Herbicide effects on Chlamydomonas reinhardtii assessed on physiological and molecular levels

Author(s):
Nestler, Holger

Publication Date:
2012

Permanent Link:
https://doi.org/10.3929/ethz-a-007578753

Rights / License:
In Copyright - Non-Commercial Use Permitted
Herbicide effects on *Chlamydomonas reinhardtii* assessed on physiological and molecular levels

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

HOLGER NESTLER

Dipl.-Geoökol., Technische Universität Bergakademie Freiberg

born on August 11, 1981 in Zschopau

citizen of Germany

accepted on the recommendation of

Prof. Dr. Rik I. L. Eggen, examiner
Dr. Marc J.-F. Suter, co-examiner
Prof. Dr. Jukka Jokela, co-examiner
Dr. Mechthild Schmitt-Jansen, co-examiner

2012
# Table of contents

List of figures .......................................................................................................................................................... III
List of tables ........................................................................................................................................................... IV
Summary ................................................................................................................................................................. V
Zusammenfassung ..................................................................................................................................................... VII

Chapter 1: General introduction ............................................................................................................................ 1
  1.1 Pesticides and their effects on green algae in freshwater systems................................................................. 2
  1.2 Approaches in the ecotoxicology of green algae .............................................................................................. 3
  1.3 Scope of the thesis ............................................................................................................................................ 7
  1.4 Outline of the thesis ........................................................................................................................................ 13
  1.5 References ..................................................................................................................................................... 14

Chapter 2: Multiple-endpoint assay provides a detailed mechanistic view of responses to herbicide exposure in *Chlamydomonas reinhardtii* ......................................................................................... 19
  2.1 Abstract ......................................................................................................................................................... 20
  2.2 Introduction .................................................................................................................................................. 21
  2.3 Materials and methods ................................................................................................................................. 24
  2.4 Results and discussion ................................................................................................................................. 28
  2.5 Concluding remarks ..................................................................................................................................... 39
  2.6 References ..................................................................................................................................................... 40

Supplementary material to chapter 2 ..................................................................................................................... 43

Chapter 3: Establishment of a proteomics pipeline for differential proteome analysis using MudPIT, OMSSA, spectral counts and G-test statistics .......................................................................................... 49
  3.1 Introduction .................................................................................................................................................. 50
  3.2 AutoHotkey scripting for automated data analysis ........................................................................................ 52
  3.3 Conversion of raw MS data to DTA and MGF format ..................................................................................... 53
  3.4 Database adaptation and search using the Open Mass Spectrometry Search Algorithm (OMSSA) .......... 54
  3.5 Optimization of OMSSA search .................................................................................................................. 56
  3.6 Correction of the systematic error of precursor mass determination by the orbitrap .................................. 59
  3.7 Spectral counts analysis ............................................................................................................................... 63
  3.8 Implementation of G-test statistics and fold change calculation ................................................................. 64
  3.9 Data analysis in biological context ............................................................................................................. 67
  3.10 References .................................................................................................................................................. 68
  3.11 Appendix .................................................................................................................................................... 71
<table>
<thead>
<tr>
<th>Chapter 4: Linking proteome responses with physiological and biochemical effects in herbicide-exposed <em>Chlamydomonas reinhardtii</em></th>
<th>85</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Abstract</td>
<td>86</td>
</tr>
<tr>
<td>4.2 Introduction</td>
<td>87</td>
</tr>
<tr>
<td>4.3 Materials and methods</td>
<td>89</td>
</tr>
<tr>
<td>4.4 Results and discussion</td>
<td>92</td>
</tr>
<tr>
<td>4.5 Concluding remarks</td>
<td>107</td>
</tr>
<tr>
<td>4.6 References</td>
<td>109</td>
</tr>
<tr>
<td><strong>Chapter 5: Conclusions and outlook</strong></td>
<td>113</td>
</tr>
<tr>
<td>5.1 Assessment of the effects of herbicides on physiological and biochemical parameters</td>
<td>114</td>
</tr>
<tr>
<td>5.2 Assessment of the effects of herbicides on the protein level</td>
<td>115</td>
</tr>
<tr>
<td>5.3 Outlook</td>
<td>117</td>
</tr>
<tr>
<td>5.4 Summarizing remarks</td>
<td>120</td>
</tr>
<tr>
<td>5.5 References</td>
<td>121</td>
</tr>
<tr>
<td><strong>Acknowledgements</strong></td>
<td>123</td>
</tr>
<tr>
<td><strong>Curriculum vitae</strong></td>
<td>125</td>
</tr>
</tbody>
</table>
List of figures

Figure 1.1: Structures, alternative names and chemical nomenclature of the herbicides paraquat, diuron and norflurazon ................................................................. 8
Figure 1.2: Target sites of the herbicides paraquat and diuron within the photosynthetic machinery and of the herbicide norflurazon in the carotenoid synthesis pathway ....................................................... 10
Figure 1.3: Reduction and oxidation of the herbicide paraquat leading to production of superoxide radicals from molecular oxygen ........................................................................... 10
Figure 1.4: Overview of the proteomics analysis pipeline including proteome profiling using MudPIT and subsequent data evaluation ........................................................................... 12
Figure 2.1: Concentration-response curves obtained in the multi-endpoint assay upon exposure of Chlamydomonas reinhardtii to the herbicide paraquat for 2, 6 and 24 h .................................................. 32
Figure 2.2: Concentration-response curves obtained in the multi-endpoint assay upon exposure of Chlamydomonas reinhardtii to the herbicide diuron for 2, 6 and 24 h .................................................. 35
Figure 2.3: Concentration-response curves obtained in the multi-endpoint assay upon exposure of Chlamydomonas reinhardtii to the herbicide norflurazon for 2, 6 and 24 h .................................................. 38
Figure S2.1: Concentration-response curves obtained for cell density and average cell volume upon exposure of Chlamydomonas reinhardtii to the herbicides paraquat, diuron and norflurazon .................................................. 44
Figure 3.1: Simplified workflow of the proteomics analysis pipeline leading to identification of significantly differently expressed proteins based on spectral counting quantitation .................................................. 51
Figure 3.2: Effect of allowed variable modifications and of different numbers of allowed missed cleavages during tryptic digestion on the resulting number of total peptide identifications and identified peptide species in a representative MudPIT run .......................................................................................... 58
Figure 3.3: Effect of the elimination of the systematic mass error produced by the orbitrap .................................................. 61
Figure 3.4: Effect of the elimination of the systematic mass error produced by the orbitrap .................................................. 62
Figure 4.1: Heat map showing proteins functionally related to the photosynthetic machinery which were significantly changed in Chlamydomonas reinhardtii in response to at least one of the exposure conditions used .......................................................................................... 97
Figure 4.2: Heat map showing redox-related proteins which were significantly changed in Chlamydomonas reinhardtii in response to at least one of the exposure conditions used .......................................................................................... 98
Figure 4.3: Heat map showing additional proteins which were significantly changed in Chlamydomonas reinhardtii in response to at least one of the exposure conditions used .......................................................................................... 99
List of tables

Table 2.1: EC50 and LOEC values of the concentration-response relationships of all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicides paraquat, diuron and norflurazon .......................................................... 29

Table S2.1: Parameters of the concentration-response curve fits based on the four parameter logistic model with their 95% confidence interval for all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicide paraquat ................................................................. 45

Table S2.2: Parameters of the concentration-response curve fits based on the four parameter logistic model with their 95% confidence interval for all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicide diuron ................................................................. 46

Table S2.3: Parameters of the concentration-response curve fits based on the four parameter logistic model with their 95% confidence interval for all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicide norflurazon ................................................................. 47

Table 3.1: Lines from a final output file with relevant parameters created during application of *G*-test statistics with multiple-testing adjustment by Benjamini-Hochberg correction to find significant differences between the counts $x_i$ of each protein found in two experimental conditions .................. 66

Table 4.1: Numbers of protein identifiers in significantly enriched functional categories of MapMan annotation in the significantly changed (TOTAL), DOWN- and UP-regulated proteins of *C. reinhardtii* after 6 h of exposure to low and high concentrations of the herbicides paraquat, diuron and norflurazon ......................................................................................................................... 93
Pesticides are widely used to control the growth and proliferation of unwanted target organisms, such as weed, fungi or insects. Following application, particularly in agriculture, substantial amounts of these anthropogenic compounds are introduced into freshwater ecosystems which raises concerns about potentially detrimental effects on non-target organisms. Since herbicides constitute the largest proportion of pesticides used, freshwater algae, which are part of the basis of the ecologically highly relevant aquatic food web, might be particularly at risk. To assess this risk, ecotoxicological testing procedures are typically employed. This involves standardized toxicity testing based on growth inhibition of unicellular algae. Since this endpoint is highly integrative, no information regarding mechanisms of toxic action can be obtained. This understanding, however, is crucial for addressing current and future challenges in ecotoxicology. To obtain it, it is necessary to analyze physiological, biochemical and subcellular molecular effects caused by the investigated compounds which underlie the impacts on the whole-organism level.

The overall goal of this thesis was to investigate the effects of herbicides on green algae with a focus on assessing and linking physiological and biochemical endpoints as well as variations on the protein level, to elucidate the exposure effects and toxic mechanisms of these compounds. The unicellular model organism *Chlamydomonas reinhardtii* was selected as a representative of the green algae. It was exposed to the model herbicides paraquat, diuron and norflurazon, each representing specific toxic mechanisms of action.

Specific physiological and biochemical endpoints were selected. These were pigment content, maximum and effective photosystem II quantum yield, ATP content, esterase and oxidative activity, thus covering a wide range of cellular parameters in the alga. They were investigated in fast and simple setups and compiled in a novel multiple-endpoint assay, in combination with the assessment of growth (chapter 2). The endpoints were measured at different time points during exposure of up to 24 h. For all endpoints and exposure durations effective concentrations inducing 50% response (EC50s), as well as lowest observable effect concentrations (LOECs) were determined when possible. The multiple-endpoint assay provided a highly detailed view of the concentration and time-dependent effects induced by the selected herbicides, thereby confirming and expanding our knowledge about their toxic mechanisms. Furthermore, a number of physiological and biochemical endpoints responded earlier or stronger to the exposures than the integrative endpoint growth. As the various endpoints responded highly specific to exposure to the three herbicides, unique patterns were observed which reflected the different mechanisms of toxicity.

Since the proteome of an organism sensitively shows specific alterations upon exposure, a differential proteome profiling analysis was performed for the assessment of herbicide effects on the subcellular level.
For this purpose a shotgun proteomics method, the multidimensional protein identification technology (MudPIT), was applied and combined with spectral counting quantification and G-test statistics to reveal significant changes in protein abundance. The required computational analysis steps were implemented in a pipeline developed for this work, using public domain software and automation via scripting (chapter 3). Based on the results of the multiple-endpoint assay, two exposure concentrations of each herbicide representing high and low stress levels were selected for analysis. Under these conditions significant changes were detected for 149-254 proteins of diverse cellular function and location (chapter 4). Among the variations found were expected ones, such as glutathione-S-transferases or various thioredoxins, as well as novel candidate markers of exposure, including the photosystem II subunit PsbR and the VIPP1 protein. Functional enrichment analysis on the lists of significantly changed proteins allowed identification of functional protein groups that responded to the exposures. Overall, the results showed activation or suppression in a variety of metabolic pathways and protein complexes, such as various antioxidant defense systems, pigment synthesis and the photosystems. Part of these effects could be linked to observations made using the multiple-endpoint assay, including those reflecting the mechanisms of action of the model herbicides. Furthermore, significant and complex proteome responses to the low concentration conditions demonstrated the high sensitivity of the proteome level compared to physiological and biochemical endpoints. The different exposure conditions induced some common variations on the protein level. Many changes, however, were specific to particular conditions, indicating the high specificity of the proteome response.

In conclusion, this thesis contributed to the understanding of adverse effects induced by the herbicides paraquat, diuron and norflurazon in green algae, and of the underlying toxic mechanisms. The particular importance and usefulness of linking different biological levels influenced by exposures was demonstrated. Both approaches, the multiple-endpoint assay as well as proteome profiling, are promising tools for future ecotoxicological research on toxic mechanisms in green algae, as well as risk assessment of herbicides and other phytotoxicants.
Zusammenfassung


Zusammenfassung

 einige physiologische und biochemische Endpunkte frühere oder stärkere Effekte an als der integrative Wachstumsendpunkt. Da die verschiedenen Endpunkte hochspezifisch auf die Exposition mit den drei Herbiziden reagierten, wurden Effektmuster beobachtet, welche die verschiedenen toxischen Mechanismen widerspiegelten.


Diese Arbeit hat zum Verständnis der schädigenden Effekte sowie der toxischen Wirkmechanismen, der Herbizide Paraquat, Diuron und Norflurazon in Grünalgen beigetragen. Die Bedeutung und Nützlichkeit der Berücksichtigung verschiedener biologischer Ebenen, die durch die Expositionen beeinflusst werden, wurde demonstriert. Die Testbatterie aus mehreren Endpunkten sowie die Proteomanalyse sind vielversprechende Grundlagen für zukünftige ökotoxikologische Forschung in Bezug auf Wirkmechanismen in Grünalgen und die Risikoabschätzung von Herbiziden und anderer phytotoxischer Verbindungen.

VIII
Chapter 1:

General introduction
1.1 Pesticides and their effects on green algae in freshwater systems

1.1.1 Pesticides in freshwater systems

Pesticides are a diverse group of anthropogenic compounds designed to hinder growth and proliferation of their target organisms, which include, among others, weeds, fungi and insects. Annually, large amounts of pesticides are being produced and applied – the global usage totalling to 2.4 million tons in both 2006 and 2007 (Grube et al., 2011). Such products are used in households and gardens, the commercial sector, by industry and governmental institutions. By far the largest and most important user, however, is agriculture, where roughly 80% of the used pesticide amount in the USA are applied (Grube et al., 2011). Unfortunately, most of the pesticides distributed in the fields do not reach the target organisms (Pimentel, 1995) and substantial amounts may be introduced into rivers, lakes and other freshwater bodies due to leaching, surface run-off and other processes. This pollution may affect water quality and produce unwanted impacts on a variety of non-target organisms. While inhibiting effects of pesticides are often due to their interference with crucial metabolic pathways, a wide diversity of organisms, ranging from microorganisms, freshwater algae and aquatic plants to invertebrates and fish, may become affected due to the presence of pesticide-sensitive sites in their physiology. The largest proportion of the globally applied pesticides are in fact herbicides (Grube et al., 2011), that is, compounds applied to control the growth of unwanted plants and algae. Therefore, a major concern regarding potential non-target organisms in freshwater systems is about photosynthetic organisms.

1.1.2 Unicellular green algae: ecological role and major physiological features

Unicellular green algae share many cellular components with the plants targeted by herbicides. These organisms are an important component of the phytoplankton, where they account for a substantial part of the oxygen and biomass production. Being primary producers, unicellular algae reside at the basis of the aquatic food web and thus bear high ecological relevance. Adverse effects on these organisms would also negatively influence higher trophic levels, including zooplankton and fish, thus disturbing the whole ecosystem. Hence, effects of anthropogenic compounds on green algae need to be assessed thoroughly.

As already mentioned, unicellular green algae are photosynthetic organisms. They are able to grow photoautotrophically using carbon dioxide as carbon source and photosynthetically active radiation originating from the sun as energy source. The light-driven process of photosynthesis occurs in specialized organelles, the chloroplasts. In the sequence of light reactions, first the water is split into protons and molecular oxygen at photosystem (PS) II. From there, the free electrons are flowing through several multi-protein complexes known as cytochrome-b$_{6}$f-complex and PS I. The main outcomes of linear electron transport chain are reducing equivalents (NADPH) as well as ATP, whose production relies on the generation of a proton gradient across the thylakoid membrane and activity of ATP synthase. Both molecules, NADPH and ATP, are required for driving the dark reactions, which include the Calvin cycle, in which CO$_{2}$
is fixed in organic metabolites, and subsequent production of sugars and starch. The light necessary to drive this whole process is collected by pigments which are synthesized through complex pathways specific to photosynthetic organisms. The most important photosynthetic pigments in green algae are chlorophylls and carotenoids.

A major issue that photosynthetic organisms have to deal with is the generation of reactive species during photosynthesis and other cellular processes. In particular, reactive oxygen species (ROS), including superoxide radicals, hydrogen peroxide, singlet oxygen, and hydroxyl radicals, are continuously created during the light reactions and hence need to be continuously eliminated. For this purpose, antioxidant defense mechanisms have evolved which balance the redox state (the ratio of cellular oxidants to antioxidants) in the cells. These systems are comprised of carotenoids and several other non-enzymatic antioxidants as well as of numerous enzymes including superoxide dismutase, ascorbate and glutathione peroxidases and glutathione reductase (Apel and Hirt, 2004; Foyer and Noctor, 2009).

1.2 Approaches in the ecotoxicology of green algae

1.2.1 Ecotoxicology and the role of standardized tests

The interdisciplinary field of ecotoxicology seeks to explore and predict effects of toxicants and other stressors on organisms, looking at all the different biological levels ranging from the subcellular to the whole ecosystem. Exploratory studies produce a basis for risk assessment of anthropogenic compounds released into the environment.

To facilitate the development of regulations which are needed to protect the environment, efficient and standardized approaches are needed for assessing the potential risks imposed by pesticides on non-target organisms. When looking at phytotoxicity, unicellular green algae are often used as test organisms. Growth is the most common endpoint measured in the standardized tests used for evaluation of concentration-response relationships, for example according to OECD or ISO guidelines (ISO, 2012; OECD, 2011). The results of such tests are often reported as toxicant concentration values where specific thresholds of effect on the investigated endpoint are reached. These include, among others, the effective concentration causing 50% inhibition (EC50) as well as the lowest observable effect concentration (LOEC). Values like these are used as measures of the toxic potential of chemicals and to predict the ecological effects they might exert when introduced into the environment.

1.2.2 Relevance of mechanistic understanding of toxicity

One of the key aspects of ecotoxicological research ultimately required for successful risk assessment is the thorough understanding of the mechanisms by which toxicants affect organisms and which ultimately cause the inhibition of growth in green algae, that is, effects on the whole-organism level. This understanding
is not only helpful for analyzing new or unknown toxicants but is crucial for addressing current challenges in ecotoxicology. These include assessment, modelling and prediction, among others, of the effects of low toxicant concentrations, complex toxicant mixtures and of multiple stressors (the simultaneous exposure to toxicants and environmental stressors) (Eggen et al., 2004). These are research questions, where it is impractical or even impossible to test all possible scenarios and combinations of toxicants, stressors etc. individually in the traditional way.

To explore toxic mechanisms it is not sufficient to study only the effects on growth as performed in standard toxicity testing procedures. Instead, the underlying processes have to be elucidated using mechanism of action oriented analyses. This requires the investigation of biological levels below the whole cell in addition to growth, thus targeting specific components and parameters of the cellular machinery’s physiology, biochemistry and molecular levels (transcripts, proteins, metabolites). Following the uptake of a toxicant, a variety of physiological, biochemical and molecular disturbances are expected to be induced in cells through the interaction with the target site(s) and subsequent downstream effects. The exact pattern and magnitudes of these effects depend on the toxicant’s mechanisms of action, the stresses caused by the respective exposure concentration and other factors. Measurements of diverse cellular components and parameters and their changes in response to exposure provide access to the complexity of effects inside the cell. Moreover, many of these parameters tend to respond earlier and stronger than growth, thus exhibiting higher sensitivity and possibly constituting markers of early response. The sensitivity is one of the prime advantages of using diverse physiological and biochemical endpoints as well as as indicators of exposure, complementary to or even instead of growth.

1.2.3 Physiological and biochemical parameters of green algae as ecotoxicological endpoints

In green algae and other photosynthetic organisms a diversity of physiological and biochemical endpoints was studied and evaluated, as reviewed elsewhere (Brain and Cedergreen, 2009; Ralph et al., 2007; Torres et al., 2008; Valavanidis et al., 2006). In the following, a few examples of cellular parameters are provided which may be used as endpoints in the ecotoxicological investigations of toxicant exposure effects in unicellular green algae.

One of the most important cellular systems in green algae is the photosynthetic machinery. Its functionality can be assessed by estimating photosynthesis rate or by measuring PS II chlorophyll fluorescence parameters. Connected to this are the levels of photosynthetic pigments, particularly of chlorophylls and carotenoids. Furthermore, the energy metabolism and energetic state of the algal cell, which can be addressed by measuring ATP or NAD(P)H levels, might be of interest. Also enzyme activity, for example, that of esterases, may be assessed as a general measure of metabolic activity. Other, more specific enzymes potentially responding to toxicant exposure are those of the xenobiotic defense mechanisms possibly involved in the elimination of toxic compounds. Another highly important parameter is the cellular
redox state. To assess the extent of oxidative stress and its consequences, one can determine the levels of ROS, the activity of enzymes of the antioxidative defense or the amount of oxidation products among cellular components such as lipids, proteins or DNA. Since the effect of different toxicants and stress conditions on such parameters may vary, it is advantageous to assess not a single but an array of such endpoints covering a wider range of components and aspects of the cellular physiology and biochemistry. This may reveal exposure markers, that is, endpoints showing a higher sensitivity and specificity, and possibly lead to toxic mechanism and stress specific response patterns, both of which being a great asset for a mechanism oriented analysis.

A downside of using such alternative endpoints compared to determination of growth alone is an often much higher labor intensity as well as requirement for more complex and costly instrumentation, which among others may include flow cytometers, pulse-amplitude-modulated fluorometers and oxygen sensitive electrodes. Complicated measurement workflow may limit the number of toxicants, concentrations and time points, as well as the diversity of endpoints that can be investigated in a given time frame using given resources. However, a simplification of such assays may help minimizing or overcoming such obstacles. A further crucial issue is the possibility for the lack of obvious connections between these parameters and the integrative final effects such as proliferation or growth, which may limit the overall ecological relevance of the tests based on such endpoints. However, if the link to growth would be established, prediction of ecologically relevant toxicant effects important for risk assessment may be achieved. For this and to generally assess their usefulness, physiological and biochemical endpoints have to be evaluated, making it necessary to investigate the effects they indicate upon exposure to toxic model compounds which are well characterized in their mechanisms of action allowing to examine corresponding hypotheses regarding the effects they induce.

1.2.4 Proteomics and its potential applications in ecotoxicology

The molecular level of biological organization, comprising genes, mRNA transcripts, proteins, and metabolites, largely defines what happens on the biochemical and physiological levels. Over the last decades, sophisticated analytical methods for studying gene transcripts, proteins, and metabolites have been developed, opening up a range of new possibilities for many research fields, including ecotoxicology. The analysis of transcripts and proteins has received particular attention as a means for exploration of effects on the subcellular level. Molecular effects could then be linked to higher levels of biological organization, exposures, and to mechanisms of toxicity (Brain and Cedergreen, 2009; Lemos et al., 2010; Schirmer et al., 2010).

The term “proteome” is defined as the total complement of proteins present in an organism (Wasinger et al., 1995). Analyzing the proteome may provide more reliable information about actual gene expression and function compared to the transcriptome, as proteins are the prime functional units and workhorses of the cellular machinery. Furthermore, the connection between individual mRNA and protein abundance may be
rather loose, thus limiting the prediction of the protein expression based on transcript information only (Garcia-Reyero and Perkins, 2011; Schwanhäusser et al., 2011).

As representative of the “omics” approaches, proteomics includes high-throughput methods aimed at global profiling of as many proteins as possible, while “targeted” proteomics approaches may be used to measure individual ones and to analyze subproteomes (Nesatyy and Suter, 2008). At the initial stage, the global profiling is essential for obtaining an overview over whole biological networks. In principle, the comprehensive nature of a profiling analysis may allow covering all components of the proteome in the organism of interest. However, the analytical capabilities are currently limited, due, among other reasons, to sample preparation as well as the insufficient resolution and difficulties in detecting low abundance proteins.

The use of proteomics methods in ecotoxicology is, among other aspects, based on the notion that the exposure of an organism to toxicants provokes characteristic responses within its proteome (Lemos et al., 2010; Monsinjon and Knigge, 2007; Nesatyy and Suter, 2008). Using recent proteomics methods, a large number of proteins can be assessed simultaneously. Since the proteome has a complex multi-component characteristic, responses highly specific to the exposures can be expected. Furthermore, many components may change significantly to even subtle stresses, showing the high sensitivity of the proteome response to toxicant exposures (Lemos et al., 2010; Monsinjon and Knigge, 2007; Nesatyy and Suter, 2008). High specificity and high sensitivity may both help understanding the toxic mechanisms affecting the organism and determining its response to the exposure. Novel protein markers of exposure may also be found, even when no previous knowledge about their connection to the mechanisms of action is available. This can support formulating hypotheses about toxicity mechanisms of a particular chemical, which can then be tested in further investigations (Lemos et al., 2010; López-Barea and Gómez-Ariza, 2006).

Based on the detailed overview provided by proteomics, it offers unique advantages over endpoints based on higher biological levels such as the physiological and biochemical parameters outlined in the previous section. Still, the need for thorough anchoring of proteome variations in responses on these levels is of high importance and should not be neglected (Lemos et al., 2010).
1.3 Scope of the thesis

The overarching objective of this thesis was a mechanism oriented analysis of herbicide exposure effects on green algae by assessing different endpoints in addition to growth. The focus was on examining the responses to herbicides looking at (1) growth as well as several physiological and biochemical endpoints and (2) the protein level. Major attempts have been made to link the two. The advantages of using such approaches were evaluated.

(1) The effects observed on different assessed endpoints and growth upon exposure were compared to each other and related to the mechanisms of action of the investigated herbicides. Furthermore, the development of the concentration-dependent responses over time as well as endpoint response patterns delivered by the multiple-endpoint assay were evaluated, and the most sensitively responding endpoints were identified.

(2) A proteome profiling analysis was applied to elucidate subcellular exposure responses by unraveling the complex changes on the protein level caused by the investigated herbicides. These analyses led to the identification of responding subcellular systems and identified candidate protein markers for the exposure conditions. Links to growth and other physiological and biochemical endpoints and growth as well as the mechanisms of action were proposed.

For both approaches, one particular goal was to explore responses at low concentrations and correspondingly mild stress levels as well as at short exposure durations to better address the question of the sensitivity of the applied endpoints. Overall, the analyses carried out in the frame of this thesis work have led to an improved understanding of herbicide effects in unicellular green algae.

1.3.1 Model organism

The unicellular green alga *Chlamydomonas reinhardtii* was selected as test organism. This species has been used as model organism in many fields of cell physiology and biochemistry, including studies of photosynthetic, cell cycle and flagellar processes (Harris, 2001). Its widespread use led to the availability of the complete sequenced genome (Merchant et al., 2007) as well as a variety of genome annotations including Gene Ontology and MapMan (Lopez et al., 2011; May et al., 2008). The availability of this information greatly supports proteomics analyses and subsequent interpretations of the obtained data. Furthermore, *C. reinhardtii* has also been used in numerous ecotoxicological studies with natural toxicants such as metals (Szivák et al., 2009) as well as with anthropogenic ones such as herbicides (Fischer et al., 2010; Jamers and De Coen, 2010). The proteome of this organism has also been specifically looked at in differential proteomics studies which included stressors such as high light (Förster et al., 2006), heat (Mühlhaus et al., 2011), and cadmium (Gillet et al., 2006).
1.3.2 Model herbicides

Herbicides act via a wide variety of toxic mechanisms affecting diverse cellular targets (Duke, 1990; Wakabayashi and Böger, 2004a; Wakabayashi and Böger, 2004b). To take this into account and to analyze the response of the chosen endpoints and the proteome in regard to different mechanisms of action, three different herbicides were selected for this work. These herbicides, paraquat (PQ), diuron (DR), and norflurazon (NF), may be regarded as model compounds since their phytotoxic target sites and mechanisms of action are very well established through many studies. Though the mechanisms of action of these three herbicides are essentially different, they also share major commonalities such as the chloroplast targeting, the strong dependence of the toxicity on light as well as the involvement of reactive oxygen species and photooxidative processes (Hess, 2000). Figure 1.1 shows chemical structures of these herbicides and provides the exact chemical designations as well as alternative common names used for them in the literature.

![Chemical structures of paraquat, diuron, and norflurazon](Figure 1.1)

The bipyridilium compound **PQ** (CAS 1910-42-5) is an inducer of superoxide radical generation at PS I (Hess, 2000) and, presumably to much lower extent, at the mitochondrial respiration chain (Cochemé and Murphy, 2008). The mechanism involves the reduction of the PQ molecule by diverting electrons from linear photosynthetic electron transport at the iron-sulfur clusters of PS I (Fujii et al., 1990; Parrett et al., 1989) (Figure 1.2A) and re-oxidizing by transferring them to molecular oxygen (Figure 1.3). This provokes a subsequent increase of levels of other reactive oxygen species (ROS) such as hydrogen peroxide and
hydroxyl radical due to various reactions which follow superoxide radical generation (Hess, 2000). The consequence is the induction of oxidative stress - an imbalance between levels of cellular oxidants and antioxidants (Halliwell, 2006). Once this stress becomes sufficiently intense, the antioxidant defense of the organisms gets overwhelmed and there occurs a widespread oxidation of biomolecules, particularly of lipids, proteins, DNA, leading to loss of vital cellular functions and, finally, to death.

The phenylurea herbicide **DR** (CAS 330-54-1) leads to reduction in photosynthetic activity. It acts through the reversible inhibition of linear electron transport at the Q₈ binding pocket of PS II by preventing plastoquinone from accepting electrons (Figure 1.2A), a trait shared by many other economically relevant herbicides (Hess, 2000). The resulting inhibition of the light reactions leads to starvation of the organism due to disturbance of the photoautotrophic metabolism. The blockage of linear electron flow may also lead to lipid peroxidation and subsequent membrane damages due to prevention of discharge of excitation energy collected by the PS II light harvesting complex (Hess, 2000).

The main target site of the third model herbicide, the pyridazinone **NF** (CAS 27314-13-2), is located on the enzyme phytoene desaturase (Sandmann et al., 1984), a protein which catalyzes a crucial step in the synthesis of carotenoids (Grossman et al., 2004) (Figure 1.2B). NF competes with the enzyme cofactor plastoquinone and not with its substrate phytoene (Breitenbach et al., 2001). Carotenoids have a number of crucial biological functions in photosynthetic organisms such as supplementing the capacity for light harvesting as well as protecting the photosynthetic machinery by quenching excited chlorophylls as well as the ROS singlet oxygen (Grossman et al., 2004; Hess, 2000). Since there is a turnover of carotenoids, the NF induced inhibition of carotenoid synthesis prevents their replacement, thus causing their depletion (Hess, 2000). This results in the loss of protection within the photosynthetic machinery leading to damage to photosynthetic membranes as well as bleaching, i.e. photooxidative destruction of chlorophylls (Hess, 2000; Wakabayashi and Böger, 2004a). Furthermore, the loss of capability for quenching singlet oxygen and triplet chlorophylls may lead to the formation of lipid peroxides (Hess, 2000). In addition, NF can bind to the DR target site, although with much lower affinity than DR (Tischer and Strotmann, 1977).
Figure 1.2: Panel A: target sites of the herbicides paraquat (red dotted box) and diuron (blue dotted box) within the photosynthetic machinery (scheme based on work by I. Szivák). Panel B: target site of the herbicide norflurazon (green dotted box) in the carotenoid synthesis pathway (redrawn after Grossman et al. (2004)). Cyt f: cytochrome f, LHC: light harvesting complex, PC: plastocyanine, Pheo: pheophytin, PQN: plastoquinone, PS: photosystem

Figure 1.3: Reduction and oxidation of the herbicide paraquat leading to production of superoxide radicals from molecular oxygen (redrawn after Cochemé and Murphy (2008)).
1.3.3 Proteome profiling analysis using multidimensional protein identification technology

Various methods are available for performing a differential global proteomics analysis (Nesatyy and Suter, 2008). Two-dimensional gel electrophoresis, a traditional gel-based technique, is frequently used to date. However, “gel-free” approaches offer several advantages including lower labor-intensity, the possibility for automation and a potentially higher proteome coverage due to less bias against hydrophobic proteins as well as by better ability for detecting lower abundance proteins (Nesatyy and Suter, 2008). One of these methods is known as multidimensional protein identification technology (MudPIT) (Delahunty and Yates, 2005; Washburn et al., 2001), a “shotgun proteomics” method. It was specifically designed to allow for an extensive coverage of proteins in complex samples like those originating from whole cell lysates. MudPIT was proposed to have potential applications in ecotoxicology for exploring chemical mechanisms of action (Uleberg et al., 2010).

The outline of proteomics analysis based on MudPIT is shown in Figure 1.4. Similar to other liquid chromatography (LC) - tandem mass spectrometry (MS/MS) approaches, the protein mixture to be analyzed with MudPIT is first proteolytically digested and the resulting peptides are loaded on a nanocolumn. The specialty of MudPIT is the application of a three-phasic column filling. This allows applying a 2D-LC gradient that separates the peptides according to their binding strength to a strong cation exchanger first and subsequently to a reversed phase C18 material. Eluted peptides are then ionized and introduced into an MS/MS instrument by an electrospray ion source. For as many ions as possible, MS/MS fragmentation spectra are collected which are then used to search protein databases with suitable matching algorithms. This leads to the identification of peptides and the corresponding proteins that were present in the original sample. The numbers of peptide identifications can be used as so-called “spectral counts”. They form the basis for a label-free quantification approach which can be used for comparing protein expression levels across different samples (Nesatyy and Suter, 2008) since spectral counts correlate well with the amount of protein present in a sample (Liu et al., 2004; Old et al., 2005). Using appropriate statistical tests such as G-test, the significance of observed differences can be further evaluated (Zhang et al., 2006).
Figure 1.4: Overview of the proteomics analysis pipeline including proteome profiling using MudPIT and subsequent data evaluation (photograph: M. J.-F. Suter). ESI: electrospray ionization, LC: liquid chromatography, MS: mass spectrometry, MS/MS: tandem mass spectrometry.
1.4 Outline of the thesis

The work for the assessment of exposure effects of the herbicides PQ, DR, and NF on *C. reinhardtii* has been performed and is presented in three main parts:

1. **Application of a multiple-endpoint assay (Chapter 2)**

   Several physiological and biochemical endpoints, covering a broad spectrum of algal cell parameters, were selected and combined with the assessment of growth to establish a novel multiple-endpoint assay. In order to make this assay battery feasible for simultaneous measurement of several endpoints, simple and fast setups were used. With this approach, the response of the endpoints could be examined at a wider range of concentrations as well as at different exposure durations. This study resulted in a comprehensive concentration- and time-resolved picture of the herbicide exposure effects. Furthermore, the different endpoints could be compared in regard to their sensitivity. The interrelationships between these endpoints and growth were explored. Finally, the observed responses were related to the herbicide mechanisms of action. The results of this work have been published in Aquatic Toxicology (2012) 110-111: 214-224.

2. **Establishment of the proteomics analysis pipeline (Chapter 3)**

   In order to evaluate the proteomics profiling data collected with MudPIT, a number of analysis steps needed to be carried out, which start with the MS raw data and finally result in the lists of proteins significantly changed in exposed organisms as compared to control. Thus, an important part of this thesis was the development of a computational analysis pipeline implementing all these steps. This workflow was based on the usage of public domain software and the automation of individual steps via scripting, particular in regard to the processing of multiple MudPIT runs.

3. **Proteomics analysis of *C. reinhardtii* exposed to herbicides (Chapter 4)**

   The proteome of *C. reinhardtii* was analyzed to explore the changes in protein abundance upon exposure to selected concentrations of the herbicides whose effects were already characterized using the multiple-endpoint assay. The goal was to determine which individual proteins and functional protein groups were affected by the exposures and how these changes are connected to the disturbances observed among the physiological and biochemical parameters and to the underlying mechanisms of action. For the proteome profiling analysis necessary to address these objectives, the outlined MudPIT approach was combined with label-free spectral counts quantitation and statistical analysis carried out within the established proteomics analysis pipeline. Enrichment analysis was used to determine the responding functional protein groups. This chapter is in press in the Journal of Proteomics.
1.5 References


Chapter 2:

Multiple-endpoint assay provides a detailed mechanistic view of responses to herbicide exposure in *Chlamydomonas reinhardtii*

Holger Nestler, Ksenia J. Groh, René Schoenenberger, Renata Behra, Kristin Schirmer, Rik I. L. Eggen, Marc J.-F. Suter

Aquatic Toxicology (2012) 110-111: 214-224
2.1 Abstract

The release of herbicides into the aquatic environment raises concerns about potential detrimental effects on ecologically important non-target species, such as unicellular algae, necessitating ecotoxicological risk assessment. Algal toxicity tests based on growth, a commonly assessed endpoint, are integrative, and hence do not provide information about underlying toxic mechanisms and effects. This limitation may be overcome by measuring more specific biochemical and physiological endpoints. In the present work, we developed and applied a novel multiple-endpoint assay, and analyzed the effects of the herbicides paraquat, diuron and norflurazon, each representing a specific mechanism of toxic action, on the single celled green alga *Chlamydomonas reinhardtii*. The endpoints added to assessment of growth were pigment content, maximum and effective photosystem II quantum yield, ATP content, esterase and oxidative activity. All parameters were measured at 2, 6 and 24 h of exposure, except for growth and pigment content, which were determined after 6 and 24 h only. Effective concentrations causing 50% of response (EC50s) and lowest observable effect concentrations (LOECs) were determined for all endpoints and exposure durations where possible. The assay provided a detailed picture of the concentration- and time-dependent development of effects elicited by the analyzed herbicides, thus improving the understanding of the underlying toxic mechanisms. Furthermore, the response patterns were unique to the respective herbicide and reflected the different mechanisms of toxicity. The comparison of the endpoint responses and sensitivities revealed that several physiological and biochemical parameters reacted earlier or stronger to disturbances than growth. Overall, the presented multiple-endpoint assay constitutes a promising basis for investigating stressor and toxicant effects in green algae.

Keywords
toxicity test, multiple-endpoint assay, *Chlamydomonas reinhardtii*, paraquat, diuron, norflurazon

Highlights
We developed a novel multiple-endpoint assay.
We analyzed concentration- and time-dependent effects of herbicides on *C. reinhardtii*.
We used the model herbicides paraquat, diuron and norflurazon.
Response patterns were unique for each herbicide, reflecting the different toxic mechanisms.
Several parameters reacted earlier or stronger to disturbances than growth.
2.2 Introduction

Herbicides constitute the largest proportion of the pesticides used in agriculture. However, only a small amount of applied chemicals actually reaches the target plants while an overwhelmingly larger portion is introduced into the environment as side effect (Pimentel, 1995). This raises concerns about the potentially detrimental effects on ecologically relevant non-target species, such as microalgae. Because these organisms are a significant part of the aquatic food web, herbicide effects on algae could also influence higher trophic levels. Ecotoxicological risk assessment of such compounds is traditionally performed using standardized tests (e.g. according to ISO and OECD guidelines) which focus on growth inhibition in unicellular algae. Although it is highly relevant ecologically, inhibition of growth is an integrative and rather unspecific endpoint, because it can be caused by toxicant interference with a wide range of cellular processes. Hence, current growth-based standard tests do not provide information about underlying toxic mechanisms and effects. Without mechanistic knowledge, however, prediction of toxicity outcome remains elusive, thus constituting a real bottleneck in risk assessment given the wide variety of chemicals that need to be assessed. Supplementing conventional tests on growth inhibition with a number of more specific physiological and biochemical endpoints may help overcome this limitation. Furthermore, inclusion of additional endpoints is of particular importance when effects must be understood and predicted under more dynamic and complex conditions reflecting real environmental situations.

Several studies have employed the multiple-endpoint approach by using different sets of parameters to analyze the response of unicellular algae to herbicides such as paraquat (Jamers and De Coen, 2010; Prado et al., 2009), flumioxazin (Geoffroy et al., 2004), isoproturon and terbutryn (Rioboo et al., 2002). Among the assessed parameters were chlorophyll content and fluorescence, and metabolic and antioxidant enzyme activity, as well as DNA damage. Several benefits of such an approach could be demonstrated. First, important insights into the mechanisms of toxicity were gained. Second, information on the relative sensitivity of the various endpoints to toxicant concentration and / or exposure duration could be obtained. This allowed identifying specific endpoints which effectively detect disturbances caused by a particular phytotoxicant. Yet, these studies were somewhat limited in that they focused on one specific toxicant or mechanism of action only, or by a rather narrow range of tested concentrations, thus for instance hampering fitting of concentration-response relationships. However, applied more broadly, a multiple-endpoint approach may allow distinguishing modes of toxic action of phytotoxicants as was demonstrated in a study using various flow cytometric fluorescence markers (Adler et al., 2007), or employing an extended bioassay (Neuwoehner et al., 2008). Taken together, all these studies have demonstrated that multiple-endpoint assays have a great potential to aid risk assessment of toxicants by providing important insights into their mechanisms of action.

In the present study, we assembled and tested a novel multiple-endpoint assay to analyze the effects of three herbicides on the single celled green alga Chlamydomonas reinhardtii. C. reinhardtii, which for decades served as a model in molecular biology and photosynthesis research (Harris, 2001), has recently
Chapter 2

gained importance in ecotoxicology where it is used to analyze the effects of stressors, including various herbicides (Fischer et al., 2010; Jamers and De Coen, 2010; Reboud, 2002). Three compounds were selected from the available diversity of herbicides as models for a specific toxic mechanism: paraquat (PQ), catalyzing the generation of superoxide radicals thereby causing oxidative stress, diuron (DR), inhibiting photosystem (PS) II, and norflurazon (NF), inhibiting carotenoid synthesis. Today, application of these compounds is partially restricted, but all three of them are still in use in parts of the world (see e.g. http://www.pesticideinfo.org/, access date: October 2011). Maximum concentrations found in polluted waters were 1 nM (Ibáñez et al., 1997) and 340 nM (Amondham et al., 2006) for PQ, 52 nM for DR and 96 nM for NF (Wightwick and Allinson, 2007).

The physiological and biochemical endpoints, which we selected in addition to growth, covered a wide range of cellular parameters expected to be affected by the exposure. The first parameter was the loss of photosynthetic pigments, i.e. bleaching, which is caused by many common classes of herbicides (Moreland, 1980; Wakabayashi and Böger, 2004). Because the effects of a particular compound on different pigments may vary, we decided to assess the three major pigment pools, chlorophyll $a$, chlorophyll $b$, and the carotenoids. We further included maximum quantum yield ($F_{v}/F_{m}$) as a measure for maximal PS II photochemical efficiency and effective quantum yield ($\Phi_{PSII}$) of PS II as a measure of photochemical efficiency of PS II in the light and of overall photosynthesis (Maxwell and Johnson, 2000). These physiological parameters, strongly influenced by various toxicants, are frequently used in bioassays (Juneau et al., 2007; Ralph et al., 2007). Likewise, the content of ATP, the central energy transfer metabolite, can be expected to change in response to organic toxicants (Cid et al., 2008; Hourmant et al., 2009; Nyberg, 1989). Hence, this biochemical endpoint was included in our assay to characterize the general energetic state of the algae. In photosynthetic organisms, ATP content is mainly determined by photophosphorylation, the main source of ATP, and carbon fixation, the major sink for the energy delivered by this compound. Another biochemical parameter added was esterase activity, which can be considered as a general metabolic indicator. It was also important for the assessment of oxidative activity, because the method used required the activity of esterases (Bartosz, 2006). The assessment of oxidative activity was included because toxic actions of herbicides often involve direct or indirect induction of oxidative stress and damage (Hess, 2000). In previous ecotoxicological studies in unicellular algae, the oxidative activity measurement in response to metals and PQ indicated mechanisms of toxic action involving reactive oxygen species (ROS) formation and oxidative stress (Jamers and De Coen, 2010; Knauert and Knauer, 2008; Szivák et al., 2009). Furthermore, it was found that oxidative activity reacted very sensitively, underlining its potential usefulness for toxicity assessment.

The goal of this study was to simultaneously obtain measurements of different endpoints in order to directly compare responses and sensitivities to the herbicide exposures and to unravel the links between them and in relation to growth. Additionally, we investigated the influence of exposure duration on the observed responses to see how these develop over time. Because we expected the physiological and biochemical parameters to respond earlier than an impact on growth became measurable, two relatively early time points
that have rarely been considered before, were assessed as well. The inclusion of three herbicides in this study served to evaluate our test system in regard to its ability to distinguish between differing mechanisms of toxic action.
2.3 Materials and methods

2.3.1 Test organism and culture conditions

The unicellular freshwater alga *Chlamydomonas reinhardtii* wild type strain CC-125 (mt+) was obtained from the Chlamydomonas Genetics Center (Duke University, NC). The algae were grown autotrophically in the liquid medium “Talaquil”, which contains nutrient concentrations typical for Swiss hard waters (BAFU, 2009). It was prepared as described previously (Le Faucheur et al., 2005), except that NaNO$_3$, which *C. reinhardtii* cannot utilize as nitrogen source, was substituted by NH$_4$Cl. Cultivation was performed in glass Erlenmeyer flasks on a High Technology Infors shaker (Infors, Bottmingen, Switzerland) at 90 rpm and 25 °C under continuous illumination of 120 ± 10 µmolm$^{-2}$s$^{-1}$ photosynthetically active radiation (PAR) provided by cool-white fluorescent lamps.

2.3.2 Test substances, test solutions and chemical analysis

Stock solutions of paraquat (1,1'-dimethyl-4,4’-bipyridinium, CAS 1910-42-5; Sigma Aldrich, Buchs, Switzerland), diuron (3-(3,4-dichlorophenyl)-1,1-dimethylurea; CAS 330-54-1; Sigma Aldrich), and norflurazon (4-chloro-5-methylamino-2-(α,α,α-trifluoro-m-tolyl)pyridazin-3(2H)-one; CAS 27314-13-2; Chem Service Inc, West Chester, PA, USA) were prepared directly in the culture medium, thus avoiding solubilization in an organic solvent. After sterilizing through a Millipore membrane filter (0.45 µm), they were stored at 4 °C until use. Analytical validation of the herbicide concentrations used in the experiments was achieved by high performance liquid chromatography coupled to an API 4000 tandem mass spectrometer (Applied Biosystems, Rotkreuz, Switzerland). No significant differences between nominal and effective concentrations were found. The same method was applied to confirm stability of the herbicide concentrations over the course of a 24 h experiment without presence of the test organism.

2.3.3 Herbicide exposure experiments

If not stated otherwise, the experiments were performed three times independently as described below. *C. reinhardtii* cells in late exponential phase were harvested and added to the exposure vessels (Schott Duran glass crystallization dishes with cover) at an initial cell density of 5.5 × 10$^5$ cells × mL$^{-1}$ in a volume of 30 mL culture medium (determination of cell density see below). For the exposures, we selected 12 different concentrations of the three herbicides, distributed around the EC50 values determined in preliminary experiments. The concentration ranges for PQ, DR and NF were 1 nM – 10 µM, 0.1 nM – 3.3 µM, and 3.3 nM – 80 µM, respectively. Additionally, three chemical-free replicate controls were included in each experiment. Exposure took place with a moderate PAR intensity of 80 ± 1 µmolm$^{-2}$s$^{-1}$ provided by cool-white fluorescent lamps at a temperature of 25 °C. The cultures were stirred continuously. At 2, 6, and 24 h after start of exposure the physiological parameters were measured as outlined below. Cell density and
average cell size were additionally determined at the beginning of each experiment, while pigment analysis was performed only after 6 and 24 h.

2.3.4 Measurement of the endpoints

All procedures described in this section were tested and optimized in preliminary investigations. The main feature of most assays was a small sample volume which allowed the use of 96 well plates and a multiwell plate reader for fast read-out. Absorbance, luminescence and fluorescence were measured on an Infinite M200 plate reader (TECAN, Männedorf, Switzerland). In our batch culture setup, the control values changed over time due to growth. However, these changes were reproducible between experiments. The variability between control replicates was not significant compared to the effects of actual exposures as evaluated by statistical analysis described below. The validation criteria for experiment performance were that the cell density of controls should have roughly doubled after 24 h, and Fv/Fm should be about 0.72 at all time points.

Cell density (mL\(^{-1}\)) and average cell volume in µm\(^3\) (= fL) were determined using an electronic particle counter (orifice 50 µm; Multisizer II; Beckman Coulter, Fullerton, CA, USA). Pigment content, ATP content, esterase activity and oxidative activity (see below) were normalized to cell density. Cell volume density (µm\(^3\)mL\(^{-1}\)) of the cultures was calculated by multiplying the average cell volume by cell density. The increase in cell volume density of each culture after 6 and 24 h of exposure was calculated by subtracting the initial cell volume density (0 h). The obtained value is referred to as growth in the present study. The use of this growth parameter integrates effects of toxicants on cell density and average cell volume. It has been shown that integration of information on changes in cell volume are necessary to accommodate for complex growth patterns of unicellular algae in batch cultures (Altenburger et al., 2008).

Maximum quantum yield (Fv/Fm) and effective quantum yield (Φ\(_{\text{PSII}}\)) were determined using a pulse-amplitude-modulated (PAM) chlorophyll fluorometer (Maxi-Imaging-PAM, Heinz Walz GmbH, Effeltrich, Germany). Aliquots of the cultures were transferred into black 96-well plates (Greiner Bio-One) and subsequently dark adapted for 5 min with only the measuring light activated prior to application of five saturation pulses 1 min apart. Following that, this scheme of adaptation and measurement was repeated with concomitant illumination of the same samples with 82 µmolm\(^{-2}\)s\(^{-1}\) of actinic light. Five measurements of Fo, Fm, F' and Fm' each were averaged. Fo stands for the fluorescence yield of PS II in the dark adapted state, F' is the fluorescence yield of PS II under actinic irradiation, and Fm and Fm' are the maximal fluorescence yields in the dark and light adapted state, respectively. Average values were used for calculating Fv/Fm = (Fm-Fo)/Fm (Kitajima and Butler, 1975), and Φ\(_{\text{PSII}}\) = (Fm'-F')/Fm' (Genty et al., 1989). 5 min of dark adaptation were found to be sufficient since Fv/Fm values around 0.72 were consistently obtained in the controls, and longer dark adaptation periods did not result in higher Fv/Fm values. Furthermore, different time intervals as well as higher than reported numbers of consecutive saturation pulses were tested, and it was found that the saturation pulses did not cause photoinhibition.
For the estimation of the chlorophyll \(a\), chlorophyll \(b\), and total carotenoid contents, the pigments were extracted based on the procedure published previously (Arnon, 1949). This method was the first to use 80% acetone and proved to be highly effective for extracting pigments from \(C\). \textit{reinhardtii}. Briefly, cells from sampled aliquots were pelleted by centrifugation at 16,000 g, resuspended in 80% cold acetone and incubated for 5 min on ice in the dark, thus allowing sufficient release of the pigments. Following another centrifugation step, the supernatants were transferred into a transparent 96-well plate (Greiner Bio-One). In order to reduce sample volume and measuring time, a microplate assay was used, adapted from the common photometry based procedure employing cuvettes. Absorption was measured at 663, 647, and 470 nm for calculating concentrations of, respectively, chlorophyll \(a\) and \(b\) and total carotenoids, as specified earlier (Lichtenthaler, 1987). The absorption at 750 nm served as null value (Porra et al., 1989).

ATP levels were measured using a bioluminescence based assay kit (BacTiter-Glo™ Microbial Cell Viability Assay; Promega, Madison, WI, USA) in opaque white 96-well plates (Greiner Bio-One). Deviating from the manufacturer’s protocol, the ratio of algal culture sample volume to reagent volume used was 2 : 1, because our preliminary experiments have shown that 50% less reagent could be used without change in assay performance. The background luminescence value of the reagent was determined by adding a cell-free control in every measurement, and the resulting luminescence subtracted from values gathered for wells with algae.

Esterase activity in algal cells was measured using the reporter dye 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA-AM, Invitrogen, Molecular Probes Inc., Eugene, OR, USA), which was stored as 2 mM stock solution in DMSO at -80 °C before use. This assay is based on the conversion of the CFDA-AM to the fluorescent product 5-carboxyfluorescein by cellular esterases. The procedure used was the same as for the assessment of oxidative activity (see below) with the following exceptions: the dye concentration was 2 µM, the concentration of EDTA was 1.5 mM, the incubation time was 30 min, and excitation and emission wavelengths were 492 and 525 nm, respectively.

Oxidative activity was determined using the reporter dye carboxy-2',7'-difluorodihydrofluorescein diacetate (H\textsubscript{2}DFFDA, Invitrogen), which was stored as 2 mM stock solution in DMSO at -80 °C before use. H\textsubscript{2}DFFDA first needs to enter cells where it is hydrolyzed by esterases to carboxy-2',7'-difluorodihydrofluorescein. This intermediary product then reacts with a variety of ROS, including organic hydroperoxides, such as lipid peroxides (Bartosz, 2006; LeBel et al., 1992), forming the oxidation product carboxy-2',7'-difluorofluorescein (DFF). The assay therefore requires a) sufficient transfer of the H\textsubscript{2}DFFDA into cells and b) a cellular environment in which esterases function properly. The applied protocol was based on a procedure previously described (Szivák et al., 2009), because we have also found that addition of EDTA as permeabilization agent is necessary to achieve sufficient fluorescence signal intensity in an acceptable incubation time. Thus, samples taken directly from the experimental cultures were incubated in the presence of 10 µM H\textsubscript{2}DFFDA and 2.5 mM Na\textsubscript{2}EDTA in black 96-well plates for 45 min at room temperature in the dark. Subsequently, the fluorescence of DFF was read with excitation and emission wavelengths of 488 and 520 nm, respectively. The background fluorescence value of the dye was determined by adding a cell-free
control in every measurement, and subtracted from values gathered for wells with algae. Parallel
determination of esterase activity was carried out to be able to exclude that a variation of oxidative activity
was due to impaired esterase activity.

2.3.5 Statistical analysis

Data from independent experiments were pooled for analysis. They are expressed as percentage of the
respective control values and presented as mean ± standard error of the mean. The software used for the
analyses was Prism 4 (Graph Pad, San Diego, CA, USA). Concentration-response curves were fitted
assuming a four parameter log logistic model, i.e. a sigmoidal curve with variable slope, according to
Equation (1):

$$y = y_{\text{min}} + \frac{(y_{\text{max}} - y_{\text{min}})}{1 + \exp(m \times (\text{Log EC50} - x))}$$  (1)

where $y$ = endpoint value, $x$ = herbicide concentration, $y_{\text{min}}$ = bottom, $y_{\text{max}}$ = top, $m$ = Hill slope and
EC50 = herbicide concentration for which the effect on the respective endpoint is 50% between the top and
bottom value. This model is commonly used because it has the advantages of directly delivering EC50 values
with confidence intervals. Furthermore, the flexibility of the top and bottom parameters make this model
particularly suitable for our datasets. For parameters that by definition cannot have more than 100% inhibition, i.e. all endpoints except growth, the bottom parameter was restricted to values greater than zero.

In addition to EC50s, Lowest Observed Effect Concentrations (LOECs) were determined. The general
usefulness of LOEC values for risk assessment may be limited as they strongly depend on the biological and
analytical variability of the underlying endpoint (Laskowsky, 1995). Nevertheless, LOECs provide valuable
supporting information on the sensitivities of the various endpoints towards the exposure regime. In the
context of this study, this feature was of particular importance for concentration-response relationships
where no EC50 values could be calculated. For LOEC calculation ($p < 0.05$), one-way analysis of variance
was performed followed by Dunnett’s post-hoc test. Depending on the shape of the corresponding
concentration-response curve, the LOEC represents either the lowest stimulatory or the lowest inhibitory
concentration.
2.4 Results and discussion

In this study we used physiological and biochemical analyses for exploring subcellular effects in the green algae *C. reinhardtii* exposed to herbicides, with a goal to shed light on the mechanisms underlying the commonly measured inhibition of growth. Our multiple-endpoint assay was based on measurements with relatively quick and simple analytical setups requiring small sample volumes and short measuring times. This, together with the multiwell plate format, allows high throughput and simultaneous determination of a higher number of endpoints compared to standard growth inhibition assays, and this at higher sensitivity. Furthermore, it enables use of extended time and toxicant concentration ranges. Concentration-response relationships are depicted in Figures 2.1-2.3 for paraquat (PQ), diuron (DR) and norflurazon (NF), respectively. All effect values (EC50, LOEC) are summarized in Table 2.1; the corresponding fitted values as well as their CIs are listed in the supplementary material, Tables S2.1-2.3. In several cases no fit could be achieved. In other instances the selected model did not result in an optimal fit, which was characterized by a low $R^2$ or excessively wide confidence intervals of one or more parameters. In many cases, the effect of the highest concentration leveled out at a bottom value significantly higher than 0%, indicating a maximal, yet not complete, inhibition of the respective endpoint. Due to this, in some cases the determined LOEC was higher than the fitted EC50. In a few instances, a stimulation of the endpoint was seen, resulting in both a positive Hill slope and an interchanged position of the top and bottom values.
Multiple-endpoint assay

EC50 values were calculated based on the four parameter logistic model. LOEC values were determined by ANOVA followed by Dunnett’s test.

The effect on the endpoint is expressed as mean percentage of the respective control level.

No EC50 was fitted within the tested concentration range of norflurazon.

not converging: No fit to the four parameter logistic model was achieved for these data.

Hyphens denote cases in which no significant difference to the respective control with p < 0.05 was found over the whole concentration range.

Table 2.1: EC50 and LOEC values of the concentration-response relationships of all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicides paraquat (PQ), diuron (DR) and norflurazon (NF)

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Time point</th>
<th>EC50 PQ (µM)</th>
<th>LOEC PQ (µM)</th>
<th>% of control</th>
<th>EC50 DR (µM)</th>
<th>LOEC DR (µM)</th>
<th>% of control</th>
<th>EC50 NF (µM)</th>
<th>LOEC NF (µM)</th>
<th>% of control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>6h</td>
<td>1.0</td>
<td>3.3</td>
<td>166</td>
<td>0.10</td>
<td>0.1</td>
<td>47</td>
<td>&gt;80</td>
<td>80</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.41</td>
<td>0.33</td>
<td>60</td>
<td>0.17</td>
<td>0.033</td>
<td>116</td>
<td>28</td>
<td>33</td>
<td>75</td>
</tr>
<tr>
<td>Chlorophyll a content</td>
<td>6h</td>
<td>1.2</td>
<td>0.33</td>
<td>87</td>
<td>0.098</td>
<td>0.066</td>
<td>91</td>
<td>nc</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.32</td>
<td>0.33</td>
<td>48</td>
<td>0.19</td>
<td>0.33</td>
<td>57</td>
<td>0.89</td>
<td>0.66</td>
<td>87</td>
</tr>
<tr>
<td>Chlorophyll b content</td>
<td>6h</td>
<td>0.81</td>
<td>0.33</td>
<td>85</td>
<td>0.087</td>
<td>0.066</td>
<td>91</td>
<td>nc</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.3</td>
<td>0.33</td>
<td>45</td>
<td>0.23</td>
<td>0.33</td>
<td>59</td>
<td>0.7</td>
<td>0.66</td>
<td>86</td>
</tr>
<tr>
<td>Carotenoids content</td>
<td>6h</td>
<td>1.2</td>
<td>0.66</td>
<td>86</td>
<td>0.083</td>
<td>0.33</td>
<td>84</td>
<td>0.79</td>
<td>0.66</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.33</td>
<td>0.33</td>
<td>53</td>
<td>0.25</td>
<td>0.33</td>
<td>69</td>
<td>0.72</td>
<td>0.66</td>
<td>75</td>
</tr>
<tr>
<td>Maximum quantum yield</td>
<td>2h</td>
<td>6.0</td>
<td>1</td>
<td>94</td>
<td>0.047</td>
<td>0.01</td>
<td>96</td>
<td>&gt;80</td>
<td>6.6</td>
<td>99</td>
</tr>
<tr>
<td>Fv/Fm</td>
<td>6h</td>
<td>0.85</td>
<td>0.66</td>
<td>71</td>
<td>0.080</td>
<td>0.033</td>
<td>90</td>
<td>18</td>
<td>3.3</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.23</td>
<td>0.33</td>
<td>18</td>
<td>0.97</td>
<td>0.033</td>
<td>92</td>
<td>4.3</td>
<td>3.3</td>
<td>79</td>
</tr>
<tr>
<td>Effective quantum yield</td>
<td>2h</td>
<td>2.8</td>
<td>0.66</td>
<td>95</td>
<td>0.077</td>
<td>0.0033</td>
<td>97</td>
<td>&gt;80</td>
<td>3.3</td>
<td>97</td>
</tr>
<tr>
<td>ΦPSII</td>
<td>6h</td>
<td>0.67</td>
<td>0.33</td>
<td>90</td>
<td>0.11</td>
<td>0.01</td>
<td>93</td>
<td>&gt;80</td>
<td>3.3</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.17</td>
<td>0.1</td>
<td>83</td>
<td>0.12</td>
<td>0.01</td>
<td>94</td>
<td>4.4</td>
<td>3.3</td>
<td>66</td>
</tr>
<tr>
<td>ATP content</td>
<td>2h</td>
<td>2.7</td>
<td>3.3</td>
<td>35</td>
<td>nc</td>
<td>-</td>
<td>-</td>
<td>1.4</td>
<td>1</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.55</td>
<td>0.33</td>
<td>73</td>
<td>0.10</td>
<td>0.33</td>
<td>84</td>
<td>0.57</td>
<td>6.6</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.16</td>
<td>0.1</td>
<td>83</td>
<td>0.31</td>
<td>0.33</td>
<td>80</td>
<td>1.4</td>
<td>1</td>
<td>124</td>
</tr>
<tr>
<td>Esterase activity</td>
<td>2h</td>
<td>6.4</td>
<td>6.6</td>
<td>71</td>
<td>0.082</td>
<td>0.1</td>
<td>80</td>
<td>&gt;80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>1.1</td>
<td>1</td>
<td>64</td>
<td>0.14</td>
<td>0.33</td>
<td>62</td>
<td>&gt;80</td>
<td>80</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.32</td>
<td>0.33</td>
<td>42</td>
<td>0.13</td>
<td>0.01</td>
<td>84</td>
<td>&gt;80</td>
<td>80</td>
<td>68</td>
</tr>
<tr>
<td>Oxidative activity</td>
<td>2h</td>
<td>6.1</td>
<td>3.3</td>
<td>127</td>
<td>0.11</td>
<td>0.00033</td>
<td>86</td>
<td>&gt;80</td>
<td>0.01</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>1.0</td>
<td>0.33</td>
<td>51</td>
<td>0.042</td>
<td>0.0001</td>
<td>87</td>
<td>&gt;80</td>
<td>0.33</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.13</td>
<td>0.01</td>
<td>80</td>
<td>0.2</td>
<td>0.01</td>
<td>71</td>
<td>0.11</td>
<td>0.1</td>
<td>76</td>
</tr>
</tbody>
</table>

*a*EC50 values were calculated based on the four parameter logistic model. LOEC values were determined by ANOVA followed by Dunnett’s test.

*b*The effect on the endpoint is expressed as mean percentage of the respective control level.

*c*No EC50 was fitted within the tested concentration range of norflurazon.

*d*not converging: No fit to the four parameter logistic model was achieved for these data.

*e*Hyphens denote cases in which no significant difference to the respective control with p < 0.05 was found over the whole concentration range.
2.4.1 Paraquat

Upon entering photosynthetic cells, the bipyridinium compound PQ diverts electrons primarily from PS I and transfers them to molecular oxygen. The superoxide radicals produced by this process give rise to subsequent generation of other ROS, notably hydrogen peroxide and hydroxyl radicals (Hess, 2000). Thus, PQ generates oxidative stress which can be assumed to be concentration-dependent and accumulating over time. When a certain threshold has been surpassed, the cellular antioxidant defense mechanisms are no longer able to respond sufficiently. At this point, oxidative stress increases rapidly, leading, among other effects, to membrane disruption and ultimately cell death (Hess, 2000).

Growth of *C. reinhardtii* was inhibited after 24 h at PQ concentrations >0.1 µM (Figure 2.1E) with a mean EC50 value (Table 2.1) being in good agreement with previously reported values of, 0.26 µM after 72 h for *C. reinhardtii* (Jamers and De Coen, 2010) and 0.17 and 0.11 µM after 24 and 48 h, respectively, in *C. moewusii* (Prado et al., 2009). However, after a shorter period of 6 h a stimulated growth was observed at the highest PQ concentrations (Figure 2.1C). Likely explanations for this observation could be the premature release or accelerated production of daughter cells, which could result in a slightly elevated number of particles in the size range of *C. reinhardtii* cells (Supplement Figure S2.1A). However, at the time of measurement, these cultures should be considered dead as indicated by the almost completely inhibited esterase activity at concentrations > 0.66 µM (Figure 2.1C). The latter endpoint exhibited a relatively low sensitivity to PQ, indicating that esterases were deactivated and degraded only after the other cellular parameters had already been strongly affected by PQ. After 24 h, esterase activity correlated with inhibition of growth. A similar correlation with growth was exhibited by the pigment content (Figure 2.1D, F). The concentration-response profiles of chlorophylls and carotenoids implied pigment synthesis to be impeded, already present pigments to be (photo)oxidatively destroyed, or both.

An earlier and stronger effect of PQ exposure was detected for the maximum (Fv/Fm) and effective (ΦPSII) PS II quantum yields (Figure 2.1B, D, F). At all time points, ΦPSII was affected more strongly than Fv/Fm, which is a commonly observed phenomenon (Juneau and Popovic, 1999; Ralph et al., 2007). The ATP content responded sensitively to PQ exposure and correlated well with ΦPSII. Since in the algal cell the chloroplast is both the primary site of ATP synthesis and harboring PS II, the high sensitivity of cellular ATP content and ΦPSII to PQ-induced toxicity is easily explained. Additionally, PQ exposure may cause an elevated metabolic demand for ATP, thus contributing to its depletion. This was previously suggested in a study that also found ATP content of a *Chlamydomonas* species to sensitively respond to PQ, however, after a much longer exposure period of 96 h (Cid et al., 2008).

Relationships between growth and other endpoints were examined. Of particular interest were the aforementioned sensitive responses of ΦPSII and ATP content. Cultures exposed to 0.33 µM PQ showed the first slight but significant inhibition of these two endpoints after 6 h of exposure with no concurrent significant effect on growth (Figure 2.1C, D, Table 2.1). After 24 h, the inhibition of ΦPSII and ATP content was nearly complete, yet growth was inhibited only to about 60% (Figure 2.1E, F). Therefore it can be
concluded that inhibition of ΦPSII and ATP content (although both being crucial for growth) did not directly correlate with growth inhibition, and could serve as an early indicator of PQ exposure.

A stimulation of oxidative activity was expected for PQ as well as for the two other herbicides based on either direct (PQ) or indirect (DR, NF) involvement of oxidative stress in the respective toxicity mechanism (Hess, 2000). However, throughout all our exposure experiments we found oxidative activity to be either not affected or not distinctly reduced relative to the controls, for instance after 24 exposure to PQ (Figure 2.1E). We believe that the explanation for these observations lies in a) the occurring stimulation of the antioxidant defense mechanisms, such as antioxidant enzyme activities, as reported before for green algae exposed to herbicides (Dewez et al., 2005; Geoffroy et al., 2004), and b) the incubation of the cells with the oxidation sensitive reporter dye in the dark. The latter is a common procedure, used because of the known light sensitivity of the respective fluorescent reporter dyes (Jamers et al., 2009; Jamers and De Coen, 2010; Szivák et al., 2009). However, by transferring the algae into the dark, photochemical generation of ROS can be assumed to be largely and abruptly eliminated. Furthermore, an excessive removal of oxidative species by the stimulated antioxidant defense mechanisms might take place in such an environment, thus preventing the reaction of those species with the reporter dye during the incubation period. Then the determined oxidative activity parameter would depend not only on the presence of species with oxidative potential, but also on the state of the antioxidantive defense mechanisms. It therefore should be interpreted as an indicator of disturbances of the redox system. Furthermore, the suppression of oxidative activity might be characteristic of short-term exposures (less than 48 h) or of stress conditions where no substantial accumulation of more stable reactive species such as lipid peroxides has (yet) occurred. In the studies cited above an increase of oxidative activity was commonly observed in PQ and metal exposed C. reinhardtii cultures. However, specific conditions causing a decrease similar to the one observed in our study were also found for Cd (Jamers et al., 2009) and for 0.05 µM PQ at 48 h, the lowest concentration used at the earliest time point assessed in that study (Jamers and De Coen, 2010). The endpoint oxidative activity used in our study thus provided only indirect evidence for the ROS generating activity of PQ in C. reinhardtii. However, it constituted one of the most sensitive markers for PQ exposure as is also reflected in EC50 and LOEC values determined at 24 h (Table 2.1). A particularly dynamic, but also highly reproducible, concentration response was observed at the 6 h time point. At one concentration (0.33 µM, see arrow in Figure 2.1C) the oxidative activity was roughly 50% of the control value and considerably lower than that at the neighboring concentrations of 0.1 and 0.66 µM. Thus, 0.33 µM PQ could be a condition where an especially strong stimulation of the antioxidant defense mechanisms occurred. The increased oxidative activity at the next higher concentration (0.66 µM) coincided with a significant reduction of most other endpoints. This pointed towards overloaded defense mechanisms and the progressive damaging of other cellular systems induced by the oxidative stress.
Figure 2.1: Concentration-response curves obtained in the multi-endpoint assay upon exposure of *Chlamydomonas reinhardtii* to the herbicide paraquat for 2 (A, B), 6 (C, D) and 24 h (E, F). Data points represent the mean ± standard error of the mean (n = 3). Fits of the data, indicated by lines, were obtained based on the four parameter logistic model. $\Phi_{\text{PSII}}$: effective quantum yield; Fv/Fm: maximum quantum yield. The arrow in C points to a concentration of a particular oxidative activity suppression discussed in the text.
These damages were reflected in the fact that the higher concentrations of PQ reduced some within 2, the majority within 6, and all of the endpoints within 24 h to around 0% control level. These results demonstrated the expected lethal effect of this herbicide, which, as we show, may occur even within very short exposure durations. Furthermore, we observed rather steep concentration-response curves (see Hill slopes in Table S2.1) and a very clear time-dependent increase of PQ impact (Figure 2.1, Table 2.1), the latter being in agreement with a previous study which assessed PQ exposure in *C. moewusii* (Prado et al., 2009). Both observations indicated the ability of *C. reinhardtii* to sustain the accumulation of damage until a certain threshold of stress was surpassed, upon which cellular deterioration and death prevailed. In summary, it can be concluded that the mechanism of toxic action of PQ was well reflected in the response of the various endpoints in the green alga.

### 2.4.2 Diuron

DR is a phenylurea herbicide that blocks electron transfer at the Qh binding pocket of the D1 protein located in PS II. This binding leads to inhibition of photosynthesis causing the organism to starve. Furthermore, photooxidative stress is induced leading to chlorosis and lipid peroxidation mediated membrane damage (Hess, 2000).

Growth was completely inhibited by exposure to the highest DR concentrations at 6 and 24 h (Figure 2.2C, E). The calculated mean EC50 values (Table 2.1) were in the range of published values determined for growth endpoints in algae, such as 0.39 µM for *C. reinhardtii* (Reboud, 2002) and 0.050 µM for *Scenedesmus vacuolatus* after 24 h (Neuwoehner et al., 2008). The inhibition of growth was mainly caused by reduced average cell volume and less so by reduced cell density (Figure S2.1C, D). This effect was already detectable after 2 h of exposure, indicating an early impact on cell size. Furthermore, growth showed a significant stimulation at 0.033 µM DR after 24 h of exposure (see arrow in Figure 2.2E, Table 2.1), which we presume to be a hormesis-type effect (Jäger and Krupa, 2009).

The early and strong response of ΦPSII (Figure 2.2B, D, F) reflected the known direct inhibition of PS II performance by DR (Maxwell and Johnson, 2000; Ralph et al., 2007) and the rapid uptake of this herbicide into the cells causing ΦPSII steady state within 1 to 2 min (data not shown). Similar EC50 values determined after 24 h exposure to DR on PSII quantum yield in *S. vacuolatus* and on oxygen production in *S. subspicatus* were 0.073 (Neuwoehner et al., 2008) and 0.056 µM (Nendza and Wenzel, 2006), respectively. The decline of Fv/Fm was most pronounced after 24 h of exposure to DR concentrations higher than 0.1 µM (Figure 2.2B, D, F). At this time point, all other endpoints were characterized by a largely constant maximum inhibition over the highest three or four concentrations (Figure 2.2E, F), making the clear decreasing trend of Fv/Fm over these concentrations a unique response. A decline of Fv/Fm generally suggests stress-induced damages to PS II causing photoinhibition (Maxwell and Johnson, 2000), thus allowing to differentiate these effects from direct PS II inhibition by DR which is effectively reflected in ΦPSII (Ralph et al., 2007). The response of Fv/Fm observed at 2 and 6 h could indicate an early onset of such
DR induced damages to PS II. However, it might also be attributable to a residual actinic effect of the applied measuring light of the chlorophyll fluorometer (Schreiber et al., 2007).

The concentration responses of most other endpoints, among them the pigment content (Figure 2.2D, F), appeared to be linked to that of $\Phi_{\text{PSII}}$. The reduction of the pigment levels after 6 and 24 h of exposure to the higher DR concentrations was probably the result of a smaller average cell volume of the respective cultures (Figure S2.1D) and not of photooxidative destruction. ATP content was only slightly affected by DR, as reflected in the mean EC50 and LOEC values and the incomplete inhibition at the highest concentrations (Figure 2.2A, C, E, Table 2.1). This was unexpected since DR inhibits linear electron transport. However, the ATP concentration might have been maintained by increased mitochondrial activity or other adaptive mechanisms such as cyclic electron transport (Rumeau et al., 2007). The effect on esterase activity was more pronounced than that on ATP content and apparent already after 2 h of exposure. Interestingly, it was particularly strong after 24 h of exposure showing a significant reduction at conditions where growth appeared unaffected or even stimulated (Figure 2.2E). This suggested a specific susceptibility of esterases to DR exposure, possibly caused by downregulation of these enzymes.

Oxidative activity responded quickly and sensitively to DR exposure (Figure 2.2A, C, E), as also reflected in the LOEC values (Table 2.1) for both early time points. This outcome indicates disturbances within the redox system. At concentrations resulting in $\Phi_{\text{PSII}}$ inhibition, a further decrease of oxidative activity relative to the controls was observed (Figure 2.2B, D, F). While photooxidative stress is known to be involved in DR toxicity (Hess, 2000), this observation may be linked to an increased activity of the antioxidative defense mechanisms, as discussed in detail in the PQ section (3.1).

*C. reinhardtii* showed the potential to at least partially recover from intermediate effects induced by DR. This was reflected by the majority of the concentration-response curves somewhat shifting to higher concentrations over time (Figure 2.2, Table 2.1). At 0.1 µM, for example, it was observed that $\Phi_{\text{PSII}}$ recovered from ~40% after 2 h to ~60% control level after 24 h. Recovery may be due to responding cellular detoxification mechanisms such as DR transformation. Alternatively, it could be a “dilution” effect caused by the increase of biovolume over time in relation to the constant amount of the herbicide, leading to reduced intracellular levels of DR. The latter explanation was supported by the observation that recovery appeared to be limited to concentrations resulting in only partial $\Phi_{\text{PSII}}$ and growth inhibition. Overall, we saw that inhibition of $\Phi_{\text{PSII}}$ correlated fairly well with the inhibition of growth (Figure 2.2C, D, E, F). This suggested that the reduction of overall photosynthesis inhibition led to a corresponding growth inhibition, which in turn evidenced the expected DR-induced starvation.
Figure 2.2: Concentration-response curves obtained in the multi-endpoint assay upon exposure of *Chlamydomonas reinhardtii* to the herbicide diuron for 2 (A, B), 6 (C, D) and 24 h (E, F). Data points represent the mean ± standard error of the mean (n = 3). Fits of the data, indicated by lines, were obtained based on the four parameter logistic model. For the 2 h concentration-response of ATP content, no fit could be achieved. $\Phi_{\text{PSII}}$: effective quantum yield; $\text{Fv/Fm}$: maximum quantum yield. The arrow in E points to a concentration where growth was significantly stimulated compared to the control.
In the upper range of DR concentrations (>0.33 µM), $\Phi_{\text{PSII}}$ was inhibited completely at all time points. Growth stopped correspondingly and no recovery occurred (Figure 2.2). However, Fv/Fm, pigment and ATP content, and esterase activity, which are endpoints at least indirectly indicating viability, were never inhibited as severely with respect to controls. This observation is in line with reports for atrazine, a herbicide having the same mechanism of toxic action as DR, for which stronger effects on growth than viability of *Chlamydomonas* were also demonstrated (Fischer et al., 2010; Reboud, 2002). Thus, contrary to PQ, exposure to DR appeared to be non-lethal implying the presence of effective defense mechanisms capable of dealing with the imposed stress levels.

### 2.4.3 Norflurazon

The mechanism of NF toxicity is inhibition of the phytoene desaturase enzyme, which catalyzes a key step in the synthesis of carotenoids (Sandmann et al., 1984). Subsequently, the depletion of carotenoids leads to loss of chlorophylls, resulting in bleaching (Wakabayashi and Böger, 2004), as well as oxidative damage of cellular components (Hess, 2000).

NF had only slight effects on both growth and esterase activity at all measured time points (Figure 2.3A, C, E, Table 2.1), causing significant inhibition of these parameters only at the two highest concentrations. This finding revealed a very limited impact of NF on both growth and viability, thereby corroborating previous findings for exposure of *C. reinhardtii* under comparable conditions, including the EC50 values of 51 and 74 µM obtained for growth endpoints (Fischer et al., 2010).

According to its target site NF caused a selective reduction of the carotenoid pool, which leveled off at approximately 80% and 40% of control level after 6 and 24 h of exposure, respectively (Figure 2.3D, F). The incomplete reduction of carotenoid content was probably attributable to a maximal inhibition of phytoene desaturase and the stability of the remaining carotenoids. Although the carotenoid levels decreased significantly after 6 h, no other endpoint displayed a correlating response, suggesting that the residual amount of these pigments sufficed to maintain the carotenoid-dependent functions at this time point. The further reduction in carotenoid contents at 24 h of exposure reflected the progression of carotenoid synthesis inhibition.

The successive depletion of the carotenoid pool appeared to be linked to effects found for several other endpoints, most notably the levels of chlorophyll a and b (Figure 2.3F), illustrating the bleaching effect of NF. Maximum and effective quantum yields were also reduced between 6 and 24 h (Figure 2.3D, F), indicating photoinhibition and a progressive deterioration of photosynthetic performance (Maxwell and Johnson, 2000). This can be attributed to a then critical lack of carotenoids which are essential components of the photosynthetic apparatus and play a crucial role in photoprotection of the photosystems and prevention of photooxidative stress (Grossman et al., 2004). However, the initial inhibition of $\Phi_{\text{PSII}}$ seen after 2 and 6 h of exposure (Figure 2.3B, D), was not due to a lack of carotenoids. Instead, it was caused by a secondary mechanism of NF action, direct inhibition of PS II similar to DR but characterized by a substantially lower
affinity to the target site (Tischer and Strotmann, 1977). The observed responses of other endpoints should be partially attributed to this. Growth inhibition in particular appeared to be linked to the initial direct PS II inhibition reflected by $\Phi_{\text{PSII}}$ and not to the reduction of carotenoid content.

Surprisingly, we observed a significant stimulatory effect of NF on the ATP level for all time points, which appeared to be inversely correlated to the reduction of carotenoids (Figure 2.3A, C, E), revealing energetic disbalances. Similar effects upon exposure of a diatom and a unicellular red algal species to the PS II inhibiting herbicide bentazon (Hourmant et al., 2009) and to surfactants (Nyberg, 1989) were previously reported. As in those studies, the specific cause for the ATP accumulation remains elusive in our case. Two explanations for the increased ATP levels are conceivable. First, the inhibition of a metabolic pathway such as carotenoid synthesis, that under normal circumstances would consume a significant amount of ATP, may have led to its accumulation, as was proposed previously (Nyberg, 1989). Second, ATP production by mitochondrial or cyclic electron transport (Rumeau et al., 2007) could have been amplified to meet an elevated ATP demand caused by the herbicide induced disturbances.

Oxidative activity responded highly sensitive to NF at all exposure durations as indicated by the LOEC values (Table 2.1), though the reduction never exceeded approx. 50% control level (Figure 2.3A, C, E). As explained in the PQ section, we interpret this behavior as the induction of antioxidative defense mechanisms. The concentration-response relationship at 24 h exhibited a particularly dynamic behavior (Figure 2.3E). First, a maximum inhibition of oxidative activity was found at concentrations coinciding with enhanced reduction of carotenoid contents. With increasing concentration of NF, the trend of oxidative activity reversed. This finding was linked to the maximum reduction of carotenoid levels thus possibly indicating the expected development of (photo)oxidative stress or exhaustion of antioxidative defense mechanisms.

The known primary mechanism of NF toxicity, inhibition of carotenoid synthesis followed by depletion of these pigments, was clearly demonstrated in our assay. However, this effect was preceded by a range of other cellular responses, including increased levels of ATP and a response of the redox system. Still, especially at the early time points, very limited responses of most endpoints were observed, often allowing only for LOEC but not EC50 value determination (Table 2.1). In fact, none of the NF concentrations applied reduced any endpoint to below 20% control level over 24 h of exposure. This led to the overall conclusion that short-term exposure to NF imposed a rather moderate stress on algae and had little impact on growth and viability. However, as indicated by the observed deteriorative trends between 6 and 24 h, *C. reinhardtii* cannot be expected to overcome the long term impact of NF exposure.
Figure 2.3: Concentration-response curves obtained in the multi-endpoint assay upon exposure of *Chlamydomonas reinhardtii* to the herbicide norflurazon for 2 (A, B), 6 (C, D) and 24 h (E, F). Data points represent the mean ± standard error of the mean (n = 3). Fits of the data, indicated by lines, were obtained based on the four parameter logistic model. For the 6 h concentration-response of chlorophyll *a* and *b* content, no fit could be achieved. $\Phi_{PSII}$: effective quantum yield; Fv/Fm: maximum quantum yield.
2.5 Concluding remarks

The multiple-endpoint assay designed and applied in this study provided a detailed concentration- and time-dependent view of the effects of the model herbicides PQ, DR, and NF in C. reinhardtii, ranging from biochemical to physiological endpoints and on to growth inhibition. The herbicide concentrations found to cause significant effects are occasionally found in the environment. (Amondham et al., 2006; Wightwick and Allinson, 2007). The response patterns displayed by the various endpoints over the assessed exposure durations were unique to the respective herbicide. They clearly reflected the different mechanisms of toxic action and thus allowed distinguishing between them. Our work has demonstrated the rapidity of toxicant uptake and induction of subcellular effects shortly after exposure starts. The highly dynamic time-dependent response of many endpoints was revealed. Various physiological and biochemical parameters responded stronger and / or earlier than growth. This confirmed the superiority of assays based on several subcellular endpoints over those relying solely on growth inhibition. The effects on individual parameters may vary strongly depending on the chemical tested and its mechanism of action. Therefore, combination and simultaneous assessment of several endpoints provides important benefits valuable for risk assessment purposes, which are particularly crucial for yet untested chemicals. Among the different endpoints assessed in this work, oxidative activity constituted a rather sensitive indicator for herbicide induced disturbances, even though it was measured as an integrative parameter only partially reflecting the underlying highly complex processes. Other studies have shown that ROS formation and accumulation (Jamers and De Coen, 2010; Knauert and Knauer, 2008; Szivák et al., 2009) as well as enzyme activities of the antioxidant defense (Dewez et al., 2005; Geoffroy et al., 2004) are highly sensitive endpoints in unicellular algae exposed to toxicants. Thus, more detailed insight into the toxicant-induced formation of oxidative species and the induced antioxidative defense could be helpful for characterizing exposures and elucidating toxic mechanisms in the future.

Overall, the presented multiple-endpoint approach improved our understanding of the mechanisms underlying the toxicity of PQ, DR and NF. Further, it could be readily applied to assess the effects on additional cellular pathways affected by other organic toxicants, including the impacts on cell division, disturbance in chlorophyll synthesis, and uncoupling. This assay will provide a solid basis for future investigation of the effects of diverse stressors, alone or in combination, on green algae. From the ecotoxicological risk assessment point of view, application of an array of specific physiological and biochemical endpoints increases the likelihood of detecting disturbances early and sensitively. Therefore, such an approach may be recommended for a wider use in unicellular algae-based assays.

Acknowledgements

The authors wish to acknowledge Bettina Wagner for support in algal cultivation, and Ilona Szivák, Katrin Tanneberger, Marion Junghans and Robert Kase for helpful discussions.
2.6 References


Laskowsky, R., 1995. Some good reasons to ban the use of NOEC, LOEC and related concepts in ecotoxicology. Oikos 73, 140-144.


Supplementary material to chapter 2
Figure S2.1: Concentration-response curves obtained for cell density (A, C, E) and average cell volume (B, D, F) upon exposure of *Chlamydomonas reinhardtii* to the herbicides paraquat (A, B), diuron (C, D) and norflurazon (E, F) after 2, 6 and 24 h. Data points represent the mean ± standard error of the mean (n = 3). Fits of the data, indicated by lines, were obtained based on the four parameter logistic model. In cases where no line is drawn no fit could be achieved.
Table S2.1: Parameters of the concentration-response curve fits based on the four parameter logistic model with their 95% confidence interval (CI) for all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicide paraquat.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Time point</th>
<th>EC50 (µM)</th>
<th>95% CI</th>
<th>Top (%)</th>
<th>Bottom (%)</th>
<th>Hill slope</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>6h</td>
<td>1.0</td>
<td>146.4</td>
<td>[133.1 – 159.8]</td>
<td>110.5</td>
<td>[101.7 – 119.2]</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.41</td>
<td>[0.35 – 0.47]</td>
<td>103.9</td>
<td>[99.6 – 108.1]</td>
<td>-5.6</td>
<td>[-11.3 – -0.1]</td>
</tr>
<tr>
<td>Cell density</td>
<td>2h</td>
<td>9.0</td>
<td>[0.052 – 1585]</td>
<td>129.2</td>
<td>[28.14 – 230.2]</td>
<td>102.2</td>
<td>[100.3 – 104.0]</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.85</td>
<td>[0.50 – 1.5]</td>
<td>116.8</td>
<td>[111.8 – 121.8]</td>
<td>105.2</td>
<td>[101.9 – 108.5]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>1.1</td>
<td>[0.62 – 2.1]</td>
<td>101.5</td>
<td>[98.9 – 104.1]</td>
<td>60.1</td>
<td>[50.7 – 69.6]</td>
</tr>
<tr>
<td>Average cell volume</td>
<td>2h</td>
<td>&gt;10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.38</td>
<td>[0.34 – 0.43]</td>
<td>101.3</td>
<td>[99.0 – 103.5]</td>
<td>50.8</td>
<td>[47.8 – 53.7]</td>
</tr>
<tr>
<td>Chlorophyll a content</td>
<td>6h</td>
<td>1.2</td>
<td>[0.77 – 1.8]</td>
<td>95.4</td>
<td>[92.2 – 98.6]</td>
<td>39.7</td>
<td>[30.2 – 49.1]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.32</td>
<td>[0.26 – 0.35]</td>
<td>102.6</td>
<td>[98.2 – 107.0]</td>
<td>2.4</td>
<td>[0 – 8.0]</td>
</tr>
<tr>
<td>Chlorophyll b content</td>
<td>6h</td>
<td>0.81</td>
<td>[0.54 – 1.2]</td>
<td>94.7</td>
<td>[91.4 – 98.0]</td>
<td>52.5</td>
<td>[45.8 – 59.2]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.3</td>
<td>[0.26 – 0.35]</td>
<td>101.4</td>
<td>[96.7 – 106.1]</td>
<td>2.3</td>
<td>[0 – 8.0]</td>
</tr>
<tr>
<td>Carotenoids content</td>
<td>6h</td>
<td>1.2</td>
<td>[0.87 – 1.6]</td>
<td>96.3</td>
<td>[93.5 – 99.0]</td>
<td>41.2</td>
<td>[34.8 – 47.6]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.33</td>
<td>[0.30 – 0.35]</td>
<td>103.0</td>
<td>[98.9 – 107.1]</td>
<td>5.3</td>
<td>[0.0 – 10.5]</td>
</tr>
<tr>
<td>Maximum quantum yield</td>
<td>2h</td>
<td>6.0</td>
<td>[4.7 – 7.6]</td>
<td>99.1</td>
<td>[98.3 – 99.9]</td>
<td>0.0</td>
<td>[0 – 18.0]</td>
</tr>
<tr>
<td>Fv/Fm</td>
<td>6h</td>
<td>0.85</td>
<td>[0.83 – 0.87]</td>
<td>99.6</td>
<td>[98.7 – 100.5]</td>
<td>0.0</td>
<td>[0 – 1.3]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.23</td>
<td>[0.20 – 0.27]</td>
<td>100.1</td>
<td>[98.0 – 102.2]</td>
<td>0.0</td>
<td>[0 – 2.1]</td>
</tr>
<tr>
<td>Effective quantum yield</td>
<td>2h</td>
<td>2.8</td>
<td>[2.7 – 2.9]</td>
<td>99.0</td>
<td>[98.1 – 99.8]</td>
<td>0.0</td>
<td>[0 – 3.1]</td>
</tr>
<tr>
<td>ΦPSII</td>
<td>6h</td>
<td>0.67</td>
<td>[0.65 – 0.68]</td>
<td>98.9</td>
<td>[98.0 – 99.8]</td>
<td>0.0</td>
<td>[0 – 1.3]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.17</td>
<td>[0.15 – 0.18]</td>
<td>100.7</td>
<td>[99.1 – 102.3]</td>
<td>0.0</td>
<td>[0 – 1.5]</td>
</tr>
<tr>
<td>ATP content</td>
<td>2h</td>
<td>2.7</td>
<td>[2.4 – 3.0]</td>
<td>100.1</td>
<td>[98.0 – 102.2]</td>
<td>0.0</td>
<td>[0 – 6.1]</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.55</td>
<td>[0.50 – 0.61]</td>
<td>99.3</td>
<td>[96.8 – 101.7]</td>
<td>0.0</td>
<td>[0 – 3.6]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.16</td>
<td>[0.14 – 0.19]</td>
<td>102.5</td>
<td>[99.1 – 106.0]</td>
<td>1.3</td>
<td>[0 – 4.5]</td>
</tr>
<tr>
<td>Esterase activity</td>
<td>2h</td>
<td>6.4</td>
<td>103.3</td>
<td>[100.2 – 106.4]</td>
<td>57.8</td>
<td>[0 – 815.8]</td>
<td>-24.1</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>1.1</td>
<td>[1.0 – 1.2]</td>
<td>95.4</td>
<td>[93.0 – 97.7]</td>
<td>3.2</td>
<td>[0 – 6.9]</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.32</td>
<td>[0.12 – 0.86]</td>
<td>99.6</td>
<td>[96.3 – 103.0]</td>
<td>2.1</td>
<td>[0 – 6.2]</td>
</tr>
<tr>
<td>Oxidative activity</td>
<td>2h</td>
<td>6.1</td>
<td>95.3</td>
<td>[89.38 – 101.2]</td>
<td>9.9</td>
<td>[0 – 158.1]</td>
<td>-21.9</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>1.0</td>
<td>82.4</td>
<td>[75.8 – 89.1]</td>
<td>1.7</td>
<td>[0 – 11.9]</td>
<td>-28.7</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.13</td>
<td>[0.080 – 0.20]</td>
<td>90.1</td>
<td>[80.2 – 100.0]</td>
<td>0.0</td>
<td>[0 – 9.5]</td>
</tr>
</tbody>
</table>

aMissing values behind given parameters indicate cases where no CI could be calculated.
bTop and bottom values are expressed as percentage of the respective control level.
cNo EC50 could be fitted within the tested concentration range.
dnot converging: No fit to the four parameter logistic model was achieved for these data.
Table S2.2: Parameters of the concentration-response curve fits based on the four parameter logistic model with their 95% confidence interval (CI) for all endpoints measured in *Chlamydomonas reinhardtii* after 2, 6 and 24 h of exposure to the herbicide diuron

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Time point</th>
<th>EC50 (µM)</th>
<th>95% CI</th>
<th>Top (%)</th>
<th>95% CI</th>
<th>Bottom (%)</th>
<th>95% CI</th>
<th>Hill slope</th>
<th>95% CI</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>6h</td>
<td>0.10</td>
<td>[0.071 – 0.15]</td>
<td>105.0</td>
<td>[94.7 – 115.2]</td>
<td>-1.9</td>
<td>[-15.7 – 11.9]</td>
<td>-2.48</td>
<td>[-4.83 – -0.136]</td>
<td>0.864</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.17</td>
<td>[0.14 – 0.20]</td>
<td>108.9</td>
<td>[104.7 – 113.0]</td>
<td>-8.8</td>
<td>[-15.9 – -1.8]</td>
<td>-2.34</td>
<td>[-2.97 – -1.71]</td>
<td>0.976</td>
</tr>
<tr>
<td>Cell density</td>
<td>2h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell density</td>
<td>6h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.11</td>
<td>[0.086 – 0.14]</td>
<td>105.6</td>
<td>[103.6 – 107.6]</td>
<td>64.5</td>
<td>[61.4 – 67.6]</td>
<td>-1.96</td>
<td>[-2.77 – -1.14]</td>
<td>0.961</td>
</tr>
<tr>
<td>Average cell volume</td>
<td>2h</td>
<td>0.094</td>
<td>[0.062 – 0.14]</td>
<td>100.7</td>
<td>[99.6 – 101.8]</td>
<td>87.8</td>
<td>[86.0 – 89.5]</td>
<td>-1.56</td>
<td>[-2.55 – -0.565]</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.10</td>
<td>[0.075 – 0.14]</td>
<td>102.4</td>
<td>[100.5 – 104.3]</td>
<td>69.2</td>
<td>[65.8 – 72.5]</td>
<td>-1.44</td>
<td>[-2.04 – -0.83]</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.26</td>
<td>[0.22 – 0.32]</td>
<td>100.2</td>
<td>[98.2 – 102.3]</td>
<td>49.2</td>
<td>[45.5 – 52.9]</td>
<td>-3.27</td>
<td>[-5.22 – -1.32]</td>
<td>0.965</td>
</tr>
<tr>
<td>Chlorophyll a content</td>
<td>6h</td>
<td>0.098</td>
<td>[0.065 – 0.15]</td>
<td>105.5</td>
<td>[103.1 – 107.9]</td>
<td>75.0</td>
<td>[71.0 – 79.0]</td>
<td>-1.52</td>
<td>[-2.43 – -0.621]</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.19</td>
<td>[0.14 – 0.25]</td>
<td>96.6</td>
<td>[94.4 – 98.8]</td>
<td>50.7</td>
<td>[47.3 – 54.2]</td>
<td>-3.38</td>
<td>[-4.86 – -1.90]</td>
<td>0.956</td>
</tr>
<tr>
<td>Chlorophyll b content</td>
<td>6h</td>
<td>0.087</td>
<td>[0.059 – 0.13]</td>
<td>105.2</td>
<td>[102.7 – 107.8]</td>
<td>78.7</td>
<td>[75.2 – 82.3]</td>
<td>-1.84</td>
<td>[-3.26 – -0.425]</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.23</td>
<td>[0.15 – 0.33]</td>
<td>96.7</td>
<td>[94.3 – 99.2]</td>
<td>51.4</td>
<td>[47.6 – 55.2]</td>
<td>-4.10</td>
<td>[-7.87 – -0.343]</td>
<td>0.943</td>
</tr>
<tr>
<td>Carotenoids content</td>
<td>6h</td>
<td>0.083</td>
<td>[0.053 – 0.13]</td>
<td>105.2</td>
<td>[102.6 – 107.7]</td>
<td>82.6</td>
<td>[79.1 – 86.1]</td>
<td>-1.91</td>
<td>[-3.67 – -0.142]</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.25</td>
<td>[0.158 – 1.1]</td>
<td>97.3</td>
<td>[95.3 – 99.2]</td>
<td>63.2</td>
<td>[60.1 – 66.4]</td>
<td>-6.27</td>
<td>[-39.1 – 26.5]</td>
<td>0.933</td>
</tr>
<tr>
<td>Maximum quantum yield</td>
<td>2h</td>
<td>0.047</td>
<td>[0.041 – 0.054]</td>
<td>99.6</td>
<td>[98.6 – 100.5]</td>
<td>68.4</td>
<td>[67.3 – 69.5]</td>
<td>-1.35</td>
<td>[-1.60 – -1.10]</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.080</td>
<td>[0.062 – 0.10]</td>
<td>99.6</td>
<td>[98.1 – 101.2]</td>
<td>63.1</td>
<td>[60.2 – 65.9]</td>
<td>-1.08</td>
<td>[-1.38 – -0.779]</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.97</td>
<td>[0.60 – 1.6]</td>
<td>101.0</td>
<td>[99.8 – 102.2]</td>
<td>5.4</td>
<td>[0 – 19.5]</td>
<td>-0.656</td>
<td>[-0.746 – -0.565]</td>
<td>0.995</td>
</tr>
<tr>
<td>Effective quantum yield</td>
<td>2h</td>
<td>0.077</td>
<td>[0.073 – 0.082]</td>
<td>98.8</td>
<td>[97.6 – 100.1]</td>
<td>0.0</td>
<td>[0 – 1.9]</td>
<td>-1.36</td>
<td>[-1.48 – -1.24]</td>
<td>0.998</td>
</tr>
<tr>
<td>ΦPSII</td>
<td>6h</td>
<td>0.011</td>
<td>[0.098 – 0.11]</td>
<td>98.3</td>
<td>[96.9 – 99.7]</td>
<td>0.0</td>
<td>[0 – 2.4]</td>
<td>-1.47</td>
<td>[-1.63 – -1.32]</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.12</td>
<td>[0.11 – 0.14]</td>
<td>99.2</td>
<td>[97.3 – 101.0]</td>
<td>0.0</td>
<td>[0 – 3.4]</td>
<td>-1.50</td>
<td>[-1.70 – -1.30]</td>
<td>0.994</td>
</tr>
<tr>
<td>ATP content</td>
<td>2h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.10</td>
<td>[0.061 – 0.17]</td>
<td>102.5</td>
<td>[99.9 – 105.0]</td>
<td>86.1</td>
<td>[82.9 – 89.3]</td>
<td>-3.03</td>
<td>[-8.34 – -2.29]</td>
<td>0.706</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.31</td>
<td>[0.014 – 6.9]</td>
<td>101.8</td>
<td>[98.8 – 104.7]</td>
<td>68.9</td>
<td>[62.9 – 74.9]</td>
<td>-8.84</td>
<td>[-360 – 343]</td>
<td>0.824</td>
</tr>
<tr>
<td>Esterase activity</td>
<td>2h</td>
<td>0.082</td>
<td>[0.046 – 0.15]</td>
<td>100.9</td>
<td>[96.3 – 105.6]</td>
<td>67.3</td>
<td>[60.8 – 73.9]</td>
<td>-1.71</td>
<td>[-3.54 – -0.119]</td>
<td>0.768</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.14</td>
<td>[0.088 – 0.22]</td>
<td>104.1</td>
<td>[100.0 – 108.2]</td>
<td>58.1</td>
<td>[51.5 – 64.7]</td>
<td>-2.36</td>
<td>[-4.19 – -0.534]</td>
<td>0.869</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.13</td>
<td>[0.083 – 0.22]</td>
<td>91.8</td>
<td>[87.3 – 96.3]</td>
<td>23.6</td>
<td>[13.1 – 34.1]</td>
<td>-1.12</td>
<td>[-1.61 – -0.621]</td>
<td>0.930</td>
</tr>
<tr>
<td>Oxidative activity</td>
<td>2h</td>
<td>0.11</td>
<td>[0.064 – 0.17]</td>
<td>84.6</td>
<td>[81.0 – 88.2]</td>
<td>55.4</td>
<td>[50.4 – 60.5]</td>
<td>-2.37</td>
<td>[-5.26 – 0.514]</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.042</td>
<td>[0.021 – 0.084]</td>
<td>85.9</td>
<td>[81.2 – 90.6]</td>
<td>49.0</td>
<td>[42.6 – 55.4]</td>
<td>-0.923</td>
<td>[-1.54 – -0.312]</td>
<td>0.868</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.20</td>
<td>[0.090 – 0.45]</td>
<td>73.9</td>
<td>[66.8 – 80.9]</td>
<td>27.4</td>
<td>[15.3 – 39.5]</td>
<td>-2.57</td>
<td>[-5.61 – 0.469]</td>
<td>0.882</td>
</tr>
</tbody>
</table>

*a* Missing values behind given parameters indicate cases where no CI could be calculated.

*b* Top and bottom values are expressed as percentage of the respective control level.

*c* not converging: No fit to the four parameter logistic model was achieved for these data.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Time point</th>
<th>EC50 (µM)</th>
<th>95% CI</th>
<th>Top (%)</th>
<th>95% CI Bottom (%)</th>
<th>95% CI</th>
<th>Hill slope</th>
<th>95% CI</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>6h</td>
<td>&gt;80c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>28</td>
<td>[1.3 – 606]</td>
<td>103.1</td>
<td>[96.6 – 109.7]</td>
<td>40.7</td>
<td>[-44.8 – 126.2]</td>
<td>-1.09</td>
<td>0.599</td>
</tr>
<tr>
<td>Cell density</td>
<td>2h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>6.6</td>
<td>[3.1 – 14]</td>
<td>101.3</td>
<td>[98.7 – 103.9]</td>
<td>83.9</td>
<td>[77.8 – 90.0]</td>
<td>-1.68</td>
<td>0.636</td>
</tr>
<tr>
<td>Average cell volume</td>
<td>2h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a content</td>
<td>6h</td>
<td>0.89</td>
<td>[0.80 – 1.0]</td>
<td>96.6</td>
<td>[96.8 – 100.4]</td>
<td>60.8</td>
<td>[59.0 – 62.5]</td>
<td>-3.16</td>
<td>0.970</td>
</tr>
<tr>
<td>Chlorophyll b content</td>
<td>6h</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.7</td>
<td>[0.59 – 0.84]</td>
<td>98.5</td>
<td>[96.0 – 100.9]</td>
<td>71.4</td>
<td>[69.0 – 73.7]</td>
<td>-4.11</td>
<td>0.897</td>
</tr>
<tr>
<td>Carotenoids content</td>
<td>6h</td>
<td>0.79</td>
<td>[0.66 – 0.94]</td>
<td>100.5</td>
<td>[98.7 – 102.3]</td>
<td>79.9</td>
<td>[78.2 – 81.6]</td>
<td>-3.74</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.72</td>
<td>[0.67 – 0.78]</td>
<td>99.6</td>
<td>[97.6 – 101.5]</td>
<td>44.4</td>
<td>[42.5 – 46.2]</td>
<td>-3.38</td>
<td>0.984</td>
</tr>
<tr>
<td>Maximum quantum yield</td>
<td>2h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fv/Fm</td>
<td>6h</td>
<td>18</td>
<td>[6.4 – 50]</td>
<td>100.3</td>
<td>[100.0 – 100.6]</td>
<td>90.8</td>
<td>[87.6 – 94.0]</td>
<td>-0.744</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>4.3</td>
<td>[3.8 – 4.8]</td>
<td>100.9</td>
<td>[99.8 – 102.0]</td>
<td>56.1</td>
<td>[54.0 – 58.2]</td>
<td>-1.56</td>
<td>0.988</td>
</tr>
<tr>
<td>Effective quantum yield</td>
<td>2h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΦPSII</td>
<td>6h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>4.4</td>
<td>[3.6 – 5.3]</td>
<td>101.2</td>
<td>[98.8 – 103.5]</td>
<td>30.8</td>
<td>[26.0 – 35.7]</td>
<td>-1.30</td>
<td>0.980</td>
</tr>
<tr>
<td>ATP content</td>
<td>2h</td>
<td>1.4</td>
<td>[0.50 – 3.7]</td>
<td>121.6</td>
<td>[115.8 – 127.5]</td>
<td>99.1</td>
<td>[95.3 – 102.9]</td>
<td>0.961</td>
<td>0.758</td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>0.57</td>
<td>[0.032 – 10]</td>
<td>128.6</td>
<td>[104.9 – 152.3]</td>
<td>91.9</td>
<td>[69.0 – 114.7]</td>
<td>0.487</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>1.4</td>
<td>[0.77 – 2.5]</td>
<td>148.1</td>
<td>[140.9 – 155.4]</td>
<td>102.8</td>
<td>[97.2 – 108.3]</td>
<td>1.55</td>
<td>0.832</td>
</tr>
<tr>
<td>Esterase activity</td>
<td>2h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative activity</td>
<td>2h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6h</td>
<td>&gt;80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24h</td>
<td>0.11</td>
<td>[0.022 – 0.55]</td>
<td>82.3</td>
<td>[72.5 – 92.2]</td>
<td>67.7</td>
<td>[61.8 – 73.7]</td>
<td>-3.11</td>
<td>0.204</td>
</tr>
</tbody>
</table>

*Missing values behind given parameters indicate cases where no CI could be calculated.
*Top and bottom values are expressed as percentage of the respective control level.
*No EC50 could be fitted within the tested concentration range.
*Not converging: No fit to the four parameter logistic model was achieved for these data.
Chapter 3:

Establishment of a proteomics pipeline for differential proteome analysis using MudPIT, OMSSA, spectral counts and G-test statistics
3.1 Introduction

The proteome constitutes a central level of biological organization determining the physiological state of an organism. The advent of mass spectrometry-based analytical tools allowed high throughput analysis of not only individual proteins but whole proteomes in a multitude of research settings. The profiling of proteomes can be accomplished using a variety of methods (Nesatyy and Suter, 2008). Initially, gel-based methods such as 2D gel electrophoresis were dominating proteomics, using mass spectrometry (MS) for the identification of resolved proteins. Non-gel based methods using 1- or 2-dimensional liquid chromatography (LC) of digested peptide mixtures coupled to tandem mass spectrometry (MS/MS) are more recent developments. While both approaches can be used for differential proteomics, i.e. detecting differences in protein expression between different samples, non-gel based techniques usually provide a better coverage of the proteome.

For the differential proteomics study on the green alga Chlamydomonas reinhardtii exposed to herbicides presented in chapter 4, a method called multidimensional protein identification technology (MudPIT) (Delahunty and Yates, 2005; Washburn et al., 2001) was used. With this procedure, the peptide mixture resulting from the trypsin digest of a full protein extract is loaded onto a polyphasic column containing stacked strong cation exchanger (SCX) and C18 phases. By running increasingly concentrated salt pulses stepwise through this setup, the cationic peptides are released from the SCX (1st dimension), and then separated according to their molecular weight and lipophilicity (2nd dimension). Typically, the duration of one step of this two-dimensional separation is two hours. The peptides eluting from the tip of the column are then desorbed into the gas phase by electrospray and transferred to the mass spectrometer where they are detected in full scan mode. Based on the acquired full scan spectra, subsequent MS/MS experiments are triggered. Once fragment ions of an intense precursor ion have been determined, they will be excluded from further selection for a specified exclusion duration, which in the present case was 60 s. This scan-dependent analysis, in combination with an exclusion list, allows for lower intensity ions to be analyzed by MS/MS. A typical 2-D analysis is composed of 11 steps, with the salt step going from 0 to 100% 500 mM NH₄Ac, and lasts 22 hours. Over that time roughly 150,000 mass spectra are acquired. Thus, MudPIT, same as all other proteome profiling methods, produces large amounts of MS data that need to be analyzed. The 150,000 spectra acquired correspond to roughly 50,000 identified peptides that code for a few thousand proteins. To achieve sufficient proteome coverage at least three replicates have to be acquired (Nesatyy et al., 2009). For the analysis of such huge datasets, a dedicated analysis pipeline with multiple steps executed sequentially had to be developed. The pipeline which was built using the utility “AutoHotkey” (AHK) for scripting (see section 3.2), other freely available software tools, as well as Xcalibur (Thermo Fisher Scientific, Bremen, Germany) is as follows (Figure 3.1):
Figure 3.1: Simplified workflow of the proteomics analysis pipeline outlined in this chapter leading to identification of significantly differently expressed proteins based on spectral counting quantitation. The RAW format MS/MS data of each MudPIT run are treated independently until experimental replicates are merged to the final datasets required for statistical analysis of differences between cases and control via \( G \)-test. The AutoHotkey (AHK) script numbers refer to the scripts given in the Appendix in which the respective steps and operations are implemented. FDR: false discovery rate
First, the MS/MS data (filtered according to number of sequence ions in a given mass range) are extracted from the raw files and stored as so-called peak lists, which are text files with standardized format (section 3.3). The Open Mass Spectrometry Search Algorithm (OMSSA) then matches the measured MS/MS information of each precursor and potential peptide with theoretical spectra produced in silico from a prepared FASTA format database (sections 3.4 and 3.5). This database contains all protein sequences of interest for a given species. It has been complemented by commonly found contaminants, and a decoy database (reversed) of the same size, for determining false discovery rates (FDR). The results of this first data analysis is then used to determine and correct the systematic mass error of precursor mass determination, observed on the LTQ-Orbitrap XL (section 3.6). After mass error correction in the peak list files, the search algorithm is re-run on the corrected data with an appropriately reduced mass window, typically 0.005 Dalton (Da). This search provides the list of identified peptides that is subsequently used for generating spectral counts data, i.e. the number of reliable and discrete peptide identifications for a specific protein (section 3.7). Following this, data sets obtained for different exposure conditions are evaluated statistically by pairwise comparisons using G-test statistics, which allows to determine significant protein expression changes (section 3.8). Finally, these results can be used for evaluation in a biological context (section 3.9 and chapter 4).

3.2 AutoHotkey scripting for automated data analysis

Performing all the steps of the analysis pipeline mentioned above manually is highly impractical and tedious. High throughput data analysis is the only way to treat the enormous amount of data acquired with MudPIT. Scripting allows to process and manipulate the data so that it can be passed from one module of the proteomics pipeline to the next. This chapter describes how this can be achieved with the freely available open source scripting utility AHK (http://www.autohotkey.com/). AHK was developed by Chris Mallett and Steve Gray and runs under Windows. Its main purpose is the automation of repetitive tasks, the creation of macros on the operating system level, and the definition of hotkeys. It further allows simulating keystrokes and mouse clicks, defining variables, using loops for iterations, working with text files and executing programs. Other programming or scripting tools exist which are capable of performing some or all of the tasks required for the analysis pipeline such as Python and PERL. AHK’s main advantage is that its syntax and handling are relatively easy to learn. Furthermore, scripts can be created and modified using simple text editors such as Windows notepad. Comprehensive tutorials, details on commands and syntax, as well as numerous example scripts are available online (http://www.autohotkey.com/). AHK is not so well established in the scientific community as other scripting languages mentioned above, especially PERL in proteomics applications. Still, several publications demonstrate the capabilities of AHK, such as automation of mass spectrometer control (Henderson et al., 2009; McQuinn et al., 2008), data collection and storage from medical scanners (Kucewicz et al., 2008), for converting tab delimited data files into files suitable for curve
focusing (Fedulova and Mannervik, 2011), for automating satellite image-registration (Gill et al., 2010), or viewing of radiological images (Lee, 2012).

A selection of AHK scripts important for the presented analysis pipeline are described in some detail in the following sections. The full scripts are provided in the appendix to this chapter. They intentionally have a linear structure, making them more easy to comprehend, use and modify. The only instance where it was necessary to divert from this rule was the implementation of the recursive quicksort algorithm (adapted from http://rosettacode.org/), which was needed to be able to efficiently sort tens of thousands of identified peptides according to their E-value (see section 7 of this chapter). This was done by integrating quicksort as a function, i.e. a subordinate piece of code that is called by the main script.

3.3 Conversion of raw MS data to DTA and MGF format

As mentioned before, a MudPIT analysis is a combination of two separation principles, one based on interaction with an SCX, the other with a lipophilic C18 material. One entire MudPIT analysis is composed of 11 analytical steps with a run time of 2 h each, adding up to several gigabytes of raw data. This data needs to be extracted to make it accessible for proteomics database search algorithms. In the presented pipeline, the mass spectrometric information (m/z and intensity) is extracted from the RAW files and written to one DTA text file per fragment spectrum using the command line program extract_msn.exe (Thermo Fisher Scientific). The following command line was used in this work:

```
extract_msn.exe -G 1 -S 0 -I 15 -B 500 -T 4000 InputFile.raw
```

where

- **-G** defines the number of related scans that need to be grouped for DTA file creation
- **-S** the number of intermediate scans that are allowed for grouping of scans
- **-I** the number of ions that need to be present in an MS/MS spectrum for DTA file creation
- **-B** lowest (bottom) molecular weight allowed
- **-T** highest (top) molecular weight allowed

-15, -B500 and -T4000 are arguments which remove precursor ions and their corresponding fragment spectra that are expected to be useless for the later analysis. -G1 and -S0 prevent extract_msn.exe from merging scans which is unwanted for spectral counts analysis.

Generally, programs being executed via command line, such as extract_msn, are particularly suited for execution and automation with AHK because this can be achieved with the simple “Run” or “RunWait” commands. “RunWait” (script 1) requires the external program to close or exit before going to the next step in the script. In the present work extract_msn was integrated into a file-loop which retrieves files having a RAW extension, one at a time, from a specified folder and its subfolders. In every loop iteration the name of the currently retrieved RAW file together with its location is stored in the variable “A_LoopFileFullPath”. Once peak listing is done on a single RAW file, which typically takes about 3 min, thousands of DTA files have been created in the folder containing the original file. These are then merged to the standardized mascot
Chapter 3

generic format (MGF) using the PERL script merge.pl (Matrix Science, Boston, MA, USA). This script is run when extract_msn.exe has finished. The AHK script line “FileMove, %A_LoopFileDir%merge.mgf, %A_LoopFileDir%\%Name%.mgf” renames the created MGF file to the name of the respective RAW file and thus prevents overwriting when the loop continues with the next source file. Throughout the analysis pipeline, the “Sleep” command was used in places where it seemed necessary to provide a time buffer for the operating system to complete a time-consuming operation, such as deleting all the current DTA files, before the script proceeds. Since MGF is a text format, the files generated from all MudPIT steps can be merged into a single file using AHK. This simplifies the subsequent database search with OMSSA. However, it was found that the searchable file size must stay below a total size of 1 Gigabyte, possibly due to limitations of the OMSSA executable.

3.4 Database adaptation and search using the Open Mass Spectrometry Search Algorithm (OMSSA)

A number of different database search algorithms were developed for matching MS/MS data with sequence database information. Two major types of algorithms can be distinguished. Heuristic algorithms calculate the score of a match based on the similarity between a theoretical MS/MS spectrum and an experimentally acquired one. Examples for these are SEQUEST and X!Tandem. Probabilistic algorithms, such as MASCOT and OMSSA, evaluate the probability of producing a peptide sequence based on an experimental spectrum by chance by modeling the process of peptide fragmentation (Kapp and Schütz, 2007).

Geer et al. published a detailed description of OMSSA and its functionality (Geer et al., 2004). OMSSA can be downloaded free of charge from http://pubchem.ncbi.nlm.nih.gov/omssa/. The quality of the OMSSA output in terms of specificity and sensitivity was shown to be as good as or better than that of other search algorithms, including the commercially available SEQUEST and MASCOT (Bakalarski et al., 2007; Balgley et al., 2007; Geer et al., 2004). In order to identify peptides acquired during a MudPIT run, a FASTA format protein sequence database is searched for matches with the experimental MS/MS spectra and their corresponding precursor ions. For this, each protein in the database is digested and all tryptic peptides are fragmented in silico, generating theoretical fragment ion spectra of b- and y-ions typically observed for collision-induced dissociation as used in this study. These are then compared to the experimental spectra. The quality of the matches (the peptide hit confidences) is expressed by a score that OMSSA reports as E-value, which is a measure of the probability of the hit being identified by chance (Geer et al., 2004).

The Chlamydomonas JGI (Joint Genome Institute) protein sequence database version 3.1 containing ~14,600 nuclear encoded proteins (“frozen gene catalog”) (Merchant et al., 2007) was used in this work. It required some adjustment to optimally work with the OMSSA executable. This included the removal of occasionally appearing asterisks from the peptide sequence, as well as the transformation of the sequence entry headers to conform to NCBI standard. This was performed using the file-reading loop of AHK which retrieves specified text line by line and thus, in the case of FASTA sequence databases, makes it possible to
extract the protein ID from all original headers. The new header lines were written by placing the ID where the NCBI format locates the accession number, and by filling the other NCBI format components with placeholders. The newly created FASTA database was further supplemented with the proteins encoded in chloroplast and mitochondria (downloaded from the ChlamyCyc homepage provided by the University of Potsdam and MPI of Molecular Plant Physiology, Potsdam, Germany), and expected contaminants, that is, digestion enzymes, keratins, and BSA (downloaded from the NCBI, http://www.ncbi.nlm.nih.gov/protein). The accession numbers of the contaminants were marked as “Contaminant_” to make hits for these proteins easily recognizable in the data evaluation steps.

The final step in sequence database creation takes into account an issue that is critical for all database search algorithms - the occurrence of false discoveries. False positive peptide identifications are random and occur due to the huge numbers of MS/MS spectra and potential target peptides in the database. This problem is usually addressed by estimating FDRs, the rate of false discoveries found in the total number of identifications. One method of estimating the FDR is the target-decoy strategy (Elias and Gygi, 2010). With this method, the raw data is searched against a database that, in addition to the biologically meaningful proteins, contains the same number of proteins but having a reversed sequence, thus retaining the original length and amino acid composition of each protein. This so called “decoy” database then allows to reliably estimate the rate of false discoveries produced by searching the “target” database. When using this strategy, the resulting list of peptide identifications contained a certain number of such “reversed” sequences with frequency increasing towards the lower end of identification confidence, i.e. higher OMSSA E-values. It is thus common practice to define an allowed threshold FDR above whose respective E-value all peptide hits are excluded from further analysis (Elias and Gygi, 2010). For the work presented, 0.5% FDR was selected, which means that 0.5% of all peptide hits were reversed sequences, indicating a similar number of false discoveries among the hits for the biologically relevant original database. Though the same could be done on the protein level, only FDRs for peptides were considered, because peptide spectral counts were used for quantitation. Finally, the complete curated FASTA sequence database has to be converted into a format that can be accessed by OMSSA, using the NCBI supplied software tool formatdb.exe.

The omssacl executable which performs the OMSSA database search is also a command line program. Information on the variety of input parameters (arguments) available to set up and modify the search, such as handling of charge states, missed cleavages, or inclusion of post-translational modifications (PTM), can be accessed by calling the help information of the executable in a command prompt (“omssacl.exe -help”), or can be obtained directly from the OMSSA site. For automated searching of the data generated by the various MudPIT runs, omssacl.exe was integrated into an AHK script using a file-loop which retrieves all MGF files and provides them as input for the OMSSA command line as shown by script 2. The search on one MGF file comprising a full MudPIT run took about 7 min. Since the used computer was a dual-core machine, its CPUs were completely occupied by two parallel OMSSA searches. To prevent the script from starting more than two searches at the same time (which would likely lead to main memory overflow), any second search was executed using the already mentioned “RunWait” command.
Script 2 represents the first OMSSA search of the pipeline. The reason for there being three searches in total on the same data is the need for correcting the systematic precursor mass error, so that the precursor ion mass tolerance can be significantly reduced (see section 3.6). The OMSSA command line ultimately run to produce the output data files used for the spectral counting was

```
omssaclexe -to 0.3 -te 0.005 -mf 3 -mv 10,110 -v 3 -zcc 1 -zt 2 -he 1E-01 -d LibraryName -fm InputFile.mgf -oc OutputFile.csv
```

where

- `-to` corresponds to the product ion mass tolerance in Da
- `-te` to the precursor ion mass tolerance in Da
- `-mf` lists the fixed modifications
- `-mv` the variable modifications (10 being N-terminal acetylation, 110 pyro-glu from N-terminal Q)
- `-v` defines the number of missed cleavages allowed
- `-zcc` defines that the precursor charge is determined based on the input file
- `-zt` minimum precursor charge to start considering multiply charged products
- `-he` the maximum E-value allowed in the hit list
- `-d` sequence library to search
- `-fm` name of mfg formatted input file
- `-oc` filename for comma separated value (CSV) formatted search results.

The “-o” argument determines the search output format by the letter that succeeds it, in this case “c”, which leads to CSV type files, a simple list presenting only rudimentary information for the peptide identifications. “-op” results in the pepXML output which is a highly detailed XML type text file output. A third output format, occasionally generated in this work, OMS, can be produced using “-ob”. The latter files can, if necessary, be merged using omssamerge.exe (included in the OMSSA package) via command line and opened with the OMSSA browser (http://pubchem.ncbi.nlm.nih.gov/omssa/browser.htm) for evaluation.

3.5 Optimization of OMSSA search

The goal of analyzing experimental MS/MS data by OMSSA was to identify as many peptides as possible, while stabilizing FDR. Hence, the number of identified peptide species was viewed as the best parameter for evaluating database search outcome quality, since it is a measure of the achieved proteome coverage. Therefore, using a protein sequence database which covers as much of the potential proteome as possible is crucial. Also, OMSSA search arguments may be varied to maximize the number of peptide hits. For that all potentially relevant search arguments were tested and, if found to improve the result significantly, included in the command line of the OMSSA search. In the following, two selected search arguments (“-mv” and “-v”) are discussed in view of their effect on the search result, using a representative MudPIT run as an example.
3.5.1 Search argument “-mv”

A highly important group of command line arguments defines the handling of post-translational modifications (PTMs) by OMSSA. PTMs are protein modifications that are biologically induced, for example for signal transduction or to modify protein function and stability. These include, among others, phosphorylation, whose appearance in the *C. reinhardtii* proteome was analyzed in detail previously (Wagner et al., 2006), hydroxylations e.g. on proline 151 of the large chain of RuBisCO (RbcL) in *C. reinhardtii* (Taylor et al., 2001), or the common acetylation of the protein N-terminus (Mann and Jensen, 2003). In addition, other modifications are willingly or unwillingly caused by sample preparation. Among these are the carboxyamidomethylation of cysteine residues induced by the iodoacetamide treatment during sample workup (the only fixed modification “-mf 3” in the command line), carbamylations of peptide N-termini and lysine residues caused by the urea buffer, or formation of pyro-glutamic acid from exposed N-terminal glutamine residues (Hunyadi-Gulyás and Medzihradsky, 2004). All of these modifications cause a specific shift of peptide precursor and fragment masses. These modified peptides are identified by OMSSA, if the corresponding mass shifts are considered during database search. In order to determine the importance of the various variable modifications, which in contrast to the fixed modifications are not expected to appear in all cases, MudPIT runs from different experimental exposure conditions were screened. This screening was done based on the “mods.xml” file included in the OMSSA package, in which more than 150 different modifications and their corresponding mass shifts are defined. The great number of OMSSA searches needed to analyze the experimental data acquired for the different herbicides and exposure conditions made automation by scripting necessary. Specifically, a loop was applied in which the variable modifications (defined by the “-mv” argument) were tested one after the other. No treatment-specific changes in the number of modifications could be observed. It was found, however, that two variable modifications, acetylation of the protein N-terminus and formation of pyro-glutamic acid on glutamines at the peptide N-terminus, increased the total number of identified peptide species and thus protein coverage (Figure 3.2A). Henceforward, “-mv 10,110” was included in the standard search command line. Phosphorylation is one of the biologically most relevant modifications and thus of special interest. Due to the low degree of phosphorylation in the proteome and the properties of the phosphate group, special enrichment and MS analysis techniques usually need to be applied (Boersema et al., 2009). Even so, a diversity of phosphorylated peptides were found while screening the MudPIT data. This included possible novel ones, such as a phosphorylated peptide from plastidic terminal oxidase (PTOX) which appeared in the samples of high concentration paraquat exposure. Most found peptides, however, have been reported previously (Turkina et al., 2006; Wagner et al., 2006). This demonstrated that MudPIT can be used to find initial evidence for phosphorylation sites on abundant proteins. However, the inclusion of many PTMs (such as phosphorylations) in the OMSSA standard search can also negatively affect the search results, or cause doubtful identifications. For example, the more additional modifications are considered, the higher the FDR tends to be and the lower the peptide identification confidence reported by OMSSA will be. The latter is true for basically all input parameters that make OMSSA less strict. Identifications of PTMs can also be
ambiguous, particularly when they share mass shifts. This is for example the case for various oxidative modifications such as carbonylation versus hydroxylation.

Figure 3.2: Effect of allowed variable modifications (panel A) and of different numbers of allowed missed cleavages (panel B) during tryptic digestion on the resulting number of total peptide identifications (white bars) and identified peptide species (black bars) in a representative MudPIT run. Only peptides and peptide species that were discrete hits for *Chlamydomonas reinhardtii* proteins at 0.5% false discovery rate were counted. OMSSA modification numbers: 10 - acetylation of the N-terminus, 110 - formation of pyro-glutamic acid on glutamines at the N-terminus.
### 3.5.2 Search argument “-v”

During treatment of the extracted proteins with proteases such as trypsin, a complete digestion is rarely achieved, if ever. As a result, peptides appear in MS/MS data sets which contain arginines or lysines at positions other than the C-terminus. Often these are peptides where these amino acids are in very close proximity to each other, such as EVTLGFVDLMRDDYVEKDR of RbcL, or LMPDKKDFGYSFPCDGPG of the PS I subunit PsaB. The allowed number of those “missed cleavages” which OMSSA considers are defined by the “-v” argument. Figure 3.2B shows the impact of allowing zero to three missed cleavages on the unique as well as total peptide hits. An increasing number of total and unique peptides was observed, reaching a plateau at three. Hence, three missed cleavages were used in the following searches.

### 3.6 Correction of the systematic error of precursor mass determination by the orbitrap

One of the most important variables for the confidence of all peptide identifications in OMSSA is the deviation between the theoretical and measured precursor mass (Geer et al., 2004). In this study, the mass of the precursor ion was determined in the orbitrap of the LTQ-Orbitrap XL used, which provides a very high mass accuracy, typically below 0.005 Da. In comparison, fragment ion spectra were acquired with the linear ion trap (LTQ) of this instrument, which is much less accurate. Accordingly, for the final database search, the cutoff for the maximally allowed mass deviation was set to 0.005 Da for the precursor ion (“-te”) and 0.3 Da for the product ions (“-to”), respectively. However, when subjecting the experimental data to OMSSA database searches, many peptide hits, particularly those with higher masses, did not meet that strict mass deviation criterion due to the occurrence of systematic mass errors and would thus have been discarded during database search. These mass errors are caused by the drift of the orbitrap calibration with time, which will eventually require recalibration of the instrument or possibly the use of lock masses, for instance of ubiquitous contaminants. The much more stable LTQ analyzer is generally unaffected. Thus, only the precursor ion mass and not fragment ion mass had to be corrected. Occurrence and correction of systematic mass errors, e.g. by offline-recalibration, is a well-known issue in mass spectrometry. A recalibration procedure related to the one outlined in this section, based on “internal standards” in the proteomics data set, i.e. peptide identifications of high confidence and contaminants, was proposed by Zubarev and Mann (2007). In this work the theoretical mass of all identified peptides was used to determine and correct the systematic experimental mass error in the data stored in the MGF files. The first step in this procedure was performing an OMSSA search which allowed a greater mass deviation for the precursor ion, e.g. 0.05 Da, to accept peptides even if they deviate significantly from the theoretical mass. The deviation between theoretical and measured peptide mass is reported in the OMSSA pepXML output delivered by script 2.
As pepXML is a structured text format containing key terms in its lines, it can easily be processed by AHK (see script 3). This is done by using a file-reading loop which checks for key terms of interest line by line using the function “IfInString”. Furthermore, only peptide hits with highest confidence are used for this procedure by including “hit_rank=1” as additional selection criterion. Note that false discoveries or hits beyond a certain FDR threshold were not eliminated, since this would have required the implementation of additional sorting and filtering. The effect of keeping these hits in the mass error correction, however, was negligible. When a line passes the filter criteria, it is processed further using the “StringSplit” function. That function creates an array of new variables containing the strings that before were separated by the user defined delimiter, here quotation marks, in the pepXML output. The respective variables in these arrays that contain the measured peptide mass \( M \), the charge state of its original precursor ion \( z \) and the mass deviation between theoretical and measured mass \( \Delta M \) are then used to calculate the \( m/z \) originally measured in the orbitrap \( m/z = M/z + H^+ \) as well as the deviation from the mass of the matched theoretical peptide in the orbitrap \( \Delta M/z \). Plotting \( \Delta M \) as a function of \( M \) observed in a MudPIT run produces a set of data points with linear correlation (see typical example in Figure 3.3A). The variation around the regression line (Figure 3.3A) represents the random error of the precursor mass measurement which cannot be corrected. The parameters of the linear correlation \( y = ax + b \) (where \( y = \Delta M/z \) and \( x = m/z \)), slope \( a \) and intercept \( b \) as well as the correlation coefficient \( r \), were determined with AHK by implementation of simple linear regression.

The script then continues with the actual correction of the precursor mass data in every MGF file. For this again a file-reading loop is used, which recognizes and retrieves every line containing the measured precursor masses and then corrects them using the regression parameters \( a \) and \( b \). The lines with the corrected \( m/z \) are written into a new MGF file with the appendix “_Corrected” using the command “FileAppend”. All other lines were written into this file unchanged. A second OMSSA search with XML output was then performed on those corrected files, using the wide mass window as in the first search, to check the effect of the mass error correction. The data points now are centered around the x-axis as shown in Figure 3.3B, the average mass deviation thus being very close to 0 Da, reducing the parameters of the regression line, \( a \) and \( b \), by orders of magnitude. The effect can also be visualized using histograms (Figure 3.4) where, after the correction, the greatest number of peptide hits should be sorted into the 0.000 Da bin with the remaining hits distributed evenly around it.

In conclusion, despite the occurrence of systematic mass errors during data acquisition, the high mass accuracy of the orbitrap MS can now be exploited by applying a much stricter cutoff for the precursor mass deviation (i.e. 0.005 Da). A final OMSSA search on the corrected data then produces the CSV type output to be used for the next step of the pipeline, the spectral counting. Using the corrected MGF data set for OMSSA search with the stricter cutoff retained most of the original identifications found with wider mass window. However, the E-values of peptide hits generally shifted towards lower values, thus indicating a general increase in significance of the results due to the stricter search criterion. This leads to an increased number of
Figure 3.3: Effect of the elimination of the systematic mass error produced by the orbitrap in MudPIT data. Deviations between theoretical and measured mass of peptides identified by OMSSA are plotted as function of the corresponding measured mass. The points correspond to the peptide hits found with a representative MudPIT run. Panel A shows the situation before and panel B after systematic mass error correction as described in section 3.6. Hits from the target database and from the reverse sequence database are plotted differently colored. The values for the parameters of the linear regression lines best fitting the plotted data are provided in the boxes.
Figure 3.4: Effect of the elimination of the systematic mass error produced by the orbitrap in MudPIT data. The histograms show the number of peptides identified by OMSSA sorted into bins representing their mass deviation. The bin width is 0.001 Da mass deviation. The data are derived from all peptide hits observed for a representative MudPIT run. Panel A shows the situation before and panel B after systematic mass error correction as described in section 3.6.
peptide hits, because OMSSA also filters according to E-values (“-he”). Furthermore, false discoveries have a tendency for a relatively high mass error of the precursor ion, therefore appearing in areas more distant from the central regression line as shown in Figure 3.3. Thus, performing recalibration allows these scattered false discoveries to be eliminated by the stricter cutoff of 0.005 thus further improving the OMSSA database search results.

### 3.7 Spectral counts analysis

Different approaches have been developed for the label-free determination of protein expression levels based on MS/MS data. One method is the usage of spectral counts, that is the sum of spectra of all identified tryptic peptides of a given protein (Nesatyy and Suter, 2008). These counts have been shown to correlate well with the amounts of proteins in complex samples (Liu et al., 2004; Old et al., 2005). In the context of spectral counting applications OMSSA was recommended due to its peptide identification capabilities (Balgley et al., 2007). An OMSSA search on a complete MudPIT run results in the identification of roughly 50,000 tryptic peptides generated from *Chlamydomonas reinhardtii*, but also from contaminants such as keratins and trypsin, and includes false discoveries from the reversed database. To prepare these results for spectral counting and to perform the counting itself, a number of steps needed to be implemented in the analysis pipeline (script 4), some of which require considerable computing power and time to complete. Script 4 contains an overarching file-loop which can be used for batch-processing of a sequence of files generated by OMSSA using the corrected MGF files. The rather long script is divided into sections, each of which is introduced by a header line, a non-executable comment identified as such by a semicolon. Also, due to its length, only a rough outline of the script will be provided, focusing on details relevant for the final result, the filtered spectral count values of all proteins identified in each MudPIT run.

In sections 1 and 2 of the script the CSV file is rewritten into a consistent format to simplify further processing. Section 3 of the script marks discrete and ambiguous hits in the CSV output. A discrete hit is a peptide match which can only be produced by one protein in the FASTA database. Ambiguous (non-discrete) hits, on the other hand, are peptide hits for a given fragment spectrum with identical E-value that could originate from several proteins. These hits were ultimately excluded from the spectral counts and statistical analysis. Furthermore, for spectra resulting in multiple hits with varying confidence only the hit(s) with lowest E-value were considered. This removes many detected false discoveries which, as already mentioned, tend to appear as hits for the same spectra as true positive hits, albeit with lower confidence. Section 4 was added to correct for the very rare cases in which a reversed sequence peptide had the same E-value as a true hit. Section 3 of the script would then falsely label otherwise discrete true positives as ambiguous. This may be caused by a reversed sequence peptide having the same sequence as one from the real database (Elias and Gygi, 2010) which was, for instance, true for a peptide originating from the PsbA protein.

Section 5 of the script then sorts the data by E-value using an adaptation of the recursive quicksort algorithm. For detailed information on this algorithm the reader is referred to common programming sources.
Given at the end of the script (section 11), it is first called by the main body of the program (going through the current peptide hit list as input variable), and consecutively calls itself as often as needed until complete sorting has been achieved. To automatically obtain a general assessment of the MudPIT data at this point, the total number of peptides and peptide species of \textit{C. reinhardtii} proteins identified, as well as the numbers of contaminant and reversed peptide hits found are determined in section 6. All these values are subsequently written into a tab delimited “SampleReport” file for manual inspection.

In section 7 the current E-value sorted results are read from top to bottom. For every peptide match the reversed sequences found up to this point are counted. Based on this, the respective FDRs are calculated and added to each line in the new CSV file. The script also determines the last peptide match below an FDR of 0.5%. All peptide matches with E-values lower than that of this peptide are then discarded in section 8, thus ensuring that the remaining true positives have an estimated FDR of <0.5%. All contaminant and reversed peptide matches are removed from the list. Following that, section 9 separates the discrete and ambiguous peptide matches and saves them in two different files.

These steps prepared the now filtered list of discrete \textit{C. reinhardtii} peptide hits for the actual spectral counting for every identified protein (section 10). This section furthermore creates a summarizing view of the results and adds the following information to the spectral counts of every identified protein: the number of tryptic peptide species, how often each of these was identified, the peptides’ sequence coverage over the theoretical protein sequence and, finally, the sum of the logarithm of the E-values. The latter is a number which indicates the overall confidence of all peptide hits of a protein and thus is a measure of the confidence of the protein identification itself. The next step of the analytical pipeline was to add the spectral counts from three replicate MudPIT runs. The resulting lists could then be used for the statistical analysis.

### 3.8 Implementation of \(G\)-test statistics and fold change calculation

For the study presented in chapter 4 a statistical method was required that allows comparison of expression levels of proteins, in our case represented by spectral counts. Generalized \(G\)-test statistics were used for this, which allows comparing counts of specific elements (proteins) in different total populations (sum of all proteins). It was initially evaluated for application in spectral counting by Zhang et al. (2006) in combination with multiple-testing adjustment by the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995). This study provided all necessary formulas for implementing the procedure in the presented work. Generally, \(G\)-test statistics can be used to find significant differences between more than two experimental conditions (Zhang et al., 2006). In this pipeline, however, only pairwise comparisons are done.

For \(G\)-test analysis by script 5, the data has to be organized as a tab delimited text file containing a spectral counts data table compiled from the two experimental conditions to be compared (e.g. “control” and “case”). In cases where a protein identifier appeared in only one of the two conditions, the count for this protein in the other condition was set to 0. Furthermore, all count pairs which had less than 2 counts in each of the two conditions were removed from the table to eliminate the so-called “one-hit-wonders” whose
identification is generally considered not trustworthy. In the first part of script 5 the identifiers as well as their corresponding spectral count pairs \((x_1, x_2)\) are saved in array-like variables. Subsequently, the further necessary values are calculated for each identifier and saved in an output text file, the final state of which is shown in Table 3.1: \(x (= x_1 + x_2)\), \(n_i\) (sums \(n_1\) and \(n_2\) of all counts \(x_i\) for all identifiers in each of the conditions 1 and 2), \(n (= n_1 + n_2)\), \(y (= n - x)\), and \(y_i (= n_i - x_i)\). The determination of the \(G\) values for all count pairs follows next. Since logarithms are involved in the equation

\[
G = 2 \times \left( \sum_{i=1}^{m} x_i \ln x_i + \sum_{i=1}^{m} y_i \ln y_i - \sum_{i=1}^{m} n_i \ln n_i - x \ln x - y \ln y + n \ln n \right)
\]

the script needs to deal with cases where one of the two counts is 0. In these instances \(0 \times \ln(0)\) is defined as 0 (B. Zhang, personal communication). \(m\) is the number of conditions to be compared, in our case two. Furthermore, the William’s correction \((w)\) is calculated according to

\[
w = 1 + \left( \frac{\sum_{i=1}^{m} \frac{n}{n_i} - 1}{\sum_{i=1}^{m} \frac{1}{y_i}} \right) \left( \frac{n + n - 1}{x + y} \right) \frac{6 n (m - 1)}{m}
\]

The previously determined \(G\) value is then corrected to \(G_{adj} = G/w\) to adjust for higher false positive rates caused by the approximation that the distribution of the \(G\)-statistic follows a \(\text{Chi}^2\) distribution (Zhang et al., 2006). The \(G_{adj}\) values are used to obtain corresponding \(p\)-values based on the assumed \(\text{Chi}^2\) distribution. For this, the R variant Rscript.exe was used (http://www.R-project.org), which requires a text input file containing all needed commands. These commands can be found in the section of the script that is not executed, i.e. marked by “/*” and “*/”. After the calculation, R writes a new data file with a column of \(p\)-values added, which indicates the significance of the observed count differences. This list is also sorted according to \(p\)-value and can thus be used directly for the Benjamini-Hochberg correction. Briefly, for every protein identifier, the \(p\)-value is compared to its predecessor and, when higher, the rank \((r)\) of that count comparison is increased by 1. Using the rank \(r\) of a comparison, the total number of comparisons \((c_{tot})\), and the desired FDR (0.05), the value \(BH\) is calculated according to \(BH = 0.05 \times r / c_{tot}\). Afterwards, \(BH\) is compared to the original \(p\)-value. As long as this value is higher or equal than the \(p\)-value, the respective count difference is considered significant. Both, \(BH\) and a “yes” or “no” indicating the significance are added to the results table.

The final section of the script adds two more columns: the fold change and the log 2 fold change between case and control. For this calculation the respective protein counts are normalized to their respective sum of all counts from a condition. Furthermore, counts of 0 in any condition are replaced with 1 to avoid dividing 0 and dividing by 0. Since all one-hit-wonders were removed prior to this analysis, the fold change contains a systematic error which, however, is negligible due to the very small number of these counts within the total count populations \(n_1\) and \(n_2\).
Table 3.1: Lines from a final output file with relevant parameters\(^1\) created during application of G-test statistics with multiple-testing adjustment by Benjamini-Hochberg (BH) correction to find significant differences between the counts \(x_i\) of each protein (ID) found in two experimental conditions. The shown lines represent the development of the G-test-statistics over the list of IDs from top (highest significance) to bottom (lowest significance). The central block contains the cutoff between significant and non-significant differences according to BH correction.

<table>
<thead>
<tr>
<th>ID</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x)</th>
<th>(y_1)</th>
<th>(y_2)</th>
<th>(y)</th>
<th>(n_1)</th>
<th>(n_2)</th>
<th>(n)</th>
<th>(G)</th>
<th>(w)</th>
<th>(G_{adj})</th>
<th>(p)-value</th>
<th>BH</th>
<th>Significant</th>
<th>FC</th>
<th>Log 2 FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>108283</td>
<td>107</td>
<td>14</td>
<td>121</td>
<td>130029</td>
<td>127306</td>
<td>257335</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>79.051912</td>
<td>1.004133</td>
<td>78.726544</td>
<td>0.000000</td>
<td>1.665E-05</td>
<td>yes</td>
<td>0.13</td>
<td>-2.94</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>148386</td>
<td>13</td>
<td>32</td>
<td>45</td>
<td>130123</td>
<td>127288</td>
<td>257411</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>8.701916</td>
<td>1.011113</td>
<td>8.606275</td>
<td>0.003350</td>
<td>0.003463</td>
<td>yes</td>
<td>2.52</td>
<td>1.33</td>
</tr>
<tr>
<td>137163</td>
<td>345</td>
<td>266</td>
<td>611</td>
<td>129791</td>
<td>127054</td>
<td>256845</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>8.608280</td>
<td>1.000818</td>
<td>8.601240</td>
<td>0.003359</td>
<td>0.003480</td>
<td>yes</td>
<td>0.79</td>
<td>-0.34</td>
</tr>
<tr>
<td>195634</td>
<td>8</td>
<td>24</td>
<td>32</td>
<td>130128</td>
<td>127296</td>
<td>257424</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>8.726861</td>
<td>1.015627</td>
<td>8.592580</td>
<td>0.003375</td>
<td>0.003497</td>
<td>yes</td>
<td>3.07</td>
<td>1.62</td>
</tr>
<tr>
<td>131102</td>
<td>25</td>
<td>49</td>
<td>74</td>
<td>130111</td>
<td>127271</td>
<td>257382</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>8.462610</td>
<td>1.006758</td>
<td>8.405805</td>
<td>0.003740</td>
<td>0.003513</td>
<td>no</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>101282</td>
<td>1</td>
<td>10</td>
<td>11</td>
<td>130135</td>
<td>127310</td>
<td>257445</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>8.745771</td>
<td>1.045462</td>
<td>8.365462</td>
<td>0.003824</td>
<td>0.003530</td>
<td>no</td>
<td>10.22</td>
<td>3.35</td>
</tr>
<tr>
<td>172202</td>
<td>1</td>
<td>10</td>
<td>11</td>
<td>130135</td>
<td>127310</td>
<td>257445</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>8.745771</td>
<td>1.045462</td>
<td>8.365462</td>
<td>0.003824</td>
<td>0.003530</td>
<td>no</td>
<td>10.22</td>
<td>3.35</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78954</td>
<td>103</td>
<td>115</td>
<td>218</td>
<td>130033</td>
<td>127205</td>
<td>257238</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>0.950288</td>
<td>1.002294</td>
<td>0.948113</td>
<td>0.330200</td>
<td>0.013020</td>
<td>no</td>
<td>1.14</td>
<td>0.19</td>
</tr>
<tr>
<td>149843</td>
<td>65</td>
<td>75</td>
<td>140</td>
<td>130071</td>
<td>127245</td>
<td>257316</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>0.950926</td>
<td>1.003572</td>
<td>0.947541</td>
<td>0.330346</td>
<td>0.013037</td>
<td>no</td>
<td>1.18</td>
<td>0.24</td>
</tr>
<tr>
<td>132530</td>
<td>57</td>
<td>46</td>
<td>103</td>
<td>130079</td>
<td>127274</td>
<td>257353</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>0.949062</td>
<td>1.004855</td>
<td>0.944767</td>
<td>0.331129</td>
<td>0.013054</td>
<td>no</td>
<td>0.82</td>
<td>-0.29</td>
</tr>
<tr>
<td>194541</td>
<td>157</td>
<td>171</td>
<td>328</td>
<td>129979</td>
<td>127149</td>
<td>257128</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>0.944459</td>
<td>1.001525</td>
<td>0.940312</td>
<td>0.331502</td>
<td>0.013070</td>
<td>no</td>
<td>1.11</td>
<td>0.15</td>
</tr>
<tr>
<td>189615</td>
<td>67</td>
<td>55</td>
<td>122</td>
<td>130069</td>
<td>127265</td>
<td>257334</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>0.934760</td>
<td>1.004099</td>
<td>0.930944</td>
<td>0.334618</td>
<td>0.013087</td>
<td>no</td>
<td>0.84</td>
<td>-0.25</td>
</tr>
<tr>
<td>156470</td>
<td>20</td>
<td>14</td>
<td>34</td>
<td>130116</td>
<td>127206</td>
<td>257422</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>0.937321</td>
<td>1.014708</td>
<td>0.923734</td>
<td>0.336496</td>
<td>0.013104</td>
<td>no</td>
<td>0.72</td>
<td>-0.47</td>
</tr>
<tr>
<td>Rps18</td>
<td>94</td>
<td>92</td>
<td>186</td>
<td>130042</td>
<td>127205</td>
<td>257270</td>
<td>130136</td>
<td>1327320</td>
<td>257456</td>
<td>6.38E-06</td>
<td>1.002689</td>
<td>0.000006</td>
<td>0.997988</td>
<td>0.01961372</td>
<td>no</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^1\) \(x_i\): counts of all peptides for one ID in condition \(i\), \(x = x_1 + x_2\), \(y_i = n_1 - x_i\), \(y = n - x\), \(n_i\): sum of all counts \(x_i\) of all IDs in condition \(i\), \(n = n_1 + n_2\); \(G\): values of the G-test statistic, \(w\): values of the William’s correction, \(G_{adj}\): G-values adjusted according to \(G_{adj} = G/w\); \(p\)-value: \(p\)-values corresponding to \(G_{adj}\) based on an assumed Chi\(^2\) distribution, BH: calculated values of the Benjamini-Hochberg correction, FC: fold change
3.9 Data analysis in biological context

The proteomics profiling data for which this pipeline was developed finally resulted in 149-254 proteins significantly differentially expressed depending on exposure condition. Analyzing such large lists protein by protein is generally rather tedious. Therefore, software tools have been developed by many research groups and organizations which allow the efficient extraction of biologically relevant information. These are, for example, mapping programs that put gene / transcript / protein identifiers into the context of their biological functions, or software platforms which allow enrichment analysis thus indicating functional terms over-represented in a submitted list of identifiers. The basis for such analyses is the functional annotation of the genome of interest, which provides prediction of the functions of genes and proteins based on their sequences. Examples for annotation databases are the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and MapMan which was specifically designed for plants (Thimm et al., 2004; Usadel et al., 2009). In the present work the lists of identifiers of significantly differently expressed proteins were subjected to a functional term enrichment analysis provided by the Algal Functional Annotation Tool (Lopez et al., 2011). Since *C. reinhardtii* is a photosynthetic organism the respective MapMan annotation (May et al., 2008) was selected as basis.

The detailed discussion of these and other results of the proteome profiling and differential proteomic analysis are presented in the following chapter of this thesis.
3.10 References


Proteomics analysis pipeline


3.11 Appendix

Script 1. Peak listing from all RAW type files in a selected folder and its subfolders using extract_msn.exe and merging to MGF type files using merge.pl

FileSelectFolder, Input, E:\, Select folder containing the subfolders containing the raw files to process
Loop, %Input%*.RAW,0,1
   
   SetWorkingDir, %A_LoopFileDir%
   RunWait c:\xcalibur\system\programs\extract_msn.exe -G1 -S0 -I15 -B500 -T4000 %A_LoopFileFullPath%
   Sleep 2000
   RunWait C:\Perl\merge.pl,, Min
   Sleep 2000
   FileDelete %A_LoopFileDir%*.dta
   StringTrimRight, Name, A_LoopFileName, 4
   FileMove, %A_LoopFileDir%/merge.mgf, %A_LoopFileDir%\%Name%.mgf
   FileDelete %A_LoopFileDir%*.txt
   Sleep 2000

Script 2. Performing OMSSA database search on all MGF files containing MS/MS data of complete MudPIT runs

FileSelectFolder, Input, E:\, Select folder containing the subfolders containing the SUM mgf files

Count = 0

Loop, %Input%*_SUM.mgf,0,1
   
   ++Count
   SplitPath, A_LoopFileFullPath, OutFileName, OutDir,, OutNameNoExt,
   StringSplit, OutDir, OutDir, \n   Name := OutDir%OutDir0%
   
   If (Count = 1)
      Run, C:\OMSSA\omssacl.exe -to 0.3 -te 0.05 -mf 3 -mv 10`,110 -v 3 -zcc 1 -zt 2 -he 1E-01 -d SequenceDatabase -
      fm %A_LoopFileFullPath% -op %OutDir%\%Name%_SUMXML.xml, C:\OMSSA\,Min
   
   If (Count = 2)
      
      {RunWait, C:\OMSSA\omssacl.exe -to 0.3 -te 0.05 -mf 3 -mv 10`,110 -v 3 -zcc 1 -zt 2 -he 1E-01 -d SequenceDatabase -
      fm %A_LoopFileFullPath% -op %OutDir\%Name%_SUMXML.xml, C:\OMSSA\,Min
      Count = 0
      }
Chapter 3

Script 3. Correction of the systematic mass error of precursor masses in MGF type MS/MS data located in subfolders each representing one sample. The parameters of the linear correlation \( y = ax + b \), slope \( a \), intercept \( b \) and correlation coefficient \( r \) are calculated according to the following equations implemented in the script:

\[
a = \frac{n \sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n \sum_{i=1}^{n} x^2 - \left( \sum_{i=1}^{n} x \right)^2},
\]

\[
b = \frac{\sum_{i=1}^{n} y - a \sum_{i=1}^{n} x}{n}
\]

and

\[
r = \frac{n \sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{\sqrt{\left( n \sum_{i=1}^{n} x^2 - \left( \sum_{i=1}^{n} x \right)^2 \right) \left( n \sum_{i=1}^{n} y^2 - \left( \sum_{i=1}^{n} y \right)^2 \right)}}
\]
Proteomics analysis pipeline

SetBatchLines, -1
SetFormat, Float, 0.10

FileSelectFolder, Input, E:\, Select folder to process,

Loop, %Input%\*_SUM.xml,0,1
{
    Sum_x = 0
    Sum_y = 0
    Sum_xy = 0
    Sum_x2 = 0
    Sum_y2 = 0
    n = 0

    Loop, Read, %A_LoopFilePath%
    {
        IfInString, A_LoopReadLine,<spectrum_query spectrum
        {
            StringSplit, Part, A_LoopReadLine,""
            PCMass := Part8
            CS:= Part10
        }
        IfInString, A_LoopReadLine,<search_hit hit_rank="1"
        {
            ++n
            x_%n% := PCMass / CS + 1.007277
            y_%n% := Part20 / CS
        }
    }

    Loop, %n%
    {
        Sum_x := Sum_x + x_%A_Index%
        Sum_y := Sum_y + y_%A_Index%
        Sum_xy := Sum_xy + x_%A_Index%*y_%A_Index%
        Sum_x2 := Sum_x2 + x_%A_Index%*x_%A_Index%
        Sum_y2 := Sum_y2 + y_%A_Index%*y_%A_Index%
    }

    a := (n*Sum_xy - Sum_x*Sum_y) / (n*Sum_x2 - Sum_x*Sum_x)
    b := (Sum_y - a*Sum_x) / n
    r := (n*Sum_xy - Sum_x*Sum_y) / sqrt((n*Sum_x2-Sum_x*Sum_x) * (n*Sum_y2-Sum_y*Sum_y))
    r2 := r*r

    FileAppend, r^2 = %r2%`n a = %a%`nb = %b%`n# of peptides = %n%`n%A_LoopFilePath%RegressionReport.csv

    Loop, %A_LoopFilePath%\*.mgf,0,0
    {
        StringTrimRight, MGFName, A_LoopFilePath, 4
        Loop, Read, %A_LoopFilePath%, %MGFName%_Corrected.mgf
        {
            IfInString, A_LoopReadLine, PEPMASS=
            {
                CalculatedMass := MeasuredMass - a * MeasuredMass - b
                FileAppend, PEPMASS=%CalculatedMass%`n
            }
        }
    }
}

Else
{
    FileAppend, %A_LoopReadLine%`n
}
}
Script 4. Generation of spectral counts data from results of individual MudPIT runs. The script executes an auxiliary script (SpectralCountsFromCSVOutput_Thread2.ahk) at the beginning of section 3 which contains the same commands as that section. The two scripts work in parallel on the halves of the current peptide list file to speed up the data processing.

SetBatchLines, -1
StringCaseSense, Off

FileSelectFolder, StartFolder, E:\,, Select folder to process the sum csv files in its subfolders

Loop, %StartFolder%\*SUMCSV.csv, 0, 1
{
  Input := A_LoopFileFullPath
  SplitPath, Input, OutFileName, OutDir, OutExtension, OutNameNoExt
  IfExist, Input.txt
  FileDelete, Input.txt
  FileAppend, %OutDir%, Input.txt

  ; ***Section 1: Remove header***
  LC = 0
  Loop, Read, %Input%, %OutDir%\Step_1.csv
  {
    IfInString, A_LoopReadLine, dta
    {
      FileAppend, %A_LoopReadLine%`n
      ++LC
    }
  }

  ; *** Section 2: Split E-value, removal of several columns and splitting of result file***
  Stop := Round(LC/2)
  Switch = 0
  File1 =
  File2 =

  Loop, Read, %OutDir%\Step_1.csv
  {
    StringSplit, Part, A_LoopReadLine,`
    If (Part4 = 0)
    Part4 = 1.0e-310
    IfInString, Part4, e-
    StringSplit, NumberPart, Part4,e
    Else
    {
      Count = 0
      While-loop Part4 < 1
      {
        Part4 := Part4 * 10
        ++Count
      }
      NumberPart1 := Part4
      NumberPart2 := -1 * Count
    }
    If (Part11 = "")
    Part11 = NoMods

    If (Part2 <> LastDTA AND A_Index >= Stop)
    Switch = 1
  }
If (Switch = 0)
File1
   = %File1%\%Part1%\,\%Part2\,\%Part3\,\%NumberPart1\,\%NumberPart2\,\%Part6\,\%Part7\,\%Part10\`
   ,\%Part11\,\%Part8\,\%Part9\n
Else
File2
   = %File2%\%Part1%\,\%Part2\,\%Part3\,\%NumberPart1\,\%NumberPart2\,\%Part6\,\%Part7\,\%Part10`
   ,\%Part11\,\%Part8\,\%Part9\n
LastDTA := Part2
}

StringReplace, File1, File1,\,,1
StringReplace, File2, File2,\,,1

FileAppend, %File1%, %OutDir%\Step_2_1.csv
FileAppend, %File2%, %OutDir%\Step_2_2.csv

FileAppend, %File1 File2, %OutDir%\Step_2.csv

;***Section 3: Removal of ambiguous peptide hits with lower E-Value and decision about discrete or not discrete status of
the remaining hits***

Run, SpectralCountsFromCSVOutput_Thread2.ahk
AC = 0

Loop, Read, %OutDir%\Step_2_1.csv
{
   StringSplit, Part_%A_Index%_, A_LoopReadLine,`,
   Line_%A_Index% := A_LoopReadLine
   ++AC
}

DnD = Discrete
FinalList =

Loop, %AC%
{
   ToolTip, %Part_%A_Index%_2, 10, 10
   B_Index := A_Index + 1
   If (Part_%A_Index%_2 = Skip)
      Continue

   If (Part_%A_Index%_2 = Part_%B_Index%_2)
      { If (Part_%A_Index%_4 * 10**Part_%A_Index%_5 < Part_%B_Index%_4 * 10**Part_%B_Index%_5)
         { FinalList = %FinalList Line_%A_Index% `," DnD " n"
            DnD = Discrete
            Skip := Part_%A_Index%_2
         }
      } Else
         { FinalList = %FinalList Line_%A_Index% `,"NotDiscrete\n"
            DnD = NotDiscrete
         }
   } Else
      { FinalList = %FinalList Line_%A_Index% `," DnD " n"
         DnD = Discrete
      }
   }

FileAppend, %FinalList%, %OutDir%\Step_3_1.csv
ToolTip,, 10, 10
Found = 0

While Found = 0
{
  IfExist, %OutDir%/Step_3_2.csv
  Found = 1
  Sleep 2000
}

FileRead, File1, %OutDir%/Step_3_1.csv
FileRead, File2, %OutDir%/Step_3_2.csv

FileAppend, % File1 File2, %OutDir%/Step_3.csv

;***Section 4: Elimination of reversed sequences ambiguously identified with hits to the real database and according change of discrete status of those true hits***

Discretes =
NonDiscretes =

Loop, Read, %OutDir%/Step_3.csv,
{
  IfInString, A_LoopReadLine, NotDiscrete
  NonDiscretes = %NonDiscretes%%A_LoopReadLine%`n
  Else
  Discretes = %Discretes%%A_LoopReadLine%`n
}

FileAppend, %Discretes%, %OutDir%/Step_3_D.csv
FileAppend, %NonDiscretes%, %OutDir%/Step_3_ND.csv

SC = 0
Part_0_2 = X

Loop, Read, %OutDir%/Step_3_ND.csv
{
  StringSplit, Part_%A_Index%_, A_LoopReadLine,`
  Before := A_Index - 1

  If (Part_%A_Index%_2 <> Part_%Before%_2)
  {
    ++SC
    Spectrum_%SC% = %A_LoopReadLine%`n
  }
  Else
    Spectrum_%SC% = % Spectrum_%SC% A_LoopReadLine "`n"
}

ReturnList =

Loop, %SC%
{
  RegExReplace(Spectrum_%A_Index%, \.dta, "", ReplacementCount1)
  RegExReplace(Spectrum_%A_Index%, "Reverse sequence", "", ReplacementCount2)
  TotalNumber_%A_Index% := ReplacementCount1
  ReverseNumber_%A_Index% := ReplacementCount2
}
Loop, %SC%
{
If (TotalNumber_%A_Index% = ReverseNumber_%A_Index%)
ReturnList = % ReturnList Spectrum_%A_Index% "n"
Else
If (ReverseNumber_%A_Index% = 0)
ReturnList = % ReturnList Spectrum_%A_Index% "n"
Else
{
TotalNumber := TotalNumber_%A_Index%
ReverseNumber := ReverseNumber_%A_Index%
Loop, Parse, Spectrum_%A_Index%, ´n
{
IfNotInString, A_LoopField, Reverse sequence
{
If (TotalNumber - ReverseNumber = 1)
{
StringReplace, Corrected, A_LoopField, NotDiscrete, Discrete
ReturnList = %ReturnList%Corrected% n
}
Else
ReturnList = %ReturnList%A_LoopField% n
}
}
}

FileAppend, %ReturnList%, %OutDir%ReturnList1.csv
Loop, Read, %OutDir%ReturnList1.csv, %OutDir%Step_3_D.csv
If (A_LoopReadLine <> "")
FileAppend, %A_LoopReadLine% n
FileMove, %OutDir%Step_3_D.csv, %OutDir%Step_3a.csv,1

FileDelete, %OutDir%Step_3_D.csv
FileDelete, %OutDir%Step_3_ND.csv
FileDelete, %OutDir%ReturnList1.csv

;***Section 5: Sorting of peptides with quicksort according to E-value***

FileRead, Content, %OutDir%Step_3a.csv
StringReplace, Content, Content, n,";",";",1

ToolTip, Sorting in progress..., 10, 10
Sorted := quicksort(Content)
ToolTip,, 10, 10
StringReplace, Sorted, Sorted,",",";",1
StringReplace, Sorted, Sorted,.;;n,1
StringReplace, Sorted, Sorted,.;;",",1
FileAppend, %Sorted%;n, %OutDir%Step_4.csv

Loop, Read, %OutDir%Step_4.csv, %OutDir%Step_41.csv
{
If (A_LoopReadLine <> "")
FileAppend, %A_LoopReadLine% n
}

FileRead, Content, %OutDir%Step_41.csv
StringTrimRight, Content, Content, 1
FileAppend, %Content;., %OutDir%Step_42.csv
FileMove, %OutDir%Step_42.csv, %OutDir%Step_4.csv,1
FileDelete, %OutDir%Step_41.csv
Chapter 3

***Section 6: Determination of number of peptide species, total peptides, reversed sequence decoy database hits and contaminant peptides***

```
TP = 0
ListTP =
ListTPComplete =
Cont = 0
FD = 0

Loop, Read,%OutDir%\Step_4.csv
{
  IfInString, A_LoopReadLine, Reverse sequence
  ++FD
  Else, IfInString, A_LoopReadLine, Contaminant_
  ++Cont
  Else
  {
    ++TP
    StringSplit, Part, A_LoopReadLine, `;
    ListTP = %ListTP%`n%Part3%`n
  }
}

ListUP := ListTP
Sort, ListUP, U
StringSplit, Part, ListUP, `n
If (Part%Part0% = "")
  UP := Part0-1
Else
  UP := Part0
UPR := Round(UP/TP*100, 1)

FileAppend, Unique Peptides Chlr =%A_Tab%%UP%`nTotal Peptides Chlr =%A_Tab%%TP%`nUnique Ratio
  =%A_Tab%%UPR%%A_Tab%Percent`nPeptides Contaminants =%A_Tab%%Cont%`nFalse Discoveries
  =%A_Tab%%FD%`, %OutDir%\SampleReport.tab
```

***Section 7: Calculation of FDR***

```
FDNum = 0
FDR = 0

Loop, Read,%OutDir%\Step_4.csv, %OutDir%\Step_5.csv
{
  StringSplit, Part, A_LoopReadLine, `;
  IfInString, Part8, Reverse Sequence
  ++FDNum
  FDR := FDNum / A_Index * 1000
  If (FDR < 5)
    Acceptable := A_Index
  FileAppend, %A_LoopReadLine%`%;FDR%`n
}
```

***Section 8: Removal of contaminants and reversed sequence hits, cutoff at FDR of 0.5%***

```
Loop, Read,%OutDir%\Step_5.csv, %OutDir%\Step_6.csv
{
  IfNotInString, A_LoopReadLine, Reverse sequence
  IfNotInString, A_LoopReadLine, Contaminant_
  FileAppend, %A_LoopReadLine%`n
  If (A_Index = Acceptable)
    Break
}
```

78
Proteomics analysis pipeline

***Section 9: Splitting of data into lists of discrete and ambiguous peptide hits***

ListDiscretes =
ListNotDiscretes =

Loop, Read, %OutDir%\Step_6.csv
{ IfInString, A_LoopReadLine, NotDiscrete
 ListNotDiscretes = %ListNotDiscretes% %A_LoopReadLine%`n
 Else
 ListDiscretes = %ListDiscretes% %A_LoopReadLine%`n
}

FileAppend, %ListNotDiscretes%, %OutDir%\%OutNameNoExt%_ND.csv
FileAppend, %ListDiscretes%, %OutDir%\%OutNameNoExt%_D.csv

***Section 10: Distribution of discrete counts to their respective protein IDs, determination of number of peptide species, calculation of sum log E-values, calculation of protein sequence coverages***

IDC = 0

FileRead, IDsLength, E:\Material\ChlRe3CpMtContREV_noUniq_IDs&Length.txt
Loop, Parse, IDsLength, `r`n
{ StringSplit, Part, A_LoopField, %A_Tab%
 IDFasta_%A_Index% := Part1
 LengthFasta_%A_Index% := Part2
 ++$IDC
}

AC = 0

Loop, Read, %OutDir%\%OutNameNoExt%_D.csv, %OutDir%\ListIDs_D.txt
{ StringSplit, Part_%A_Index%_, A_LoopReadLine, `;`
 FileAppend, % Part_%A_Index%_7 `n`
 ++$AC
}

FileRead, Content, %OutDir%\ListIDs_D.txt
Sort, Content, U
FileAppend, %Content%, %OutDir%\ListIDs_Uniques_D.txt
Loop, Read, %OutDir%\ListIDs_Uniques_D.txt, %OutDir%\%OutNameNoExt%_Summary_D.csv
{ ID := A_LoopReadLine
 Count = 0
 Peptides =
 StartStop =
 ToolTip, %ID%, 10, 10
 Loop, %AC%
 { If (Part_%A_Index%_7 = ID)
 { If (Count = 0)
 { Peptides := Part_%A_Index%_3
 ++$Count
 SumLogEValue := LOG(Part_%A_Index%_4 * 10**Part_%A_Index%_5)
 StartStop = % Part_%A_Index%_10 `n`, % Part_%A_Index%_11 `n`
 } Else
 { Peptides = % Peptides `n`, % Part_%A_Index%_3
 ++$Count

}
Chapter 3

SumLogEValue := SumLogEValue + LOG(Part_%A_Index%_4 * 10**Part_%A_Index%_5)
StartStop = % StartStop Part_%A_Index%_10 "", Part_%A_Index%_11 "n"
}
}

StringUpper, Peptides, Peptides
UniquePeptides := Peptides
Sort, UniquePeptides, U D`
StringSplit, UP_, UniquePeptides,`
SumLogEValue := Round(SumLogEValue, 2)
Sort, StartStop, U
PeptideList =

Loop, %UP_0%
{
RegExReplace(Peptides, UP_%A_Index%, Replacement = "", FindCount)
If (A_Index < UP_0)
PeptideList = % PeptideList FindCount "** UP_%A_Index% "",
Else
PeptideList = % PeptideList FindCount "** UP_%A_Index%"
}

Loop, %IDC%
If (ID = IDFasta_%A_Index%){
Length := LengthFasta_%A_Index%
Break
}

Loop, %Length%
Pos_%A_Index% = 0
Loop, Parse, StartStop,`n{
StringSplit, Pair, A_LoopField,`
Loop, %Length%
{
If (Pair1 <= A_Index)
Pos_%A_Index% = 1
If (Pair2 = A_Index)
Break
}

Hits = 0
Loop, %Length%
Hits := Hits + Pos_%A_Index%
Coverage := Round(Hits / Length *100, 2)
FileAppend, %ID%`;%Count%`;%UP_0%`;%PeptideList%`;%SumLogEValue%`;%Coverage%`n
}

FileDelete, Input.txt
FileDelete, %OutDir%Step_2_1.csv
FileDelete, %OutDir%Step_2_2.csv
FileDelete, %OutDir%Step_3_1.csv
FileDelete, %OutDir%Step_3_2.csv
FileDelete, %OutDir%ListIDs_D.txt
FileDelete, %OutDir%ListIDs_Uniques_D.txt
FileDelete, %OutDir%ListIDs_NT.txt
FileDelete, %OutDir%ListIDs_Uniques_NT.txt
}
Proteomics analysis pipeline

;***Section 11: Quicksort function***

quicksort(list)
{
    StringSplit, list, list, `;

    If (list0 < 2)
    {
        list = `;%list%
        Return list
    }

    StringSplit, Part, List1, `, pivot := Part4*10**Part5

    Loop, Parse, list, `;
    {
        StringSplit, Part, A_LoopField, `, EValue := Part4*10**Part5
        If (EValue < pivot)
            less = %less%`;%A_LoopField%
        Else If (EValue > pivot)
            more = %more%`;%A_LoopField%
        Else
            pivotlist = %pivotlist%`;%A_LoopField%
        }
    }

    StringTrimLeft, less, less, 1
    StringTrimLeft, more, more, 1
    If (less <>"")
        less := quicksort(less)
    If (more <>"")
        more := quicksort(more)
    Return less . pivotList . more
}
Script 5. Implementation of G-test statistics and Benjamini-Hochberg correction for determining the significance of differences between protein spectral counts in two conditions

SetBatchLines, -1
SetFormat, Float, 0.20

FileSelectFile, Input, , E:\, Select tab delimited text file with control column left and case column right

SplitPath, Input, OutFileName, OutDir, OutExtension, OutNameNoExt, OutDrive

n1 := 0
n2 := 0
Loop, Read, %Input%
{
    Stringsplit, Part, A_LoopReadLine, %A_Tab%
    ID_%A_Index% := Part1
    x1_%A_Index% := Part2
    x2_%A_Index% := Part3
    x_%A_Index% := Part2 + Part3
    n1 := n1 + x1_%A_Index%
    n2 := n2 + x2_%A_Index%
    AC := A_Index
}

n := n1 + n2
Loop, %AC%
{
    y1_%A_Index% := n1 - x1_%A_Index%
    y2_%A_Index% := n2 - x2_%A_Index%
    y_%A_Index% := y1_%A_Index% + y2_%A_Index%
}

***Calculation of G values***

FileAppend, ID %A_Tab% x1 %A_Tab% x2 %A_Tab% x %A_Tab% y1 %A_Tab% y2 %A_Tab% y %A_Tab%
    n1 %A_Tab% n2 %A_Tab% n %A_Tab% G %A_Tab% w %A_Tab%
    Gadji, %OutDir%|%OutNameNoExt%_GTest.txt
    Loop, %AC%
    {
        If (x1_%A_Index% = 0)
            G_%A_Index% := % 2 * (x2_%A_Index% * LN(x2_%A_Index%) + y1_%A_Index% * LN(y1_%A_Index%) +
                               y2_%A_Index% * LN(y2_%A_Index%) - n1 * LN(n1) - n2 * LN(n2) - x_%A_Index% * LN(x_%A_Index%) -
                               y_%A_Index% * LN(y_%A_Index%) + n * LN(n))
        Else If (x2_%A_Index% = 0)
            G_%A_Index% := % 2 * (x1_%A_Index% * LN(x1_%A_Index%) + y1_%A_Index% * LN(y1_%A_Index%) +
                               y2_%A_Index% * LN(y2_%A_Index%) - n1 * LN(n1) - n2 * LN(n2) - x_%A_Index% * LN(x_%A_Index%) -
                               y_%A_Index% * LN(y_%A_Index%) + n * LN(n))
        Else
            G_%A_Index% := % 2 * (x1_%A_Index% * LN(x1_%A_Index%) + x2_%A_Index% * LN(x2_%A_Index%) +
                               y1_%A_Index% * LN(y1_%A_Index%) + y2_%A_Index% * LN(y2_%A_Index%) -
                               n1 * LN(n1) - n2 * LN(n2) - x_%A_Index% * LN(x_%A_Index%) -
                               y_%A_Index% * LN(y_%A_Index%) + n * LN(n))
        w_%A_Index% := % 1 + ((n/n1 + n/n2 - 1) * (n/x_%A_Index% + n/y_%A_Index% - 1)) / (6 * n)
        Gadji_%A_Index% := % G_%A_Index% / w_%A_Index% 
    }

FileAppend, ID %A_Index% A_Tab x1 %A_Tab% A_Tab x2 %A_Tab% A_Tab x %A_Tab% A_Tab y1 %A_Tab% A_Tab y2 %A_Index% A_Tab y %A_Index% A_Tab n1 A_Tab n2 A_Tab n A_TAB
    G_%A_Index% A_Tab w %A_Index% A_Tab Gadji_%A_Index% "n", %OutDir%\%OutNameNoExt%\_GTest.txt
}
***Calculation of p-values using Rscript***

FileCopy, %OutDir%\%OutNameNoExt%_GTest.txt, E:\Material\DataForR.txt, 1
Sleep 1000
Run, C:\Program Files\R\R-2.13.0\bin\i386\Rscript.exe E:\Material\Perform_G-Test_In_R.txt, C:\Program Files\R\R-2.13.0\bin\i386, Min
Sleep 1000
FileMove, E:\Material\DataForR.txt, %OutDir%\%OutNameNoExt%_GTest.txt, 1

/*
R-Commands in Perform_G-Test_In_R.txt:
data <- read.table("E:Material/DataForR.txt", header = TRUE, sep = "\t")
data[,14] <- 1-pchisq(data[,13],1)
colnames(data)[14] <- "p-value"
sorted_data <- data[order(data[,14]),]
write.table(sorted_data, file = "E:Material/DataForR.txt", append = FALSE, quote = FALSE, sep = "\t", col.names = TRUE)
*/

***Benjamini-Hochberg procedure to account for multiple testing***

Loop, Read, %OutDir%\%OutNameNoExt%_GTest.txt
{
  If (A_Index = 1)
  {
    Header := A_LoopReadLine
  }
  Else
  {
    AC := A_Index - 1
    Line_%AC% := A_LoopReadLine
    StringSplit, Part, A_LoopReadLine, %A_Tab%
    Parts2to14_%AC% =
    Loop, 14
    {
      If (A_Index > 1)
      If (A_Index = 2)
        Parts2to14_%AC% := Part2
      Else
        Parts2to14_%AC% = % Parts2to14_%AC% A_Tab Part%A_Index%
    }
    PValue_%AC% := Part15
  }
}
Rank = 1
FileAppend, %Header%\%A_Tab%Benjamini-Hochberg%A_Tab%Significant\n, %OutDir%\%OutNameNoExt%_GTest_Sorted&BHC.txt

Loop, %AC%
{
  If (A_Index > 1)
  If (PreviousPValue < PValue_%A_Index%)
    ++Rank
  BHValue := Rank / AC * 0.05
  If (PValue_%A_Index% <= BHValue)
    Significant = yes
  Else
    Significant = no
  FileAppend, % Parts2to14_%A_Index% A_Tab PValue_%A_Index% A_Tab BHValue A_Tab Significant\n  "\n  PreviousPValue := PValue_%A_Index%
}
;***Fold change calculation***

Loop,
    Read, %OutDir%\%OutNameNoExt\_GTest_Sorted&BHC.txt, %OutDir%\%OutNameNoExt\_GTest_Sorted&BHC_Log2FoldChange.txt
    
    IfNotInString, A_LoopReadLine, Benjamini-Hochberg
    
    StringSplit, Part, A_LoopReadLine,%A_Tab%
    
    If (Part2 = 0)
        Part2 = 1
    If (Part3 = 0)
        Part3 = 1
    
    FoldChange := Round((Part3/n2)/(Part2/n1), 2)
    Log2FoldChange := Round(LOG(FoldChange)/LOG(2), 2)

    FileAppend, %A_LoopReadLine%%A_Tab%%FoldChange%%A_Tab%%Log2FoldChange`n
    
    Else
    FileAppend, %A_LoopReadLine%%A_Tab%%FoldChange%%A_Tab%%Log2FoldChange`n
    
    FileDelete, %OutDir%\%OutNameNoExt\_GTest_Sorted&BHC.txt
    FileMove, %OutDir%\%OutNameNoExt\_GTest_Sorted&BHC_Log2FoldChange.txt, %OutDir%\%OutNameNoExt\_GTest.txt, 1
Chapter 4:

Linking proteome responses
with physiological and biochemical effects
in herbicide-exposed *Chlamydomonas reinhardtii*

Holger Nestler, Ksenia J. Groh, René Schönenberger, Rik I.L. Eggen, Marc J.-F. Suter

*Journal of Proteomics, in press*
4.1 Abstract

Exposure to a toxicant causes proteome alterations in an organism. In ecotoxicology, analysis of these changes may allow linking them to physiological and biochemical endpoints, providing insights into subcellular exposure effects and responses and, ultimately mechanisms of action. Based on this, useful protein markers of exposure can be identified. We investigated the proteome changes induced by the herbicides paraquat, diuron, and norflurazon in the green alga *Chlamydomonas reinhardtii*. Shotgun proteome profiling and spectral counting quantification in combination with G-test statistics revealed significant changes in protein abundance. Functional enrichment analysis identified protein groups that responded to the exposures. Significant changes were observed for 149-254 proteins involved in a variety of metabolic pathways. While some proteins and functional protein groups responded to several tested exposure conditions, others were affected only in specific cases. Expected as well as novel candidate markers of herbicide exposure were identified, the latter including the photosystem II subunit PsbR or the VIPPI protein. We demonstrate that the proteome response to toxicants is generally more sensitive than the physiological and biochemical endpoints, and that it can be linked to effects on these levels. Thus, proteome profiling may serve as a useful tool for ecotoxicological investigations in green algae.

Keywords

*Chlamydomonas reinhardtii*, herbicides, differential proteome analysis, MudPIT, label-free quantitation

Highlights

The proteome of herbicide-exposed *Chlamydomonas reinhardtii* was analyzed. Spectral counting was combined with G-test statistics to identify significant changes. 149-254 significantly changed proteins were identified, depending on exposure. Certain proteome variations were linked to physiological and biochemical endpoints. Proteomic profiling was demonstrated to be a useful tool for algal ecotoxicology.
4.2 Introduction

Herbicides constitute a significant part of the man-made organic toxicants released into the aquatic environment that adversely affect non-target species, such as the ecologically relevant green algae. In the frame of toxicity investigations and risk assessment integral endpoints such as growth and reproduction are traditionally monitored. However, before those endpoints show a significant change, the effects of exposure may already be manifest on the subcellular level, also at much lower toxicant concentrations.

Proteomics, which analyzes the expression of proteins within an organism, can be used to detect and characterize responses to external stimuli on the subcellular level. Currently, it is increasingly applied in ecotoxicology [1] and other fields of biological science. Stressors such as anthropogenic toxicants are expected to provoke variations in protein expression and degradation. Detection of these changes by means of differential proteomics provides a better understanding of the stress effects and mechanisms of toxic action, leading to improved toxicity predictions [2,3].

In several studies the effects of herbicides on green algae and other photosynthetic organisms were explored using “omics” techniques. These included transcriptome analyses of the effects of paraquat [4] on the freshwater green alga *Chlamydomonas reinhardtii*, and of glyphosate [5] and atrazine/bentazone [6] on soybean. Furthermore, the grape vine proteome changes upon flumioxazin exposure [7], and the alterations in the *Scenedesmus* metabolome following prometryn treatment [8] were characterized. Up to now, little information is available on the effects of herbicides on the proteome of green algae. Differential proteomic studies have a great potential for investigating subcellular stress mechanisms and responses affecting growth and other physiological and biochemical endpoints. This has been demonstrated by studies analyzing the effects of various stressors [9] including high light [10], heat [11], iron deficiency [12], and cadmium exposure [13] in *C. reinhardtii*. The observed variations were highly complex involving many cellular processes such as photosynthesis, induction of common stress related proteins, and antioxidant defense mechanisms, pigment and protein synthesis as well as cellular signaling.

Most previous work focused on strong stresses leading to severe perturbations on cellular endpoints and growth which prevented investigating the potentially highly sensitive and specific subcellular responses. Another shortcoming was that often only one stressor was used, which does not allow direct comparison of different mechanisms of action in the same biological system. The current study addressed these issues by looking at the effects of several herbicides at concentrations causing both milder and stronger stress conditions, with the overall goal to gain insights into the short-term changes in the proteome of exposed *C. reinhardtii*. Three model herbicides each representing a specific mechanism of action were used: paraquat (PQ), an inducer of superoxide radical generation at photosystem (PS) I, diuron (DR), an inhibitor of photosynthetic electron flow at PS II, and norflurazon (NF), an inhibitor of carotenoid synthesis [14,15]. This allowed distinguishing between more common herbicide-induced responses, and those which are more specific to particular herbicide exposure conditions.
Various methods are available for performing differential proteomic analyses and for quantifying and comparing protein expression levels across different samples [16]. Here, we utilized a gel-free method, known as multidimensional protein identification technology (MudPIT) [17,18]. A label-free spectral count approach [19] was used for quantitation, followed by statistical evaluation by G-test [20]. Finally, enrichment analysis allowed identifying functional groups of proteins and cellular systems that responded significantly to the different herbicide exposures.
4.3 Materials and methods

4.3.1 Chemicals and reagents

Tris(2-carboxyethyl)-phosphine hydrochloride (TCEP), iodoacetamide (IAA), and CHAPS were obtained from Sigma–Aldrich, Switzerland. Endoproteinase Lys-C (sequencing grade) and trypsin (recombinant, proteomic grade) were purchased from Roche Applied Science. Other conventional chemicals were either from Sigma-Aldrich or Fluka, Switzerland and HPLC solvents from Acros Organics, Belgium. Fused silica tubing was purchased from BGB Analytik AG, Switzerland. Reversed phase (RP, C18) and strong cation exchanger (SCX) resins were obtained from Macherey-Nagel AG, Switzerland.

4.3.2 Algae cultivation and herbicide exposure

Culturing and exposure conditions were as described previously [21]. Briefly, photoautotrophically grown *C. reinhardtii* wild type strain CC-125 (mt+) in the late exponential phase were cultivated in three replicates with a starting cell density of 5.5 × 10^5 cells × mL⁻¹ for each condition including the herbicide free controls. The herbicide exposure concentrations were 0.066 µM and 0.66 µM PQ, 3.3 nM and 0.1 µM DR, and 0.1 and 1 µM NF. Algae were exposed for 6 h under 80 ± 1 µmol m⁻² s⁻¹ of photosynthetically active radiation at 25 °C.

4.3.3 Protein extraction and preparation of tryptic digests

At the end of the exposure period, cells were harvested by centrifugation. The cell pellets were frozen in liquid nitrogen and stored at -80°C until further processing. The pellets were placed in a lysis buffer (0.1% CHAPS, 25 mM KCl, 25 mM Tris-HCL, 25 mM MgCl₂, pH 7.5) and disrupted on ice by four 30 s sonication cycles (Labsonic M, Sartorius Stedim Biotech, Goettingen, Germany) with 10 s pauses. Cell debris was removed by centrifugation and proteins were precipitated from the supernatant using a methanol/chloroform procedure [22]. The protein pellets were washed in methanol and then redissolved in a resolubilization buffer (9 M urea, 2 M thiourea, 50 mM Tris-HCl, pH 8.5). Total protein concentration was determined using the Bradford assay with BSA as a standard. The proteins were then reduced with TCEP and alkylated with IAA. The digestion with endoproteinase Lys-C for 6 h was followed by overnight incubation with trypsin at a ratio of enzyme to total protein of 1:100 for both digests [23]. The reaction was quenched with formic acid.

4.3.4 LC separation and MS

The 2D-LC-MS/MS method MudPIT was chosen for proteome profiling. Due to an additional separating dimension, this method delivers a significantly higher analytical coverage of proteins in complex
samples, as compared to 1D-LC approaches [18]. While the peptide fractionation occurs on-line, MudPIT is also less labor-intensive than off-line fractionation techniques. Each of the three replicates for every condition was analyzed, leading to a sufficient coverage of the detectable proteins as reported previously [24]. For each MudPIT run 10 µg of peptides was pressure loaded onto a triphasic column (5 µm C18-SCX-3 µm C18) made in-house from a 100 µm ID fused-silica (prepared using a Sutter-2000 needle puller). After desalting, the peptides were eluted using an 11-step MudPIT protocol [18]. Each analysis step had a gradient duration of 120 min, resulting in completion of one MudPIT run after 22 h. The LTQ-Orbitrap XL (Thermo Scientific, Bremen, Germany) used for data acquisition was operated in nano-ESI positive ion mode at 1.5 kV spray voltage. The ion transfer capillary temperature was set to 200°C, the tube lens to 110 V. Full scan FT mass spectra (300-2000 m/z, resolution of 60,000) were followed by seven data-dependent MS/MS scans (ions selected based on peak intensity) acquired in the linear ion trap with normalized 35% collision energy. This sequence of experiments was carried out throughout all MudPIT steps. Dynamic exclusion, after two repetitive MS/MS scans, was set to 60 s and the automatic gain control (FT) to $10^6$.

4.3.5 Analysis of LC-MS/MS data

Automation of data processing and analysis was performed using AutoHotkey scripts (version 1.0.48.05, http://www.autohotkey.com/) written in-house. Peak listing from MS raw files was done with extract_msn.exe supplied with the XCalibur package (Thermo Fisher Scientific) followed by merging of the MS/MS peak lists to the mgf format by the perl script merge.pl (Matrix Science, Boston, MA, USA). The data were searched for peptide hits with the Open Mass Spectrometry Search Algorithm (OMSSA, version 2.1.7) [25] against *Chlamydomonas JGI* (Joint Genome Institute) protein sequence database version 3.1 containing ~14,600 nuclear encoded proteins (“frozen gene catalog”) [26] supplemented with 76 organelle encoded proteins (obtained from ChlamyCyc, University of Potsdam and MPI of Molecular Plant Physiology, Potsdam, Germany) and common contaminants such as digestion enzymes, human keratins and BSA (obtained from NCBI). Sequences of less than 10 amino acid length were removed from the database. Furthermore, for estimation of false discovery rates (FDR) a reversed version of this database [27] created with decoy.pl (Matrix Science) was added.

During database search carbamidomethyl-cysteine was set as a fixed modification. Variable modifications were protein N-terminal acetylation and peptide N-terminal formation of pyroglutamic acid. Tryptic specificity was set to a maximum of three missed cleavages. Product ion m/z tolerance was set to 0.3 Da. The precursor ion mass accuracy was initially set to 0.05 Da and finally to 0.005 Da after removal of systematic mass errors [28,29]. The minimum precursor charge to consider multiply charged products was set to 2. For individual MS/MS spectra only the peptide hits found with best confidence (lowest E-value) and only discrete hits with FDR less than 0.5% were considered for further statistical analysis.

Results from the triplicate runs were merged, and spectral counts were calculated from the number of discrete peptides observed for each identified protein [19]. Subsequently, two-column count tables were
created for the control and all exposure conditions. Count pairs having less than two peptides in the control and exposure conditions, were removed from the list. The resulting lists for control and exposures were analyzed by G-test as described previously [20]. The statistical platform R (http://www.r-project.org/) was used to calculate the p-values, and the Benjamini-Hochberg FDR control (<0.05) was applied to account for multiple testing [30]. For each protein, the proportion of its counts in the total counts of all proteins was determined for all experimental conditions. The protein fold changes were then estimated by dividing this value in the exposure conditions by the corresponding value in the control condition. For this calculation, 0 counts in either condition were substituted with the minimum observable value 1 to avoid calculating fold changes with 0. Lists of significantly different protein IDs were submitted for enrichment analysis based on MapMan ontology of the C. reinhardtii genome [31,32] performed by the Algal Functional Annotation Tool [33]. This analysis was performed for the complete list of significantly changed IDs as well as for separate lists of those being up- and down-regulated. Results were filtered by requiring at least 3 protein IDs to be present in a bin with a score < 0.05. Manual cleanup was performed to remove occasional redundancy among daughter and mother terms. Heat map graphics were generated with Mayday [34].
4.4 Results and discussion

In this study MudPIT was used to analyze the changes in the proteome of *C. reinhardtii* exposed for 6 h to the herbicides PQ, DR, and NF, at a low and high concentration each (LPQ and HPQ, LDR and HDR, LNF and HNF, respectively). Previously, we thoroughly characterized the physiological and biochemical effects of these herbicides on *C. reinhardtii* at 2, 6 and 24 h [21]. Suitable herbicide concentrations for the proteomic profiling were selected based on the results of that study. The higher concentration was close to the EC50 values obtained for the physiological and biochemical endpoints reflecting the mechanism of action of the respective toxicant. The lower concentration was selected around the LOEC values of the various physiological and biochemical endpoints as well as growth. Investigating low concentration conditions was of particular interest since protein variations found under these conditions could be useful indicators of cellular responses to mild disturbances and stresses, and thus be early protein markers of effects. Furthermore, this also allows examining dose-related responses between exposure concentrations and protein expression.

Searching against the database leads to ambiguous hits, that is, hits of equal confidence to more than one possible protein sequence, for around 8% of the collected spectra. These hits were removed from the lists processed for the spectral counting. In all control and exposure MudPIT runs combined, roughly 3,900 proteins were identified with at least two discrete peptide hits at a peptide false discovery rate of 0.5%. Thus, the proteome coverage can be estimated to be 27% when relating the number of identifications to the approximately 14,600 sequences of the database used [35]. The depth of the proteome characterization in *C. reinhardtii* obtained by MudPIT-based profiling is comparable to other recent studies characterizing the proteome of this organism with other MS/MS based methods [11,35]. For each experimental condition the complete lists of identified proteins with their spectral counts can be found in supplementary file 1.

The statistical analysis of the spectral counts of control and treated algae samples using G-test revealed 149 to 254 significantly up- and down-regulated proteins (supplementary file 2). Subjecting the separate lists of all significantly changed, the down- as well as up-regulated protein identifiers ("Total", "Down" and "Up" in Table 4.1) to enrichment analysis based on MapMan annotation, identified functional groups enriched relative to the total proteome. All groups with a score < 0.05 and at least 3 identifiers present, are listed in Table 4.1. Figures 4.1-3 show the response to the six different conditions of significantly changed proteins discussed below.
Table 4.1: Numbers of protein identifiers in significantly enriched functional categories of MapMan annotation in the significantly changed (TOTAL), DOWN- and UP-regulated proteins of *C. reinhardtii* after 6 h of exposure to low and high concentrations of the herbicides paraquat (LPQ, HPQ), diuron (LDR, HDR) and norflurazon (LNF, HNF).

<table>
<thead>
<tr>
<th></th>
<th>LPQ TOTAL</th>
<th>LPQ DOWN</th>
<th>LPQ UP</th>
<th>HPQ TOTAL</th>
<th>HPQ DOWN</th>
<th>HPQ UP</th>
<th>LDR TOTAL</th>
<th>LDR DOWN</th>
<th>LDR UP</th>
<th>HDR TOTAL</th>
<th>HDR DOWN</th>
<th>HDR UP</th>
<th>LNF TOTAL</th>
<th>LNF DOWN</th>
<th>LNF UP</th>
<th>HNF TOTAL</th>
<th>HNF DOWN</th>
<th>HNF UP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS</td>
<td>19</td>
<td>11</td>
<td>8</td>
<td>17</td>
<td>13</td>
<td>4</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>19</td>
<td>5</td>
<td>14</td>
<td>23</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>PS: calvin cycle</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS: light reaction</td>
<td>12</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>PS: light reaction</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS: light reaction</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS: light reaction</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS: light reaction</td>
<td>4</td>
<td>3</td>
<td></td>
<td>3</td>
<td>4</td>
<td></td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetrapyrrole synthesis</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amino acid metabolism</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amino acid metabolism</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>3</td>
<td>4</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>49</td>
<td>25</td>
<td>24</td>
<td>63</td>
<td>32</td>
<td>31</td>
<td>29</td>
<td>19</td>
<td>10</td>
<td>44</td>
<td>22</td>
<td>22</td>
<td>70</td>
<td>42</td>
<td>28</td>
<td>33</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>protein, aa activation</td>
<td>5</td>
<td></td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, degradation</td>
<td>16</td>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td></td>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, degradation</td>
<td>5</td>
<td></td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, folding</td>
<td>3</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, synthesis</td>
<td>32</td>
<td>18</td>
<td>14</td>
<td>31</td>
<td>26</td>
<td>5</td>
<td>20</td>
<td>13</td>
<td>7</td>
<td>29</td>
<td>16</td>
<td>13</td>
<td>47</td>
<td>26</td>
<td>21</td>
<td>30</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>protein, synthesis</td>
<td>3</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, synthesis, elongation</td>
<td>20</td>
<td>9</td>
<td>11</td>
<td>22</td>
<td>21</td>
<td>10</td>
<td>7</td>
<td>18</td>
<td>9</td>
<td>9</td>
<td>31</td>
<td>17</td>
<td>14</td>
<td>21</td>
<td>8</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, synthesis, elongation</td>
<td>9</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, synthesis, elongation</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein, targeting</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Count 1</td>
<td>Count 2</td>
<td>Count 3</td>
<td>Count 4</td>
<td>Count 5</td>
<td>Count 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>major CHO metabolism</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>major CHO metabolism.synthesis</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minor CHO metabolism.others</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glycolysis</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mitochondrial electron transport/ATP synthesis</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mitochondrial electron transport/ATP synthesis.F1-ATPase</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lipid metabolism</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lipid metabolism. FA synthesis and FA elongation</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cell</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cell.cycle</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cell.organization</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>redox.ascorbate and glutathione</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>redox.ascorbate and glutathione.ascorbate</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>redox.perioxidoxin</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>redox.thioredoxin</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stress.abiotic</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hormone metabolism</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-metabolism.ammonia metabolism</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nucleotide metabolism</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nucleotide metabolism.synthesis</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA.RNA binding</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>signaling.calcium</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transport</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transport.ABC transporters and multidrug resistance systems</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transport.p- and v-ATPases</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>misc</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>misc.short chain dehydrogenase/reductase (SDR)</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not assigned.no ontology.pentatricopeptide (PPR)repeat-containing protein</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.4.1 Common responses induced by low and high concentration exposure conditions

Although the specific mechanisms of action of the three investigated herbicides are different, they all act primarily on the chloroplast and target the photosystems and overall photosynthesis. A further commonality among the toxicity mechanisms of our three model herbicides is the disturbance of cellular redox homeostasis, which can lead to oxidative stress [14]. Hence, a number of common responses were expected across several or all of the investigated exposure conditions.

Regarding the components of the photosynthetic machinery, we observed wide changes in the expression of key proteins. These included various subunits of PS I and II, their light harvesting complex proteins as well as of the enzymes of the Calvin cycle, with patterns differing between the investigated exposure conditions (Table 4.1). Because PS II, particularly its D1 protein, is a well-known target for oxidative stress and photoinhibition [36], it was expected that it would be more severely affected than PS I. However, this was not the case, as shown by the fact that many subunits of both PS I and PS II were changed across the various exposures (Table 4.1). The possible explanation could be that an extensive remodulation of both photosystems occurs, or that more effective mechanisms for protection and/or repair of PS II exist.

For example, the D1 protein of PS II (PsbA, Figure 4.1) had significantly increased counts in the LPQ, HPQ, HDR and LNF conditions. For PQ and NF an increased synthesis of the protein could play a role which was hinted upon by the PsbA translation factor (TBA1, Figure 4.1, supplementary file 1) that showed a tendency to increase in the PQ and NF exposures, which, however, was only significant for LPQ at $p < 0.05$. For HDR, on the other hand, this could be caused by the retardation of its photoinactivation and degradation [37,38]. Furthermore, in every exposure condition, significantly increased counts of the PS II subunit PsbR were observed which could be a response to the general stress encountered by PS II, as this protein is involved in the stabilization of other PS II subunits [39]. This finding is supported by a previously found up-regulation of the PsbR transcript in response to the copper-induced oxidative stress in green algae [40]. Another notable increase was observed for the electron carrier plastocyanine PCY1 in LPQ, LDR, HNF, and, most strongly, LNF, the exposure conditions where the effective quantum yield, and with that photosynthesis, was not yet significantly affected [21]. The uncharacterized luminal peptide ULP1 had a comparatively high spectral count value in controls, but was significantly reduced in all treatments, most strongly in HDR and HNF. This matches the findings of Hahn et al. who found that this protein was suppressed in a non-photosynthetic Chlamydomonas mutant [41], and further corroborates that it might be functionally linked to PS I [11], which together with ULP1, was also widely suppressed. This suppression was for example seen in the downregulation of the light-harvesting protein LHCA4 in all exposures. Generally, the herbicide exposures, as expected, widely affected the photosynthetic machinery, which is known to respond strongly to various environmental stimuli [42]. These responses could indicate acclimation of photosynthetic processes to cope with the imposed disturbances.

The pathways leading to the formation of the photosynthetic pigments, such as “tetrapyrrole synthesis” and “isoprenoids”, responded significantly in all high concentration conditions with up- and down-regulation
patterns depending on the herbicide used (Table 4.1). The low concentrations did not show this behavior, or only to a lesser extent (chlorophyll synthesis in LPQ), thus allowing us to distinguish between the lesser and more severe exposure conditions and the concomitant cellular stress observed.

The involvement of oxidative stress was reflected by the frequent up-regulation of various redox and antioxidant defense related proteins, which has already been observed previously in herbicide-exposed photosynthetic organisms on the transcript and protein levels [4,6,7]. These proteins are expressed at different subcellular locations and belong to several diverse systems and enzyme classes, including various thioredoxins, peroxiredoxins, and glutathione metabolism enzymes. Many of them were detected with comparatively high counts, and, depending on the particular exposure, responded in a very differentiated way. For example, glutathione S-transferases were up-regulated in the extreme HPQ condition, but not in HDR and HNF (Figure 4.2) thus indicating that in the two latter conditions no comparably strong induction of oxidative stress took place. Diversity of the covered cellular systems, reliability of detection, and specificity of the response to particular exposure conditions make redox related proteins excellent targets for further research on exposure effects in photosynthetic organisms.

A crucial functional protein group, widely affected at both low and high herbicide concentrations, was “protein synthesis” containing various cytosolic, plastidic and mitochondrial ribosomal proteins. In addition, a number of transcription, translation initiation and translation elongation factors (supplementary file 2), as well as several proteins belonging to the “amino acid metabolism” group were responding frequently. These included, e.g the chloroplast-encoded translation factor TufA which was significantly up-regulated in HNF and all low concentration conditions (Figure 4.3), similar to the pattern observed for PCY1 (see above). Hence, the herbicides, both at low and high concentrations, strongly affect protein synthesis. The influence on protein synthesis exerted by the investigated herbicides probably contributed to the overall proteome variations observed, since protein expression is known to be greatly influenced at the translation stage [43]. Furthermore, modulation of protein synthesis may provide a basis for cellular acclimation to herbicide induced disturbances.

Another notable common response included the thiazole biosynthetic enzyme THI4a, a protein involved in thiamine synthesis. This protein had high spectral counts which were decreased significantly in all conditions except LNF (Figure 4.3). The down-regulation was strongest (>14-folds) in the HPQ condition, suggesting a particularly high sensitivity of THI4a to oxidative stress. On the other extreme, the acyl carrier protein ACP2, involved in fatty acid synthesis, showed increased levels in LPQ and, more pronounced, in the LDR and HNF conditions while in HPQ it was significantly reduced. These two examples illustrate that metabolic pathways other than those commonly expected to respond can be heavily influenced by herbicides, even at low concentrations.

In conclusion, common proteome alterations were indeed observed in *C. reinhardtii* when exposed to the different concentrations of the herbicides PQ, DR and NF. These variations may highlight cellular components and pathways of general importance for the response to chloroplast-targeting herbicides.
4.4.2 Specific responses among the tested exposure conditions

Paraquat

The bipyridinium herbicide PQ diverts electrons primarily from PS I and transfers them to molecular oxygen, thereby producing superoxide radicals which subsequently give rise to hydrogen peroxide, hydroxyl radicals, and other ROS [14]. Thus, HPQ exposure leads to the generation of severe oxidative stress resulting in detrimental effects on most physiological and biochemical endpoints causing cellular death between 6 and 24 h [21]. In this regard, HPQ was the most extreme condition used in this study. The exposure effects and response to stress were reflected in the differential enrichment of functional groups as well as in altered protein profiles of individual proteins (Figures 4.1-3, Table 4.1).
Figure 4.2: Heat map showing redox-related proteins which were significantly changed in *Chlamydomonas reinhardtii* in response to at least one of the exposure conditions used (PQ, DR, and NF at two different concentrations L and H). Statistical evaluation was performed with G-test, fold changes are presented as log 2 values.

Based on the mechanism of action of PQ, an up-regulation of the proteins involved in the antioxidant defense mechanisms could be expected. However, the reporter dye-based measurements of oxidative activity at the HPQ condition suggested an overall decline of the antioxidant defense activity, possibly due to overloading [21]. Likewise, we saw variable responses in the proteome, both induction and suppression. Important antioxidant enzymes were strongly up-regulated (Figure 4.2). These included the glutathione S-transferases GSTS1 and GSTS2, which are induced by oxidative stress and detoxify oxidized substrates such as lipid peroxides, and the ascorbate peroxidase APX1, which, using ascorbate as electron donor, degrades peroxides such as hydrogen peroxide, thereby producing dehydroascorbate. Interestingly, we also observed an up-regulation of dehydroascorbate reductase (DHAR), which regenerates ascorbate from
dehydroascorbate at the expense of glutathione, and of glutamate-cysteine ligase (GSH1), which catalyzes the first step in glutathione synthesis. These findings indicate that ascorbate and glutathione systems play a central role in the defense against the oxidative stress provoked by HPQ. Besides these, thioredoxin o (TRXo), a mitochondrial thioredoxin, and nucleoredoxin (NRX2), a nucleus localized thioredoxin related protein, were also increased. This indicates that PQ induces oxidative stress not only in the chloroplast but also in the mitochondria [44]. This oxidative stress also affects the nucleus.

Figure 4.3: Heat map showing additional proteins which were significantly changed in *Chlamydomonas reinhardtii* in response to at least one of the exposure conditions used (PQ, DR, and NF at two different concentrations L and H). Statistical evaluation was performed with *G*-test, fold changes are presented as log 2 values.
On the other hand, down-regulation in response to herbicide exposure has also been observed. This was the case for some of the redox related proteins, including TRXm, a crucial and highly abundant chloroplastic thioredoxin, as well as the peroxiredoxins PRX2 (cytosolic) and PRX6 (chloroplastic), belonging to another group of major antioxidants. TRXm interacts with Calvin cycle enzymes [45], several of which were decreased as well (Table 4.1, and below). Furthermore, although expected, we did not observe an increase in counts for chloroplastic superoxide dismutase FSD1 which transforms superoxide into \( \text{H}_2\text{O}_2 \) and thus would be the first line of defense against the PQ produced superoxide radicals. This could be due to inactivation and degradation of FSD1 by \( \text{H}_2\text{O}_2 \) and other ROS [46]. Thus, the described down-regulations point toward a depletion of antioxidative capacity during HPQ exposure, particularly in the chloroplast, its primary target. A contributing factor could also be a decreased synthesis due to mRNA template depletion, as down-regulation of various antioxidant protein transcripts in response to PQ exposure has been reported previously [4].

The observed disturbances of the redox system clearly reflected the strong stress induced by the HPQ condition. This was further supported by the up-regulation of a number of proteins belonging to the “abiotic stress” group which was significantly affected only in this exposure (Table 4.1). These included the small heat shock proteins HSP22C and HSP22F, which are a type of chaperones preventing denaturation and aggregation of proteins in plastids [47]. Other chaperones were also up-regulated, including CLPB3 (HSP100 family) and HSP70A (Figure 4.3). Several of these proteins responded similarly during strong heat stress [11]. Another response to the severe oxidative stress caused by HPQ was the up-regulation of the protein degradation pathway (Table 4.1). The up-regulated proteins included five components of the proteasome complex, the ubiquitin-activating enzyme UBA1 involved in marking proteins for degradation via the proteasome, and a number of proteases and peptidases. The induction of the proteasome would enable the removal of greater amounts of oxidized and otherwise damaged proteins, which are the prime target for oxidative modifications in cells [48], in addition to lipids. Within the functional group “protein synthesis”, the overwhelming majority of significantly changed ribosomal proteins were decreased in their expression (Table 4.1). This indicates a disturbance of the protein synthesis accompanying the general cellular decline caused by PQ exposure.

HPQ resulted in a strong inhibition of photosynthesis due to damage of the photosynthetic apparatus. This was indicated by the severe reduction of the PS II effective quantum yield and reduced contents of photosynthetic pigments [21]. On the protein level, this impact on photosynthetic performance was reflected in the down-regulation of many proteins involved in the light reactions and the Calvin cycle. Furthermore, all significantly changed proteins annotated to the pathways of chlorophyll and carotenoid synthesis were down-regulated (Table 4.1) indicating that the observed loss of pigments was probably due not only to photooxidative destruction but also to blocked pigment synthesis. Some of the observations may indicate that the cells actively tried to maintain photosynthetic functions, in particular the activity of PS II. These responses included the up-regulation of the D1 protein (PsbA), a central component of PS II, and of the peptidyl-prolyl cis-trans isomerase FKB20-2, whose orthologue in *Arabidopsis thaliana* participates in the
accumulation of the PS II supercomplex in the chloroplast thylakoid lumen (Figure 4.1) [49]. Furthermore, the up-regulation of the serine protease DEG1C may be seen as a response to damaged photosystem proteins because it is similar to an A. thaliana protein which was shown to be involved in the turnover of photodamaged PS II [50].

Taken together, the proteome changes which occurred in response to HPQ were typical for a strong oxidative stress condition and quite similar to those seen after the exposure of C. reinhardtii to a comparable stressor, a high concentration of cadmium [13]. Exposure to this metal led to oxidative stress and also resulted in the down-regulation of proteins involved in photosynthesis and chlorophyll synthesis, as well as in up-regulation of chaperones, antioxidant defense and glutathione synthesis pathways.

The LPQ condition was characterized by the lack of significant responses of the physiological and biochemical endpoints at the 6 h time point. After 24 h, however, measurements of oxidative activity indicated a substantial induction of antioxidant defense, while ATP content and effective quantum yield showed a decreasing tendency [21]. The protein profiling revealed extensive variations affecting many cellular systems (Table 4.1, supplementary file 2).

As suggested by the PQ mechanism of action and the observations at the 24 h time point, a positive response of antioxidant defense proteins was expected. However, a significant increase was observed only for some less important redox related proteins such as the plastidic thioredoxin-like protein CDSP32, and the chloroplastic glutaredoxin GRX6 (Figure 4.2). These observations may suggest a specialized role played by these proteins under the conditions of slightly stimulated superoxide production and induction of low level oxidative misbalances. Interestingly, the expression of the mitochondrial superoxide dismutase MSD2 was significantly reduced. Taken together, these variations support the hypothesis that the antioxidant defense mechanisms were easily able to cope with the oxidative stress induced by LPQ after 6 h of exposure.

Although the oxidative damages caused by LPQ were not pronounced, several cellular systems displayed strong alterations, among them a large number of proteins related to photosynthesis (Table 4.1). These included various photosystem subunits, light harvesting complex proteins as well as Calvin cycle enzymes, most of which were down-regulated. Further, the changes in the expression level of proteins belonging to the functional groups “mitochondrial electron transport” and “glycolysis” may indicate an influence of LPQ on cellular energy metabolism. The same is indicated by the observed down-regulation of the plastidic ADP/ATP carriers, AAA1 and AAA2 (Figure 4.3), and of PGR5 (Figure 4.1), a protein involved in the ATP production during cyclic electron transport [51]. Furthermore, extensive perturbations of protein synthesis occurred, as many ribosomal subunit proteins were significantly up- or down-regulated (Table 4.1). Other observations such as the up-regulation of the mitogen-activated kinase MAPKKK13 (Figure 4.3), which was the only significant up-regulation of a MAP kinase observed in this study, could be an indication for a yet unknown signaling pathway which is crucial for the response to the low concentration of PQ. Overall, proteome variations provoked by LPQ demonstrate adaptive responses such as remodeling of the photosystems, energy metabolism and protein synthesis, which could allow the exposed cultures to
maintain cellular homoeostasis, thus explaining the lack of significant responses on the integrative physiological and biochemical levels.

The LPQ condition was 10 times less concentrated than the HPQ and was expected to generate much lower concentrations of superoxide radicals and only mild if any oxidative stress. In agreement with this, the proteome responses to LPQ greatly differed from those observed for HPQ. In particular, the indicators of strong stress, such as up-regulation of the small HSPs, facilitators of protein degradation or certain antioxidant defense proteins, remained unaffected in LPQ. Among those, GSTS2 was even significantly down-regulated, in contrast to its dramatic up-regulation by the HPQ condition (Figure 4.2). Overall, the proteome variations in HPQ clearly reflected the cellular decline caused by severe oxidative stress induced by PQ, while LPQ evoked responses that could serve as indicators of mild redox misbalances and highlight the pathways involved in maintenance of viability.

**Diuron**

The phenylurea herbicide DR is a model compound for blockage of linear electron transfer at the Qb binding pocket of the PS II D1 protein [14]. Reflecting this mechanism of action, the HDR condition was characterized by a roughly 50% inhibition of effective quantum yield of PS II [21]. An inhibition of electron flow can generally be expected to cause strong responses of the proteome of photosynthetic cells as, for example, the expression of numerous genes appears to be influenced by the redox state of the plastoquinone pool and b6f complex [42].

As was expected, this herbicide had a significant effect on the expression of a number of proteins involved in photosynthesis, although apparently not as strong as that caused by HPQ and HNF for example (Table 4.1). Among those changes was a significant increase of the spectral count of the DR target site itself, the D1 protein (PsbA), compared to the control (Figure 4.1). This could be expected because DR has been shown to retard degradation of this specific protein [37,38]. Other PS II subunits, PsbC, PsbD (D2 protein) and PsbE, were affected negatively and much stronger than in any other condition tested in this study. Thus, they could be useful indicators of exposure to high levels of DR. Further downstream of the electron flow, HDR had a significant impact as well (Table 4.1, supplementary file 2). For example, several subunits of PS I were decreased, including the core subunit PsbB. Furthermore, the nearly 5-fold up-regulation of the light harvesting complex protein LHCA1 (Figure 4.1) was accompanied by the down-regulation of four other LHCA proteins. Overall, the observed impact of HDR on PS I and II was largely negative. Also, considering the fact that the PS II quantum yield was reduced, no explanation could be found for the increased expression levels of seven Calvin cycle enzymes involved in carbon fixation.

The photosynthesis inhibition caused by DR also results in starvation of the organism. Among other indications, this was reflected by the roughly 50% inhibition of *C. reinhardtii* cell volume growth at the HDR condition which was largely due to the smaller average cell size in the exposed cultures [21]. On the protein level, the enrichment of the functional group “cell cycle” was observed (Table 4.1), which contained
three peptidyl-prolyl cis-trans isomerases, proteins involved in cellular signal transduction. Furthermore, toposomerase SPO11a, a protein participating in meiosis, and the septin-like protein SEP1, playing a role in cell division, were up-regulated (Figure 4.3), indicating various disturbances induced by DR on growth and cell development.

Finally, the inhibition of PS II by DR causes photooxidative stress which leads to induction of membrane damage by lipid peroxidation [14]. Connected to that, an approximately 30% inhibition of oxidative activity was seen at HDR compared to the control, hinting at the induction of the antioxidant defense mechanisms [21]. Indeed, several proteins of this group were significantly up-regulated, including an ascorbate peroxidase (JGI ID 165193), FSD1, as well as peroxiredoxin PRX1 (Figure 4.2), which is a chloroplastic 2-Cys peroxiredoxin protecting the photosynthetic machinery from oxidative damage [52]. Furthermore, the chloroplastic thioredoxins TRXf2 and TRXx, and the thioredoxin-like protein disulfide isomerase PDI5 were up-regulated in this condition. The response of the latter two was specific for HDR in the presented study. On the other hand, the observed down-regulation of the cytosolic GRX1 and DHAR could indicate defense systems playing a lesser role in the HDR exposed algae.

Unlike for HDR, a significant albeit very small reduction of the effective quantum yield in response to LDR was seen only after 2 h. At the 6 h time point, however, this inhibition had disappeared, and the only other endpoint that remained significantly changed was a slightly reduced oxidative activity [21]. Interestingly, these rather subtle disturbances of cellular parameters caused by LDR were accompanied by some strong responses on the protein level (Table 4.1, supplementary file 2), even though the total number of significantly affected proteins was smaller when compared to all other conditions, including HDR.

Regarding the number of significantly changed photosynthesis related proteins, LDR had an even greater effect than the HDR condition, in that a total of 20 annotated proteins were affected, compared to the 15 at HDR (Table 4.1). The observed changes included the up- and down-regulation of many photosynthesis-related proteins such as several subunits of PS I and PS II, including the unique response of Psb28 which was up-regulated (Figure 4.1), as well as PS I light harvesting complex proteins, several of which were down-regulated.

Within the redox related proteins, the expression of several important enzymes increased, including PRX1 and the mitochondrial superoxide dismutase MSD1 (Figure 4.2). The latter also shows that LDR impacts on the antioxidant defense system were not restricted to the chloroplast. Further changes included the increase in the levels of the general regulatory factor 14-3-3 like protein FTT1 (Figure 4.3), which is involved in key physiological processes and modulates the activity of many target proteins upon stress [53]. Two predicted transcription factors (JGI IDs 143178 and 190755) showed a strong up-regulation in LDR but not in any other exposure and thus could serve as indicators for this condition.

Overall, the responses of the proteome to LDR were not as substantial as those to HDR, with the exception of effects on the photosynthesis related proteins. For example, the ribosomal subunits (Table 4.1) or redox related proteins were only slightly affected by LDR. The observed proteome variations also showed...
that the stress caused by HDR was not as severe as that caused by HPQ, although both conditions corresponded to the EC50s for the inhibition of effective quantum yield by the respective herbicide. The severity of the effects on the biochemical and physiological endpoints decreased over time for both, HDR and LDR, probably due to growth-caused dilution of the herbicide [21]. In addition, some of the observed proteome variations such as the up-regulation of PRX1 or remodulation of the photosystems could be involved in the recovery process from DR-caused perturbations.

**Norflurazon**

The pyridazinone herbicide NF inhibits the enzyme phytoene desaturase, which catalyzes a critical step of carotenoid synthesis [15]. This inhibition is followed by a depletion of carotenoids which subsequently leads to a loss of chlorophylls [54] and reduction of photosynthetic activity. Accordingly, on the physiological and biochemical levels the exposure to HNF after 6 h was manifested mainly in the depletion of the carotenoid pool which after 24 h was accompanied by a reduction of the chlorophyll levels [21]. In addition to the inhibition of carotenoid synthesis, the herbicide NF is also known to inhibit PS II through a DR-similar mechanism [55]. Although in the HNF condition this inhibition was not measurable on the physiological and biochemical levels, proteome alterations influenced by this secondary mechanism of action cannot be excluded.

On the protein level, no significant response of the primary NF target, phytoene desaturase PDS1, was observed, although increasing the levels of this enzyme in response to the accumulating phytoene could possibly allow the organism to counteract the NF-caused inhibition. Apparently, no such regulation feedback exists in *C. reinhardtii*, as can be seen by the fact that the carotenoid levels remained low between 6 and 24 h [21]. However, the functional protein group “isoprenoids” was enriched (Table 4.1), containing several other significantly changed enzymes involved in carotenoid synthesis, including the up-regulated 1-deoxy-D-xylulose 5-phosphate synthase DXS, which catalyzes the initial step of carotenoid synthesis, and geranylgeranyl reductase GGR (Figure 4.1), which provides a crucial metabolite for chlorophyll synthesis [56]. Interestingly, three enzymes of the functional group “tetrapyrrole synthesis” were also significantly up-regulated, indicating an increased synthesis of chlorophyll a and b. Up-regulation of chlorophyll production provides a plausible explanation of the observed stability of these pigments in the presence of an already significant loss of carotenoids after 6 h of exposure to HNF [21].

Because carotenoids constitute a major component of the photosynthetic machinery, their depletion can be assumed to cause disturbances within this system. Indeed, strong perturbations were observed in both, PS I and II (Table 4.1). In particular, all nine light harvesting complex proteins of PS I were reduced, seven of them significantly (supplementary file 2). Down-regulation of these proteins could thus serve as early and strongly responding indicator of exposure to NF. The ATP content, which depends heavily on photosynthetic electron transport, was increased significantly at the HNF condition [21]. Accordingly, proteins involved in cyclic electron transfer such as PCY1 or the b6f complex subunits PetA and PetO (Figure 4.1) were up-
regulated, which could point toward the involvement of this ATP generating mechanism. Another link to the observed accumulation of ATP could be the enriched functional group “mitochondrial electron transport/ATP synthesis” as well as the up-regulation of the mitochondrial adenine nucleotide translocator ANT1, which enables the antiport of ATP and ADP over the inner mitochondrial membrane.

An important component of the NF mechanism of action is the oxidative damage of cellular components, particularly of lipids, caused by the absence of carotenoids which are crucial antioxidants [14]. Similar to PQ and DR, and indicated by the observed response of the oxidative activity endpoint, NF exposure also resulted in an up-regulation of antioxidant defense mechanisms [21]. On the protein level, an up-regulation of the chloroplastic PRX1 and TRXf2 as well as an ascorbate peroxidase (JGI ID 165193) was seen (Figure 4.2). On the other hand, the cytosolic PRX2 was down-regulated, possibly showing that less antioxidant capacity was needed in the cytosol compared to the chloroplast. Further interesting variations included the up-regulation of the regulatory factor 14-3-3 like protein FTT1, which is involved in diverse signaling processes, the topoisomerase SPO11a and the septin-like protein SEP1 (Figure 4.3). These indicate that HNF may also have an effect on cell development processes, similar to that observed in DR exposures.

In the LNF condition the only significant response found on the physiological and biochemical levels was a slight reduction of the oxidative activity, while no reduction of carotenoid content was detected, even after the longer exposure period of 24 h [21]. Due to this, the observed strong impact on many cellular systems and processes on the proteome level (Table 4.1) was not expected. The detected disturbances suggest that even without depletion of carotenoids, the exposure still produced significant alterations in the proteome that matched or even exceeded those of the more severe exposure to a higher herbicide concentration.

From the enzymes involved in carotenoid synthesis, only two were significantly affected by the LNF exposure: up-regulated DXS and down-regulated 4-hydroxy-3-methylbut-2-enyl diphosphate reductase IDS (Figure 4.1). The two-fold up-regulation of DXS was similar to that observed in HNF, while a significant change of the IDS expression level was seen only in LNF. Other major changes included the up-regulation of various components of photosynthesis, particularly within the light reaction group where several subunits of PS I and PS II, as well as light-harvesting proteins were affected (Table 4.1, supplementary file 2). The increased levels of the carbonic anhydrases CAH1 and CAH4 (Figure 4.3) may indicate a response of the carbon concentrating mechanism. The observation of significantly increased counts of VIPP1 (vesicle inducing protein in plastids) further suggests that LNF exposure may have an effect on thylakoid formation [57].

The functional group “protein synthesis” was widely affected as well. A particularly large number of ribosomal subunit proteins showed a differential expression upon exposure to the low NF concentration (Table 4.1). Furthermore, five proteins annotated to the “protein targeting” group were down-regulated, including SEC23A and SEC24A (Figure 4.3), which are subunits of a protein complex involved in protein
trafficking from the endoplasmic reticulum to the Golgi apparatus [58]. Direct perturbation of protein synthesis and trafficking by LNF could be an important cause of the extensive proteome variations observed.

As expected based on the previous study, LNF had a significant stimulating impact on a number of redox related proteins, particularly on the chloroplastic thioredoxins TRXm and TRXf2 as well as the cytosolic TRXH (Figure 4.2). TRXf2 is known to interact with the Calvin cycle enzyme fructose-1,6-bisphosphate phosphatase FBP1 [45], which was one of the significantly up-regulated proteins involved in carbon fixation (Figure 4.1). Also FSD1 and PRX1 levels were elevated (Figure 4.2) showing an increased antioxidative capacity within the herbicide targeted chloroplast. Other changes such as the up-regulation of the mitochondrial PRX7 and the down-regulation of MSD2 indicate that NF exposure had an effect on the redox condition of other cellular compartments as well.

Among the low concentration conditions used in this study, LNF produced the strongest impact on the proteome, showing that even very small disturbances on higher biological levels can be accompanied by extensive changes in the underlying molecular pathways. In both conditions investigated for the carotenoid synthesis inhibitor NF, no particularly intense response of the carotenoid synthesis pathway was observed compared to the effects on the same system observed in PQ and DR exposures. Therefore, the newly identified proteins and functional groups that did respond to the NF exposures may represent more suitable targets for further exploration of the subcellular effects of this herbicide.
4.5 Concluding remarks

In this study we used MudPIT as a tool for investigating ecotoxicologically relevant questions by characterizing protein abundance changes caused by toxicants. A great complexity of proteome variations in *C. reinhardtii* exposed to the herbicides PQ, DR, and NF was revealed, which provided novel insights into the organism’s response to herbicide exposure. We found that not only the chloroplast, which is the primary target site of the applied herbicides, but also many other cellular systems were strongly affected.

These substantial impacts on the proteome were observed at a relatively early time point, and also at the low herbicide concentrations, where biochemical and physiological endpoints as well as growth were not yet affected at all or were only slightly inhibited. This demonstrates the responsiveness of the proteome to stressors and highlights the great sensitivity of the technique. Similar observations were made when analyzing the effects of dissolved silver on the proteome of *C. reinhardtii* (Smitha Pillai, personal communication). When comparing the results of control, low and high concentration conditions of each herbicide, only occasional dose-related responses were observed. In fact, each exposure condition appeared to be rather unique, probably due to both, the differences in the mechanisms of action as well as the different exposure concentrations applied, resulting in mild or severe stress conditions. Common trends between the different herbicide exposure conditions were found among higher level functional protein groups such as the components of photosynthesis or ribosomal subunits, and rarely among individual proteins.

The proteins which responded to one or several exposure conditions could potentially serve as markers of herbicide exposure. We have observed and confirmed expected responses for known markers such as glutathione S-transferases, the various thioredoxins, or the components of the Calvin cycle. Interesting candidates for novel markers identified in this study were, for example, the PS II subunit PsbR, the VIPP1 protein or certain predicted transcription factors. These proteins will have to be validated and studied further, to explore their role in stress responses and mechanisms of toxicity in green algae. This could be done, for example, by using targeted proteomics, which provides greater sensitivity and higher sample throughput, thus allowing us to monitor many individual proteins over more concentrations, time points, toxicants, or species. Due to the inherent increase in sensitivity, this method could also help in investigating potentially interesting lower abundance proteins barely covered by the limited dynamic range of global proteomics, such as certain components of pigment synthesis pathways or certain transcription factors. A part of the molecular effects found on the proteome level could be linked to responses observed on the physiological and biochemical levels. This was pronounced most extensively in the HPQ condition, where the mechanism of action was thus reflected best. However, especially in the low concentration conditions, a great part of the observed proteome responses needs yet to be investigated further, particularly in respect to their magnitude and time-dependence, since many responses could be transient. So far, the effects of herbicides on the transcriptome and proteome of green algae have only scarcely been investigated. The issue is further complicated by the variety of toxicants, experimental exposure conditions, concentrations and durations as well as test species that have been used. Also, contrary to the responses to
high light, metals and other naturally occurring stressors which have developed through evolution, exposure
to (anthropogenic) herbicides can be assumed to be unnatural, possibly with the exception of the PQ induced
rise of superoxide radical levels. This may contribute to the difficulty of evaluating observed responses. The
complex interactions within the proteins and with other cellular components are another crucial factor that
could make it difficult to explain observed effects in the context of a response to exposures as direct and
downstream responses within this network cannot be distinguished. Clarification of these issues could be
achieved by integrating proteomics and other molecular data into system biology models [1].

We showed that proteome profiling using MudPIT is a promising approach for molecular research in
ecotoxicology, with a potential for screening for protein markers of exposure in *C. reinhardtii*, ultimately
leading to characterization of toxic mechanisms of action. However, future research is needed to deepen the
understanding of the responses and dynamics of the proteome patterns in order to link changes in protein
expression to mechanisms of action and physiological/biochemical responses, facilitating the elucidation and
prediction of toxicity in green algae.

**Acknowledgments**

We would like to thank Bettina Wagner for her support in *Chlamydomonas* cultivation, Victor J.
Nesatyy for technical support and Smitha Pillai for helpful discussions. This work was funded by Eawag.

**Appendix: Supplementary data**

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2012.06.017.
4.6 References


112


Chapter 5:

Conclusions and outlook
Anthropogenic compounds, such as pesticides, may exert unwanted effects in the environment by negatively affecting a variety of potential non-target organisms, such as the ecologically important green algae. Therefore, it is important to carry out a comprehensive assessment of the associated risks. To investigate the potentially detrimental effects of man-made toxicants on green algae, standardized toxicity tests based on growth inhibition (ISO, 2012; OECD, 2011) have been used traditionally. This endpoint is highly relevant ecologically but does not deliver any information regarding underlying effects and toxic mechanisms which is required to deal with the current challenges in ecotoxicology. To address this limitation, toxicant exposure effects need to be explored in more detail on lower levels of biological organization. This includes the assessment of specific cell physiology and biochemistry parameters, as well as characterization of changes occurring on the molecular level. The latter may be accessed through “omics” methods, which are now increasingly applied in ecotoxicology (Lemos et al., 2010; Schirmer et al., 2010).

The general aim of this thesis was to perform a mechanism-oriented exploration of herbicide exposure effects in green algae, and to evaluate the usefulness and informative value added to the more common simple measurement of growth inhibition by examination of a range of other, physiological and biochemical, parameters, as well as proteome alterations. For this purpose, two distinct research approaches were used to study the effects of the selected model herbicides PQ, DR and NF on the chosen model organism, the green alga *Chlamydomonas reinhardtii*: (1) a novel multiple-endpoint assay and (2) differential proteomic profiling.

### 5.1 Assessment of the effects of herbicides on physiological and biochemical parameters

For the multiple-endpoint assay, determination of growth was supplemented with the measurement of photosynthetic pigments (chlorophylls *a* and *b* and carotenoids), maximum and effective quantum yield of PS II, ATP levels, esterase and oxidative activity. Simple and fast setups, mainly based on multiwell plates, were used to assess these different physiological and biochemical endpoints simultaneously and with high throughput. This allowed obtaining a detailed time- and concentration-resolved view of the herbicide effects on algal cells and thorough evaluation of the selected cellular parameters.

The use of a wider range of toxicity endpoints based on different cellular parameters was expected to increase the likelihood of detecting exposure induced disturbances at earlier time points as well as at lower toxicant concentrations. Indeed, certain physiological and biochemical endpoints responded to the presence of herbicides very early and / or more strongly as compared to other endpoints as well as to growth which proved the benefit of combining the assessment of different cellular parameters in an assay battery. Due to their high sensitivity, some endpoints may serve as suitable markers of early responses to exposures, for example oxidative activity and effective quantum yield, and may become part of a screening system providing early warning on the presence and effects of toxicants. The assessed endpoints showed characteristic response patterns specific to each tested herbicide, which reflected the mechanisms of action of PQ, DR and NF. Changes in the patterns of responses over time provided further insights into the
toxicodynamics of these herbicides, their toxic mechanisms and into the development of underlying effects over the three sampling time points. With respect to the overall impact on growth, connections to underlying and preceding responses of physiological and biochemical endpoints could be drawn. This confirmed the usefulness of the presented multiple-endpoint approach for prediction of growth effects. The outlined points overall highlight the added value, also in regard to risk assessment of toxicants, brought by assessment of various physiological and biochemical endpoints, allowing for a mechanism-oriented analysis in contrast to tests solely relying on single endpoints such as growth. Thus, the presented multiple-endpoint assay was shown to be a valuable mechanism-based tool for studying toxicant as well as other stressor effects. It should be realized, however, that this work only covered a relatively small number of compounds and endpoints. As there is a great diversity of stressors, toxicants and mechanisms of action, already within the herbicides (Wakabayashi and Böger, 2004a; Wakabayashi and Böger, 2004b), it is expected that further physiological and biochemical endpoints will be needed to cover the whole range of potential effects, related with the variety of toxic mechanisms and not covered by the selected compounds.

5.2 Assessment of the effects of herbicides on the protein level

The proteomics profiling by MudPIT performed in the frame of this thesis required the establishment of an extensive computational pipeline for the analysis of the generated data as part of this work which ultimately allowed data evaluation in biological context. The pipeline was largely founded in a comparatively simple scripting language, which was essential to perform and / or automate the various steps in the analysis pipeline from raw MS/MS data workup to statistical evaluation. MudPIT-based profiling resulted in a coverage of the *C. reinhardtii* proteome comparable to other recent studies characterizing the whole proteome of this organism (Mühlhaus et al., 2011; Wienkoop et al., 2010). Thus, the desired comprehensive overview of the biological network was achieved which provided the sound basis for the analysis of toxicant-induced variations. In combination with spectral counts quantitation, G-test and functional enrichment analysis, significantly affected proteins as well as functional protein groups were identified for the low- and high-concentration exposure conditions for each of the three herbicides PQ, DR, and NF. Generally, highly complex effects of the various exposures were detected on the protein level represented by 149-254 significantly changed proteins. These included many proteins localized outside the chloroplast highlighting the cell-wide downstream impacts of the tested herbicides whose primary targets are the chloroplast and photosynthesis (Duke, 1990; Hess, 2000).

Two major traits are characteristic for the proteome response to exposures: high sensitivity, meaning that the proteome is a strong and fast responding indicator for occurring exposures even at minor exposure induced stress conditions, and high specificity (Lemos et al., 2010; Monsinjon and Knigge, 2007; Nesatyy and Suter, 2008), which would allow differentiating between varying stress conditions provoked by different toxic mechanisms and toxicant concentrations. In the present study, significant proteome changes were observed even in the low herbicide concentration conditions, which provoked no or only subtle effects on the
various parameters compiled in the multiple-endpoint assay. Since the sampling for the protein analysis also took place after a relatively short exposure duration, both aspects of the proteome sensitivity in response to exposures were demonstrated. In all conditions various proteins were identified which responded with either very high or very low fold changes thus suggesting these as sensitive potential marker proteins that possibly play crucial roles in the cellular response to the investigated herbicide exposures. Some variations on functional protein groups as well as individual proteins were found across all or several of the exposure conditions, such as proteins and protein groups involved in photosynthesis or protein synthesis. These could indicate general responses to toxicant induced stress or chloroplast targeting herbicides. In contrast to common shared alterations, for many other proteins changes in their abundance were specific to particular tested conditions revealing effect / response patterns in the present study which could be seen as complementary to those provided by the multiple-endpoint assay. These included the redox-related proteins and proteins regulating transcription or translation. Thus, the expected high specificity of the proteome response was also demonstrated.

Proteome profiling as used in this work does not require *a priori* hypotheses and instead can lead to discovery of yet unknown candidate exposure marker proteins and functional protein groups, e.g. certain metabolic pathways. Such findings may support elucidation of toxic mechanisms (Lemos et al., 2010; López-Barea and Gómez-Ariza, 2006). Indeed, besides expected proteins and pathways, a number of novel ones, which have not been implicated in response to the exposures to the selected herbicides before, have been found to change significantly upon exposure to herbicides in the present study. Overall, this provided new insights into the exposure induced responses on the molecular level of the green alga.

Links between the known mechanisms of action of the respective herbicides, their effects on the physiological and biochemical parameters on one side, and their effects on the proteome on the other, were proposed or could be established in some instances. Thus, the potential of proteome profiling for exploration of underlying effects and toxic mechanisms of the investigated herbicides was demonstrated. However, a substantial proportion of the observed alterations in the proteome could not be explained based on the current knowledge, and thus requires further research. Future studies, building upon the findings of the present one, should shed light onto the cellular functions and regulation of these proteins as well as their specific roles in stress response and in mechanisms of toxicity of PQ, DR, NF or different stressors in *C. reinhardtii*, possibly validating them as markers of exposure.
5.3 Outlook

Based on the demonstrated sensitivity and specificity aspects and the obtained insight into toxic mechanisms and effects of the tested compounds underlying growth inhibition, further research using the multiple-endpoint assay approach should focus on broadening the array of the investigated stressors. This will provide an overview over the effects of agents acting through different mechanisms. The stressors to investigate may include organic phytotoxicants of different structure, heavy metals and other chemicals acting via diverse, possibly yet uncharacterized, mechanisms of action. The effects of environmental stressors such as nutrient deprivation, temperature, high light or UV radiation, may also be investigated. The application of the presented assay may be particularly useful for research on chemicals for which the underlying toxic mechanisms and effects on green algae are unknown, at least at the initial stage where it may help formulating hypotheses to be investigated further. Furthermore, the assay also provides an excellent basis for mechanistic investigation of the effects caused by combinations of stressors, including chemical mixtures or multiple stressors of different nature such as chemicals and UV radiation. This may allow a better understanding of real-world exposure scenarios as single stressor conditions typically assessed in the laboratory setting up to now are untypical for natural freshwater systems (Eggen et al., 2004).

Another direction of future work could be an in-depth investigation of exposure-caused disturbances picked up through the multiple-endpoint assay. This would provide a more detailed characterization of toxic mechanisms underlying the responses of specific endpoints of interest. This kind of investigations may, however, require more complicated experimental setups. Some of the possible approaches are briefly outlined in the following two paragraphs.

Among the parameters tested in the multiple-endpoint assay in the present work, cellular oxidative activity proved to be the most sensitive. Since the method used for assessment of oxidative activity was essentially an integrative measurement, specific underlying processes, particularly formation of the various reactive species (particularly ROS) as well as their elimination by diverse antioxidant defense mechanisms, remained uncharacterized, although some hints into this direction were delivered by the proteomics analysis. Resolving the actual activity of antioxidative defense enzymes may prove useful, as it was already shown to contribute to the understanding of herbicide induced disturbances in photosynthetic organisms (Dewez et al., 2005; Geoffroy et al., 2004; Radetski et al., 2000). The characterization of formation of specific reactive species, which addresses the other side of the equation, does not appear to be widely used in algal research so far, although it may further add to the elucidation of mechanisms of toxicity. For example, using such approaches in previous studies, singlet oxygen and superoxide radicals were measured upon exposure to herbicides, including PQ and DR, as well as to high light and UV radiation (Fischer et al., 2006; Hideg and Schreiber, 2007; Hideg et al., 2002).

Analysis of chlorophyll fluorescence signals of PS II has significantly more potential for elucidating the effects on the photosynthetic machinery than that used in the present study (Juneau et al., 2007; Maxwell and Johnson, 2000; Ralph et al., 2007). For more detailed measurements, however, an instrumentation more
sensitive than the imaging-PAM, whose prime advantage is the fast measuring of the large numbers of samples simultaneously, is required. Then, several parameters other than the simple but robust PS II quantum yields may be considered, such as non-photochemical quenching. Similarly, ATP levels as simple indicator for the energetic state of the cells could be the starting point from where the additional information on processes influencing the cellular energy situation could be gathered. Further research may include determination of the levels of the nucleotides AMP, ADP, ATP followed by calculation of an “energy charge” parameter and the ATP/ADP ratio or monitoring of NADP and NADPH as performed previously (Forti et al., 2003; Peixoto et al., 2008).

Similarly to the strategy outlined for the multiple-endpoint assay, the high sensitivity and specificity of the proteome responses detectable by proteome profiling should be exploited by broadening the spectrum of investigated herbicides and other toxicants and stressors. Future studies should also address the effects of additional exposure concentrations thus allowing a reliable exploration of dose-response-relationships. The development of response patterns over time also needs to be addressed by performing the proteome profiling after shorter and longer exposure durations. This may allow exploration of the dynamics of the proteome responses and distinguishing between direct and downstream responses which, in the field of transcriptomics, have been termed “initial signature of exposure” and “signature of effect”, respectively (van Straalen and Feder, 2012). Direct responses occur early after exposure and thus are characteristic of the toxicant itself, while downstream responses occur later and reflect the induced effects. Analyzing the protein response patterns obtained in such future studies would further strengthen the linking between protein responses and the exposure induced stress as well as the mechanisms of action and advance the mechanistic understanding of subcellular exposure effects. This could also lead to generalizations of observed proteome responses. that is, elucidate which proteins and protein groups are key players in these toxic responses. Ultimately, characterization of such patterns of protein expression changes under standardized experimental conditions may at some point support inferring mechanisms of action and thus may greatly help in elucidating toxic mechanisms of yet uncharacterized toxicants.

Extending the understanding of molecular effects of herbicides in their entirety could further profit from integration of proteomic and other molecular data obtained from transcriptomics and metabolomics in system biology models (Garcia-Reyero and Perkins, 2011; Jamers et al., 2009). The result would be an extensive data repository which includes the known relations and interactions between the various biomolecules and cellular processes and thus provides a comprehensive picture and model of the complex effects within the whole cellular network.

It has to be noted that global proteomic profiling analyses still suffer from several important limitations. For instance, a complete proteome profiling cannot be achieved as of yet (Lemos et al., 2010), meaning that detection and quantification (and thus exploration of significant differences) can be reliably performed only for the relatively highly abundant proteins. Although these may also be directly influenced by the toxicant, lower abundance proteins which are masked by higher abundance ones could have been more relevant to the exposure response, for example when considering transcription and translation
regulation, or cellular signaling. Another important limiting factor is the amount of labor and time required by most proteome profiling approaches, including MudPIT, which strongly limits the number of samples that can be processed. Such analytical bottlenecks and other issues could be addressed by the application of “targeted” proteomics (Nesatyy and Suter, 2008; Yocum and Chinnaiyan, 2009). This approach does not attempt to profile the whole proteome but instead concentrates on monitoring selected peptides for a limited number of proteins of interest thus allowing a much higher sample throughput. Due to the increased sensitivity, this technique has a much higher potential for investigation of lower abundance proteins which often are barely covered by the global proteomics methods.

Other proteomics approaches have also emerged which go beyond the mere determination of protein abundance levels. These methods have not been applied in ecotoxicological research so far, but may have great potential in addressing certain questions. One such example could be assessment of the protein turnover rates. Only recently, genome-wide assessment of turnover rates of proteins additional to levels was reported (Schwanhäusser et al., 2011). Since the levels of the various proteins in an organism are determined by the rates of synthesis and degradation, it is possible that these underlying processes may reflect the molecular response of an exposed organism much better than the abundance levels alone. Another useful approach could be the so-called “activity based proteome profiling” where low-molecular weight and active-site directed probes are used to determine the activity of functional enzymes in biological systems (Lemos et al., 2010). Applying this approach would be highly interesting for assessing the response of organisms to exposure, particularly since it is independent of available sequence information and thus will not be restricted to model organisms only.
5.4 Summarizing remarks

The toxic mechanism oriented approaches applied and evaluated in this work, the multiple-endpoint assay and proteome profiling using MudPIT, both provided the desired highly detailed picture of underlying effects induced by the model herbicides PQ, DR and NF in *C. reinhardtii*. Links between growth, different physiological and biochemical endpoints and the protein level could be established. Furthermore, connections between a part of these effects and the toxic mechanisms of action of the three compounds were found or proposed. Inclusion of low herbicide concentrations inducing only mild stresses allowed identification of sensitively responding endpoints and markers of exposures on the physiological and biochemical as well as the protein level, demonstrating one of the major benefits of these approaches over growth based toxicity assays. The higher specificity of such endpoints, particularly of the proteome level, was shown as well which allowed the differentiation of investigated stress conditions and, to some extent, toxic mechanisms of action based on observed variations. Thus, both approaches applied in this work – particularly when being combined – have demonstrated their merits and potentials as tools of a mechanism oriented analysis to advance the characterization and understanding of toxicant effects and toxicity mechanisms in green algae.

The highlighted points also suggest the wider application of the presented tools for addressing the current challenges in ecotoxicology that require a mechanistic understanding of toxicant effects. Ultimately, the added information gained by such mechanism-oriented approaches will support risk assessment of anthropogenic toxicants affecting green algae. Particularly proteomics, when firmly anchored in effects on higher levels of biological organization and utilized for further investigations in ecotoxicological context, can be expected to increasingly contribute to exploration of toxic mechanisms and prediction of toxicant and other stressor effects in green algae as well as in other organisms.
5.5 References


Acknowledgements

In these years of my doctoral studies at Eawag a lot of people were supportive of the project and me in many ways - in the lab, the learning and writing process, were giving moral support as well as friendship, and provided fun and distraction. Many thanks to all of them for making this time an unforgettable and valuable experience.

First of all I want to thank Marc Suter and Rik Eggen for their supervision, guidance, feedback, discussions and patience, and for giving me the opportunity to do this work. Furthermore, Jukka Jokela is thanked for being referee to this thesis and for the helpful discussions particularly in statistical issues. I am grateful to Mechthild Schmitt-Jansen for having accepted to be the external referee on the examination board. Martin Ackermann is acknowledged for chairing the defense.

A great deal of the lab work was done with the dedicated and active support of René Schönenberger. Thank you for that and for everything else! I am greatly indebted to Ksenia Groh for fruitful discussions and especially for guidance during the final writing up of the chapters. Renata Behra and Kristin Schirmer are thanked for many helpful discussions and particularly for their contributions to chapter 2 of this thesis. Two important foundations of this work were to keep *Chlamydomonas* alive and happy, as well as to get MudPIT to work. I want to acknowledge Bettina Wagner’s and Victor Nesatyy’s support for these tasks.

Many more current and former members of the Utox department provided help in the lab, with everyday problems, contributed ideas, helpful discussions or feedback: Adrian Ammann, Roman Ashauer, Nadine Bramaz, Regine Dayer, Beate Escher, Beat Fischer, Stephan Fischer, David Kistler, Muris Korkaric, Alexandra Kroll, Danielle Madureira, Enrique Navarro, Judith Neuwöhner, Christina Otto, Franziska Pfister, Flavio Piccapietra, Smitha Pillai, Karin Rüfenacht, Christian Scheidegger, Laura Sigg, Ilona Szivák, Katrin Tanneberger and Nathalie Vallotton. The Utox department was a great working environment as well as great company to be around with.

I also want to acknowledge the help and support of members of the Uchem department, particularly of Martin Krauss and Heinz Singer in the various orbitrap-related issues. From Umik, Thomas Fleischmann, Christoph Werlen and Hans-Ueli Weilenmann are thanked for lab-work advice and running the autoclave, and Marius Vital for providing the initial procedure and material for the ATP assay. Jafet Andersson helped me with implementing R in the statistical evaluation of the spectral counts. I am also grateful to Marion Junghans and Robert Kase of the Ecotox Centre for their advice regarding toxicity test evaluation and related topics. For supporting René and me with constructing the algae exposure facility (as well as various other tasks) I thank Peter Gäumann and his colleagues at the Eawag workshop as well as Erich Eschmann from the technical support at EMPA. Eawag is thanked for funding the project. Carsten Krantz from Thermo Fisher Scientific is acknowledged for advice on various Xcalibur and MS related issues. Marc Moniatte from EPFL gave the crucial hint on how to get cysteine containing peptides into the analysis.
Furthermore, I want to thank some people who were not connected to the project itself but nonetheless greatly contributed to make being in and around Eawag worthwhile (roughly in order of acquaintance): Simón, Miguel, Andreas, Tobias, Mirjam, Niksa, Franziska, Merle, Corine, Eva, Mauricio, Colette, Stefan, Ines, Lukas, Thomas, Anna, Ruedi, Johannes, Aurea, Thea, Anita, Melanie, Jerry, Verena, Anni, Julita, Severin, Lena, Carmen, Carl, Fred, Yang, Petra, Barbara, Michal and Anita. A great time was also playing with the EMPA-Eawag unihockey people as well as joining the fire brigade for the, unfortunately, last year of its existence. More thanks go to many Bergstrasse 55 dwellers, to the guys from z&h for many enjoyable evenings, to the ZVV - especially the bus drivers of line 751 - for getting me to Eawag, back home and anywhere else in and around Zurich, as well as to the ASVZ for offering such excellent possibilities for sport.

Last but not least, I want to express my thanks to my parents for welcoming me back to their home every time I returned to the Heimat, and to my brother Falk and cousin Sandy for helping to keep up my morale.
Curriculum vitae

Holger Nestler born on August 11, 1981 in Zschopau, Germany

2006-2012 PhD in the Department of Environmental Toxicology at the Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf / Swiss Federal Institute of Technology (ETHZ), Zurich, Switzerland

2005-2006 Diploma thesis

Untersuchungen zum mikroaeroben Abbau von Monochlorbenzen

2004-2006 Intern, diploma student, research assistant at the Department of Hydrogeology, Helmholtz Centre for Environmental Research - UFZ, Halle / Saale, Germany

2000-2006 Study of geocology at the TU Bergakademie Freiberg, Germany

Specialization in environmental analytics / environmental geochemistry, environmental microbiology / biotechnology, pedology