expensive and energy-intensive due to the high demands of algae harvesting and the biofuel extraction/ transesterification procedures — nowadays the cost of biofuel production from algae might be competitive compared to other biofuel sources, but it still way more expensive than fossile fuels. An interesting strategy to make these processes more sustainable is to combine in the production of different substances in common facilities denominated **Biorefineries** [190] that produce not only biofuels and biomass but also power and high-value chemicals with pharmaceutical or other industrial interest, either simultaneously or in sequence, thus maximizing the profitability of biomass feedstock [62] [223] [287] [313] [190] [276]. Many of these substances are secondary metabolites accumulated by algae under different conditions that can be extracted from the algal biomass, including among others antioxidants (Vitamins E and C, BHT, Gluthathione, Astaxanthine and other carotenoids), fatty acid derivates, polysaccharides, lectins or antibiotics. The de-

sired substance depends on the algae and the biological conditions they are cultivated (as e.g. Chlamydomonas producing Hydrogen under S-deprivation [199], the mesophilic *Haematococcus sp.* accumulating Astaxanthin under elevated temperatures [300], or *Dunaliella sp.* accumulating β-carotene in highly saline environments [20]). An overview of different algae, cultivation conditions and produced substances can be found in Skjånes 2013 [276].

The standard steps for Biorefinery production are algae selection and cultivation, biomass harvesting, cell disruption, fractionation and compound extraction through mild extraction techniques, purification of the extracted substances, and finally drying of the rest material for combustion. The major challenge at this point is the separation and mild extraction of the different fractions from algae, since each compound frequently requires processing methods that might damage the other fractions. The biorefinery techniques appropriate for mild extraction are relatively new and should therefore be studied thoroughly before commercial use is possible [313].

2 Publications

2.1 Functional analysis of proteins and protein species using shotgun proteomics and linear mathematics

One of the elements that confer the proteome its high variability is the possibility to extend the functions of the protein (e.g. activate or inactivate an enzyme), control its behaviour or modify its structure through transient posttranslationational modifications (PTMs) e.g. block expression factors or signal transduction pathways. Even though well known PTMs such as Proline Methylation or Cysteine Carbamoylation can be included in Protein Databases and new resources for identification of PTMs in Mass Spectrometry measurements such as dbPTM have been developed recently ([343], [192]) many modifications or also unidentified peptide fragments that could be useful for EST confirmation might still remain undetected.

Considering the fact that the high mass accuracy from Orbitrap Mass Spectrometers permits peptide identification from precursor ions, the MAPA (Mass Accuracy Precursor Alignment) method was developed by **Hoehenwarter et al. 2011 [126]**. This procedure is based on the reducing proteome data to two dimensions (m/z for peptide identification and spectral counting for ion quantification) and building a matrix in which the measured peptides are arranged according to these two parameters. This is achieved by the ProtMax-Algorithm (Hoehenwarter et al. 2008). This algorithm allows unbiased, database-independent peptide analysis, in which previously unknown peptides can be sorted, statistically analysed, and finally processed through different strategies, such as testing different databases, matching with EST libraries, or even De Novo-Sequencing [127]. Therefore, the MAPA method is a very powerful tool for performing shotgun proteomics analysis on organisms that haven't been fully sequenced yet, and to search peptides from posttranslationally modified proteins that might act as biomarkers [127].

For the implementation of the MAPA strategy data from many organisms were used — among others, also from my studies on *Chlamydomonas reinhardtii*.

2.1.1 Declaration of authorship

The results of this chapter are presented in the form of a manuscript published in the journal "Amino Acids". For this work, I contributed to the experimental work relative to different measurements on *Chlamydomonas reinhardtii* and the subsequent data mining that led to some of the datasets used during the elaboration of this publication.

2.1.2 Published manuscript

Amino Acids (2011) 41:329–341
DOI 10.1007/s00726-010-0669-1

ORIGINAL ARTICLE

Functional analysis of proteins and protein species using shotgun proteomics and linear mathematics

Wolfgang Hoehenwarter · Yanmei Chen · Luis Recuenco-Munoz · Stefanie Wienkoop · Wolfram Weckwerth

Received: 1 April 2010 / Accepted: 16 June 2010 / Published online: 3 July 2010

Abstract Covalent post-translational modification of proteins is the primary modulator of protein function in the cell. It greatly expands the functional potential of the proteome compared to the genome. In the past few years shotgun proteomics-based research, where the proteome is digested into peptides prior to mass spectrometric analysis has been prolific in this area. It has determined the kinetics of tens of thousands of sites of covalent modification on an equally large number of proteins under various biological conditions and uncovered a transiently active regulatory network that extends into diverse branches of cellular physiology. In this review, we discuss this work in light of the concept of protein speciation, which emphasizes the entire post-translationally modified molecule and its interactions and not just the modification site as the functional entity. Sometimes, particularly when considering complex multisite modification, all of the modified molecular species involved in the investigated condition, the protein species must be completely resolved for full understanding. We present a mathematical technique that delivers a good approximation for shotgun proteomics data.

Keywords Mass spectrometry · Proteomics · Protein function · Protein species · PTM

W. Hoehenwarter (✉) · Y. Chen · L. Recuenco-Munoz · S. Wienkoop · W. Weckwerth
Department of Molecular Systems Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria
e-mail: wolfgang.hoehenwarter@univie.ac.at

Proteomes and proteomics: an update

The dynamic total protein complement of a biological system, the proteome, is of enormous complexity and size. A well-defined concept to deal with this situation is the speciation of the proteome (Jungblut et al. 2008). Each and every polypeptide is defined by the sum of its covalent chemical bonds meaning its primary structure and in addition any covalently bonded moieties. Thus, every protein, every expressed member of a multigene family, i.e. protein isoform, and post-translationally modified form of a protein is understood as a unique protein species. It is not clear if every protein species has its own function, the paradigm, however, captures the extent of function in the proteome. Newest insights, however, are that one primary structure can result in multiple secondary structures and conformations and thus possible multiple functions due to bias towards synonymous, slowly translated codons (Zhang et al. 2009).

The meta-analysis of large scale genomics, transcriptomics and proteomics studies and the use of central data repositories such as National Center for Biotechnology Information (NCBI) and The Arabidopsis Information Resource (TAIR) allow a relatively good survey of the proteome of higher eukaryotes (Fig. 1). The lower proteome entries (red) intersect with the upper genome entries (black) on their common axis reflecting the manifold greater complexity of the proteome. The human being, mouse and rat as well as the small flowering plant *Arabidopsis thaliana* all have around 25,000 protein encoding genes (Collins et al. 2004; Gibbs et al. 2004; Kaul et al. 2000; Waterston et al. 2002). Transcriptome analysis of 32 human tissues revealed that only about half of these are expressed at one time (Jongeneel et al. 2005). Transcriptome-based monitoring of more than 10,000 multi-exon

genes in 52 human tissues and cell lines showed that nearly every gene is alternatively spliced (Johnson et al. 2003) and other studies suggest many genes will generate multiple, sometimes up to a hundred or thousand alternative mRNAs (Schmucker et al. 2000; Tress et al. 2008). Therefore, the amount of translated expression products present in a cell at any one time was estimated to be approximately 30,000. More than 200,000 post-translational modifications (PTM) of proteins are reported in the literature (Seo and Lee 2004) and numerous proteomics studies have identified tens to hundreds of modifications of a single protein (Garcia et al. 2007; Larsen et al. 2001; Scheler et al. 1997). The average number of differently processed expressed gene products is estimated at 5–7 in humans (Humphery-Smith 2004), ultimately allowing a conservative estimate that the proteome of the higher eukaryotic cell comprises 150,000 protein species at one time. Others put the number at 1 million (Humphery-Smith 2004). As an afterthought, we note that we have not considered the dynamic range of protein abundance which has been shown to be four to five orders of magnitude in *E. coli* and yeast (around <100 to 2,000,000 copies per cell) (de Godoy et al. 2008; Lu et al. 2007; Usaite et al. 2008) and ten or more in human plasma (Anderson and Anderson 2002).

The current proteomics technologies have had some success in dealing with this complexity. Shotgun proteomics is the most popular application for unbiased, discovery type analysis of the proteome. The proteins are first digested into peptides followed by an online combination of liquid chromatography (LC) for peptide separation and electrospray ionization mass spectrometry (ESI-MS) for the measurement of peptides mass to charge ratios (*m/z*). It has been used to quantify all of the translated expression products of yeast (de Godoy et al. 2008) and major coverage of all of the open reading frames of *Drosophila* (Brunner et al. 2007), mouse (Graumann et al. 2008) and *Arabidopsis* (Baerenfaller et al. 2008) has also been achieved. The large scale identification and quantification of PTMs has had some encouraging results as well. Several groups have reported thousands of phosphorylation sites (Beausoleil et al. 2004; Olsen et al. 2006, 2010; Schmidt et al. 2008; Sugiyama et al. 2008; Villen et al. 2007), acetylation sites (Choudhary et al. 2009) and other modifications (MacCoss et al. 2002; Tsur et al. 2005) and hundreds of ubiquitination sites (Igawa et al. 2009; Maor et al. 2007; Peng et al. 2003). Nevertheless, the proteome-wide mapping of PTMs remains a difficult task because the modifications are transient, modified peptides are often not very abundant and require specific enrichment strategies and their tandem mass spectra (MS/MS spectra) are difficult to interpret.

Two-dimensional gel electrophoresis (2DE), tried and tested for more than 30 years, still provides the best resolution of the proteome on the protein level. It is capable of separating the heterogenous mixture into up to 10,000 components according to their isoelectric points (PI) and molecular weights (M_w) (Klose and Kobalz 1995). The protein species remain intact giving researchers access to the entire molecule. Analysis with liquid chromatography for further separation of the proteins at distinct positions in the gel matrix (2DE spots) and the highly sensitive new generation of mass spectrometry machines has revealed that generally a few protein species co-migrate (Jungblut et al. 2010). This precludes accurate quantification of individual protein species by measuring polypeptide staining intensity; they can, however, be quantified by labeling them with stable isotopes either directly or by covalently attaching chemical moieties and then comparing their mass spectrometric signals (Schulze and Usadel 2010). Despite these technological advances, the maximum entry on the proteomics axis (gray) in Fig. 1 is still more than one order of magnitude below the maximum entry for proteomes highlighting the need for further innovation and that proteomics science will remain challenging for some years to come.

PTMs: switches that control protein function

Post-translationally modified proteins have a vast array of functions in the cell, which are often quite different from those of their unmodified counterparts. Covalent modification of primary structure can alter protein conformation thereby producing new recognition motifs, hiding existing motifs or making previously hidden motifs accessible. PTMs can have allosteric effects, controlling protein function via modification of non-active sites. Initial modification can recruit subsequent additional modifications by either the same or different chemical moieties which may act in concert. Two or more different PTMs can compete for modification of the same residue in protein primary structure. All of these mechanisms affect protein interaction via specific PTM-binding domains which defines function (Seet et al. 2006).

The covalent attachment of phosphate moieties donated by ATP and catalyzed by kinases via esterification of side chain hydroxyl-groups of primarily serine, threonine and tyrosine residues is the most widely studied PTM. It occurs on approximately one-third of eukaryotic proteins and is mostly a monoadduct (Cohen 2000; Hubbard and Cohen 1993). The phosphoryl-group can be removed by phosphatase catalyzed hydrolysis so protein phosphorylation is reversible and in most cases transient. It is a central switch of protein function. It controls the activity of the components of signaling cascades such as the canonical mitogen-associated protein kinase (MAPK) cascade downstream of the tyrosine auto-phosphorylated epidermal growth factor

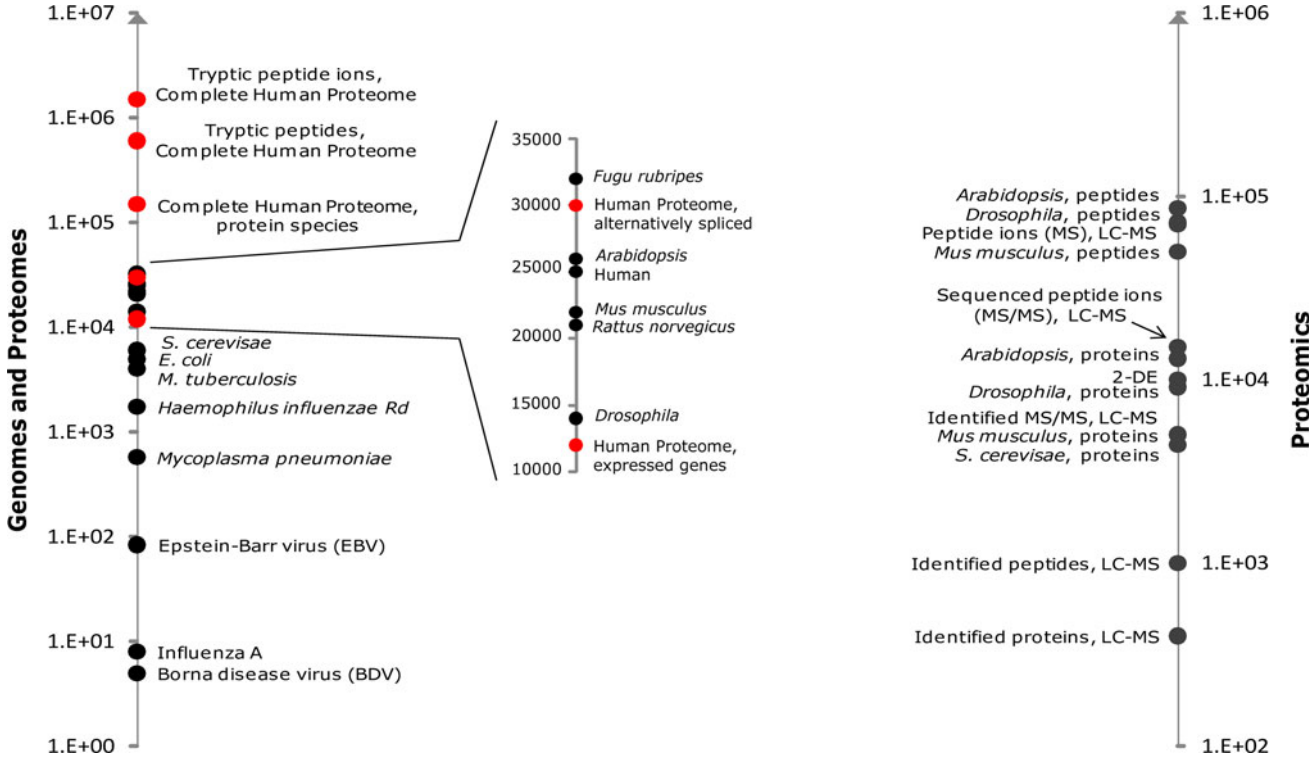

Fig. 1 Some important reference values regarding genomes in *black*, proteomes in *red* and proteomics in *gray*. The axes are in logarithmic scale. The values are from pertinent publications (Adams et al. 2000; Ahn et al. 2007; Aparicio et al. 2002; Baer et al. 1984; Baerenfaller et al. 2008; Bogdanov and Smith 2005; Brunner et al. 2007; Collins et al. 2004; Cox and Mann 2008; Cubitt et al. 1994; de Godoy et al. 2008; Fleischmann et al. 1995; Garcia et al. 2007; Ghaemmaghami et al. 2003; Gibbs et al. 2004; Goffeau et al. 1996; Graumann et al. 2008; Johnson et al. 2003; Jongeneel et al. 2005; Jungblut et al. 2008; Kaul et al. 2000; Klose and Kobalz 1995; Larsen et al. 2001; Misra et al. 2002; Myers et al. 2000; Ono et al. 2006; Scheler et al. 1997; Schmucker et al. 2000; Tress et al. 2008; Waterston et al. 2002) and from our experience. The RNA viruses have very small genomes with less than ten genes; the Epstein–Barr virus has a double-stranded DNA genome of 84 open reading frames. Unicellular organisms all have less than 10,000 genes. Peptide ions (MS), LC–MS refers to all recorded *m/z* in a 1D shotgun proteomics analysis, sequenced peptide ions denotes the total number of recorded MS/MS spectra

(EGF) receptor (Pawson and Nash 2000) that consists of the MAPK kinase kinase (MAPKKK) RAF, the MAPK kinase (MAPKK) MEK/MKK1 and the MAPK ERK (Marshall 1994). It also controls the MAPK cascade MEKK1-MKK4/MKK5-MPK3/MPK6 that administers response to bacterial pathogens in plants (Asai et al. 2002) following perception of the flg22 epitope of bacterial flagellin by the leucine rich repeat receptor kinase (LRR-RK) FLS2 (Chinchilla et al. 2006). MAP kinases are modular as well as promiscuous. This means that cascades can be activated by diverse stimuli, that components of different cascades are interchangeable and act on several substrates to communicate a host of different signals and that they produce a wide range of functional responses (Nakagami et al. 2005). The MAPKs are activated and deactivated by phosphorylation and dephosphorylation at multiple sites which gives rise to a large number of functional protein species whose abundance and very existence is in a high state of flux.

Phosphorylation can determine the translocation of proteins to organelles and subcellular compartments for instance in the case of protein phosphatase-1 (PP1) which is released from glycogen particles to the cytosol upon phosphorylation of one of its subunits, G_M (Hubbard and Cohen 1993). Apoptosis is inhibited by binding and sequestering of the phosphorylated proapototic BAD by 14-3-3 protein (Lizcano et al. 2000). Gene transcription is initiated by phosphorylation of the active subunit of the transcription factor NFκB, p65–p50 (Wang and Baldwin 1998), phosphorylation of members of the Jun, Fos and ATF transcription factor families (Karin et al. 1997; Murphy et al. 2002; Ventura et al. 2003) as well as other DNA-associated proteins. Transcription is inhibited by phosphorylation of HY5, a b-ZIP transcription factor that controls light induced gene expression in plants (Hardtke et al. 2000).

Reversible protein phosphorylation ties in with ubiquitination, and indeed, there is extensive regulatory crosstalk between these two PTMs (Hunter 2007). Protein species such as the inhibitory subunit of NFκB, IκBα, phosphorylated at serine residues 32 and 36 (Karin 1999; Yaron et al. 1998), phosphorylated HIV-1-Vpu (Margottin et al.

1998), cyclins and cyclin-dependent kinase (CDK) inhibitors (Koepp et al. 1999) as well as dephosphorylated β-catenin (Kitagawa et al. 1999) and the anti-apototic Bcl-2 protein (Dimmeler et al. 1999) are targeted for ubiquitination and ultimately degradation by the 26S proteasome. The phosphate moiety and surrounding amino acids, which has generally become known as a phosphodegron, is the recognition signal for the ubiquitin protein ligase E3. The enzyme binds the phosphodegron on the substrate protein via an F-box subunit that contains the WD40 or LRR-binding domains and catalyzes the covalent attachment of ubiquitin to an ε-amino group of an internal lysine residue or to the α-amino group (Breitschopf et al. 1998). Alternatively, the two main types of E3 ubiquitin protein ligases, the single or multisubunit RING finger ligases that lack catalytic activity and function by recruiting the E2-ubiquitin conjugated intermediate and the target protein substrate and the catalytic HECT domain E3 ligases can themselves be controlled by phosphorylation. The anaphase promoting complex type E3 enzyme is phosphorylated conferring substrate specificity and recognition properties (Shteinberg et al. 1999) and the Itch E3 ligase is allosterically activated by phosphorylation (Gallagher et al. 2006).

Protein ubiquitination is reversible through nucleophile attack of the peptide bond between ubiquitin and the substrate catalyzed by deubiquitinating enzymes (Dubs) (Sowa et al. 2009; Ventii and Wilkinson 2008). Reversible ubiquitination controls cellular processes and in particular kinase activity and protein phosphorylation primarily by controlling the stoichiometry of their components via degradation. The regulatory phosphate moiety on the active kinase becomes a phosphodegron that is recognized by the enzymes of the ubiquitination machinery leading to covalent attachment of polyubiquitin moieties branched at lysine residue 48 of ubiquitin followed by degradation by the proteasome. In some cases activated kinases are targeted to the lysosome/vacuole or subcellular compartments, which is mediated by lysine 63 branched polyubiquitination and multiubiquitination (Huang et al. 2006). Polyubiquitin of kinases branched at lysine residues 29 or 33 can directly inhibit kinase activity and kinases can be activated by polyubiquitin promoted transphosphorylation (Chen 2005). In plants, the ubiquitin proteasome system may play a role in sucrose related processes possibly via substrate proteins phosphorylated by Snf-1-related protein kinases (SnRKs) (Ellis et al. 2002). It has also been implicated in activation of the auxin response by degradation of Aux/IAA proteins and in suppression of transcription of light responsive genes by degradation of unphosphorylated HY5 (Ellis et al. 2002).

Acetylation of lysine as well as serine and threonine residues is another PTM that has far reaching impact on cellular functions. The formation of the amide bond between the lysine ε-amino group and the hydroxyl group of the acetate moiety that is donated by Acetyl-CoA is catalyzed by histone acetyltransferases (HAT, also known as lysine acetyltransferases [KAT]). Its hydrolysis is mediated by histone deacetylases (HDAC or in analogy KDAC for lysine acetyltransferase). As for phosphorylation and sumoylation but not for ubiquitination, primary structure consensus sequences have been identified and it has been ascertained that lysine acetylation is favored in ordered secondary structure as well as in macromolecular complexes (Choudhary et al. 2009; Kim et al. 2006) such as the nuclear HAT complexes themselves (Thompson et al. 2004) or the major actin nucleation complex ARP2/3.

There is ample evidence that protein function is regulated by acetylation in conjunction with phosphorylation, ubiquitination and other PTMs. The abundance of phosphorylated protein species is influenced by acetylation exercising control over kinases, for example the phosphoinositide-3-kinase related protein kinases (PIKK) that are integral to DNA damage repair (Jiang et al. 2006; Sun et al. 2005). The activity of CDC2, a kinase involved in cell cycle progression and mitosis, and CDK9 may be abolished by acetylation in their kinase domains (Sabo et al. 2008). Competitive acetylation of serine and threonine residues in the activation loop of MAPKKs is used by *Yersinia* species to shut down signal transduction via the MAPK cascade and overcome the immune response (Mukherjee et al. 2006). Binding of phosphorylated primary structure motifs by the 14-3-3 domain is inhibited by its acetylation (Choudhary et al. 2009) and acetylation and methylation in cis as well as in trans affects the interaction of binding domains and recognition motifs on histones (Fischle et al. 2003; Latham and Dent 2007). Ubiquitination, methylation, SUMOylation and acetylation all compete for modification of lysine residues and thereby modulate the activity of such prominent examples as the tumor suppressor p53, the transcriptional cofactor p300 and the nuclear transport protein RANGAP1 (Bouras et al. 2005; Mahajan et al. 1997; Yang and Seto 2008). Ubiquitin ligases and DUBs are themselves extensively acetylated which may affect their function and the ubiquitination of other proteins (Choudhary et al. 2009).

In summary, phosphorylated, ubiquitinated and acetylated protein species regulate nearly all aspects of cell life (Choudhary et al. 2009; Ciechanover et al. 2000; Cohen 2000). This includes metabolism (Cohen 1999; Kempa et al. 2007; Kim et al. 2006; Polge and Thomas 2007), the cell cycle (Brooks and Gu 2003; Choudhary et al. 2009; Glotzer et al. 1991; Inze and De Veylder 2006; O'Connell et al. 2000; Zhang et al. 2008) and cytokinesis (Takahashi et al. 2004) and growth and death (Haas et al. 1995; Kim et al. 2006). Signal transduction and concurrently the response to infection (Kotlyarov et al. 1999; Rock and

Goldberg 1999), pathogens (Asai et al. 2002; Chisholm et al. 2006; Devoto et al. 2003; Kunkel and Brooks 2002) and abiotic stress (Ichimura et al. 2000; Kempa et al. 2007; Nakagami et al. 2005) is also highly regulated by PTMs.

Reversible, PTM at multiple sites by different moieties produces a tremendous amount of transiently abundant protein species that are the primary conveyors of function in the cell. These protein species greatly amplify the functional repertoire inherent in the proteome beyond the unmodified expressed genome. Indeed, the unmodified proteins and protein complexes may be seen as the molecular machinery, the PTM as the switches used to operate it.

Shotgun proteomics unravels the cellular signaling network

The function of phosphorylated protein species was traditionally investigated with reverse genetic approaches combined with enzyme activity and protein interaction assays, particularly in the case of signaling cascades and the phosphorylation status of their components (Ahlfors et al. 2004; Asai et al. 2002; Chinchilla et al. 2007; Meskiene et al., 1998, 2003; Nuhse et al. 2000; Schweighofer et al. 2007; Takahashi et al. 2007). Recent advances in shotgun proteomics have made the detection and quantification of site-specific phosphorylation on a proteome-wide scale feasible (Amanchy et al. 2005; Beausoleil et al. 2004; Benschop et al. 2007; Chen et al. 2010; Chi et al. 2007; Nuhse et al. 2004, 2007; Olsen et al. 2006, 2010; Schmidt et al. 2008; Schulze 2010; Sugiyama et al. 2008; Thelemann et al. 2005; van Bentem and Hirt 2007; Villen et al. 2007).

A landmark is the study of the phosphorylation events following EGF stimulation of HeLa cells that employed subcellular fractionation to analyze the nuclear and cytosolic protein complement, strong cation exchange and titanium dioxide affinity chromatography to enrich phosphorylated peptides and stable isotope labeling (SILAC) and shotgtun proteomics to identify and quantify the phosphorylation of amino acid residues over time (Olsen et al. 2006). It describes the kinetics of 6,600 phosphorylation sites mapped to 2,244 proteins, around 20% of the expressed open reading frames, and of EGF signaling in unprecedented detail, including the entire MAPK cascade, a large number of transcription factors and associated proteins that were not known to be involved in growth factor signaling, cytoskeletal proteins such as actin and GTP-associated proteins, the ubiquitination machinery and RNA-binding proteins. Phosphorylation at multiple sites in many cases with different kinetics was shown to predominate and to regulate protein function as exemplified by the early, activating tyrosine autophosphorylation of the EGF receptor that declined concomitantly with later phosphorylation of serine and threonine residues, which is known to attenuate the signal via negative feedback (Schlessinger 2000).

Two studies of the plasma membrane-associated proteome that applied similar methodologies as the work on the EGF receptor describe the early response to perception of flg22 in *Arabidopsis* cell culture (Benschop et al. 2007; Nuhse et al. 2007). Like EGF receptor signaling it is hallmarked by transient protein phosphorylation. Benschop et al. (2007) demonstrated that the measured changes in the phosphorylation levels of amino acid residues were due to reaction kinetics and not changes in protein abundance by normalizing the abundance ratios of phosphorylated peptides to peptides that did not contain a phosphorylation site on the same protein. Both studies identified phosphorylation sites on numerous receptor-like kinases (RLK), implicating phosphorylation in sensitizing the plant to further pathogen challenge and priming of defense response (Conrath et al. 2002; Zipfel et al. 2004). Phosphorylation of the components of the MAPK cascade, regulatory protein phosphatases such as PP2C (Schweighofer et al. 2007) and auxin signaling proteins was also found. The latter connects MAPKs with the down regulation of auxin signaling which is known to play a role in the plant immune response (Kovtun et al. 1998; Navarro et al. 2006).

The induced phosphorylation of many sites on a single protein was also reported by both authors. Differential kinetics of several residues of the H^+-ATPases AHA1 and AHA2 and reduction of phosphorylation of the penultimate threonine residue 948 which directly controls protein activity (Palmgren 2001) was detected following treatment with flg22. More pronounced, the respiratory oxidative burst protein RBOHD, an NADPH oxidase involved in the production of reactive oxygen intermediates in the apoplast following pathogen perception and integral to defense response was found to be differentially phosphorylated at up to seven distinct residues. This suggests the protein is controlled by various interacting partners and that it administers distinct functions in manifold physiological pathways in a complex regulatory network.

The widespread incidence and far reaching affects of acetylation in all cellular compartments was discovered for the first time by two shotgun proteomics studies (Choudhary et al. 2009; Kim et al. 2006). The more recent work describes the kinetics of 3,600 sites on 1,750 proteins in three human cell lines including one affected by acute myeloid leukemia in response to two KDAC inhibitors, one of which (SAHA) is in clinical use. The studies show that multisite acetylation of proteins is a central regulator of all nuclear and DNA-associated processes. This includes chromatin remodeling and DNA replication, transcription,

splicing, DNA damage repair and nuclear transport. They detected the hitherto unknown acetylation of numerous cytoplasmic metabolic enzymes including aldolase which also binds actin and provide evidence for a major role of acetylated proteins in cytoskeleton architecture as actin-, micro- and intermediate filament proteins were all functionally acetylated (Anastasiadis et al. 2000; Posern et al. 2004; Zhang et al. 2007). They also showed that the modification is prominent in the mitochondrion and uncovered a strong link to its regulation of energy metabolism (Kim et al. 2006). Many of the proteins of the TCA cycle, oxidative phosphorylation, lipid, carbohydrate, amino acid and nucleotide metabolism as well as the mitochondrial dehydrogenase protein complexes were found to be acetylated. The activity of the dehydrogenase complexes is controlled by NAD^+ to NADH ratios and Acetyl-CoA levels which also serve as substrates or cofactors for HATs and HDACs (Blander and Guarente 2004) suggesting a regulatory feedback loop.

Acetylation of multiple lysine residues of the same protein was prevalent like in the studies of phosphorylation. The herpesvirus-associated ubiquitin-specific protease (HAUSP), a DUB that influences the nuclear cytoplasmic partitioning of the tumor suppressor PTEN via deubiquitination was acetylated at five different residues. The cytoskeletal protein cortactin was modified at 7 sites and 14 acetyl-lysine residues were identified for the HSP90-α subunit. This further indicates that intricate combinatorial switches that consist of reversible modification at distinct residues by different interchangeable moieties are common controls of protein function.

The largest work on PTM sites to date was published recently. It defines the kinetics of over 20,000 phosphorylation sites on 6,000 proteins in the cell cycle and highlights what is possible with mass spectrometry-based proteomics today (Olsen et al. 2010). While not yet fully comprehensive, shotgun proteomics has provided substantial information on the in vivo post-translationally modified proteome, recording the state of thousands of residues on thousands of proteins under various conditions. Shotgun proteomics studies have elucidated an extensive regulatory network based on the interactions of transiently abundant highly modified protein species that is universally active in the cell and given a global perspective on the function and profound importance of PTMs.

Shotgun proteomics and protein species function

The measurements of the state of modification of individual amino acid residues reflect the effects of the experimental condition under investigation such as exposure to a growth factor or temperature stress. In most cases they will also reflect the effects on individual proteins. An example is the differential phosphorylation of the EGF receptor described above. The observed kinetics are in agreement with the known changes in the phosphorylation state of several residues in response to EGF perception. In cases of more complex multisite modification, such as the differential phosphorylation of the RBOHD protein at seven or the acetylation of HSP90-α at 14 different residues, the effects of the experimental condition on the protein may not be so clear. Although the induced changes in the modification state at multiple sites can undoubtedly be brought into causal connotation with the condition, it is not straightforward to assay if all of the sites are truly localized on one protein, or more precisely on one protein species, or are shared by several protein species that administer function in concert.

This point must not be ignored. As we know, the polypeptide molecule which has its conformation defined by primary structure and which in many cases is a subunit of higher order macromolecular assembly is the basic functional entity and not the peptide or amino acid residue (we will not consider the point of multiple secondary structures with the same primary structure mentioned in the introduction).To truly pinpoint function in some cases it may be necessary to have the knowledge of the full primary structure of all polypeptides involved in a physiological process in addition to modification kinetics. This can be achieved experimentally by separating the proteome into its protein species using top-down technologies such as 2-DE, or, which may be the next true breakthrough in proteomics research, top-down LC-ESI-MS proteomics (Chait 2006; Siuti and Kelleher 2007).

Mathematical procedures are very powerful and can give good approximations of protein species function from shotgun proteomics data. The proteome is digested into peptides so the protein species are not separated prior to analysis. Nevertheless, we routinely use the variance and covariance of the abundance of the identified proteins, which is inferred by the measured abundance of the peptides assigned to them by comparative search of translated nucleic acid sequences or de novo spectral interpretation to explore a multitude of biological questions. We do this primarily with a combination of the canonical principal component analysis (PCA) (Pearson 1901) and independent component analysis (ICA) (Comon 1994). These techniques make it possible to determine correlations between the shotgun proteomics analyses and the phenotypes under different experimental conditions (Scholz et al. 2004).

Data consisting of multiple observations of a large number of variables can be seen as a number of vectors or points in space. PCA and ICA compress and project this data into a lower number of dimensions, in a direction so

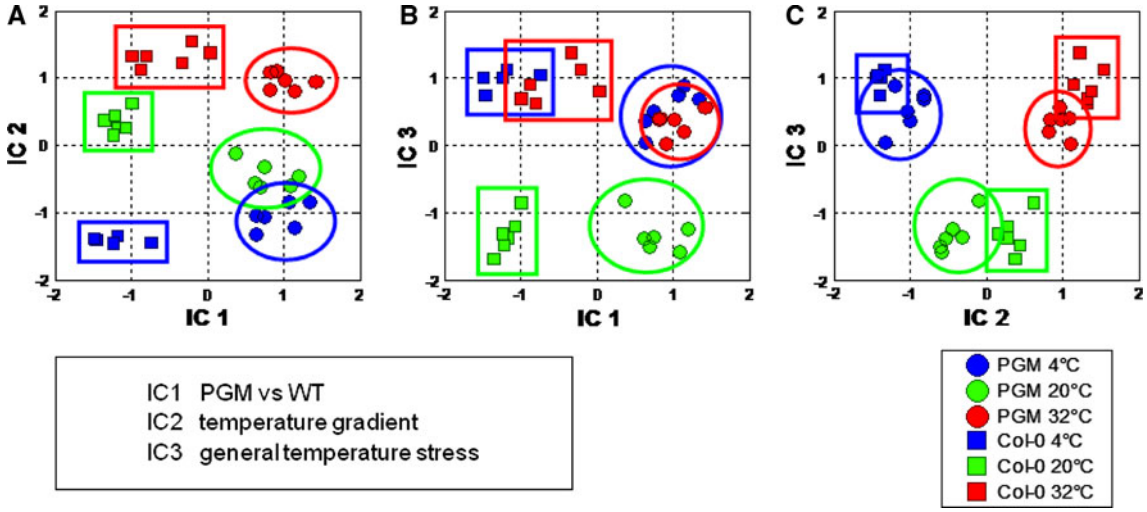

Fig. 2 Visualization of shotgun proteomics data with ICA. Thirty-six combined shotgun proteomics and metabolomics analyses of two genotypes of *Arabidopsis thaliana*, the wild type Col-0 and a starch-deficient mutant PGM, adapted to three temperatures, 4, 20 and 32°C, for 3 days, are plotted in three 2-dimensional plots (**a**–**c**) of an optimal three dimensional space using combined principal and independent component analysis (PCA/ICA). Four principal components from the PCA which was applied first were used as the input for subsequent ICA. The first dimension, IC1 contains information on the proteins that are discriminatory for the genotypes, the second, IC2, on proteins indicative of temperature adaption, the third, IC3, on proteins expressed under temperature stress. The sample pattern in all three dimensions, i.e. the complete data structure nearly perfectly reflects the applied experimental conditions. The figure was taken from Wienkoop et al. (2008) and is reprinted here with permission

that the maximum amount of information is retained, the principal or independent components (PC or IC). The data is visualized in up to three dimensions and the user can judge structure and inherent patterns that can shed light on relationships that may be interesting. The PC or ICs themselves can be expressed in terms of the original variables making it clear which parts of the original data are prominent and determine the observed structure and relationships. These can then be interpreted in a biological context.

In more mathematical terms, PCA can be explained as an eigenvalue decomposition of the covariance matrix. The principal components are the orthonormal eigenvectors that indicate the directions of maximal variance in decreasing order of their eigenvalues and span the lower dimensional space of linear combinations of the original variables for mapping with the minimum reconstruction error. Their coordinates in the original higher dimensional space are the loadings that place a scalar value on the contribution of the variables, in our case the identified proteins, to the data structure. ICA is an extension of PCA where the components are independent, a stronger condition than uncorrelation in PCA, and no longer restricted to orthonormality. The independent components (IC) are sorted according to their kurtosis, the fourth auto-cumulant of a distribution of parameter values following the mean, variance/standard deviation and skewness and which is an indicator of its Gaussian fit. A sub-Gaussian or flat distribution can be the result of a single or few significant differences in parameter values and is therefore often of biological interest, i.e. a significant difference in protein abundance under two or more conditions. As in PCA, the loadings are the coordinates of the independent components in the higher dimensional space.

Figure 2 taken from a recent study of ours (Wienkoop et al. 2008), shows a plot of 36 combined shotgun proteomics and metabolomics analyses of two *A. thaliana* genotypes, the Col-0 wild type and a plastidic PGM mutant, under cold (4°C), normal ambient (20°C) and heat (32°C) conditions reduced to three dimensions using PCA/ICA. Figure 2a, b both contain the first dimension, IC1, and looking at the data from this angle, it is clear that IC1 separates the two investigated genotypes. Figure 2a additionally shows the second dimension, IC2, of the compressed data which separates the applied temperatures in a gradient that goes from low to high. Figure 2b contains the third dimension, IC3, in addition to the first, which discriminates the temperature stressed conditions of 4 and 32°C from the normal environmental condition of 20°C. Figure 2c combines the second and third dimensions; the former again shows the gradient from low to high temperature, the latter again distinguishes the stressed conditions from the normal ambient temperature. These two independent components do not contain any information regarding the genotypes so they are congruent.

The plotted reduced data contains 95% of the extent (to be mathematically precise, the variance) and also 95% of the information of the original data. Therefore, the

correlation between the shotgun proteomics analyses which are fully quantitative and the investigated phenotypes is as equally nearly perfect as the correlation between the lower dimensional data and the phenotypes as seen in the plot. This means that the analyses can be used to extrapolate function. It can also be shown analytically, that there is a fundamental relationship between the covariance of the measured proteomics data and the fluctuating concentrations of protein species in physiological pathways which can be expressed as entries in the Jacobian matrix (Steuer et al. 2003). In the given example low-temperature induced protein 78 and cold-regulated protein COR6.6 (KIN2) as well as several RNA-binding proteins were shown to regulate the plants' adaption to temperature (Kim et al. 2005; Smallwood and Bowles 2002).

To rapidly determine the sites of functional PTMs, we have developed an application for comparison of hundreds of shotgun proteomics analyses called mass accuracy precursor alignment (MAPA) (Hoehenwarter et al. 2008). It works by measuring the *m/z* of peptides with very high accuracy (average error ~1.5 ppm, SD ~1 ppm) which is a unique identifier and allows the abundance of every measured peptide in any number of shotgun proteomics analyses to be aligned in a quantification matrix. It is done with a program we developed called ProtMAX. It places the *m/z* ratios of all peptides ions for which at least one MS/MS spectrum is available in all of the compared analyses in the rows, the analysis identifiers in the columns and the number of MS/MS spectra for each *m/z* ratio in each analysis in the cells. The number of MS/MS spectra is the spectral count of each peptide ion which correlates linearly with peptide abundance over two orders of magnitude (Abdi et al. 2006; Old et al. 2005; Wienkoop et al. 2006) so the matrix contains a quantitative value for each peptide in each analysis. The strength of MAPA is that it produces an accurate quantitative comparison of any number of shotgun proteomics measurements in minimal time with minimal computing power.

PCA/ICA as well as clustering or supervised classification schemes can be employed to find the peptides that are especially characteristic of the experimental conditions. The quantification matrix is unbiased. It is inclusive of all of the measured peptides and does not omit post-translationally modified peptides that are difficult to identify by searching databases of translated DNA sequences. Therefore, post-translationally modified peptides that are correlated with the different experimental conditions can be readily detected and then identified by de novo spectral interpretation.

We employed this strategy to explore the processes involved in hormone signaling in *A. thaliana* cell culture, 1, 3 and 6 h after exposure to the phytohormones abscisic acid (ABA), gibberellic acid (GA), auxin (IAA), jasmonate (JA) and kinetin (Chen et al. 2010). In total, 152 phosphorylated peptides were identified and quantified using mass spectrometry. These peptides contained 170 phosphorylation sites. They were all differentially responsive to at least one of the hormones and could be mapped to 130 proteins. Many of the sites were induced by several hormones which is indicative of activated protein species transmitting signals via several pathways in the regulatory network. The abundance of the phosphorylated peptides at all of the time points after exposure to the phytohormones was used to produce the MAPA quantification matrix. This quantitative data was then used to model the phosphorylation kinetics with the combined PCA/ICA.

The activation of MAPK and calcium-dependent kinase (CDPK) cascades in response to phytohormone perception is well known (Alonso and Stepanova 2004; Kovtun et al. 1998; Ludwig et al. 2005; Navarro et al. 2006; Schweighofer et al. 2007; Takahashi et al. 2007). As a small example from the study, Fig. 3 shows the phosphorylation state of four sites on the central modules of cell signaling from the membrane to the nucleus 3 h after treatment with each of the hormones. The RLK is phosphorylated in response to the continued perception of IAA, JA and kinetin. The cytosolic CDPK is induced under all of the conditions which can be expected in light of its promiscuous nature. The exclusive phosphorylation of the ABA responsive element-binding factors (AREB) in response to ABA indicates the signal has been transmitted through the network and reached its highly specific destination. The sites on the AREBs had not been described previously showing the power of MAPA and PCA/ICA to uncover novel regulatory modifications because of their causal and potentially functional correlation to the phenotype.

Concluding remarks

The proteome is vastly more complex and dynamic than the genome. PTM in particular expand its functional potential immensely. Shotgun proteomics has measured the modification kinetics of a very large number of amino acid residues on an equally substantial number of proteins and uncovered a regulatory network that extends into all aspects of cellular physiology.

Transient modifications of multiple sites by different interchangeable moieties are recognized by PTM-binding domains that are organized as modules in proteins (Seet et al. 2006; Yang 2005). The dynamic interaction of recognition motifs and binding domains executes a powerful regulatory program that is flexible, sensitive and far reaching. These insights, where the modular-binding domains "read" the state of modification of proteins that

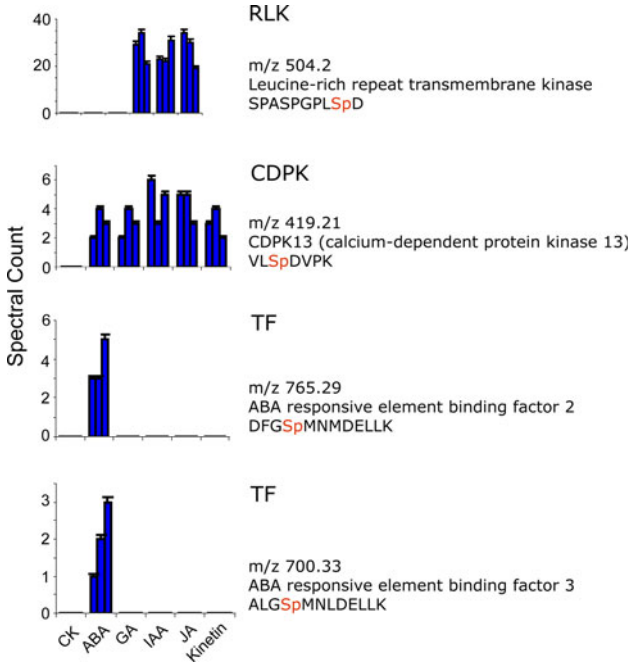

Fig. 3 Phosphorylation state of key components of the signal transduction cascade 3 h after induction with five phytohormones in *Arabidopsis thaliana* cell culture. The phosphorylation kinetics of four serine residues on a receptor-like kinase (RLK), a CDPK and transcription factors (TF) were determined with MAPA. The *m/z* and peptide primary structure are shown, the phosphorylation site is colored *red*

controls function are expressed in the emerging concept of the protein code (Gimona 2006; Sims and Reinberg 2008). Considering the large number of known PTMs, the common phenomenon of dynamic multisite modification and the equally large number of interaction domains, this "protein code", if one wishes to use the term, may be much more complex than the genetic code and constitute a primary functional dimension in the molecular biology of the cell.

To truly understand the mechanisms of molecular interaction and function in some cases it may be necessary to resolve the proteome on a molecular level, the protein species level. Shotgun proteomics studies in conjunction with higher mathematical procedures can give good approximations of protein species function. Ideally they should be employed to gain a more or less comprehensive overview of the question at hand. They should be complemented with top-down technologies such as 2-DE to focus on the protein species of interest and genetic and biochemical techniques to validate the functional hypothesis.

Acknowledgments We thank Peter R Jungblut for his kind solicitation of the manuscript and for the many years of stimulating discussion. WH is generously supported by EraNet Plant Genomics.

References

Abdi F, Quinn JF, Jankovic J, McIntosh M, Leverenz JB, Peskind E, Nixon R, Nutt J, Chung K, Zabetian C et al (2006) Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. J Alzheimers Dis 9:293–348

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF et al (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195

Ahlfors R, Macioszek V, Rudd J, Brosche M, Schlichting R, Scheel D, Kangasjarvi J (2004) Stress hormone-independent activation and nuclear translocation of mitogen-activated protein kinases in Arabidopsis thaliana during ozone exposure. Plant J 40:512–522. doi:10.1111/j.1365-313X.2004.02229.x

Ahn NG, Shabb JB, Old WM, Resing KA (2007) Achieving in-depth proteomics profiling by mass spectrometry. Acs Chem Biol 2:39–52

Alonso JM, Stepanova AN (2004) The ethylene signaling pathway. Science 306:1513–1515. doi:10.1126/science.1104812

Amanchy R, Kalume DE, Iwahori A, Zhong J, Pandey A (2005) Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). J Proteome Res 4:1661–1671. doi:10.1021/Pr050134h

Anastasiadis PZ, Moon SY, Thoreson MA, Mariner DJ, Crawford HC, Zheng Y, Reynolds AB (2000) Inhibition of RhoA by p120 catenin. Nat Cell Biol 2:637–644. doi:10.1038/35023588

Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1:845–867

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A et al (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310. doi:10.1126/science.1072104

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. Nature 415:977–983. doi:10.1038/415977a

Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, Hatfull G, Hudson GS, Satchwell SC, Seguin C et al (1984) DNA-sequence and expression of the B95–8 Epstein–Barr virus genome. Nature 310:207–211

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science 320:938–941

Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li JX, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci USA 101:12130–12135. doi:10.1073/pnas.0404720101

Benschop JJ, Mohammed S, O'Flaherty M, Heck AJR, Slijper M, Menke FLH (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics 6:1198–1214. doi:10.1074/mcp.M600429-MCP200

Blander G, Guarente L (2004) The Sir2 family of protein deacetylases. Annu Rev Biochem 73:417–435. doi:10.1146/annurev.biochem.73.011303.073651

Bogdanov B, Smith RD (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. Mass Spectrom Rev 24:168–200. doi:10.1002/mas.20015

Bouras T, Fu M, Sauve AA, Wang F, Quong AA, Perkins ND, Hay RT, Gu W, Pestell RG (2005) SIRT1 deacetylation and

repression of p300 involves lysine residues 1020/1024 within the cell cycle regulatory domain 1. J Biol Chem 280:10264–10276. doi:10.1074/jbc.M408748200

Breitschopf K, Bengal E, Ziv T, Admon A, Ciechanover A (1998) A novel site for ubiquitination: the N-terminal residue, and not internal lysines of MyoD, is essential for conjugation and degradation of the protein. EMBO J 17:5964–5973. doi:10.1093/emboj/17.20.5964

Brooks CL, Gu W (2003) Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol 15:164–171

Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de Lichtenberg U, Rinner O et al (2007) A high-quality catalog of the Drosophila melanogaster proteome. Nature Biotechnology 25:576–583

Chait BT, Chait BT (2006) Chemistry. Mass spectrometry: bottom-up or top-down? Science 314:65–66. doi:10.1126/science.1133987

Chen ZJ (2005) Ubiquitin signalling in the NF-kappaB pathway. Nat Cell Biol 7:758–765. doi:10.1038/ncb0805-758

Chen Y, Hoehenwarter W, Weckwerth W (2010) Comparative analysis of phytohormone—responsive phosphoproteins in *Arabidopsis thaliana* using TiO_2 phosphopeptide enrichment and MAPA. Plant J

Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JEP, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci USA 104:2193–2198. doi:10.1073/pnas.0607084104

Chinchilla D, Bauer Z, Regenass M, Boller T, Felix G (2006) The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell 18:465–476. doi:10.1105/tpc.105.036574

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nurnberger T, Jones JDG, Felix G, Boller T (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448:497. doi:10.1038/Nature05999

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124:803–814. doi:10.1016/j.cell.2006.02.008

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325:834–840. doi:10.1126/science.1175371

Ciechanover A, Orian A, Schwartz AL (2000) Ubiquitin-mediated proteolysis: biological regulation via destruction. Bioessays 22:442–451

Cohen P (1999) The Croonian Lecture (1999). Identification of a protein kinase cascade of major importance in insulin signal transduction. Philos T Roy Soc B 354:485–495

Cohen P (2000) The regulation of protein function by multisite phosphorylation—a 25 year update. Trends Biochem Sci 25:596–601

Collins FS, Lander ES, Rogers J, Waterston RH, Conso IHGS (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945. doi:10.1038/Nature03001

Comon P (1994) Independent component analysis, a new concept. Signal Process 36:287–314

Conrath U, Pieterse CMJ, Mauch-Mani B (2002) Priming in plant-pathogen interactions. Trends Plant Sci 7:210–216

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372

Cubitt B, Oldstone C, Delatorre JC (1994) Sequence and genome organization of borna-disease virus. J Virol 68:1382–1396

de Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455:1251–1260

Devoto A, Muskett PR, Shirasu K (2003) Role of ubiquitination in the regulation of plant defence against pathogens. Curr Opin Plant Biol 6:307–311. doi:10.1016/S1369-5266(03)00060-8

Dimmeler S, Breitschopf K, Haendeler J, Zeiher AM (1999) Dephosphorylation targets Bcl-2 for ubiquitin-dependent degradation: a link between the apoptosome and the proteasome pathway. J Exp Med 189:1815–1822

Ellis C, Turner JG, Devoto A (2002) Protein complexes mediate signalling in plant responses to hormones, light, sucrose and pathogens. Plant Mol Biol 50:971–980

Fischle W, Wang Y, Allis CD (2003) Binary switches and modification cassettes in histone biology and beyond. Nature 425:475–479. doi:10.1038/nature02017

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Gallagher E, Gao M, Liu YC, Karin M (2006) Activation of the E3 ubiquitin ligase itch through a phosphorylation-induced conformational change. Proc Natl Acad Sci USA 103:1717–1722. doi:10.1073/pnas.0510664103

Garcia BA, Pesavento JJ, Mizzen CA, Kelleher NL (2007) Pervasive combinatorial modification of histone H3 in human cells. Nat Methods 4:487–489. doi:10.1038/Nmeth1052

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425:737–741. doi:10.1038/nature02046

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521. doi:10.1038/Nature02426

Gimona M (2006) Protein linguistics—a grammar for modular protein assembly? Nat Rev Mol Cell Biol 7:68–73. doi:10.1038/nrm1785

Glotzer M, Murray AW, Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. Nature 349:132–138

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M et al (1996) Life with 6000 genes. Science 274:546, 563–567

Graumann J, Hubner NC, Kim JB, Ko K, Moser M, Kumar C, Cox J, Scholer H, Mann M (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. Mol Cell Proteomics 7:672–683

Haas AL, Baboshina O, Williams B, Schwartz LM (1995) Coordinated induction of the ubiquitin conjugation pathway accompanies the developmentally programmed death of insect skeletal-muscle. J Biol Chem 270:9407–9412

Hardtke CS, Gohda K, Osterlund MT, Oyama T, Okada K, Deng XW (2000) HY5 stability and activity in Arabidopsis is regulated by phosphorylation in its COP1 binding domain. EMBO J 19:4997–5006

Hoehenwarter W, van Dongen JT, Wienkoop S, Steinfath M, Humme J, Erban A, Sulpice R, Regierer B, Kopka J, Geigenberger P et al (2008) A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. Proteomics 8:4214–4225

Huang F, Kirkpatrick D, Jiang X, Gygi S, Sorkin A (2006) Differential regulation of EGF receptor internalization and

degradation by multiubiquitination within the kinase domain. Mol Cell 21:737–748. doi:10.1016/j.molcel.2006.02.018
Hubbard MJ, Cohen P (1993) On target with a new mechanism for the regulation of protein phosphorylation. Trends Biochem Sci 18:172–177
Humphery-Smith I (2004) A human proteome project with a beginning and an end. Proteomics 4:2519–2521. doi:10.1002/pmic.200400866
Hunter T (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol Cell 28:730–738. doi:10.1016/j.molcel.2007.11.019
Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K (2000) Various abiotic stresses vapidly activate Arabidopsis MAP kinases ATMPK4 and ATMPK6. Plant J 24:655–665
Igawa T, Fujiwara M, Takahashi H, Sawasaki T, Endo Y, Seki M, Shinozaki K, Fukao Y, Yanagawa Y (2009) Isolation and identification of ubiquitin-related proteins from Arabidopsis seedlings. J Exp Bot 60:3067–3073. doi:10.1093/jxb/erp134
Inze D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet 40:77–105. doi:10.1146/annurev.genet.40.110405.090431
Jiang X, Sun Y, Chen S, Roy K, Price BD (2006) The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. J Biol Chem 281:15741–15746. doi:10.1074/jbc.M513172200
Johnson JM, Castle J, Garrett-Engele P, Kan ZY, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302:2141–2144
Jongeneel CV, Delorenzi M, Iseli C, Zhou DX, Haudenschild CD, Khrebtukova I, Kuznetsov D, Stevenson BJ, Strausberg RL, Simpson AJG et al (2005) An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Res 15:1007–1014
Jungblut PR, Holzhutter HG, Apweiler R, Schluter H (2008) The speciation of the proteome. Chem Cent J 2:16
Jungblut PR, Schiele F, Zimny-Arndt U, Ackermann R, Schmid M, Lange S, Stein R, Pleissner KP (2010) Helicobacter pylori proteomics by 2-DE/MS, 1-DE-LC/MS and functional data mining. Proteomics 10:182–193. doi:10.1002/pmic.200900361
Karin M (1999) How NF-kappaB is activated: the role of the IkappaB kinase (IKK) complex. Oncogene 18:6867–6874. doi:10.1038/sj.onc.1203219
Karin M, Liu Z, Zandi E (1997) AP-1 function and regulation. Curr Opin Cell Biol 9:240–246
Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY et al (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815
Kempa S, Rozhon W, Samaj J, Erban A, Baluska F, Becker T, Haselmayer J, Schleiff E, Kopka J, Hirt H et al (2007) A plastid-localized glycogen synthase kinase 3 modulates stress tolerance and carbohydrate metabolism. Plant J 49:1076–1090. doi:10.1111/j.1365-313X.2006.03025.x
Kim YO, Kim JS, Kang H (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. Plant J 42:890–900. doi:10.1111/j.1365-313X.2005.02420.x
Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, Cheng T, Kho Y, Xiao H, Xiao L et al (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell 23:607–618. doi:10.1016/j.molcel.2006.06.026
Kitagawa M, Hatakeyama S, Shirane M, Matsumoto M, Ishida N, Hattori K, Nakamichi I, Kikuchi A, Nakayama K (1999) An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of beta-catenin. EMBO J 18:2401–2410. doi:10.1093/emboj/18.9.2401
Klose J, Kobalz U (1995) 2-Dimensional electrophoresis of proteins—an updated protocol and implications for a functional-analysis of the genome. Electrophoresis 16:1034–1059
Koepp DM, Harper JW, Elledge SJ (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle. Cell 97:431–434
Kotlyarov A, Neininger A, Schubert C, Eckert R, Birchmeier C, Volk HD, Gaestel M (1999) MAPKAP kinase 2 is essential for LPS-induced TNF-alpha biosynthesis. Nat Cell Biol 1:94–97
Kovtun Y, Chiu WL, Zeng W, Sheen J (1998) Suppression of auxin signal transduction by a MAPK cascade in higher plants. Nature 395:716–720. doi:10.1038/27240
Kunkel BN, Brooks DM (2002) Cross talk between signaling pathways in pathogen defense. Curr Opin Plant Biol 5:325–331
Larsen MR, Larsen PM, Fey SJ, Roepstorff P (2001) Characterization of differently processed forms of enolase 2 from *Saccharomyces cerevisiae* by two-dimensional gel electrophoresis and mass spectrometry. Electrophoresis 22:566–575
Latham JA, Dent SY (2007) Cross-regulation of histone modifications. Nat Struct Mol Biol 14:1017–1024. doi:10.1038/nsmb1307
Lizcano JM, Morrice N, Cohen P (2000) Regulation of BAD by cAMP-dependent protein kinase is mediated via phosphorylation of a novel site, Ser155. Biochem J 349:547–557
Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25:117–124. doi:10.1038/Nbt1270
Ludwig AA, Saitoh H, Felix G, Freymark G, Miersch O, Wasternack C, Boller T, Jones JD, Romeis T (2005) Ethylene-mediated cross-talk between calcium-dependent protein kinase and MAPK signaling controls stress responses in plants. Proc Natl Acad Sci USA 102:10736–10741. doi:10.1073/pnas.0502954102
MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A et al (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. Proc Natl Acad Sci USA 99:7900–7905
Mahajan R, Delphin C, Guan T, Gerace L, Melchior F (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell 88:97–107
Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics 6:601–610. doi:10.1074/mcp.M600408-MCP200
Margottin F, Bour SP, Durand H, Selig L, Benichou S, Richard V, Thomas D, Strebel K, Benarous R (1998) A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. Mol Cell 1:565–574
Marshall CJ (1994) MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. Curr Opin Genet Dev 4:82–89
Meskiene I, Bogre L, Glaser W, Balog J, Brandstotter M, Zwerger K, Ammerer G, Hirt H (1998) MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. Proc Natl Acad Sci USA 95:1938–1943
Meskiene I, Baudouin E, Schweighofer A, Liwosz A, Jonak C, Rodriguez PL, Jelinek H, Hirt H (2003) Stress-induced protein phosphatase 2C is a negative regulator of a mitogen-activated protein kinase. J Biol Chem 278:18945–18952. doi:10.1074/jbc.M300878200
Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE et al (2002) Annotation of the *Drosophila melanogaster*

euchromatic genome: a systematic review. Genome Biol 3:RESEARCH0083
Mukherjee S, Keitany G, Li Y, Wang Y, Ball HL, Goldsmith EJ, Orth K (2006) Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation. Science 312:1211–1214. doi:10.1126/science.1126867
Murphy LO, Smith S, Chen RH, Fingar DC, Blenis J (2002) Molecular interpretation of ERK signal duration by immediate early gene products. Nat Cell Biol 4:556–564. doi:10.1038/ncb822ncb822
Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA et al (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204. doi:8395
Nakagami H, Pitzschke A, Hirt H (2005) Emerging MAP kinase pathways in plant stress signalling. Trends Plant Sci 10:339–346. doi:10.1016/j.tplants.2005.05.009
Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312:436–439. doi:10.1126/science.1126088
Nuhse TS, Peck SC, Hirt H, Boller T (2000) Microbial elicitors induce activation and dual phosphorylation of the *Arabidopsis thaliana* MAPK 6. J Biol Chem 275:7521–7526
Nuhse TS, Stensballe A, Jensen ON, Peck SC (2004) Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. Plant Cell 16:2394–2405. doi:10.1105/tpc.104.023150
Nuhse TS, Bottrill AR, Jones AM, Peck SC (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. Plant J 51:931–940. doi:10.1111/j.1365-313X.2007.03192.x
O'Connell MJ, Walworth NC, Carr AM (2000) The G2-phase DNA-damage checkpoint. Trends Cell Biol 10:296–303
Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4:1487–1502
Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635–648
Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA et al (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3:ra3. doi:10.1126/scisignal.2000475
Ono M, Shitashige M, Honda K, Isobe T, Kuwabara H, Matsuzuki H, Hirohashi S, Yamada T (2006) Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. Mol Cell Proteomics 5:1338–1347
Palmgren MG (2001) PLANT PLASMA MEMBRANE H+-ATPases: Powerhouses for Nutrient Uptake. Annu Rev Plant Physiol Plant Mol Biol 52:817–845. doi:10.1146/annurev.arplant.52.1.817
Pawson T, Nash P (2000) Protein-protein interactions define specificity in signal transduction. Genes Dev 14:1027–1047
Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572
Peng JM, Schwartz D, Elias JE, Thoreen CC, Cheng DM, Marsischky G, Roelofs J, Finley D, Gygi SP (2003) A proteomics approach to understanding protein ubiquitination. Nat Biotechnol 21:921–926. doi:10.1038/Nbt849
Polge C, Thomas M (2007) SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control? Trends Plant Sci 12:20–28. doi:10.1016/j.tplants.2006.11.005
Posern G, Miralles F, Guettler S, Treisman R (2004) Mutant actins that stabilise F-actin use distinct mechanisms to activate the SRF coactivator MAL. EMBO J 23:3973–3983. doi:10.1038/sj.emboj.7600404
Rock KL, Goldberg AL (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. Annu Rev Immunol 17:739–779
Sabo A, Lusic M, Cereseto A, Giacca M (2008) Acetylation of conserved lysines in the catalytic core of cyclin-dependent kinase 9 inhibits kinase activity and regulates transcription. Mol Cell Biol 28:2201–2212. doi:10.1128/MCB.01557-07
Scheler C, Muller EC, Stahl J, Muller-Werdan U, Salnikow J, Jungblut P (1997) Identification and characterization of heat shock protein 27 protein species in human myocardial two-dimensional electrophoresis patterns. Electrophoresis 18:2823–2831
Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. Cell 103:211–225
Schmidt A, Gehlenborg N, Bodenmiller B, Mueller LN, Campbell D, Mueller M, Aebersold R, Domon B (2008) An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. Mol Cell Proteomics 7:2138–2150
Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101:671–684
Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 20:2447–2454
Schulze WX (2010) Proteomics approaches to understand protein phosphorylation in pathway modulation. Curr Opin Plant Biol. doi:10.1016/j.pbi.2009.12.008
Schulze WX, Usadel B (2010) Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol 61:491–516. doi:10.1146/annurev-arplant-042809-112132
Schweighofer A, Kazanaviciute V, Scheikl E, Teige M, Doczi R, Hirt H, Schwanninger M, Kant M, Schuurink R, Mauch F et al (2007) The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in Arabidopsis. Plant Cell 19:2213–2224. doi:10.1105/tpc.106.049585
Seet BT, Dikic I, Zhou MM, Pawson T (2006) Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol 7:473–483. doi:10.1038/nrm1960
Seo J, Lee KJ (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. J Biochem Mol Biol 37:35–44
Shteinberg M, Protopopov Y, Listovsky T, Brandeis M, Hershko A (1999) Phosphorylation of the cyclosome is required for its stimulation by Fizzy/cdc20. Biochem Biophys Res Commun 260:193–198. doi:10.1006/bbrc.1999.0884
Sims RJ, Reinberg D (2008) Is there a code embedded in proteins that is based on post-translational modifications? Nat Rev Mol Cell Bio 9:815–820. doi:10.1038/Nrm2502
Siuti N, Kelleher NL (2007) Decoding protein modifications using top-down mass spectrometry. Nat Methods 4:817–821. doi:10.1038/nmeth1097
Smallwood M, Bowles DJ (2002) Plants in a cold climate. Philos Trans R Soc Lond B Biol Sci 357:831–847. doi:10.1098/rstb.2002.1073
Sowa ME, Bennett EJ, Gygi SP, Harper JW (2009) Defining the human deubiquitinating enzyme interaction landscape. Cell 138:389–403. doi:10.1016/j.cell.2009.04.042
Steuer R, Kurths J, Fiehn O, Weckwerth W (2003) Observing and interpreting correlations in metabolomic networks. Bioinformatics 19:1019–1026

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol Syst Biol 4:193. doi:10.1038/Msb.2008.32

Sun Y, Jiang X, Chen S, Fernandes N, Price BD (2005) A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. Proc Natl Acad Sci USA 102:13182–13187. doi:10.1073/pnas.0504211102

Takahashi Y, Soyano T, Sasabe M, Machida Y (2004) A MAP kinase cascade that controls plant cytokinesis. J Biochem 136:127–132. doi:10.1093/Jb/Mvh118

Takahashi F, Yoshida R, Ichimura K, Mizoguchi T, Seo S, Yonezawa M, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K (2007) The mitogen-activated protein kinase cascade MKK3-MPK6 is an important part of the jasmonate signal transduction pathway in Arabidopsis. Plant Cell 19:805–818. doi:10.1105/tpc.106.046581

Thelemann A, Petti F, Griffin G, Iwata K, Hunt T, Settinari T, Fenyo D, Gibson N, Haley JD (2005) Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. Mol Cell Proteomics 4:356–376. doi:10.1074/mcp.M400118-MCP200

Thompson PR, Wang D, Wang L, Fulco M, Pediconi N, Zhang D, An W, Ge Q, Roeder RG, Wong J et al (2004) Regulation of the p300 HAT domain via a novel activation loop. Nat Struct Mol Biol 11:308–315. doi:10.1038/nsmb740

Tress ML, Bodenmiller B, Aebersold R, Valencia A (2008) Proteomics studies confirm the presence of alternative protein isoforms on a large scale. Genome Biol 9:R162

Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA (2005) Identification of post-translational modifications by blind search of mass spectra. Nat Biotechnol 23:1562–1567

Usaite R, Wohlschlegel J, Venable JD, Park SK, Nielsen J, Olsson L, Yates JR (2008) Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *Saccharomyces cerevisiae* strains: the comparison of two quantitative methods. J Proteome Res 7:266–275. doi:10.1021/Pr700580m

van Bentem SD, Hirt H (2007) Using phosphoproteomics to reveal signalling dynamics in plants. Trends Plant Sci 12:404–411. doi:10.1016/j.tplants.2007.08.007

Ventii KH, Wilkinson KD (2008) Protein partners of deubiquitinating enzymes. Biochem J 414:161–175. doi:10.1042/Bj20080798

Ventura JJ, Kennedy NJ, Lamb JA, Flavell RA, Davis RJ (2003) c-Jun NH(2)-terminal kinase is essential for the regulation of AP-1 by tumor necrosis factor. Mol Cell Biol 23:2871–2882

Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci USA 104:1488–1493. doi:10.1073/pnas.0609836104

Wang D, Baldwin AS Jr (1998) Activation of nuclear factor-kappaB-dependent transcription by tumor necrosis factor-alpha is mediated through phosphorylation of RelA/p65 on serine 529. J Biol Chem 273:29411–29416

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562. doi:10.1038/Nature01262

Wienkoop S, Larrainzar E, Niemann M, Gonzalez EM, Lehmann U, Weckwerth W (2006) Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics. J Sep Sci 29:2793–2801

Wienkoop S, Morgenthal K, Wolschin F, Scholz M, Selbig J, Weckwerth W (2008) Integration of metabolomic and proteomic phenotypes: analysis of data covariance dissects starch and RFO metabolism from low and high temperature compensation response in *Arabidopsis thaliana*. Mol Cell Proteomics 7:1725–1736. doi:10.1074/mcp.M700273-MCP200

Yang XJ (2005) Multisite protein modification and intramolecular signaling. Oncogene 24:1653–1662. doi:10.1038/sj.onc.1208173

Yang XJ, Seto E (2008) Lysine acetylation: codified crosstalk with other posttranslational modifications. Mol Cell 31:449–461. doi:10.1016/j.molcel.2008.07.002

Yaron A, Hatzubai A, Davis M, Lavon I, Amit S, Manning AM, Andersen JS, Mann M, Mercurio F, Ben-Neriah Y (1998) Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. Nature 396:590–594. doi:10.1038/25159

Zhang X, Yuan Z, Zhang Y, Yong S, Salas-Burgos A, Koomen J, Olashaw N, Parsons JT, Yang XJ, Dent SR et al (2007) HDAC6 modulates cell motility by altering the acetylation level of cortactin. Mol Cell 27:197–213. doi:10.1016/j.molcel.2007.05.033

Zhang J, Shi X, Li Y, Kim BJ, Jia J, Huang Z, Yang T, Fu X, Jung SY, Wang Y et al (2008) Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol Cell 31:143–151. doi:10.1016/j.molcel.2008.06.006

Zhang G, Hubalewska M, Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol 16:274–280. doi:10.1038/Nsmb.1554

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, Boller T (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. Nature 428:764–767. doi:10.1038/Nature02485

2.2 The different proteomes of *Chlamydomonas reinhardtii*

One of the main features of Systems Biology is reflected in the development of the "omics"-approaches, which aim to identify the complete set of biomolecules of a certain kind in an organism at a certain time point. Thus, high-throughput measuring devices are needed — and just as necessary are computer-based strategies that can analyze the correspondently huge amount of data. For a classical bottom-up LC-MS shotgun proteomics approach, a vital step is the "protein inference": the association of measured ions to peptide sequences belonging to certain proteins, mostly from an organism-specific protein database, with the help of a search algorithm.

Since the development of proteomics, great advances have been made within the development of algorithms (as BLAST) and search engines (SEQUEST, MASCOT or the open source OMSSA), mostly regarding parameters about the quality of the identification (Xcorr, FDR). Nevertheless, there hasn't been as much work done on the other key element of this procedure: the databases — from the registered 35.000 cultivated plant species, there are only 37 fully sequenced and functionally annotated plant genomes ([48]) — for untargeted studies on the rest of the species, protein databases must be translated from e.g. EST, PUT assembly and contig libraries; another alternative is to perform homology studies through BLAST with databases from other organisms (mainly the NCBI-Databases from *Arabidopsis thaliana*).

Through a comparative study among different databases/annotations and a graphical display of the corresponding results, we want to remark the need for a standardization of databases and performance evaluation, considering that the results of a same dataset being analyzed with different databases or even different releases of the same database can differ considerably, thus affecting not only protein identification, but also protein quantification and the interpretation of protein fuctions in a biological context.

2.2.1 Declaration of authorship

The results of this chapter are presented in the form of a manuscript published in the journal "Journal of Proteomics". For this work, I contributed to the experimental design and carried out the experiments that led to the datasets used for the elaboration of this manuscript.

2.2.2 Published manuscript

JOURNAL OF PROTEOMICS 75 (2012) 5883–5887

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Technical note

The different proteomes of *Chlamydomonas reinhardtii*

*Luis Valledor, Luis Recuenco-Munoz, Volker Egelhofer, Stefanie Wienkoop, Wolfram Weckwerth**

Department of Molecular Systems Biology, University of Vienna, Althanstrasse 14, 1090, Vienna, Austria

ARTICLE INFO

Article history:
Received 23 April 2012
Accepted 30 July 2012
Available online 7 August 2012

Keywords:
Proteogenomics
Genome annotation
Functional annotation
Systems biology
Plant systems biology
PROMEX

ABSTRACT

Protein identification and proteome mapping mostly rely on the combination of tandem mass spectrometry and sequence database searching. Despite constant improvements achieved in instrumentation, search algorithms, and genome annotations, little effort has been invested in estimating the impact of different genome annotation releases on the final results of a proteome study. We have used a large dataset of mass spectra obtained using an Orbitrap LTQ XL instrument, covering different growth situations of the model species *Chlamydomonas reinhardtii*. More than one million spectra were analyzed employing the SEQUEST algorithm and four different databases corresponding to the major *Chlamydomonas* genome assemblies. In total more than 3000 proteins and about 11,000 peptides were identified. 238 proteins were exclusively detected in assembly 3.0 in contrast to 1222 missing proteins only detectable in other databases. The comparison of the results demonstrates that the database selection affects not only the number of identified proteins but also label free quantitation and the biological interpretation of the results. Lists of protein accessions exclusively assigned to individual *C. reinhardtii* genome assemblies and annotations are provided as a resource for proteogenomic studies.

Nowadays the identification of peptides and the mapping of the proteomes rely on the combination of tandem mass spectrometry and sequence database searching. In a typical proteomics pipeline the identification of proteins is based on peptide-centric proteomics, that identifies peptides rather than proteins [1]. Peptides are identified by matching the acquired MS/MS spectra against a protein sequence database, while proteins are inferred after peptide identification. A multitude of search engines are available for identifying the different peptides in the sample, and new tools are constantly designed for improving the quality of the analysis in terms of increasing the positive identifications while reducing the number of false positives [2]. However, in spite of these advances, the importance of the employed database has not been specifically addressed in most cases.

The contents of the database are paramount for protein identification. The first step of protein identification is the *in silico* digestion of all of the sequences of the database, generating theoretical MS/MS spectra for every possible candidate. The generated spectra are then compared to each experimental spectrum and a score is calculated for each peptide [3]. In a second step, the set of all identified peptides is compared to the undigested protein database and used to infer which proteins may have been present [4]. Consequently, the annotated genome database is the foundation of the whole process including peptide identification and assembly of the corresponding proteins.

Large sequence databases containing sequences of different species, like NCBI or Uniprot, are classically used for protein

* *Corresponding author at:* Department of Molecular Systems Biology, Faculty of Life Sciences, University of Vienna, Althanstr. 14, 1090, Vienna, Austria. Tel.: +43 1 4277 577 00, +43 664 60277 577 00 (mobile); fax: +43 1 4277 9 577.
E-mail addresses: luis.valledor@univie.ac.at (L. Valledor), wolfram.weckwerth@univie.ac.at (W. Weckwerth).

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2012.07.045

identification. But the annotation of these databases is not always good, and their large sizes make the searches slow, also increasing the possibility of misidentifications. On the other hand, advances in next generation sequencing (NGS) technologies have increased the number of organisms that have been fully sequenced [5], making it possible to utilize organism-specific databases. The assembly and functional annotation of these databases is constantly changed and improved since the genomes of most species are re-sequenced and reassembled to increase the reliability of the final sequence [6]. This fact, together with the advances in the algorithms employed in the bioinformatics-processing pipeline for predicting new genes and proteins, leads to a regular release of new assemblies and/or annotations. The different releases of a genome assembly and annotation can include more sequences and functional annotation, but also the elimination or changes of some genes. Obviously these changes will be reflected in different protein numbers and sequence modifications, which will change the *in silico* digests employed by the search engines. Here we addressed the question, how do these changes affect the identification, quantification, and the functional characterization of the proteome?

We demonstrate that different assemblies and annotations of the *Chlamydomonas reinhardtii* genome lead to dramatically different results. Differences are not only limited to the number of detected peptides, but also found in the variable identification of proteins and coverage of functional units in the metabolic network.

For this study 280 Orbitrap-LTQ (Thermo Scientific, USA) runs (>130 Gb of raw data, 1087447 spectra), available at MOSYS (University of Vienna) [7–9] were selected, aiming to cover different *C. reinhardtii* growth situations and cellular fractions. The objective of using this high number of experimental situations and spectra was to not over-represent some proteins or introduce a sample dependent bias. We have employed four different databases covering two versions of the genome assembly: Chlamydomonas 3.0 (Chlr_3) and Chlamydomonas 4.0. Three annotations corresponding to Chlamydomonas 4.0 were considered: Augustus v.5 (Aug_5), Augustus v.10.1 (JGI_153) and Augustus v10.2 (JGI_169). Files with raw spectra were processed using the SEQUEST algorithm available in Proteome Discoverer 1.3 (Thermo Scientific, USA) and the four described databases. The following settings were used: precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.8 Da. Only charge states +2 or greater were used. Identification confidence was set to a 5% FDR and the variable modifications were set to: acetylation of N-terminus, oxidation of methionine and carbamidomethyl cysteine formation. No fixed modifications were set. A maximum of two missed cleavages were set for all searches. We have tested independently two different thresholds for considering a confident protein identification: i) having at least one unique peptide (peptide that only appears once in the entire database) with a X-Correlation value 0.5 greater than the charge state (i.e. 2.5 for peptides with charge +2); ii) having at least two peptides (one of them being unique) with a X-Correlation value 0.25 greater than the charge state. In our hands, the application of these thresholds leads to a higher peptide assignment to proteins, reducing the coefficient of variation of replicates, and, in consequence, improving final quantitation. Similar thresholds have been previously proposed for improved protein identification and quantitation [10,11].

The four different files obtained after processing each batch of LC/MS analyses, one per database, were merged with Proteome Discoverer© (Thermo-Fisher-Scientific) into a single file. This file includes the different identifications obtained with the different databases and shows most of the overlapping accessions between the different databases. These lists included redundancies, corresponding to peptides and proteins present in more than one intermediate file. For example some photosynthesis-related proteins were present in all of the batches. In a second step these redundancies were removed, keeping only the best-scoring peptide entry for each protein accession. This procedure was done twice, considering independently the two different thresholds described above (Supplementary Fig. 1). The next step of the analysis was to compare the different lists of peptides and proteins obtained with each database. Since each database has different accession names we have used the capabilities of Proteome Discoverer© (Thermo-Fisher-Scientific) for comparing sequences, the "Algal Functional Annotation Tool" [12] for comparing annotations, and BLAST algorithm for establishing the correspondences between databases when the other approaches failed.

Our first aim was to make an estimation of the influence of the different genome assemblies and annotations over the detection of peptides employing a 5% FDR and X-Correlation value greater than charge state +0.25 as thresholds (Fig. 1a). A total of 10,957 different peptides were identified employing the four databases (List of identified peptides is available in Supplementary Table 1). Despite most of the peptides (8485) being shared by all of the genome annotations, two different groups corresponding to each major genome assembly can be established: Assembly release 3.0 (Chlr_3) which detected the lowest number of peptides (8760) and release 4.0 (Aug_5-Augustus 5, JGI_153-Augustus 10.1,JGI_169-Augustus 10.2) that detected 10,680±62 depending on the version of the annotation. These differences can be explained by the fact that the different assemblies do not share the same genome sequence. Release 4.0 is larger and more accurate, which is reflected in the high number of identified peptides, but it lacks some sequences, with 150 peptides that only can be found employing the release 3.0. Even the differences between the three annotations based on the 4.0 assembly are not trivial. In fact, despite sharing the same genome sequence, 138 peptides are differently identified between Augustus 5 and JGI_169 (Augustus 10.2) annotations.

Theoretically the increased number of peptides between the releases can be explained by the fact that there are big technical gaps between them, with improved sequence reads and computer algorithms. Our data agree with this hypothesis, since the newest release revealed the highest number of detected peptides. However, the different annotations of assembly 4.0 showed strong unexpected differences. The most probable answer is that the differences between transcript prediction algorithms employed for the release of the different annotations may result in some incorrectly predicted, truncated or removed proteins. The gene model Cre09.g410550, N5-Phosphoribosyl-ATP transferase, has different protein sequences in the different databases. This protein has been truncated (more than 80 amino acids) in the Chlamy 4.0 release of the genome (Supplementary

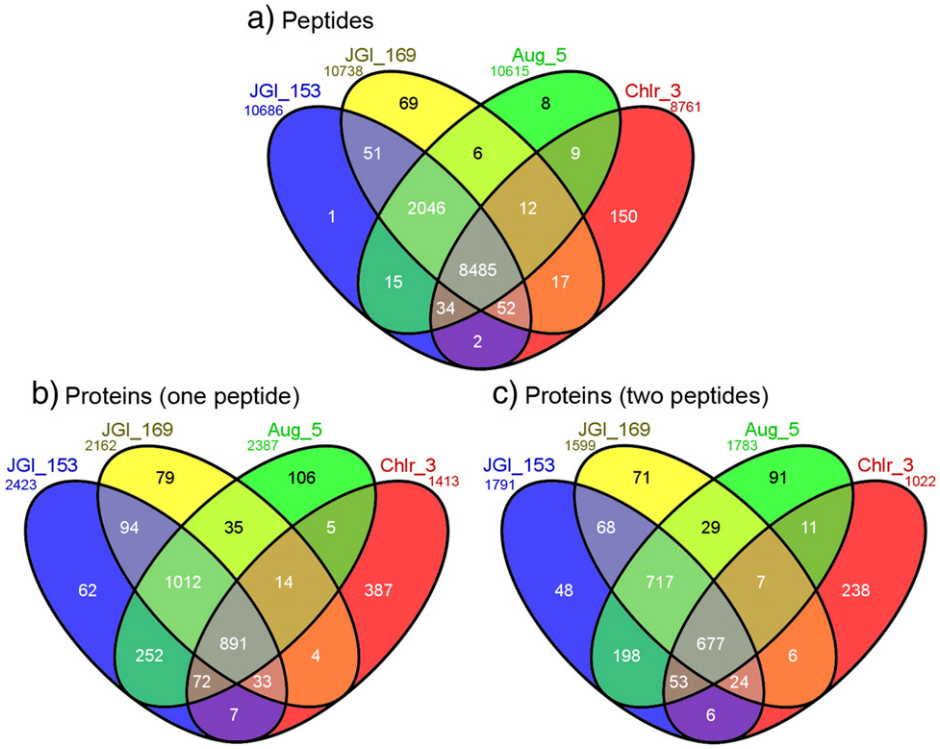

Fig. 1 – Venn diagrams of non-redundant peptides and proteins identified with the different databases and thresholds. Chlr_3: Assembly 3, Annotation 3; Aug_5: Assembly 4.0, Annotation Augustus 5; JGI_153: Assembly 4.0, Annotation Augustus 10.1; JGI_169: Assembly 4.0, Annotation: Augustus 10.2. Non-redundant peptides identified with each database (a). Proteins identified with at least 1 (b) or 2 (c) peptides. Diagrams were plot employing the on-line tool Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Fig. 2a). This problem can also be observed in closely related annotations, i.e. the gene Cre09.g405500 (Supplementary 2b). Considering only the closest annotations JGI_153 (Augustus 10.1) and JGI_169 (Augustus 10.2) we have detected 3 extra amino acids that validate the "old" JGI153 sequence rather than the updated JGI_169 (Supplementary Fig. 2b). It must be also noticed that changes in the database sequence may change the pool of peptides obtained after *in silico* digestion and their probabilistic distribution, altering the X-Correlation score [13]. To discard this technical effect we have studied the distribution of the X-Correlations between the identified peptides and databases (Supplementary Fig. 3a), showing that Chlr_3 has slightly lower values than assembly 4 in which no significant differences can be established between annotations. The peptides that are exclusive from the different databases have the same X-Correlation distribution as the whole population (Supplementary Fig. 3b), with high confidence values (Supplementary Fig. 3c) indicating that the missed detection of peptides is related to the protein database.

At protein level, as in the case of the peptides, the databases based on assembly 4.0 detected more proteins than those based on 3.0 (Fig. 1b, c) (Supplementary Table 2). The differences in the number of proteins that can be identified exclusively in each genome release are remarkable: 238 proteins were exclusively detected using assembly 3.0 while 1222 proteins were exclusively found in assembly 4.0. These ratios are similar to those obtained for peptides. Supplementary Table 4 summarizes Fig. 1b, c showing the protein accessions that were found only in one database.

Some remarkable differences between identified peptides and proteins can be observed. For example, considering the database JGI_153 (Augustus 10.1) we detected only 1 peptide ("MSTSEFETVVLTPSRLR") that is not present in the other databases (Fig. 1a). But this number increases to 48 in the case of the proteins (2 peptides per protein) (Fig. 1b, c). The appearance of 48 exclusive proteins using two peptide threshold, means that at least 48 unique peptides are required for its identification, so this is the minimum number of spectra not assigned, or assigned to other proteins when using other database releases. This can only be explained by the principles of the search algorithm. The basis of the identification is usage of all peptides for explaining the minimum number of proteins [1]. Consequently, depending on the protein sequences present in the different databases some spectra will be assigned to different peptides-protein accessions, having a high impact over the final result.

The assignment of peptides-spectra to different proteins is also important for label free quantitation methods, since both precursor ion areas or spectral count methods are based on the assignment of the discriminatory peptides to individual proteins [14,15]. Table 1 shows some protein accessions for which the peptide to spectra match (PSM) changes depending on the employed database.

Table 1 – Impact of the different tested databases on spectral counting (∑PSM) employed for label-free quantitation. PSM value was directly obtained from Proteome Discoverer results file when using 2 peptides per protein threshold. Five samples corresponding to cold stress and nitrogen depletion datasets were pooled to get this table.

Accession	ID	∑PSM Chlr3	∑PSM Aug5	∑PSM Aug10.1	∑PSM Aug10.2
Cre09.g405500		64	64	64	61
Cre07.g344400	3-Phosphoglycerate dehydrogenase	199	250	250	250
Cre28.g776100	Carbamoyl phosphate synthase	178	232	232	232
Cre18.g749750	PSI LHCA3	166	167	167	167
Cre10.g436550	Low-CO2 inducible protein	95	141	141	141
Cre11.g477950		17	184	184	184
Cre02.g076250	Chloroplast elongation factor G	111	110	111	111
Cre06.g263250		17	41	42	42
Cre01.g021250	Argininosuccinate lyase (ARG7)	18	35	29	35
Cre10.g439550		5	17	18	18
Cre12.g496000	Peptidyl-prolyl cis-trans isomerase, cyclophilin-type	43	43	43	8
Cre03.g193750	Glycine cleavage system, (GCST)	45	43	43	43
Cre13.g595250	Ribosomal protein L21	11	11	11	4

Furthermore there are notable differences between the different annotations based on assembly 4.0. Focusing on the more conservative threshold of 2 peptides per protein our data shows that only 1394 identifications, about 80%, are shared by all the different annotations of assembly 4.0. Moreover there are 48, 71 and 91 proteins that can only be detected employing release Augustus 10.1(JGI_153), 10.2 (JGI_169) or Aug_5, respectively. These differences dramatically change the biological interpretation of the results, since components of these pathways are differentially identified (Supplementary Table 3).

Comparing the differences between databases at this level is complex, and should be done case by case. In Fig. 2 an overview of these results is shown. There are significant differences between the numbers of proteins assigned to each MapMan category depending on the database. We have employed MapMan categories for representation since they are less complex and consequently easier to visualize than KEGG [16] or Gene Onthology [17]. However, KEGG classification showed a higher number of differences than MapMan as 224 categories were analyzed, as detailed in Supplementary Table 3.

The discrepancies between databases, i.e. the loss of proteins between releases or truncated sequences, highlight the need for new algorithms of gene prediction and implementation strategies for proteomics data. In this sense the integration of proteomic technologies into the annotation process will lead to improved accuracy and experimental validation of the *in silico* gene models [7,18–20].

In conclusion the comparison of the performance of the different databases demonstrates that the database selection affects not only the protein identification, but also its quantification and interpretation in a biological context. This fact increases the importance of the database design in order to maximize the number of identifications while reducing the artifacts, and also hinder the direct comparison of datasets

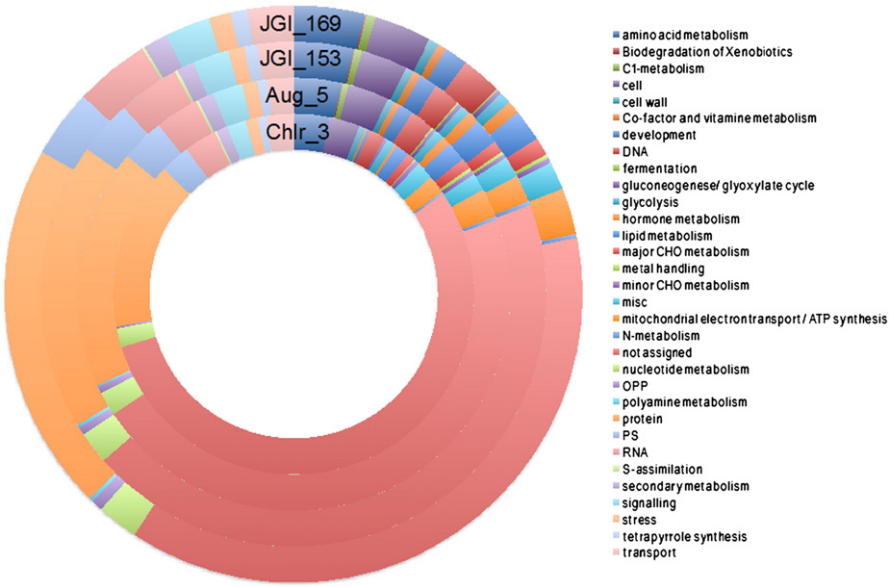

Fig. 2 – Doughnut chart representing the identified proteins (at least two peptides per protein) according to level 1 MapMan categories employing the indicated *Chlamydomonas* databases. Proteins were classified employing Algal Functional Annotation Tool [12].

analyzed with different databases and even releases of the same assembly. Further work is in progress towards the standardization and evaluation of the performance of the databases, since they play a significant role as essential as the identification algorithms.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2012.07.045.

Acknowledgments

This work and LV fellowship were supported by a Marie Curie IEF grant (European Union).

REFERENCES

[1] Cottrell JS. Protein identification using MS/MS data. J Proteomics 2011;74:1842-51.
[2] Sun CS, Markey MK. Recent advances in computational analysis of mass spectrometry for proteomic profiling. J Mass Spectrom 2011;46:443-56.
[3] Sadygov RG, Cociorva D, Yates JR. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat Methods 2004;1:195-202.
[4] Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data. Mol Cell Proteomics 2005;4:1419-40.
[5] Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods 2008;5:16-8.
[6] Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res January 2012;40(D1):D571-9.
[7] May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J, et al. Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. Genetics 2008;179:157-66.
[8] Hummel J, Niemann M, Wienkoop S, Schulze W, Steinhauser D, Selbig J, et al. ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. BMC Bioinformatics 2007;8:216.
[9] Wienkoop S, Wei, May P, Kempa S, Irgang S, Recuenco-Munoz L, et al. Targeted proteomics for *Chlamydomonas reinhardtii* combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. Mol Biosyst 2010;6:1018-31.
[10] Zhou J-Y, Schepmoes AA, Zhang X, Moore RJ, Monroe ME, Lee JH, et al. Improved LC-MS/MS spectral counting statistics by recovering low-scoring spectra matched to confidently identified peptide sequences. J Proteome Res 2010;9:5698-704.
[11] Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2002;2:43-50.
[12] Lopez D, Casero D, Cokus S, Merchant S, Pellegrini M. Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. BMC Bioinformatics 2011;12:282.
[13] Eng JK, Mccormack AL, Yates JR. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J Am Soc Mass Spectrom 1994;5:976-89.
[14] Elliott MH, Smith DS, Parker CE, Borchers C. Current trends in quantitative proteomics. J Mass Spectrom 2009;44:1637-60.
[15] Hoehenwarter W, Wienkoop S. Spectral counting robust on high mass accuracy mass spectrometers. Rapid Commun Mass Spectrom 2010;24:3609-14.
[16] 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:D109-14.
[17] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000;25:25-9.
[18] Weckwerth W. Green systems biology, from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. J Proteomics 2011;75:284-305.
[19] Specht M, Stanke M, Terashima M, Naumann-Busch B, Janßen I, Höhner R, et al. Concerted action of the new Genomic Peptide Finder and AUGUSTUS allows for automated proteogenomic annotation of the *Chlamydomonas reinhardtii* genome. Proteomics 2011;11:1814-23.
[20] 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;6:6207-20.

2.3 System Level Network Analysis of Nitrogen Starvation and Recovery in *Chlamydomonas reinhardtii* reveals potential new targets for increased lipid accumulation

Nitrogen starvation has been widely studied on *Chlamydomonas reinhardtii*, ranging from nutritional regulation of its sexuality in the 1950s to the role of Nitrogen in flagellar coordination. Nevertheless, none of these findings might be as economically relevant as its increased lipid accumulation -mostly TAGs stored in lipid droplets / oil bodies (LBs) — during N starvation. This phenomenon has been very well documented in the latest years. Chlamydomonas has also been widely investigated for hydrogen photoproduction (mostly under sulphur deprivation**1, 4**), and its capability to photosynthesis on both autotrophic and heteroautotrophic conditions enhances its appeal as a biomass producer (especially considering how easy it is to reproduce this conditions in Chlamydomonas cultures e.g. adding Acetate as a carbon source to the cultivation medium) Furthermore, having *Chlamydomonas reinhardtii* a fully sequenced genome, many proteomics and metabolomics protocols and databases, pathway annotations, and a wide range of molecular biology and transcriptomics tools, the possibilites of further improvement of the lipidic production via Bioengineering makes this model algae one of the most appropiate organisms for further investigation in the field of microalgae-related biofuel production.

There have already been many studies about TAG-Accumulation in *Chlamydomonas reinhardtii* ([273], [327], [35]), more recently also with starch mutants [345], [165]. However, even if some of these studies already integrate transcript, metabolites and even proteomic data, all of them deal only with the immediate response to N-depletion. In this publication, the CR stress reaction has been followed over a longer period of time — 4 days including N-repletion after 72 hours, thus following the recovery of the algae for a complete day. To gain a more comprehensive insight into the CR adaptation to N-starvation and its recovery at a cellular level, Metabolomics and Proteomics analyses (featuring qualititative and semiquantitative analysis on all proteins, and absolute quantification of proteins from several metabolic pathways) were performed and correlated with physiological measurements (including Fv/Fm, Biomass and lipid content) and an additional *Chlamydomonas reinhardtii* transcript dataset. Besides, we also aimed at a narrower search of transcription factors and biomarkers for these metabolic changes — hence, enriched nuclei fractions were prepared at every time point for Shotgun Proteomics analysis. To enhance our biomarker search, we additionally used the database-independent MAPA-approach, as described in **Hoehenwarter et al. 2011** and Egelhofer et al. 2013 **[126]**, [75].

2.3.1 Declaration of authorship

The results of this chapter are presented in the form of a manuscript published in the journal "Biotechnology for Biofuels". For this work, I contributed to the experimental design and carried out the experiments relative to the sampling, the cell nuclei purification and proteomics analysis - I also cowrote the Materials and Methods section, and collaborated in the data mining and evaluation.

2.3.2 Published manuscript

Valledor et al. Biotechnology for Biofuels (2014) 7:171
DOI 10.1186/s13068-014-0171-1

Biotechnology for Biofuels

RESEARCH ARTICLE Open Access

System-level network analysis of nitrogen starvation and recovery in *Chlamydomonas reinhardtii* reveals potential new targets for increased lipid accumulation

Luis Valledor[1,2,3*], Takeshi Furuhashi[1], Luis Recuenco-Muñoz[1], Stefanie Wienkoop[1] and Wolfram Weckwerth[1*]

Abstract

Background: Nitrogen starvation is known to cause drastic alterations in physiology and metabolism leading to the accumulation of lipid bodies in many microalgae, and it thus presents an important alternative for biofuel production. However, despite the importance of this process, the molecular mechanisms that mediate the metabolic remodeling induced by N starvation and especially by stress recovery are still poorly understood, and new candidates for bioengineering are needed to make this process useful for biofuel production.

Results: We have studied the molecular changes involved in the adaptive mechanisms to N starvation and full recovery of the vegetative cells in the microalga *Chlamydomonas reinhardtii* during a four-day time course.
High throughput mass spectrometry was employed to integrate the proteome and the metabolome with physiological changes. N starvation led to an accumulation of oil bodies and reduced Fv/Fm.. Distinct enzymes potentially participating in the carbon-concentrating mechanism (CAH7, CAH8, PEPC1) are strongly accumulated. The membrane composition is changed, as indicated by quantitative lipid profiles. A reprogramming of protein biosynthesis was observed by increased levels of cytosolic ribosomes, while chloroplastidic were dramatically reduced. Readdition of N led to, the identification of early responsive proteins mediating stress recovery, indicating their key role in regaining and sustaining normal vegetative growth.
Analysis of the data with multivariate correlation analysis, Granger causality, and sparse partial least square (sPLS) provided a functional network perspective of the molecular processes. Cell growth and N metabolism were clearly linked by the branched chain amino acids, suggesting an important role in this stress. Lipid accumulation was also tightly correlated to the COP II protein, involved in vesicle and lysosome coating, and a major lipid droplet protein. This protein, together with other key proteins mediating signal transduction and adaption (BRI1, snRKs), constitute a series of new metabolic and regulatory targets.

Conclusions: This work not only provides new insights and corrects previous models by analyzing a complex dataset, but also increases our biochemical understanding of the adaptive mechanisms to N starvation in *Chlamydomonas*, pointing to new bioengineering targets for increased lipid accumulation, a key step for a sustainable and profitable microalgae-based biofuel production.

* Correspondence: luis.valledor@univie.ac.at; wolfram.weckwerth@univie.ac.at
[1]Department of Ecogenomics and Systems Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria
[2]Cyanoteam, Global Change Research Center-Czechglobe. Academy of Sciences of the Czech Republic, Belidla 4, 603 00 Brno, Czech Republic
Full list of author information is available at the end of the article

Background

The need for an alternative to fossil fuels, with rising prices and declining reserves, has led to renewed interest in microalgae as a potential source for biofuel production [1-3]. Nutrient availability, temperature, and light intensity are the major environmental determinants of algal growth, reproduction, and morphology, including the accumulation of lipids in the form of triacylglycerols (TAGs) [4-6]. Photoautotrophically grown algae offer better solar energy conversion efficiency and a range of technical and ethical advances compared to traditional oil crops [7-9]. Considering their increasing importance as bioproducers and the need to achieve an optimized balance between lipid production and growth, thorough analyses of the underlying molecular mechanisms that mediate stress-induced accumulation of lipids in microalgae are necessary. Nevertheless, these analyses are still in a very early stage [10].

Growth-inhibiting conditions and an imbalance between carbon and some macro- and micro- elements such as Fe, S, Zn, or N [11] lead to metabolic rearrangements modulating cell division, morphology, and photosynthetic capacity [12] and the accumulation of starch [13] and/or lipids [14] to protect cellular structures and increase the microalgae survival probability under adverse circumstances. The accumulation of lipid bodies in *Chlamydomonas reinhardtii* under N deficiency has been recently documented in detail [4,14,15], establishing a well-known environment in which changes in morphology and some key genes are defined.

These studies together with the availability of a sequenced genome [16], proteomics and metabolomics protocols and databases [17-27], pathway annotations [28-30], and a wide range of molecular biology [31] and transcriptomics tools [32] make *C. reinhardtii* the premier molecular model for research in microalgae.

The employment of recent advances in high throughput profiling methodologies has allowed the system-level characterization of *C. reinhardtii* at transcriptomic [11,15,33], proteomic, and metabolomic levels [19,21,34]. In the present study, we have added a further layer of investigation, in particular distinguishing short- and long-term adaptive mechanisms and the recovery phase of the cells from N starvation to normal vegetative growth. In contrast to previous studies on differential gene expression [33,35] or metabolomics analyses [34], we have studied N starvation and the following recovery process after N readdition during a four-day experiment. Using classical physiological measurements, mass spectrometry for quantitative proteomics (GeLC-LTQ-Orbitrap-MS) and metabolomic (GC-MS) changes, and mining available datasets we depicted the responses of *C. reinhardtii* to available N, showing the dynamic behavior of the biochemical pathways and metabolism to the N availability and providing new potential bioengineering targets for increased lipid accumulation.

Results

Physiological responses to nitrogen starvation and recovery in *Chlamydomonas reinhardtii* cells

Chlamydomonas reinhardtii cells show a high ability to adapt dynamically to environmental conditions. The stress adaption process is based in short- and long-term changes in metabolism affecting the morphological phenotype. Therefore, we have selected controls (0 h), three sampling times under N starvation (5 h-N5h, 24 h-N24h, 72 h-N72h), and two further samplings after N replenishment (77 h + N5h, 96 h + N24h), aiming to cover both short- and long-term responses for acclimation and recovery.

N starvation leads to a block of growth (Figure 1), which is significantly slower than that in control cultures (Additional file 1: Figure S1). The fresh weight (FW) of the cultures was sustained during initial N starvation, and reduced after 72 h. Also, chlorophylls were affected, with a turn in the culture color from green to yellow under N starvation with a 30% decrease in Fv/Fm (Figures 1 and 2). As expected, N starvation induced a quick accumulation of lipids, a 1.75-fold increase in 72 h (Figure 1), most of them in the form of lipid bodies (Figure 2). N starvation is affecting the normal physiological behavior of the cells, blocking cell growth and reducing photosynthesis. Cells stop dividing probably because there is no available N for sustaining protein and nucleotide biosynthesis. Under these circumstances intracellular N should be recycled to support critical life-supporting pathways. Photosynthesis and antennas are reduced to avoid oxidative cellular damages and changes in cellular pools. The excess of available energy and carbon is then channeled into an increased production of lipids. Lipids act not only as an energy and carbon sink, but also in membrane stabilization.

N replenishment quickly reverted these physiological adaptations, with total lipids, Fv/Fm, and chlorophylls returning to normal levels in 24 h (Figures 1 and 2). The FW of cell cultures was increased 1.5- and 2.5-fold at 5 h or 24 h after N replenishment. The degradation of TAGs and also the recovered photosynthetetic activity probably provided the energy needed for this fast recovery of vegetative cell growth. These results support the hypothesis that there are sensing mechanisms that regulate not only N uptake and N assimilation, but also many other metabolic pathways (ranging from glycolysis to secondary metabolism) defining physiology and cell growth.

An initial functional overview of the changes in the proteome (Additional file 2: Tables S1, S2) and metabolome (Additional file 2: Table S3) showed that N starvation affects most of the cellular metabolism (Figure 3, Additional file 3: Figure S2, Additional file 4: Figure S3, Additional file 5: Figure S4, Additional file 6: Figure S5, Additional file 7: Figure S6), that is, ammonia transport and

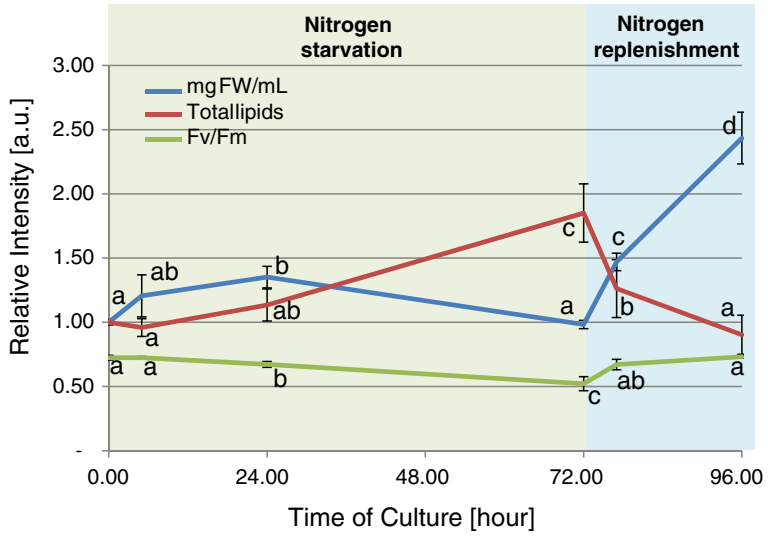

Figure 1 Physiological measurements. Variations in the photosynthetic rate (Fv/Fm) and in the accumulation of lipids, fatty acids, and fresh weight during the experimental time course. Different letters within series indicate significant differences (ANOVA followed by a TukeyHSD, $P < 0.05$). Fresh weight and fatty acid values were normalized as a relative abundance considering the average contents in T0. Fv/Fm values were plotted without transformation.

fixation proteins, photorespiration, or oxidative pentose phosphate pathways. Differential pathways are depicted below.

We applied multivariate analyses for integrating physiological measurements with protein and metabolite datasets, reducing the complexity of the data, and performing untargeted correlation network analysis [22,36-38]. Principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA) resulted in a similar classification of the samples (Figure 4A; Additional file 8: Figure S7a, b; Additional file 2: Tables S4,S5) while sparse partial least square (sPLS) analysis gave a slightly different picture (Additional file 8: Figure S7c; Additional file 2: Table S6). PC1 seems to explain the adaption to N starvation: ammonia transport and assimilation enzymes are positively correlated to this component, while N demanding activities, like polyamine and protein biosynthesis, and amino acid degradation are negatively correlated.

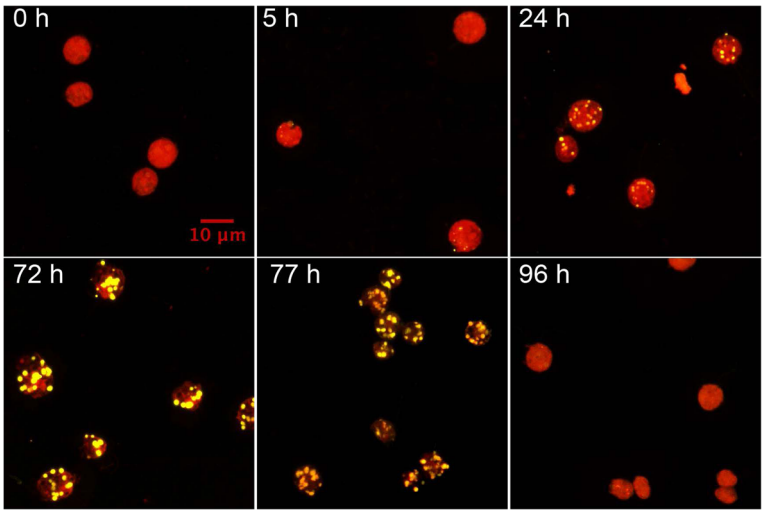

Figure 2 Lipid bodies accumulation during the time course experiment. Nitrogen starvation causes a quick accumulation of lipid bodies (yellow dots) with a maximum density obtained after 72 h of culture. Notice that the autofluorescence of the cells is dramatically reduced at 72 to 77 h, indicating a sharp decrease in the pigment content. Lipid bodies disappear completely 24 h after the addition of ammonia (96 h). Cells were fixed in 1% formaldehyde and stored at 4°C until analysis. Cells were stained with Nile red to identify the lipid bodies.

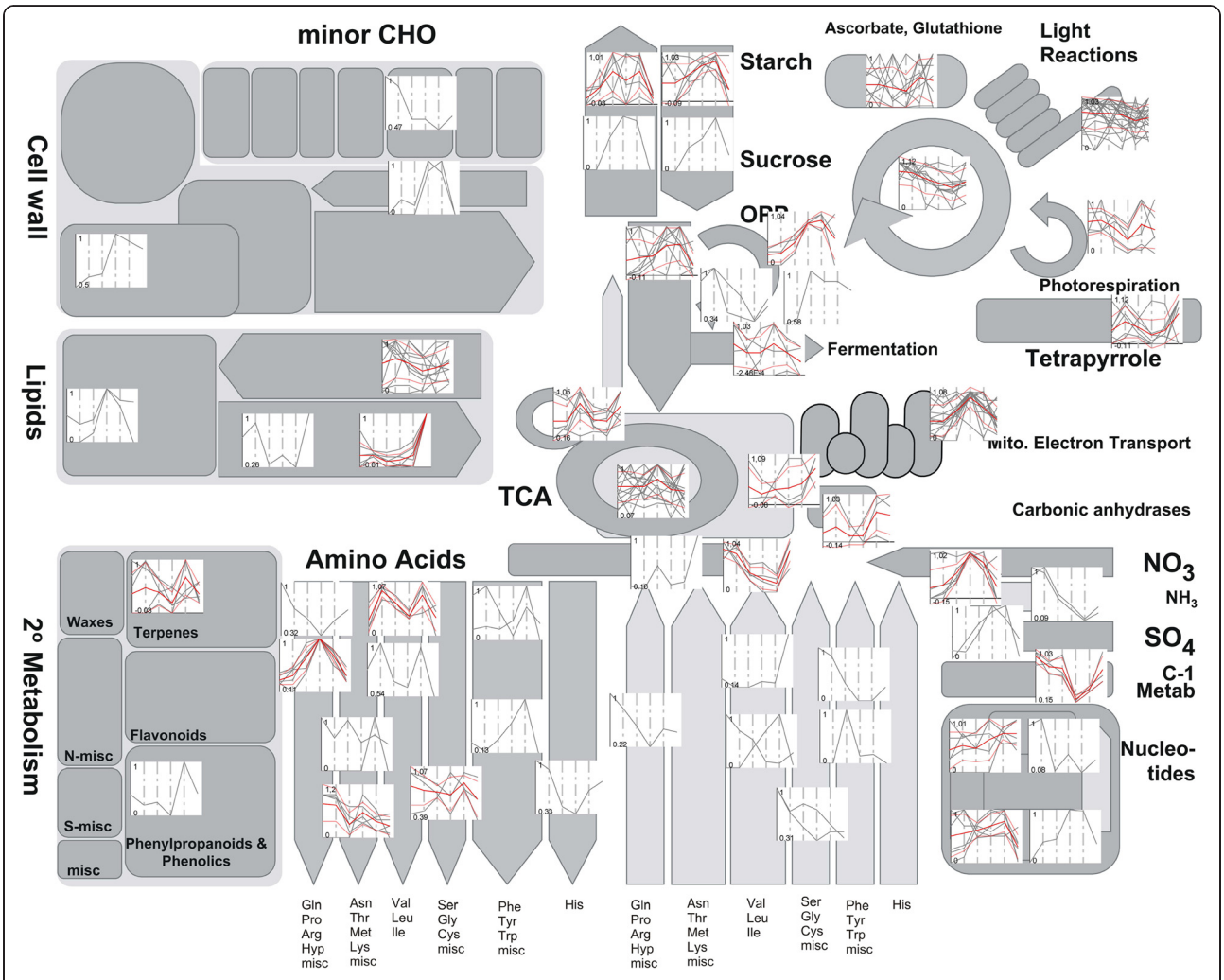

Figure 3 Representation of N starvation- and recovery-induced changes in major pathways and processes using MapMan visualization platform and our *Chlamydomonas reinhardtii* mapping. Lines in gray represent individual proteins that were differentially accumulated during the experiment ($P < 0.05$) at each category, thick red lines represent the average value within all of the clustered proteins at each time point, and thin red lines represent the average ± one standard deviation. Protein abundances were normalized as a percentage of the maximal value in the time series.

Furthermore, the high correlation of central metabolism enzymes and cGMP-dependent kinases showed that N starvation also affects normal respiration. The high correlation of glycerol pathways and C18:2 indicated the importance of lipid biosynthesis during N starvation. PC2, on the other hand, explains the N-starvation recovery and growth on the basis of an increased energy production through lipid degradation, and the recovery of the pigments biosynthesis. These findings support our previous hypothesis, based on physiological parameters, of how N starvation affects the cell function since the previously described effects can be explained at a biochemical level based on PC1 and PC2 (plotting other components does not improve the separation between groups (Additional file 8: Figure S7d, e, f)). Biclustering of functional categories of protein changes (Figure 4B) against the time points of N starvation and N readdition demonstrated impressively that the recovered cells are more similar to the control cells and that there is a complete remodeling of cellular processes during starvation and recovery. Clusters of short- and long-term responsive functional processes can be identified. Furthermore, we have used sPLS networks in which protein and metabolites were used as predictive variables to explain the physiological parameters aiming to establish interactions between proteins, metabolites, and phenotype (Figure 4C, Additional file 6: Figure S5D, Additional file 9: Figure S8, Additional file 10: Figure S9).

Dynamics of 1,658 proteins and 52 metabolites during nitrogen starvation and recovery

The application of mass-spectrometry-based profiling of proteins and metabolites allowed a comprehensive analysis

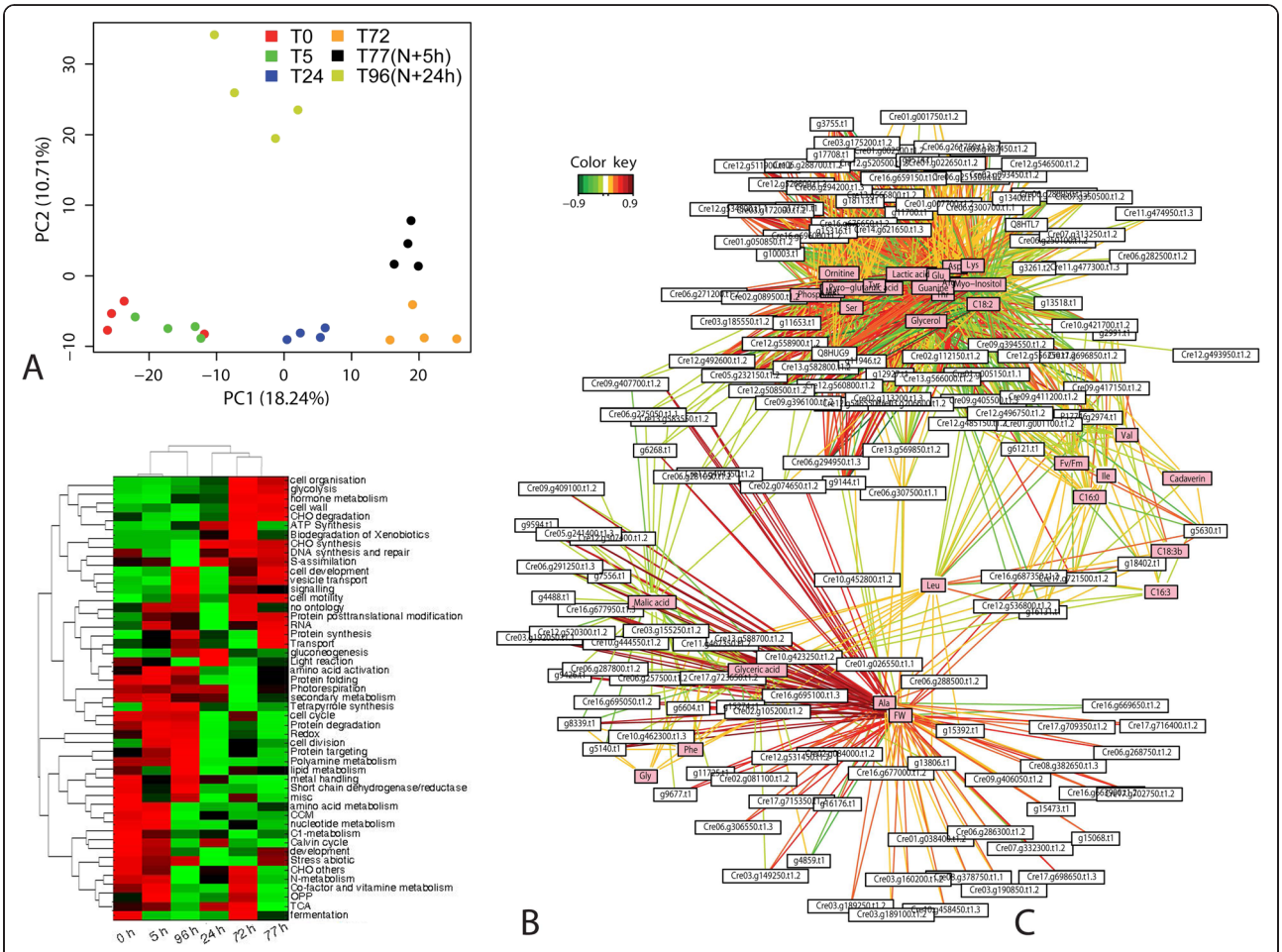

Figure 4 Classification of the different samples according to multivariate methods. (A) Principal componet analysis (PCA) of the integrated proteomic, metabolomic, and physiological datasets. Glutamate family enzymes (ASS, ASL, OCT, NAG), ammonia metabolism (Cre13.g592200.t1.2), purine biosynthesis proteins (Cre07.g318750.t1.2, Cre08.g364800.t1.2), NADH:ubiquinone oxidoreductase, cGMP-dependent kinases (Cre03.g199050.t1.2), glycolysis enzymes (PK, GAP-DH, PEPC), glyceraldehyde-3P-DH, glycerol, and C18:3 showed high correlations to PC1. Calvin cycle proteins (SBPase, PPE), chloroplastic ATPase, amino acid degradation, polyamine synthase, fatty acid elongation, catalases, and aspartic acid showed a negative correlation to PC1. Fresh weight, alanine, beta oxidation-related proteins (Acyl-CoA oxidases, HADH), oxidoreductases (Cre16.g677950.t1.3, g13806.t1, g4488.t1, g9426.t1), signal peptide and protein peptidases, and tetrapyrrole biosynthesis proteins showed a high correlation to PC2, while organic acids (fumaric and glyceric), phosphate, and photosynthesis-related enzymes (light reaction and carbon fixation), showed a negative correlation. These variables were used to infer the biological meaning of the principal components 1 and 2. Loading matrix is available in Additional file 2: Table S4. **(B)** Hierarchical clustering and heatmap of the analyzed proteins grouped by functional category according to MapMan. Three different clusters (0, 5 h, and 96 h, 24 h, and 72 and 77 h) can be distinguished, showing the different degrees of response to N starvation. The aggrupation of the 96-h and 5-h samples remarks the effect of the N repletion over the cultures. **(C)** sPLS-based network, showing the significant interaction between proteins and metabolites. Three major clusters can be distinguished, corresponding to fresh weight, glycerol/C18:2/C16, and N metabolism, which are mixed in the image. Fv/Fm and C16 are outside these groups acting as a link. For further details see Results and Discussion. This complex network is depicted in Additional file 6: Figure S5D, Additional file 10: Figures S9 and Additional file 11: Figure S10.

of the responses to N starvation and N readdition. More than 15,000 peptides and 3,200 proteins were detected in 552,529 spectra obtained from whole cell and nuclei protein extracts. 1,658 proteins were above the minimum abundance threshold for confident quantitation (Additional file 2: Tables S1 and S2). This number represents about 15% of the genes in the *C. reinhardtii* genome [16]. GC-MS allowed the unequivocal identification of 52 primary metabolites after comparison to reference standards (Additional file 2: Table S3) and quantified 60% of them with differential accumulation, at least in one of the sampling times.

The use of proteomics- and metabolomics-based approaches provides a direct readout of the metabolic and physiological adaptive mechanisms that are present in the cell and links molecular dynamics to genome-based theoretical metabolic networks [39,40]. The reconstruction of the *C. reinhardtii* network based on KEGG Orthology (KO) annotation consists in 7,330 reactions belonging to 263 pathways, which are catalyzed by 713 enzyme classes

[23]. Of the total proteins described in this study, 845 were annotated to enzyme commission (EC) numbers using Biomart, the KEGG enzyme database, and manual annotation during the curation process. These enzymes defined 447 classes of reactions belonging to 157 pathways, constituting more than 50% of the modeled class reactions and pathways in *C. reinhardtii* [23,41].

We have further analyzed the total proteome using MapMan [30,42] and a custom bin map for the release 5.3 of the *Chlamydomonas* genome. A total of 1,233 proteins were assigned to functional bins. Employing this classification tool allowed us to plot the dynamics of the differentially accumulated proteins (P <0.05) to illustrate the overall changes in the metabolome (Figure 3), and also to focus specifically on photosynthetic (Additional file 3: Figure S2), N (Additional file 4: Figure S3, Additional file 5: Figure S4), lipid (Additional file 6: Figure S5), and nucleotide metabolism (Additional file 7: Figure S6). The integration of the generated data with the available datasets with transcriptomic [33] and metabolomic [34] data in the overlapping experimental points (0 and 72 h) allowed a robust and comprehensive analysis of the molecular changes induced by N starvation. However, 166 differentially accumulated proteins of the total protein dataset were not mapped to any pathway, indicating the need for a continuous improvement of functional annotation.

Response of *Chlamydomonas reinhardtii* primary metabolism to N depletion and recovery

N starvation induces quick and profound changes in primary metabolism (Figure 3, Additional file 3: Figure S2, Additional file 4: Figure S3, Additional file 5: Figure S4, Additional file 6: Figure S5, Additional file 7: Figure S6). Glyceraldehyde-3-P-dehydrogenase (GAPDH) showed maximum peaks at 72 h. Interestingly, one pyruvate kinase isoform (Cre06.g280950.t1.2) was only detectable after N depletion, showing a drop in abundance 24 h after N repletion. At the same time pyruvate increased after N repletion. Fructose bisphosphate aldolase (FBA2) and glucose-6-P-isomerase (PGIC) showed a peak after 5 h and were then repressed until N repletion. This repression might lead to the accumulation of the glucose-6-P-pool which could be channeled into starch synthesis. Indeed, starch accumulates during N starvation [4,13]. Starch synthases (SS1, SS2, SS4, GBSS1) were upregulated during starvation, and quickly decreased after N repletion. The sugars fructose, glucose, and trehalose are accumulated during starvation (+15-, +5-, and +9-fold, respectively), being reduced after N repletion. Trehalose has been described as a substitute of N-containing compatible solutes [43] and a positive regulator of starch biosynthesis [44]. More recently a key role of trehalose in stress responses by regulating the SNF/bZIP system was postulated [45]. Glucose and fructose can also act as regulators of this system involving hexokinase (HXK) and potentially sucrose and trehalose metabolism [46]. However, sucrose metabolism is not clarified completely in *Chlamydomonas reinhardtii*. One gene which is at least assigned to sucrose metabolism is sucrose phosphate phosphatase (SPP). HXK and SPP abundances were already increased 24 h after N starvation, showing a maximum peak after 72 h, being quickly downregulated after N repletion.

On the other hand, the pool of glucose-6-P can also enter the oxidative pentose phosphate pathway. The rate-controlling enzyme of this pathway glucose-6-P-dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGDH) showed an increased abundance under N starvation with a peak 5 h after N repletion (+3 to 5fold). In contrast, nonoxidative enzymes, transaldolase and cytosolic ribose-5-P-isomerase, showed an opposite trend with a minimum presence after 24 h of stress, potentially reducing the synthesis of ribulose-5-P (see discussion below).

To extend the classical correlation analysis, we applied Granger causality analysis (see also [22,38,47,48]). Here, time-lagged correlations are exploited with the potential that pairs of precusors and products can be identified. The Granger causality function is implemented in the COVAIN toolbox [22,49]. The COVAIN function also applies a Benjamini-Hochberg correction to reveal only significant correlations (for further information see the COVAIN manual; http://www.univie.ac.at/mosys/software.html). Because of the low number of time points, only a few Granger causalities have been identified and they have to be treated carefully. For example, in Figure 5A Granger causality showed a steady increase of NADH: ubiquinone reductase (Cre09.g415850.t1.2) that was negatively correlated with one of the precursors of NAD-biosynthesis nicotinic acid (Figure 5A), pointing to an increase in mitochondrial activity. The citric acid cycle (tricarboxylic acid, TCA) enzymes were in general upregulated, with a twofold increase for most of the proteins after 72 h, and a quick return to control levels after N repletion. Both succinate dehydrogenase (SDH2) and pyruvate dehydrogenase (PDH1) behaved the same. The abundance of TCA intermediates was not consistent with a unified response of this pathway. Pyruvate concentration is reduced during N starvation while malate is accumulated. Succinate and oxalate levels are constant during the whole experiment. The malic enzyme (ME3) also remains at constant levels during N starvation, but after repletion its abundance is increased sevenfold, which is correlated to a fivefold decrease of malate abundance. The upregulation of the TCA cycle is also functional for the production of carbon backbones for amino acid biosynthesis by anaplerotic reactions.

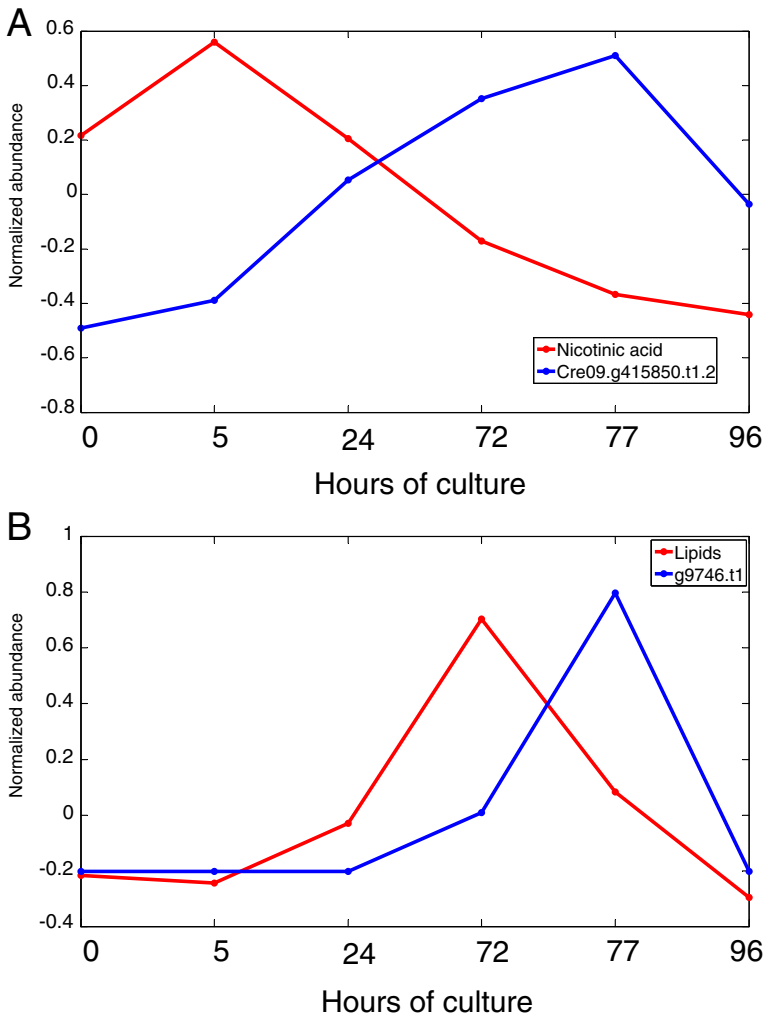

Figure 5 Examples for identified Granger causalities. Metabolite, protein, and physiological data were combined and analyzed with respect to time-shifted correlations using the toolbox COVAIN [38]. This procedure, called Granger causality, allows for the identification of directed correlations or nonlinear processes. Therefore, these associations must be carefully interpreted. **(A)** The metabolite nicotinic acid (NA) shows a time-shift behavior with respect to an NAD-reductase. Because NA is an intermediate in NAD biosynthesis, the increased levels of NAD-reductase may be involved in the consumption of NA. **(B)** A clathrin assembly protein involved in vesicle formation (COP II, see also Discussion) follows with a time-shift the accumulation of total lipids. After readdition of N, the protein concentration declines. This protein is also highly correlated to a major lipid droplet protein, glycerol, and various fatty acids, indicating its potential role in the formation of lipid bodies during N starvation (for further details see Results and Discussion sections).

Most of the proteins related to oxidative phosphorylation are accumulated during N starvation. NADH-DH isoforms, ATP synthases, and cytochrome c oxidase (COX) increased up to seven-, two-, or eightfold, respectively (Additional file 2: Table S1). This response is not shared with other nutrient depletions such as sulfur [50] or temperature stress [51].

N starvation induces a decrease in abundance of proteins involved in photosynthesis and increased carbon concentrating mechanism (CCM)

A sharp decrease of light-dependent reactions proteins (Additional file 3: Figure S2) followed N starvation. The reduction of cytochrome b6f complex (negative fivefold), ATP synthases (negative twofold), and photosystem II (PSII) proteins may explain at least partially the observed reduction in photosynthesis, confirming the observations of Majeran *et al.* [52], which showed a degradation of cytochrome b6f and light harvesting complexes (LHCs) after N starvation. Interestingly, although only two PSII proteins reached control levels, the Fv/Fm rate was fully recovered 24 h after replenishment. These results showed that, as proposed by Plumley and Schmidt [53], the efficiency of PSII during N starvation is not mainly limited by protein content but by other compounds such as chlorophylls. This was supported phenotypically and molecularly

with a sharp recovery of green cells in the culture and the upregulation of the tetrapyrrole-related biosynthetic pathways (Figure 3).

Other pathways such as the cyclic electron flow have been recently described as essential mechanisms for photosynthesis [54] to provide ATP in situations of stress [55]. Cyclic electron flow-related enzymes showed a peak after 72 h (up to threefold), being then reduced after repletion, confirming its essential role as initially suggested by Peltier and Schmidt [56].

Most of the enzymes belonging to the CO_2 fixation pathway, like RuBisCO large subunit (RBCL), FBPase, transketolase, or phosphoglycerate kinase, were reduced under N starvation (Additional file 3: Figure S2) as previously pointed out in [34]. Other proteins such as RuBisCO activase (RBCA) or phosphoribulokinase (PRKA) were increased with time, with a maximum abundance peak measured after 72 h in the case of PRKA.

Carbonic anhydrases (CAH1, CAH3, CAH7, and CAH8) belonging to CCMs are differentially regulated during N starvation. CAH3 decreased in abundance; conversely, CAH7 and CAH8 increased in abundance, peaking at 72 or 77 h. Phosphoenolpyruvate carboxylases (PEPC1, PEPC2) were not detectable when N was present in the media (0 and 96 h) and its maximum abundance was detected at around 72 h. Apparently, *C. reinhardtii* is highly flexible to maintain CO_2 fixation, triggering specific enzymes depending on the type of stress, as cold stress was characterized by increased PEPC2 and CAH3 [22].

We have not detected a good correlation between changes of photosynthesis-related proteins and its transcripts (r^2 = 0.166), showing that the specific regulation of protein species abundance does occur at the translational level, but also by posttranslational and degradation mechanisms.

Nitrogen metabolism quickly adapts to an environment without ammonia

N starvation leads to an increased abundance of specific mRNAs, proteins, and metabolites [15,33,34,57]. In our study we used the strain CC503, a *nit1 nit2* mutant lacking an active nitrate signaling pathway (*nit2*). As expected, no differences within nitrate transporters, nitrate, and nitrite reductases (Additional file 4: Figure S3) were detected. By contrast, proteins involved in NH_4^+ assimilation were overaccumulated: two ammonia transporters (g3261.t2, Cre13.g569850.t1.2) were induced by N starvation. Interestingly, Cre13.g569850.t1.2 accumulation continued with a maximum abundance 24 h after N repletion (a threefold increase compared to 24 h of N starvation), indicating a long responsive time in this element. Enzymes of the Glu-Gln (GS/GOGAT) system were also upregulated, and the abundance of all glutamine synthase (GS) isoforms increased five- to tenfold, just to decrease quickly to control levels after N repletion. On the other hand, glutamate synthase (GSN) isoforms behave differentially: GSN1 abundance is stable, while GSN2 was induced by N starvation. Glutamate dehydrogenases (GDH1, GDH2), involved in N degradation, showed a tenfold decrease starting to respond to N repletion after 24 h (slower than GS and GSN enzymes). The correlation of fold change at transcript and protein levels ($r^2 > 0.8$) points to the transcriptional regulation of this system.

The central amino acid metabolism and some degradation pathways were decreased (two- to threefold) after N starvation (Figure 3). Conversely, the glutamate family pathway quickly responded to the availability of N, being strongly upregulated by N starvation with a fourfold increase in N-acetyl-gamma-glutamyl-phosphate reductase (AGPR), arginosuccinate lyase (AGS1) and synthase (ARG7). Furthermore, three enzymes of the branched amino acid biosynthesis (acetolactate synthase 1 and 2, ALS1, ALS2; acetohydroxy acid isomeroreductase, AAI1) showed abundance peaks at 5 h and 77 h. Most of the amino acids were tightly connected in the sPLS correlation network (Additional file 11: Figure S10) with glutamate and its derivative pyroglutamic acid among the central nodes. These nodes also linked proteins involved in a wide range of activities and functions, showing the diversity of the responses triggered by N starvation.

Cytosolic ribosomes are accumulated during N starvation, while chloroplastidic ribosomes are degraded

A general ribosomal degradation and resynthesis is classically associated to N starvation [58]. These observations are partially supported, at least for chloroplast ribosomes, by a recent high-throughput transcriptomic study [33]. We detected a differential behavior between chloroplast and cytoplasmic ribosomes (Additional file 5: Figure S4). Evidence of protein accumulation was observed for cytosolic ribosomes, with a two- to tenfold increase in the proteins corresponding to 40S. In the case of chloroplastidic ribosomes, the abundance of both subunits decreased after N starvation, being downregulated by more than 40% of the quantified proteins. However, these proteins quickly responded to N replenishment, recovering its initial abundance values in 5 h. Our data refine the current model of ribosome recycling following N starvation, in contrast to previous theories pointing to an untargeted ribosomal degradation-resynthesis. Here, we rather observe a targeted accumulation of cytosolic and a degradation of chloroplastidic ribosomes. Furthermore, the differential behavior of cytosolic, increased, and chloroplast, decreased, ribosomes suggests a more active role of the nuclear encoded proteins for adapting to N starvation.

Response of lipid metabolism to N starvation

N starvation led to the accumulation of lipids (Figures 1 and 2) mainly in the form of TAGs [10]. In consequence

we expected to detect upregulated lipid biosynthetic pathways but, surprisingly, most of the enzymes of the fatty acid biosynthesis were not changed after starvation (Additional file 6: Figure S5A). The key enzymes 3-ketoacyl-CoA-synthase 1 (KAS1), which catalyzes the reaction of acetyl-ACP with chain-extending malonyl-ACP, and one isoform of enoyl-ACP-reductase (Cre06.g294950.t1.3), which catalyzes the last step of the fatty acid biosynthesis, are downregulated. This behavior was previously shown at protein level [59]. The downregulation of two triglyceride lipases (Cre01.g002400.t1.3, g9707.t1) suggests that this decrease compensates the decrease in lipogenic enzymes allowing an effective lipid accumulation. At transcript level a modest increment (less than +2-fold) of biosynthetic genes of fatty acids under N deprivation was observed [33]. An undescribed monogalactosyl diacylglycerol synthase annotated to locus g14367.t1 showed a +3-fold change following N starvation. The product of this reaction could serve as a substrate for the desaturation of oleic and linoleic acid [60] while also controlling the cell proliferation by blocking replicative DNA polymerase as described in animals [61]. The pool of glycerol, essential for the formation of TAG, was significantly increased after N starvation (+4-fold; Additional file 2: Table S3), and can be explained by the accumulation of glycerol-3-phosphate dehydrogenase/dihydroxyacetone-3-phosphate reductases (GPD2, GPD4) (Additional file 6: Figure S5B). These enzymes showed a high correlation to glycerol in the interaction network (Additional file 6: Figure S5C). This network also links acetate, glycerol, and C18:2, oxidoreductases, ribosomal proteins, and a major lipid droplet protein (Cre09.g405500.t1.3) [62,63]. This protein has a high similarity to a major protein in lipid droplets in *Haematococcus pluvialis* and *Dunaliella* (BLAST e-values of 10^{-84} and 10^{-46}, respectively, NCBI-nr database). Cre09.g405500.t1.3 was not detected at 0 h, but from 5 h to 72 h it increased its abundance up to 14-fold. Proteins for NH_4^+ transport and assimilation and specific signaling proteins (BSU1) were also part of this network. These results demonstrate the power of the integrative networks based on sPLS-correlation for discovering new interactions between proteins and metabolites, allowing the capability to also accurately associate uncharacterized proteins to functional clusters.

N starvation leads not only to the accumulation of lipids, but also to the change of the total lipid composition of the cells. We have studied 10 long chain fatty acids (Additional file 2: Table S3, Additional file 6: Figure S5D). Four of them, C16:3, C16:0, C18:2, and C18:3 (9,12,15), were significantly different in our experiment. C16:0, C16:3, and C18:2 were accumulated, while C18:3 (9,12,15) was reduced. Chloroplastidic desaturase Δ12 (CDD12, Cre13.g590500.t1.2) is eightfold reduced at transcript level, and may explain the reduction in C18:3. Stearoyl-CoA desaturase Δ9 (SCD, Cre09.g397250.t1.2) transcripts are twofold increased during N starvation and may also explain the increase of C18:2. How this system is regulated is not clear, and cannot be elucidated alone at the protein or transcript level.

N replenishment causes a quick activation of beta-oxidation pathways, with increased levels of five acyl-CoA-oxidase isoforms and also the recovery of the levels of the lipase Cre01.g002400.t1.3.

Nitrogen starvation induces significant changes in the nuclear proteome

To further investigate changes at the nuclear regulatory level we first studied nuclei-enriched fractions of *C. reinhardtii* (Additional file 2: Table S2). Secondly, we performed an *in silico* analysis of total cell and nuclear fractions focusing on nuclear proteins related to signaling and transcriptional regulation and finding 268 proteins (Additional file 2: Table S8), from which 136 were differentially expressed ($P \leq 0.01$). As expected, the levels of histones and other core proteins did not significantly change with time, whereas other chromatin-interacting proteins followed different dynamics. N starvation induced an initial decrease in RNA polymerase I and RNA helicases, which were partially recovered after repletion. On the other hand, N starvation triggered the accumulation of DNA binding proteins, such as Cre06.g252000.t1.2, a leucine/zipper transcription factor, which was quickly silenced again after N replenishment (a fivefold reduction in 5 h). An Argonaute-like protein (AGO; Cre04.g214250.t1.3) was also induced by N repletion, suggesting a role of the siRNA system [64]. The DEAD/DEAH box helicase Cre01.g021600.t1.2 is also accumulated in the absence of N in the medium, indicating that some SWI/SNF regulation pathway may be occurring. This class of enzymes and regulation has been described to be part of the abiotic stress response mechanism [65]. We have used the Plant Transcription Factor database [66] to mine the *Chlamydomonas* genome release 5.3 for finding transcription factors. However, many of these postulated transcription factors were not detected in the nuclear-enriched fractions, pointing to its low abundance and the technical challenges for high-throughput identification. To date most of these proteins are uncharacterized, and even for the known families their members can have different regulatory roles [67].

N starvation is known for triggering gametogenesis and sexual reproduction in *C. reinhardtii* [31], but we have not detected related proteins, probably because they are expressed at very low levels or they are not functionally annotated.

Discussion

C. reinhardtii cells respond to changes in the availability of ammonia by drastically changing their metabolism and normal development. In this work we analyzed the changes in the metabolome and proteome, also integrating other

available datasets, thus presenting the most complete and comprehensive overview of the response to total available nitrogen and its recovery. The design of this study, together with the employed analytical techniques, allowed us to update current knowledge, providing new insights for the understanding of the complex metabolic dynamics that follow N starvation and, for the first time, recovery of vegetative growth after N replenishment, by using high-throughput omics analyses.

Nitrogen/carbon balance defines the metabolic switch under N starvation

The sensing of the balance between N and C is transduced into a change in N- and C-responsive pathways, leading to a metabolic adjustment that dominates many pathways and defines the observed physiological and phenotypical changes. However, it has to be considered that the presence of acetate in the medium will change the cellular sensitivity towards the maintenance of favorable C/N ratios [68], causing increased responses to the oxidative stress as a side effect of the downregulation of photosynthesis and carbon assimilation [69]. N starvation also leads to increased ammonia uptake and assimilation proteins and a decreased nucleic acid and protein biosynthesis, which may be the cause for the reduced cell growth.

The reduced photosynthesis is proposed to be mediated by the degradation of LHCs and cytochromes, rather as an adaptive mechanism to stress and energy contents than as a way to recycle nitrogen. The reduction of PSII activity would lead to the production of less reactive oxygen species, but at a cost of reducing ATP and NADPH production. But even in a situation of reduced growth, as observed under N starvation, energy is still necessary to sustain metabolism. We detected an increase of cyclic electron flow proteins which pump H^+ to the thylakoid lumen in order to increase ATP production. Furthermore, and based on the metabolic reconstruction, an increased pool of glucose-6-P was predicted. Part of this glucose-6-P can join the initial reactions of the oxidative pentose pathway. The corresponding enzymes were accumulated, rather than glycolytic enzymes. In contrast, nonoxidative enzymes, transaldolase, and cytosolic ribose-5-P-isomerase showed an opposite trend with a minimum presence after 24 h of stress, potentially reducing the synthesis of ribulose-5-P. The pool of ribose-5-P could enter the chloroplastic reductive pentose phosphate pathway to produce 3-phosphoglycerate, which can continue with the glycolytic pathway. The result of this reprogramming could be the production of enough pyruvate and $NADPH + H^+$ to support biosynthetic processes at a cost of half ATP payback compared to glycolysis. This hypothesis was supported by the accumulation of phosphoribulokinase (PRK1), which followed the same abundance trend as G6PDH and 6PGDH. These changes in protein abundance are coincident with the changes in the mRNA expression levels given by [33] in an N-depletion study; furthermore, [34] recently described the increased accumulation in the metabolites 3-phosphoglycerate and ribose-5-P during N starvation. This remodeling, although energetically unfavorable, can be used by the cell to maintain adequate levels of $NADPH + H^+$, which are required to maintain an increased lipid production. However, future studies applying flux analyses are necessary to investigate these processes in more detail.

The functional role of cyclic electron flow and the quick recovery of PSII efficiency (correlated rather to pigment biosynthesis than to the recovery of PSII proteins) indicate that photosystems and electron transfer chain were fully functional during N starvation, even though the abundance of some subunits and cytochromes was reduced.

Athough CO_2 fixation enzymes were in general reduced, the abundance of CCM proteins CAH7, CAH8, and PEPC1 increased. CAH7 was only detected after a long exposure to N depletion (72 h) and may explain why this enzyme was not detected in previous studies limited in time to not more than 48 h [33,34]. CAH7 and CAH8 are similar and both of them contain a hydrophobic chain which localizes them in periplasmic space associated to the plasma membrane, probably covering decreased CAH1 and 2 activities and transferring inorganic CO_2 to a transport system or pore [70]. The specific role of these CAHs under nutrient deprivation remains undetermined and cannot be easily explained at this time. By contrast, the kinases belonging to the Snrk2-SNF2 family have an essential role in the control of gene expression through the activation of bZIP transcription factors and SWI/SNF chromatin remodeling complexes [71,72]. These were quickly responsive to N starvation in our experimental system. Kinases of this family control the lipid accumulation in yeast [73], and have been linked to increased tolerance to N deficiency in *Arabidopsis* [74] or S in *C. reinhardtii* [50]. SNF families, also implied in energy sensing and gene regulation [75], control the metabolic response to other stresses like cold acclimation, inducing the accumulation of sugars and starch [51]. The accumulation of the signaling molecules glucose-6-P and trehalose [45,76] and the specific dynamics of HEX and SPP found in the N starvation, provide an extra support to this hypothesis.

Branched chain amino acids linked N metabolism and cell growth

The unavailability of N caused a rearrangement of amino acid metabolism at the same time. The pool of glutamate, essential for ammonia assimilation, was drastically reduced after starvation, triggering aspartate aminotransferases and citrate synthases from the oxoglutarate pathway. Enzymes

belonging to branched chain amino acids (BCAs) were negatively correlated to the abundance of Val, Ile, and Leu, which indicates that enzyme abundance is increased to compensate the lower abundance of BCAs. Leu and Ile were significantly correlated ($r^2 > 0.9$) to Cre16.g687350.t1.2, an acyl-CoA oxidase and mediator of amino acid degradation, and a monofunctional catalase which may be controlling the oxidative damage. Interestingly, the same interaction cluster was found when analyzing the sPLS-based network; it was the bridge that linked FW and amino acid-nitrogen clusters.

The increased abundance of carbamoyl phosphate synthetase subunits (CMP1, CMP2) and the dynamics shown by arginine and ornithine suggest the importance of the urea cycle for recycling the ammonia originated in the protein breakdown.

Are glycerol metabolism, MLDP, and COP II proteins key elements for lipid droplet formation and TAG accumulation?

The increased abundance of total lipids and lipid droplets following N-deprivation could indicate that the TAG of the droplets is synthesized de novo. The expected increase of the fatty acid biosynthetic pathway was, in general, not present but two Acetyl-CoA-Synthetases (g1224.t1, Cre12.g507400.t1.2) required for the activation of fatty acids and forming part of lipid droplets [62] were significantly accumulated. Glycerol metabolism, needed to provide the backbone for the glycerolipid biosynthesis, was significantly responsive, with a combined accumulation of the proteins of this pathway of more than 50-fold. Cre09.g405500, a major lipid droplet protein (MLDP) [62,63], increased 14-fold under N starvation and was linked to C18:2 and glycerol, showing further correlation (>0.93) with clathrin and COP II, which forms vesicle coats and allows liposome fusion [77]. The MLDP protein was identified by Moellering and Benning and by James *et al.* after analyzing lipid bodies associated proteins [62,78]. Recently Wase *et al.* described a similar behavior of this protein under N deprivation, showing a 4.1-fold change after six days of culture without nitrogen [59]. MLDP is highly correlated with lipid accumulation during N starvation and dropped completely after N replenishment. MLDP and other proteins linked by sPLS and Granger causality analysis such as COP II provide hub proteins in the adaptation process to N starvation and revovery (see Figure 5B). These proteins are involved in vesicle formation, suggesting their strong role in lipid body formation, thereby confining the TAGs into storage structures and promoting their accumulation. Lipid body formation can be also influenced by GTPases as COP II protein is inactivated by GTPases [79], which we also found to be decreased during N starvation. Correlation networks need to be further analyzed since they also reach fatty acid and starch biosynthetic enzymes, pointing to key processes and promising targets for strain engineering.

The increased C18:2 and the consequent alteration of the different fatty acids suggest a remodeling of the membranes, and thus fatty acids in membranes seem to be recycled into TAGs. TAG lipases and phospholipases were downregulated.

N repletion led to an increased beta-oxidation for rapid lipid degradation and also increased the abundance of five acyl-CoA-oxidases. These were not correlated to enzymes of the BCA metabolism, demonstrating differential regulatory pathways. The lipase Cre01.g002400.t1.3 that was recovered after N replenishment was negatively correlated (>|0.9|) to the translation initiation inhibitor (Cre12.g551350.t1.2) and a protein showing a DNA methyltransferase domain (Cre12.g508050.t1.2). This might indicate that the transcriptional response to N starvation is regulated not only by transcription factors, but also by epigenetic mechanisms, as suggested by [50].

Can N recovery help us to understand lipid accumulation or metabolism and provide new targets for biotechnological improvement of oil production?

The study of specific changes occurring 5 h and 24 h after N resupply provides specific targets that could have potential use for bioengineering applications. Some of them have been depicted above (PEPC, CAH); however, there are specific targets that can be highlighted. One of the possible ways of intervention involves the enzymes and proteins related to lipid body stabilization and degradation with MLDP a key protein in the lipid biogenesis network. MLDP has been studied extensively and, despite its silencing by RNAi, does not lead to an increased lipid accumulation; however, the droplet size is significantly increased [62]. This is a surprising result since our data, and also that of [54], points out that the abundance of this protein is positively correlated to lipid accumulation. The combined manipulation of this protein together with beta oxidation, silencing acyl-CoA oxidases (especially Cre16.g687350.t1.2, Cre11.g467350.t1.2) and lipase (Cre01.g002400.t1.3), could result in an increased accumulation of lipids. Lipid transporters also play a key role in lipid metabolism, with Cre15.g641200.t1.2 a mitochondrial fatty acid transporter as a responsive candidate. g13764.t1, Cre13.g573150.t1.3 (hydroxylases) and Cre11.g467350.t1.2 (acyl-CoA oxidase) were accumulated (hydroxylases) or repressed (acyl-CoA oxidase) under N starvation, following the same trend as its transcripts.

Interestingly, these genes were not reported to be significantly expressed under other stresses such as C, S, or Fe deprivation or oxidative stress according to AlgaePath [80]. However, finding central metabolism enzymes related to lipid biogenesis and only responsive to N starvation (when

comparing with the available NGS/proteomics datasets) is particularly difficult. Furthermore, modifying the gene expression of a number of genes is tricky, and a perfect flux model should be available in order to avoid unintended effects. Previous attempts to increase lipid accumulation by targeting specific genes have not been completely successful [10]. This indicates that lipid accumulation is the result of a complex regulatory network linking cellular processes such as vesicle formation as well as processes of central metabolism such as G6P, branched chain amino acids, and energy metabolism. This is clearly demonstrated by the system-level analysis of our study.

In this sense, our results and other proteomic and transcriptomic datasets [23,33-35,50,59] show many complementary genes/protein clusters responsive to stress [80]. This suggests that there are major points of metabolic regulation based on common signaling elements that can be considered as potential targets for increasing lipid accumulation. Numerous DNA and RNA binding proteins and helicases showed differential accumulations in our assay (described above), but these candidates could not be easily proposed as biotechnological targets since their interaction network is unknown. On the other hand, the BRI1 suppressor (Cre01.g050850.t1.2), an inhibitor of BRI1, a receptor-like kinase which is responsive to brassinosteroids located both in plasma and in nuclear membranes [81], was correlated to glycerol and C18:2, and it is known to initiate a signaling cascade leading to regulation of gene expression in the nucleus through BZR/BES proteins [82]. BZR is known to block the metabolic switch in response to P deprivation [83] and other abiotic stresses [82] in *Arabidopsis* and also inhibits chloroplast development [84], so a similar effect could be expected in *Chlamydomonas*. Thus, blocking not only the receptor, but also the transcription factor BZR by 14-3-3 proteins may be needed to respond to N starvation. In our dataset a 14-3-3 protein responding to N deprivation and recovery was quantified (Cre06.g257500.t1.2). The BRI1 suppressor was only detected when N was not present in the medium, showing a quick adaptive response to environment.

FtsH chloroplast metalloproteases are closely related to development [85], stress responses [86], and chloroplast function [87], also regulating lipid degradation in bacteria [88]. We found three metalloproteases (Cre12.g485800.t1.2, Cre17.g720050.t1.2, g14586.t1) that were downregulated when N was absent from the media. Interestingly, and according to [74] and considering all available datasets, Cre17.g720050.t1.2 responds only to N starvation, making this enzyme a potential candidate for further study.

Conclusions

The comprehensive analysis of systemic responses to N starvation and recovery in *C. reinhardtii* demonstrated that metabolism and growth are significantly affected at a system level. A complex network of stress-responsive proteins, metabolites, and physiological parameters was established, expanding our current understanding of physiological processes driven by a small set of proteins. Many uncharacterized proteins were identified by multivariate correlation network analysis to be involved in the response to N starvation. By N readdition it was possible to extract a list of proteins that showed a fast recovery effect, suggesting that they are highly involved in the re-establishment of vegetative cell growth. This study provides new insights and alterations to previous models and offers a complex dataset, which will be further analyzed towards increasing our biochemical understanding of the adaptive mechanisms to N starvation and recovery in *Chlamydomonas*.

Methods

Strains and cultures

Chlamydomonas reinhardtii CC-503 *cw92*, mt+, agg1+, *nit1*, *nit2* (available at the Chlamydomonas Culture Collection, Duke University) cultures were grown in HEPES-Acetate-Phosphate medium supplemented with 7 mM NH_4Cl (HAP + N; TAP medium in which Tris was replaced by 5 mM HEPES) at 25°C with shaking (120 rpm) in a 14:8 light:dark photoperiod (85 μmol m^{-2} s^{-1}; Sylvania GroLux lamps). Cultures were pelleted down by centrifugation and resuspended in HAP -N (NH_4Cl was replaced by 7 mM KCl) media to a final density of $1\text{-}3 \times 10^5$ cells mL^{-1}. Cells were sampled at times 0, 5, 24, and 72 h. After this sampling, NH_4Cl was added to the HAP-N cultures to a final concentration of 7 mM, and then the cultures were sampled after 5 h (77 h) and 24 h (96 h).

Physiological measurements

At each harvesting time the cell density was measured by employing a Thoma counting chamber and the fresh weight was determined gravimetrically. The photosynthetic rate was measured with an imaging/pulse-amplitude modulation fluorimeter (OS1-FL, Opti-Sciences).

Total lipids were extracted from frozen pellets with 200 μL of a mixture of chloroform:isopropanol (1:1) and vigorous vortexing for 3 min. The samples were centrifuged (14.000 × g, 5 min, room temperature) and the supernatants were transferred to a new tube. The pellet was re-extracted with 500 μL of hexane and vigorous vortexing for 3 min. The samples were centrifuged, and the combined supernatants were dried in a speed vac. The amount of lipids was determined gravimetrically.

Microscopy

The *Chlamydomonas* cultures were fixed in 3% (v/v) formaldehyde and kept at 4°C until staining. The cells were stained in a mixture of HEPES-Acetate-Phosphate medium with 3% (v/v) formaldehyde and 5 μg mL^{-1} of Nile red and incubated

15 min in the dark before imaging. Nile red was freshly added to the staining solution from a concentrated stock (0.1 g mL^{-1} in acetone). The stained cultures were directly observed by adding 15 μL to each slide under an LSM 780 confocal microscope (Zeiss, Germany) and Z series covering all of the cell height were captured. DIC, Nile red, and autofluorescence were recorded in different channels. From these stacks the four images closer to the middle section were selected and used for obtaining a maximum projection using the software Fiji (Figure 2).

Nuclei isolation

Nuclei were isolated following the protocol described in [89] starting from 250 mg of fresh weight. Isolated nuclei were separated from cell debris by the use of a Percoll cushion.

Quantitative proteome analysis (GeLC-LTQ-Orbitrap-MS)

Proteins were extracted from frozen pellets (40 to 50 mg fresh weight), gel fractionated, and trypsin digested following a previously described protocol [18]. Ten micrograms of digested peptides were loaded per injection into a one-dimensional nano-flow LC-Orbitrap/MS and resolved in a 90-min gradient from 5 to 40% (v/v) acetonitrile/0.1% (v/v) formic acid using a monolithic C18 column (Chromolith RP-18r, Merck, Darmstadt, Germany). MS analysis was performed on an Orbitrap LTQ XL mass spectrometer [22].

The raw data were searched with the SEQUEST algorithm present in Proteome Discoverer version 1.3 (Thermo, Germany) as described by Valledor *et al.* [17] using the *Chlamydomonas* genome v.5.3 (17,737 accessions), *Chlamydomonas* mitochondria (8 accessions), and chloroplast (76 accessions) databases. Only highly confident proteins, defined by at least two peptides with XCorr value greater than charge state +0.25 and 5% FDR, were considered for this work. Protein functions were identified using the BioMart tool available at Phytozome (http://www.phytozome.org), Mercator (http://MapMan.gabipd.org/web/guest/app/mercator), and the Algal Functional Annotation Tool [28].

The identified proteins were quantitated by a label-free approach based on total ion count followed by an NSAF normalization strategy:

$$(NSAF)_k = (PSM/L)_k \Big/ \sum_{i=1}^{N} (PSM/L)_i$$

in which the total number of spectra counts for the matching peptides from protein k (PSM) was divided by the protein's length (L), then divided by the sum of PSM/L for all N proteins within each sample. Before quantitation a second filtering step was performed to retain only those proteins with enough abundance (minimum abundance of 0.001) for a low-biased quantitation [90,91]. Statistical analyses were conducted according to [92,93]. Proteins were accounted for quantification only if they were present in all of the biological replicates (n = 4) at least one sampling time, or in five samples corresponding to different times. Missing spot volumes were determined from the dataset employing a sequential K-nearest neighbor algorithm. Univariate comparison between treatments was performed by the nonparametric Kruskal-Wallis test over log-transformed data for obtaining *P*-values (Additional file 2: Tables S1, S2, S8). Protein abundance values were scaled and subjected to principal component analysis (PCA), discriminant analysis (PLS-DA), sparse partial least square (sPLS), and heatmap-clustering analyses (Additional file 2: Tables S4-S6).

GC-MS polar metabolite and fatty acid methyl ester (FAME) analyses

Polar metabolites were extracted from frozen pellets (15 to 20 mg fresh weight) as described in [94]. Fatty acids were extracted from frozen cell pellets (15 to 20 mg) previously disrupted with liquid nitrogen and glass beads in a homogenizer (15 seconds, maximum speed, Retsch Mixer Mill MM400). One mL of cyclo-hexane:water (1:0.4) was quickly added to the homogenized tissue, and mixing was repeated with the same conditions. The organic phase was recovered after centrifugation (24,000 × g, 10 min) and transferred to a new tube. Pellets were re-extracted following the same procedures, and both organic fractions were combined in a 2-mL tube and dried by speed vac.

Sample derivatization (polar metabolites), methylesterification (lipids), and GC-MS measurements were carried out following the procedure previously developed in our group [95] on a triple quad (TSQ Quantum GC; Thermo) instrument. Metabolites were identified based on their mass spectral characteristics and GC retention times, by comparison with retention times of reference compounds in an in-house reference library. Only metabolite peaks that were detected in all of the biological replicates for at least one sampling point or in more than 18 samples were further considered. Normalized abundance values for each metabolite were obtained by dividing peak areas by the total peak areas of each sample (sample-centric normalization). Significant differences between sampling times were assessed by One-way ANOVA (Additional file 2: Table S3), and multivariate analyses were conducted together with proteins and physiological data as described above.

Bioinformatic tools

All statistical procedures described above were performed using the software R v2.15.2 [96] core functions plus the packages SeqKNN, mixOmics, and gplots. Confocal microscopy images were processed and analyzed using the last version of the software Fiji [97] employing

core packages. MapMan 3.5.1R2 [42] was used to make a functional classification of the proteins. A map for the proteome of the JGI_236 version was specifically created using Mercator [23] and the Algal Functional Annotation Tool [28]. The accuracy of the classification was supervised by comparison to local databases, also adding some manual annotations. Experimental data were classified according to this newly developed map.

We have used the dataset of transcriptomic data provided in [33]. This dataset was generated using version 4 of the *Chlamydomonas* genome. To compare this dataset to ours, we have BLASTed Chlamy 4 against Chlamy 5.3 considering significant all hits with e-values lower than 10^{-80}. The equivalences between Chlamy 4 and Chlamy 5, as the Illumina-based normalized abundance employed for this work, are given as supplementary material (Additional file 2: Table S9).

Granger causality was tested with the statistical toolbox COVAIN [38]. Parameter settings for COVAIN were Granger lag time 1 and Granger significance *P*-value 0.05 including a Benjamini-Hochberg correction. A list of all identified Granger causalities can be found in Additional file 2: Table S10.

Additional files

Additional file 1: Figure S1. Comparison of cell growth (fresh weight) of N repleted and depleted cultures of *C. reinhardtii* CC503 grown under the same conditions employed in this study. Cultures were grown in triplicate.

Additional file 2: Table S1. List of the 1,534 quantified proteins in whole cell fractions. Protein abundance was determined following an NSAF approach. The mean abundance ± SD for each sampling time is indicated, as well as *P*- and q-values. Percentage of coverage and number of unique peptides used for identification and score are indicated. Deflines, E.C. number, and MapMan bins were manually curated. **Table S2.** List of proteins extracted from nuclei-enriched extracts. Spectra were processed as previously described. Deflines, E.C. number, and MapMan bins were manually curated. **Table S3.** List of the 52 quantified metabolites. Metabolite abundance was estimated from the peak areas of the indicated characteristic ions. The mean normalized abundances with respect to control ± SD for each sampling time are indicated as well as *P*-values. **Table S4.** PCA summary, sample scores, and variable loadings of the integrated (proteins, metabolites, physiological measurements) dataset. **Table S5.** PLS-discriminant analysis correlations of the integrated dataset. **Table S6.** sPLS analysis correlations of the integrated dataset employing proteins-metabolites as predictive variables and physiological measurements as responsive variables. **Table S7.** Values of individual replicates of the datasets employed for multivariate analyses. (a) Normalized abundances for proteins; (b, c) normalized abundances for polar metabolites and FAME; (d) physiological measurements. **Table S8.** List of nuclear proteins obtained after the joint analysis of total protein and nuclear extracts. Spectra were processed as previously described, but one step of *in silico* fractionation was added to remove non-nuclear contaminants. Deflines, E.C. number, and MapMan bins were manually curated. **Table S9.** Correspondence between 454 dataset (Miller *et al.* [33]), Chlamydomonas 5.3, and MapMan analyses. Abundace of 454 detected transcripts is also provided. **Table S10.** Granger causality analyzed by COVAIN (Sun and Weckwerth, 2012 [38]) between metabolites, proteins, and physiological data.

Additional file 3: Figure S2. Representation of nitrogen starvation- and recovery-induced changes in photosynthetic pathways using MapMan. Individual plots show the variations in protein abundance at the six studied time points. Only differential proteins ($P < 0.05$) were plotted. Protein abundances were normalized as a percentage of the maximal value in the time series.

Additional file 4: Figure S3. Representation of simplified N transport, fixation, and assimilation pathways showing N starvation- and recovery-induced changes using MapMan. Individual plots show the variations in protein abundance at the six studied time points. Only differential proteins ($P < 0.05$) were plotted. Protein abundances were normalized as a percentage of the maximal value in the time series.

Additional file 5: Figure S4. Representation of ribosomal proteins using MapMan. Individual plots show the variations in protein abundance at the six studied time points. Only differential proteins ($P < 0.05$) were plotted. Protein abundances were normalized as a percentage of the maximal value in the time series.

Additional file 6: Figure S5. Metabolic changes in lipid biosynthetic pathways induced by N starvation. (A) MapMan-based representation of the evolution of the enzymes related to fatty acid biosynthesis and elongation. (B) Variations in the protein abundance of the functional groups involved in lipid metabolism. (C) Evolution of the cellular content of the indicated fatty acids during the time course experiment. (D) sPLS-based network showing the most correlated proteins and metabolites to glycerol and C18:2 nodes. C18:2 was connected to ammonia transport and assimilation proteins, whereas glycerol was mainly connected to glycerol biosynthesis proteins and NADH:ubiquinone oxidoreductases. Interestingly, Cre09.g405500.t1.3 (major lipid droplet protein, with a high importance in the model) links C18:2 and glycerol. This protein was manually annotated to major lipid droplet protein. Other proteins that link both metabolites are a ribosomal protein and g13518.t1 (annotated as amine oxidase). Network interpretation: edge color indicates correlation between nodes, circle diameter indicates the importance of each node within the network, and node color its radiality. Figure was plotted employing Cytoscape 2.8.3.

Additional file 7: Figure S6. Representation of nitrogen starvation- and recovery-induced changes in nucleotide metabolism pathways. Individual plots show the variations in protein abundance at the six studied time points. Only differential proteins ($P < 0.05$) were plotted. Protein abundances were normalized as a percentage of the maximal value in the time series.

Additional file 8: Figure S7. Classification of the integrated dataset employing different multivariate methods such as principal component analysis, PCA (a,d), partial least squares discriminant analysis, PLS-DA (b,e), and sparse partial least squares, s-PLS (c,f), for which proteins and metebolites were used as predictive variables and physiological measurements as explanatory. For all analyses the first two components allowed an effective classification of the samples based on the sample time.

Additional file 9: Figure S8. Simplified sPLS-based network constructed considering only highly correlated nodes (correlation > |0.85|). The three main nodes established different correlations among them, with the glycerol-C18:2 cluster negatively correlated to the others. Interestingly, the fresh weight (FW) cluster is positively linked to the amino acids-myo-inositol cluster by a bridge constituted by the branched amino acids. Val, Leu, and Ile are positively correlated to a monofunctional catalase grouping with amino acids, and on the other side to an acyl-CoA oxidase linked to the FW. Surprisingly one enoyl-ACP-reductase, which is correlated to the amino acids cluster, is negatively correlated to glycerol-C18:2. Glycerol is also negatively correlated to Cre12.g536800.t1.2, a protein showing an uncharacterized domain which is related to FW. Network interpretation: edge color indicates correlation between nodes, circle diameter indicates the importance of each node within the network, and node color its radiality. Figure was plotted employing Cytoscape 2.8.3.

Additional file 10: Figure S9. Detailed representation of the fresh weight (FW). FW is correlated to different kinds of enzymes: signaling enzymes, mainly GTPases and Ca2+ dependent, endopeptidases, oxidoreductases, and acyl-CoA synthases, oxidases, and dehydrogenases. Node size is dependent on stress value and node color on radiality. Edge

color indicates correlation (red, positive, green negative). Figures were plotted employing Cytoscape 2.8.3. Only highly correlated nodes (correlation > |0.85|) were plotted.

Additional file 11: Figure S10. Detailed representation of the amino acid cluster. This cluster showed complex interactions, in which nitrogen assimilation and amino acid metabolism enzymes interact with lipid biosynthesis proteins. Based on this network, some relation between N metabolism and lipid biosynthesis can be established, for example, at the level of one enoyl-ACP-reductase, which correlated to Asp, Arg, Glu, and myo-inositol, or a 3-oxoacyl-ACP reductase linked to Thr and myo-inositol. Thus, myo-inositol is a metabolite that needs to be further investigated since it is in the center of this cluster, linking amino acids, lipids, and sugars. Node size is dependent on stress value and node color on radiality. Edge color indicates correlation (red, positive, green negative). Figures were plotted employing Cytoscape 2.8.3. Only highly correlated nodes (correlation > |0.85|) were plotted.

Abbreviations

CAH: carbonic anhydrase; CCM: carbon concentrating mechanism; GC-MS: gas chromatography coupled to mass spectrometry; GeLC-LTQ-Orbitrap-MS: gel electrophoresis - liquid cromatography coupled to LTQ-Orbitrap mass spectrometer; KEGG: Kyoto Encyclopedia of Genes and Genomes; LHC: light harvesting complex; MLDP: major lipid droplet protein; PC: principal component; PCA: principal component analysis; sPLS: sparse partial least square; sPLS-DA: sPLS - discriminant analysis; TAG: triacylglycerol.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LV designed and performed the research, analyzed the data, and wrote the paper. TF contributed new GC-MS and FAME analysis procedures and performed research. LR-M performed research. SW contributed methods. WW designed the research, analyzed the data, and wrote the paper. All authors read, corrected, and approved the final manuscript.

Acknowledgements

The authors wish to thank Dr. Monica Meijón for her invaluable help with confocal microscopy and Fiji. This research and LV work were supported by the European Union through a Marie Curie IEF grant (FP7-PEOPLE-2009-IEF-255109).

Author details

[1]Department of Ecogenomics and Systems Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria. [2]Cyanoteam, Global Change Research Center-Czechglobe, Academy of Sciences of the Czech Republic, Belidla 4, 603 00 Brno, Czech Republic. [3]Present address: Plant Physiology, University of Oviedo, Catedrático Rodrígo Uría s/n, E-33006 Oviedo, Spain.

Received: 1 July 2014 Accepted: 17 November 2014
Published online: 24 December 2014

References

1. Mata TM, Martins AA, Caetano NS: **Microalgae for biodiesel production and other applications: a review.** *Renew Sustain Energy Rev* 2010, **14:**217–232.
2. 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.** *Bioenerg Res* 2008, **1:**20–43.
3. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A: **Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances.** *Plant J* 2008, **54:**621–639.
4. Siaut M, Cuine S, Cagnon C, Fessler B, Nguyen M, Carrier P, Beyly A, Beisson F, Triantaphylides C, Li-Beisson Y, Peltier G: **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.
5. Vítová M, Bišová K, Umysová D, Hlavová M, Kawano S, Zachleder V, Čížková M: ***Chlamydomonas reinhardtii*: duration of its cell cycle and phases at growth rates affected by light intensity.** *Planta* 2011, **233**(1):75–86.
6. Hema R, Senthil-Kumar M, Shivakumar S, Chandrasekhara Reddy P, Udayakumar M: ***Chlamydomonas reinhardtii*, a model system for functional validation of abiotic stress responsive genes.** *Planta* 2007, **226:**655–670.
7. Gouveia L, Oliveira A: **Microalgae as a raw material for biofuels production.** *J Ind Microbiol Biotechnol* 2009, **36:**269–274.
8. Weyer K, Bush D, Darzins A, Willson B: **Theoretical maximum algal oil production.** *Bioenerg Res* 2010, **3:**204–213.
9. Weckwerth W: **Green systems biology - from single genomes, proteomes and metabolomes to ecosystems research and biotechnology.** *J Proteomics* 2011, **75:**284–305.
10. Merchant SS, Kropat J, Liu B, Shaw J, Warakanont J: **TAG, You're it! Chlamydomonas as a reference organism for understanding algal triacylglycerol accumulation.** *Curr Opin Biotechnol* 2012, **23:**352–363.
11. Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, Merchant SS, Malasarn D: **A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*.** *Plant J* 2011, **66:**770–780.
12. Grossman A: **Acclimation of *Chlamydomonas reinhardtii* to its nutrient environment.** *Protist* 2000, **151:**201–224.
13. Ball SG, Dirick L, Decq A, Martiat J-C, Matagne R: **Physiology of starch storage in the monocellular alga *Chlamydomonas reinhardtii*.** *Plant Sci* 1990, **66:**1–9.
14. Wang ZT, Ullrich N, Joo S, Waffenschmidt S, Goodenough U: **Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*.** *Eukaryot Cell* 2009, **8:**1856–1868.
15. Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, Cokus SJ, Hong-Hermesdorf A, Shaw J, Karpowicz SJ, Gallaher SD, Johnson S, Benning C, Pellegrini M, Grossman A, Merchant SS: **Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*.** *J Biol Chem* 2012, **287:**15811–15825.
16. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, Marshall WF, Qu L-H, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen C-L, Cognat V, Croft MT, Dent R, *et al*: **The *Chlamydomonas* genome reveals the evolution of key animal and plant functions.** *Science* 2007, **318:**245–250.
17. Valledor L, Recuenco-Munoz L, Egelhofer V, Wienkoop S, Weckwerth W: **The different proteomes of *Chlamydomonas reinhardtii*.** *J Proteomics* 2012, **75:**5883–5887.
18. Valledor L, Weckwerth W: **An improved detergent compatible gel fractionation LC-LTQ-Orbitrap workflow for plant and microbial proteomics.** *Methods Mol Biol* 2013, **1072:**347–358.
19. Bolling C, Fiehn O: **Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation.** *Plant Physiol* 2005, **139:**1995–2005.
20. Hummel J, Niemann M, Wienkoop S, Schulze W, Steinhauser D, Selbig J, Walther D, Weckwerth W: **ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites.** *BMC Bioinformatics* 2007, **8:**216.
21. Wienkoop S, Weiss J, May P, Kempa S, Irgang S, Recuenco-Munoz L, Pietzke M, Schwemmer T, Rupprecht J, Egelhofer V, Weckwerth W: **Targeted proteomics for *Chlamydomonas reinhardtii* combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses.** *Mol Biosyst* 2010, **6:**1018–1031.
22. Valledor L, Furuhashi T, Hanak AM, Weckwerth W: **Systemic cold stress adaptation of *Chlamydomonas reinhardtii*.** *Mol Cell Proteomics* 2013, **12:**2032–2047.
23. May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J, Weiss J, Recuenco-Munoz L, Ebenhoh O, Weckwerth W, Walther D: **Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*.** *Genetics* 2008, **179:**157–166.
24. Wolschin F, Weckwerth W: **Combining metal oxide affinity chromatography (MOAC) and selective mass spectrometry for robust identification of *in vivo* protein phosphorylation sites.** *Plant Methods* 2005, **1**(1):9.
25. 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.
26. Rolland N, Atteia A, Decottignies P, Garin J, Hippler M, Kreimer G, Lemaire SD, Mittag M, Wagner V: **Chlamydomonas proteomics.** *Curr Opin Microbiol* 2009, **12:**285–291.

27. Kempa S, Hummel J, Schwemmer T, Pietzke M, Strehmel N, Wienkoop S, Kopka J, Weckwerth W: **An automated GCxGC-TOF-MS protocol for batch-wise extraction and alignment of mass isotopomer matrixes from differential 13C-labelling experiments: a case study for photoautotrophic-mixotrophic grown *Chlamydomonas reinhardtii* cells.** *J Basic Microbiol* 2009, **49:**82–91.
28. Lopez D, Casero D, Cokus S, Merchant S, Pellegrini M: **Algal functional annotation tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data.** *BMC Bioinformatics* 2011, **12:**282.
29. 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.
30. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M: **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant J* 2004, **37:**914–939.
31. Harris EH: *The Chlamydomonas Sourcebook: Introduction to Chlamydomonas and Its Laboratory Use.* Oxford, UK: Academic Press; 2009.
32. Toepel J, Albaum S, Arvidsson S, Goesmann A, Ia Russa M, Rogge K, Kruse O: **Construction and evaluation of a whole genome microarray of *Chlamydomonas reinhardtii*.** *BMC Genomics* 2011, **12:**579.
33. Miller R, Wu G, Deshpande RR, Vieler A, Gaertner K, Li X, Moellering ER, Zauner S, Cornish A, Liu B, Bullard B, Sears BB, Kuo M-H, Hegg EL, Shachar-Hill Y, Shiu S-H, Benning C: **Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen-deprivation predict diversion of metabolism.** *Plant Physiol* 2010, **154:**1737–1752.
34. Lee DY, Park J-J, Barupal DK, Fiehn O: **System response of metabolic networks in *Chlamydomonas reinhardtii* to total available ammonium.** *Mol Cell Proteomics* 2012, **11:**973–988.
35. Msanne J, Xu D, Konda AR, Casas-Mollano JA, Awada T, Cahoon EB, 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.
36. Glinski M, Weckwerth W: **The role of mass spectrometry in plant systems biology.** *Mass Spectrom Rev* 2006, **25:**173–214.
37. Weckwerth W: **Metabolomics in systems biology.** *Annu Rev Plant Biol* 2003, **54:**669–689.
38. Sun X, Weckwerth W: **COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data.** *Metabolomics* 2012, **8:**81–93.
39. Patti GJ, Yanes O, Siuzdak G: **Innovation: Metabolomics: the apogee of the omics trilogy.** *Nat Rev Mol Cell Biol* 2012, **13**(4):263–269.
40. Weckwerth W: **Unpredictability of metabolism - the key role of metabolomics science in combination with next-generation genome sequencing.** *Anal Bioanal Chem* 2011, **400:**1967–1978.
41. Boyle NR, Morgan JA: **Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii.** *BMC systems biology* 2009, **3**(1):4.
42. Ramšak Ž, Baebler Š, Rotter A, Korbar M, Mozetič I, Usadel B, Gruden K: **GoMapMan: integration, consolidation and visualization of plant gene annotations within the MapMan ontology.** *Nucleic Acids Res* 2014, **42**(Database issue):D1167–1175.
43. Galinski E, Herzog R: **The role of trehalose as a substitute for nitrogen-containing compatible solutes (*Ectothiorhodospira halochloris*).** *Arch Microbiol* 1990, **153:**607–613.
44. Kolbe A, Tiessen A, Schluepmann H, Paul M, Ulrich S, Geigenberger P: **Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of ADP-glucose pyrophosphorylase.** *Proc Natl Acad Sci U S A* 2005, **102:**11118–11123.
45. Zhang Y, Primavesi LF, Jhurreea D, Andralojc PJ, Mitchell RAC, Powers SJ, Schluepmann H, Delatte T, Wingler A, Paul MJ: **Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose-6-phosphate.** *Plant Physiol* 2009, **149:**1860–1871.
46. Nägele T, Weckwerth W: **Mathematical modeling reveals that metabolic feedback regulation of SnRK1 and hexokinase is sufficient to control sugar homeostasis from energy depletion to full recovery.** *Front Plant Sci* 2014, **5:**365.
47. Doerfler H, Lyon D, Nagele T, Sun X, Fragner L, Hadacek F, Egelhofer V, Weckwerth W: **Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism.** *Metabolomics* 2013, **9:**564–574.
48. Walther D, Strassburg K, Durek P, Kopka J: **Metabolic pathway relationships revealed by an integrative analysis of the transcriptional and metabolic temperature stress-response dynamics in yeast.** *Omics* 2010, **14:**261–274.
49. Sun X, Weckwerth W: **Using COVAIN to analyze metabolomics data.** In *The Handbook of Plant Metabolomics.* ; 2013:305–320.
50. Gonzalez-Ballester D, Casero D, Cokus S, Pellegrini M, Merchant SS, Grossman AR: **RNA-seq analysis of sulfur-deprived *Chlamydomonas* cells reveals aspects of acclimation critical for cell survival.** *Plant Cell Online* 2010, **22:**2058–2084.
51. Valledor L, Furuhashi T, Hanak AM, Weckwerth W: **Systemic cold stress adaptation of *Chlamydomonas reinhardtii*.** *Mol Cell Proteomics* 2013, **2032–2047:**12(8).
52. Majeran W, Wollman F-A, Vallon O: **Evidence for a role of ClpP in the degradation of the chloroplast cytochrome b6f complex.** *Plant Cell Online* 2000, **12:**137–149.
53. Plumley GF, Schmidt GW: **Light-harvesting chlorophyll a/b complexes: interdependent pigment synthesis and protein assembly.** *Plant Cell Online* 1995, **7:**689–704.
54. Munekage Y, Hashimoto M, Miyake C, Tomizawa K-I, Endo T, Tasaka M, Shikanai T: **Cyclic electron flow around photosystem I is essential for photosynthesis.** *Nature* 2004, **429:**579–582.
55. Rumeau D, Peltier G, Cournac L: **Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response.** *Plant Cell Environ* 2007, **30:**1041–1051.
56. Peltier G, Schmidt GW: **Chlororespiration: an adaptation to nitrogen deficiency in *Chlamydomonas reinhardtii*.** *Proc Natl Acad Sci* 1991, **88:**4791–4795.
57. Fernandez E, Llamas A, Galvan A: **Nitrogen assimilation and its regulation.** In *The Chlamydomonas Sourcebook. Volume 2: Organellar and Metabolic Processes.* 2nd edition. Edited by Harris E. Dordrecht: Elsevier; 2009:69–113.
58. Martin NC, Chiang K-S, Goodenough UW: **Turnover of chloroplast and cytoplasmic ribosomes during gametogenesis in *Chlamydomonas reinhardi*.** *Dev Biol* 1976, **51:**190–201.
59. Wase N, Black PN, Stanley BA, DiRusso CC: **Integrated quantitative analysis of nitrogen stress response in *Chlamydomonas reinhardtii* using metabolite and protein profiling.** *J Proteome Res* 2014, **13:**1373–1396.
60. Sato N, Seyama Y, Murata N: **Lipid-linked desaturation of palmitic acid in monogalactosyl diacylglycerol in the blue-green alga (Cyanobacterium) *Anabaena variabilis* studied in vivo.** *Plant Cell Physiol* 1986, **27:**819–835.
61. Iijima H, Musumi K, Hada T, Maeda N, Yonezawa Y, Yoshida H, Mizushina Y: **Inhibitory effect of monogalactosyldiacylglycerol, extracted from spinach using supercritical CO_2, on mammalian DNA polymerase activity.** *J Agric Food Chem* 2006, **54:**1627–1632.
62. Moellering ER, Benning C: **RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*.** *Eukaryot Cell* 2010, **9:**97–106.
63. Nguyen HM, Baudet M, Cuine S, Adriano JM, Barthe D, Billon E, Bruley C, Beisson F, Peltier G, Ferro M, Li-Beisson Y: **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.
64. Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang X-J, Qi Y: **A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*.** *Genes Dev* 2007, **21:**1190–1203.
65. Owttrim GW: **RNA helicases and abiotic stress.** *Nucleic Acids Res* 2006, **34:**3220–3230.
66. Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B: **PlnTFDB: updated content and new features of the plant transcription factor database.** *Nucleic Acids Res* 2010, **38:**D822–D827.
67. Riano-Pachon DM, Correa LGG, Trejos-Espinosa R, Mueller-Roeber B: **Green transcription factors: a Chlamydomonas overview.** *Genetics* 2008, **179:**31–39.
68. Yuan J, Doucette CD, Fowler WU, Feng X-J, Piazza M, Rabitz HA, Wingreen NS, Rabinowitz JD: **Metabolomics-driven quantitative analysis of ammonia assimilation in *E. coli*.** *Mol Syst Biol* 2009, **5:**302.
69. Martin T, Oswald O, Graham IA: **Arabidopsis seedling growth, storage lipid mobilization, and photosynthetic gene expression are regulated by carbon: nitrogen availability.** *Plant Physiol* 2002, **128:**472–481.

70. Ynalvez RA, Xiao Y, Ward AS, Cunnusamy K, Moroney JV: **Identification and characterization of two closely related β-carbonic anhydrases from *Chlamydomonas reinhardtii*.** *Physiol Plant* 2008, **133:**15–26.
71. Baena-González E, Sheen J: **Convergent energy and stress signaling.** *Trends Plant Sci* 2008, **13:**474–482.
72. Fujii H, Verslues PE, Zhu J-K: ***Arabidopsis* decuple mutant reveals the importance of SnRK2 kinases in osmotic stress responses *in vivo*.** *Proc Natl Acad Sci* 2011, **108:**1717–1722.
73. Kamisaka Y, Tomita N, Kimura K, Kainou K, Uemura H: ***DGA1* (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the delta-*snf2* disruptant of *Saccharomyces cerevisiae*.** *Biochem J* 2007, **408:**61–68.
74. Aukerman M, Simmons CR, Allen SM, Loussaert D, Luck S, Sakai H, Tingey SV: **Plants having altered agronomic characteristics under nitrogen limiting conditions and related constructs and methods involving genes encoding SNF2 domain-containing polypeptides.** United States Patent WO 2010111568 A3.
75. Robaglia C, Thomas M, Meyer C: **Sensing nutrient and energy status by SnRK1 and TOR kinases.** *Curr Opin Plant Biol* 2012, **15:**301–307.
76. Hanson J, Smeekens S: **Sugar perception and signaling.** *Curr Opin Plant Biol* 2009, **12:**562–567.
77. Futai E, Hamamoto S, Orci L, Schekman R: **GTP/GDP exchange by Sec12p enables COPII vesicle bud formation on synthetic liposomes.** *EMBO J* 2004, **23:**4286–4296.
78. James GO, Hocart CH, Hillier W, Chen H, Kordbacheh F, Price GD, Djordjevic MA: **Fatty acid profiling of *Chlamydomonas reinhardtii* under nitrogen deprivation.** *Bioresour Technol* 2011, **102:**3343–3351.
79. Antonny B, Madden D, Hamamoto S, Orci L, Schekman R: **Dynamics of the COPII coat with GTP and stable analogues.** *Nat Cell Biol* 2001, **3:**531–537.
80. Zheng H-Q, Chiang-Hsieh Y-F, Chien C-H, Hsu B-KJ, Liu T-L, Chen C-NN, Chang W-C: **AlgaePath: comprehensive analysis of metabolic pathways using transcript abundance data from next-generation sequencing in green algae.** *BMC Genomics* 2014, **15:**196.
81. Wang Z-Y, Nakano T, Gendron J, He J, Chen M, Vafeados D, Yang Y, Fujioka S, Yoshida S, Asami T: **Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis.** *Dev Cell* 2002, **2:**505–513.
82. Gampala SS, Kim T-W, He J-X, Tang W, Deng Z, Bai M-Y, Guan S, Lalonde S, Sun Y, Gendron JM: **An essential role for 14-3-3 proteins in brassinosteroid signal transduction in *Arabidopsis*.** *Dev Cell* 2007, **13:**177–189.
83. Singh AP, Fridman Y, Friedlander-Shani L, Tarkowska D, Strnad M, Savaldi-Goldstein S: **Activity of the brassinosteroid transcription factors BZR1 and BES1/BZR2 blocks developmental reprogramming in response to low phosphate availability.** *Plant Physiol* 2014, **166:**678–688.
84. Yu X, Li L, Zola J, Aluru M, Ye H, Foudree A, Guo H, Anderson S, Aluru S, Liu P: **A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in *Arabidopsis thaliana*.** *Plant J* 2011, **65:**634–646.
85. Sakamoto W, Zaltsman A, Adam Z, Takahashi Y: **Coordinated regulation and complex formation of YELLOW VARIEGATED1 and YELLOW VARIEGATED 2, chloroplastic FtsH metalloproteases involved in the repair cycle of photosystem II in Arabidopsis thylakoid membranes.** *Plant Cell Online* 2003, **15:**2843–2855.
86. Nixon PJ, Barker M, Boehm M, de Vries R, Komenda J: **FtsH-mediated repair of the photosystem II complex in response to light stress.** *J Exp Bot* 2005, **56:**357–363.
87. Zaltsman A, Ori N, Adam Z: **Two types of FtsH protease subunits are required for chloroplast biogenesis and Photosystem II repair in *Arabidopsis*.** *Plant Cell Online* 2005, **17:**2782–2790.
88. Ogura T, Inoue K, Tatsuta T, Suzaki T, Karata K, Young K, Su L-H, Fierke CA, Jackman JE, Raetz CRH, Coleman J, Tomoyasu T, Matsuzawa H: **Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in *Escherichia coli*.** *Mol Microbiol* 1999, **31:**833–844.
89. Lodha M, Schroda M: **Analysis of chromatin structure in the control regions of the *Chlamydomonas HSP70A* and *RBCS2* genes.** *Plant Mol Biol* 2005, **59:**501–513.
90. Hoehenwarter W, Wienkoop S: **Spectral counting robust on high mass accuracy mass spectrometers.** *Rapid Commun Mass Spectrom* 2010, **24:**3609–3614.
91. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP: **Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors.** *Proc Natl Acad Sci* 2006, **103:**18928–18933.
92. Valledor L, Jorrin J: **Back to the basics: maximizing the information obtained by quantitative two dimensional gel electrophoresis analyses by an appropriate experimental design and statistical analyses.** *J Proteomics* 2011, **74:**1–18.
93. Valledor L, Romero C, Jorrín-Novo J: **Standardization of data processing and statistical analysis in a 2-DE-based comparative plant proteomics experiment.** *Methods Mol Biol* 2013, **1072:**51–60.
94. Valledor L, Escandón M, Meijón M, Nukarinen E, Cañal MJ, Weckwerth W: **A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms.** *Plant J* 2014, **79:**173–180.
95. Furuhashi T, Fragner L, Furuhashi K, Valledor L, Sun X, Weckwerth W: **Metabolite changes with induction of *Cuscuta haustorium* and translocation from host plants.** *J Plant Interact* 2011, **7:**84–93.
96. R Development Core Team: **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing; 2012.
97. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A: **Fiji: an open-source platform for biological-image analysis.** *Nat Meth* 2012, **9:**676–682.

2.4 Targeted quantitative analysis of a diurnal RuBisCO subunit expression and translation profile in *Chlamydomonas reinhardtii* introducing a novel Mass Western approach

My previous research to this PhD-Thesis includes co-authoring the first systematic target proteomics strategy for absolute quantification of proteins in Chlamydomonas reinhardtii (**Wienkoop et al, 2010**), measuring 89 proteins from different metabolic pathways with a dynamic range of 4 orders of magnitude; combined with a subcellular fractionation and two different growth conditions (heteroautotrophic / heteroautotrophic), interesting insights on were achieved on the activation of mitochondrial metabolism in air-adapted and CO_2-limited autotrophic cells or the distribution and relevance of certain enzyme isoforms, such as the CAHs. However, an unexpected conclusion captured our interest: the measured values of both RuBisCO subunits RbLS and RbSS differed considerably from the classical 1:1 stoichiometry of the holoenzyme, as it had been previously suggested/assumed **[338]**.

For the elucidation of the stoichiometry of the RuBisCO synthesis on *Chlamydomonas reinhardtii*, different proteomics approaches are performed on algae grown on heteroautotrophic conditions with a dark/light photoperiod and with samples collected every 3h over a 24h gradient. The Mass Western approach used for absolute protein quantification has been also implemented: critical factors like equimolarity and accurate quantity of the standard peptides have been improved through standard peptides designed with cross-concatenated peptide sequences from both RuBisCO subunits and an equalizer peptide, and including an additional quantifier peptide. To exclude any kind of artifacts in our protein data, the results of the cell extracts were confirmed with additional analyses: prefractionation with SDS-PAGE for both LC-MS/MS, densitometric analysis and Western Blotting, and also through LC-MS/MS measurements on RuBisCO holoenzyme, purified both with BN-PAGE and FPLC. For a more accurate description of the diurnal oscillation of both subunits, the mRNA content for RbcLS and RbcSS was also measured, and finally correlated to our proteomics data.

2.4.1 Declaration of authorship

The results of this chapter are presented in the form of a manuscript published in the journal "Journal of Proteomics". For this work, I contributed to the experimental design, carried out the sampling and all the experiments relative to the proteomics section, collaborated in the data mining and wrote the corresponding "Materials and Methods" section on the manuscript.

2.4.2 Published manuscript

JOURNAL OF PROTEOMICS 113 (2015) 143–153

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Targeted quantitative analysis of a diurnal RuBisCO subunit expression and translation profile in *Chlamydomonas reinhardtii* introducing a novel Mass Western approach

*Luis Recuenco-Muñoz, Pierre Offre, Luis Valledor, David Lyon, Wolfram Weckwerth, Stefanie Wienkoop**

Department of Ecogenomics and Systems Biology, Faculty of Life Sciences, University of Vienna, Vienna, Austria

ARTICLE INFO

Article history:
Received 4 November 2013
Accepted 26 September 2014

Keywords:
RuBisCO
SRM
Mass Western
Chlamydomonas reinhardtii
Targeted protein absolute quantification
Protein complex stoichiometry

ABSTRACT

RuBisCO catalyzes the rate-limiting step of CO_2 fixation in photosynthesis. Hypothetical mechanisms for the regulation of *rbcL* and *rbcS* gene expression assume that both large (LSU) and small (SSU) RuBisCO subunit proteins (RSUs) are present in equimolar amounts to fit the 1:1 subunit stoichiometry of the holoenzyme. However, the actual quantities of the RSUs have never been determined in any photosynthetic organism. In this study the absolute amount of *rbc* transcripts and RSUs was quantified in *Chlamydomonas reinhardtii* grown during a diurnal light/dark cycle. A novel approach utilizing more reliable protein stoichiometry quantification is introduced. The *rbcL*:*rbcS* transcript and protein ratios were both 5:1 on average during the diurnal time course, indicating that SSU is the limiting factor for the assembly of the holoenzyme. The oscillation of the RSUs was 9 h out of phase relative to the transcripts. The amount of *rbc* transcripts was at its maximum in the dark while that of RSUs was at its maximum in the light phase suggesting that translation of the rbc transcripts is activated by light as previously hypothesized. A possible post-translational regulation that might be involved in the accumulation of a 37-kDa N-terminal LSU fragment during the light phase is discussed.

Biological significance
A novel MS based approach enabling the exact stoichiometric analysis and absolute quantification of protein complexes is presented in this article. The application of this method revealed new insights in RuBisCO subunit dynamics.

1. Introduction

During the last decade, quantitative analysis of proteins and their relative changes in abundance in response to environmental perturbation has become routine. Absolute quantification, however, remains challenging. The most accurate method to date is the utilization of the stable isotope dilution technique [1] using isotopic labeled standard peptides of known amounts in combination with the SRM approach. This method has initially been used by Barnidge and colleagues [2] for the targeted

* *Corresponding author.*
E-mail address: stefanie.wienkoop@univie.ac.at (S. Wienkoop).

http://dx.doi.org/10.1016/j.jprot.2014.09.026
1874-3919/

absolute protein quantification of G protein-coupled receptor rhodopsin using SDS-PAGE and was called Mass Western [3,4] for the targeted absolute protein quantification of complex in-solution digests. Different protein isoforms can be distinctively quantified to yield absolute amounts from a complex sample with a single MS analysis [3,5]. However, equimolarity and accurate quantity of the standard peptides remain critical factors for determination of precise protein complex stoichiometry. Holzmann and colleagues [6] described the equimolarity through equalizer peptide (EtEP) strategy based on mTRAQ. For this method, each standard peptide is coupled to an equalizer peptide (EP). Labeling is introduced to peptides after synthesis using mTRAQ. Alternatively, labeled synthesized peptides with amino acid quantitative analysis are available. Here, we introduce a novel Mass Western strategy that uses cross-concatenated synthetic peptides from different RSU protein subunits including an EP and introducing a quantifier peptide (QP).

The Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is the key CO_2-fixing enzyme in photosynthetic eukaryotes. RuBisCO has been described as one of the most inefficient enzymes due to a very low catalytic turnover rate and competing carboxylase and oxygenase activities. RuBisCO therefore limits photosynthetic CO_2 fixation and it is generally accepted that chloroplasts must accumulate large quantities of RuBisCO to sustain sufficient rates of carbohydrate synthesis [7]. Besides its role in carbon assimilation this enzyme complex became of main focus for its plant nitrogen (N) use efficiency [8]. RuBisCO degradation into amino acids leads to N re-utilization for the de novo protein synthesis.

RuBisCO consists of eight large and eight small protein subunits in both higher and lower plants. The large subunit (LSU) is encoded by the *rbcL* gene localized in the chloroplastic genome while several variants of the small subunit (SSU) are encoded in the nuclear genome by the *rbcS* gene family. The LSU and SSU subunits are assumed to be present in equimolar quantities within plant cells to match the one-to-one RuBisCO subunit stoichiometry [9]. The localization of RuBisCO genes in two different cellular compartments and the assumed one-to-one stoichiometric balance of its subunits suggest the existence of mechanisms to tightly coordinate the expression of *rbcL* and *rbcS* genes [10]. Although the coordinated synthesis of the subunits has been observed [11], the absolute concentrations of LSU and SSU have never been determined. The absolute concentration of LSU and SSU at a certain time point integrates several parameters such as the stability as well as the rates of synthesis and decay of each polypeptide and is therefore crucial to a quantitative understanding of RuBisCO gene expression.

Diurnal cycling of LSU and SSU cellular concentration with peak abundance levels during the photoperiod has been recently suggested by an analysis of protein relative expression change in rice leaves [12]. However, a transcript study [13] reported that in *Arabidopsis*, grown in a light/dark photoperiod, *rbcS* mRNA exhibits a diurnal pattern of expression, with peak abundance occurring soon after onset of light and minimum levels at the end of the light period. Thus, diurnal oscillation of relative *rbcS* mRNA levels in *Arabidopsis* occurred in an inverse time frame to the SSU oscillation in rice leaves. Consistently with this observation, several studies of *rbcL* gene expression in *Chlamydomonas reinhardtii* [14–16] indicated that *rbcL* mRNA levels are not directly correlated to the amount of LSU. However, the diurnal variations of the absolute amount of *rbcL* and *rbcS* transcript and protein have never been characterized altogether in any plant model.

The aim of this study was to determine the absolute amounts of *rbcL* and *rbcS* transcripts and proteins in *C. reinhardtii* grown during a diurnal time course. The green algae *C. reinhardtii* is an interesting model system for the study of the photosynthetic apparatus, because mixotrophic cells under low light have been shown to achieve growth rates similar to photoautotrophic counterparts when supplied with acetate [17]. The particular strain CC503 was used to investigate phototaxis and biofuel production and its genome sequence [18] was used to reconstruct a genome-scale metabolic network [19].

We present a novel conceptual approach of the Mass Western [5] that allowed us to determine absolute protein concentrations with accurate stoichiometry as well as diurnal protein translation changes of the different RuBisCO subunit proteins (RSUs).

2. Materials and methods

2.1. *Chlamydomonas* growth conditions and cell sampling

The cell wall deficient strain *C. reinhardtii* CC503 *cw92 agg1+*, *nit1*, *nit2*, mt+ (originally derived from strain CC-125), was obtained from the *Chlamydomonas* Center. Synchronic *C. reinhardtii* CC503 cultures were cultivated in 250 mL flasks at 24 °C, stirred at 110 rpm on an orbital shaker (New Brunswick 44R) and exposed to 100 mmol m^2 s^1 light (Sylvania GroLux) and a light/dark cycle of 14/10 h. Cells were grown in fresh HAP medium (TAP media in which TRIS was replaced by 5 mM HEPES, and supplemented with 8 mM NH_4Cl) for four days and supplied with air without additional CO_2 before sampling. For mixotrophic growth condition HAP medium supplemented with 10 mM NaAc was employed [4]. Samples were taken at early log phase (cell density ~ 1.5×10^6 cells/mL) every 3 h for 24 h. For an overview of the experimental setup see also Fig. 1.

2.2. Total protein extraction and in solution digestion

C. reinhardtii cell pellets (1.0 g fresh weight each) were ground in a chilled mortar using liquid N. Extraction buffer (For in solution digestion: 50 mM Tris–HCl, pH 8.0, 5 mM dithiothreitol, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF) - For SDS-PAGE: 62,5mM TrisHCl, pH 6.8, 2% SDS (w/v), 10% Glycerol 10% (v/v), 0,01M DTT, 2mM PMSF, PI 0,2%) was added and crude protein extracts were then centrifuged at 10,000 ×*g* for 10 min. For in solution digestion, the supernatants were immediately desalted on a Sephadex G-25 column (1.5 cm), previously equilibrated with 50 mL Tris–HCl, pH 8.0. For SDS-PAGE, the supernatants were precipitated using overnight incubations in 10 volumes of chilled Acetone (0.5%-ß-Mercaptoethanol), washed three times with chilled methanol (0.5%-ß-Mercaptoethanol), air-dried, and kept at –20°C until further SDS-PAGE analysis. Concentration of total protein extracts was measured according to Bradford [20].

Aliquots containing 25 μg of soluble *C. reinhardtii* proteins were digested overnight at 37 °C with Poroszyme immobilized trypsin

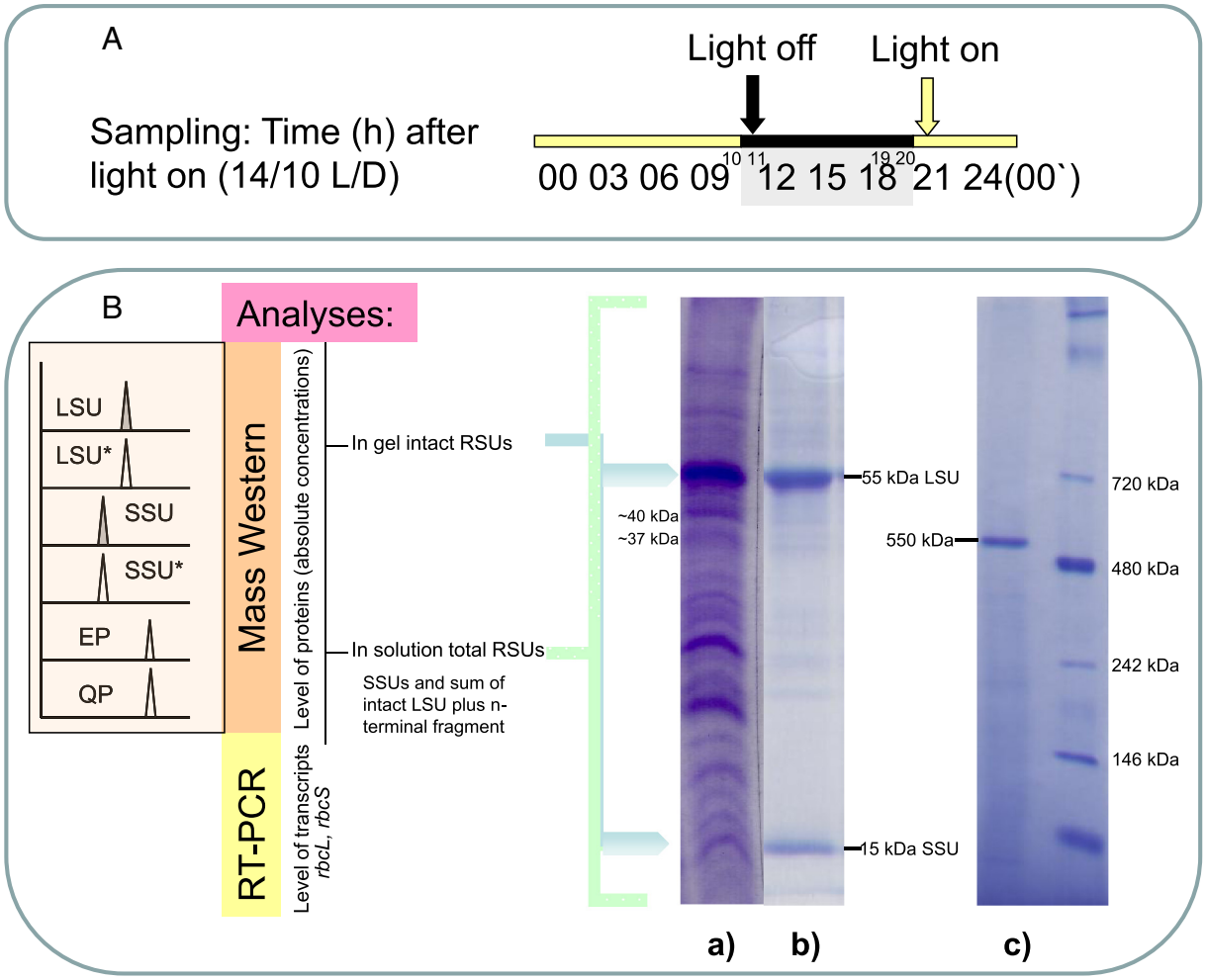

Fig. 1 – Overview of the experimental setup. A. Sampling strategy B. Analytical structure; The Mass Western is schematically depicted and used for abs. quant. of the 55- and 15 kDa RuBisCO subunit gel-bands of gel a) SDS gel of total protein extract and b) SDS gel of RuBisCO holoenzyme after size exclusion (gel lane was chosen from time point 12 as an average example); and the abs. quant. of the 550 kDa gel-band of c) the BN gel of the RuBisCO holoenzyme plus NativeMarker™ (Novex®). Gel-band at 37 kDa (a) was separately extracted for protein Identification on MS (Supplemental Table 2). Mass Western was additionally used for gel-free abs. quant. of the RuBisCO subunits within the crude extract. EP = equalizer peptide; QP = quantifier peptide; LSU = large subunit; SSU = small subunit; * = isotopic labeled.

beads (1:10, v/v, Applied Biosystems, Darmstadt, Germany). The peptide mixtures were centrifuged at 6000 ×*g* and desalted using SPEC C18 columns (Varian, Darmstadt, Germany) according to the manufacturer's instructions. Desalted digest solutions were dried and pellets stored at −20 °C until use. Three biological replicates were analyzed per time point each replicate consisting of 1 μg total spiked with 0.5 pmol standard peptides. Protein concentration of the extract was measured using Bradford's assay [20].

2.3. RuBisCO holoenzyme extraction

For the isolation of the intact RuBisCO holoenzyme complex, a Size Exclusion (SE) FPLC and a Blue Native (BN) gel approach were used.

The SE protocol of Olinares et al. [21] was used with following modifications: The cell pellets were ground in a chilled mortar using liquid N under addition of extraction buffer (50 mM HEPES, pH 8.0, 5 mM $MgCl_2$, 1 mM NaCl, 1 mM PMSF) and then homogenized with a 7 mL glass homogenizer. Extracts were centrifuged at 4 °C for 10 min (10,000 ×*g*) to pellet cell debris and supernatant protein concentration was measured according to Bradford [20]. Supernatants were sterile-filtered through 45 μm cut-off cellulose acetate filters (VWR) and desalted with Mini-Trap G-25 Sephadex columns (GE-Healthcare, previously equilibrated with 10 volumes of extraction/elution buffer). Total protein concentration was measured a second time using Bradford's assay [20] and extracts were centrifuged for 5 min at 13,000 ×*g* before Fast Protein Liquid Chromatography (FPLC) SE.

The extract (400 μg each) was loaded onto a Superdex 200 10/300 GL column (GE Healthcare) at a flow rate of 0.5 mL min^{-1} with an Äkta FPLC System (Amersham). The elution was monitored at an absorbance of 280 nm and 27 fractions of 300 μL were collected. The fractions containing the RuBisCO (~550 kDa) were immediately precipitated using overnight

incubations in 10 volumes of chilled Acetone (0.5%-ß-Mercaptoethanol), washed once with chilled methanol (0.5%-ß-Mercaptoethanol), air-dried, and kept at −20 °C until further SDS-PAGE analysis.

The BN protocol of Fiala and colleagues [22] was used with following modifications: after the first centrifugation (1500 ×*g*, 4 °C), the cell pellets were washed once in TBS-buffer, and homogenized with a 7 mL glass homogenizer after adding 500 μL of extraction buffer (20 mM BisTris, 500 mM Aminocaproic acid, 20 mM NaCl, 2 mM EDTA, 10% Glycerol, 1 mM PMSF). Extracts were centrifuged at 4 °C for 15 min (13,000 ×*g*) to pellet cell debris, filtered with 45 μm and kept at 4 °C until subsequent dialysis over night at 4 °C with 10 μm-cutoff membranes (extraction buffer was used as dialysis buffer). After dialysis, protein concentration was measured according to Bradford [20]. Finally, 10 μg of each sample was loaded and separated in a 4%–15% gradient BN-Gel, according to Fiala et al. [22].

2.4. SDS-PAGE and digestion of the RuBisCO subunits and holoenzyme

Pellets of total protein extracts and FPLC size exclusion fractions were re-dissolved in SDS-loading buffer and separated by SDS-PAGE on 12.5% (w/v) polyacrylamide gels. After electrophoresis, proteins were visualized using Gel-Code Blue Stain reagent (Pierce Biotechnology, Rockford, USA) and densitometric analysis of the RuBisCO subunit stoichiometry was performed with ImageJ software (http://rsb.info.nih.gov/ij/index.html). Gel bands corresponding to a molecular weight of approximately 55 kDa for the LSU and 15 kDa for the SSU as well as the 560 kDa BN band of the RuBisCO holoenzyme were cut out and digested for determining the absolute amount of the RuBisCO subunits. In-gel digestion was carried out as previously described [3]. Digests of bands taken from the same gel lane were combined and spiked with standard synthetic peptides prior to desalting. A total number of three biological replicates were analyzed per time point.

2.5. Western Blot analysis

The above-mentioned SDS-gels were additionally used for Western Blot analysis of the RSUs. For the immunostaining of both RuBisCO subunits, acetone-precipitated pellets of total protein extracts as well as holoenzyme fractions were re-dissolved in SDS-loading buffer and separated by SDS-PAGE on 12.5% (w/v) polyacrylamide gels as described before. After electrophoresis, proteins were electrotransferred to a Millipore Immobilon PVDF membrane (16 V, 80 mA, 1 h), blocked with 5% milk on TBS-T buffer (no previous Ponceau-staining used for band preview), and incubated over night at 4 °C with two primary antibodies: against RbcS (AS07 259, Agrisera) diluted 1:2500 in milk, and against RbcL (AS03 037, Agrisera) diluted 1:5000 in milk, both of them being rabbit polyclonal and KLH-conjugated. Subsequently, the membranes were incubated with conjugate Rabbit F(ab′)2 Fragment ECL-HRP antibodies (NA9340, GE Healthcare) at 1:5000 dilution in milk for 1 h at room temperature. Blots were developed with the ECL Prime Western Blotting Detection Reagent RPN2232 (GE Healthcare) and exposed to X-ray film (Fuji) (Supplemental Fig. 1).

2.6. Enhanced Mass Western approach for protein stoichiometry analysis

The strategy to determine subunit stoichiometry of multi-protein complexes as reported by Holzmann et al. [6] was adapted to the Mass Western approach [5] and further optimized as follows: Stoichiometric quantification of the absolute amounts of subunits is achieved by the use of cross-concatenated internal standards that deliver equimolar amounts of synthetic peptide standards for all subunits of the investigated protein complex upon tryptic digestion [6]. Two cross-concatenated proteotypic standards, labeled with heavy isotopes ($^{13}C/^{15}N$) (AQUAPoP 3H Ultimate, Thermo Electron, Ulm, Germany) were used in this study (Table 1). Each standard consisted of one RuBisCO LSU-specific mono-labeled peptide, one RuBisCO SSU-specific mono-labeled peptide (present in both SSU1 and SSU2) and one double-labeled non-target (i.e. absent from the proteome of *C. reinhardtii*) peptide. The two concatenated standards were made of different LSU- and SSU-specific peptides (cross-concatenated) but identical non-target peptides. Non-target peptide moieties were used as equalizer peptide (EP), allowing normalization between the two concatenated internal standards.

Absolute quantification of the concatenated standards and protein complex subunits is enabled by direct comparison to a third internal standard (i.e. quantifier peptide), consisting of a mono-labeled version of the EP, which was quantified with high accuracy (AQUA Ultimate, Thermo Electron, Ulm, Germany).

The concatenated peptide standards (approximately 0.5 pmol), and the quantitative standard peptide (0.5 pmol) were added to every sample, as this amount is within the linear range of detection [23] and in accordance within the estimated (high) abundance of the RuBisCO subunits.

2.7. LC-Orbitrap-MS RuBisCO holoenzyme analysis

A one-dimensional (1D) nano-flow LC-MS/MS system (UltiMate 3000, Thermo Fisher Scientific, Austria) was employed to detect and quantify the amount of RuBisCO subunits. Peptide mixtures (RuBisCO holoenzyme from SE plus standard peptides) were eluted using an Easy-Spray RSLC PEPMAP® C18 column (15 cm × 50 μm, 2 μm; Thermo Scientific) during a 20 min gradient from 2% to 50% (v/v) acetonitrile containing 0.1% (v/v) FA with a controlled flow rate of 0.3 μL min^{-1}. MS analysis was performed on an Orbitrap Elite mass spectrometer (Thermo Scientific) operated in the positive profile mode within a mass range from 350 to 800. One dependent (MS2) scan was enabled for each full scan with collision energy of 35. The minimal signal threshold was set to 10,000. Spray voltage was adjusted to 1.6 kV and temperature of heated transfer capillary

Table 1 – Average ratio between LSU and SSU protein and rbcL and rbcS transcript levels across a diurnal time course. a) total amount of crude extract; b) intact RSU gel bands; c) RuBisCO holoenzyme (BN); d) transcripts; Standard error n = 3, t-test p values given when p < 0.05.

a) LSU/SSU of crude samples	4.8 ± 1.5; p = 0.00043
b) LSU/SSU of crude SDS gel samples	1.3 ± 0.4; ns
c) LSU/SSU of BN gel samples	1.5 ± 0.08; ns
d) rbcL/rbcS	5.0 ± 0.9; p = 0.000012

to 200 °C. Same settings were used for the 37 kDa SDS gel digest of the crude extract (Supplemental Table 2).

2.8. LC-triple quadrupole (TripleQuad)-mass spectrometry (MS) and selective reaction monitoring (SRM)

LC system and gradient were as described above with some modifications: Peptide mixtures were eluted using a Peptide ES-18 column (15 cm × 0.1 mm, 2.7 μm; Sigma-Aldrich) during a 20 min gradient from 2% to 50% (v/v) acetonitrile containing 0.1% (v/v) FA with a controlled flow rate of 0.4 μL min^{-1}. MS analysis was performed on a TSQ Quantum Vantage mass spectrometer (Thermo Scientific) operated in the positive mode and tuned to its optimum sensitivity for each standard peptide as previously described [4]. Scan width for all SRMs was 0.7 mass units, and resolution was set to 0.3 and 0.7 mass units for Q1 and Q3, respectively. The dwell time per transition was 50 ms, collision gas pressure was set to 1.5 mTorr, the declustering voltage to 5 V and the chrome filter peak width to 20 s. Spray voltage was adjusted to 1.8 kV and temperature of heated transfer capillary to 270 °C. Collision energies used for recording transitions are summarized in Supplemental Table 1.

2.9. Transcript analysis

Nucleic acids were extracted from *C. reinhardtii* cell pellets (1.0 g fresh weight) using a protocol based on the method of Tourna and colleagues [24]. Cells were disrupted using a bead beating approach, an SDS-based lysis buffer and a phenol/chloroform/isoamyl-alcohol solution, as described previously. Precipitation of nucleic acids was performed for 2 h at 4 °C with the addition of one volume of isopropanol, 5 M NaCl (final salt concentration 0.2 M) and glycogen. Precipitated nucleic acids were resuspended in nuclease free water (Qiagen, Hilden, Germany) and then purified with an RNeasy minikit (Qiagen). DNA was removed from the nucleic acid extracts by digestion with 2.5 U of DNase I (Fermentas, St. Leon-Rot, Germany) in 10 μL reactions which were incubated for 1 h at 37 °C. The extracts were further purified using an RNeasy minikit (Qiagen). Reverse transcription of the purified RNA templates was performed with random hexamers using SuperScript® III reverse transcriptase (LifeTech — Invitrogen, Wien, Austria) according to the manufacturer's instructions. Control reactions containing *C. reinhardtii* RNA templates and all reagents except the reverse transcriptase were performed to ensure that no DNA was present in the purified RNA extracts. An additional control reaction without template was performed to check for the presence of contaminants in the reagents.

Two new primer pairs were designed to target fragments of *C. reinhardtii rbcL* gene sequence and another two primer pairs, fragments of the concatenated *rbcS* gene exon sequences. PCR primers were designed using Primer3-web 0.4 [25] and *C. reinhardtii* CC503 cw92 mt+ *rbcL*, *rbcS1* and *rbcS2* gene sequences as input sequences. Specificity of the primer pairs was confirmed by BLAST searches against *C. reinhardtii* genome sequence and by characterizing the length of PCR-generated amplicons using agarose gel electrophoresis analysis. Primer pairs designed to match *rbcS* exon sequences target both *rbcS1* and *rbcS2* sequences.

Quantification of reverse transcribed *rbcL* and *rbcS* mRNAs (*rbcL* and *rbcS* cDNA) was performed using a SybrGreen I quantitative PCR approach. Amplification of *rbcL* cDNA fragments was performed with primer Cr-rbcL-1059F (5′-CGTTGAAAAGACCGTAGCC-3′) and Cr-rbcL-1215R (5′-ACCACCACCGAACTGAAGAC-3′) and duplicate dilution series (10^2–10^7 *rbcL* gene copies) of a 938 bp *C. reinhardtii rbcL* gene fragment generated with primer Cr-rbcL-441F (5′-ATTCGTAGGTCCTCCACACG-3′) and Cr-rbcL-1378R (5′-CACATGCAGCAGCAAGTTCT-3′) was used as a standard with an efficiency of 93% and R^2 value of 0.995. Amplification of *rbcS* cDNA fragments was carried out with primer Cr-rbcS-205F (5′-GACAACCGCTACTGGACCAT-3′) and Cr-rbcS-353R (5′-ATCTGCACCTGCTTCTGGTT-3′) and duplicate dilution series (10^2–10^7 *rbcS* gene copies) of a 416 bp *C. reinhardtii rbcS1* gene fragment and a 450 bp *C. reinhardtii rbcS2* gene fragment, both generated with primer Cr-rbcS-1069F (5′-ACAAGGCCTACGTGTCCAAC-3′) and Cr-rbcS-1444R (5′-GATCTGCACCTGCTTCTGGT-3′), were used as standard with an efficiency of 86% and R^2 value of 0.997. All reactions (25 μL) contained 12.5 μL of 2× Quantifast SYBR Green PCR Master Mix (Qiagen), 0.2 mg mL^{-1} BSA, 1 μM of each primer and 5 μL of cDNA template. Amplification conditions were 95 °C for 15 min, followed by 35 cycles of 15 s at 95 °C, 30 s at 60 °C and plate read at 78 °C and a final extension step of 10 min at 60 °C. Temperature cycling of the reactions was carried out in a Mastercycler® RealPlex^2 thermocycler (Eppendorf, Wien, Austria) and the specificity of the amplification was assessed by melting curve analysis at the end of each PCR run and agarose gel electrophoresis. For normalization, total protein sample amounts of 5.3 mg ± 12% (standard deviation) and 5.5 mg ± 17% during light and dark phase were used, respectively.

2.10. Statistics

For regression analysis, calculation of standard deviation, standard error and Student's t-test Excel (Microsoft Office 2007) was used. Settings for the t-test were 2 sites, type 1. Anova was carried out using the MatLab tool Covain [26]. Statistical significance was defined with a p value ≤ 0.05.

3. Results

3.1. RuBisCO holoenzyme stoichiometry after BN and SE

The fully sequenced *C. reinhardtii* genome revealed one putative transcript of the *rbcL* and two for the *rbcS*. The two SSU coding sequences are highly similar (98%). Thus, no proteotypic peptides could be selected for the differentiation of the two SSUs. The protein abundance of SSU is therefore a sum of both isoforms (if both were present). For technical verification, two RuBisCO holoenzyme purification strategies were conducted prior to stoichiometric analysis (Fig. 1B) using either a BN or SE approach [21,22]. For the highly abundant and purified RuBisCO holoenzyme, a direct LC–MS/MS Orbitrap based Mass Western approach (without SRM) was conducted. Ion traces of the target peptides were extracted from the chromatogram (see also Fig. 1B) and concentrations calculated

from the peak areas. The ratio for the LSU vs. SSU was found to be around 1:1 (Table 1c).

Additionally, densitometric analyses of the subunit stoichiometry were performed using the 2D SE × SDS-gel images of the purified intact RSUs (Fig. 1B). Here also an average ratio of about 1:1 (1.3 ± 0.03) was observed without further normalization to the protein size. No diurnal oscillation was observed for the purified holoenzyme (data not shown).

3.2. *RuBisCO large and small subunit transcript and protein abundances are not stoichiometric equivalent across the diurnal cycle: large subunit on average fivefold more abundant*

For absolute protein quantification of the RuBisCO subunits of a crude (total protein) extract the novel SRM based Mass Western approach of improved accuracy was chosen [27] and two different digestions were performed (Fig. 1a): i) an in-solution digestion of the crude extract and ii) an in-gel digest of the two intact RuBisCO subunit (15 and 55 kDa) bands of the crude extract. Independently, transcript (Fig. 3) as well as in solution protein absolute amounts (Fig. 2A) showed a statistically significant average 5:1 fold ratio between the large vs. small subunit(s) of RuBisCO (RSU and *rbc* ratio t-test $p < 0.05$) across the diurnal time course (Table 1). In contrast the ratio of intact RSUs of the in-gel digest of the crude extract showed nearly a 1:1 (1.3:1) stoichiometry (Fig. 2B, Table 1) as well as 1.5:1 for the BN-gel digest of the holoenzyme (Table 1).

Mass Western showed that LSU reached a maximum amount of 192 and 902 fmol/µg total protein extract of the in-gel (intact 55-kDa subunit) and in-solution digest (presumably intact 55-kDa subunit plus breakdown fragments), respectively. Note: 100 fmol equals 5 ng and 1.5 ng of LSU and SSU, respectively. Thus, the difference between intact amounts of LSU (in-gel 15 and 55 kDa bands) to total LSU amount (in-solution) may explain the amount of LSU N-terminal fragment(s), although two different extractions are compared and the 37 kDa-band has not been quantified separately. SSU amounts only slightly oscillated between 51 and 212 fmol/µg total protein extract both in-gel and in-solution (Fig. 2A and B).

3.3. *The diurnal transcript-to-protein pattern of RuBisCO small and large subunits shows best correlation within a 9 h shift*

The protein pattern analyses demonstrated a pronounced change in abundance during the diurnal time course for the in-solution total LSU protein (intact plus fragment) (Fig. 2). LSU abundance of the complex protein extract revealed an increase in protein abundance during the light phase and a second increase during the end of the dark phase (statistically significant compared to SSU). For the SSU the diurnal changes seem to be less pronounced (Fig. 2A). Here RSU oscillations do correlate with the regression $R^2 = 0.664$. Protein patterns of intact LSU and SSU of the crude extract with 1:1 stoichiometry exhibit a trend to positive correlation of $R^2 = 0.568$ (Fig. 2B). Data of the diurnal BN holoenzyme analysis are almost identical and thus not presented.

The transcript analysis revealed possible changes with a general decrease during the light phase and an increase during the dark phase, however statistically not significant (Fig. 3). Nevertheless, compared to the time course of the protein levels no good correlation was observed, the best transcript-to-protein fit was found by shifting protein abundance of the total RSUs for 9 h (Fig. 4). Here, the regression for

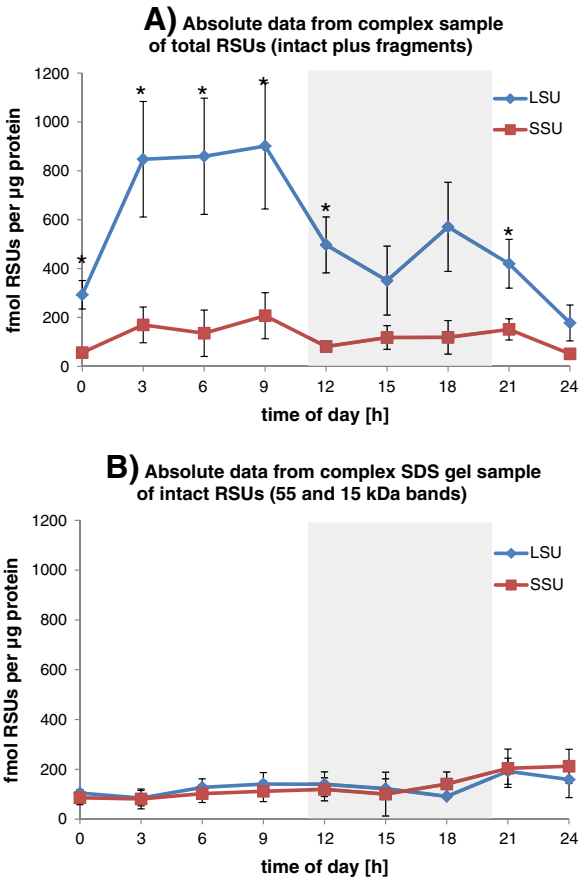

Fig. 2 – Diurnal time course of intact and complete RuBisCO subunit concentration levels of the crude extract. Mass Western analysis of (A) RSUs from total soluble protein extract, (B) of intact RSU protein bands from SDS-polyacrylamide gel electrophoresis of the crude extract. Standard error n = 3; Anova $p < 0.05$ marked with asterisks.

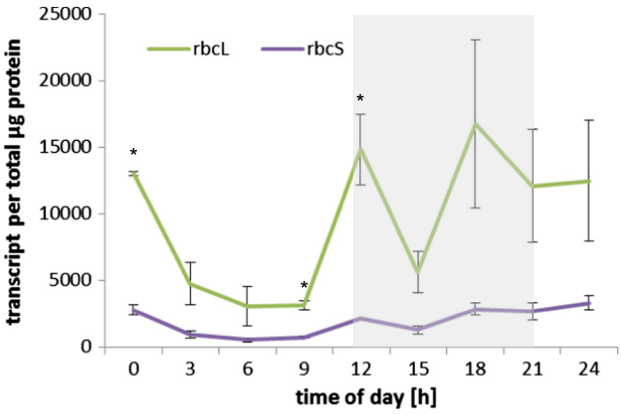

Fig. 3 – Diurnal time course of RuBisCO transcripts normalized to the total protein content of the sample. Standard error n = 3; Anova $p < 0.05$ marked with asterisks.

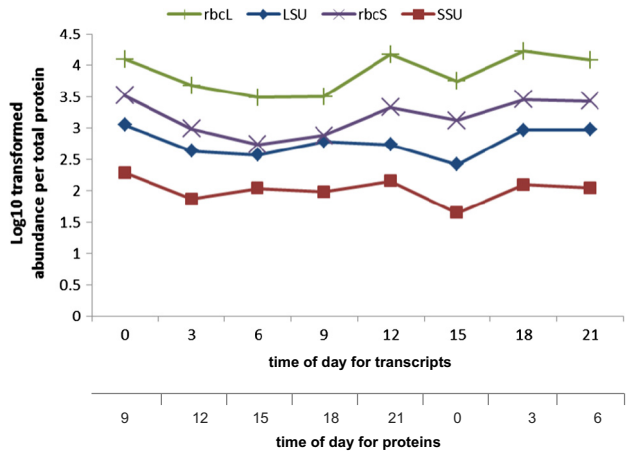

Fig. 4 – RuBisCO transcript-to-protein correlation. Normalized diurnal transcript and protein data were log_{10} transformed and in-all plotted with a time shift of 9 h of best data fitting (no significant correlation).

rbcS vs. SSU was $R^2 = 0.474$ and for rbcL vs. LSU $R^2 = 0.464$. There was no correlation (best pos. fit 0.251) found between transcripts and intact RSUs. This might be due to the fact that intact RSUs seem not to have a significant diurnal oscillation. The average transcript-to-protein conversions were determined as 60 amol intact RSU per transcript and 85 amol total RSU per transcript with a high variance of around 50% between day and night phases.

4. Discussion

4.1. *Technical challenges for the correct calculation of RuBisCO subunit dynamics*

Understanding the dynamics of proteins and protein complexes is becoming more and more important. This includes protein turnover analysis and the ability to accurately measure the stoichiometry of protein complexes and frequency of their related subunits. The dual role of RuBisCO, both as the enzyme responsible for CO_2-fixation and as N storage protein [8,28], is the reason why the quantification of changes in RuBisCO pool size, and the determination of underlying mechanisms, are of major interest [29]. The most common hexadecameric form of the RuBisCO holoenzyme with a subunit stoichiometry of the LSU and SSU of 1:1 is also present in *C. reinhardtii* [30]. It consists of eight 55-kDa chloroplast-encoded catalytic LSUs and eight 15-kDa nuclear encoded non-catalytic SSUs that can be separated by gel-electrophoresis. In various proteomic studies, the authors calculated RuBisCO subunit stoichiometry using different methods. Peltier and colleagues [31] calculated the relative abundance of the subunits by normalizing the spot densities after 2D gel analysis of enriched *Arabidopsis thaliana* chloroplast proteins by dividing it with the molecular masses of the proteins. In fact, this way they reached a stoichiometry of one-to-one for the RuBisCO subunits. However, density calculations of our 2D (SE × SDS-PAGE) purified holoenzyme without further normalization to molecular masses already resulted in a ratio (LSU/SSU) of about one-to-one (Fig. 1Bb). Thus we assume that further normalization of 2D gel spot or band densities or density calculations for stoichiometric analyses in general appears critical. Also recently, Mastrobuoni and colleagues [32] proposed the stoichiometry of RuBisCO protein of a *C. reinhardtii* crude extract close to 1:1. They used a modified emPAI (Exponentially Modified Protein Abundance Index; [33]) for the relative quantification of the protein. The emPAI is based on the normalization of the number of identified peptides (spectral count) of a protein by its number of possible peptides. This is similar to the NSAF (Normalized Spectral Abundance Factor) [34]. Based on the spectral count, both are relatively good approximations for relative stoichiometry calculations of different proteins within a sample. However, in the above-mentioned study, peptide intensities were used for normalization. Due to the diverse ionization properties of each peptide, intensities (or areas) of peptide spectra – especially when comparing different proteins – can only be used for relative or absolute quantification if combined with the stable isotope dilution technique [1]. It is the incorporation of heavy standard peptides (usually of known amounts) into native samples that enables normalization of different peptide signals for the same or different proteins. Consequently, only this allows for the peptide signal intensity based comparison of different proteins within the same or different samples.

Because of the above-mentioned examples, uncertainty seems to persist on the correct interpretation of protein abundance data. Summarizing the methods described the absolute quantification of proteins from complex samples using direct shotgun-LC-TripleQuadMS and selective reaction monitoring (SRM) without pre-fractionation, also called Mass Western [4,5,35] is to the best of our knowledge the most accurate approach to date. Nevertheless, some challenges remain concerning accurate stoichiometric analysis using Mass Western or related techniques. Due to diverse solubilization and digestion properties, exact absolute quantification and stoichiometry of proteins may remain deficient. To overcome this, we developed an enhanced Mass Western approach for improved accuracy of stoichiometric analysis.

Novelty of the Mass Western: Compared to the previous Mass Western approach for absolute quantification, the novel Mass Western comprises of several new features providing the means for very accurate stoichiometric protein analysis. We now introduced a) concatenated synthetic peptides, b) a singly labeled EP, and c) a QP. The Mass Western is different to QconCAT [36] since concatenated signature peptides are not encoded by genes (and expressed as recombinant proteins) but simply synthesized. They are cross-concatenated as explained above. In comparison to Holzmann and colleagues [6], Mass Western uses stable isotope labeled peptides of differentially labeled amino acids. This way, the same peptide sequence can be used as equalizer for several concatenated peptides. For price efficiency, only the quantifier peptide is precisely quantified by amino acid analysis. This peptide can be used for several hundreds of analyses and peptide combinations. A drawback might be the peptide length. Only a limited number of amino acids (<40) with suitable hydrophobicity can be used for peptide synthesis (synthesis feasibility check with the company is recommended). It should be mentioned, that the novel described improvements of the

Mass Western are directed to enhance accuracy of the stoichiometry between different proteins not the absolute concentrations itself. Sample handling such as desalting might still cause unspecific sample loss, leading to a loss in absolute concentration. However, for stoichiometry analysis sample handling is identical since protein complex is measured from the very same sample. Thus, differences due to technical artefacts may be excluded. Comparing Mass Western and Western Blot, both techniques are challenging: While less sample handling is necessary for Western Blotting, antibody availability/specificity and reliability of densitometric analysis of quantities are a major drawback compared to stable isotope dilution based MS approaches [37].

Noteworthy, most studies on RuBisCO dynamics are based on the determination of RuBisCO content and turnover analysis using a colorimetric measurement of protein stained with Coomassie brilliant blue R on SDS gel electrophoresis by eluting with formamide [38–41]. They used stained RuBisCO subunit bands from SDS gel, using calibration curves made with RuBisCO purified from wheat or rice leaves. Our data, however, reveal a difference in analyzing intact subunit ratios compared to the total amount in crude sample.

A recent study on the absolute quantities of *C. reinhardtii* proteins using the Mass Western revealed strong differences in the translation stoichiometry between the RuBisCO subunits [4]. Unexpected, the LSU was found several folds higher in abundance compared to the SSU. In the present work, we want to refine and try to explain these findings using Mass Western. In addition, we analyzed absolute transcript and protein levels to test previously described non-correlative or inverse diurnal transcript-to-protein patterns [13].

4.2. The differential RuBisCO subunit dynamics

It has been suggested that interplays between LSU and the SSUs are crucial for the function of the enzyme by maintaining its active conformation [42,43]. Thus, for assembly of the RuBisCO holoenzyme, translations of the RSUs are discussed to have concerted regulation. On the other hand, the lack of SSU in *Rhodospirillum*, led to the conclusion that LSU translational regulation must be SSU independent [9].

Data show a RuBisCO holoenzyme ratio of nearly 1:1 (Fig. 2b and Table 1). Nevertheless, deviation between our 1.5:1 holoenzyme stoichiometry to the general known 1:1 stoichiometry can either be explained by a selective loss of SSU during sample handling or by another yet to identify reason. Note: an almost identical stoichiometric result of the holoenzyme can be obtained with and without purification. RuBisCO LSU always shows a trend to higher ratios.

Interestingly, ratios between transcripts as well as large and small protein subunits of the crude extract are on average about 5 fold higher for the large subunit, independently (Figs. 2a, 3 and Table 1). The significantly higher LSU level of the crude extract can, at least to a high extent be explained by the presence of a previously described 37-kDa fragment [28–31]. A mass spectrometry based gel band digest analysis of the crude extract that revealed a large peptide sequence coverage of RuBisCO at 37 kDa (see Supplemental Table 2). Western Blot analysis of crude extract as well as the purified holoenzyme fraction (Supplemental Fig. 1), however, did not visualize the 37-kDa fragment. This is in agreement with several other publications, where Western Blot analysis of *C. reinhardtii* did not capture any specific (or artificial) breakdown products at all [32–34].

The Coomassie blue stained 37 kDa as well as 15 kDa gel bands of the crude extracts appear much weaker than the 55 kDa band. This observation is well known and explained by the fact that smaller proteins bind less dye and more importantly because several proteins also co-resolve with e.g. RuBisCO in a 1D-gel band [35]. This is the reason, why Coomassie blue staining may not be valid for stoichiometric analysis. Our results thus suggest a strong degradation of LSU that may either be caused artificially during sample preparation or might be post-translationally regulated and thus of biological relevance. MS based analyses of in-gel digests of the 37 kDa bands, however, only identified RuBisCo in the crude extracts — not in the gels of purified in vivo samples. These results may suggest a specific rather than artificially caused breakdown. Furthermore, transcript and protein data suggests that hexadecameric holoenzyme formation is limited by the SSUs which has never been shown before. These data are from mixotrophic algae and there are indications for similar ratios in photoautotrophically grown *C. reinhardtii* cells [4], although, the absolute amount of SSU was significantly higher in photoautotrophic compared to mixotrophic cells.

Pyrenoids are specific algae organelles located in the plastids. It is the site of elevated CO_2 channeled by the carbon concentration mechanism (CCM). The SSU has been shown to be involved in the formation of the pyrenoid [10]. The rather slow catalysis and confused specificity [44] of RuBisCO seem to be improved by two surface-exposed α-helices of SSU targeting the holoenzyme to the pyrenoid of *C. reinhardtii* [45]. A higher concentration of SSU might thus be crucial for photoautotrophic organisms.

The diurnal time course revealed a possible inverse oscillation with maximum values for transcripts during the dark phase and a maximum increase especially for the total LSU abundance during the light phase with a slight increase also during dark phase (Fig. 2). In chloroplasts of *C. reinhardtii*, Ryo and colleagues [46] showed a circadian regulation of transcripts with a daily peak during the end of the dark phase. Herrin and colleagues [47] studied the regulation of *rbcL* and other genes during the cell cycle of *C. reinhardtii*. They also found accumulation of the *rbcL* mRNA during the last half of the light and during the dark period with evidence for a differential translational regulation of LSU synthesis initiated during the light phase. Increased transcript degradation of *rbcL* upon light was described by Salvadort and colleagues [48].

The *rbc* mRNAs on one side and the RSU proteins on the other seemed not well correlated (Fig. 4). Best transcript-to-protein correlation across the day, however, was found with a nine hour time shift where the 12 h dark point of the transcripts correlates positively with the 21 h light point for the total RSUs. Noticeably, intact RSUs (55- and 15-kDa gel bands) did not correlate with corresponding transcripts as they do not show a diurnal oscillation compared to the total RSU and transcript abundances. These findings suggest that 1) total holoenzyme abundance may not significantly change during the day while 2) turnover of LSU synthesis and breakdown seems to increase since 3) total but not intact LSU abundance increases with

maximum CO_2-fixation during light phase and may not be controlled by epistasy of synthesis as previously described in higher plants [49]. These results may confirm the hypothesis of Kim and Mullet [50] that translation initiation complexes for *rbcL* are formed in the dark, but the transition step of translation initiation complexes entering the elongation phase of protein synthesis and/or the elongation step might be inhibited. This translational block seems to be released upon illumination and contribute to light-activated translation of *rbcL* transcript [51].

Nassoury and colleagues [51] demonstrated a circadian CO_2 fixation rhythm in Dinoflagellates. They found that CO_2-fixation rate changes were not related to differing RuBisCO holoenzyme levels, which matches with our results. In addition, a Western Blot analysis revealed no significant changes in RuBisCO holoenzyme levels in rice leaves [12]. Interestingly, however, several degradation products of the LSU of RuBisCO were initially found in pea chloroplasts after incubation in light [52]. Desimone and colleagues [53] reported that oxidative treatment stimulated partial degradation of the LSU of RuBisCO in isolated chloroplasts of barley. LSU was found to be degraded into an N-terminal side fragment of 37-kDa and a C-terminal side fragment of 16-kDa by the hydroxyl radical in chloroplasts in light [54]. Light seems an essential requirement for the fragmentation due to the production of hydrogen peroxide and generation of the reducing power at thylakoid membranes in light [55]. Ishida and colleagues [56] described that the C-terminus of the 37-kDa fragment was Ser-328 (Ser-327 in *C. reinhardtii*). They found the cleavage site positioned at the N-terminal end of the flexible loop (loop 6) within the b/a-barrel domain, constituting the catalytic site of RuBisCO. CO_2 fixation is related to the level of phosphorylation of both the large and small subunits such that phosphorylation accompanies CO_2 assimilation [57]. If phosphorylation plays a regulatory role for this site-specific cleavage is unknown. So far the explanation for the specific degradation of the LSU in the light has been explained by reactive oxygen species generated at its catalytic site by a Fenton-type reaction [56]. The exact function of the 37-kDa fragment of the LSU remains yet to be understood. Taken all data into consideration, it becomes evident that the high amount of total LSU is explained by an increase in LSU degradation during the light phase along with higher synthesis rates derived by possible increased mRNA levels in the dark. A hypothesis may be that, due to limited SSU availability, most of the LSU is unstable and rapidly fragmented upon light [55] or that LSU is to a large extent only partially translated [58]. Despite its rapid degradation, LSU fragments might have catalytic function explaining higher CO_2 fixation rates in light. This may further suggest a rapid degradation of LSU after catalytic reaction that needs to be replaced more frequently than SSU. Besides all hypotheses, it should also be considered that degradation of the LSU may as well have no biologically relevant function. Nevertheless, a significantly higher ratio of large subunit on both, transcript and protein level and trends of diurnal dynamics suggests biological relevance. Indications for a specific translational regulation of the LSU may also be underlined by the fact that changes in protein abundance in response to environmental perturbations have so far only been observed for the LSU and not SSU for instance upon salt stress [59] and drought [60].

5. Conclusions and outlook

We found that,

1) The ratios of transcripts as well as proteins are 5 fold on average for the large subunit in crude extracts, and around 1:1 for the purified holoenzyme.
2) The presented data suggest that LSU translation regulation is SSU independent and holoenzyme assembly SSU limited.
3) Diurnal oscillation of the LSU translation seems controlled by light and post translational regulation.
4) The data further indicate that LSU accumulation in the light phase may play a role during increased CO_2-fixation.

Further work needs to be done to clarify if and as to how the 37-kDa LSU fragment plays a biological relevant role. In this respect, it will be especially interesting to study the diurnal changes and RuBisCO abundance of photoautotrophically grown cells in the future.

We believe that in future, Mass Western combined with additional ^{15}N metabolic labeling experiments for RuBisCO turnover calculations will allow gaining better understanding of the role of RuBisCO specifically in the regulation of N use efficiency not only at the whole plant level but also at organ and subcellular level. The Mass Western may be adapted to most protein(s) of interest.

Sources of funding

L.V. was supported by a Marie Curie IEF Fellow (FP7-PEOPLE-2009-IEF-255109, European Union).

P.O. is funded by a research grant (09-EuroEEFG-FP-034) from the European Science Foundation.

D.L. is funded by the Austrian Science Foundation (FWF); project [P23441-B20].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.09.026.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgment

We thank Christian Molitor (Institute of Biophysical Chemistry, University of Vienna) for providing us with column and FPLC for SE. We would also like to thank M. Angeles Castillejo and Christiana Staudinger for useful comments.

REFERENCES

[1] Barr JR, Maggio VL, Patterson DG, Cooper GR, Henderson LO, Turner WE, et al. Isotope dilution—mass spectrometric

quantification of specific proteins: model application with apolipoprotein A–I. Clin Chem 1996;42:1676–82.
[2] Barnidge DR, Dratz E a, Martin T, Bonilla LE, Moran LB, Lindall A. Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. Anal Chem 2003;75:445–51.
[3] Wienkoop S, Larrainzar E, Glinski M, Gonzalez EM, Arrese-Igor C, Weckwerth W. Absolute quantification of *Medicago truncatula* sucrose synthase isoforms and N-metabolism enzymes in symbiotic root nodules and the detection of novel nodule phosphoproteins by mass spectrometry. J Exp Bot 2008;59:3307–15.
[4] Wienkoop S, Weiss J, May P, Kempa S, Irgang S, Recuenco-Munoz L, et al. Targeted proteomics for *Chlamydomonas reinhardtii* combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. Mol Biosyst 2010;6:1018–31.
[5] Lehmann U, Wienkoop S, Tschoep H, Weckwerth W. If the antibody fails—a Mass Western approach. Plant J 2008;55: 1039–46.
[6] Holzmann J, Pichler P, Madalinski M, Kurzbauer R, Mechtler K. Stoichiometry determination of the MP1-p14 complex using a novel and cost-efficient method to produce an equimolar mixture of standard peptides. Anal Chem 2009;81:10254–61.
[7] Johnson X. Manipulating RuBisCO accumulation in the green alga, *Chlamydomonas reinhardtii*. Plant Mol Biol 2011;76: 397–405.
[8] Feller U, Anders I, Mae T. Rubiscolytics: fate of RuBisCo after its enzymatic function in a cell is terminated. J Exp Bot 2008; 59:1615–24.
[9] Wostrikoff K, Stern D. RuBisCO large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. Proc Natl Acad Sci U S A 2007;104:6466–71.
[10] Khrebtukova I, Spreitzer RJ. Elimination of the *Chlamydomonas* gene family that encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci U S A 1996;93:13689–93.
[11] Mishkind ML, Schmidt GW. Posttranscriptional regulation of ribulose 1,5-bisphosphate carboxylase small subunit accumulation in *Chlamydomonas reinhardtii*. Plant Physiol 1983;72:847–54.
[12] Wang Z, Wang T. Dynamic proteomic analysis reveals diurnal homeostasis of key pathways in rice leaves. Proteomics 2011;11:225–38.
[13] Pilgrim ML, McClung CR. Differential involvement of the circadian clock in the expression of genes required for ribulose-1,5-bisphosphate carboxylase/oxygenase synthesis, assembly, and activation in *Arabidopsis thaliana*. Plant Physiol 1993;103:553–64.
[14] Winder TL, Anderson JC, Spalding MH. Translational regulation of the large and small subunits of ribulose bisphosphate carboxylase/oxygenase during induction of the CO(2)-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Physiol 1992;98:1409–14.
[15] Shapira M, Lers a, Heifetz PB, Irihimovitz V, Osmond CB, Gillham NW, et al. Differential regulation of chloroplast gene expression in *Chlamydomonas reinhardtii* during photoacclimation: light stress transiently suppresses synthesis of the RuBisCO LSU protein while enhancing synthesis of the PS II D1 protein. Plant Mol Biol 1997;33: 1001–11.
[16] Cohen I, Sapir Y, Shapira M. A conserved mechanism controls translation of RuBisCO large subunit in different photosynthetic organisms 1. Plant Physiol 2006;141:1089–97.
[17] Genkov T, Spreitzer RJ. Highly conserved small subunit residues influence RuBisCO large subunit catalysis. J Biol Chem 2009;284:30105–12.
[18] Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 2007; 318:245–50.
[19] Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, et al. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. Mol Syst Biol 2011;7:518.
[20] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Anal Biochem 1976;72: 248–54.
[21] Olinares PBD, Ponnala L, Van Wijk KJ. Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry and hierarchical clustering. Mol Cell Biol 2010;9:1594–615.
[22] Fiala GJ, Schamel WW a, Blumenthal B. Blue native polyacrylamide gel electrophoresis (BN-PAGE) for analysis of multiprotein complexes from cellular lysates. J Vis Exp 2011: 1–8.
[23] Wienkoop S, Weckwerth W. Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in *Arabidopsis thaliana*. J Exp Bot 2006;57:1529–35.
[24] Tourna M, Freitag TE, Prosser JI. Stable isotope probing analysis of interactions between ammonia oxidizers. Appl Environ Microbiol 2010;76:2468–77.
[25] Roze S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000;132: 365–86.
[26] Sun X, Weckwerth W. COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. Metabolomics 2012;8: 81–93.
[27] Lyon D, Weckwerth W, Wienkoop S. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S, editors. Mass Western for Absolute Quantification of Target Proteins and Considerations About the Instrument of Choice. , Plant Proteomics Methods Protoc. Methods Mol. Biol.Totowa, NJ: Humana Press; 2014. p. 199–208.
[28] Millard P. The accumulation and storage of nitrogen by herbaceous plants. Plant Cell Environ 1988;11:1365–504.
[29] Irving LJ, Robinson D. A dynamic model of RuBisCO turnover in cereal leaves. New Phytol 2006;169:493–504.
[30] Taylor TC, Backlund a, Bjorhall K, Spreitzer RJ, Andersson I. First crystal structure of RuBisCO from a green alga, *Chlamydomonas reinhardtii*. J Biol Chem 2001;276(48):159–64.
[31] Peltier J-B, Cai Y, Sun Q, Zabrouskov V, Giacomelli L, Rudella A, et al. The oligomeric stromal proteome of *Arabidopsis thaliana* chloroplasts. Mol Cell Proteomics 2006;5: 114–33.
[32] Mastrobuoni G, Irgang S, Pietzke M, Aßmus HE, Wenzel M, Schulze WX, et al. Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*. BMC Genomics 2012;13.
[33] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 2005;4:1265–72.
[34] Zybailov B, Florens L, Washburn M. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. Mol Biosyst 2007;3: 354–60.
[35] Larrainzar E, Wienkoop S, Scherling C, Kempa S, Ladrera R, Arrese-Igor C, et al. Carbon metabolism and bacteroid functioning are involved in the regulation of nitrogen fixation in *Medicago truncatula* under drought and recovery. Mol Plant Microbe Interact 2009;22:1565–76.
[36] Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ. Multiplexed absolute quantification for

proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protoc 2006;1:1029–43.
[37] Aebersold R, Burlingame AL, Bradshaw Ra. Western blots versus selected reaction monitoring assays: time to turn the tables? Mol Cell Proteomics 2013;12:2381–2.
[38] Makino A, Mae T, Ohira K. Colorimetric measurement of protein stained with Coomassie brilliant blue R on sodium dodecyl sulfate-polyacrylamide gel electrophoresis by eluting with formamide. Agric Biol Chem 1986;50:1911–2.
[39] Makino A, Shimada T, Takumi S, Kaneko K, Matsuoka M, Shimamoto K, et al. Does decrease in ribulose-1,5-bisphosphate carboxylase by antisense RbcS lead to a higher N-use efficiency of photosynthesis under conditions of saturating CO_2 and light in rice plants? Plant Physiol 1997;114:483–91.
[40] Makino A, Osmond B. Effects of nitrogen nutrition on nitrogen partitioning between chloroplasts and mitochondria in pea and wheat. Plant Physiol 1991;96:355–62.
[41] Suzuki Y, Makino A, Mae T. Changes in the turnover of RuBisCO and levels of rbcL and rbcS in rice leaves from emergence to senescence. Plant Cell Environ 2001;24: 1353–60.
[42] Spreitzer RJ. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. Arch Biochem Biophys 2003;414:141–9.
[43] Van Lun M, van der Spoel D, Andersson I. Subunit interface dynamics in hexadecameric RuBisCO. J Mol Biol 2011;411: 1083–98.
[44] Tcherkez GGB, Farquhar GD, Andrews TJ. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc Natl Acad Sci U S A 2006;103:7246–51.
[45] Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ, et al. RuBisCO small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. Proc Natl Acad Sci U S A 2012;109:19474–9.
[46] Kawazoe R, Mahan KM, Venghaus BE, Carter ML, Herrin DL. Circadian regulation of chloroplast transcription in *Chlamydomonas* is accompanied by little or no fluctuation in RPOD levels or core RNAP activity. Mol Biol Rep 2012;39: 10565–71.
[47] Herrin DL, Michaels AS, Paul GA. Regulation of genes encoding the large subunit of ribulose-l, 5-bisphosphate carboxylase and the photosystem II polypeptides D-1 and D-2 during the cell cycle of *Chlamydomonas reinhardtii* preparation of thylakoid membranes isolation of thylakoid membrane. J Cell Biol 1986;103:1837–45.
[48] Salvadort ML, Klein U, Bogorad L. Light-regulated and endogenous fluctuations of chloroplast transcript levels in *Chlamydomonas*. Regulation by transcription and RNA degradation. Plant J 1993;3:213–9.
[49] Choquet Y, Vallon O. Synthesis, assembly and degradation of thylakoid membrane proteins. Biochimie 2000;82:615–34.
[50] Kim J, Mullet JE. A mechanism for light-induced translation of the rbcL mRNA encoding the large subunit of ribulose-1,5-bisphosphate carboxylase in barley chloroplasts. Plant Cell Physiol 2003;44:491–9.
[51] Nassoury N, Fritz L, Morse D. Circadian changes in ribulose-1,5-bisphosphate carboxylase/oxygenase distribution inside individual chloroplasts can account for the rhythm in dinoflagellate carbon fixation. Plant Cell 2001; 13:923–34.
[52] Mitsuhashi W, Feller U. Effects of light and external solutes on the catabolism of nuclear-encoded stromal proteins in intact chloroplasts isolated from pea leaves. Plant Physiol 1992;100:2100–5.
[53] Desimone M, Henke a, Wagner E. oxidative stress induces partial degradation of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase in isolated chloroplasts of barley. Plant Physiol 1996;111:789–96.
[54] Ishida H, Nishimori Y, Sugisawa M, Makino a, Mae T. The large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase is fragmented into 37-kDa and 16-kDa polypeptides by active oxygen in the lysates of chloroplasts from primary leaves of wheat. Plant Cell Physiol 1997;38:471–9.
[55] Ishida H, Shimizu S, Makino a, Mae T. Light-dependent fragmentation of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase in chloroplasts isolated from wheat leaves. Planta 1998;204:305–9.
[56] Ishida H, Makino a, Mae T. Fragmentation of the large subunit of ribulose-1,5-bisphosphate carboxylase by reactive oxygen species occurs near Gly-329. J Biol Chem 1999;274:5222–6.
[57] Foyer CH. Stromal protein phosphorylation in spinach (*Spinacia oleracea*) chloroplasts. Biochem J 1985;231:97–103.
[58] Blowers A, Klein U, Ellmore G, Bogorad L. Functional in vivo analyses of the 3′ flanking sequences of the *Chlamydomonas* chloroplast rbcL and psaB genes. Mol Gen Genet MGG 1993; 238:339–49.
[59] Pang Q, Chen S, Dai S, Chen Y, Wang Y, Yan X. Comparative proteomics of salt tolerance in *Arabidopsis thaliana* and *Thellungiella halophila*. J Proteome Res 2010;9:2584–99.
[60] Aranjuelo I, Molero G, Erice G, Christophe Avice J, Nogues S. Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). J Exp Bot 2011;62:111–23.

3 Concluding Discussion

3.1 Consolidation of publications, delineating the preamble and comprehensive methodology

The increasing importance of algae *Chlamydomonas reinhardtii* both as a model organism with a completely sequenced genome and as a potential bioproducer has led to many comprehensive studies in a systems biology context. The work within this thesis continues the line of work initiated by me and my current group in 2008 with the publication of the first comprehensive analysis on this model algae integrating genome annotation techniques, systematic high-resolution shotgun proteomics, systematic metabolite profiling and structural modeling **[195]**. In this experiment, 4202 unique peptide sequences were identified — these sequences could be mapped to 1069 proteins, finally reconstructing a draft metabolic network that allowed the identification of missing enzymatic links (e.g. galactinol synthase). A second integrative study on Chlamydomonas reinhardtii was published by me and my group in 2010 **[338]**. Here, we aimed to elucidate basal growth processes (especially the interplay of photosynthetic apparatus, mitochondrial proteins and metabolites) in both autotrophical and mixotrophical growth conditions — through integration of metabolomics, stable flux measurements, untargeted proteomics and absolute protein quantification, a comprehensive dataset with 25% proteome coverage, high complementarity and strong correlations between intracellular protein and metabolite concentrations (most notably in the TCA-cycle) was produced. By means of a previous subcellular fractionation and the introduction of the Mass Western technique, the exact amount of different enzyme isoforms could be measured in their respective locations (e.g. CAH), reaching a dynamic range of four orders of magnitude. This enabled the measurement of low-abundance proteins that would be barely detectable on relative quantification approaches, such as SBPase.

Nevertheless, despite the wide scope of techniques and the achievements reached in these two previous publications, many newer questions have arised both from a technical and from a biological standpoint.

For instance, since many approaches rely on protein inference from ion matches to tryptic peptides (obtained through *in silico* digestion of a protein database), it is key to determine the degree in which a certain database influences every measurement — not only on qualitative measurements, but also on quantitative ones. To evaluate the degree of variability coming from different databases on a proteomics measurement, a *Chlamydomonas reinhardtii* MS-dataset was analysed using the same algorithm, FDR value and correlation score with four different databases containing different annotations from two different genome assemblies. These analyses were performed both for peptide and protein identification to test the effect of different databases on the identification of discriminatory peptides and thus protein assignment. To avoid any sample-dependent bias the dataset covered different growth situations and cell fractions, and the distribution of X-Correlations between identified peptides and databases was also studied to evaluate the variation of score factors according to the different database sizes. This study has been published in **Valledor et al. 2012 [310]**.

The identification and relative quantification of previously unknown ions by an untargeted proteomics approach has been used to enhance tryptic peptide databases, but it could also help define the role of posttranslational modifications on protein function (or "protein speciation", as denominated by Jungblut et al. [142]), protein interaction (e.g. acetylation, ubiquitination). With this goal in mind, the **Mass Accuracy Precursor Alignment** (**MAPA**) workflow was developed. First published in 2008, MAPA allows the alignment of precursor ions of an exact m/z with their respective spectral counts from an MS measurement in a single matrix. Through normalisation of these values, data reduction through PCA and consequent unsupervised independent component analysis, the characteristic peptides for certain experimental conditions can be clustered and further identified by de novo peptide sequence prediction. This allows quick and reliable identification of previously unknown PTMs, and by extension the localisation of functional PTMs, making MAPA a powerful tool for proteins as functional entities. This procedure has been tested by this group on Chlamydomonas sets, and also on hormone signaling experiments on *Arabidopsis thaliana* in which phosphorylation kinetics were correlated to phytohormone exposure [49], as published on **Hoehenwarter et al 2011 [126]**.

Another possible application of absolute peptide quantification and functional protein analysis is the study of protein complexes regarding subunit localisation and exact stoichiometry. Even though the highly abundant/relevant PS-related enzyme RuBisCO presents a 1:1 stoichiometry as a holoenzyme, the measured ratio between LS and SS in **Wienkoop et al. 2010** was actually several fold **[338]** — besides, recent models on other species have also noted the accumulation of a 37 kDa (oxidative) degradation N-fragment from LS [135]. To determine the stoichiometry of RuBisCO accurately, the amounts of the single protein subunits should be measured both on raw cell extracts and in their isolated fractions. Protein separation on SDS-PAGE would also allow to look for LSU peptide sequences in the 37kDa protein band from every sample. The presence of the RuBisCo subunits on the SDS-PAGE should be additionally visualized with Western Blots. To exclude any effects from sample manipulation on the final subunit ratio, further quantification of both subunits should be performed on previously purified RuBisCo holoenzyme extracts (final result should be 1:1). To describe the dynamics of the synthesis of both subunits, samples were collected every 3 hours during a light/dark cycle of 24h. For absolute peptide quantification, a new version of the Mass Western (Lehmann 2008) was developed: to guarantee constant stoichiometry between the different standard peptides and also reflect the variability brought by digestion efficiency, standard peptides featuring cross-concatenated peptide sequences were used. Furthermore, an additional, double-labeled equaliser peptide was used for higher accuracy. This study has been published on **Recuenco-Muñoz et al. 2015 [253]**.

A further study regarding the different connections between C- and N-metabolism in microalgae are an increasingly relevant topic due to the high potential of microalgae as potential bioproducers for the growing biofuel industry. Since lipid accumulation is significantly higher during N starvation, many studies related to Nitrogen stress have been performed on *Chlamydomonas reinhardtii* — however, most of them had been limited to a 24h interval of N-depletion, instead of considering not only short but also long-term adaptive responses to N-depletion. Aiming to elucidate these physiological changes and to search for potential target genes to increase TAG accumulation through bioengineering, samples were obtained over a 96 hour interval consisting of a 72h starvation phase and a 24h recovery phase. The dynamic behaviour of the biochemical pathways and metabolism to N availability was investigated through a systems biology approach at a protein level, at a metabolite level, and enhanced with classical physiology measurements to track cell density, Fm/Fv rate, and total lipid and biomass content. At a pro-

tein level, qualitative and semiquantitative LC-MS/MS proteomics were performed on whole cell extracts — additionally, cell nuclei were isolated to look for significant changes in transcription factors that would otherwise go under higher concentrated cytosolic and membrane proteins. The identified proteins were then functionally classified and mapped. The changes on a metabolic level were followed through GC-MS polar metabolite analysis, whereas lipid composition was studied through fatty acid methyl ester (FAME) analysis. Finally, the correlations between all the datasets were tested/established with multivariate statistics, Granger causality and sparse partial least square (sPLS) analyses. For a more robust and comprehensive analysis, the generated data were also mined with previously existing transcriptomic and metabolomic datasets at their overlapping experimental points (0h and 72h). This study has been published on **Valledor et al. 2014 [309]**.

3.2 Contribution to the scientific progress

There have already been comprehensive studies on *Chlamydomonas reinhardtii* in a Systems Biology context due to its importance as a model organism with a completely sequenced genome, its unique phyisiological properties and also due to its growing importance as a potential bioproducer.

My PhD-Thesis further explores different aspects and connections of C- and N-metabolism of model algae Chlamydomonas reinhardtii, both as a systemic response to external stress conditions (enhanced lipid accumulation during N starvation) and through a specific stoichiometry study of RuBisCO, a key enzyme of C-metabolism which also acts as a major N-storage protein. For this purpose, a wide range of proteomics methods were applied: qualitative and quantitative analysis, database dependent and independent protein identification, absolute and relative quantification, or spectral counting and AUC. Many of these techniques were also coupled to different subfractionation techniques (SDS-PAGE, BN-PAGE, FPLC, nuclei isolation) and complemented with both more classical strategies (physiological measurements, Western Blotting) and other omics-related methods (metabolomics, transcriptomics, genome annotation). This wide scope of different proteomics methods makes this PhD-Thesis a very comprehensive overview of the current approaches to protein analysis in a systems biology context. The data obtained from these experiments were also used to evaluate and refine other aspects of proteomic analysis such as PTM localization through MAPA and how different databases can alter the results of both qualitative and quantitative analysis.

From a biological standpoint, the relevance of this PhD-Thesis relates to the two already mentioned questions within C- and N-metabolism of *Chlamydomonas reinhardtii*:

(a) **RuBisCO-stoichiometry study**: the results previously published in **Wienkoop et al. 2010** were confirmed, having a LSU:SSU stoichiometry of 5:1 on both complete cell extracts and prefractionated RSU fractions. This difference is not only due to sample preparation effects, since our results from purified holoenzyme extracts were way closer to the classical 1:1 ratio ($\sim$ 1.5:1). This data are also supported by transcriptomics results, featuring a similar stoichiometry of 5:1, what suggests that the translation of both subunits might be regulated independently. Whereas SSU levels remained relatively constant during the whole 24h, the correlation between protein and mRNA values for the LSU diurnal oscillation was only apparent after considering a 9 hour time shift, with higher mRNA accumulation during the dark phase and higher protein accumulation during the light phase — this could be evidence of light being a key factor at the regulation of LSU synthesis. The presence of an 37kDa N-fragment from the

LSU could not only be confirmed but also correlated with higher CO_2-fixation. Therefore, this fragment may not only play a role on RuBisCO synthesis or as a marker for oxidative stress, but also have catalytic activity as well. Since these conclusions are still very speculative (and controversial) further research should be performed. On a technical level, the new Mass Western manages to compensate many of the challenges of peptide quantification, giving the most accurate peptide quantification to date — nonetheless, new improvements on MS-technology should enhance its accuracy. Besides, the addition of further equilizer peptides might help overcome further issues, such as compensating for protein loss during sample treatment, or increase the limited size of the cross-concatenated standard peptides. On a biological level,this experimental design could be applied also on related enzymes (such as RuBisCO activase), subsequently studying their correlation with both RuBisCO subunits and the 37kDa fragment. Besides, further information about the dynamics of RuBisCO synthesis might be won from applying this approach on different cultivation conditions, e.g. photoautotrophic grown algae.

(b) **Enhanced TAG accumulation as a result of N-deprivation**: this experimental design turned out decisive to gain new insights on TAG accumulation by *Chlamydomonas reinhardtii* under N-starvation. First of all, a complex network of stress-responsive proteins, metabolites and physiological parameters was established and complemented with older transcriptomes and metabolomics datasets. This network showed many complementary gene/protein complexes responsive to stress, suggesting major points of metabolic regulation based on common signaling elements. In general, it could be confirmed that the adaptation mechanisms to N-depletion might be switched by the C/N balance in the cell. Another distinctive mechanism was the differential ribosome accumulation in the cell (decrease in the chloroplasts, increase in the cytosol), hinting at a more active role of nuclear encoded proteins during N-stress. This can be supported by major changes found in the nuclear proteome (136 proteins were differentially expressed), even though, ironically, none of them were related to the major physiological changes N-starvation on *Chlamydomonas reinhardtii* is known for (gametogenesis and sexual reproduction). The adjustment of lipid metabolism to N-starvation could be confirmed both on the a higher TAG accumulation as also in a change of the total lipid composition of the cell (with an increased relative amount of C18-2 fatty acids). Significant changes on a protein level were detected mainly on proteins related to oxidative phosphorylation (most notably the increase on NADH-synthases, ATP-synthases and COX) and a sharp decrease of the light-dependent reaction proteins. These changes are especially relevant since they are not shared with other depletions or abiotic stresses. Another novelty of this study is the inclusion of a recovery phase after N-deprivation: most significant was the instant biomass increase after N-replenishment and the significantly higher expression of proteins related to the β-oxidation pathways. At this stage, proteins related to glycerol metabolism, MLDP and COP were found to be relevant for cell growth processes. Finally, we also managed to find target genes that could potentially bring higher lipid accumulation through bioengineering, both through iRNA silencing (e.g. MDLP) or overexpression (e.g. BRI1 suppressor, related to glycerol and C18:2 and only detected at N deprivation). Nevertheless, even though these genes have not been reported to be significant in other stresses, these results still should be complemented by a perfect flux model in order to correct and avoid unintended effects. Besides, this line of work could be also continued through other approaches already used in this thesis, such as the study of individual enzymes (e.g. FtsH metalloproteases) or the application of MAPA on nuclear proteome analysis to find hitherto unidentified transcription factors.

In general, the findings contained in my PhD-Thesis would also profit from further technical enhancements. From a technical standpoint, new improvements in LC-MS/MS technology

should bring better chromatographic peptide separation and MS measurements with higher speed and accuracy. The incorporation of new devices to mass spectrometers has also led to the development of further quantification techniques (e.g. **Parallel Reaction Monitoring** (**PRM**) by Thermo). These advances have also been reflected on the development of many (open source) platforms like Skyline [185] or MaxQuant [53] that offer both more machine compatibility and more complex data mining workflows. Also previously existing workflows like MAPA have been improved, adding new parameters such as RT. Nevertheless, besides powerful identification algorithms and versatile analysis tools, protein databases are a capital element for reliable protein identification and quantification that cannot be taken for granted — in fact, this work gives evidence of the difficulties in keeping track of proteins from the same species, showing very disparate protein identifications coming out of a same dataset, no matter if using different genome assemblies, different annotations from the same assembly, or even identifying proteins through detection of one or two peptides. These results remark how continous update of the currently available, manually curated protein databases with novel experimental data and newer genome annotation models are capital for accurate and reliable protein identification and quantification, and how important it is not only to implement these databases but also to develop new strategies for database quality control.

4 Summary

Systems Biology can be described as a holistic approach striving to analyze an organism as a biological system, integrating the characterization of its components and the description of their interactions and dynamics. This strategy led to the development of the -omics studies, which aim at collecting the complete set of a determined kind of biomolecules from an organism at a certain time point (e.g. transcriptomics: mRNA; proteomics: proteins; metabolomics: metabolites, etc). Thus, high-throughput technologies that allow highly reproducible measurements and large data production must be applied: a good example is the use of Mass Spectrometry (MS) for Proteomics studies. My PhD thesis continues our previous work with two comprehensive -omics publications on *Chlamydomonas reinhardtii*, exploring different applications of LC-MS/MS for Proteomics studies on this organism. *Chlamydomonas reinhardtii*, a unicellular algae with high ability to adapt dynamically to environmental conditions, great relevance as a model plant, and also increasingly important as a potential bioproducer of fuel and biomass. This line of work has led to two publications previous to this PhD-Thesis, both of which I co-authored: **May et al 2008 [195]** and **Wienkoop et al. 2010 [338]**).

This PhD-Thesis comprehends four different topics within the context of the proteomic study of *Chlamydomonas reinhardtii*:

1- The data gained from this and also previous studies were also used to implement the **Mass Accuracy Precursor Alignment** (**MAPA**) strategy, which is based on the combination of the ProtMax algorithm and multivariate statistics (PCA/ICA). On **Hoehenwarter et al.2011** the application of this workflow for search of functional PTMs for protein speciation in samples from different species is discussed. The MAPA approach was successfully applied on a study of the phosphorylation state of key components of signal induction cascades after phytohormone treatment in *Arabidopsis thaliana* — not only were the phosphorylated peptides identified, but also their phosphorylation site **[126]**.

2- Proteomics data can also be used to enhance genomic annotation and implement the currently existing protein databases, since databases made of different genome assemblies and even from the same genome assembly but different annotations can still significantly influence the results of a proteomics study. Hence, four different databases made from different annotations of two different genome assemblies were tested for the same proteomics dataset. On **Valledor et al. 2012** these results are displayed graphically, mapping the identified proteins, evaluating the overlaps between different databases, and also comparing the outcomes of single peptide and two-peptide protein identification for each database, showing significant differences for both qualitative and quantitative analysis —– in fact, different databases can significantly alter both qualitative analyses (different IDs in different annotations) and quantitative protein analysis (with key peptides for protein validation not being identified at all). **[310]**.

3- A qualitative/semiquantitative experiment was developed for a Nitrogen deprivation study during a four day cycle, comprising three days of Nitrogen depletion and 24h of Nitrogen repletion for recovery. To establish the connection between N-metabolism, lipid accumulation, cell growth and eventual changes in membrane composition, proteomic and metabolomic data were gained from whole cell extracts and isolated nuclear fractions, correlated with pyhsiological

measurements and mapped. Subsequently, these data were further correlated with an existing transcriptomic dataset, and associated to a metabolic network. This approach led to new insights into physiological mechanisms for both adaptation to N-starvation and its recovery after N-repletion, finding significant changes at organelle (e.g. differential ribosome accumulation), protein and lipid levels (e.g. downregulation of PS-relative proteins and changes in both lipid amount and lipid composition during N-starvation), and finding putative target genes to increase lipid accumulation through bioengineering (e.g. overexpression of the BRI1 suppressor, as published in **Valledor et al. 2014 [309]**.

4- After having found evidence of different stoichiometry for the large and the small RuBisCO subunit in our previous work **[338]**, a new strategy was developed, including enhanced Mass Western via cross-concatenated peptides for absolute quantification and combining it with different extraction and prefractionation methods to exclude any kind of technical artefacts. Furthermore, the protein data were correlated with transcript data over a diurnal cycle to follow the dynamic of RuBisCO-Expression during a light/dark cycle and discuss possible physiological effects of the diurnal oscillation of the large subunit. The results of this study, published on **Recuenco-Munoz et al. 2015** do not only confirm the 5:1 stoichiometry for LSU:SSU and a possible light regulation of LSUs diurnal cycle, but also bring evidence for the accumulation of a 37kDa N-fragment from LSU and its possible physiological functions — the positive correlation of C-assimilation and the amount of 37kDa N-fragment suggests a catalytic activity of its own **[253]**.

As a result, my PhD-Thesis not only covers a wide range of applications and challenges of the current Proteomics plant research (different strategies for quantitative/qualitative analysis and a comparative study on protein databases), but also deals with relevant aspects of the *Chlamydomonas reinhardtii* metabolism, bringing new insights on key aspects of its C- and N-metabolism (RuBisCO, N-starvation and recovery) and offering new tools and suggestions for both functional studies (MAPA) and industrial application (bioengineering targets for improved lipid production).

Zusammenfassung

Systembiologie kann als ein holistischer Ansatz beschrieben werden, laut dessen ein Organismus nicht als Summe seiner Bestandteile betrachtet wird, sondern auch deren Interaktion und Dynamik berücksichtigt. Diese Strategie führte zur Entwicklung der -Omics Studien. Ziel der Omics-Studien ist es, den gesamten Satz einer gewissen Makromolekülenklasse eines Organismus zu einem bestimmten Zeitpunkt zu detektieren, identifizieren und wenn möglich quantifizieren.

In der vorliegenden Arbeit wurden vier zentrale Fragestellungen der Proteomanalyse in der einzelligen Modellalge *Chlamydomonas reinhardtii* behandelt:

1- Die Datensätze aus diesen und auch aus anderen Proteomics-Studien an verschiedenen Pflanzensorten wurden benutzt, um den MAPA (Mass Accuracy Precursor Alignment) Workflow zu implementieren. Diese Strategie umfasst die Kombination des ProtMax Algorithmus mit multivariater Statistik. In **Hoehenwarter et al. 2011** wird die Anwendung von MAPA für funktionellen Studien und Proteinspezifikation diskutiert **[126]** und erfolgreich an einer Studie über Signalinduktionskaskaden nach Phytohormonbehandlung an *Arabidopsis thaliana* angewendet: hier wurden nicht nur phosphorylierte Peptide identifiziert, sondern auch ihre jeweilige Phosphorylierungsstelle.

2- Experimentelle Datensätze von Proteomics-Untersuchungen können auch für Genomannotationsstudien und Verbesserung der bestehenden, manuell gewarteten Proteindatenbanken dienen. Proteindatenkbanken können sowohl aus verschiedenen Genomrekonstruktionen als auch aus verschiedenen Annotationen der gleichen Genomrekonstruktion stammen — dabei können verschiedene Proteindatenbanken signifikante Unterschiede bei der Analyse des gleichen Datensatzes erzeugen. Dies wurde gestestet, indem ein bestimmter *Chlamydomonas reinhardtii* Datensatz mit vier verschiedenen Datenbanken analysiert wurde. Diese Datenbanken stammen aus mehreren Annotationen von zwei verschiedenen Genomrekonstruktionen. Die Ergebnisse aus diesem Vergleich wurden in **Valledor et al. 2012** grafisch dargestellt, die entsprechenden Überlappungen veranschaulicht, und die Ergebnisse aus verschiedenen Identifizierungsstrategien analysiert (Einzelpeptidbestimmung, Proteinbestimmung mit 2 Peptiden). Hier wurden signifikante Unterschiede zwischen den verschiedenen Datenbanken gefunden — tatsächlich wird nicht nur die qualitative Analyse (z.B. verschiedene Genomannotationen ergeben verschiedene Protein-IDs) sondern auch die quantitative Proteinanalyse (z.B. durch Nichtidentifizierung notwendiger Peptide für die Validierung ihrer ursprünglicher Proteine) entscheidend beeinfluβt **[310]**.

3- Um den Zusammenhang zwischen N-Streβ und erhöhter TAG-Akkumulation in CR zu untersuchen, wurde ein Experiment geplant, welches die Anpassung über 72 Stunden N-Mangel und die Erholung 24 Stunden nach N-Zugabe verfolgen sollte. Dabei wurden Proteomics und Metabolomics Untersuchungen an komplette Zellextrakte und auch aufgereinigte Zellkernfraktionen durchgeführt und mit klassischen physiologischen Messungen kombiniert — die entstehenden Ergebnisse wurden zusätzlich mit bereits bestehenden Metabolomics und Transcriptomics-Datensätzen verglichen. Um die Zusammenhänge zwischen N-Metabolismus, Lipidakkumulation, Zellwachstum und etliche Veränderungen in der Lipidzusammensetzung

der Zelle zu veranschaulichen wurden die gewonnenen Daten mithilfe multivariater Statistik prozessiert und schlieβlich in einem metabolischen Netzwerk visualisiert. Durch diese Strategie gelang es **Valledor et al. 2014**, neue Erkenntnisse über die physiologischen Mechanismen von *Chlamydomonas reinhardtii* zur Anpassung an N-Mangel und von N-Mangel zurück zum Ausgangszustand zu erlangen. Dabei konnten signifikante Veränderungen auf Organell-, Protein- und Lipidniveau gefunden werden, genauso wie potentielle Target-Gene, dessen Manipulation durch Bioengineering zu einer Erhöhung der Lipidakkumulation der Alge führen könnte **[309]**.

4- Nachdem Evidenz aus unseren vorigen Untersuchungen auf eine von 1:1 abweichende Stöchiometrie von beiden RuBisCO Untereinheiten hinwies [338], wurde eine Studie durchgeführt, um diese Hypothese gründlich zu untersuchen. Dafür wurden beide Untereinheiten mittels einer neuen Entwicklung des Mass Western (Lehmann 2008) untersucht, dessen Standardpeptide aus aneinandergereihten, querverknüpften Peptidsequenzen aus beiden RuBisCO-Untereinheiten bestanden — dazu wurde auch einen zusätzlichen, doppelt-markierten Equalizer-Peptid zu genaueren Quantifizierung benutzt. Um etliche präparativ erzeugten Artefakte auszuschlieβen, wurden die Mengen der beiden Untereinheiten in Messproben aus (a) rohen Zellextrakten (b) präfraktionierten Proteinextrakten und (c) purifiziertem RuBisCO-Holoenzym quantifiziert. Zusätzlich wurden diese Proteindaten mit Transkriptdaten über einen 24stündigen Zeitintervall korreliert, um die Dynamik der RSU-Expression über ein Licht/Dunkel Tageszyklus zu verfolgen. Die von mir in **Recuenco-Munoz et al. 2015** publizierten Ergebnisse nicht nur bestätigen die 5:1 Stöchiometrie für LSU:SSU, sondern zeigen auch eine Oszillation der LSU-Konzentration über den Tageszyklus. Da diese Oszillation eine Verschiebung von 9h zwischen der höchsten mRNA-Konzentration (im Dunkel) und der höchsten LSU-Konzentration zeigt, könnte man eine lichtbedingte Regulation der LSU-Synthese annehmen. Zudem wurde eine Akkumulation eines 37-kDa N-Proteinfragmentes der LSU gemessen, dessen positiven Korrelation mit der C-Assimilation auf eigene katalytische Aktivität hinweisen könnte **[253]**.

Die vorliegende Arbeit deckt somit nicht nur eine grosse Bandbreite an Techniken und Herausforderungen der heutigen Proteomics-Forschung in der molekularen Pflanzenphysiologie ab, sondern bringt neue Ansichten über den C- und N-Stoffwechsel einer der relevantesten Modellalgen, *Chlamydomonas reinhardtii* (RuBisCO, Lipidproduktion unter N-Mangel). Schlieβlich wurden neue Methoden und Strategien entwickelt, die sowohl in funktionellen (z.B. MAPA, Mass Western) als auch in industriellen Studien (Steigerung der Lipidproduktion durch Bioengineering) Anwendung finden könnten.

Bibliography

[1] Christian H Ahrens, Erich Brunner, Ermir Qeli, Konrad Basler, and Ruedi Aebersold. Generating and navigating proteome maps using mass spectrometry. *Nature reviews. Molecular cell biology*, 11(11):789–801, nov 2010.

[2] Jens Allmer. Algorithms for the de novo sequencing of peptides from tandem mass spectra. *Expert Review of Proteomics*, jan 2014.

[3] Jens Allmer, Bianca Naumann, Christine Markert, Monica Zhang, and Michael Hippler. Mass spectrometric genomic data mining: Novel insights into bioenergetic pathways in Chlamydomonas reinhardtii. *Proteomics*, 6:6207–6220, 2006.

[4] Alma Lorena Almaraz-Delgado, Jose Flores-Uribe, Victor Hugo Perez-Espana, Edgar Salgado-Manjarrez, and Jesus Agustin Badillo-Corona. Production of therapeutic proteins in the chloroplast of Chlamydomonas reinhardtii. *{AMB} Express*, 4:57, 2014.

[5] A F Maarten Altelaar, Danny Navarro, Jos Boekhorst, Bas van Breukelen, Berend Snel, Shabaz Mohammed, and Albert J R Heck. Database independent proteomics analysis of the ostrich and human proteome. *Proceedings of the National Academy of Sciences*, 109(2):407–412, 2012.

[6] N Leigh Anderson, Norman G Anderson, Lee R Haines, Darryl B Hardie, Robert W Olafson, and Terry W Pearson. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *Journal of proteome research*, 3(2):235–44, 2004.

[7] Vasilios M. E. Andriotis, Marilyn J. Pike, Baldeep Kular, Stephen Rawsthorne, and Alison M. Smith. Starch turnover in developing oilseed embryos. *New Phytologist*, 187(3):791–804, aug 2010.

[8] Charles Ansong, Si Wu, Da Meng, Xiaowen Liu, Heather M Brewer, Brooke L Deatherage Kaiser, Ernesto S Nakayasu, John R Cort, Pavel Pevzner, Richard D Smith, Fred Heffron, Joshua N Adkins, and Ljiljana Pasa-Tolic. Top-down proteomics reveals a unique protein S-thiolation switch in Salmonella Typhimurium in response to infection-like conditions. *Proceedings of the National Academy of Sciences of the United States of America*, 110(25):10153–8, jul 2013.

[9] E Asamizu, Y Nakamura, S Sato, H Fukuzawa, and S Tabata. A Large Scale Structural Analysis of cDNAs in a Unicellular Green Alga, Chlamydomonas reinhardtii. I. Generation of 3433 Non-redundant Expressed Sequence Tags. *DNA Research*, 6(6):369–373, 1999.

[10] Ariane Atteia, Annie Adrait, Sabine Brugière, Marianne Tardif, Robert van Lis, Oliver Deusch, Tal Dagan, Lauriane Kuhn, Brigitte Gontero, William Martin, Jérôme Garin,

Jacques Joyard, and Norbert Rolland. 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. *Molecular biology and evolution*, 26(7):1533–48, jul 2009.

[11] M R Badger and G D Price. The Role of Carbonic Anhydrase in Photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, 45(1):369–392, jun 1994.

[12] Ali Bahadar and M. Bilal Khan. Progress in energy from microalgae: A review. *Renewable and Sustainable Energy Reviews*, 27:128–148, nov 2013.

[13] Steven G. Ball and Matthew K. Morell. From Bacterial Glycogen to Starch: Understanding the Biogenesis of the Plant Starch Granule. *Annual Review of Plant Biology*, 54(1):207–233, jun 2003.

[14] Shibdas Banerjee and Shyamalava Mazumdar. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, 2012:1–40, 2012.

[15] Marcus Bantscheff, Markus Schirle, Gavain Sweetman, Jens Rick, and Bernhard Kuster. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry*, 389(4):1017–31, oct 2007.

[16] Sophia Barinova, Mariyo Boboev, and Hikmat Hisoriev. Freshwater algal diversity of the South-Tajik Depression in a high-mountainous extreme environment, Tajikistan. *Turkish Journal of Botany*, 39(3):535–546, 2015.

[17] W C Barker, J S Garavelli, H Huang, P B McGarvey, B C Orcutt, G Y Srinivasarao, C Xiao, L S Yeh, R S Ledley, J F Janda, F Pfeiffer, H W Mewes, a Tsugita, and C Wu. The protein information resource (PIR). *Nucleic acids research*, 28(1):41–44, 2000.

[18] J R Barr, V L Maggio, D G Patterson, G R Cooper, L O Henderson, W E Turner, S J Smith, W H Hannon, L L Needham, and E J Sampson. Isotope dilution–mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clinical chemistry*, 42:1676–1682, 1996.

[19] Benedikt Beel, Katja Prager, Meike Spexard, Severin Sasso, Daniel Weiss, Nico Müller, Mark Heinnickel, David Dewez, Danielle Ikoma, Arthur R. Grossman, Tilman Kottke, and Maria Mittag. A Flavin Binding Cryptochrome Photoreceptor Responds to Both Blue and Red Light in Chlamydomonas reinhardtii. *The Plant Cell*, 24(7):2992, 2012.

[20] a Ben-Amotz and M Avron. On the Factors Which Determine Massive beta-Carotene Accumulation in the Halotolerant Alga Dunaliella bardawil. *Plant physiology*, 72(3):593–597, 1983.

[21] Ami Ben-Amotz and Mordhay Avron. The biotechnology of cultivating the halotolerant alga Dunaliella. *Trends in Biotechnology*, 8:121–126, 1990.

[22] John R Benemann. CO_2 mitigation with microalgae systems. *Energy Conversion and Management*, 38(96):S475–S479, 1997.

[23] Robert J Beynon, Mary K Doherty, Julie M Pratt, and Simon J Gaskell. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature methods*, 2(8):587–9, aug 2005.

[24] B. Bharathiraja, M. Chakravarthy, R. Ranjith Kumar, D. Yuvaraj, J. Jayamuthunagai, R. Praveen Kumar, and S. Palani. Biodiesel production using chemical and biological methods - A review of process, catalyst, acyl acceptor, source and process variables. *Renewable and Sustainable Energy Reviews*, 38:368–382, oct 2014.

[25] Klaus Biemann. Laying the groundwork for proteomics: Mass spectrometry from 1958 to 1988. *Journal of Proteomics*, 107:62–70, 2006.

[26] Aivett Bilbao, Emmanuel Varesio, Jeremy Luban, Caterina Strambio-De-Castillia, Gérard Hopfgartner, Markus Müller, and Frédérique Lisacek. Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics*, 15(5-6):964–980, 2015.

[27] Laurence V Bindschedler and Rainer Cramer. Quantitative plant proteomics. *Proteomics*, 11(4):756–75, mar 2011.

[28] Ian K. Blaby, Crysten E. Blaby-Haas, Nicolas Tourasse, Erik F Y Hom, David Lopez, Munevver Aksoy, Arthur Grossman, James Umen, Susan Dutcher, Mary Porter, Stephen King, George B. Witman, Mario Stanke, Elizabeth H. Harris, David Goodstein, Jane Grimwood, Jeremy Schmutz, Olivier Vallon, Sabeeha S. Merchant, and Simon Prochnik. The Chlamydomonas genome project: A decade on. *Trends in Plant Science*, 19(10):672–680, 2014.

[29] Ian K Blaby, Anne G Glaesener, Tabea Mettler, Sorel T Fitz-Gibbon, Sean D Gallaher, Bensheng Liu, Nanette R Boyle, Janette Kropat, Mark Stitt, Shannon Johnson, Christoph Benning, Matteo Pellegrini, David Casero, and Sabeeha S Merchant. Systems-Level Analysis of Nitrogen Starvation-Induced Modifications of Carbon Metabolism in a Chlamydomonas reinhardtii Starchless Mutant. *Plant Cell*, 25(11):4305–4323, nov 2013.

[30] W P Blackstock and M P Weir. Proteomics: quantitative and physical mapping of cellular proteins. *Trends in biotechnology*, 17(3):121–7, mar 1999.

[31] Jillian L. Blatti, Joris Beld, Craig A. Behnke, Michael Mendez, Stephen P. Mayfield, and Michael D. Burkart. Manipulating Fatty Acid Biosynthesis in Microalgae for Biofuel through Protein-Protein Interactions. *PLoS ONE*, 7(9):e42949, sep 2012.

[32] P H Boer, L Bonen, R W Lee, and M W Gray. Genes for respiratory chain proteins and ribosomal RNAs are present on a 16-kilobase-pair DNA species from Chlamydomonas reinhardtii mitochondria. *Proceedings of the National Academy of Sciences of the United States of America*, 82(10):3340–4, 1985.

[33] Bogdan Bogdanov and Richard D Smith. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom Rev*, 24(2):168–200, 2005.

[34] HC Bold and MJ Wynne. *Introduction to the Algae: Structure and Reproduction, 2nd Edition*, volume 5. Prentice-Hall, Inc, Toronto, 2nd edition, 1985.

[35] N. R. Boyle, M. D. Page, B. Liu, I. K. Blaby, D. Casero, J. Kropat, S. J. Cokus, A. Hong-Hermesdorf, J. Shaw, S. J. Karpowicz, S. D. Gallaher, S. Johnson, C. Benning, M. Pellegrini, A. Grossman, and S. S. Merchant. Three Acyltransferases and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol Accumulation in Chlamydomonas, 2012.

[36] Nanette R Boyle and John A Morgan. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. *BMC systems biology*, 3:4, 2009.

[37] John E Boynton, Nicholas W Gillham, Elizabeth H Harris, Jonathan P Hosler, Anrra M Johnson, Allan R Jones, Barbara L Randolph-anderson, Dominique Robertson, T E D M Klein, Katherine B Shark, and John C Sanford. Chloroplast Transformation in Chlamydomonas with High Velocity Microprojectiles. *Science*, 240:1534–1537, 1988.

[38] Liam Brennan and Philip Owende. Biofuels from microalgae - A review of technologies for production, processing, and extractions of biofuels and co-products. *Renewable and Sustainable Energy Reviews*, 14(2):557–577, feb 2010.

[39] Ghislain Breton and Steve a Kay. Circadian rhythms lit up in Chlamydomonas. *Genome biology*, 7(4):215, jan 2006.

[40] PA Broady, EA Flint, WA Nelson, Cassie Coope V, MD De Winton, PM Novis, and DP Gordon. Phyla Chlorophyta and Charophyta: green algae. In *New Zealand Inventory of Biodiversity: Volume Three: Kingdoms Bacteria, Protozoa, Chromista, Plantae, Fungi*, pages 347–381. Canterbury University Press, Christchurch, New Zealand, 2012.

[41] Virginie Brun, Christophe Masselon, Jérôme Garin, and Alain Dupuis. Isotope dilution strategies for absolute quantitative proteomics. *Journal of proteomics*, 72(5):740–9, jul 2009.

[42] David A Cairns, Douglas Thompson, David N Perkins, Anthea J Stanley, Peter J Selby, and Rosamonde E Banks. Proteomic profiling using mass spectrometry–does normalising by total ion current potentially mask some biological differences? *Proteomics*, 8(1):21–7, jan 2008.

[43] M. Cambra Sánchez, J., Álvarez Cobelas, M. & Aboal Sanjurjo. Lista florística y bibliográfica de los clorófitos (Chlorophyta) de la Península Ibérica, Islas Baleares e Islas Canarias. Technical report, Asociación Española de Limnología, Burgos (Spain), 1998.

[44] I Cărăuş. *Algae of Romania. A distributional checklist of actual algae. Version 2.4*. University of Bacău - Romania, 4 edition, 2017.

[45] Natalie Castellana and Vineet Bafna. Proteogenomics to discover the full coding content of genomes: A computational perspective. *Journal of Proteomics*, 73(11):2124–2135, 2010.

[46] Madeli Castruita, David Casero, Steven J Karpowicz, Janette Kropat, Astrid Vieler, Scott I Hsieh, Weihong Yan, Shawn Cokus, Joseph a Loo, Christoph Benning, Matteo Pellegrini, and Sabeeha S Merchant. Systems biology approach in Chlamydomonas reveals connections between copper nutrition and multiple metabolic steps. *The Plant cell*, 23(4):1273–92, apr 2011.

[47] Jennifer A. Cham, Luca Bianco, and Conrad Bessant. Free computational resources for designing selected reaction monitoring transitions. *Proteomics*, 10(6):1106–1126, 2010.

[48] Antoine Champagne and Marc Boutry. Proteomics of nonmodel plant species. *PROTEOMICS*, 13(3-4):663–673, feb 2013.

[49] Yanmei Chen, Wolfgang Hoehenwarter, and Wolfram Weckwerth. Comparative analysis of phytohormone-responsive phosphoproteins in arabidopsis thaliana using tio2-phosphopeptide enrichment and mass accuracy precursor alignment. *The Plant Journal*, 63(1):1–17, 2010.

[50] Yusuf Chisti. Biodiesel from microalgae beats bioethanol. *Trends in biotechnology*, 26:126–131, 2008.

[51] Pierre Collet, Laurent Lardon, Arnaud Hélias, Stéphanie Bricout, Isabelle Lombaert-Valot, Béatrice Perrier, Olivier Lépine, Jean-Philippe Steyer, and Olivier Bernard. Biodiesel from microalgae - Life cycle assessment and recommendations for potential improvements. *Renewable Energy*, 71:525–533, nov 2014.

[52] R.G. Cooks, T. Ast, and Md.A. Mabud. Collisions of polyatomic ions with surfaces. *International Journal of Mass Spectrometry and Ion Processes*, 100:209–265, oct 1990.

[53] Jürgen Cox and Matthias Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26(12):1367–72, 2008.

[54] R. Craig, J. C. Cortens, D. Fenyo, and R. C. Beavis. Using annotated peptide mass spectrum libraries for protein identification. *Journal of Proteome Research*, 5(8):1843–1849, 2006.

[55] Robertson Craig and Ronald C. Beavis. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, 20(9):1466–1467, 2004.

[56] Robertson Craig, John P Cortens, and Ronald C Beavis. Open source system for analyzing, validating, and storing protein identification data. *Journal of proteome research*, 3(6):1234–42, jan 2004.

[57] Robertson Craig, John P. Cortens, and Ronald C. Beavis. The use of proteotypic peptide libraries for protein identification. *Rapid Communications in Mass Spectrometry*, 19(13):1844–1850, 2005.

[58] Frederick R Cross and James G Umen. The Chlamydomonas cell cycle. *The Plant journal : for cell and molecular biology*, 82(3):370–92, may 2015.

[59] Attila Csordas, Rui Wang, Daniel Ríos, Florian Reisinger, Joseph M. Foster, Douglas J. Slotta, Juan Antonio Vizcaíno, and Henning Hermjakob. From Peptidome to PRIDE: Public proteomics data migration at a large scale. *Proteomics*, 13(10-11):1692–1695, 2013.

[60] Lyris M. F. de Godoy, Jesper V. Olsen, Jürgen Cox, Michael L. Nielsen, Nina C. Hubner, Florian Fröhlich, Tobias C. Walther, and Matthias Mann. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455(7217):1251–1254, oct 2008.

[61] Marcin Dębowski, Marcin Zieliński, Anna Grala, and Magda Dudek. Algae biomass as an alternative substrate in biogas production technologies - Review. *Renewable and Sustainable Energy Reviews*, 27:596–604, nov 2013.

[62] Ayhan Demirbas. Political, economic and environmental impacts of biofuels: A review. *Applied Energy*, 86:S108–S117, 2009.

[63] Ayhan Demirbas and M. Fatih Demirbas. Importance of algae oil as a source of biodiesel. *Energy Conversion and Management*, 52(1):163–170, jan 2011.

[64] Xiao-Dong Deng, Bo Gu, Ya-Jun Li, Xin-Wen Hu, Jian-Chun Guo, and Xiao-Wen Fei. The Roles of acyl-CoA: Diacylglycerol Acyltransferase 2 Genes in the Biosynthesis of Triacylglycerols by the Green Algae Chlamydomonas reinhardtii. *Molecular Plant*, 5(4):945–947, jul 2012.

[65] F. Desiere. The PeptideAtlas project. *Nucleic Acids Research*, 34(90001):D655–D658, 2006.

[66] K Uma Devi, G Swapna, and S Suneetha. *Microalgae in Bioremediation: Sequestration of Greenhouse Gases, Clearout of Fugitive Nutrient Minerals, and Subtraction of Toxic Elements from Waters*. Elsevier Insights. Elsevier, 2014.

[67] Allison Doerr. DIA mass spectrometry. *Nature Publishing Group*, 12(1):35, 2015.

[68] L. Domokos, D. Hennberg, and B. Weimann. Computer-aided identification of compounds by comparison of mass spectra. *Analytica Chimica Acta*, 165(C):61–74, 1984.

[69] Meng-Qiu Dong, John D. Venable, Nora Au, Tao Xu, Sung Kyu Park, Daniel Cociorva, Jeffrey R. Johnson, Andrew Dillin, and John R. Yates. Quantitative Mass Spectrometry Identifies Insulin Signaling Targets in C. elegans. *Science*, 317(5838), 2007.

[70] Viktoria Dorfer, Peter Pichler, Thomas Stranzl, Johannes Stadlmann, Thomas Taus, Stephan Winkler, and Karl Mechtler. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *Journal of proteome research*, 13(8):3679–84, aug 2014.

[71] Martin et al. Dragosits. Adaptive laboratory evolution – principles and applications for biotechnology. *Microbial Cell Factories*, 12(1):64, 2013.

[72] Romain Drissi, Marie Line Dubois, and Fran??ois Michel Boisvert. Proteomics methods for subcellular proteome analysis. *FEBS Journal*, 280(22):5626–5634, 2013.

[73] Alain Dupuis, Jacques Antoine Hennekinne, Jérôme Garin, and Virginie Brun. Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics*, 8(22):4633–4636, 2008.

[74] W T Ebersold, R P Levine, E E Levine, and M A Olmsted. Linkage maps in Chlamydomonas reinhardi. *Genetics*, 47(5):531–43, may 1962.

[75] Volker Egelhofer, Wolfgang Hoehenwarter, David Lyon, Wolfram Weckwerth, and Stefanie Wienkoop. Using ProtMAX to create high-mass-accuracy precursor alignments from label-free quantitative mass spectrometry data generated in shotgun proteomics experiments. *Nature protocols*, 8(3):595–601, 2013.

[76] Jarrett D Egertson, Andreas Kuehn, Gennifer E Merrihew, Nicholas W Bateman, Brendan X MacLean, Ying S Ting, Jesse D Canterbury, Donald M Marsh, Markus Kellmann, Vlad Zabrouskov, Christine C Wu, and Michael J MacCoss. Multiplexed MS/MS for improved data-independent acquisition. *Nature methods*, 10(8):744–6, aug 2013.

[77] J E Elias, F D Gibbons, O D King, F P Roth, and S P Gygi. Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nature Biotechnology*, 22(2):214–219, 2004.

[78] Lisa Elviri. *Tandem Mass Spectrometry - Applications and Principles*. InTech, feb 2012.

[79] J Eng. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of The American Society for Mass Spectrometry*, 5(11):976–989, 1994.

[80] Jayson A. Falkner, James A. Hill, and Philip C. Andrews. Proteomics FASTA archive and reference resource. *Proteomics*, 8(9):1756–1757, 2008.

[81] Jilian Fan, Carl Andre, and Changcheng Xu. A chloroplast pathway for the de novo biosynthesis of triacylglycerol in Chlamydomonas reinhardtii. *FEBS letters*, 585:1985–1991, 2011.

[82] Jilian Fan, Chengshi Yan, Carl Andre, John Shanklin, Jörg Schwender, and Changcheng Xu. Oil accumulation is controlled by carbon precursor supply for fatty acid synthesis in Chlamydomonas reinhardtii. *Plant & cell physiology*, 53(8):1380–90, aug 2012.

[83] Jilian Fan, Chengshi Yan, Carl Andre, John Shanklin, Jörg Schwender, and Changcheng Xu. Oil accumulation is controlled by carbon precursor supply for fatty acid synthesis in Chlamydomonas reinhardtii. *Plant & cell physiology*, 53:1380–90, 2012.

[84] J. Fenn, M Mann, C. Meng, S. Wong, and C. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, oct 1989.

[85] N Focks and C Benning. wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant physiology*, 118(1):91–101, sep 1998.

[86] Barbara E. Frewen, Gennifer E. Merrihew, Christine C. Wu, William Stafford Noble, and Michael J. MacCoss. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Analytical Chemistry*, 78(16):5678–5684, 2006.

[87] Sebastien Gallien, Adele Bourmaud, Sang Yoon Kim, and Bruno Domon. Technical considerations for large-scale parallel reaction monitoring analysis. *Journal of proteomics*, 100:147–59, apr 2014.

[88] Benjamin A. Garcia, Nertila Siuti, C. Eric Thomas, Craig A. Mizzen, and Neil L. Kelleher. Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry. *International Journal of Mass Spectrometry*, 259(1-3):184–196, jan 2007.

[89] Olivier Garsmeur, James C Schnable, Ana Almeida, Cyril Jourda, Angélique D'Hont, and Michael Freeling. Two evolutionarily distinct classes of paleopolyploidy. *Molecular biology and evolution*, 31(2):448–54, feb 2014.

[90] J. W. Gauthier, T. R. Trautman, and D. B. Jacobson. Sustained off-resonance irradiation for collision-activated dissociation involving Fourier transform mass spectrometry. Collision-activated dissociation technique that emulates infrared multiphoton dissociation. *Analytica Chimica Acta*, 246(1):211–225, 1991.

[91] Lewis Y Geer, Sanford P Markey, Jeffrey A Kowalak, Lukas Wagner, Ming Xu, Dawn M Maynard, Xiaoyu Yang, Wenyao Shi, and Stephen H Bryant. Open mass spectrometry search algorithm. *J Proteome Res*, 3(5):958–964, 2004.

[92] Sa Geng, Peter De Hoff, James G. Umen, AM Callahan, and H Gruber. Evolution of Sexes from an Ancestral Mating-Type Specification Pathway. *PLoS Biology*, 12(7):e1001904, jul 2014.

[93] S. A. Gerber, J. Rush, O. Stemman, M. W. Kirschner, and S. P. Gygi. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences*, 100(12):6940–6945, 2003.

[94] Stéphanie Gérin, Grégory Mathy, and Fabrice Franck. Modeling the dependence of respiration and photosynthesis upon light, acetate, carbon dioxide, nitrate and ammonium in Chlamydomonas reinhardtii using design of experiments and multiple regression. *BMC systems biology*, 8(1):96, jan 2014.

[95] Sina Ghaemmaghami, Won-Ki Huh, Kiowa Bower, Russell W. Howson, Archana Belle, Noah Dephoure, Erin K. O'Shea, and Jonathan S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, oct 2003.

[96] Maria L Ghirardi, Paul W King, David W Mulder, Carrie Eckert, Alexandra Dubini, Pin-Ching Maness, and Jianping Yu. *Hydrogen Production by Water Biophotolysis*, volume 38. Springer Netherlands, 2014.

[97] Francis D Gibbons, Joshua E Elias, Steven P Gygi, and Frederick P Roth. SILVER helps assign peptides to tandem mass spectra using intensity-based scoring. *Journal of the American Society for Mass Spectrometry*, 15(6):910–912, 2004.

[98] L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics*, 11(6):O111.016717–O111.016717, 2012.

[99] Joshua M Gilmore and Michael P Washburn. Advances in shotgun proteomics and the analysis of membrane proteomes. *Journal of proteomics*, 73(11):2078–91, oct 2010.

[100] Mirko Glinski and Wolfram Weckwerth. The role of mass spectrometry in plant systems biology. *Mass Spectrom Rev*, 25(2):173–214, 2006.

[101] David González-Ballester, David Casero, Shawn Cokus, Matteo Pellegrini, Sabeeha S Merchant, and Arthur R Grossman. RNA-seq analysis of sulfur-deprived Chlamydomonas

cells reveals aspects of acclimation critical for cell survival. *The Plant cell*, 22:2058–2084, 2010.

[102] U Goodenough, J.E. Boynton, N.W. Gillham, E.H. Harris, J.P. Hosler, A.M. Johnson, A.R. Jones, B.L. Randolph-Anderson, D. Robertson, T.M. Klein, K.B. Shark, J.C. Sanford, E.H. Harris, B. Huang, Z. Ramanis, D.J.L. Luck, R. Kamiya, G.B. Witman, K.L Kindle, G. Kreimer, M. Melkonian, P.A. Lefebvre, J.L. Rosenbaum, R.P Levine, R.A. Lewin, M.S. Livingstone, P.P. Sziber, W.G. Quinn, D.J.L. Luck, S.M. Pasquale, U.W. Goodenough, R. Sager, and M.L. Sogin. Green yeast. *Cell*, 70(4):533–538, aug 1992.

[103] Carrie Goodson, Robyn Roth, Zi Teng Wang, and Ursula Goodenough. Structural correlates of cytoplasmic and chloroplast lipid body synthesis in Chlamydomonas reinhardtii and stimulation of lipid body production with acetate boost. *Eukaryotic Cell*, 10(12):1592–1606, 2011.

[104] Noelle M Griffin, Jingyi Yu, Fred Long, Phil Oh, Sabrina Shore, Yan Li, Jim A Koziol, and Jan E Schnitzer. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature biotechnology*, 28(1):83–9, jan 2010.

[105] Christian H. Gross, Laura P. W. Ranum, and Paul A. Lefebvre. Extensive restriction fragment length polymorphisms in a new isolate ofChlamydomonas reinhardtii. *Current Genetics*, 13(6):503–508, jun 1988.

[106] A Grossman. Acclimation of Chlamydomonas reinhardtii to its nutrient environment. *Protist*, 151:201–224, 2000.

[107] Arthur R Grossman, Martin Croft, Vadim N Gladyshev, Sabeeha S Merchant, Matthew C Posewitz, Simon Prochnik, and Martin H Spalding. Novel metabolism in Chlamydomonas through the lens of genomics. *Curr Opin Plant Biol*, 10(2):190–198, apr 2007.

[108] Arthur R Grossman, Elizabeth E Harris, Charles Hauser, Paul A Lefebvre, Diego Martinez, Dan Rokhsar, Jeff Shrager, Carolyn D Silflow, David Stern, Olivier Vallon, and Zhaoduo Zhang. Chlamydomonas reinhardtii at the crossroads of genomics. *Eukaryotic cell*, 2(6):1137–50, dec 2003.

[109] Sarman Oktovianus Gultom and Bo Hu. Review of microalgae harvesting via co-pelletization with filamentous fungus. *Energies*, 6(11):5921–5939, 2013.

[110] R.K. Gupta. *Algae of India Vol 2 - A Checklist of Chlorophyceae Xanthophyceae, Chrysophyceae and Euglenophyceae*. Botanical Survey of India, 2012.

[111] S P Gygi, G L Corthals, Y Zhang, Y Rochon, and R Aebersold. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America*, 97(17):9390–5, aug 2000.

[112] S P Gygi, B Rist, S A Gerber, F Turecek, M H Gelb, and R Aebersold. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology*, 17(10):994–9, oct 1999.

[113] Syed Haider, Benoit Ballester, Damian Smedley, Junjun Zhang, Peter Rice, and Arek Kasprzyk. BioMart central portal - Unified access to biological data. *Nucleic Acids Research*, 37(SUPPL. 2):23–27, 2009.

[114] Patrick C. Hallenbeck and John R. Benemann. Biological hydrogen production; fundamentals and limiting processes. *International Journal of Hydrogen Energy*, 27(11):1185–1193, 2002.

[115] Ben Hankamer, Florian Lehr, Jens Rupprecht, Jan H. Mussgnug, Clemens Posten, and Olaf Kruse. Photosynthetic biomass and H 2 production by green algae: from bioengineering to bioreactor scale-up. *Physiologia Plantarum*, 131(1):10–21, sep 2007.

[116] Elizabeth H. Harris. *The Chlamydomonas Sourcebook. A Comprehensive Guide to Biology and Laboratory Use*. Academic Press Inc, 1 edition, 1989.

[117] Alex G. Harrison, Alex B. Young, Christian Bleiholder, Sándor Suhai, and Béla Paizs. Scrambling of sequence information in collision-induced dissociation of peptides. *Journal of the American Chemical Society*, 128(32):10364–10365, 2006.

[118] Kevin J Hart. Software Development for Mass Spectrometric Analysis. *Rapid Communications in Mass Spectrometry*, 10(December 1995):393–398, 1996.

[119] Ralf G Hartmer, Desmond A Kaplan, Carsten Stoermer, Markus Lubeck, and Melvin A Park. Data-dependent electron transfer dissociation of large peptides and medium size proteins in a QTOF instrument on a liquid chromatography timescale. *Rapid communications in mass spectrometry : RCM*, 23(15):2273–82, aug 2009.

[120] Razif Harun, Jason W S Yip, Selvakumar Thiruvenkadam, Wan A W A K Ghani, Tamara Cherrington, and Michael K Danquah. Algal biomass conversion to bioethanol - a step-by-step assessment. *Biotechnology Journal*, 9(1):73–86, jan 2014.

[121] Dorothea Hemme, Daniel Veyel, Timo Mühlhaus, Frederik Sommer, Jessica Jüppner, Ann-Katrin Unger, Michael Sandmann, Ines Fehrle, Stephanie Schönfelder, Martin Steup, Stefan Geimer, Joachim Kopka, Patrick Giavalisco, and Michael Schroda. Systems-wide analysis of acclimation responses to long-term heat stress and recovery in the photosynthetic model organism Chlamydomonas reinhardtii. *The Plant cell*, 26(11):4270–97, nov 2014.

[122] Anja Hemschemeier, David Casero, Bensheng Liu, Christoph Benning, Matteo Pellegrini, Thomas Happe, and Sabeeha S Merchant. {COPPER} {RESPONSE} {REGULATOR}1-Dependent and -Independent Responses of the Chlamydomonas reinhardtii Transcriptome to Dark Anoxia. *Plant Cell*, 25(9):3186–3211, sep 2013.

[123] Anja Hemschemeier, Anastasios Melis, and Thomas Happe. Analytical approaches to photobiological hydrogen production in unicellular green algae. *Photosynthesis research*, 102(2-3):523–40, 2009.

[124] Johannes a. Hewel, Sadhna Phanse, Jian Liu, Nicolas Bousette, Anthony Gramolini, and Andrew Emili. Targeted protein identification, quantification and reporting for high-resolution nanoflow targeted peptide monitoring. *Journal of Proteomics*, 81:159–172, 2013.

[125] Franz Hillenkamp, Michael Karas, Ronald C. Beavis, and Brian T. Chait. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical Chemistry*, 63(24):1193A–1203A, dec 1991.

[126] Wolfgang Hoehenwarter, Yanmei Chen, Luis Recuenco-Munoz, Stefanie Wienkoop, and Wolfram Weckwerth. Functional analysis of proteins and protein species using shotgun proteomics and linear mathematics. *Amino acids*, 41(2):329–41, jul 2011.

[127] Wolfgang Hoehenwarter, Joost T van Dongen, Stefanie Wienkoop, Matthias Steinfath, Jan Hummel, Alexander Erban, Ronan Sulpice, Babette Regierer, Joachim Kopka, Peter Geigenberger, and Wolfram Weckwerth. A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. *Proteomics*, 8(20):4214–4225, 2008.

[128] Michael R Hoopmann and Robert L Moritz. Current algorithmic solutions for peptide-based proteomics data generation and identification. *Current opinion in biotechnology*, 24(1):31–8, feb 2013.

[129] Jue Liang Hsu, Sheng Yu Huang, Nan Haw Chow, and Shu Hui Chen. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Analytical Chemistry*, 75(24):6843–6852, 2003.

[130] H.J. Hu, M. Yang, L.M. Luo, and S. Chen. *Flora algarum sinicarum aquae dulcis Tomus XX Chlorophyta Chlorophyceae Order Volvocales (II) Genus Chlamydomonas.* Science Press, Beijing, 2015.

[131] Qiang Hu, Milton Sommerfeld, Eric Jarvis, Maria Ghirardi, Matthew Posewitz, Michael Seibert, Al Darzins, Gilles Peltier, Q Hu, M Sommerfeld, E Jarvis, M Ghirardi, M Posewitz, M Seibert, RH Wijffels, MJ Barbosa, DR Georgianna, SP Mayfield, AWD Larkum, IL Ross, O Kruse, B Hankamer, SG Ball, L Dirick, A Decq, JC Martiat, RF Matagne, M Siaut, S Cuine, C Cagnon, B Fessler, M Nguyen, P Carrier, F Delrue, Y Li-Beisson, PA Setier, C Sahut, A Roubaud, AK Froment, SS Merchant, J Kropat, BS Liu, J Shaw, J Warakanont, G Peltier, GW Schmidt, A Grossman, DD Wykoff, JP Davies, A Melis, AR Grossman, C Robaglia, M Thomas, C Meyer, P Coello, SJ Hey, NG Halford, D Gonzalez-Ballester, SV Pollock, W Pootakham, AR Grossman, NR Boyle, MD Page, BS Liu, IK Blaby, D Casero, J Kropat, S Aranda, A Laguna, S Luna, M Kajikawa, Y Sawaragi, H Shinkawa, T Yamano, A Ando, M Kato, V Chochois, L Constans, D Dauvillée, A Beyly, M Solivérès, S Ball, S Garrett, J Broach, H Moriya, Y Shimizu-Yoshida, A Omori, S Iwashita, M Katoh, A Sakai, W Becker, HG Joost, EH Harris, L Cournac, G Latouche, Z Cerovic, K Redding, J Ravenel, G Peltier, U Goodenough, I Blaby, D Casero, SD Gallaher, C Goodson, S Johnson, C Goodson, R Roth, ZT Wang, U Goodenough, K Bisova, DM Krylov, JG Umen, M Schulz-Raffelt, M Lodha, M Schroda, M Kajikawa, Y Sawaragi, H Shinkawa, T Yamano, A Ando, M Kato, D Kim, VO Ntui, N Zhang, L Xiong, S Kassis, T Melhuish, RS Annan, SL Chen, JC Lee, GP Livi, S Himpel, P Panzer, K Eirmbter, H Czajkowska, M Sayed, LC Packman, PA Lochhead, G Sibbet, N Morrice, V Cleghon, M Soundararajan, AK Roos, P Savitsky, P Filippakopoulos, AN Kettenbach, JV Olsen, JF Han, D Miranda-Saavedra, N Luebbering, A Singh, G Sibbet, MAJ Ferguson, SC Mithoe, PJ Boersema, L Berke, B Snel, AJR Heck, FLH Menke, HX Wang, B Gau, WO Slade, M Juergens, P Li, LM Hicks, WA Wilson, PJ Roach, M Montero, E Baroja-Fernandez, FJ Munoz, G Eydallin, P Lee, BR Cho, HS Joo, JS Hahn, M Malcher, S Schladebeck, HU Mosch, JR Rohde, R Bastidas, R Puria, ME Cardenas, S Wullschleger, R Loewith, MN Hall, E Baena-Gonzalez, J Sheen, DE Martin, A Soulard, MN Hall, CH Tsai, J Warakanont, T Takeuchi, BB Sears, ER Moellering,

C Benning, K-V Dang, J Plet, D Tolleter, M Jokel, S Cuine, P Carrier, D Tolleter, B Ghysels, J Alric, D Petroutsos, I Tolstygina, D Krawietz, D Dauvillee, O Stampacchia, J Girard-Bascou, JD Rochaix, P Berthold, R Schmitt, W Mages, KL Kindle, K Katoh, K Misawa, K Kuma, T Miyata, M Gouy, S Guindon, O Gascuel, J Felsenstein, K Tamura, D Peterson, N Peterson, G Stecher, M Nei, S Kumar, CM Liu, F Willmund, JP Whitelegge, S Hawat, B Knapp, M Lodha, Y Li-Beisson, B Shorrosh, F Beisson, MX Andersson, V Arondel, PD Bates, U Schreiber, U Schliwa, and W Bilger. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant Journal*, 54(4):621–639, may 2008.

[132] Jan Hummel, Michaela Niemann, Stefanie Wienkoop, Waltraud Schulze, Dirk Steinhauser, Joachim Selbig, Dirk Walther, and Wolfram Weckwerth. ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC bioinformatics*, 8:216, 2007.

[133] Donald F. Hunt, W. M. Bone, J. Shabanowitz, J. Rhodes, and J. M. Ballard. Sequence analysis of oligopeptides by secondary ion/collision activated dissociation mass spectrometry. *Analytical Chemistry*, 53(11):1704–1706, sep 1981.

[134] International Energy Agency (IEA). IEA Technology essentials-biofuel production. *International Energy Agency*, pages 1–4, 2007.

[135] H Ishida, S Shimizu, a Makino, and T Mae. Light-dependent fragmentation of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase in chloroplasts isolated from wheat leaves. *Planta*, 204(3):305–9, mar 1998.

[136] Gabriel O James, Charles H Hocart, Warwick Hillier, Hancai Chen, Farzaneh Kordbacheh, G Dean Price, and Michael A Djordjevic. Fatty acid profiling of Chlamydomonas reinhardtii under nitrogen deprivation. *Bioresource technology*, 102:3343–3351, 2011.

[137] K.R. Jennings. Collision-Induced Decompositions of aromatic molecular ions. *International Journal of Mass Spectrometry and Ion Physics*, 1:227–235, 1968.

[138] Li Ji, Tanya Barrett, Oluwabukunmi Ayanbule, Dennis B. Troup, Dmitry Rudnev, Rolf N. Muertter, Maxim Tomashevsky, Alexandra Soboleva, and Douglas J. Slotta. NCBI Peptidome: A new repository for mass spectrometry proteomics data. *Nucleic Acids Research*, 38(SUPPL.1):731–735, 2009.

[139] Richard S. Johnson, Stephen A. Martin, Klaus. Biemann, John T. Stults, and J. Throck. Watson. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical Chemistry*, 59(21):2621–2625, nov 1987.

[140] Xenie Johnson and Jean Alric. Central carbon metabolism and electron transport in Chlamydomonas reinhardtii: metabolic constraints for carbon partitioning between oil and starch. *Eukaryotic cell*, 12(6):776–93, jun 2013.

[141] Jesus Jorrin-Novo, Setsuko Komatsu, Wolfram Weckwerth, and Stefanie Wienkoop. *Plant Proteomics: Methods and Protocols*. Phospho- proteomics IN. Humana Press, Hatfield, Hertfordshire, AL10 9AB, UK, second edition, 2014.

[142] Peter R Jungblut, Hermann G Holzhütter, Rolf Apweiler, and Hartmut Schlüter. The speciation of the proteome. *Chemistry Central journal*, 2(1):16, jan 2008.

[143] Nadine Jungnick, Yunbing Ma, Bratati Mukherjee, Julie C Cronan, Dequantarius J Speed, Susan M Laborde, David J Longstreth, and James V Moroney. The carbon concentrating mechanism in Chlamydomonas reinhardtii: finding the missing pieces. *Photosynthesis research*, 121(2-3):159–73, sep 2014.

[144] Lukas Käll, John D Storey, Michael J Maccoss, and William Stafford Noble. Posterior Error Probabilities and False Discovery Rates : Two Sides of the Same Coin Posterior Error Probabilities and False Discovery Rates : Two Sides of the Same Coin. *Journal of Proteome Research*, pages 40–44, 2008.

[145] Anastasia Kalli, Michael J Sweredoski, and Sonja Hess. Data-dependent middle-down nano-liquid chromatography-electron capture dissociation-tandem mass spectrometry: an application for the analysis of unfractionated histones. *Analytical chemistry*, 85(7):3501–7, apr 2013.

[146] Pushpa Kathir, Matthew LaVoie, William J Brazelton, Nancy A Haas, Paul A Lefebvre, and Carolyn D Silflow. Molecular map of the Chlamydomonas reinhardtii nuclear genome. *Eukaryotic cell*, 2(2):362–79, apr 2003.

[147] Lani C Keller, Edwin P Romijn, Ivan Zamora, John R Yates, and Wallace F Marshall. Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Curr Biol*, 15(12):1090–1098, jun 2005.

[148] Karen L. Kindle, Rogene A. Schnell, Emilio Fermindez, and Paul A. Lefebvre. Stable Nuclear Transformation of Chlamydomonas Using the Chlamydomonas Gene for Nitrate Reductase. *Cell Biology*, 109(1):2589–2601, 1989.

[149] Bala Kiran, Ritunesh Kumar, and Devendra Deshmukh. Perspectives of microalgal biofuels as a renewable source of energy. *Energy Conversion and Management*, jun 2014.

[150] Torsten Kleffmann, Doris Russenberger, Anne von Zychlinski, Wayne Christopher, Kimmen Sjölander, Wilhelm Gruissem, Sacha Baginsky, J.-R. Yates, N. Rolland, G. von Heijne, and et Al. The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Current biology : CB*, 14(5):354–62, mar 2004.

[151] J Klose. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, 26(3):231–43, jan 1975.

[152] Olaf Kruse and Ben Hankamer. Microalgal hydrogen production. *Current opinion in biotechnology*, 21(3):238–43, jun 2010.

[153] Takeaki Kubo, Jun Abe, Tatsuaki Saito, and Yoshihiro Matsuda. Genealogical relationships among laboratory strains of Chlamydomonas reinhardtii as inferred from matrix metalloprotease genes. *Current Genetics*, 41(2):115–122, may 2002.

[154] M. La Russa, C. Bogen, A. Uhmeyer, A. Doebbe, E. Filippone, O. Kruse, and J.H. Mussgnug. Functional analysis of three type-2 DGAT homologue genes for triacylglycerol

production in the green microalga Chlamydomonas reinhardtii. *Journal of Biotechnology*, 162(1):13–20, 2012.

[155] Henry Lam and Ruedi Aebersold. Spectral library searching for peptide identification via tandem MS. *Methods in molecular biology (Clifton, N.J.)*, 604:95–103, jan 2010.

[156] Henry Lam and Ruedi Aebersold. Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. *Methods (San Diego, Calif.)*, 54(4):424–31, aug 2011.

[157] Henry Lam, Eric W. Deutsch, and Ruedi Aebersold. Artificial decoy spectral libraries for false discovery rate estimation in spectral library searching in proteomics. *Journal of Proteome Research*, 9(1):605–610, 2010.

[158] Henry Lam, Eric W Deutsch, James S Eddes, Jimmy K Eng, Stephen E Stein, and Ruedi Aebersold. Building consensus spectral libraries for peptide identification in proteomics. *Nature methods*, 5(10):873–5, oct 2008.

[159] Maya D Lambreva, Maria Teresa Giardi, Irene Rambaldi, Amina Antonacci, Sandro Pastorelli, Ivo Bertalan, Ivan Husu, Udo Johanningmeier, and Giuseppina Rea. A Powerful Molecular Engineering Tool Provided Efficient Chlamydomonas Mutants as Bio-Sensing Elements for Herbicides Detection. *Plos One*, 8(4):e61851, apr 2013.

[160] Vinzenz Lange, Paola Picotti, Bruno Domon, and Ruedi Aebersold. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology*, 4:222, 2008.

[161] Julia Laskin, Eduard Denisov, and Jean Futrell. A Comparative Study of Collision-Induced and Surface-Induced Dissociation. 1. Fragmentation of Protonated Dialanine. *Journal of the American Chemical Society*, 122(40):9703–9714, oct 2000.

[162] Julia Laskin and Jean H Futrell. Activation of large ions in FT-ICR mass spectrometry. *Mass spectrometry reviews*, 24(2):135–67, jan 2004.

[163] Kyle J Lauersen, Hanna Berger, Jan H Mussgnug, and Olaf Kruse. Efficient recombinant protein production and secretion from nuclear transgenes in Chlamydomonas reinhardtii. *Journal of Biotechnology*, 167(2):101–110, aug 2013.

[164] Michel Lavoie, Sebastien Sabatier, Jacqueline Garnier-Laplace, and Claude Fortin. {URANIUM} {ACCUMULATION} {AND} {TOXICITY} {IN} {THE} {GREEN} {ALGA} {CHLAMYDOMONAS} {REINHARDTII} {IS} {MODULATED} {BY} {pH}. *Environmental Toxicology and Chemistry*, 33(6):1372–1379, jun 2014.

[165] Bongsoo Lee, Gang-Guk Choi, Yoon-E. Choi, Minji Sung, Min S Park, and Ji-Won Yang. Enhancement of lipid productivity by ethyl methane sulfonate-mediated random mutagenesis and proteomic analysis in Chlamydomonas reinhardtii. *Korean Journal of Chemical Engineering*, 31(6):1036–1042, jun 2014.

[166] Seungmook Lee, Min-Seok Kwon, Hyoung-Joo Lee, Young-Ki Paik, Haixu Tang, Jae K Lee, and Taesung Park. Enhanced peptide quantification using spectral count clustering and cluster abundance. *BMC bioinformatics*, 12:423, jan 2011.

[167] Ute Lehmann, Stefanie Wienkoop, Hendrik Tschoep, and Wolfram Weckwerth. If the antibody fails–a mass Western approach. *Plant J*, 55(6):1039–1046, sep 2008.

[168] Florian Lehr and Clemens Posten. Closed photo-bioreactors as tools for biofuel production. *Current Opinion in Biotechnology*, 20(3):280–285, jun 2009.

[169] R P Levine and W T Ebersold. The Genetics and Cytology of Chlamydomonas. *Annual Review of Microbiology*, 14(1):197–216, oct 1960.

[170] Guo Zhong Li, Johannes P C Vissers, Jeffrey C. Silva, Dan Golick, Marc V. Gorenstein, and Scott J. Geromanos. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics*, 9(6):1696–1719, 2009.

[171] Jin Billy Li, Shaoping Lin, Honggui Jia, Hongmin Wu, Bruce A. Roe, David Kulp, Gary D. Stormo, and Susan K. Dutcher. Analysis of Chlamydomonas reinhardtii Genome Structure Using Large-Scale Sequencing of Regions on Linkage Groups I and III. *The Journal of Eukaryotic Microbiology*, 50(3):145–155, may 2003.

[172] Xiaobo Li, Ru Zhang, Weronika Patena, Spencer S Gang, Sean R Blum, Nina Ivanova, Rebecca Yue, Jacob M Robertson, Paul A Lefebvre, Sorel T Fitz-Gibbon, Arthur R Grossman, and Martin C Jonikas. An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in Chlamydomonas reinhardtii. *The Plant cell*, 28(2):367–87, 2016.

[173] Yanqun Li, Mark Horsman, Nan Wu, Christopher Q Lan, and Nathalie Dubois-Calero. Biofuels from microalgae. *Biotechnology progress*, 24(4):815–20, mar 2008.

[174] Yanna Liang, Nicolas Sarkany, and Yi Cui. Biomass and lipid productivities of Chlorella vulgaris under autotrophic, heterotrophic and mixotrophic growth conditions. *Biotechnology Letters*, 31(7):1043–1049, 2009.

[175] Lujian Liao, Daniel B McClatchy, and John R Yates. Shotgun proteomics in neuroscience. *Neuron*, 63(1):12–26, jul 2009.

[176] A J Link, J Eng, D M Schieltz, E Carmack, G J Mize, D R Morris, B M Garvik, and J R Yates. Direct analysis of protein complexes using mass spectrometry. *Nature biotechnology*, 17(7):676–82, jul 1999.

[177] Daniel P. Little, J. Paul. Speir, Michael W. Senko, Peter B. O'Connor, and Fred W. McLafferty. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Analytical Chemistry*, 66(18):2809–2815, sep 1994.

[178] Bensheng Liu, Astrid Vieler, Chao Li, A Daniel Jones, and Christoph Benning. Triacylglycerol profiling of microalgae Chlamydomonas reinhardtii and Nannochloropsis oceanica. *Bioresource Technology*, 146:310–316, oct 2013.

[179] Hongbin Liu, Rovshan G. Sadygov, and John R. Yates. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Analytical Chemistry*, 76(14):4193–4201, jul 2004.

[180] Martin Lohr, Chung-Soon Im, and Arthur R Grossman. Genome-based examination of chlorophyll and carotenoid biosynthesis in Chlamydomonas reinhardtii. *Plant physiology*, 138(1):490–515, may 2005.

[181] Joseph Longworth, Josselin Noirel, Jagroop Pandhal, Phillip C Wright, and Seetharaman Vaidyanathan. HILIC- and SCX-based quantitative proteomics of Chlamydomonas reinhardtii during nitrogen starvation induced lipid and carbohydrate accumulation. *Journal of proteome research*, 11:5959–71, 2012.

[182] David Lopez, David Casero, Shawn J Cokus, Sabeeha S Merchant, and Matteo Pellegrini. Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics*, 12(1):282, 2011.

[183] J. L. López. Two-dimensional electrophoresis in proteome expression analysis. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 849(1-2):190–202, 2007.

[184] Md.A. Mabud, Michael J. Dekrey, and R. Graham Cooks. Surface-induced dissociation of molecular ions. *International Journal of Mass Spectrometry and Ion Processes*, 67(3):285–294, nov 1985.

[185] Brendan Maclean, Daniela M Tomazela, Nicholas Shulman, Matthew Chambers, Gregory L Fin, Barbara Frewen, Randall Kern, David L Tabb, Daniel C Liebler, and Michael J Maccoss. Skyline : An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics*, pages 9–10, 2010.

[186] Jyoti Prakash Maity, Jochen Bundschuh, Chien-Yen Chen, and Prosun Bhattacharya. Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives - A mini review. *Energy*, apr 2014.

[187] Parag Mallick and Bernhard Kuster. Proteomics: a pragmatic perspective. *Nature biotechnology*, 28(7):695–709, jul 2010.

[188] M Mann, R C Hendrickson, and A Pandey. Analysis of proteins and proteomes by mass spectrometry. *Annual review of biochemistry*, 70:437–73, jan 2001.

[189] Matthias. Mann, Chin Kai. Meng, and John B. Fenn. Interpreting mass spectra of multiply charged ions. *Analytical Chemistry*, 61(15):1702–1708, aug 1989.

[190] Giorgos Markou and Elias Nerantzis. Microalgae for high-value compounds and biofuels production: A review with focus on cultivation under stress conditions. *Biotechnology Advances*, 31(8):1532–1542, dec 2013.

[191] Guido Mastrobuoni, Susann Irgang, Matthias Pietzke, Heike E Aßmus, Markus Wenzel, Waltraud X Schulze, and Stefan Kempa. Proteome dynamics and early salt stress response of the photosynthetic organism Chlamydomonas reinhardtii, 2012.

[192] Rune Matthiesen. *Mass Spectrometry Data Analysis in Proteomics*, volume 1007. Springer-Verlag GmbH, 2 edition, 2013.

[193] Jude E Maul, Jason W Lilly, Liying Cui, Claude W DePamphilis, Webb Miller, Elizabeth H Harris, and David B Stern. The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. *The Plant cell*, 14(11):2659–79, 2002.

[194] B.K. Maulood, F. M. Hassan, A.A. Al-Lami, J.J. Toma, and A.M. Ismail. Checklist of Algal Flora in Iraq. *Ministry of Environment.*, page 94, 2013.

[195] Patrick May, Stefanie Wienkoop, Stefan Kempa, Björn Usadel, Nils Christian, Jens Rupprecht, Julia Weiss, Luis Recuenco-Munoz, Oliver Ebenhöh, Wolfram Weckwerth, and Dirk Walther. Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of Chlamydomonas reinhardtii. *Genetics*, 179(1):157–166, 2008.

[196] F W McLafferty, D M Horn, K Breuker, Y Ge, M A Lewis, B Cerda, R A Zubarev, and B K Carpenter. Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *Journal of the American Society for Mass Spectrometry*, 12(3):245–9, mar 2001.

[197] R L Meier. Biological cycles in the transformation of solar energy into useful fuels. *Solar Energy Research, F. Daniels and J. A. Duffie, Eds*, 23:179–183, 1955.

[198] Felix Meissner and Matthias Mann. Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. *Nature immunology*, 15(2):112–7, feb 2014.

[199] Anastasios Melis, Liping Zhang, Marc Forestier, Maria L Ghirardi, and Michael Seibert. Sustained Photobiological Hydrogen Gas Production upon Reversible Inactivation of Oxygen Evolution in the Green Alga Chlamydomonas reinhardtii 1. *Plant Physiology*, 122(January):127–135, 2000.

[200] Rui L. Mendes, Helena L. Fernandes, José P. Coelho, Elsa C. Reis, Joaquim M. S. Cabral, Julio M. Novais, and Antonio F. Palavra. Supercritical CO_2 extraction of carotenoids and other lipids from *Chlorella vulgaris*. *Food Chemistry*, 53:99–103, 1995.

[201] Mariângela Menezes, Mata Atlântica, Ankistrodesmus West, and G S West. Chlorophyceae. In Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro. Janeiro, editor, *Catálogo de plantas e fungos do Brasil*, volume 1, pages 335–352. Forzza, R.C. Eds, Rio de Janeiro, 2010.

[202] Gerben Menschaert, Tom T M Vandekerckhove, Geert Baggerman, Liliane Schoofs, Walter Luyten, and Wim Van Criekinge. Peptidomics coming of age: a review of contributions from a bioinformatics angle. *Journal of proteome research*, 9(5):2051–61, 2010.

[203] Sabeeha S Merchant, Janette Kropat, Bensheng Liu, Johnathan Shaw, and Jaruswan Warakanont. TAG, you're it! Chlamydomonas as a reference organism for understanding algal triacylglycerol accumulation. *Current opinion in biotechnology*, 23(3):352–63, jun 2012.

[204] Sabeeha S Merchant, Simon E Prochnik, Olivier Vallon, Elizabeth H Harris, Steven J Karpowicz, George B Witman, Astrid Terry, Asaf Salamov, Lillian K Fritz-Laylin, Laurence Maréchal-Drouard, Wallace F Marshall, Liang-Hu Qu, David R Nelson, Anton A Sanderfoot, Martin H Spalding, Vladimir V Kapitonov, Qinghu Ren, Patrick Ferris, Erika

Lindquist, Harris Shapiro, Susan M Lucas, Jane Grimwood, Jeremy Schmutz, Pierre Cardol, Heriberto Cerutti, Guillaume Chanfreau, Chun-Long Chen, Valérie Cognat, Martin T Croft, Rachel Dent, Susan Dutcher, Emilio Fernández, Hideya Fukuzawa, David González-Ballester, Diego González-Halphen, Armin Hallmann, Marc Hanikenne, Michael Hippler, William Inwood, Kamel Jabbari, Ming Kalanon, Richard Kuras, Paul A Lefebvre, Stéphane D Lemaire, Alexey V Lobanov, Martin Lohr, Andrea Manuell, Iris Meier, Laurens Mets, Maria Mittag, Telsa Mittelmeier, James V Moroney, Jeffrey Moseley, Carolyn Napoli, Aurora M Nedelcu, Krishna Niyogi, Sergey V Novoselov, Ian T Paulsen, Greg Pazour, Saul Purton, Jean-Philippe Ral, Diego Mauricio Riaño-Pachón, Wayne Riekhof, Linda Rymarquis, Michael Schroda, David Stern, James Umen, Robert Willows, Nedra Wilson, Sara Lana Zimmer, Jens Allmer, Janneke Balk, Katerina Bisova, Chong-Jian Chen, Marek Elias, Karla Gendler, Charles Hauser, Mary Rose Lamb, Heidi Ledford, Joanne C Long, Jun Minagawa, M Dudley Page, Junmin Pan, Wirulda Pootakham, Sanja Roje, Annkatrin Rose, Eric Stahlberg, Aimee M Terauchi, Pinfen Yang, Steven Ball, Chris Bowler, Carol L Dieckmann, Vadim N Gladyshev, Pamela Green, Richard Jorgensen, Stephen Mayfield, Bernd Mueller-Roeber, Sathish Rajamani, Richard T Sayre, Peter Brokstein, Inna Dubchak, David Goodstein, Leila Hornick, Y Wayne Huang, Jinal Jhaveri, Yigong Luo, Diego Martínez, Wing Chi Abby Ngau, Bobby Otillar, Alexander Poliakov, Aaron Porter, Lukasz Szajkowski, Gregory Werner, Kemin Zhou, Igor V Grigoriev, Daniel S Rokhsar, and Arthur R Grossman. The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science (New York, N.Y.)*, 318:245–250, 2007.

[205] Xiaoling Miao and Qingyu Wu. High yield bio-oil production from fast pyrolysis by metabolic controlling of Chlorella protothecoides. *Journal of Biotechnology*, 110(1):85–93, 2004.

[206] G Michaelis, C Vahrenholz, and E Pratje. Mitochondrial DNA of Chlamydomonas reinhardtii: the gene for apocytochrome b and the complete functional map of the 15.8 kb DNA. *Molecular & general genetics : MGG*, 223(2):211–216, 1990.

[207] Ján A. Miernyk. Seed Proteomics. In *Methods in molecular biology (Clifton, N.J.)*, volume 1072, pages 361–377. Humana Press, aug 2014.

[208] A Harvey Millar, Joshua L Heazlewood, Brian K Kristensen, Hans-Peter Braun, and Ian M Møller. The plant mitochondrial proteome. *Trends in plant science*, 10(1):36–43, jan 2005.

[209] Rachel Miller, Guangxi Wu, Rahul R Deshpande, Astrid Vieler, Katrin Gärtner, Xiaobo Li, Eric R Moellering, Simone Zäuner, Adam J Cornish, Bensheng Liu, Blair Bullard, Barbara B Sears, Min-Hao Kuo, Eric L Hegg, Yair Shachar-Hill, Shin-Han Shiu, and Christoph Benning. Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant physiology*, 154(4):1737–52, dec 2010.

[210] Jonathan Minden. Comparative proteomics and difference gel electrophoresis. *BioTechniques*, 43(6):739–745, 2007.

[211] Eric R Moellering and Christoph Benning. RNA interference silencing of a major lipid droplet protein affects lipid droplet size in Chlamydomonas reinhardtii. *Eukaryotic cell*, 9:97–106, 2010.

[212] Andrew Moore. Biofuels are dead: long live biofuels(?) – part one. *New Biotechnology*, 25(1):6 – 12, 2008.

[213] Annie Moradian, Anastasia Kalli, Michael J Sweredoski, and Sonja Hess. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics*, 14(4-5):489–97, mar 2014.

[214] James V Moroney and Ruby a Ynalvez. Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. *Eukaryotic cell*, 6(8):1251–9, aug 2007.

[215] Howard R. Morris, Maria Panico, Michael Barber, Robert S. Bordoli, Robert D. Sedgwick, and Andrew Tyler. Fast atom bombardment: A new mass spectrometric method for peptide sequence analysis. *Biochemical and Biophysical Research Communications*, 101(2):623–631, jul 1981.

[216] Joseph Msanne, Di Xu, Anji Reddy Konda, J Armando Casas-Mollano, Tala Awada, Edgar B Cahoon, and Heriberto Cerutti. Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae Chlamydomonas reinhardtii and Coccomyxa sp. C-169. *Phytochemistry*, 75:50–9, mar 2012.

[217] Florence Mus, Alexandra Dubini, Michael Seibert, Matthew C Posewitz, and Arthur R Grossman. Anaerobic acclimation in Chlamydomonas reinhardtii: anoxic gene expression, hydrogenase induction, and metabolic pathways. *The Journal of biological chemistry*, 282:25475–25486, 2007.

[218] Karlie A Neilson, Naveid A Ali, Sridevi Muralidharan, Mehdi Mirzaei, Michael Mariani, Gariné Assadourian, Albert Lee, Steven C van Sluyter, and Paul A Haynes. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics*, 11(4):535–53, feb 2011.

[219] Alexey I Nesvizhskii. Protein identification by tandem mass spectrometry and sequence database searching. *Methods in molecular biology (Clifton, N.J.)*, 367:87–119, 2007.

[220] Alexey I Nesvizhskii. Proteogenomics: concepts, applications and computational strategies. *Nature Methods*, 11(11):1114–1125, 2014.

[221] Alexey I Nesvizhskii and Ruedi Aebersold. Interpretation of shotgun proteomic data: the protein inference problem. *Molecular & cellular proteomics : MCP*, 4(10):1419–40, oct 2005.

[222] Hoa M Nguyen, Mathieu Baudet, Stéphan Cuiné, Jean-Marc Adriano, Damien Barthe, Emmanuelle Billon, Christophe Bruley, Fred Beisson, Gilles Peltier, Myriam Ferro, and Yonghua Li-Beisson. Proteomic profiling of oil bodies isolated from the unicellular green microalga Chlamydomonas reinhardtii: with focus on proteins involved in lipid metabolism. *Proteomics*, 11:4266–73, 2011.

[223] B.P. Nobre, F. Villalobos, B.E. Barragán, A.C. Oliveira, A.P. Batista, P.A.S.S. Marques, R.L. Mendes, H. Sovová, A.F. Palavra, and L. Gouveia. A biorefinery from Nannochloropsis sp. microalga - Extraction of oils and pigments. Production of biohydrogen from the leftover biomass. *Bioresource Technology*, 135:128–136, 2013.

[224] I.V. Novakovskaya and E.N. Patova. Green algae in spruce forests in the north-east of European Russia. *Biologia, Section Botany*, 63(6):832–838, 2008.

[225] P H O'Farrell. High resolution two-dimensional electrophoresis of proteins. *The Journal of biological chemistry*, 250(10):4007–21, may 1975.

[226] Jesper V Olsen, Boris Macek, Oliver Lange, Alexander Makarov, Stevan Horning, and Matthias Mann. Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods*, 4(9):709–12, sep 2007.

[227] S.-E. Ong. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, 1(5):376–386, may 2002.

[228] Sandra Orchard, Henning Hermjakob, and Rolf Apweiler. The Proteomics Standards Initiative. *PROTEOMICS*, 3(7):1374–1376, jul 2003.

[229] W J Oswald and C G Golueke. Biological transformation of solar energy. *Advances in applied microbiology*, 2:223–62, 1960.

[230] Kevin G Owens. *Application of Correlation Analysis Techniques to Mass Spectral Data*, volume 27. Informa UK Limited, mar 1992.

[231] Béla Paizs and Sándor Suhai. Fragmentation pathways of protonated peptides. *Mass spectrometry reviews*, 24(4):508–48, jan 2005.

[232] Bernhard Palsson. Adaptive Laboratory Evolution. *Microbe Magazine*, 6(2):69–74, feb 2011.

[233] Alexandre Panchaud, Alexander Scherl, Scott A Shaffer, Priska D Von Haller, Hemantha D Kulasekara, Samuel I Miller, and David R Goodlett. PAcIFIC : how to dive deeper into the proteomics ocean. *Analytical Chemistry*, 81(15):6481–6488, aug 2009.

[234] D.J.C. Pappin, P. Hojrup, and A.J. Bleasby. Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, 3(6):327–332, jun 1993.

[235] Gregory J Pazour, Nathan Agrin, John Leszyk, and George B Witman. Proteomic analysis of a eukaryotic cilium. *The Journal of cell biology*, 170(1):103–13, jul 2005.

[236] Yasset Perez-Riverol, Emanuele Alpi, Rui Wang, Henning Hermjakob, and Juan Antonio Vizcaíno. Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics*, 15(5-6):930–49, mar 2015.

[237] C Periappuram, L Steinhauer, D L Barton, D C Taylor, B Chatson, and J Zou. The plastidic phosphoglucomutase from Arabidopsis. A reversible enzyme reaction with an important role in metabolic control. *Plant physiology*, 122(4):1193–9, apr 2000.

[238] D N Perkins, Darryl J C Pappin, David M Creasy, and John S Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data Proteomics and 2-DE. *Electrophoresis*, 20(18):3551–67, 1999.

[239] Marie-Mathilde Perrineau, Jeferson Gross, Ehud Zelzion, Dana C Price, Orly Levitan, Jeffrey Boyd, and Debashish Bhattacharya. Using Natural Selection to Explore the Adaptive Potential of Chlamydomonas reinhardtii. *Plos One*, 9(3):e92533, mar 2014.

[240] Amelia C Peterson, Jason D Russell, Derek J Bailey, Michael S Westphall, and Joshua J Coon. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics : MCP*, 11(11):1475–88, 2012.

[241] Brian P Piasecki, Kenneth R Diller, and Jerry J Brand. Cryopreservation of Chlamydomonas reinhardtii: a cause of low viability at high cell density. *Cryobiology*, 58(1):103–109, feb 2009.

[242] Danae Pitre, Amiel Boullemant, and Claude Fortin. Uptake and sorption of aluminium and fluoride by four green algal species. *Chemistry Central Journal*, 8:8, feb 2014.

[243] Tessa Pocock, Marc André Lachance, Thomas Pröschold, John C. Priscu, Sam Sulgi Kim, and Norman P A Huner. Identification of a psychrophilic green alga from Lake Bonney Antarctica: Chlamydomonas raudensis ETTL. (UWO 241) Chlorophyceae. *Journal of Phycology*, 40(6):1138–1148, 2004.

[244] Grace Pokoo-Aikins, Ahmed Nadim, Mahmoud M. El-Halwagi, and Vladimir Mahalec. Design and analysis of biodiesel production from algae grown through carbon sequestration. *Clean Technologies and Environmental Policy*, 12(3):239–254, jun 2010.

[245] Thomas Pröschold, Elizabeth H Harris, and Annette W Coleman. Portrait of a species: Chlamydomonas reinhardtii. *Genetics*, 170(4):1601–10, aug 2005.

[246] Thierry Rabilloud and Cécile Lelong. Two-dimensional gel electrophoresis in proteomics: A tutorial. *Journal of Proteomics*, 74(10):1829–1841, 2011.

[247] Arthur Ramazanov and Zakir Ramazanov. Isolation and characterization of a starchless mutant of Chlorella pyrenoidosa STL-PI with a high growth rate, and high protein and polyunsaturated fatty acid content. *Phycological Research*, 54(4):255–259, dec 2006.

[248] B L Randolph-Anderson, J E Boynton, N W Gillham, E H Harris, A M Johnson, M P Dorthu, and R F Matagne. Further characterization of the respiratory deficient dum-1 mutation of Chlamydomonas reinhardtii and its use as a recipient for mitochondrial transformation. *Molecular & general genetics : MGG*, 236(2-3):235–44, jan 1993.

[249] Naim Rashid, Muhammad Saif Ur Rehman, Madeha Sadiq, Tariq Mahmood, and Jong-In Han. Current status, issues and developments in microalgae derived biodiesel production. *Renewable and Sustainable Energy Reviews*, 40:760–778, dec 2014.

[250] John A Raven, John Beardall, and Mario Giordano. Energy costs of carbon dioxide concentrating mechanisms in aquatic organisms. *Photosynthesis research*, 121(2-3):111–24, sep 2014.

[251] Mamta Rawat, MargaretC. Henk, LaraL. Lavigne, and JamesV. Moroney. Chlamydomonas reinhardtii mutants without ribulose-1,5-bisphosphate carboxylase-oxygenase lack a detectable pyrenoid. *Planta*, 198(2):263–270, feb 1996.

[252] Shaikh A. Razzak, Mohammad M. Hossain, Rahima A. Lucky, Amarjeet S. Bassi, and Hugo de Lasa. Integrated CO_2 capture, wastewater treatment and biofuel production by microalgae culturing - A review. *Renewable and Sustainable Energy Reviews*, 27:622–653, nov 2013.

[253] Luis Recuenco-Muñoz, Pierre Offre, Luis Valledor, David Lyon, Wolfram Weckwerth, and Stefanie Wienkoop. Targeted quantitative analysis of a diurnal RuBisCO subunit expression and translation profile in Chlamydomonas reinhardtii introducing a novel Mass Western approach. *Journal of Proteomics*, 113:143–153, 2015.

[254] T. B K Reddy, Alex D. Thomas, Dimitri Stamatis, Jon Bertsch, Michelle Isbandi, Jakob Jansson, Jyothi Mallajosyula, Ioanna Pagani, Elizabeth A. Lobos, and Nikos C. Kyrpides. The Genomes OnLine Database (GOLD) v.5: A metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Research*, 43(D1):D1099–D1106, 2015.

[255] J D Rochaix. Chlamydomonas reinhardtii as the photosynthetic yeast. *Annual review of genetics*, 29:209–230, 1995.

[256] P Roepstorff and J Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry*, 11(11):601, nov 1984.

[257] Sergio Rosales-Mendoza. Future directions for the development of Chlamydomonas-based vaccines. *Expert Review of Vaccines*, 12(9):1011–1019, sep 2013.

[258] Philip L Ross, Yulin N Huang, Jason N Marchese, Brian Williamson, Kenneth Parker, Stephen Hattan, Nikita Khainovski, Sasi Pillai, Subhakar Dey, Scott Daniels, Subhasish Purkayastha, Peter Juhasz, Stephen Martin, Michael Bartlet-Jones, Feng He, Allan Jacobson, and Darryl J Pappin. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics : MCP*, 3(12):1154–69, dec 2004.

[259] Jens Rupprecht. From systems biology to fuel - Chlamydomonas reinhardtii as a model for a systems biology approach to improve biohydrogen production. *Journal of Biotechnology*, 142(1):10–20, 2009.

[260] R Sager and M R Ishida. Chloroplast DNA in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, 50(4):725–30, oct 1963.

[261] R. Sager and Z. Ramanis. Chloroplast genetics of chlamydomonas. I. Allelic segregation ratios. *Genetics*, 83(2):303–321, 1976.

[262] G. A. Scheele. Two-Dimensional Gel Analysis of Soluble Proteins. *The Journal of Biological Chemistry*, 250(14):5375–5385, 1975.

[263] Peer M. Schenk, Skye R. Thomas-Hall, Evan Stephens, Ute C. Marx, Jan H. Mussgnug, Clemens Posten, Olaf Kruse, and Ben Hankamer. Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. *BioEnergy Research*, 1(1):20–43, mar 2008.

[264] Melanie Schmidt, Gunther Gessner, Matthias Luff, Ines Heiland, Volker Wagner, Marc Kaminski, Stefan Geimer, Nicole Eitzinger, Tobias Reissenweber, Olga Voytsekh, Monika Fiedler, Maria Mittag, and Georg Kreimer. Proteomic analysis of the eyespot of Chlamydomonas reinhardtii provides novel insights into its components and tactic movements. *The Plant cell*, 18(8):1908–30, aug 2006.

[265] Stefan Schmollinger, Timo Muehlhaus, Nanette R Boyle, Ian K Blaby, David Casero, Tabea Mettler, Jeffrey L Moseley, Janette Kropat, Frederik Sommer, Daniela Strenkert, Dorothea Hemme, Matteo Pellegrini, Arthur R Grossman, Mark Stitt, Michael Schroda, and Sabeeha S Merchant. Nitrogen-Sparing Mechanisms in Chlamydomonas Affect the Transcriptome, the Proteome, and Photosynthetic Metabolism. *Plant Cell*, 26(4):1410–1435, apr 2014.

[266] Michel Schneider, The Uniprot Consortium, and Sylvain Poux. UniProtKB amid the turmoil of plant proteomics research. *Frontiers in Pant Science*, 3(December):270, 2012.

[267] Thomas Schulze, Katja Prager, Hannes Dathe, Juliane Kelm, Peter Kiessling, and Maria Mittag. How the green alga Chlamydomonas reinhardtii keeps time. *Protoplasma*, 244:3–14, 2010.

[268] Waltraud X Schulze and Björn Usadel. Quantitation in mass-spectrometry-based proteomics. *Annual review of plant biology*, 61:491–516, 2010.

[269] JC Schwartz, JEP Syka, and ST Quarmby. Improving the fundamentals of MSn on 2D linear ion traps: new ion activation and isolation techniques. In *The 53rd ASMS Conference on Mass Spectrometry and Allied Topic*, St. Antonio, TX, USA, 2005.

[270] Melissa A Scranton, Joseph T Ostrand, Francis J Fields, and Stephen P Mayfield. Chlamydomonas as a model for biofuels and bio-products production. *The Plant journal : for cell and molecular biology*, 82(3):523–31, may 2015.

[271] Joerg Seidler, Nico Zinn, Martin E Boehm, and Wolf D Lehmann. De novo sequencing of peptides by MS/MS. *Proteomics*, 10(4):634–49, feb 2010.

[272] Y Shen, W Yuan, Z J Pei, Q Wu, and E Mao. Microalgae Mass Production Methods. *Transactions of the ASABE - American Society of Agricultural Engineers.*, 52(4):1275–1287, 2009.

[273] Magali Siaut, Stéphan Cuiné, Caroline Cagnon, Boris Fessler, Mai Nguyen, Patrick Carrier, Audrey Beyly, Fred Beisson, Christian Triantaphylidès, Yonghua Li-Beisson, and Gilles Peltier. Oil accumulation in the model green alga Chlamydomonas reinhardtii: characterization, variability between common laboratory strains and relationship with starch reserves. *BMC biotechnology*, 11:7, 2011.

[274] Simone Sidoli, Shu Lin, Kelly R Karch, and Benjamin A Garcia. Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative abundance and stoichiometry. *Analytical chemistry*, 87(6):3129–33, mar 2015.

[275] Deborah M Simpson and Robert J Beynon. QconCATs: design and expression of concatenated protein standards for multiplexed protein quantification. *Analytical and bioanalytical chemistry*, 404(4):977–89, sep 2012.

[276] Kari Skjånes, Céline Rebours, and Peter Lindblad. Potential for green microalgae to produce hydrogen, pharmaceuticals and other high value products in a combined process. *Critical Reviews in Biotechnology*, 33(2):172–215, jun 2013.

[277] Raphael Slade and Ausilio Bauen. Micro-algae cultivation for biofuels: Cost, energy balance, environmental impacts and future prospects. *Biomass and Bioenergy*, 53(0):29–38, 2013.

[278] Lekha Sleno and Dietrich A. Volmer. Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry*, 39(10):1091–1112, 2004.

[279] Damian Smedley, Syed Haider, Steffen Durinck, Luca Pandini, Paolo Provero, James Allen, Olivier Arnaiz, Mohammad Hamza Awedh, Richard Baldock, Giulia Barbiera, Philippe Bardou, Tim Beck, Andrew Blake, Merideth Bonierbale, Anthony J Brookes, Gabriele Bucci, Iwan Buetti, Sarah Burge, Cédric Cabau, Joseph W Carlson, Claude Chelala, Charalambos Chrysostomou, Davide Cittaro, Olivier Collin, Raul Cordova, Rosalind J Cutts, Erik Dassi, Alex Di Genova, Anis Djari, Anthony Esposito, Heather Estrella, Eduardo Eyras, Julio Fernandez-Banet, Simon Forbes, Robert C Free, Takatomo Fujisawa, Emanuela Gadaleta, Jose M Garcia-Manteiga, David Goodstein, Kristian Gray, José Afonso Guerra-Assunção, Bernard Haggarty, Dong-Jin Han, Byung Woo Han, Todd Harris, Jayson Harshbarger, Robert K Hastings, Richard D Hayes, Claire Hoede, Shen Hu, Zhi-Liang Hu, Lucie Hutchins, Zhengyan Kan, Hideya Kawaji, Aminah Keliet, Arnaud Kerhornou, Sunghoon Kim, Rhoda Kinsella, Christophe Klopp, Lei Kong, Daniel Lawson, Dejan Lazarevic, Ji-Hyun Lee, Thomas Letellier, Chuan-Yun Li, Pietro Lio, Chu-Jun Liu, Jie Luo, Alejandro Maass, Jerome Mariette, Thomas Maurel, Stefania Merella, Azza Mostafa Mohamed, Francois Moreews, Ibounyamine Nabihoudine, Nelson Ndegwa, Céline Noirot, Cristian Perez-Llamas, Michael Primig, Alessandro Quattrone, Hadi Quesneville, Davide Rambaldi, James Reecy, Michela Riba, Steven Rosanoff, Amna Ali Saddiq, Elisa Salas, Olivier Sallou, Rebecca Shepherd, Reinhard Simon, Linda Sperling, William Spooner, Daniel M Staines, Delphine Steinbach, Kevin Stone, Elia Stupka, Jon W Teague, Abu Z Dayem Ullah, Jun Wang, Doreen Ware, Marie Wong-Erasmus, Ken Youens-Clark, Amonida Zadissa, Shi-Jian Zhang, and Arek Kasprzyk. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic acids research*, 43(W1):W589–98, 2015.

[280] Bryan E Smith, James A Hill, Mark A Gjukich, and Philip C Andrews. Tranche distributed repository and ProteomeCommons.org. *Methods in molecular biology (Clifton, N.J.)*, 696:123–45, jan 2011.

[281] Lloyd M Smith and Neil L Kelleher. Proteoform: a single term describing protein complexity. *Nature methods*, 10(3):186–7, mar 2013.

[282] R D Smith, J A Loo, C G Edmonds, C J Barinaga, and H R Udseth. Sensitivity considerations for large molecule detection by capillary electrophoresis-electrospray ionization mass spectrometry. *Journal of chromatography*, 516(1):157–65, sep 1990.

[283] Martin H Spalding. Microalgal carbon-dioxide-concentrating mechanisms: Chlamydomonas inorganic carbon transporters. *J Exp Bot*, 59(7):1463–1473, 2008.

[284] O Stemmann, H Zou, S A Gerber, S P Gygi, M W Kirschner, T.J. Yen, G.J. Gorbsky, M.J. Weber, C. Lengauer, C. Ginther, and et Al. Dual inhibition of sister chromatid separation at metaphase. *Cell*, 107(6):715–26, dec 2001.

[285] Wendy A. Stirk, Vince Ördög, Ondřej Novák, Jakub Rolčík, Miroslav Strnad, Péter Bálint, and Johannes van Staden. Auxin and cytokinin relationships in 24 microalgal strains. *Journal of Phycology*, 49(3):459–467, jun 2013.

[286] Emma Suali and Rosalam Sarbatly. Conversion of microalgae to biofuel. *Renewable and Sustainable Energy Reviews*, 16(6):4316–4342, aug 2012.

[287] Bobban G. Subhadra. Sustainability of algal biofuel production using integrated renewable energy park (IREP) and algal biorefinery approach. *Energy Policy*, 38(10):5892–5901, 2010.

[288] Danielle L Swaney, Graeme C McAlister, and Joshua J Coon. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature methods*, 5(11):959–64, nov 2008.

[289] Hanifa Taher, Sulaiman Al-Zuhair, Ali H. Al-Marzouqi, Yousef Haik, and Mohammed Farid. Effective extraction of microalgae lipids from wet biomass for biodiesel production. *Biomass and Bioenergy*, 66:159–167, jul 2014.

[290] Koichi Tanaka, Hiroaki Waki, Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida, Tamio Yoshida, and T Matsuo. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2(8):151–153, 1988.

[291] Lothar Täuscher. Algen (Cyanobacteria et Phycophyta) - Stand: 2013. In D Frank and P Schnitter, editors, *Pflanzen und Tiere in Sachsen-Anhalt Ein Kompendium der Biodiversität*, pages 63–112. Westermann Druck Zwickau, Rangsdorf, 2016.

[292] Chris F Taylor, Norman W Paton, Kathryn S Lilley, Pierre-Alain Binz, Randall K Julian, Andrew R Jones, Weimin Zhu, Rolf Apweiler, Ruedi Aebersold, Eric W Deutsch, Michael J Dunn, Albert J R Heck, Alexander Leitner, Marcus Macht, Matthias Mann, Lennart Martens, Thomas A Neubert, Scott D Patterson, Peipei Ping, Sean L Seymour, Puneet Souda, Akira Tsugita, Joel Vandekerckhove, Thomas M Vondriska, Julian P Whitelegge, Marc R Wilkins, Ioannnis Xenarios, John R Yates, and Henning Hermjakob. The minimum information about a proteomics experiment (MIAPE). *Nature Biotechnology*, 25(8):887–893, aug 2007.

[293] G. Tell. *Catálogo de las algas de agua dulce de la República Argentina*, volume 7. Vaduz: Cramer, bibliotheca phycologica edition, 1985.

[294] Mia Terashima, Michael Specht, and Michael Hippler. The chloroplast proteome: a survey from the Chlamydomonas reinhardtii perspective with a focus on distinctive features. *Current genetics*, 57:151–168, 2011.

[295] Mia Terashima, Michael Specht, and Michael Hippler. The chloroplast proteome: a survey from the Chlamydomonas reinhardtii perspective with a focus on distinctive features. *Current genetics*, 57(3):151–68, jun 2011.

[296] Tobias Ternent, Attila Csordas, Da Qi, Guadalupe G??mez-Baena, Robert J. Beynon, Andrew R. Jones, Henning Hermjakob, and Juan Antonio Vizca??no. How to submit MS proteomics data to ProteomeXchange via the PRIDE database. *Proteomics*, 14(20):2233–2241, 2014.

[297] Jesse B Therien, Oleg A Zadvornyy, Matthew C Posewitz, Donald A Bryant, and John W Peters. Growth of Chlamydomonas reinhardtii in acetate-free medium when co-cultured with alginate-encapsulated, acetate-producing strains of Synechococcus sp. PCC 7002. *Biotechnology for biofuels*, 7(1):154, 2014.

[298] Andrew Thompson, Jürgen Schäfer, Karsten Kuhn, Stefan Kienle, Josef Schwarz, Günter Schmidt, Thomas Neumann, R Johnstone, A Karim A Mohammed, and Christian Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry*, 75(8):1895–904, apr 2003.

[299] Alexander N Tikhonov. The cytochrome b6f complex at the crossroad of photosynthetic electron transport pathways. *Plant physiology and biochemistry : PPB / Société française de physiologie végétale*, 81:163–83, aug 2014.

[300] Agus Eko Tjahjono, Yachiyo Hayama, Toshihide Kakizono, Yoshio Terada, Naomichi Nishio, and Shiro Nagai. Hyper-Accumulation of Astaxanthin in a Green Alga Haematococcus pluvialis at elevated temperatures. *Biotechnology Letters*, 16(2):133–138, 1994.

[301] Jörg Toepel, Stefan P Albaum, Samuel Arvidsson, Alexander Goesmann, Marco la Russa, Kristin Rogge, and Olaf Kruse. Construction and evaluation of a whole genome microarray of Chlamydomonas reinhardtii, 2011.

[302] Miller Tran, Christina Van, Daniel J Barrera, Pär L Pettersson, Carlos D Peinado, Jack Bui, and Stephen P Mayfield. Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1):E15–22, jan 2013.

[303] Chia-Hong Tsai, Jaruswan Warakanont, Tomomi Takeuchi, Barb B Sears, Eric R Moellering, and Christoph Benning. The protein Compromised Hydrolysis of Triacylglycerols 7 (CHT7) acts as a repressor of cellular quiescence in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44):15833–8, nov 2014.

[304] Chih-Chiang Tsou, Dmitry Avtonomov, Brett Larsen, Monika Tucholska, Hyungwon Choi, Anne-Claude Gingras, and Alexey I Nesvizhskii. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods*, 12(3):258–264, 2015.

[305] Takashi Tsuchiya, Tsukasa Osaki, Naoto Minamino, and Kazuki Sasaki. Peptidomics for studying limited proteolysis. *Journal of Proteome Research*, 14(11):4921–4931, 2015.

[306] Christoph W. Turck, Arnold M. Falick, Jeffrey A. Kowalak, William S. Lane, Kathryn S. Lilley, Brett S. Phinney, Susan T. Weintraub, H. Ewa Witkowska, and Nathan A. Yates. The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 Study: Relative Protein Quantitation. *Mol. Cell. Proteomics*, 6(8):1291–1298, 2007.

[307] James G. Umen. Evolution of sex and mating loci: An expanded view from Volvocine algae. *Current opinion in microbiology*, 14(6):634, 2011.

[308] C Vahrenholz, G Riemen, E Pratje, B Dujon, and G Michaelis. Mitochondrial DNA of Chlamydomonas reinhardtii: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. *Current genetics*, 24(3):241–7, sep 1993.

[309] Luis Valledor, Takeshi Furuhashi, Luis Recuenco-Muñoz, Stefanie Wienkoop, and Wolfram Weckwerth. System-level network analysis of nitrogen starvation and recovery in Chlamydomonas reinhardtii reveals potential new targets for increased lipid accumulation. *Biotechnology for biofuels*, 7:171, jan 2014.

[310] Luis Valledor, Luis Recuenco-Munoz, Stefanie Wienkoop, and Wolfram Weckwerth. The different proteomes of Chlamydomonas reinhardtii. *Journal of Proteomics*, 2012.

[311] Olivier Vallon and Susan Dutcher. Treasure hunting in the Chlamydomonas genome. *Genetics*, 179(1):3–6, 2008.

[312] Robert van Lis, Ariane Atteia, Guillermo Mendoza-Hernández, and Diego González-Halphen. Identification of novel mitochondrial protein components of Chlamydomonas reinhardtii. A proteomic approach. *Plant physiology*, 132(1):318–30, may 2003.

[313] Marieke Vanthoor-Koopmans, Rene H. Wijffels, Maria J. Barbosa, and Michel H.M. Eppink. Biorefinery of microalgae for food and fuel. *Bioresource Technology*, 135:142–149, 2013.

[314] T. Veen, A., Hof, C.H.J., Kouwets, F.A.C. & Berkhout. Rijkswaterstaat Waterdienst, Informatiehuis Water [Taxa Watermanagement the Netherlands (TWN). Technical report, The Netherlands: Laboratory for Hydrobiological Analysis, Rijkswaterstaat (Consulted March 2017), 2015.

[315] Natarajan Velmurugan, Minji Sung, Sung Yim, Min S Park, Ji Yang, and Ki Jeong. Systematically programmed adaptive evolution reveals potential role of carbon and nitrogen pathways during lipid accumulation in Chlamydomonas reinhardtii. *Biotechnology for Biofuels*, 7(1):117, 2014.

[316] John D Venable, Meng-Qiu Dong, James Wohlschlegel, Andrew Dillin, and John R Yates. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods*, 1(1):39–45, 2004.

[317] Nathan C. VerBerkmoes, Vincent J. Denef, Robert L. Hettich, and Jillian F. Banfield. Systems Biology: Functional analysis of natural microbial consortia using community proteomics. *Nature Reviews Microbiology*, 7(3):196–205, mar 2009.

[318] M M Vestling, M A Kelly, and C Fenselau. Optimization by mass spectrometry of a tryptophan-specific protein cleavage reaction. *Rapid communications in mass spectrometry : RCM*, 8(9):786–90, oct 1994.

[319] Helene Vigeolas, Torsten Möhlmann, Norbert Martini, H Ekkehard Neuhaus, and Peter Geigenberger. Embryo-specific reduction of ADP-Glc pyrophosphorylase leads to an inhibition of starch synthesis and a delay in oil accumulation in developing seeds of oilseed rape. *Plant physiology*, 136(1):2676–86, sep 2004.

[320] Milada Vítová, Kateřina Bišová, Dáša Umysová, Monika Hlavová, Shigeyuki Kawano, Vilém Zachleder, and Mária Cížková. Chlamydomonas reinhardtii: duration of its cell cycle and phases at growth rates affected by light intensity. *Planta*, 233(1):75–86, jan 2011.

[321] Juan Antonio Vizcaíno, Joseph M Foster, and Lennart Martens. Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research. *Journal of proteomics*, 73(11):2136–46, oct 2010.

[322] Björn Voss, Linda Meinecke, Thorsten Kurz, Salim Al-Babili, Christoph F Beck, and Wolfgang R Hess. Hemin and magnesium-protoporphyrin IX induce global changes in gene expression in Chlamydomonas reinhardtii. *Plant physiology*, 155(2):892–905, feb 2011.

[323] Robert D. Voyksner and Heewon Lee. Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry*, 13(14):1427–1437, jul 1999.

[324] Jian Wang, Monika Tucholska, James D R Knight, Jean-philippe Lambert, Stephen Tate, Brett Larsen, Anne-claude Gingras, and Nuno Bandeira. MSPLIT-DIA : sensitive peptide identification for data-independent acquisition. *Nature Publishing Group*, 12(12):1106–1108, 2015.

[325] Yingjun Wang and Martin H Spalding. LCIB in the Chlamydomonas CO_2-concentrating mechanism. *Photosynthesis research*, 121(2-3):185–92, sep 2014.

[326] Zi Teng Wang, Nico Ullrich, Sunjoo Joo, Sabine Waffenschmidt, and Ursula Goodenough. Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless Chlamydomonas reinhardtii. *Eukaryotic cell*, 8(12):1856–68, dec 2009.

[327] Zi Teng Wang, Nico Ullrich, Sunjoo Joo, Sabine Waffenschmidt, and Ursula Goodenough. Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless Chlamydomonas reinhardtii. *Eukaryotic cell*, 8(12):1856–68, dec 2009.

[328] Nishikant Wase, Paul N Black, Bruce A Stanley, and Concetta C DiRusso. Integrated Quantitative Analysis of Nitrogen Stress Response in Chlamydomonas reinhardtii Using Metabolite and Protein Profiling. *Journal of Proteome Research*, 13(3):1373–1396, mar 2014.

[329] M P Washburn, D Wolters, and J R Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology*, 19(3):242–7, mar 2001.

[330] Wolfram Weckwerth, Marcelo Ehlers Loureiro, Kathrin Wenzel, and Oliver Fiehn. Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci U S A*, 101(20):7809–7814, 2004.

[331] Wolfram Weckwerth, Kathrin Wenzel, and Oliver Fiehn. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*, 4:78–83, 2004.

[332] Chad R Weisbrod, Jimmy K Eng, Michael R Hoopmann, Tahmina Baker, and James E Bruce. Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. *Journal of proteome research*, 11(3):1621–32, mar 2012.

[333] Haworth E Y Whitton B A, John D M, Kelly M G. Freshwater algae in Britain and Ireland. Technical report, Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, Oxon, OX10 8BB, 2014.

[334] Stefanie Wienkoop, Mirko Glinski, Nobuo Tanaka, Vladimir Tolstikov, Oliver Fiehn, and Wolfram Weckwerth. Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. *Rapid Commun Mass Spectrom*, 18(6):643–650, 2004.

[335] Stefanie Wienkoop, Estibaliz Larrainzar, Michaela Niemann, Esther M Gonzalez, Ute Lehmann, and Wolfram Weckwerth. Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics. *J Sep Sci*, 29(18):2793–2801, 2006.

[336] Stefanie Wienkoop, Christiana Staudinger, Wolfgang Hoehenwarter, Wolfram Weckwerth, and Volker Egelhofer. ProMEX – a mass spectral reference database for plant proteomics. *Frontiers in Plant Science*, 3, 2012.

[337] Stefanie Wienkoop, Julia Weiss, Patrick May, Stefan Kempa, Susann Irgang, Luis Recuenco-Munoz, Matthias Pietzke, Thorsten Schwemmer, Jens Rupprecht, Volker Egelhofer, and Wolfram Weckwerth. Targeted proteomics for Chlamydomonas reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. *Molecular bioSystems*, 6(6):1018–31, jun 2010.

[338] Stefanie Wienkoop, Julia Weiss, Patrick May, Stefan Kempa, Susann Irgang, Luis Recuenco-Munoz, Matthias Pietzke, Thorsten Schwemmer, Jens Rupprecht, Volker Egelhofer, and Wolfram Weckwerth. Targeted proteomics for Chlamydomonas reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. *Molecular bioSystems*, 6(6):1018–31, jun 2010.

[339] Patrick E. Wiley, J. Elliott Campbell, and Brandi McKuin. Production of Biodiesel and Biogas from Algae: A Review of Process Train Options. *Water Environment Research*, 83(4):366–338, 2011.

[340] M R Wilkins, K Ou, R D Appel, J C Sanchez, J X Yan, O Golaz, V Farnsworth, P Cartier, D F Hochstrasser, K L Williams, and a a Gooley. Rapid protein identification using N-terminal "sequence tag" and amino acid analysis. *Biochemical and biophysical research communications*, 221(3):609–13, apr 1996.

[341] M R Wilkins, C Pasquali, R D Appel, K Ou, O Golaz, J C Sanchez, J X Yan, A A Gooley, G Hughes, I Humphery-Smith, K L Williams, and D F Hochstrasser. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology (Nature Publishing Company)*, 14(1):61–5, jan 1996.

[342] Flavia V Winck, Diego M Riaño-Pachón, Frederik Sommer, Jens Rupprecht, and Bernd Mueller-Roeber. The nuclear proteome of the green alga Chlamydomonas reinhardtii. *Proteomics*, 12:95–100, 2012.

[343] Eric Witze. Mapping Protein Post-Translational Modifcications With Mass Spectrometry. *Nature methods*, 4(10):798–806, 2007.

[344] Dirk A. Wolters, Michael P. Washburn, and John R. Yates. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Analytical Chemistry*, 73(23):5683–5690, dec 2001.

[345] Victoria H Work, Randor Radakovits, Robert E Jinkerson, Jonathan E Meuser, Lee G Elliott, David J Vinyard, Lieve M L Laurens, G Charles Dismukes, and Matthew C Posewitz. Increased lipid accumulation in the Chlamydomonas reinhardtii sta7-10 starchless isoamylase mutant and increased carbohydrate synthesis in complemented strains. *Eukaryotic cell*, 9(8):1251–61, aug 2010.

[346] Jennifer R Wortman, Brian J Haas, Linda I Hannick, Roger K Smith, Rama Maiti, Catherine M Ronning, Agnes P Chan, Chunhui Yu, Mulu Ayele, Catherine A Whitelaw, Owen R White, and Christopher D Town. Annotation of the Arabidopsis genome. *Plant physiology*, 132(2):461–8, jun 2003.

[347] Chaochao Wu, Tujin Shi, Joseph N Brown, Jintang He, Yuqian Gao, Thomas L Fillmore, Anil K Shukla, Ronald J Moore, David G Camp, Karin D Rodland, Wei-Jun Qian, Tao Liu, and Richard D Smith. Expediting SRM assay development for large-scale targeted proteomics experiments. *Journal of proteome research*, 13(10):4479–87, oct 2014.

[348] Christine C Wu and Michael J MacCoss. Shotgun proteomics: tools for the analysis of complex biological systems. *Current opinion in molecular therapeutics*, 4(3):242–50, jun 2002.

[349] Cong Wu, John C Tran, Leonid Zamdborg, Kenneth R Durbin, Mingxi Li, Dorothy R Ahlf, Bryan P Early, Paul M Thomas, Jonathan V Sweedler, and Neil L Kelleher. A protease for 'middle-down' proteomics. *Nature methods*, 9(8):822–4, aug 2012.

[350] Yin-Hu Wu, Hong-Ying Hu, Yin Yu, Tian-Yuan Zhang, Shu-Feng Zhu, Lin-Lan Zhuang, Xue Zhang, and Yun Lu. Microalgal species for sustainable biomass/lipid production using wastewater as resource: A review. *Renewable & Sustainable Energy Reviews*, 33:675–688, may 2014.

[351] Vicki H Wysocki, Katheryn A Resing, Qingfen Zhang, and Guilong Cheng. Mass spectrometry of peptides and proteins. *Methods (San Diego, Calif.)*, 35(3):211–22, mar 2005.

[352] Wenqiang Yang, Claudia Catalanotti, Matthew C Posewitz, Jean Alric, and Arthur R Grossman. Insights into Algal Fermentation. In J T VanDongen and F Licausi, editors, *Low-Oxygen Stress in Plants: Oxygen Sensing and Adaptive Responses to Hypoxia*, volume 21, pages 135–163. Springer Vienna, 2014.

[353] J R Yates, S F Morgan, C L Gatlin, P R Griffin, and J K Eng. Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Analytical chemistry*, 70(17):3557–3565, sep 1998.

[354] John R Yates. SPECIAL FEATURE : Mass Spectrometry and the Age of the Proteome. *Journal of Mass Spectrometry*, 33(November 1997):1–19, 1998.

[355] John R Yates. The revolution and evolution of shotgun proteomics for large-scale proteome analysis. *Journal of the American Chemical Society*, 135(5):1629–40, mar 2013.

[356] Shuiyan Yu, Quanyu Zhao, Xiaoling Miao, and Jiping Shi. Enhancement of lipid production in low-starch mutants Chlamydomonas reinhardtii by adaptive laboratory evolution. *Bioresource Technology*, 147:499–507, nov 2013.

[357] Zhongqi Zhang. Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal Chem*, 76:3908–3922, 2004.

[358] Han-Qin Zheng, Yi-Fan Chiang-Hsieh, Chia-Hung Chien, Bo-Kai Justin Hsu, Tsung-Lin Liu, Ching-Nen Nathan Chen, and Wen-Chi Chang. AlgaePath: comprehensive analysis of metabolic pathways using transcript abundance data from next-generation sequencing in green algae. *BMC genomics*, 15(1):196, mar 2014.

[359] L.D. Zhu, E. Hiltunen, E. Antila, J.J. Zhong, Z.H. Yuan, and Z.M. Wang. Microalgal biofuels: Flexible bioenergies for sustainable development. *Renewable and Sustainable Energy Reviews*, 30:1035–1046, feb 2014.

[360] Wenhong Zhu, Jeffrey W Smith, and Chun-Ming Huang. Mass spectrometry-based label-free quantitative proteomics. *Journal of biomedicine & biotechnology*, 2010:840518, jan 2010.

[361] Roman A. Zubarev, Neil L. Kelleher, and Fred W. McLafferty. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society*, 120(13):3265–3266, apr 1998.

5 Affirmation

I hereby declare to have written this Diploma Thesis on my own, having used only the listed resources and tools.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass die Arbeit selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken, gegebenenfalls auch elektronischen Medien, entnommen sind, durch Angabe der Quelle als Entlehnung kenntlich gemacht wurden.