ABSTRACT


Mass spectrometry (MS) based proteomics is an emerging field in analyzing biological systems. Fast scan cycles, high mass measurement accuracies and low detection limits elevate mass spectrometers to be the instrument of choice in dealing with the overwhelming complexities of proteomes. Front-end fractionation/separation methods aid in facilitating sample complexities and automated data analysis software assist in obtaining thousands of protein identifications at the back-end and come in handy especially for global proteome analyses. The work presented herein focuses on the improvement of label free proteomic strategies while examining an agriculturally important disease causing fungus *Magnaporthe oryzae* (*M. oryzae)*.

Offline reverse phase (RP) chromatography in combination with online RP-MS was found to increase proteome coverage by 61.6% for the model fungal organism yeast (*Saccharomyces cerevisiae*) when compared to one dimensional (1D) RP-MS. The simplicity of this fractionation method lies in the same mobile phases applied in both dimensions consisting of acetonitrile/water mixtures and the ion pairing reagent formic acid. Ten fractions were created from a whole yeast digest while a 30 min short gradient was applied offline. Those ten fractions were then run each with individual gradients for 60 min on the online *nanoLC*-MS. Protein identifications from the fractionated sample (1028) were compared to protein identifications yielded by running a 60 min gradient for the whole digest (636). The fact that more total sample could be loaded on column in combination with the narrow gradients allowed for higher proteome coverage with the 2D-LC method.
A more established approach 1D-gel electrophoresis combined with *nanoLC-MS* (GeLC) was applied on *M. oryzae conidia* and three different normalization methods were evaluated to correct for analytical variation in three *spectral counting* data sets. A total of 1511 proteins were identified in this study. Myoglobin and ovalbumin, two exogenous proteins, were spiked into the *M. oryzae* protein mixture to perform normalization to specific proteins (NSP). The normalization methods normalized spectral abundance factor (NSAF) and total spectral count normalization (TSpCN) performed well and could account for the analytical variation between the samples; however NSP was not effective in correcting for the variation. Furthermore fold-changes and t-testing were investigated for detecting real biological change and it was shown that fold-changes were more difficult to observe for highly abundant proteins (high spectral counts), whereas statistically significant change was more difficult to observe for low abundant proteins (low spectral counts). A combination of both approaches with appropriate cutoff-values for spectral count bin width, rather than single cutoff-values for the entire data set were proposed.

Filter aided sample preparation (FASP) in combination with anion StageTip fractionation (6 fractions) was applied on *M. oryzae conidia* and yielded identification of 2912 proteins. Long 210 min 5-35% B gradients were run and an in-depth characterization of the proteins identified was performed. The same approach was applied for the longitudinal study from conidia over germinated conidia to appressorium formed conidia. Differences of wild-type to the cpkA mutant were also analyzed. Protein regulations were observed between time points in the life cycle. Further absolute quantification on 4 selected proteins with interesting behavior was performed by protein cleavage isotope dilution mass spectrometry (PC-IDMS) and confirmed the patterns detected in the global study.
Lastly the ubiquitome was investigated as post translational modification in *M. oryzae mycelium*. Agarose coupled TUBEs (Tandem Ubiquitin Binding Entities) were used for ubiquitinated protein enrichment and enriched samples were digested using the filter aided sample preparation (FASP) procedure. Under nitrogen starvation 63 candidate polyubiquitinated proteins were identified including overrepresentation of proteins involved in translation, transport and protein modification and suggesting that ubiquitination of target proteins plays an important role in nutrient assimilation, development and pathogenicity of *M. oryzae*. 
Development, Evaluation and Application of Proteomic Strategies Towards Understanding the Biology of the Rice Blast Fungus

Magnaporthe oryzae

by

Emine Gokce

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chemistry

Raleigh, North Carolina

2012

APPROVED BY:

__________________________________
David C. Muddiman
Professor, Chemistry
Committee Chair

_______________________________
Ralph A. Dean
Professor, Plant Pathology

________________________________
Edmond F. Bowden
Professor, Chemistry

_______________________________
Morteza G. Khaledi
Professor, Chemistry
DEDICATION

I would like to dedicate this thesis to my one and only, beloved and caring husband Bilal Gokce. Mein Schatz, your love gives me strength and your patience with me is what makes “us” possible. Together there won’t be a thing we will start but not finish inshallah.

Thank you Allah’im for making Bilal my kismet. I always knew what I wanted from life, but he is a gift that I could have never dreamed of. I am so thankful for him…
Emine Gokce is the daughter of Sefika and Samih Alkan, born on October 17th 1986 in Krefeld, Germany. Having the privilege of being raised between the Turkish and the German culture has made Emine the open minded and social person she is today. Although being the baby of the family came with the price of high expectations regarding her performances, Emine always delivered. The love she developed for chemistry started with Mr. Popovic at Fichte Gymnasium in Krefeld and ingrained itself during her studies at Heinrich Heine University in Dusseldorf. Emine graduated with her Diplom degree in June 2009 from Dusseldorf. Dr. Weinkauf, Dr. Schaper, Dr. Schmidt, Dr. Bettermann, Dr. Beutner, Dr. Mayer and Dr. Kläui, are all professors that have taught her very well and prepared her at best for the journey of the doctoral degree in the United States that she started together with her husband Dipl. Physicist Bilal Gokce. With Dr. David C. Muddiman as an outstanding advisor at the Chemistry Department of North Carolina State University Emine performed research in the field of proteomics during her graduate school career.
ACKNOWLEDGMENTS

Thank you Dr. Muddiman, for giving me the opportunity to be part of your magnificent research group. Thank you for guiding me throughout graduate school and thank you for being the great mentor that you are. I have learned so much from you and from all the members in the group: Tim, Chris, Genna, Hunter, Shan, Angelito, Jeremy, Guillaume, Amber, Zhi-Chang and Phil. Thank you all so much…

Thank you Dr. Dean, for your support. Your encouragement has always excited me and led me to work harder. Thank you Dr. Oh and Dr. Franck for insights about the biological part of the project, all the culturing work and the informative meetings.

Thank you Dr. Bowden and thank you Dr. Khaledi for being part of my committee and the lively discussions about my project.

And finally: Thank you to my family in Germany for all the trust they had in me.
# TABLE OF CONTENTS

**LIST OF TABLES** .......................................................................................................................... x

**LIST OF FIGURES** .......................................................................................................................... xi

**LIST OF PUBLICATIONS** ................................................................................................................. xii

**CHAPTER 1: Introduction**

1.1 The Rice Blast Disease *Magnaporthe oryzae* .............................................................................. 1

1.2 Shotgun Proteomics .......................................................................................................................... 6

   1.2.1 Fractionation and Separation Methods Prior to Mass Spectrometry .............................. 6

   1.2.2 Electrospray Ionization Mass Spectrometry .................................................................... 13

      1.2.2.1 Hybrid Mass Spectrometer ....................................................................................... 15

   1.2.3 Data Analysis ......................................................................................................................... 19

1.3 Synopsis of Completed Research ................................................................................................. 22

1.4 References ..................................................................................................................................... 24

**CHAPTER 2: Increasing Proteome Coverage with Offline RP HPLC Coupled to Online RP NanoLC–MS** ....................................................................................................................... 37

2.1 Introduction ..................................................................................................................................... 37

2.2 Experimental Section ..................................................................................................................... 38

   2.2.1 Sample Preparation ............................................................................................................... 38

   2.2.2 Offline RP HPLC .................................................................................................................. 39

   2.2.3 NanoLC-MS ........................................................................................................................ 41

   2.2.4 Data Analysis ....................................................................................................................... 42
### CHAPTER 3: Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation During Proteomic Workflows 

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>52</td>
</tr>
<tr>
<td>3.2 Experimental Section</td>
<td>55</td>
</tr>
<tr>
<td>3.2.1 Sample Preparation</td>
<td>55</td>
</tr>
<tr>
<td>3.2.2 NanoLC-MS/MS</td>
<td>57</td>
</tr>
<tr>
<td>3.2.3 Data Analysis</td>
<td>58</td>
</tr>
<tr>
<td>3.3 Results and Discussion</td>
<td>59</td>
</tr>
<tr>
<td>3.4 Conclusions</td>
<td>71</td>
</tr>
<tr>
<td>3.5 References</td>
<td>72</td>
</tr>
</tbody>
</table>

### CHAPTER 4: In-Depth Analysis of the *Magnaporthe oryzae* Conidial Proteome

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>76</td>
</tr>
<tr>
<td>4.2 Experimental</td>
<td>79</td>
</tr>
<tr>
<td>4.2.1 Sample Preparation and Digestion</td>
<td>79</td>
</tr>
<tr>
<td>4.2.2 StageTip Anion Fractionation</td>
<td>80</td>
</tr>
<tr>
<td>4.2.3 NanoLC-MS/MS</td>
<td>81</td>
</tr>
<tr>
<td>4.2.4 Data Analysis</td>
<td>82</td>
</tr>
<tr>
<td>4.3 Results and Discussion</td>
<td>83</td>
</tr>
<tr>
<td>4.3.1 Gradient Study</td>
<td>83</td>
</tr>
</tbody>
</table>
4.3.2 Reproducibility Experiment ................................................................. 86
4.3.3 Subcellular Localization Analysis ......................................................... 88
4.3.4 Gene Ontology Analysis ......................................................................... 90
4.3.5 Biological Insights into Identified Proteins ............................................. 92
4.4 Conclusions .............................................................................................. 96
4.5 References ............................................................................................... 97

CHAPTER 5: Characterization of the Magnaporthe oryzae Proteome During Conidial Germination and cAMP Mediated Appressorium Formation ........................................ 103

5.1 Introduction ............................................................................................. 103
5.2 Experimental Section ................................................................................ 107
  5.2.1 Strains and Culture Conditions .............................................................. 107
  5.2.2 Generation of ΔcPKA Mutant ................................................................. 107
  5.2.3 Sample Preparation ................................................................................ 108
  5.2.4 NanoLC MS/MS Analysis ....................................................................... 109
  5.2.5 Absolute Quantification ........................................................................ 110
  5.2.6 Data Analysis ........................................................................................ 111
5.3 Results and Discussion ............................................................................. 112
  5.3.1 Identification and Label Free Quantification of Proteins During Germination and Appressorium Formation ................................................................. 112
  5.3.2 Gene Ontology Analysis of Regulated Proteins ...................................... 118
  5.3.3 Extracellular Proteins Up-Regulated in Appressoria ............................ 118
  5.3.4 Cell Wall Biosynthesis and Remodeling .............................................. 121
5.3.5 Regulation and Transport Functions in Response to Conidial Germination and Appressorium Development .................................................................127
5.3.6 Regulation of Ribosomal Proteins During Conidial Germination .............131
5.3.7 Down Regulation of Mitochondrial Proteins During Germination ...........132
5.3.8 Protease Regulation ..............................................................................134
5.3.9 Regulation of Secondary Metabolism ....................................................135
5.3.10 Validation of the Global Proteomics Data via Absolute Quantification ....136
5.3.11 Comparison of the cAMP Responsive Transcriptome and Proteome ......140
5.4 Conclusions ..............................................................................................141
5.5 References ..................................................................................................142

CHAPTER 6: Polyubiquitin is Required for Growth, Development and Pathogenicity in the Rice Blast Fungus Magnaporthe oryzae .................................................................154

6.1 Introduction .................................................................................................154
6.2 Materials and Methods ...............................................................................156

6.2.1 Effect of Proteasome Inhibitor, Bortezomib, on M. oryzae ......................156
6.2.2 Prediction of Ubiquitin Associated Proteins in M. oryzae and their Transcription Analysis .................................................................................................156
6.2.3 Targeted Gene Replacement and Complementation of MGG_01282 .......157
6.2.4 Mutant Phenotype Assays .........................................................................158
6.2.5 Sexual Reproduction ................................................................................159
6.2.6 Western Blotting ......................................................................................159
6.2.7 Ubiquitinated Protein Profiling by NanoLC-MS/MS .................................160
6.3 Results..................................................................................................................................................161

6.3.1 Inhibition of Proteasome Mediated Protein Degradation Blocks Conidial Germination, Appressorium Induction and Pathogenicity in *M. oryzae*........161

6.3.2 Components of Ubiquitin Mediated Protein Modification are Highly Conserved and Regulated in Response to Developmental and Nutritional Stimuli in *M. oryzae*.................................................................................................................................163

6.3.3 Highly Conserved but Structurally Diverse Ubiquitin Genes in *M. oryzae* Show Different Expression Patterns during Development and Starvation...........166

6.3.4 Nitrogen Starvation Results in Dramatic Increase in Protein Ubiquitination which is Mediated by the Polyubiquitin Protein, MGG_01282......................169

6.3.5 The Polyubiquitin Protein MGG_01282 is Essential for Fungal Growth, Development and Pathogenicity.........................................................................................170

6.3.6 Ubiquitinated Protein Profiling During Fungal Growth under Nitrogen Stress176

6.4 Discussion..............................................................................................................................................177

6.5 References............................................................................................................................................181

GLOSSARY................................................................................................................................................186
# LIST OF TABLES

**Table 1.1** Functional Groups for Anion and Cation Exchange Resins ...........................................10
**Table 2.1** Calculated Peptide Concentrations per Fraction.................................................................40
**Table 2.2** One Hour Gradients of the Fractions .........................................................................................41
**Table 2.3** Peptide Overlaps between Fractions .........................................................................................46
**Table 3.1** False Positive Rates for SpC Bin Widths at Different Fold-Change Cutoffs ............67
**Table 3.2** False Positive Rates for SpC Bin Widths at Different p-Value Cutoffs ...................68
**Table 3.3** False Positive Rates for combined p-Value and Fold-Change Cutoffs Applied on Different SpC Bin Widths.................................................................................................................69
**Table 4.1** Number of Full Scans, MS/MS Scans, Identified Peptides, Calculated Peptide Yield and Identified Proteins for the Different Gradients Used for the Single Fraction at pH8........................................................................................................................................................................86
**Table 4.2** False Positive Rate Table (s: spectral counts per replicate, N: number of proteins) ........................................................................................................................................................................87
**Table 5.1** Number of Differentially Expressed Proteins for Each Comparison .....................115
**Table 5.2** Fold Changes of Proteins Up-Regulated in Response to cAMP Treatment and in the Wild Type When Compared to the ΔcpkA Mutant ..................................................................................116
**Table 5.3** Proteins Down-Regulated in Response to cAMP Treatment and in the Wild Type when Compared to the ΔcpkA Mutant ............................................................................................................117
**Table 6.1** Ubiquitin Pathway Associated Proteins in *M. oryzae* .........................................................163
LIST OF FIGURES

Figure 1.1 Asexual Disease Cycle of M. oryzae .................................................................1
Figure 1.2 Tricyclazole Acting as Melanin Biosynthesis Inhibitor .........................................2
Figure 1.3 Golden Rice in Comparison to White Rice ..........................................................3
Figure 1.4 Beta Carotene Biosynthesis and Subsequent Cleavage to Retinol .........................4
Figure 1.5 Most Rice Cultivating States in the US ...............................................................5
Figure 1.6 1D-Gel Image ......................................................................................................7
Figure 1.7 Strong Anion and Strong Cation Exchange Surface Interactions .........................9
Figure 1.8 Anion Exchange Resin Used in StageTips ..........................................................11
Figure 1.9 ESI-LTQ-FTICR and Orbitrap ............................................................................15
Figure 1.10 FTMS and MS/MS Spectra ...............................................................................17
Figure 1.11 PC-IDMS Approach on a Triple Quadrupole Instrument ................................19
Figure 2.1 Experimental Workflow .....................................................................................44
Figure 2.2 Representative Data Comparing the Complexity in Chromatography ............45
Figure 2.3 Venn Diagram Presenting Protein Identifications .............................................47
Figure 3.1 Experimental Workflow .....................................................................................59
Figure 3.2 Venn Diagrams Presenting Protein Identifications ............................................60
Figure 3.3 SpC Scatter Plots from Combined Analysis of Sample 1 and 2 .........................62
Figure 3.4 Unnormalized SpC Data and NSpCs from Combined Analysis of Sample 2 and Sample 2’ ........................................................................................................63
Figure 3.5 Volcano Plots for Comparing the Normalized SpC between Sample 1 and Sample 2......................................................................................................65

Figure 4.1 Schematic of FASP and StageTip Workflow.........................................................84

Figure 4.2 Total Ion Chromatograms and Frequency of Consecutive MS/MS Events ........85

Figure 4.3 Subcellular Localization Charts.............................................................................89

Figure 4.4 Gene Ontology Analysis..........................................................................................91

Figure 5.1 cAMP Induced Appressorium Formation in *M. oryzae* ..................................112

Figure 5.2 Venn Diagrams Presenting Protein Identifications .............................................113

Figure 5.3 Predicted Subcellular Localization of the *M. oryzae* Proteome .....................119

Figure 5.4 Absolute Quantification.........................................................................................139

Figure 6.1 Bortezomib Blocks Conidia Germination, Appressorium Formation and Pathogenicity.................................................................162

Figure 6.2 Ubiquitin Associated Genes .................................................................................165

Figure 6.3 Clustering and Gene Expression............................................................................168

Figure 6.4 Western Blots .......................................................................................................169

Figure 6.5 Growth Reduction Under Nitrogen Limiting Conditions.................................171

Figure 6.6 Conidia Production in Wild Type, Mutants, Ectopic and Complemented Strains ...........................................................................................................172

Figure 6.7 Abnormal Conidia Production in Mutants............................................................173

Figure 6.8 Appressorium Formation and Pathogenicity in Rice and Barley .......................174

Figure 6.9 Fruiting Bodies Formed in Each Test Strain ..........................................................175
LIST OF PUBLICATIONS

1. Franck W. L., Gokce E., Oh Y., Muddiman D. C., and Dean R. A. Characterization of the *Magnaporthe oryzae* Proteome During Conidial Germination and cAMP Mediated Appressorium Formation, submitted to *Molecular Cellular Proteomics*.


1.1 The Rice Blast Disease *Magnaporthe oryzae*

Rice, *Oryza sativa*, is the second-most worldwide produced crop after maize. White rice, brown rice, forbidden rice and golden rice; long, medium or short rice; more than thousands of rice varieties exist and are produced yearly around the globe. Rice is the primary nutritional source for 50% of the human population and “Have you had your rice today?” is a common greeting in China. Rice is served for breakfast, lunch and dinner in many parts of Asia. The crop can grow almost anywhere, although requiring somewhat more water than other cereal crops. However, rice is not the easiest to harvest and suffers from blast, a disease known for more than 3 centuries, the filamentous fungus *Magnaporthe oryzae* (*M. oryzae*). The rice blast disease likely originated in China and has spread over the centuries around the world\textsuperscript{1-3}.

As illustrated in Figure 1.1, the life cycle of *M. oryzae* begins with the attachment of an asexually produced

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{Figure1.png}
\caption{Asexual disease cycle of *M. oryzae*\textsuperscript{1}}
\end{figure}

\textsuperscript{1}
spore, known as *conidium*, to the cell wall of the host. After germination and formation of a specialized infection structure, the appressorium, the contents of the *conidium* empty into the appressorium in preparation for infection. Enormous *turgor pressure* develops within the melanin lined appressorium, which forces an infection peg into the host cell. One or two days after penetration the fungus begins to kill host cells, making it a *hemibiotrophic* fungus. Under favorable conditions, e.g. high fertilizing rates, drought stress, aerobic soils and no wind at nights, a single lesion can generate thousands of *conidia* per night and continues to produce *conidia* for over twenty days. The conidium-conidium life cycle can be completed within a week.

**Figure 1.2** Tricyclazole acting as melanin biosynthesis inhibitor. Acetate molecules form tetrahydroxynaphthalene by undergoing pentaketide cyclization. The conversion to scytalone can be inhibited by tricyclazole as can the conversion from trihydroxynaphthalene to vermelone. The genes ALB1, RYS1 and BUF1 were shown to control the polynaphthalene forming steps.
Managing the disease is challenging and expensive. Effective fungicides, such as tricyclazole, require at least 3-4 spray treatments\textsuperscript{5,6}. Tricyclazole acts by inhibiting the reduction steps during the melanin biosynthesis (see Figure 1.2).

The secondary metabolite and polyketide melanin had been shown to be essential for the fungus to create the \textit{turgor pressure} required to penetrate into the rice plants\textsuperscript{7-9}. The responsible genes for melanin production ALB1, RSY1 and BUF1 have been named by their nonpathogenic mutant phenotypes albino, rosy and buff respectively.

Other management techniques include resistant varieties or cultural techniques such as burning infected crop areas; however such approaches cannot guarantee a healthy yield either. Molecular genetic approaches are beginning to emerge with the perspective of creating varieties resistant to diseases with increased yield and quantity. Although debates are ongoing whether genetically modified food is safe or harmful, almost all soybeans (93\%)\textsuperscript{10,11}, corn (86\%)\textsuperscript{12,13}, canola (82\%)\textsuperscript{14} and sugar beets (95\%)\textsuperscript{15} cultivated in the US are genetically modified. The European Union requires special labels for genetically modified food whereas it is not obligatory in the US or Canada. The prospects

\textbf{Figure 1.3} Golden rice in comparison to white rice\textsuperscript{ii}. 
of using genetically modified crops in Asia is uncertain. The development of golden rice offers promise\textsuperscript{16}. However, the manipulation in golden rice is not to increase resistance against herbicides (\textit{glyphosate} and \textit{glufosinate}) as is the case for many genetically modified crops; it is in fact an insertion of the genes phytoene synthase from maize and carotene desaturase from the bacterium \textit{Erwinia uredovora} to accumulate beta carotene production, which gives the rice a golden color as shown in Figure 1.3\textsuperscript{17}. Beta carotene is the metabolic precursor which can be cleaved in the human body by the enzyme beta-carotene 15, 15'-monooxygenase into two molecules of vitamin A (retinol) (Figure 1.4)\textsuperscript{18}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{carotene_biosynthesis.png}
\caption{Beta carotene biosynthesis and subsequent cleavage to retinol. Phytoene is build out of two molecules of geranylgeranyl pyrophosphate by phytoene synthase and converted to \(\zeta\)-Carotene by phytoene desaturase. \(\zeta\)-Carotene desaturase functions in forming lycopene which finally becomes \(\beta\)-Carotene through lycopene cyclase. \(\beta\)-Carotene 15, 15'-monooxygenase can cleave \(\beta\)-Carotene to two molecules of Vitamin A (retinol) in the human body.}
\end{figure}
In first trials of golden rice the gene lycopene cyclase was inserted as well; however studies showed that lycopene cyclase necessary for the beta carotene biosynthesis pathway was already expressed in rice and the insertion could be omitted. The need for golden rice came from lack of Vitamin A in rice kernels resulting in Vitamin A deficiency causing death and eye sight loss when rice is the main nutrition source.

Rice cultivation in the United Stated started 300 years ago in the Carolinas and Georgia. Today the United States is one the 10 most rice producing countries in the world. Currently Arkansas (42.1%), California (26.2%), Louisiana (14.3%), Mississippi (5.8%), Missouri (4.5%) and Texas (7.0%) are the six main rice growing states (Figure 1.5). M. oryzae was reported in the US the first time in California in 1996 and presents a threat to rice production due to the increasing demand as a consequence of increasing population. The lack of biological and chemical understanding of

Figure 1.5 Statistics from 2009 indicate the six most rice cultivating states in the US. Arkansas with 42.1% of the total rice production in the US is followed by California and Louisiana. Texas, Mississippi and Missouri present the other three rice producing states.
host-pathogen interactions is hampering our ability to find an effective cure. The genome of both rice and rice blast have been sequenced; rice being completed in 2002 by Yu et al., and *M. oryzae* being completed in 2005 by Dean et al. Microarray experiments have revealed differentially expressed genes during appressorium development of *M. oryzae* in the following years; however, the need for proteomic analyses emerged as correlation of gene expression data and proteomics data were found to be poor in many other organisms and changes in phenotype might become more traceable when transcriptome level and protein level analyses are used to complement each other. Investigating high molecular level details of proteins and post translational modifications might therefore be significant for the development of targeted downstream control strategies. Once target genes are identified gene knockout or overexpression can be performed and silencing complete gene families through RNA interference is a promising alternative approach as well.

1.2 Shotgun Proteomics

1.2.1 Fractionation and Separation Methods Prior to Mass Spectrometry

When dealing with complex mixtures such as total protein extracts for proteomic analyses the step of fractionation is indispensable to achieve extensive proteome coverage. Either performed at protein and/or peptide level, fractionation allows reducing sample complexity introduced to the mass spectrometer, aiding protein identifications.
Electrophoresis based fractionation methods are common (1D-PAGE\textsuperscript{26}, 2D-PAGE\textsuperscript{27}, OFFGEL\textsuperscript{28-32} and \textit{GelFree}\textsuperscript{33}) and for global analyses 1D-PAGE gels have established themselves as a successful method to decrease complexity of the proteome at the protein level\textsuperscript{26}. Separation on 1D-gels is carried out based on MW and is presented in Figure 1.6. Negatively charged sodium dodecyl sulfate (SDS) masks the native charge of the denatured proteins and aids to migrate through the pores of the polyacrylamide gel towards the positively charged electrode once a voltage is applied. The pore size is inversely proportional on the percent polyacrylamide used. Therefore gradient gels are preferred when fractionation for a broad range of proteins is anticipated as in the case for complex protein samples. The tris-glycine Laemmli buffering system is used, stacking the proteins at pH 6.8 and resolving them at pH 8.8\textsuperscript{34}. Glycerol is added to the sample to increase the viscosity and the bromphenol blue dye front enables tracking the migration during the run. Fluorescent dyes\textsuperscript{35} or the more conveniently coomassie dyes\textsuperscript{35} are applied to stain and image the gel lanes, each containing an entire proteome. After an initial washing step where the SDS is removed the coomassie dye can bind under acidic buffer conditions to basic amino acid residues of the proteins. The chemical modification is not permanent so that after excising the gel lanes into user-defined fractions destaining and in-gel digestion procedures for mass spectrometry (MS) analysis can be performed. The downsides
of this method are low recovery of peptides (~20%) from the gels and also tedious and time consuming sample preparation when large-scale experiments are intended.

One dimensional and 2 dimensional liquid chromatography based fractionation and separation platforms for proteomics approaches have also been developed. Ion exchange chromatography at peptide level is for example a very common technique. Fractionation based on isoelectric point is possible due to ionic interactions between stationary phase and analyte molecules when anion or cation exchange is performed. Cation exchange\textsuperscript{36-38}, positively charged molecules interacting with negatively charged resins (see Table 1.1 for functional groups on resins), is predominant in proteomics literature probably due to the occurrence of more basic peptides in nature\textsuperscript{39}. When trypsin is used for digestion, at least one basic amino acid residue is guaranteed in every peptide (lysine or arginine). But also finding a stable resin suitable for required high pressures (for 2D-LC) and high pHs for anion exchange chromatography might have been another factor. Silica based resins are stable at high pressures, however decompose at high pH’s. In contrary polymer based resins are stable at pH 0-14, however are too soft to be utilized in 2D-LC\textsuperscript{40}.

A salt gradient or a pH gradient can be applied in order to perform separation. The pH is kept constant for salt gradients where salt molecules electrostatically compete against analyte molecules bound to the charged resin (Figure 1.7).
Salt levels can be kept low when pH gradients are performed. Gradients from basic pH to acidic pH are used when applied on anion resins and the opposite direction, acidic to basic, is utilized when cation resins are utilized. Analyte molecules elute from the column when their isoelectric point is reached. Although pH elution seems to be more practical for proteomic approaches, since salts in the sample suppress ESI efficiency, it should be kept in mind that pH changes of the buffer alter charges on the resin as well as on the adsorbed molecules, especially for weak anion and cation exchange and small changes in the pH of the buffer.

Figure 1.7 Schematic of strong anion (SAX) and strong cation exchange (SCX) surface interactions. Quaternary ammonia ions are implemented as functional groups on the resin for strong anion exchange and attract negatively charged ions. For strong cation exchange sulfonic acid groups are used to attract positively charged molecules.
might result in larger pH changes at micro level where molecules are adsorbed on the resin and thus pH gradients might result in crude fractionations\textsuperscript{41}.

\textbf{Table 1.1:} Functional groups for anion and cation exchange resins

<table>
<thead>
<tr>
<th>Functional Group</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\text{CH}_2\text{N}^+(\text{CH}_3)_3)</td>
<td>Strong Anion Exchange</td>
</tr>
<tr>
<td>(-\text{N}^+(\text{CH}_3)_3)</td>
<td>Strong Anion Exchange</td>
</tr>
<tr>
<td>(-\text{CH}_2\text{N}^+\text{H}(\text{CH}_3)_2)</td>
<td>Weak Anion Exchange</td>
</tr>
<tr>
<td>(-\text{CH}_2\text{N}^+\text{H}_2(\text{CH}_3))</td>
<td>Weak Anion Exchange</td>
</tr>
<tr>
<td>(-\text{CH}_2\text{SO}_3^-)</td>
<td>Strong Cation Exchange</td>
</tr>
<tr>
<td>(-\text{SO}_3^-)</td>
<td>Strong Cation Exchange</td>
</tr>
<tr>
<td>(-\text{CH}_2\text{COO}')</td>
<td>Weak Cation Exchange</td>
</tr>
<tr>
<td>(-\text{COO}')</td>
<td>Weak Cation Exchange</td>
</tr>
</tbody>
</table>

The inability to couple strong cation exchange directly to mass spectrometer when salt gradients were applied was bypassed with MudPIT\textsuperscript{42-44} (multi-dimensional protein identification technology). MudPIT is an online combination of strong cation exchange and reverse phase chromatography developed by Yates and coworkers. A short column of strong cation exchange resin separates the peptide sample applying a salt gradient, while on the reverse phase (RP) column desalting and another level of separation based on hydrophobicity takes place. MudPIT allowed the unbiased identification of 1484 proteins across different
classes and abundance levels in the yeast proteome\textsuperscript{43}. Due to the downsides of MudPIT requiring multiple solvent combinations, limited loading capacity resulting in low abundant signal and more than 24 h for a 12 step salt gradient for a single sample\textsuperscript{45}, the engagement of MudPIT decreased in the last decade.

FASP\textsuperscript{46,47} (filter aided sample preparation) combined with StageTip (stop and go extraction tip) fractionation\textsuperscript{48-50} is a more recently developed methodology. By utilizing a 30 kDa molecular weight cutoff filter, up to 250 µg of protein can be prepared for digestion in one filter within 3 hours. Preparation of up to 30 samples simultaneously is possible. Denatured proteins with molecular weights as small as 5 kDa can be retained dependent upon the radius of gyration\textsuperscript{51,52}. The filter enables SDS depletion, which is advantageous during protein extraction but is incompatible with electrospray ionization mass spectrometry (ESI-MS). Highly concentrated Urea (8M) is utilized to weaken the SDS-protein interactions and aids in removing the SDS. Fractionation is performed by in-house made anion StageTips (Figure 1.8) with a universal pH buffer series.

![Figure 1.8](image.png)  

\textbf{Figure 1.8} Anion exchange resin used in StageTips. The beads are embedded in a teflon mesh and consist of a styrene-divinylbenzene copolymer net modified with quaternary ammonium ions as the functional group.
Universal buffers, first developed in 1924 by Prideaux and Ward\textsuperscript{52}, make pH gradients over a large range possible. While conventional buffer, e.g. Tris (tris(hydroxymethyl)methylamine) can only buffer over one or two pH ranges (7.1-9.0), universal buffers like the Britton Robinson buffer can buffer from pH 2.0 to pH 12.0\textsuperscript{53}. Britton Robinson buffer is made of equal molarities of boronic acid (H\textsubscript{3}BO\textsubscript{3}pK\textsubscript{a}=9.24), phosphoric acid (H\textsubscript{3}PO\textsubscript{4}pK\textsubscript{a}=2.148, 7.198, 12.375), and acetic acid (CH\textsubscript{3}COOHpK\textsubscript{a}=4.792). The small differences in pK\textsubscript{a} values in universal buffers, leads to pH changes proportional to the amount of base added, which is advantageous when preparing a buffer series.

RP chromatography is the most commonly used mode for peptide purification and separation and due to its compatibility usually directly coupled to mass spectrometers. Hydrophobic interactions take place between the analyte molecules and the C18 stationary phase. The elution order is hydrophilic to hydrophobic analytes and common mobile phases are water-acetonitrile mixtures with formic acid as ion pairing reagent. While at acidic pH’s acidic and neutral analytes will be retained sufficiently without ion pairing reagents, charged basic analytes might be too polar to be retained. Therefore ion pairing reagents are added to form neutral species which can then be retained on the hydrophobic reverse phase column\textsuperscript{54}. However, for ESI strong ion pairing reagents \textit{e.g.} trifluoroacetic acid should be avoided as they suppress ionization of analyte molecules at the ESI source\textsuperscript{55-57}.

\textit{Gradient} elution is performed for complex mixtures rather than \textit{isocratic} elution, because of higher sensitivity of late eluting peaks, enhanced resolution of early and late eluting peaks and higher peak capacity. The rule of thumb to calculate the peak capacity is:
This rule is based on the assumption that all peaks have the same width as it is mostly the case for gradient elution. For completely orthogonal 2D separations the total peak capacity is simply the product of both dimensions\textsuperscript{58}.

RP-RP at two different pH values\textsuperscript{59-61} has become an option in separation of peptides for proteomic analyses as it was shown to be identical in orthogonality to SCX-RP\textsuperscript{62}. The orthogonality in RP (high pH)-RP (low pH) arises only from the difference in the pH values since the column used in both dimensions are C18 columns, whereas secondary hydrophobic interactions on a SCX column decrease the orthogonality to RP.

1.2.2 Electrospray Ionization Mass Spectrometry

The development of ESI by John B. Fenn and coworkers\textsuperscript{63}, a soft ionization technique for large biomolecules had tremendous impact on the area of biological mass spectrometry and empowered liquid chromatography (LC) to be coupled online to mass spectrometers\textsuperscript{63}. The continuous ESI source produces multiply charged ions on the emitter tip without fragmentation. Depending on operation in positive or negative mode either an oxidation process or a reduction process occurs on the emitter tip. To analyze proteins and peptides the system is usually operated in positive mode. The oxidation process is as follows:

\[
2 \text{H}_2\text{O} \rightarrow \text{H}_2(g) + 4 \text{H}^+ + 4\text{e}^-
\]  

\textbf{Equation 2}
High voltage (2-5 kV) is applied to the tip and the liquid sample becomes nebulized and forms charged droplets while emitting. The droplets begin to desolvate and two different theories exist describing the theory of the gas ion formation process: Charge Residue Model (CRM) proposed by Dole et al.\textsuperscript{64,65} and Ion Desorption Model (IDM) proposed by Iribarne and Thompson\textsuperscript{66-68}. The CRM states that droplets undergo coulomb explosion and multiple smaller droplets (estimated 20\textsuperscript{69,70}) composed of solvent and analyte molecules are produced\textsuperscript{71} after reaching the Rayleigh limit\textsuperscript{72}: 

\[ q_r = \sqrt{64\pi^2 \varepsilon_0 \gamma r^3} \] \hspace{1cm} \textbf{Equation 3}

where \( \varepsilon_0 \) is the dielectric constant, \( \gamma \) the surface tension and \( r \) the radius of the droplet. This process repeats itself until single analyte molecules with some charge are left and all solvent is evaporated. The theory of the IDM is that the process of coulomb explosion is repeated until the electric field on the droplet can lift the inner analyte molecule onto the surface and pushes it out. IDM, which was supported later by Fenn\textsuperscript{73}, explains the selectivity of ion charge states better, since more hydrophobic analytes would leave the droplet very early with low charge states, whereas more hydrophilic analytes can stay longer inside the droplets and have higher charge states. The CRM theory however is accepted for large molecules, but the debate about both theories is still going on. A recent review by Kebarle and Verkerk\textsuperscript{74} hypothesized even a possible combination of both mechanisms going on in ESI.
1.2.2.1 Hybrid Mass Spectrometer

Combinations of different mass analyzers are referred to as hybrid analyzers. Linear ion traps for example can be coupled to high resolving power mass spectrometers such as Orbitraps or Fourier Transform Ion Cyclotron (FTICR) (Figure 1.9). These hybrid instruments provide high quality tandem MS spectra and lead to extremely confident peptide identifications in proteomic approaches.

Figure 1.9 ESI-LTQ-FTICR and Orbitrap. After electrospray ionization ions can enter the hybrid mass analyzers through the heated capillary inlet and are transferred to the linear ion trap (LTQ) which is coupled to either the ICR cell or the Orbitrap. Both are high resolving power mass analyzers but have autonomous ways of detecting ion frequencies.

High resolving powers for the precursor spectrum can be achieved with FTICR\textsuperscript{75,76} instruments depending on the strength of the magnetic field surrounding the ICR-cell. Ions are excited by an electric field perpendicular to the magnetic field (B) and start moving in
circles inside the cell due to the Lorentz force. Mass to charge ratios \((m/z)\) are determined by the detector plates based on the cyclotron frequency \(\omega_c\) of the ions:

\[
\omega_c = \frac{z e B}{m}
\]

Equation 4

Another high resolving power mass analyzer is the Orbitrap which can also be coupled easily to a linear ion trap. In the LTQ-Orbitrap\textsuperscript{77-80} ions are transferred from the linear ion trap to a C-trap and from there to the Orbitrap. The C-trap has the function of storing the ions and squeezing them into a small ion cloud, allowing them to enter the Orbitrap fast and start moving in orbits around the inner spindle like electrode. The ion currents are measured on the outer detector electrodes and the time-domain signal is converted by fast Fourier transformation into a frequency domain, which is further converted according to Equation 5 into a mass to charge \((m/z)\) spectrum, where \(\omega_z\) is the axial oscillation frequencies and \(k\) is a constant coming from the field curvature\textsuperscript{81}.

\[
\omega_z = \sqrt{\frac{k}{m/z}}
\]

Equation 5

The advantage of the Orbitrap compared to FTICR is that no magnetic fields are needed, and it functions using solely the electrostatic field. Expensive cryogen cooling of the magnets is therefore not necessary. The resolving power of an Orbitrap is routinely 60,000 at \(m/z = 400\), and leads to mass measurement accuracies \(<3\) ppm with external calibration. Internal
calibration by enabling the lock mass (polydimethylcyclosiloxane \( m/z = 445.120025 \)) present in ambient laboratory air results in a mass measurement accuracy (MMA) of \( \leq 1 \) ppm\(^79\).

Besides transferring ions to the Orbitrap the linear ion trap can also store, accumulate and fragment selected precursor ions. Collision induced dissociation (CID) is performed by colliding neutral gas molecules with the ions in the linear ion trap. The kinetic energy is then transferred into internal energy and fragmentation of peptides occurs at the CO-N peptide bonds, yielding b and y ions (Figure 1.10). The energy applied for the collision is in linear correlation with the selected peptide mass. Therefore the collision energy is set as percent normalized collision energy (\( \sim 25\text{-}35\% \)) and accounts with that for the mass dependency.

\[ \text{Figure 1.10} \] The isotopic distribution for a selected m/z value (848.95 m/z) is shown in a high resolving power FTMS spectrum acquired in an Orbitrap as well as the MS/MS spectrum, subsequently acquired in the linear quadrupole.
One MS scan is generally followed by 8 MS/MS scans and a whole scan cycle is completed within 3 seconds. The most abundant 8 precursor ions are selected for MS/MS; however, singly charged precursors are excluded due to poor fragmentation. Dynamic exclusion of previously selected precursors is performed for about 3 minutes to give lower abundant ions opportunities to be selected as well. Triggers between MS and MS/MS events are the automatic gain control (AGC) and the maximum ion injection time. In cases where the not enough ions could be accumulated to reach the AGC limit the ion injection times represent the limiting factor.

Although linear ion traps provide low MMA, the information obtained from the LTQ combined with the accurate precursor mass from the Orbitrap yields unambiguous peptide identifications in complex biological mixtures. The real strength of linear ion traps however lays in the various scan modes when three linear quadrupoles are coupled back to back to a triple quadrupole instrument. Precursor ion scan, product ion scan, neutral loss scan and selected reaction monitoring (SRM) are possible. The SRM scan mode is particularly used for absolute quantification purposes in combination with stable isotope labeled synthetic peptides and termed protein cleavage isotope dilution mass spectroscopy (PC-IDMS) as shown in Figure 1.11. In SRM mode a precursor ion mass is selected in the first quadrupole, dissociated in the second quadrupole (RF only) and specific fragment ion masses are monitored in the third quadrupole. This approach is highly sensitive, selective and is performed with a fast LC-gradient to take advantage of retention time (co-elution of stable isotope labeled peptide and natural peptide) and avoid quantification of non-target ions with
similar masses and transitions. After determining relative changes in proteins of interest from a global study, PC-IDMS allows studying these proteins in higher resolution, since no fractionation is needed and much more samples (e.g. time points) can be processed and run in a much shorter timeframe.

Figure 1.11 PC-IDMS approach on a triple quadrupole instrument. The stable isotope labeled synthetic peptide is added concurrently with trypsin to the protein sample. The peptide mixture is introduced to the triple quadrupole mass spectrometer operating in SRM mode. Precursor masses are selected in Q1 while fragmentation happens in Q2 (RF only) and fragment ions are detected in Q3. Absolute quantification can be performed based on the intensities of the stable isotope labeled peptide and the natural peptide of interest under the assumption $c_{\text{peptide}} \approx c_{\text{protein}}$.

1.2.3 Data Analysis

A variety of search engines for proteomic approaches exist currently and are used in different labs; MASCOT$^{83}$, SEQUEST$^{84}$, MaxQuant$^{85}$, X-Tandem$^{86}$, OMSSA$^{87}$, ProteinPilot$^{88}$, Protein Prospector$^{89}$, MyriMatch$^{90}$ are only some examples. Even though the goal of these programs
to match \textit{in-silico} created peptides from a predicted proteome to the MS and \textit{MS/MS} spectra, every algorithm delivers slightly different protein list results\textsuperscript{91}.

MASCOT is one of the most popular search algorithms due to its compatibility with data files from mass spectrometers from a broad range of different manufacturers (Thermo, Agilent, Bruker, Shimadzu, Waters, AB SCIEX). Four different peak picking algorithms are available in MASCOT for tandem MS data with high resolution MS and low resolution \textit{MS/MS} data and are described here beginning with the most time consuming option to the fastest. Option 1 allows for a high resolution re-gridding (400 points/Da) in the precursor scan and searches for precursor and fragment ions up to charge state +7. Option 2 performs a 200 points/Da re-gridding, searches for precursors for up to +4 and fragment ions up to +2. Peak picking of the precursor scan can be skipped and precursor m/z and charge determined the raw file can be adopted with option 3 or instead peak picking of the precursor scan can be performed but the centroid data can be taken directly from the raw file for the \textit{MS/MS} scans with option 4.

Once a peak list is created, experimentally determined mass values are then compared to the calculated peptide and fragment ion masses. The best matches are reported with probability based scores dependent amongst other things on the spectrum qualities. Important search space limiting parameters are mass tolerance, enzyme specificity, allowed missed cleavages and fixed and variable modifications.

To account for misassignments, a probability based false discovery rate (FDR) is calculated in the ProteoIQ software\textsuperscript{92}. In contrary to the false positive rate (FPR) the number of false
positives is not known when calculating the FDR\textsuperscript{92-94}, but estimated by using peptide matches to either a reverse- or a shuffle- or a random-database. These databases are created based on the predicted forward-sequences. The FDR is defined as the expected quotient of the number of FPs and the total number of identified sequences (sum of FPs and true positives (TPs)):

\[
FDR = \frac{FP}{FP + TP}
\]

Label free measurements facilitated relative quantification enormously. No tagging or label incorporations are needed for the approach and studies have proven that spectral counts\textsuperscript{95} and ion abundances\textsuperscript{96-100} can be used as a measure of relative abundance.

To account for variation caused by sample handling and run to run variance as well as to get more accurate quantification, spectral count normalization is performed. Total spectral count normalization\textsuperscript{101} (TSpCN), Normalized Spectral Abundance Factor\textsuperscript{102,103} (NSAF) and Normalization to specific proteins (NSP) are some methods used to normalize spectral counting data.

TSpCN is performed by choosing the replicate with the highest number of TSpCs (t) and normalize remaining replicates to it. Normalization is performed within a sample as well as across different samples. The normalization factor (NF) is applied on each protein (k) individually:

\[
TSpCN = SpC_k \times NF_t
\]
To account for the fact that larger proteins can produce more tryptic peptides and thus get a higher number of SpCs Washburn and co-workers developed NSAF. SpCs for a given protein are divided by its number of amino acids to yield a SAF, which is then divided by the sum of all SAFs to yield a NSAF.

\[
NSAF = \frac{(SpC_{i}/Length)_{k}}{\sum_{i=1}^{N}(SpC_{i}/Length)_{k}}
\]

Equation 8

In NSP an endogenous house-keeping protein or an exogenous spike-in protein is chosen and the assumption is made that SpCs for those proteins should be conserved throughout runs. The difference in SpCs for the chosen protein determines the NF. If more than one standard protein is utilized, the NF becomes an average NF incorporating SpC pattern of all standard proteins p:

\[
NSP = SpC_{k} \times ANF_{p}
\]

Equation 9

1.3 Synopsis of Completed Research

This dissertation presents the approach of proteomic characterization of a biological system: The rice blast fungus *Magnaporthe oryzae*. Chapter 2 describes a 2 dimensional RP-RP liquid chromatography fractionation and separation methodology developed and applied on the model fungal organism *Saccharomyces cerevisiae*. Improvement in protein identifications were observed despite any orthogonality between both chromatography dimensions. Solely
the fact that more material per fraction could be loaded on the second dimension column and extended gradients could be run per fraction increased the number of protein identifications by 61.6%. A more common fractionation technique of 1D-Gels combined with nanoLC-MS/MS applied on *M. oryzae* is described in Chapter 3. 1511 conidial proteins were identified in this study. Normalization techniques applied on the label-free data set showed that TSpCN and NSAF performed superior over NSP. FASP combined with StageTips anion fractionation was employed in Chapter 4 and a doubling in protein identifications in *M. oryzae* conidia could be achieved with this methodology. Biological insights are given about the conidial proteome. Chapter 5 presents a temporal study throughout the life cycle of wild-type and mutant (Δcpka) *M. oryzae*. Protein up and down regulations were examined during different time points monitoring the physiological change the conidia undergoes while forming the infective appressoria. Absolute quantification of 4 proteins with interesting expression profiles were performed and absolute concentrations were found to be in agreement with the relative concentrations determined by spectral counting in the global study. Finally the post translational modification ubiquitination was found to be highly increased in *M. oryzae* under nitrogen starvation and is described in Chapter 6. Mycelium grown under nitrogen starvation conditions revealed 63 significantly enriched polyubiquitinated candidate proteins suggesting that ubiquitination plays a role in redirecting pathways in *M. oryzae* for survival under extreme conditions. The polyubiquidin encoding gene MGG_01282 displayed up regulation in gene expression during appressorium formation and targeted gene deletion of MGG_01282 disabled the fungus to cause disease.
1.4 References


\[\text{\textsuperscript{i} Modified with permission from: Dean et al. The Genome Sequence of the Rice Blast Fungus *Magnaporthe grisea*, *Nature* **2005**, 434, 980-986.}\]

\[\text{\textsuperscript{ii} The file is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Source: http://flickr.com/photos/86712369@N00/5516789000.}\]

36
CHAPTER 2
Increasing Proteome Coverage with Offline RP HPLC Coupled to Online RP NanoLC–MS


2.1 Introduction

Mass spectrometry (MS) is the central technology for proteomic analyses. High sensitivity, rapid duty cycles, parts-per-million mass accuracy and high resolving power of hybrid mass spectrometers have made MS a valuable tool in proteomics. However, MS still has limitations in dynamic range (3-4 orders of magnitude), and detecting the entire proteome of a species is a target that has not yet been accomplished. The difficulties herein reside mostly in the wide dynamic range of proteins and the sheer complexity of the proteomes, which even the rapid duty cycles of the most recently developed mass spectrometers cannot handle. The development of improved MS platforms is still in progress and is essential to increase the proteome coverage.

The fungus Saccharomyces cerevisiae (S. cerevisiae) is commonly used as a model organism in technology and methodology evaluation studies and has to date the most extensively characterized proteome. Mann and coworkers1 were the first in combining two dimensional polyacrylamide gel electrophoresis (2D-PAGE) fractionation2 and a matrix assisted laser desorption ionization–time of flight mass spectrometer identifying 150 yeast proteins in their study in 1996. Twelve years later, again Mann and coworkers3 identified 3639 proteins using 1D-PAGE4,5 and 3987 yeast proteins using OFFGEL6-10 both coupled to an online liquid chromatography (LC) linear trap quadrupole (LTQ)-Orbitrap mass spectrometer.
A variety of analytical strategies exists for the fractionation of complex mixtures prior to MS analysis and is either performed at the protein or peptide level, each with their own advantages and disadvantages. Aside from electrophoresis based fractionation methods (2D-PAGE and 1D-PAGE, OFFGEL and GelFree\textsuperscript{11}), chromatography based methods have also been developed. Offline strong cation exchange (SCX)\textsuperscript{12-14} prior to nanoLC-MS and multi-dimensional protein identification technology (MudPIT)\textsuperscript{15-17}, combining SCX and reversed phase (RP) online, are both common methods used in proteomics.

RP-RP at two different pH values\textsuperscript{18-20} has become an option in separation techniques as it was found to be identical in orthogonality to SCX-RP\textsuperscript{21}. The orthogonality in RP-RP arises only from the difference in the pH values since the column used in both dimensions are C18 columns. In this study we performed peptide level offline RP high pressure liquid chromatography (HPLC) fractionation at micro-flow/min rates of a \textit{S. cerevisiae} whole digest prior to RP nanoLC-MS (LC-LC-MS), using the same mobile phase and pH and compared this method to the performance of only nanoLC-MS analysis of the whole digest. The total number of proteins and protein groups identified from each analysis were compared.

2.2 Experimental Section

2.2.1 Sample preparation

The \textit{S. cerevisiae} strain Y15696 (BY4742; MaTα; his3D1; leu2D0; lys2D0; ura3D0; YIR034c::kanMX4), with a \textit{lys}1 gene deletion purchased from EuroScarf (Frankfurt, Germany) was inoculated in liquid yeast peptone dextrose and harvested during log phase. A yeast cell pellet was collected after centrifugation of the culture at 5,000 rpm for 10 min at
4 °C, washing with 50 mM Tris-HCl (Sigma Aldrich, St. Louis, MO) buffer and followed by a second centrifugation step. By grinding the pellet with liquid nitrogen the cells were lysed and resuspended in 50 mM Tris-HCl. Cell debris was removed with a third centrifugation step at same conditions mentioned above.

Protein concentration was determined by a BCA and Bradford assay. To denature the protein, urea (Sigma Aldrich, St. Louis, MO) was added to the yeast lysate to a final concentration of 8 M. Dithiothreitol (DTT) (Biorad, Hercules, CA) was added to a final of 5 mM and the sample was incubated for 30 min at 56 °C to reduce the protein disulfide bonds. Iodoacetamide (Sigma Aldrich, St. Louis, MO) was added to a final of 20 mM and the sample was incubated for 30 min in the dark at room temperature for alkylation of the free thiols. DTT was added again to quench the alkylation reaction. In order to dilute the urea to 2 M, 50 mM Tris-HCL was added. Digestion was performed overnight with the addition of trypsin (Sigma Aldrich, St. Louis, MO) at a 1:50 enzyme: protein ratio. Formic acid (FA) (Sigma Aldrich, St. Louis, MO) was added to stop the digestion.

2.2.2 Offline RP HPLC

An aliquot of the yeast digest (63.5 μg), was reconstituted in 50 μl mobile phase A (98 % HPLC-Grade H2O (Burdick and Jackson, Muskegon, MI), 2 % Acetonitrile (ACN) (Burdick and Jackson, Muskegon, MI) and 0.2 % FA). Peptide level fractionation was performed on a Shimadzu HPLC system (CBM-20A prominence communications bus module, DGU-20A5 prominence degasser, two LC-20AD prominence pumps, CTO-20A prominence column oven at 40 °C, SPD-20A prominence ultraviolet-visible (UV-VIS) detector) (Shimadzu,
Columbia, MD) equipped with a 100 μl sample loop. The sample was loaded onto a 5 μm, 200 Å, Magic C18AQ, 2.0 mm × 150 mm column (Michrom Bioresources Inc., Auburn, CA). The pumps were operated at a combined flow rate of 200 μl/min. After one minute equilibration time and a 10 minute column wash at 5% mobile phase B (98 % ACN, 2 % H₂O and 0.2 % FA) a 30 minute gradient was performed from 5 % - 50 % B.

Table 2.1 Calculated Peptide Concentrations per Fraction

<table>
<thead>
<tr>
<th>fraction #</th>
<th>conc. [μg/ml]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.23</td>
</tr>
<tr>
<td>2</td>
<td>3.23</td>
</tr>
<tr>
<td>3</td>
<td>4.84</td>
</tr>
<tr>
<td>4</td>
<td>7.26</td>
</tr>
<tr>
<td>5</td>
<td>8.87</td>
</tr>
<tr>
<td>6</td>
<td>9.68</td>
</tr>
<tr>
<td>7</td>
<td>8.87</td>
</tr>
<tr>
<td>8</td>
<td>7.26</td>
</tr>
<tr>
<td>9</td>
<td>4.84</td>
</tr>
<tr>
<td>10</td>
<td>3.23</td>
</tr>
</tbody>
</table>

The gradient was ramped up to 95 % B over one minute and held for 4 minutes to wash the column. The gradient was then ramped down to 5 % B in two minutes and maintained for three minutes. A Gilson FC 203B fraction collector (Gilson Inc., Middleton, WI) was used to collect fractions every 3 minutes during the 30 minute gradient for a total of 10 fractions. Protein concentrations of the fractions were calculated using the UV-VIS spectrum and the Scope’s method²² (Table 2.1). The fractions were dried down and reconstituted in mobile phase A to a concentration of approximately 100 ng/μl.
2.2.3 NanoLC-MS

The nanoLC prior to MS/MS was performed on a nanoLC-1D system from Eksigent (Dublin, CA) at room temperature. A 75 μm i.d. IntegraFrit capillary (New Objective, Woburn, MA) was packed in house to 5 cm with Magic C18AQ packing material (Michrom BioReasources, Auburn, CA) and operated as a trap. A 75 μm i.d. PicoFrit capillary column (New Objective, Woburn, MA) was packed 15 cm with the same packing material. Separations were carried out using a continuous, vented column configuration as previously described by our group.23

<table>
<thead>
<tr>
<th>fraction #</th>
<th>gradient [%B]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 - 9</td>
</tr>
<tr>
<td>2</td>
<td>8 - 12</td>
</tr>
<tr>
<td>3</td>
<td>10 - 14</td>
</tr>
<tr>
<td>4</td>
<td>13 - 17</td>
</tr>
<tr>
<td>5</td>
<td>15 - 19</td>
</tr>
<tr>
<td>6</td>
<td>18 - 22</td>
</tr>
<tr>
<td>7</td>
<td>21 - 25</td>
</tr>
<tr>
<td>8</td>
<td>25 - 29</td>
</tr>
<tr>
<td>9</td>
<td>29 - 33</td>
</tr>
<tr>
<td>10</td>
<td>33 - 37</td>
</tr>
</tbody>
</table>

A 2 μl (200 ng) sample was injected into the 10 μl loop and loaded onto the trap column with approximately ten column washes prior to analytical separation. For peptide separation on the analytical column, the flow rate was set to 350 nl/min. A 5 minute column wash was performed at 2 % B followed by a 1 hour linear gradient. The gradient was ramped up to 90 % B in 1 minute and maintained for 10 minutes. Two minutes were required to establish 2 % B and this was maintained for two minutes. Fraction 1 was subjected to a 1 hour gradient of 5 – 9 % B and the gradients for the following fractions were determined based on the
elution end time of the previous fraction. All gradients consisted of a 4 % increase in B over the 1 hour time period. The gradient for each fraction is listed in Table 2.2. In addition, a whole yeast digest sample (200 ng) was analyzed by nanoLC-MS/MS and a 1 hour gradient of 10 – 40 % B was applied. Three technical replicates of the whole yeast sample were run. MS analysis was performed using a hybrid LTQ-Orbitrap MS (Thermo Fischer Scientific, Bremen, Germany). The automatic gain control (AGC) limit for the Fourier Transform MS (FTMS) was set to 1×106 and the maximum injection time was 500 ms. For the ion trap the AGC limit was 8×103 and the maximum injection time was 80 ms. The resolving power was set to 30,000fwhm at m/z 400 and 8 data dependent MS/MS events were performed for ions with charge states ≥+2. Singly charged ions were rejected for MS/MS. Dynamic exclusion was enabled, and ions selected for MS/MS interrogation were excluded for 180 seconds. The normalized collision energy was 35% and lock mass calibration using polydimethylcyclosiloxane present in ambient laboratory air (m/z 445.120025) was enabled. External calibration was also performed following manufacturer instructions and using manufacturer’s calibration mix.

2.2.4 Data analysis

Data analysis was performed by converting the .raw files into .mgf files through MASCOT Distiller version 2.3.01 (Matrix Science Inc., Boston, MA) and searching the .mgf files against the target reverse yeast ORF database (orf_trans_all.fasta.gz) from the Stanford University web page: http://www.yeastgenome.org in MASCOT. Parameters used in MASCOT were ± 5 ppm peptide ion tolerance, ± 0.6 Da MS/MS fragment ion tolerance and
2 allowed missed cleavages. A fixed modification was set to be carbamidomethylation of cysteine and variable modifications were oxidation of methionine as well as deamidation of glutamine and asparagine. The output of the database search in the form of .dat files were analyzed in ProteoIQ version 2.1.01_SILAC_beta08 (BioInquire, Athens, GA) at 1 % FDR.24

2.3 Results & Discussion
The experimental workflow is diagramed in Figure 2.1. The S. cerevisiae strain was grown in liquid culture and harvested during log phase. BCA and Bradford assays were performed to estimate the concentration of the lysate and the sample was digested in solution using trypsin. An aliquot of the digest was run directly on the nanoLC-MS and another was collected as 10 fractions during offline RP HPLC prior to nanoLC-MS.
Figure 2.1 Experimental workflow. *S. cerevisiae* was inoculated, harvested and lysed. After a tryptic digestion, 200 ng of the digest were analyzed in triplicate by nanoLC–MS. A sample containing 63.5 μg of yeast peptides was loaded onto the offline HPLC and 10 fractions were collected; narrower gradients were applied to the fractions on the nanoLC–MS. The data was searched in MASCOT and combined for comparative analysis in ProteoIQ.

The *S. cerevisiae* sample fractionated offline was detected and approximately quantified by UV-VIS at 205 nm (see Figure 2.2A and B). The total recovery was around 60% and sufficient to perform further nanoLC-MS analysis with the fractions. The total ion chromatogram (TIC) for the nanoLC-MS analysis, increasing only 4% B over a 60 min linear gradient, is shown in Figure 2.2B. The TIC in Figure 2.2C emphasizes the complexity of the *S. cerevisiae* whole sample digest applying a 60 min gradient for peptide elution of 10-40% B.
Figure 2.2 Representative data comparing the complexity in chromatography. (A) UV–vis spectrum at 205 nm for the fractionation on the offline HPLC. Fractions 1–10 were collected during the time periods marked with dashed black lines. The gradient is implied with the dashed line in red. As an example fraction 6 was collected at 25–29% B and is highlighted in violet. (B) TIC of fraction 6. (C) TIC of whole digest.

By applying narrow gradients to the already fractionated *S. cerevisiae* digest the need for an additional level of separation is demonstrated through the chromatogram and more MS/MS spectra for that gradient range are obtained. Reducing the complexity of the sample in addition to increasing the total amount of material loaded yields the potential for more protein identifications, in particular, lower abundant species, which are frequently suppressed in data-dependent LC-MS experiments of complex samples. As an example, the protein: RPL31B SGDID:S000004398, Chr XII from 931754-931698, 931348-931064, reverse complement was identified with one peptide in the whole digest sample (21.05% sequence
coverage), whereas in the LC-LC approach, with more opportunity for detection, the same protein was identified with 6 unique peptides (55.26% sequence coverage).

The peptide overlaps between the fractions are described in Table 2.3. As expected early fractions have no peptide overlap with late fractions offering evidence that the separation was carried out successfully. The large overlap of 283 peptides between fractions 6 and 7 is due to the number of peptides eluting at the same time and thus bleeding out. Carryover is also attributable to peptides that are eluting at the time boundary as the fraction collector switches between fractions.

The Venn diagram in Figure 2.3 presenting the results, show that evaluating the 10 offline RP-LC fractions of the S. cerevisiae digest in narrow gradients by nanoLC-MS increased protein identification by 61.6% versus direct analysis of the whole digest by nanoLC-MS. In the LC-LC-MS 1028 yeast proteins were identified at 1 % FDR compared to 636 proteins resultant of nanoLC-MS analysis. The majority of proteins identified from the whole digest

<table>
<thead>
<tr>
<th>fraction#</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>247</td>
<td>29</td>
<td>6</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>257</td>
<td>25</td>
<td>10</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>25</td>
<td>461</td>
<td>43</td>
<td>30</td>
<td>17</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>10</td>
<td>43</td>
<td>631</td>
<td>137</td>
<td>96</td>
<td>42</td>
<td>10</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>12</td>
<td>30</td>
<td>137</td>
<td>1063</td>
<td>227</td>
<td>123</td>
<td>37</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>4</td>
<td>17</td>
<td>96</td>
<td>227</td>
<td>1227</td>
<td>283</td>
<td>108</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>42</td>
<td>123</td>
<td>283</td>
<td>1164</td>
<td>246</td>
<td>89</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>37</td>
<td>108</td>
<td>246</td>
<td>977</td>
<td>196</td>
<td>105</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>28</td>
<td>89</td>
<td>196</td>
<td>684</td>
<td>169</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>15</td>
<td>30</td>
<td>105</td>
<td>169</td>
<td>461</td>
</tr>
</tbody>
</table>
sample analyzed by nanoLC-MS were also identified within the total population of protein identifications from the LC-LC-MS experiment.

The analysis time for the fractionation must be considered as analyzing a sample in a timely manner is also very important. The time required for this fractionation is about an hour for the initial offline RP separation. Influenced by the number of fractions collected, 10 fractions here, nanoLC-MS analysis of all fractions requires 10 hours. Although LC-LC-MS analysis time compared to nanoLC-MS analysis time is a considerable increase, 1D separations are known to be insufficient for complex mixtures.

De Godoy et al.\(^3\) analyzed 24 OFFGEL-fractions using 2 h gradients on the nano-LC and identified 3987 yeast proteins (83 proteins/hour). In comparison, in our LC-LC-MS study 10 fractions yielded 1028 proteins using a non-orthogonal method and only 1 h gradients (102proteins/h). The same database and the MASCOT search engine were used in both studies. The only difference in the bioinformatic platform was the analysis software; ProteoIQ employed for our LC-LC-MS analysis whereas de Godoy et al. used MaxQuant. Considering the rate of protein identification per hour, vide supra, it is demonstrated that LC-LC-MS has an acceptable analysis time in order to continue the pursuit of characterizing an entire proteome through MS analysis.

**Figure 2.3** Venn diagram for comparison of proteins identified with each method. The majority of proteins identified in the whole digest (636) are within the identified proteins from the LC–LC fractions (1028).
2.4 Conclusions

Fractionation prior to LC-MS is essential in order to overcome the dynamic range and complexity of biological samples. We have shown that offline HPLC fractionation at the peptide level coupled to online nanoLC-MS, a non-orthogonal method, is a simple and effective way of reducing the sample complexity and thus increasing the number of proteins identified. The increase of instrument time from 3 hours (triplicate of whole digest) to 10 hours (10 fractions), is a disadvantage (212 protein IDs/h vs. 102 protein IDs/h). However, fractionation allows a greater amount of sample to be loaded onto the nano-LC-column (2 μg versus 200 ng) and thus yields a 61.6% increase in number of proteins identified compared to employing nano-LC-MS without pre-fractionation.
2.5 References


CHAPTER 3

Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation During Proteomic Workflows


3.1 Introduction

Label-free quantification for proteomic analyses has gained popularity throughout the last decade. Advantages of label-free approaches compared to label incorporated methods (e.g. SILAC\textsuperscript{1,2}) include simplicity of sample preparation and applicability to any organism. Additionally, reduced sample complexity allows for an increase in the number of peptides sequenced, which results in a greater dynamic range and more comprehensive proteome coverage\textsuperscript{3,4}. Spectral counting\textsuperscript{5} and ion abundance\textsuperscript{6-10} have been used for label-free quantification and are known to correlate with protein abundance\textsuperscript{11}. Furthermore, an algorithm combining spectral counting and ion abundance measurements was developed by Feener and co-workers\textsuperscript{12}, enabling quantification for an increased number of proteins.

The number of spectral counts (SpCs) for a protein is simply the number of MS/MS spectra that result in identification of its proteolytic peptides. In bottom-up proteomic strategies, data-dependent MS/MS acquisition software selects peptide ions based upon their abundance and charge state, which typically favors identification of more abundant peptides/proteins. Applying dynamic exclusion for previously selected peptides limits the number of SpCs for those abundant peptides and enables the selection of peptides of lower abundance, resulting


in higher protein sequence coverage, more confident protein identifications and increased depth of proteome coverage\textsuperscript{13}.

Normalization of SpCs is performed to reduce the variance observed between samples and replicates. The variability in SpCs can be caused by numerous factors including sample preparation, gel-to-gel variance (if a gel-based approach is, part of the proteomic workflow), and changes in chromatography. Carvalho \textit{et al.}\textsuperscript{14} pointed out the importance of normalizing spectral counting data in order to quantify proteins; however, the most effective normalization method for label-free spectral counting has yet to be elucidated.

For total spectral count (TSpC) normalization\textsuperscript{15}, the technical replicate with the highest number of TSpC is chosen and the remaining technical replicates for that sample are normalized to it. Subsequently, the values across different samples are normalized to the sample with the highest technical replicate TSpC. The normalization is done for each protein individually and comparisons of average normalized SpCs of the same proteins are made between samples. Comparison of absolute or normalized SpCs between proteins to determine their relative abundance is generally precluded by that fact that longer proteins yield a higher number of spectral counts on average than shorter proteins. To account for this, Washburn and co-workers\textsuperscript{16-18} developed a method, termed normalized spectral abundance factor (NSAF), in which the SpC for a given protein are divided by its length (L) to give a spectral abundance factor (SAF). To account for variations between runs, the SAF for a given protein (\textit{e.g.}, SpC/L) is subsequently normalized to the sum of all SAFs for proteins identified within that run to create a normalized SAF (\textit{i.e.}, NSAF) that can be used to compare the relative abundance of proteins both between and within samples. In the latter method, the authors are
using the assumption that the sum of all SAF should be conserved between replicates to correct for differences in sampling rates. In TSpC normalization, the assumption is the sum of all SpC (i.e., the TSpC) should be conserved between runs/samples. In comparison, normalization to selected proteins (NSP) does not assume the sampling rate should be conserved between replicates/samples; rather, it relies on the premise that the total SpC for a standard protein should be conserved between replicates/samples if it is present at the same concentration. In practice, the standard protein(s) can be an endogenous house-keeping protein or an exogenous protein; the latter has the advantage of knowing precisely the amount of protein added to each sample. Any change observed in the SpC for the standard protein is assumed to reflect the variation between replicates and samples for the entire identified proteome. Consequently, the relative changes in SpC for the standard protein between replicates/samples are used as correction factors to normalize the SpC of all proteins.

Significant biological change in protein abundance is determined either by fold-change\textsuperscript{19,20}, by significance testing\textsuperscript{21-23} or by using a combination of both\textsuperscript{14}. Even though SpC datasets do not necessarily meet the criteria for a normal distribution, student’s t-tests are often applied in such experiments to determine statistical significance. Zhang \textit{et al.}\textsuperscript{21} performed a control experiment with yeast samples and 6 spike-in proteins at 3 different concentrations to calculate the false positive rates (FPR) for different significance tests. When only one replicate was utilized and thus the assumption of a normal distribution was obviously false, the G-test provided the lowest FPRs. However, when three replicates were utilized, t-testing performed similarly.
We are utilizing label-free quantification to gain proteomic insights related to pathogenicity of the fungus *Magnaporthe oryzae* (*M. oryzae*). *M. oryzae* causes rice blast disease destroying millions of hectares of rice each year, resulting in losses valued at billions of dollars\(^{24}\). Since half of the world’s human population relies on rice as a nutrition source\(^{25}\), understanding fungal development as it relates to disease progression is important for development of control strategies. Dean and coworkers sequenced the whole genome of *M. oryzae* in 2005\(^{26}\) providing a reliable database for proteomic approaches. Using 1D gel and liquid chromatography-mass spectrometry (GeLC-MS) we would like to study the development of the fungus *M. oryzae*, wild type and mutants, at different time points in its life cycle. Since the production of biological samples over a wide range of conditions (e.g. time, treatments, mutants) is difficult and downstream proteomic workflow is very time consuming, establishing a confident normalization method is critical. Thus, we have used *M. oryzae* conidia spectral counting data to compare the normalization methods TSpC, NSAF and NSP in their ability to account for variance between samples due to differences in sample preparation and chromatographic performance.

3.2 Experimental Section

3.2.1 Sample Preparation

*M. oryzae* conidia were harvested from 8 day old minimal medium plates. Three biological replicates, each containing 2 million conidia, were pooled to account for biological variance. Conidia were lysed by bead beating in a 1X PBS (Fisher Scientific, Pittsburgh, PA), 2 M Urea (Sigma Aldrich, St. Louis, MO), and 0.1 % SDS buffer (Bio-Rad, Hercules, CA) to
create the biological sample from which all experiments were derived. Protein concentration was determined via a BCA assay (Thermo Fisher Scientific, Rockford, IL). Samples 1 and 2, derived from the same biological sample, were prepared and processed on different days. Chicken myoglobin (Sigma Aldrich) and equine ovalbumin (Sigma Aldrich) were chosen as spike-in proteins and 25 ng of each was added to 50 µg of total protein for each sample. The samples were loaded onto 10-20% gradient 1D-SDS PAGE gels (Bio-Rad). It is necessary to start out with more material in regards to the number of fractions, the recovery, and the number of injections of a complex protein/peptide mixture. 50 µg is the amount of material loaded for the size gels utilized and it allows for adequate amounts of peptide material to be recovered (post digestion). After Coomassie staining (Bio-Rad), 10 fractions of gel bands were excised and in-gel digestion was performed on each fraction. Briefly, each gel-fraction was destained with 100 µl of 50:50 ammonium bicarbonate (Sigma Aldrich)/acetonitrile (ACN) (Burdick and Jackson, Muskegon, MI). Reduction was performed with 100 µl of 10 mM dithiothreitol (Sigma Aldrich) at 56 °C for 30 min, alkylation with 100 µl of 90 mM iodoacetamide (Sigma Aldrich) in the dark for 30 min and digestion overnight at 37 °C with trypsin (protein: protease ratio of 5:1). Acetonitrile was added and discarded between each step to dehydrate the gel pieces. To extract the peptides, 200 µl of 5% formic acid (Sigma Aldrich) in ACN was added to each fraction and incubated for 15 min at 37 °C. The supernatants were transferred into new tubes. ACN (100 µl) was added to the gel pieces and supernatants were combined; this was repeated once more for each fraction. Sample 2’ was produced by pooling one-third of the volume of adjacent in-gel digested fractions of sample 2 together to give a sample with only 5 gel fractions, yet having
undergone the same sample processing as the 10 gel fraction sample. All samples were dried down and stored at -20 °C until nanoLC-MS/MS analysis.

3.2.2 NanoLC-MS/MS

A 75 μm i.d. IntegraFrit capillary (New Objective, Woburn, MA) trap was packed to 5 cm with Magic C18AQ packing material (Michrom BioReasources, Auburn, CA). A 75 μm i.d. PicoFrit capillary column (New Objective, Woburn, MA) was packed 15 cm with the same packing material. Separation was carried out using a nanoLC-1D+ system from Eksigent (Dublin, CA) with a continuous, vented column configuration as previously reported by our group. A 2 μl (200 ng) sample was aspirated into a 10 μl loop and loaded onto the trap. Only 200 ng were analyzed per injection so as not to overload the nanoLC column. The flow rate was set to 350 nl/min for separation on the analytical column. Mobile phase A was composed of 98% H₂O (Burdick and Jackson), 2% ACN and 0.2% formic acid (Sigma) and mobile phase B was composed of 98% ACN, 2% H₂O and 0.2% formic acid. A 1 hour linear gradient from 5% to 50% B was performed. All measurements were performed at room temperature and 3 technical replicates of each sample were run to allow for statistical comparisons between samples, which are necessary for label-free quantification.

A hybrid LTQ-Orbitrap XL MS (Thermo Fisher Scientific, Bremen, Germany) was used to perform MS analysis. For data dependent acquisition, the parameters recently published by our group to be optimal for achieving maximum proteome coverage were used verbatim. External calibration was performed following manufacturer instructions and using the
manufacturer’s calibration mix and lock mass internal calibration using polydimethylcyclosiloxane (m/z 445.120025) was enabled\textsuperscript{30}.

3.2.3 Data Analysis

Data analysis was performed by searching each .RAW file, independently, against a concatenated target-reverse \textit{M. oryzae} database (MG8_GeneCall10.fasta) from the Broad Institute using MASCOT Distiller version 2.3.01 (Matrix Science Inc., Boston, MA). MASCOT parameters were ±5 ppm peptide ion tolerance, ±0.6 Da MS/MS fragment ion tolerance and 2 allowed missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine and deamidation of glutamine and asparagine were variable modifications. Peptide lists (.dat files) were created for each .RAW file by MASCOT. ProteoIQ version 2.1.01_SILAC_beta08 (BioInquire, Athens, GA) was used to create 5 different label-free spectral counting projects: a) Sample 1, b) Sample 2, c) Sample 2’, d) combination of sample 1 and 2 and e) combination of sample 2 and 2’. A 1% protein FDR was applied to each project, independently (\textit{i.e.}, the FDR was calculated based on the cumulative results of the sample files included in that particular project)\textsuperscript{31}. Log\textsubscript{2} SpC ratios were calculated and a pairwise t-test was performed on the proteins identified in sample 1 and sample 2.
3.3 Results and Discussion

The experimental workflow is shown in **Figure 3.1** Sample 1 and sample 2 are derived from the same biological sample but were prepared and processed after protein extraction on different days.

**Figure 3.1** Experimental workflow. The samples were prepared and processed on two different days. Myoglobin (25 ng) and ovalbumin (25 ng) were added to 50 µg of *M. oryzae* conidial protein. One dimensional SDS-PAGE separation and in-gel digestion were performed. Ten fractions from day 1 and 10 fractions from day 2 were analyzed in triplicate by nanoLC-MS using different traps and columns for each sample set. Additionally, adjacent fractions from day 2 sample were pooled and also analyzed in triplicate by nanoLC-MS.

*M. oryzae* conidial protein spiked with equine myoglobin and chicken ovalbumin was loaded onto 1D-SDS-PAGE gels and in-gel digestion was performed. The 10 fractions of each sample were analyzed by nanoLC-MS/MS using different traps and analytical columns for
each sample. Additionally, one-third of the final volume of adjacent in-gel digested fractions of sample 2 was combined to produce sample 2’ (five fractions) and were also analyzed by nanoLC-MS.

**Figure 3.2** Venn diagrams showing the TSpCs and the total protein numbers identified at 1% protein FDR in each sample. Sample 2 yielded the highest number of identified proteins with 1,477. While a difference of 19% in the number of TSpCs between sample 1 and 2 was observed, a reduction in TSpCs by 50% was noticed when adjacent fractions were combined and only half the number of fractions were analyzed.

Sample 1 yielded 76,638 TSpC (the sum of 3 technical replicates) and 1,185 proteins identified (see **Figure 3.2**). The TSpC number for sample 2 was 95,025 and 1,477 proteins were identified. The number of shared proteins between sample 1 and 2 was 1,121. Sample 1 and sample 2 contained 64 and 356 unique proteins, respectively. The differences in protein identifications (24%) and number of TSpCs (19%) between sample 1 and 2 were caused by sample processing on different days, reagent quality, gel-to-gel variance and use of a different trap and column. A slightly higher percentage of difference in the number of proteins identified was observed in a study by Cooper et al.\textsuperscript{22}. They reported differences up to 30% in the number of proteins identified between 9 replicate soybean peptide samples.
spiked with different amounts (0.005 to 2.5 pmol) of tryptic peptides from bovine apotransferrin tryptic digest, separated by MudPIT and analyzed on a LTQ-Orbitrap XL mass spectrometer.

In order to determine if normalization methods can recover from variables such as large differences in sample complexity, we mimicked such a situation by doubling the sample complexity of sample 2 by pooling adjacent gel fractions to create sample 2’. The number of TSpC for sample 2’ was 49,067, about half of the TSpC of sample 2, which was expected due to the decrease in the number of fractions (10 fractions for sample 2 and only 5 fractions for sample 2’). Moreover only 1,087 proteins were identified from sample 2’ and were a subset of the total population of protein identifications from sample 2.

SpC scatter plots for proteins from combined analysis of sample 1 and 2 are shown in Figure 3.3. The first plot shows the unnormalized SpCs for each protein (sum of the three technical replicates) versus the average SpCs of the two samples. A regression line slope of 1 is anticipated in the absence of biological variation as is the case here. The unnormalized scatter plots show slopes of 0.901 and 1.099 for samples 1 and 2 respectively, indicating that sample handling on different days, gel to gel variance, difference in reagent quality and slightly different chromatography have some effect on SpC reproducibility. Normalizing the data with TSpC normalization corrected the slopes to 1.0093 for sample 1 and 0.9907 for sample 2, while normalization with NSAF corrected the slopes to 1.001 and 0.9878, respectively. NSP yielded some improvements by normalization to the spike-in proteins myoglobin and/or ovalbumin. NSP to myoglobin and ovalbumin yielded slopes of 1.0262 for sample 1 and 0.9738 for sample 2. Slopes of 0.968 for sample 1 and 1.032 for sample 2 were
observed by normalizing to myoglobin and normalizing to ovalbumin resulted in slopes of 1.0696 for sample 1 and 0.9304 for sample 2.

Figure 3.3 SpC scatter plots from combined analysis of sample 1 and 2. Unnormalized SpCs and normalized SpC (NSpC) data for each protein are plotted versus the average SpCs for the protein derived from both samples. NSAF normalization with slopes of 1.001 and 0.999 for sample 1 and 2 corrects best in comparison to NSP and TSpC normalization.

The sample complexity was doubled in the case of sample 2’ to simulate a drastic change in sample complexity and to evaluate the ability of normalization methods to compensate for that. In the unnormalized scatter plots, shown in Figure 3.4, the slope of sample 2’ (0.6968) was almost half of the slope of sample 2 (1.3032). The normalized plots show that normalization can correct even for such drastic differences in sample complexity.
The slope for sample 2’ gets corrected to 1.0188 with TSpC normalization and to 1.0122 with NSAF. In spite of good sequence coverage, and the use of two spike-in proteins with different attributes, NSP did not perform as well as TSpC and NSAF normalization; however, better performance was observed for ovalbumin compared to myoglobin. Slopes of 0.9791 for sample 2 and 1.0209 for sample 2’ were observed for normalization to ovalbumin, whereas normalization to myoglobin yielded slopes of 1.2636 for sample 2 and 0.7364 for sample 2’. Myoglobin was identified with 21 SpCs (8.98 NSpCs) in sample 1, 21 SpCs...
(7.18 NSpCs) sample 2, and 20 SpCs (13.24 NSpCs) sample 2’ across all three replicates. In comparison, ovalbumin was identified with 63 SpCs (26.66 NSpCs) in sample 1, 88 SpCs (30.1 NSpCs) in sample 2, and 45 SpCs (29.81 NSpCs) in sample 2’ correlating with the pattern of TSpCs in each sample (76,638, 95,025, 49,067). This observations suggests that as a larger protein ovalbumin had a greater number of SpC and, thus, was more sensitive to the variations between samples.

As an additional metric, we evaluated which normalization method gives rise to the lowest variance across the technical replicates. TSpC normalization and NSAF resulted in lower median coefficient of variations (CV) for the samples, while the median CVs using NSP were significantly higher when normalized to both spike-in proteins. These data indicate that TSpC and NSAF are the superior normalization techniques compared to NSP. This observation is most likely due to the fact that former methods utilize the entire identified proteome for normalization, which allows for better correction of variability within the similar biological samples. NSP may be better suited to instances in which global protein expression differences exist between two biological samples.

After identifying the optimal normalization method(s) we wanted to determine the best means for detecting true biological change between two samples. In previous concepts, one specific threshold has been applied to an entire data set in order to define what proteins are changing significantly. However, results from this study and others have shown higher SpC proteins may require separate criteria for detecting significant change due to their lower variance; conversely, that the high variance of low SpC proteins should preclude them from quantification. Moreover, various methods such as significance testing or applying fold-
change thresholds have been utilized with various criteria that are often selected arbitrarily and with disregard to their true predictive value.

**Figure 3.5** Volcano plots for comparing the normalized SpC between sample 1 and sample 2. The log$_2$ expression ratio is plotted versus the –log$_{10}$ of the p-value obtained from significance testing (pairwise t-test). A) Plot comparing all proteins identified between samples 1 and 2. Proteins outside the given fold-change limits or above the p-value cutoffs are considered to have significantly changed. The absolute number of proteins meeting each criteria are given in parenthesis. The Venn diagram shows the overlap in the number of proteins deemed to have changed when applying either a 2-fold change cutoff or p-value cutoff or 0.05. B) Plots comparing proteins at different SpC levels. The proteins in each plots are defined by S, the SpC obtained per replicate injection.

To demonstrate this we produced the volcano plots, shown in Figure 3.5, with the data obtained here between samples 1 and 2. Volcano plots have long been utilized in the genomic microarray analysis to quickly identify species that have both large and highly significant
changes and were more recently applied to proteomic spectral counting data sets by Yates and co-workers\textsuperscript{14}. In these plots, the expression change for a given protein is plotted on the x-axis while the corresponding statistical significance is plotted on the y-axis. In Figure 3.5A, the plot shows when applying a standard p-value cutoff of 0.05, 276 of 1511 proteins identified (~18%) between sample 1 and 2 would be falsely discovered to have changed in abundance. Similarly, 290 such proteins (~19%) would be falsely discovered to have changed if the traditional 2-fold threshold for expression change was utilized. Interestingly, we also found these two methods lead to false discovery of different proteins – in particular, different level proteins. As indicated by the Venn diagram in Figure 3.5A, only 100 of the same proteins were falsely discovered between the 0.05 p-value cutoff and 2-fold expression change threshold. Upon closer inspection, nearly all of proteins falsely discovered by the 2-fold cutoff were low abundance (\textit{i.e.}, low SpC) proteins, while those discovered by the 0.05 p-value cutoff were slightly biased towards higher abundance proteins. This phenomenon can be seen in Figure 3.5B, which depicts volcano plots for different SpC levels: S\leq3, 3<S\leq10 and S>10, where S indicates the SpC per replicate (6 in this case, 3 per sample). These plots show the fold-change distribution narrows as S increases, indicating more abundant proteins have lower variance data than less abundant proteins.

In Table 3.1 and Table 3.2, narrower SpC per replicate (S) bins were utilized to calculate the false positive rates (FPR) achieved at different SpC levels for different stringency criteria. Given we could be certain no biological change had occurred between samples 1 and 2, we were able to define the FPR for a given criterion as the number of proteins discovered (falsely) in that SpC bin by the total number of proteins (N) falling within that same bin.
When comparing different fold-change cutoffs (Table 3.1), there again appears to be a strong propensity for low SpC proteins to be falsely discovered. Across all fold-change cutoffs, the proteins with lower SpC have the higher FPR. Although a higher fold change cutoff could be applied to reduce the overall FPR for all proteins, these data suggest little confidence could be placed in results for low abundance even when higher stringency cutoffs are applied. For example, if a 2.5-fold cut-off was utilized the FPR higher abundance proteins (S>5) would be acceptable, but the FPR for lower proteins (S≤5) would be no better than 10% and would be particularly poor, 39%, for very low abundance proteins (S<1.67).

<table>
<thead>
<tr>
<th>Fold-Change Cutoff</th>
<th>SpC/Rep</th>
<th>S&lt;1.67</th>
<th>1.67≤S≤2</th>
<th>2&lt;S≤5</th>
<th>5&lt;S≤10</th>
<th>10&lt;S≤15</th>
<th>15&lt;S≤20</th>
<th>20&lt;S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td></td>
<td>80.9</td>
<td>73.7</td>
<td>68.1</td>
<td>53.5</td>
<td>51.8</td>
<td>34.7</td>
<td>24.8</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>77.3</td>
<td>59.6</td>
<td>45.0</td>
<td>27.1</td>
<td>21.3</td>
<td>8.3</td>
<td>6.1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>59.4</td>
<td>28.9</td>
<td>19.6</td>
<td>8.0</td>
<td>2.1</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>44.2</td>
<td>14.9</td>
<td>10.1</td>
<td>4.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>39.4</td>
<td>13.2</td>
<td>6.7</td>
<td>2.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>251</td>
<td>114</td>
<td>404</td>
<td>299</td>
<td>141</td>
<td>72</td>
<td>230</td>
</tr>
</tbody>
</table>

In Table 3.2, which shows the FPR for different stringency p-values, the opposite trend is observed. Excluding proteins with a low number of SpC (S<1.67), the FPR increases slightly as S proceeds from low values to higher values, indicating that more abundant proteins are more apt to yield a false positive at a given p-value. If a single p-value was selected, such as 0.01, this data shows both high abundance proteins (S>20) and very low abundance proteins (S<1.67) would yield the majority of false positives. The importance of these observations is
two-fold: first, the large variance of low SpC proteins increases their probability to have erroneously large fold-changes and, secondly, high SpC proteins have a greater propensity to yield low p-values simply as a result of their lower variance.

Table 3.2 False Positive Rates for SpC Bin Widths at Different p-Value Cutoffs

<table>
<thead>
<tr>
<th>p-value Cutoff</th>
<th>S&lt;1.67</th>
<th>1.67≤S≤2</th>
<th>2≤S≤5</th>
<th>5≤S≤10</th>
<th>10≤S≤15</th>
<th>15≤S≤20</th>
<th>20≤S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>30.5</td>
<td>25.4</td>
<td>27.7</td>
<td>31.1</td>
<td>36.9</td>
<td>30.6</td>
<td>36.5</td>
</tr>
<tr>
<td>0.05</td>
<td>19.8</td>
<td>14.0</td>
<td>14.9</td>
<td>18.1</td>
<td>21.3</td>
<td>16.7</td>
<td>24.3</td>
</tr>
<tr>
<td>0.01</td>
<td>10.7</td>
<td>2.6</td>
<td>3.7</td>
<td>4.3</td>
<td>5.7</td>
<td>6.9</td>
<td>9.6</td>
</tr>
<tr>
<td>0.005</td>
<td>10.7</td>
<td>0.9</td>
<td>1.7</td>
<td>2.7</td>
<td>3.5</td>
<td>5.6</td>
<td>3.9</td>
</tr>
<tr>
<td>0.001</td>
<td>7.8</td>
<td>0.0</td>
<td>0.2</td>
<td>0.7</td>
<td>1.4</td>
<td>0.0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Given the different protein levels are uniquely affected by the two testing methods, we sought to apply dual constraints in order to better control the FPR for all protein levels. In Table 3.3 are shown different combinations of p-value and fold-change cutoffs applied to the entire dataset as well as to different SpC levels. In general, researchers seek to maintain a low FPR; however, using constraints that yield too low of an FPR will result in a high false negative rate (FNR). Thus we think it reasonable that a FPR of 10% be sought, so as not to exclude too many true positives that would occur during future experiments. The combination that best accomplishes this for the entire dataset, regardless of SpC, is a p-value and fold-change cutoff of 0.1 and 2.5, respectively. Notice, this combination utilizes stringent fold-change cutoffs and lax p-value cutoffs. This is because the majority of proteins in the
dataset have a relatively low SpC (median $S = 5.3$) and are more strongly affected by fold-change cutoffs. Consequently, this combination results in undesirably low FPR higher SpC proteins ($S>5$) and undesirably high FPR for lower SpC proteins ($S\leq2$). Even when excluding very low abundant proteins from the data set (i.e., when considering only proteins having $S\geq1.67$), similar outcomes are reach. It is apparent, when looking at the other combinations there is no single combination that will satisfy the 10% target FPR across all SpC levels. As a result, we suggest applying different constraints to the different SpC levels.

<table>
<thead>
<tr>
<th>p-value</th>
<th>fold-change</th>
<th>ALL</th>
<th>$S\geq1.67$</th>
<th>$1.67\leq S \leq 2$</th>
<th>$2 &lt; S \leq 5$</th>
<th>$5 &lt; S \leq 10$</th>
<th>$10 &lt; S \leq 15$</th>
<th>$15 &lt; S \leq 20$</th>
<th>$20 &lt; S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.25</td>
<td>25.7</td>
<td>25.7</td>
<td>25.5</td>
<td>23.7</td>
<td>26.7</td>
<td>28.4</td>
<td>29.1</td>
<td>25.0</td>
</tr>
<tr>
<td>0.1</td>
<td>1.5</td>
<td>18.8</td>
<td>17.5</td>
<td>25.5</td>
<td>22.8</td>
<td>22.3</td>
<td>20.4</td>
<td>17.0</td>
<td>8.3</td>
</tr>
<tr>
<td>0.1</td>
<td>2</td>
<td>10.5</td>
<td>7.7</td>
<td>24.7</td>
<td>16.7</td>
<td>12.6</td>
<td>7.4</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
<td>2.5</td>
<td>7.4</td>
<td>4.4</td>
<td>22.3</td>
<td>11.4</td>
<td>7.2</td>
<td>4.0</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
<td>3</td>
<td>6.4</td>
<td>3.4</td>
<td>21.1</td>
<td>10.5</td>
<td>5.9</td>
<td>2.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>1.25</td>
<td>15.0</td>
<td>15.0</td>
<td>15.1</td>
<td>12.3</td>
<td>13.9</td>
<td>17.4</td>
<td>17.0</td>
<td>13.9</td>
</tr>
<tr>
<td>0.05</td>
<td>1.5</td>
<td>11.3</td>
<td>10.5</td>
<td>15.1</td>
<td>11.4</td>
<td>13.1</td>
<td>12.4</td>
<td>9.9</td>
<td>6.9</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>6.6</td>
<td>5.1</td>
<td>14.3</td>
<td>9.6</td>
<td>8.7</td>
<td>5.4</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>2.5</td>
<td>5.0</td>
<td>3.3</td>
<td>13.5</td>
<td>7.9</td>
<td>5.2</td>
<td>3.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>3</td>
<td>4.4</td>
<td>2.6</td>
<td>13.5</td>
<td>7.9</td>
<td>4.2</td>
<td>2.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>1.25</td>
<td>5.2</td>
<td>4.1</td>
<td>10.4</td>
<td>1.8</td>
<td>3.7</td>
<td>4.0</td>
<td>5.0</td>
<td>4.2</td>
</tr>
<tr>
<td>0.01</td>
<td>1.5</td>
<td>3.8</td>
<td>2.5</td>
<td>10.4</td>
<td>1.8</td>
<td>3.5</td>
<td>2.7</td>
<td>2.1</td>
<td>1.4</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>2.7</td>
<td>1.3</td>
<td>9.6</td>
<td>1.8</td>
<td>2.7</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>2.5</td>
<td>2.3</td>
<td>1.0</td>
<td>8.8</td>
<td>1.8</td>
<td>2.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>3</td>
<td>2.3</td>
<td>1.0</td>
<td>8.8</td>
<td>1.8</td>
<td>2.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.005</td>
<td>1.25</td>
<td>3.4</td>
<td>2.1</td>
<td>10.4</td>
<td>0.9</td>
<td>1.7</td>
<td>2.3</td>
<td>2.8</td>
<td>4.2</td>
</tr>
<tr>
<td>0.005</td>
<td>1.5</td>
<td>2.9</td>
<td>1.4</td>
<td>10.4</td>
<td>0.9</td>
<td>1.5</td>
<td>2.0</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>0.005</td>
<td>2</td>
<td>2.0</td>
<td>0.5</td>
<td>9.6</td>
<td>0.9</td>
<td>1.0</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.005</td>
<td>2.5</td>
<td>1.7</td>
<td>0.3</td>
<td>8.8</td>
<td>0.9</td>
<td>0.5</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.005</td>
<td>3</td>
<td>1.7</td>
<td>0.3</td>
<td>8.8</td>
<td>0.9</td>
<td>0.5</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.001</td>
<td>1.25</td>
<td>1.5</td>
<td>0.3</td>
<td>7.6</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>0.001</td>
<td>1.5</td>
<td>1.5</td>
<td>0.2</td>
<td>7.6</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>0.001</td>
<td>2</td>
<td>1.3</td>
<td>0.1</td>
<td>7.6</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.001</td>
<td>2.5</td>
<td>1.3</td>
<td>0.1</td>
<td>7.6</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.001</td>
<td>3</td>
<td>1.3</td>
<td>0.1</td>
<td>7.6</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>N</td>
<td>1511</td>
<td>1260</td>
<td>251</td>
<td>114</td>
<td>404</td>
<td>299</td>
<td>141</td>
<td>72</td>
<td>230</td>
</tr>
</tbody>
</table>
Since large fold-change cutoffs likely result in a large FNR for higher abundance proteins, we suggest applying more stringent p-value cutoffs and less stringent fold-change cutoffs for this set of proteins. For instance, we would not set the fold-change cutoff any higher than 1.5-fold for proteins having S>10 as the chances of observing a larger fold-change, particularly a fold increase, for high abundance proteins is reduced due to the low linear dynamic range of spectral counting. Conversely, we would suggest utilizing more stringent fold-change cutoffs and less stringent p-value cutoffs for low SpC proteins (i.e., S≤10). For very low SpC proteins (S<1.67) we would consider excluding these from consideration or applying both stringent p-value and fold-change cutoffs due to their disposition to yield false positives. Here we defined very low SpC proteins as those with less than 1.67 SpC per replicate injection, or an average of 5 total spectral counts between two samples. Earlier studies by Old et al.\textsuperscript{11} and Collier et al.\textsuperscript{3} also proposed a cutoff of 5 or more total spectral counts across 2 samples, both having triplicate injections, to ensure accurate quantification. Gammulla et al.\textsuperscript{32} utilized even more stringent criteria, allowing for quantification of proteins having 6 or more spectral counts in each sample when triplicate injections were performed. We chose to define our constraints and bin widths using SpC per replicate (S), such that they would be independent of the number injections and samples. Consequently, comparisons will be able to be drawn in future experiments regardless of the number of samples or replicates. It should be emphasized that these results were obtained in the context of a Gel-based proteomics experiments, which may have led to higher false positive rates than other sample preparation techniques (\textit{e.g.} MudPIT) due to inherent differences in their reproducibility. Additionally, the use of technical replicates rather than biological replicates here may have
contributed to the low variation observed for higher SpC proteins. As such, we would caution readers when applying these same criteria to their data. Instead, we would recommend performing similar control experiments in order to define the variability specific to their lab, protocol, and sample type.

3.4 Conclusions

TSpC normalization, NSAF and NSP for label-free spectral counting data was investigated on the in-gel tryptic digest of *M. oryzae*. Normalization to TSpC and NSAF normalization revealed very good correlations and low variance for all data sets. With normalization, correcting for variance caused by sample preparation, gel to gel variance, chromatographic performances, and even drastic changes in sample complexity was possible. We evaluated further that accurate quantification is dependent on the number of SpCs. When applying different constraints for significance tests and/or fold-change cutoffs, we observed biases in the FPR across different SpC levels. In particular, we observed higher SpC proteins to have lower variance data and, as a result, required less stringent fold-change cutoffs to achieve accurate quantification. Conversely, lower SpC proteins showed less reproducibility and required higher fold-change cutoffs in combination with significance testing to ensure accurate quantification. Consequently, we suggest applying different constraints to different SpC levels in order to circumvent these biases and maintain a constant FPR for all proteins.
3.5 References


(8) Bondarenko, P. V.; Chelius, D.; Shaler, T. A.: Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-


CHAPTER 4

In-Depth Analysis of the *Magnaporthe oryzae* Conidial Proteome

The following work was reprinted with permission from: Gokce E., Franck W. L., Oh Y., Dean R. A., and Muddiman D. C. *J. Proteome Res.* 2012, in print. Copyright © 2012 American Chemical Society.

4.1 Introduction

The filamentous fungus *Magnaporthe oryzae* (*M. oryzae*) is responsible for the destruction of millions of hectares of rice every year and impacts all rice growing regions of the world\(^1-3\). In many Asian countries rice is a primary source of nutrition and yield losses associated with rice blast disease pose significant current and future risks to rice production. The lack of durable resistance genes in rice necessitates more effective control strategies to prevent fungal infection of rice crops. More in depth studies of the pathosystem, including application of proteomics will not only provide deeper insight in the biology of the pathogen, but may reveal new opportunities for disease control.

Asexual conidia are the primary source of inocula initiating pathogenesis. *M. oryzae* conidia typically contain three cells in a pyriform structure which adhere strongly to rice plant surfaces via a spore tip mucilage that is generated upon hydration of the conidium\(^4\). Following attachment, a germ tube emerges from the conidium and a dome shaped, melanized appressorium is formed\(^5\). Turgor pressure generated within the appressorium facilitates invasion of the leaf via a penetration peg from which invasive hyphae develop\(^6\). As a hemibiotrophic fungus, *M. oryzae* colonizes the outer cell layers of the leaf in its biotrophic phase and eventually proceeds to a necrotrophic phase characterized by lesion development.
Aerial hyphae emerge from these lesions and produce conidiophores that release conidia to initiate additional infections\(^7\).

The genome of \textit{M. oryzae} was recently sequenced by Dean \textit{et al.}\(^8\) and is composed of 7 chromosomes, 41.7 million bases (Mb) and 11043 annotated genes. The latest genome annotation (MG8, Broad Institute) predicts the presence of 12991 proteins derived from 12827 protein coding genes. The availability of a high-quality genome sequence and annotation provides a reliable foundation for proteomic analyses.

Global proteomic analyses provide a framework towards an in depth understanding of an organism’s cellular processes and organization but are subject to numerous challenges. The complexity of the proteome far exceeds that of the genome when protein isoforms, alternative splicing variants and post translational modifications are considered. Although genetic information encoded by the genome remains constant across different tissue types, the composition of the cellular proteome can vary amongst tissues, time and space. Analysis of the transcriptome via DNA microarray analysis or RNA sequencing captures a portion of this variation. However, measurement of mRNA levels does not capture regulation at the levels of protein translation and degradation and as a result mRNA levels show limited correlation to protein levels\(^9\). Furthermore, the complexity of protein and peptide chemistry complicates global proteome analyses, and no one technology is currently able to capture the complete diversity of a proteome.

Mass spectrometry (MS) is to date the most versatile and comprehensive tool for characterization of proteomes. Despite the fast duty cycles of modern hybrid mass spectrometers the complexity of proteomes requires fractionation methods to be applied prior
to tandem MS analysis. One dimensional sodium dodecyl sulfate polyacrylamide (1D-SDS-PAGE) gels are a common way of performing protein level fractionation prior to proteomic analyses\textsuperscript{10,11}. However, for large scale analyses encompassing multiple biological conditions and replicates this method proves to be very time-consuming and labor-intensive. The filter aided sample preparation method (FASP)\textsuperscript{12,13} combined with stop and go extraction tip (StageTip) anion exchange fractionation was recently published\textsuperscript{14-17} as an alternative peptide level fractionation strategy. Utilizing this approach Mann and coworkers were able to identify 4206 proteins of the hippocampal membrane proteome\textsuperscript{14}. The ability of this approach to facilitate the sample processing is impressive.

Previous proteomic analyses of \textit{M. oryzae} were limited in scope. Kim \textit{et al.} reported the identification of four proteins induced during appressorium formation\textsuperscript{18}. Two additional studies targeted proteins secreted in response to nitrogen starvation\textsuperscript{19} (89 differentially expressed proteins) or appressorium formation on artificial surfaces\textsuperscript{20} (53 secreted proteins identified). Bhadauria \textit{et al.} found 31 proteins to be differentially regulated in the conidial proteome of Δcom1 mutant when compared to the wild type\textsuperscript{21}. Furthermore, in a previous study we reported identification of 1500 proteins from \textit{M. oryzae} conidia using a GeLC based approach\textsuperscript{11}. Herein we extend our analysis of \textit{M. oryzae} conidial proteome and report the identification and analysis of 2912 proteins identified at 1\% false discovery rate (FDR) from conidia of \textit{M. oryzae} using FASP digestion and StageTip fractionation followed by nanoLC-MS/MS analysis.
4.2 Experimental Section

4.2.1 Sample Preparation and Digestion

All chemicals were purchased from Sigma Aldrich, St. Louis, MO, unless otherwise stated. Conidia were harvested from 8 day old minimal medium agar as published in Gokce et al.11 Briefly, mycelia were scraped from the surface of the medium and gently agitated in water to release conidia from the conidiophores prior to filtration through miracloth (EMD Biosciences, La Jolla, CA) to remove mycelium. Conidia were pelleted by centrifugation at 3,000 × g for 10 min at 4°C and the resulting pellets were flash frozen in liquid nitrogen and stored at -80°C. Lysis buffer (100 µl per 2 million conidia) containing 2M urea, 1X phosphate buffered saline and 0.1% SDS was used to extract the proteins by bead beating with 100 mg of 0.5 mm Zirconia/Silica beads (BioSpec Products, Inc., Bartlesville, OK) followed by boiling for 5 min. Debris and beads were spun down by centrifugation at 10,000 × g for 10 min at 4°C. Protein concentrations were determined by Bicinchoninic acid assay. One sample was prepared for preliminary analysis (gradient study) and two samples (same biological replicate) of 50 µg protein each were prepared for the reproducibility experiment. The samples were digested and fractionated on different days following the filter aided sample preparation (FASP) method protocol published by Mann and coworkers13 with minor changes. Samples were dried down to 27 µl and 3 µl of 50 mM Dithiothreitol (DTT) (Thermo Fischer Scientific, Rockford, IL) was added to a final DTT concentration of 5 mM. Incubation for 30 min at 56 °C followed to reduce the protein disulfide bonds. The samples were then mixed with 200 µl of 8 M urea in 0.1 M Tris/HCl (pH 8.5) on Vivacon 500 30 kDa MW cutoff filter units (Vivacon products, Littleton, MA) and centrifuged at constant 21 °C to
prevent carbamylation at 14,000 × g in a refrigerated bench-top centrifuge (Eppendorf 5810R) (Eppendorf, Hauppauge, NY) equipped with a fixed angle rotor (F45-30-11) (Eppendorf, Hauppauge, NY) for 15 min. This step was repeated one more time and the flow through solvent was discarded before 100 µl of 0.05 M iodoacetamide was added. The samples were incubated for 30 min in the dark at room temperature for alkylation of the free thiols and centrifuged afterwards for 10 min. Three 100 µl 8M urea buffer washes and three 100 µl 0.05 M ammonium bicarbonate in H2O washes were performed. The flow-through collection vials were switched out with clean ones and 80 µl ammonium bicarbonate buffer and 2.5 µl of 0.2 µg/µl stock trypsin solution were added at a 1:100 enzyme:protein ratio. Digestion was performed for 16 h at 37 °C. The peptides were eluted with 40 µl ammonium bicarbonate buffer. Nanodrop 2000c (Thermo Scientific, Wilmington, DE) readings at 280 nm with 2 µl of each sample were taken and indicated ~98% peptide recovery.

4.2.2 StageTip Anion Fractionation

A Kel-F Hub, point style 3, gauge 16 (1.19 mm) needle with a 100 µl plunger assembly was used to extract 6 anion disks from an Empore anion extraction disk (3M, St. Paul, MN) and stacked into a 200 µl pipette tip (Eppendorf, Hauppauge, NY). The same needle and plunger were used to extract 18 C18-disks from an Empore C18 extraction disk (3M, St. Paul, MN) and 6 desalting StageTips were prepared by stacking 3 C18 disks into six 200 µl pipette tips. Seven StageTips (1 anion, 6 C18) were prepared for every sample17.

A swing bucket rotor (A-2-DWP) (Eppendorf, Hauppauge, NY) was placed into the refrigerated centrifuge with 2 robust pipette tip boxes in order to perform the StageTip
conditionings. The anion StageTips were conditioned by adding 100 μl methanol (Burdick and Jackson, Muskegon, MI) from the top of the pipette tip, placing into the pipette tip box and centrifuging at 1,800 × g for 2 min. Subsequently a 100 μl 1M NaOH and two 100 μl of Britton Robinson buffer pH 11 (Ricca Chemicals, Arlington, TX) washes were performed. C18 StageTips were conditioned by sequential washes with 50 μl of methanol, 50 μl 80% (v/v) acetonitrile (Burdick and Jackson, Muskegon, MI) containing 3% (v/v) trifluoroacetic acid and 50 μl H2O (Burdick and Jackson, Muskegon, MI).

The peptide samples were diluted to ~200 μl by adding 100 μl of Britton Robinson buffer pH 11 and then loaded on top of the anion StageTips. Peptides were eluted onto a C18 StageTip placed underneath the anion StageTip by centrifugation. Additional peptide elutions from the anion StageTip were performed with Britton Robinson buffers pH 8, 6, 5, 4, and 3 into individual C18 StageTips. Sodium chloride (0.25 M) was added to buffer pH 3 in order to ensure elution of all peptides from the anion StageTips. The C18 StageTips containing the fractionated samples were then washed with 50 μl 0.1% (v/v) trifluoroacetic acid in H2O and samples eluted with 100 μl 60% (v/v) acetonitrile using a syringe. Fractions were dried to completion and stored at -80 °C until analysis.

4.2.3 NanoLC-MS/MS

The nanoLC-MS/MS was performed on a cHIPLC-Nanoflex system (Eksigent, Dublin, CA) at 26 °C. A trap column 200 μm × 0.5 mm packed with Chrom XP C18-CL 3 μm 120 Å was in-line with a 75 μm × 15 cm analytical column packed with the same packing material. The fractions were diluted in 20 μl mobile phase A (98% H2O, 2% acetonitrile, 0.1% formic
acid), each run 4 μl of the sample was injected into the sample loop and loaded onto the trap column. A 350 nl/min flow rate was used during the run. Sixty min 5-50% B gradients were run for the preliminary analysis. A 210 min 5-50% B gradient and a shallow 5-35% B gradient for 210 minutes were also explored. The reproducibility experiments were all run with the shallow 210 min gradient. Three technical replicates of each fraction were recorded. MS analysis was performed on a hybrid LTQ-Orbitrap XL MS (Thermo Fischer Scientific, Bremen, Germany). Optimized instrument parameters for achieving maximum proteome coverage recently published by our group were used verbatim\textsuperscript{22} with one minor change. The MS/MS data was collected in centroid mode and not in profile mode, due to very large file sizes (above 2 GB) that could not be handled by Xcalibur version 2.0.7 SP1.

4.2.4 Data Analysis

Raw files were analyzed in the RawMeat program version 2.0, downloaded from http://vastscientific.com/rawmeat. Raw files were also searched against a concatenated target-reverse \textit{M. oryzae} database MG8 (\textit{Magnaporthe} comparative Sequencing Project, Broad Institute of Harvard and MIT, http://www.broadinstitute.org) and converted into .dat files through MASCOT Distiller version 2.3.2 (Matrix Science Inc., Boston, MA). In MASCOT the distiller option Orbitrap_low_res_MS2_2.opt was used as peak picking algorithm and search parameters were the same as in Gokce \textit{et al.}\textsuperscript{11}. ProteoIQ version 2.3.02 (BioInquire, Athens, GA) was used to determine the protein identifications (ID’s) at 1% protein FDR\textsuperscript{23}. The WoLF PSORT program\textsuperscript{24} was utilized to assign subcellular localization to all predicted proteins in version 8 of the \textit{M. oryzae} genome.
4.3 Results & Discussion

4.3.1 Gradient Study

The FASP-digestion and StageTip fractionation workflow is shown in Figure 4.1. The procedure is rapid and amenable to processing multiple samples simultaneously. M. oryzae conidial proteins extracted by bead beating were reduced with DTT and loaded onto molecular weight cutoff filters. Utilizing a 30 kDa molecular weight cutoff filter, 150 µg of total protein per sample was processed in one filter within 3 hours. In our experience, up to 30 samples can be processed simultaneously. SDS, which is advantageous during protein extraction but is incompatible with electrospray ionization mass spectrometry (ESI-MS), was thoroughly depleted from samples during processing. Highly concentrated Urea (8M) was utilized to weaken the hydrophobic SDS-protein interactions and as an aid in removing the SDS by interacting electrostatically with the peptides. After SDS depletion, alkylation with iodoacetamide and buffer exchange with ammonium bicarbonate, digestion was performed with trypsin. Anion exchange fractionation using a pH gradient (pH 11, 8, 6, 5, 4, 3) and C18 clean-up was performed using StageTips resulting in extremely clean samples. The fractions were subjected to triplicate nanoLC-MS analysis and no trap or column clogging was experienced using FASP in combination with the StageTip clean-up.
Figure 4.1 Schematic of FASP and StageTip workflow. After protein extraction and reduction the sample is loaded onto the 30 kDa molecular weight cutoff filter. Urea was used to help deplete the SDS from the sample. Reduction, alkylation, buffer exchange to ammonium bicarbonate and trypsin digestion were performed on the filter. StageTip anion exchange fractionation and subsequent C18 clean-up was performed at the peptide level.

Running 60 min gradients (5-50% B) for the six fractions we identified 1511 proteins, comparable to our GeLC study on similar samples analyzing 10 gel-fractions as previously published. By extending our gradients to 210 min we increased protein ID’s by 60% (60 min: 687 → 210 min: 1091) for a single fraction (fraction pH 8) (Figure 4.2A). Almost all identified proteins in the 60 min gradient run were within the identifications in the 210 min run. Furthermore, running a shallower gradient 5-35% B increased our number of protein ID’s by an additional 5% (210 min: 1091 → 210 min shallow: 1145). Although the intensity of the total ion chromatograms (TIC’s) dropped with both increased gradient times and shallower gradients, the chromatograms still appeared to be very complex attributable to the fact that more material can be loaded per run. Figure 4.2B shows the number of consecutive MS² scans, allowed was a maximum of 8 MS/MS per precursor ion scan. In the 60 min gradient run those 8 MS/MS events were hit most of the time (green colored bars), indicating under-sampling of the proteome. By performing longer separations (210 min
gradients) we saw an overall increase in the total number of MS/MS events and the shallow gradient decreased the frequency of 8 consecutive MS/MS events (rose colored bars) suggesting a more efficient sampling of the proteome.

**Figure 4.2** Total ion chromatograms of a 60 min gradient (5-50% B) in comparison to 210 min (5-50% B), and 210 min with a shallow gradient (5-35% B) for the pH 8 fraction are shown (A). The frequency of consecutive MS/MS events for each gradient performed is diagrammed in B. While the maximum of 8 consecutive MS/MS events was hit most of the time in the 60 min gradient, a more uniform distribution between consecutive scans can be seen for the 210 min gradient. A drop in 8 MS/MS events was only seen for the 210 min shallow gradient, suggesting an optimized window for peptide elution and detection.

**Table 4.1** shows additional information about the three different gradient runs. An approximate three-fold increase was observed in the number of precursor spectra (60 min: 3387→ 210 min: 9427→ 210 min shallow: 9753) as well as fragment ion spectra (60 min: 5672→ 210 min: 14341→ 210 min shallow: 17185) by switching to the longer gradients. The number of identified unique peptides was doubled (60 min: 2017→ 210 min: 3977→ 210 min shallow: 4290). The peptide yield was calculated by dividing the number of identified peptides by the number of MS/MS spectra, and although the peptide yield
decreased from 36% to 29% to 25%, the number of protein ID’s still increased due to the identification of new unique peptides derived from previously undetected proteins.

Table 4.1 Number of Full Scans, MS/MS Scans, Identified Peptides, Calculated Peptide Yield and Identified Proteins for the Different Gradients Used for the Single Fraction at pH8.

<table>
<thead>
<tr>
<th>Gradient Type</th>
<th>Full Scans</th>
<th>MS/MS Scans</th>
<th>Peptides</th>
<th>Peptide Yield</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 min gradient</td>
<td>3387</td>
<td>5672</td>
<td>2017</td>
<td>36%</td>
<td>687</td>
</tr>
<tr>
<td>210 min gradient</td>
<td>9427</td>
<td>14341</td>
<td>3977</td>
<td>28%</td>
<td>1091</td>
</tr>
<tr>
<td>210 min shallow gradient</td>
<td>9753</td>
<td>17185</td>
<td>4290</td>
<td>25%</td>
<td>1145</td>
</tr>
</tbody>
</table>

4.3.2 Reproducibility Experiment

Applying 210 min LC gradients on all fractions in the reproducibility experiment resulted in the identification of 2912 proteins, corresponding to 22.4% of the predicted proteome of *M. oryzae*. The 2912 identified proteins are ranked by abundance derived from Normalized Spectral Abundance Factor (NSAF) values with spectral count data included for each protein. The FASP and StageTip approach with 6 fractions run on 210 min gradients required 1 day of instrument time for each replicate per sample. It is possible that more extensive processing and fractionation may provide additional protein identifications. For example, utilizing milligram quantities of protein, protein level fractionation by gel filtration, multi-enzyme FASP digestion and dedicating 12 days of instrument time to a single sample, Mann and coworkers were able to identify 10255 proteins from HeLA cells. Limitations based upon the availability and efficient use of instrument time as well as the ability to generate...
milligram quantities of protein sample preclude identification of this many proteins in most analyses.

Table 4.2 False Positive Rate Table (s: spectral counts per replicate, N: number of proteins)

<table>
<thead>
<tr>
<th>p-value</th>
<th>fold-change</th>
<th>ALL</th>
<th>≥1.67</th>
<th>1.66&lt;s≤2</th>
<th>2&lt;s≤5</th>
<th>5&lt;s≤10</th>
<th>10&lt;s≤15</th>
<th>15&lt;s≤20</th>
<th>s&gt;20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.25</td>
<td>17.69</td>
<td>17.28</td>
<td>19.4</td>
<td>19.1</td>
<td>19.6</td>
<td>18.6</td>
<td>21.6</td>
<td>12.0</td>
</tr>
<tr>
<td>0.1</td>
<td>1.50</td>
<td>11.61</td>
<td>9.76</td>
<td>19.4</td>
<td>19.1</td>
<td>15.7</td>
<td>8.2</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
<td>2.00</td>
<td>7.04</td>
<td>4.37</td>
<td>18.3</td>
<td>16.7</td>
<td>7.3</td>
<td>1.5</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
<td>2.50</td>
<td>5.19</td>
<td>2.38</td>
<td>17.1</td>
<td>9.8</td>
<td>3.9</td>
<td>0.8</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
<td>3.00</td>
<td>3.91</td>
<td>1.06</td>
<td>16.0</td>
<td>4.2</td>
<td>1.7</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>1.25</td>
<td>11.16</td>
<td>10.44</td>
<td>14.2</td>
<td>11.6</td>
<td>11.1</td>
<td>12.0</td>
<td>14.5</td>
<td>6.3</td>
</tr>
<tr>
<td>0.05</td>
<td>1.50</td>
<td>7.35</td>
<td>5.73</td>
<td>14.2</td>
<td>11.6</td>
<td>8.6</td>
<td>5.5</td>
<td>6.2</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>2.00</td>
<td>4.81</td>
<td>2.84</td>
<td>13.1</td>
<td>10.2</td>
<td>4.8</td>
<td>1.1</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>2.50</td>
<td>3.81</td>
<td>1.91</td>
<td>11.9</td>
<td>8.8</td>
<td>2.7</td>
<td>0.8</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>0.05</td>
<td>3.00</td>
<td>2.88</td>
<td>0.89</td>
<td>11.3</td>
<td>3.3</td>
<td>1.4</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>1.25</td>
<td>4.57</td>
<td>2.97</td>
<td>11.3</td>
<td>2.8</td>
<td>3.4</td>
<td>3.2</td>
<td>4.6</td>
<td>1.9</td>
</tr>
<tr>
<td>0.01</td>
<td>1.50</td>
<td>3.33</td>
<td>1.44</td>
<td>11.3</td>
<td>2.8</td>
<td>1.7</td>
<td>1.7</td>
<td>2.9</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>2.00</td>
<td>2.54</td>
<td>0.72</td>
<td>10.3</td>
<td>1.9</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>2.50</td>
<td>2.23</td>
<td>0.64</td>
<td>9.0</td>
<td>1.9</td>
<td>0.9</td>
<td>0.6</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>0.01</td>
<td>3.00</td>
<td>1.89</td>
<td>0.34</td>
<td>8.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

N 2912.00 2356.00 556.00 215.00 776.00 474.00 241.00 158.00 492.00

In order to determine cutoffs for the detection of real biological change in future experiments via label free quantification for the FASP StageTip sample preparation and fractionation method, a false positive rate (FPR) table (Table 4.2) was created using data from replicate analyses of the a single conidial proteome sample where no difference in protein abundance is expected. Different fold-change and p-value cutoffs were applied across spectral count bin width (s stands for spectral count per replicate) and very similar FPR and protein distributions across the bin widths (N) were seen when compared to previous GeLC data\textsuperscript{11}.
Low FPR’s are colored in blue while high FPR’s are red. The FPRs indicate that lower fold change-cutoffs can be applied on higher spectral count bin widths while maintaining FPRs consistent with low spectral count bin widths. This is especially beneficial for very abundant proteins where commonly used 2 fold-changes are rarely observed. Defining 10% FPR to be acceptable for each spectral count bin width, a p-value of 0.05 and the appropriate fold-change cutoff will be applied on the spectral bin widths in further large scale experiments where detection of differential protein expression by label free quantification is desired.

4.3.3 Subcellular Localization Analysis

An analysis of the subcellular localization of the conidial proteome is presented in Figure 4.3. Predictions of subcellular localization for the entire *M. oryzae* predicted proteome were performed using the WoLF PSORT program\textsuperscript{24}. A qualitative analysis of the identified proteins indicates that the distribution of proteins in the predicted proteome (Figure 4.3A) is highly similar to that of the conidial proteome (Figure 4.3B). A higher percentage of cytoplasmic proteins were observed in the conidial proteome relative to the whole proteome (14.8% of the whole proteome, 26.7% of the proteins identified), likely due to the high solubility of this group of proteins. Conversely, nuclear proteins were not as well represented in the conidial proteome (27.5% of the whole proteome, 21.3% of the proteins identified). Interestingly, 249 proteins predicted to be extracellular were identified from conidia without specifically targeting extracellular proteins during sample preparation. However, when compared to the whole proteome, extracellular proteins represent a smaller percentage of the identified proteins (8.6% compared to 15.8% for the entire proteome).
Figure 4.3 Subcellular localization charts for the predicted whole proteome (A), sum of the number of proteins identified in each category (B), and the sum of NSAF values for proteins identified in each category (C). NSAFs indicate that while plasma membrane proteins are represented in our study (9.2% of total protein identifications) they are of lower abundance (3.6%). The same is true for extracellular proteins. More highly abundant proteins were mitochondrial and cytoplasmic proteins.

To obtain a more quantitative view of the distribution of proteins identified from conidia, NSAF values were calculated for each protein and a sum of NSAF values was generated for all proteins within each subcellular localization category (Figure 4.3C). The results of this analysis indicate that cytoplasmic and mitochondrial proteins represent not only the largest groups of identified proteins but also include proteins of high abundance. In contrast, the nuclear and extracellular protein categories that are underrepresented qualitatively also generally contain proteins of lower abundance or in the case of extracellular proteins were not specifically targeted in this study. Finally, although membrane localized proteins were identified at a rate comparable to that observed for the whole proteome, these proteins appear
much lower in abundance. The FASP methodology was previously reported to identify membrane proteins in an unbiased manner\textsuperscript{14}. However, it is unclear if the apparent low abundance of membrane proteins in conidia is due to inefficiencies in their isolation and detection using the current methodologies or simply reflects their true biological status.

4.3.4 Gene Ontology Analysis

Further classification of the conidial proteome was performed using gene ontology (GO) analysis. GO terms were assigned to the \textit{M. oryzae} proteome using the BLAST2GO program\textsuperscript{27} and tested for enrichment relative to the entire proteome. Fisher’s Exact test was used within BLAST2GO at an FDR of 5\% and reduced for most specific terms. A total of 12 GO terms were underrepresented and are primarily related to DNA metabolism, transcription and extracellular proteins in agreement with subcellular localization analysis and further supporting the observation that nuclear and extracellular proteins are under sampled. A total of 35 GO terms were overrepresented and primarily reflect classes of highly abundant proteins. The number of proteins identified for selected GO Process categories is presented in Figure 4.4 with a calculated sum of NSAF values for all proteins associated with a specific GO term presented above. Comparison of the sum of NSAF values to the total number of identified proteins for a given GO term indicates that five GO terms disproportionately contain highly abundant proteins and include: translation, carbohydrate metabolism, generation of precursor metabolites and energy, response to stress and secondary metabolic process.
Figure 4.4 Gene ontology analysis of biological process terms of observed proteins and their relative abundance levels determined by summing NSAF values of the proteins in each category. Beige bars highlight categories with disproportionally highly abundant proteins.
4.3.5 Biological Insights into Identified Proteins

Examination of the 10% of proteins (291 proteins) with the highest NSAF values provides clues regarding the metabolic processes required for conidial germination. The translation GO term (GO:0006412) contains primarily ribosomal proteins, translation factors and tRNA synthetases and 80 of these proteins fall within the highest 10% of NSAF values. A total of 11 proteins corresponding to components of the ATP synthase complex were identified in the conidial proteome and belong to the generation of precursor metabolites and energy GO term (GO:0006091). Eight of the ATP synthase complex proteins are amongst the top 10% of NSAF values and four are among the top 20 most abundant proteins as determined by NSAF values. Glycolysis (8 proteins), the pyruvate dehydrogenase complex (4 proteins), and the citrate cycle (11 proteins) account for the highly abundant proteins in the carbohydrate metabolism GO term (GO:0005975). The response to stress GO term (GO:0006950) contains 15 chaperone proteins that account for 42% of the sum of NSAF values for that GO term.

Proteins involved in synthesis of the polyketide melanin, ALB1 (MGG_07219), RSY1 (MGG_05059), BUF1 (MGG_02252) in addition to a BUF1 homolog, 4HNR, (MGG_07216), account for 33% of the sum of NSAF values for the Secondary metabolism GO Term (GO:0019748). Three of these proteins, RSY1, BUF1, and 4HNR are within the top 10% of most abundant proteins. Previous studies have established a melanin requirement for pigmentation of *M. oryzae* cultures and also formation of functional appressoria prior to plant infection\(^{28-31}\). *M. oryzae* strains carrying mutations in the hydroxynapthalene reductase *BUFI* (MGG_02252), the polyketide synthase *ALBI* (MGG_07219), and the scytalone dehydratase *RSY1* all display abnormal culture pigmentation and are nonpathogenic due to
formation of defective appressoria. BUF1 and 4HNR were amongst the most abundant proteins in our data set. Interestingly 4HNR was shown previously to catalyze the reduction of tetrahydroxynaphthalene to scytalone, a function often associated with BUF1. \textit{M. oryzae} mutant strains lacking a functional 4HNR retained the ability to produce melanin and form functional appressoria presumably due to the ability BUF1 to catalyze the reduction of both tri- and tetrahydroxynaphthalene. The high abundance of 4HNR in conidia supports the hypothesis that, although not specifically required, it contributes to the normal pathway for melanin biosynthesis \textit{M. oryzae}.

\textit{M. oryzae} conidia utilize storage reserves including glycogen, mannitol and trehalose for germination, hyphal development and appressorium formation. Glycogen is synthesized from UDP-Glucose in part via the activity of glycogen synthase. UDP-glucose is derived from Glucose-6P, an intermediate of glycolysis, via the activities of phosphoglucomutase and UTP-glucose-1-phosphate uridylyltransferase. We identified four highly abundant conidial proteins predicted to be involved in synthesis of glycogen including a putative glycogen synthase (MGG_07289), a putative phosphoglucomutase (MGG_04495), and a putative UTP-glucose-1-phosphate uridylyltransferase (MGG_01631). A fourth protein, 1,4-alpha-glucan-branching enzyme (MGG_03186), is a homolog of yeast GLC3 (63% amino acid identity), a protein involved in the generation of 1,6-alpha-glucosidic linkages within the glycogen molecule. The utilization of conidial glycogen reserves during germination and appressorium formation was documented previously. Glycogen is broken down to glucose-1P via the activity of glycogen phosphorylase (MGG_01819), also detected at high levels in conidia. The importance of glycogen homeostasis in conidiation and conidial germination of
M. oryzae is reflected in the abundance of conidial proteins involved in glycogen metabolism.

M. oryzae conidia also contain lipid droplets that are mobilized during germination and appressorium formation\textsuperscript{37}. One of the most abundant proteins in the dataset was MGG_11916, a homolog of the Colletotrichum gloeosporioides CAP20 protein\textsuperscript{38} and the Metarhizium anisopliae MPL1 protein\textsuperscript{39}. The CAP20 gene of C. gloeosporioides is expressed in conidia and during appressorium formation and is necessary for pathogenicity. The MPL1 gene of M. anisopliae, a homolog of mammalian perilipins, was highly expressed in conidia and localized to lipid droplets. In addition, ΔMPL1 mutant conidia contain fewer lipid droplets. MPL1 is also required for generation of turgor pressure in developing appressoria and is required for full virulence in M. anisopliae. The high conidial abundance of MGG_11916 and its homology to MPL1 (47% amino acid identity) suggests a role in maintenance and hydrolysis of lipid droplets of M. oryzae conidia.

Utilization of lipid reserves requires beta oxidation of fatty acids liberated via lipase activity. A prior investigation of the M. oryzae genome indicated the presence of 16 proteins involved in β-oxidation of fatty acids\textsuperscript{40}. A total of 15 of these proteins were identified in the conidial proteome, including a peroxisomal multifunctional β-oxidation protein, MFP1, previously demonstrated to be required for appressorium-mediated plant infection. Furthermore, two putative Acyl-CoA dehydrogenases (MGG_08661 and MGG_05949) predicted to catalyze the first step in β-oxidation of fatty acids were also identified at high abundance. β-oxidation of fatty acids results in the production of acetyl-CoA which is further metabolized via the glyoxylate cycle. Isocitrate lyase is required for efficient germination of
conidia as well as pathogenicity in *M. oryzae*\(^{41}\). Both enzymes unique to the glyoxylate cycle, isocitrate lyase (MGG\(\_\)04895) and a putative malate synthase (MGG\(\_\)02813) were identified and their relative abundances are similar to other enzymes of the citrate cycle indicating that *M. oryzae* conidia contain a functional glyoxylate bypass and are capable of utilizing Acetyl-CoA derived from fatty acid \(\beta\)-oxidation.

Several proteins involved in the detoxification of reactive oxygen species (ROS) were also detected at high levels in conidia. The most abundant of these is the peroxiredoxin, HYR1 (MGG\(\_\)07460). HYR1 is a homolog of the yeast glutathione-dependent peroxidase HYR1p and *M. oryzae ΔHYRI* mutants display growth impairments in the presence of \(H\_2O\_2\)\(^{42}\). HYR1 was demonstrated to be localized to vacuoles during the earliest stages of conidial germination on the surfaces of barley leaves and is required for efficient response to plant-derived ROS during infection. Additional abundant ROS detoxification proteins identified include a putative superoxide dismutase (MGG\(\_\)02625), a putative monofunctional catalase CATA (MGG\(\_\)10061), and the bifunctional catalase/peroxidases CPXA/KATG1\(^{43}\) (MGG\(\_\)04337) and CPXB\(^{44,45}\) (MGG\(\_\)09834). Catalase activity was previously reported in supernatants of both conidial suspensions and hyphal cultures. CPXB was described as an extracellular catalase/peroxidase and \(ΔCPXB\) mutants have reduced extracellular catalase activity and increased sensitivity to \(H\_2O\_2\) in culture. The detection of several highly abundant proteins involved in ROS detoxification suggests that *M. oryzae* conidia are subject to oxidative stress and maintain high levels of proteins capable of mitigating the negative effects of ROS exposure.
Overall, the analysis of identified proteins suggests that conidia are primed for germination and rich in proteins required for basic housekeeping functions. Proteins required for utilization of storage reserves during germination are also highly abundant in conidia. Furthermore, proteins with known roles in the biology of the conidium were often identified in high abundance.

4.4 Conclusions
We successfully identified 2912 proteins of the _M. oryzae_ proteome by applying FASP and StageTip fractionation combined with long and shallow nanoLC gradients to conidial protein samples. In comparison to GeLC we achieved a doubling in the number of protein identifications with this new methodology and observed marginally lower FPR values for the reproducibility experiment. Subcellular localization analysis of the proteome revealed in general a comparable distribution of our data to the predicted proteome. Gene ontology analysis indicates that housekeeping proteins and components of central metabolic processes are well represented in the highly abundant proteins of the _M. oryzae_ conidial proteome. In addition, analysis of the conidial proteome indicates that proteins important to the physiology of conidia such as those involved in glycogen metabolism and β-oxidation of fatty acids are present and often highly abundant. The depth of proteome coverage for _M. oryzae_ conidia presented here provides an excellent foundation for future investigations of pathogenic development of _M. oryzae_ including germination of conidia and development of appressoria.
4.5 References


(43) Zamocky, M.; Furtmüller, P. G.; Bellei, M.; Battistuzzi, G.; Stadlmann, J.; Vlasits, J.; Obinger, C.: Intracellular Catalase/Peroxidase from the Phytopathogenic Rice Blast Fungus


CHAPTER 5
Characterization of the *Magnaporthe oryzae* Proteome During Conidial Germination and cAMP Mediated Appressorium Formation

The following work was part of a collaborative effort and is submitted to *Molecular Cellular Proteomics*.

5.1 Introduction

The rice blast pathogen, *Magnaporthe oryzae*, is a filamentous fungus belonging to the Ascomycota. Rice blast is the most serious disease of rice worldwide and its presence has been documented in more than 85 countries including all major rice producing regions\(^1\). Yield losses in cultivated rice associated with infection by *M. oryzae* can exceed 50% with typical losses ranging from 10 to 30%\(^1\). *M. oryzae* also infects other grass hosts including wheat, barley and millet\(^2\). The importance of rice as a staple food, the genetic tractability of *M. oryzae*, and the availability of genome sequences for both the fungus and the plant host have made the rice-*M. oryzae* pathosystem a predominant model for the study of plant-fungal interactions.

Rice blast disease is initiated upon contact of an asexual conidium with the plant surface. Attachment to the leaf surface is mediated by spore tip mucilage produced at the conidial apex upon hydration of the conidium\(^3\). Germination of the conidium results in the production of a germ tube from which development of a penetration structure known as an appressorium occurs in response to plant-derived cues. Appressoria of *M. oryzae* are characterized by a heavily melanized cell wall that facilitates generation of a large internal turgor pressure via accumulation of compatible solutes that promote influx of water into the appressorium\(^4\). This
pressure produces the mechanical force required for penetration of the plant cuticle by a penetration peg formed at the appressorium pore. Upon gaining entry into the plant, a primary invasive hypha develops and gives rise to secondary bulbous hyphae. The early stages of infection are characterized by a biotrophic phase in which host cell death is not evident and *M. oryzae* evades host defenses by mechanisms that are currently not well understood but likely involve the delivery of effector proteins responsible for suppression of host defense responses. During later stages of infection, the pathogen enters a necrotrophic phase in which host cell death occurs and necrotic lesions develop. Conidiophores emerging from these lesions produce conidia that serve as inocula for subsequent infection.

Appressorium formation can be induced *in vitro* by germination of conidia on hard, hydrophobic surfaces that mimic the waxy cuticle of rice leaves. A requirement for nutrient limitation or starvation was also proposed for initiation of appressoria. A number of chemical elicitors are also known to stimulate appressorium formation *in vitro* on hydrophilic surfaces that do not normally support the establishment of appressoria. The plant-derived cutin monomers, *cis*-9,10-epoxy-18-hydroxyoctadecanoic acid, *cis*-9-octadecen-1-ol, and 1,16-hexadecanediol stimulate differentiation of appressoria. Furthermore, addition of the secondary messenger, cAMP, its analog 8-Bromo cAMP or the phosphodiesterase inhibitor, 3-isobutyl-1-methylxanthine, to conidial solutions on non-inductive hydrophilic surfaces is sufficient to trigger development of appressoria.

Previous studies established a role for G-protein and cAMP signaling in the initiation and development of appressoria in *M. oryzae*. Loss-of-function mutations in the adenylate cyclase, Mac1, and the G-protein alpha subunit, MagB, result in the inability to form
appressoria on inductive surfaces\textsuperscript{10,11}. Appressorium formation in the mac1 and magB mutants can be restored by addition of exogenous cAMP to germinating conidia\textsuperscript{10,11}. During appressorium development, an increase in intracellular cAMP levels mediated by adenylate cyclase activity stimulates release of the regulatory subunit of the protein kinase A holoenzyme from the catalytic subunit which is then free to phosphorylate downstream targets. Deletion of the cAMP dependent protein kinase catalytic subunit (cpkA) causes a significant delay in appressorium formation and reduction in the total number of appressorium formed resulting in a loss of pathogenicity\textsuperscript{12,13}. A point mutation was described previously in the regulatory subunit of protein kinase A, sum1, resulting in the constitutive activation of CPKA leading to the formation of appressoria on non-inductive surfaces in absence of external stimulus\textsuperscript{14}. Furthermore, roles in the maintenance of intracellular cAMP levels during pathogenic development were recently demonstrated for the high and low-affinity phosphodiesterases, PdeH and PdeL, which catalyze the hydrolysis of cAMP\textsuperscript{15,16}. A clear link between cAMP signaling, PKA activity and pathogenic development has been established in \textit{M. oryzae}. However, the biological processes regulated by the pathway remain largely uncharacterized. A yeast two hybrid assay identified seven proteins that interact directly with CPKA\textsuperscript{17}. Recently, the expression of two transcriptional regulators, CDTF1 and SOM1, were reported to be reduced in \textit{mac1} and \textit{cpkA} mutants with SOM1 also being shown to interact weakly with CPKA\textsuperscript{18}. Mutational analysis of these two genes indicated that each has multiple pleiotrophic effects on growth and development in \textit{M. oryzae} \textsuperscript{18}. Furthermore, a total of 110 cAMP responsive sequence tags were identified in a SAGE analysis of conidia following cAMP treatment of which 60 (50 up-regulated and 10 down-regulated) were
assigned to a gene or EST\textsuperscript{19}. In a DNA microarray based analysis of conidia germinated for nine hours in the presence or absence of cAMP, a total of 1014 transcripts were differentially expressed (644 up-regulated and 370 down-regulated)\textsuperscript{20}. In addition, appressorium formation on a hydrophobic surface was also investigated and set of 357 consensus appressorium genes regulated in both cAMP-induced and surface-induced appressoria was obtained. Additional transcriptomic studies have analyzed the \textit{M. oryzae} transcriptome in response to nitrogen starvation\textsuperscript{21}, appressorium formation and the Pmk1 map kinase signaling pathway\textsuperscript{22}, and the \textit{in planta} transcriptome\textsuperscript{23-25} using a variety of technologies including EST analysis, DNA microarrays, RNAseq and HT-SuperSAGE technologies.

Although extensive analysis of the \textit{M. oryzae} transcriptome has been performed to date, investigations of the \textit{M. oryzae} proteome remain limited in number and scale. A total of four proteins were identified as induced during appressorium formation on an inductive wax surface in a 2D gel based approach\textsuperscript{26}. Two studies targeting secreted proteins identified 89 and 53 differentially expressed proteins in response to nitrogen starvation and appressorium formation on an inductive surface respectively\textsuperscript{27,28}. Comparison of the conidial proteomes of wild-type \textit{M. oryzae} and a COM1 mutant whose protein product is required for normal conidial morphogenesis revealed 31 proteins that changed in abundance\textsuperscript{29}. Finally, previous studies from our group\textsuperscript{30,31} have established a comprehensive characterization of the \textit{M. oryzae} conidial proteome, the most recent having identified 2912 proteins from conidia using the filter aided sample preparation method (FASP) followed by stop and go extraction tip (StageTip) anion exchange fractionation in combination with nanoLC-MS/MS\textsuperscript{32}. 
Here we extend our proteome analysis of *M. oryzae* through characterization of the proteome during conidial germination and cAMP-induced appressorium formation. Label free quantification via spectral counting facilitated identification of proteins differentially expressed during conidial germination and appressorium formation. Additionally, comparison of the proteomes of wild-type *M. oryzae* and a ΔcpkA mutant strain offers further insight into the role of cAMP signaling during pathogenic development.

5.2 Experimental Section
5.2.1 Strains and Culture Conditions

Wild-type *M. oryzae* strain 70-15 cultures were maintained at 25°C under constant illumination on a minimal medium agar consisting of the following components per liter: 10g Sucrose, 6g NaNO₃, 0.52g KCl, 0.52g MgSO₄·7H₂O, 1.52g KH₂PO₄, 5µg Biotin, 1mg Thiamine and 1ml of 1000X Trace element solution (2.2g ZNSO₄, 1.1g H₃BO₃, 0.5g MnCl₂·4H₂O, 0.5g FeSO₄·7H₂O, 0.17g CoCl₂, 0.16g CuSO₄·5H₂O, 0.15g Na₂MoO₄·2H₂O and 5g NaEDTA per 100ml). *Escherichia.coli* strain DH5α was maintained on LB medium³³ and ampicillin was added at 100µg/ml where appropriate.

5.2.2 Generation of ΔcpkA Mutant

A gene replacement construct for deletion of CpkA (MGG_06368) was generated by PCR amplification of an approximately 6.7 kb fragment encompassing the insertion site of the hygromycin resistance cassette as well as upstream and downstream flanking sequence from the original ΔcpkA mutant described previously¹² using the primers cpkA mut F (5’-
GTCGAACCTTCGCTTCTTTGA-3’) and cpkA mut R (5’-CAATCTCGCACTCCTGAACATG-3’). The resulting PCR product was then cloned into the pGEMT-EASY vector (Promega, Madison, WI) and transformed into E. coli DH5α to generate pCPKAKO.

* M. oryzae* protoplasts were generated from mycelium grown for 3 days in Complete Medium (10g Sucrose, 6g casamino acids, 6g yeast extract and 1ml 1000X Trace elements solution per liter), harvested by filtration via Miracloth (EMD Chemicals, Inc., San Diego, CA) and washed three times with sterile water. The resulting mycelium was treated with Lysing Enzymes (Sigma, St. Louis MO) in 1M sorbitol for 2 hours with gentle shaking after which protoplasts were harvested by filtration via Miracloth, followed by centrifugation and resuspension in STC (1.0 M sorbitol, 50 mM Tris [pH 8.0], 50 mM CaCl₂). A PCR product (3µg) derived from amplification of pCPKAKO with primers cpkA mut F and cpkA mut R was transformed into protoplasts as described previously. Candidate mutants were confirmed for the absence of the wild-type CpkA allele and the presence of hygromycin cassette by PCR and southern blot analysis (data not shown).

5.2.3 Sample Preparation

Conidia were harvested from 8 day old minimal medium plates by filtration through Miracloth to separate conidia from mycelium. Aliquots of two million conidia in 20ml of H₂O (germination condition) or 50mM cAMP (Sigma Aldrich, St. Louis, MO) (appressorium formation condition) were spotted onto the hydrophilic surface of 205 x 110mm GelBond® (Lonza, Rockland, ME) sheets and incubated at room temperature in the dark. Aliquots of
500,000 conidia in 5ml of H₂O or 50mM cAMP were used for absolute quantification studies. Samples were harvested by removal of the bulk liquid and subsequent scraping of the remaining liquid from the surface with a razor blade into a 50ml reagent reservoir. Samples were collected and centrifuged at 13,000 x g and 4°C for five minutes. Supernatants were removed from the pellets and discarded prior to freezing the pellets in liquid nitrogen followed by storage at -80°C.

Total protein was collected from each sample by bead beating for two minutes in a buffer consisting of 1X phosphate-buffered saline, 0.1% SDS and 2M urea using approximately 150mg of 0.5mm zirconia/silica beads (BioSpec Products, Inc., Bartlesville, OK). The samples were boiled for 5 minutes and centrifuged at 13,000 x g and 4°C for five minutes to remove beads and cellular debris. A BCA assay was used to determine protein concentration. Protein samples were trypsin digested using the FASP methodology exactly as described previously. A total of six peptide fractions were generated by StageTip anion fractionation with successive peptide elutions in Britton Robinson buffers pH 11, 8, 6, 5, 4, and 3 followed by a cleanup on C18 StageTips as previously described.

5.2.4 NanoLC MS/MS Analysis

Separation was performed on a nanoLC-1D+ system from Eksigent (Dublin, CA, USA) equipped with a continuous, vented column configuration, a 10 μL sample loop, 3 cm trap packed with Magic C18AQ packing material (Michrom BioReasources, Auburn, CA, USA) and a 15 cm analytical column packed with the same packing material. The flow rate was set to 350 nL/min during separation and a 3.5 h linear gradient from 5% to 35% B was applied.
A hybrid LTQ-Orbitrap XL MS (Thermo Fisher Scientific, Bremen, Germany) was used to perform MS/MS analysis. Optimal instrument parameters as reported in Gokce et al. were utilized.

In the second experiment comparing wt and ΔcpkA mutant conidia germinated in the presence of 50mM cAMP, a hybrid LTQ-FTICR MS (Thermo Fisher Scientific, Bremen, Germany) was used to perform the MS/MS analysis. LC and MS instrument parameters were identical to those used for the Orbitrap described above, except the source temperature was found to yield higher signal at 200 °C and a higher resolving power of 100,000 could be applied. The Lock mass function is not available for FTICR instruments.

5.2.5 Absolute Quantification

A nanoLC-2D system equipped with a cHiPLCnanoflex system (Eksigent, Dublin, CA), utilizing a 5 cm trap (200 μm × 0.5 mm packed with Chrom XP C18-CL 3 μm 120 Å), 15 cm column (75 μm × 15 cm packed with the same packing material) and a 20 μm i.d. SilicaTip ESI emitter (New Objective, Woburn, MA) was used to perform separation for SRM analysis. A short 30 min gradient was applied and data was acquired on a TSQ Vantage triple stage quadrupole mass spectrometer (Thermo Scientific, San Jose, CA). Instrument parameters were same as reported by Shuford et al.

Conidia, 18 hour cAMP treated and untreated peptides were analyzed on the triple quadrupole mass spectrometer and Skyline software v.1.1.0.2095 was utilized to determine abundant and unique peptide sequences for the 4 proteins chosen, assuming c_{peptide}≈c_{protein}.

Two fmol of IIAAPTIGR* (MGG_05580), 10 fmol of VASDLVQLLR* (MGG_09355),
5 fmol of MLTGDAIAIAK* (MGG_04994) and 5 fmol of CFDYIGVAK* (MGG_08526) synthetic peptides (New England Peptides, Gardner, MA) were spiked into samples prior to trypsin digestion for absolute quantification. Peak areas for each transition were exported from Skyline and transition ratios were further inspected in Excel.

5.2.6 Data Analysis

RAW files obtained by LC-MS/MS were searched against a concatenated target-reverse M. oryzae database (MG8_GeneCall10.fasta) from the Broad Institute using MASCOT Distiller version 2.3.01 (Matrix Science Inc., Boston, MA, USA). MASCOT parameters were the same as reported in Gokce et al.31. ProteoIQ version 2.3.05 (Bio-Inquire, Athens, GA, USA) was used to apply a 1% protein false discovery rate (FDR) for confident protein identifications. Normalization based upon total spectral counts was performed in ProteoIQ as described previously30. Differential protein expression was determined using a pairwise Student’s t-test (p-value = 0.05) on normalized spectral counts. Proteins lacking spectral count data for one condition in a pairwise comparison were assigned a total of one spectral count for all three biological replicates to facilitate calculation of fold changes. Proteins of low abundance with fewer than five total spectral counts in at least one condition of a given pairwise comparison were removed from the analysis. Fold change cutoffs for determination of differential expression were applied based upon total spectral count bin widths as determined previously30,31.
5.3 Results and Discussion

5.3.1 Identification and Label Free Quantification of Proteins During Germination and Appressorium Formation

To investigate proteome changes occurring during *M. oryzae* conidial germination and appressorium formation following treatment with cAMP, a label free quantitative proteomics study was undertaken. Conidia were germinated on the hydrophilic surface of GelBond® film in the presence or absence of 50mM cAMP. In the first experiment total protein was collected from triplicate

![Figure 5.1](image)

**Figure 5.1** A) Light micrographs of *M. oryzae* conidia germinated in 50mM cAMP on a hydrophilic surface showing the development of appressoria at the time points used in the proteomics study. Scale bar=10 µM. B) The number of germinated conidia (blue bars) and germinated conidia forming appressoria (red bars) were counted under the conditions described in panel A. Percentages are the average with standard error shown from three biological replicates with 100 conidia counted per replicate.
samples at 4, 8, 12 and 18hr as well as from conidia. In a second experiment wild-type *M. oryzae* and *ΔcpkA* mutant conidia were germinated in the presence of 50mM cAMP and total protein was collected from triplicate samples at 8 and 18 hours. Wild-type fungal morphology at each time point as well as the germination and appressorium formation rates are depicted in Figure 5.1.

![Venn diagram](image.png)

**Figure 5.2** A) Venn diagram displaying the number and overlap of proteins identified from conidia (2977 proteins), germinating conidia (3057 proteins) and appressoria (3071 proteins) samples in the wild-type time course study (experiment 1). Germination and appressoria proteins were identified in at least one time point (4, 8, 12, or 18hr). B) Venn diagram displaying the number and overlap of proteins from multiple independent analyses of the *M. oryzae* proteome including 3170 proteins identified in the wild-type time course study (experiment 1), 1651 proteins identified in the wild type vs. *ΔcpkA* mutant study (experiment 2), and 2912 proteins from the conidial proteome reported by Gokce *et al*..

At 4 hours approximately 80% of wild-type conidia have germinated and swelling of the germ tube tips is evident, however appressoria are not yet visible. By 8 hours, approximately 60% of germinated conidia have produced appressoria which start to melanize by 12 hours and are fully melanized by 18 hours. Germination of conidia from the *ΔcpkA* mutant is
slightly delayed at 4 hours but returns to wild-type levels by 8 hours and appressoria are first observed after 12 hours with fewer than 20% of germinated conidia producing appressoria by 18 hours (data not shown).

Collectively, between the two experiments a total of 841899 tandem mass spectra were collected and assigned to peptides. In the first study analyzing wild-type conidial germination and appressorium formation at multiple time points, a total of 3170 proteins were identified at a 1% protein false discovery rate (FDR). A total of 2977 proteins were identified from the conidial samples and when combined with our previous study\textsuperscript{31} brings the number of unique protein identifications to 3364 for the conidial proteome. In this study, an additional 193 proteins were identified during conidial germination or appressorium formation but not conidia (\textbf{Figure 5.2A}). In the second experiment comparing cAMP mediated appressorium development in the wild-type and \textit{ΔcpkA} mutant strains, 1651 proteins were identified including 30 proteins not identified in the first experiment bringing the total number of unique protein identifications to 3200. Collectively, the number of unique proteins reported here (3200 proteins) and by Gokce \textit{et al.}\textsuperscript{31} (2912 proteins) totals 3540 proteins (\textbf{Figure 5.2B}), representing 27.6% of the predicted \textit{M. oryzae} proteome.

In addition to expanding the coverage of the \textit{M. oryzae} proteome, we sought to identify proteins whose abundance changed during germination and appressorium formation. Changes in protein abundance were determined by spectral counting based label free quantification. Protein regulation during conidial germination was assessed by comparing the four germination time points directly to conidia and regulation in response to cAMP was determined by comparing cAMP treated and untreated samples at each respective time point.
Likewise, in the second experiment, the impact of the \( cpkA \) mutation was determined by direct comparison of wild-type and mutant cAMP treated samples at early (8 hours) and late (18 hours) time points. In a previous study encompassing independent analyses of a single sample, a false positive rate (FPR) for detection of biological change was established\(^3\). Proteins were divided into bin widths based upon the number of spectral counts for a given protein. A smaller fold change (1.5 fold) can be applied to proteins with larger numbers of spectral counts (average number of spectral counts greater than 10 spectral counts per replicate) while maintaining an FPR below 10\%\(^3\). However, proteins at the lower end of the spectral count distribution require a higher fold change (2.0 fold) to maintain a similar FPR\(^3\).

**Table 5.1** Number of Differentially Expressed Proteins for Each Comparison

<table>
<thead>
<tr>
<th>comparison</th>
<th>4hr</th>
<th>8hr</th>
<th>12hr</th>
<th>18hr</th>
<th>Totals (unique)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+cAMP/untreated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>37</td>
<td>53</td>
<td>51</td>
<td>49</td>
<td>166</td>
</tr>
<tr>
<td>Down</td>
<td>19</td>
<td>102</td>
<td>106</td>
<td>145</td>
<td>327</td>
</tr>
<tr>
<td>untreated/conidia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>66</td>
<td>75</td>
<td>63</td>
<td>101</td>
<td>208</td>
</tr>
<tr>
<td>Down</td>
<td>163</td>
<td>158</td>
<td>206</td>
<td>167</td>
<td>383</td>
</tr>
<tr>
<td>wt/(\Delta cpkA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>16</td>
<td></td>
<td>134</td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>Down</td>
<td>98</td>
<td>28</td>
<td></td>
<td></td>
<td>125</td>
</tr>
</tbody>
</table>

**Table 5.1** presents the number of differentially expressed proteins for each comparison based upon criteria described in the materials and methods. In the cAMP treated vs. untreated comparison, 166 proteins and 327 proteins were identified as up-regulated and down-regulated in at least one time point respectively. Likewise, during conidial germination, 208 up-regulated and 383 down-regulated proteins were identified. Furthermore, 148 and 125 proteins were up and down-regulated, respectively, in the wild-type and \( \Delta cpkA \) mutant comparison.
Table 5.2 Fold Changes of Proteins Up-Regulated in Response to cAMP Treatment and Up-Regulated in the Wild Type When Compared to the ΔcpkA Mutant.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Annotation</th>
<th>+cAMP vs. untreated*</th>
<th>wild type vs. ΔcpkA*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4hr  8hr  12hr  18hr</td>
<td>8hr  18hr</td>
</tr>
<tr>
<td>MGG_08526</td>
<td>hypothetical protein</td>
<td>1.1  31.3  106.0  296.2*</td>
<td>3.4  146.9*</td>
</tr>
<tr>
<td>MGG_10036</td>
<td>phenylalanine and histidine ammonia-lyase</td>
<td>1.1  10.8*  15.3  60.5</td>
<td>2.0  21.2*</td>
</tr>
<tr>
<td>MGG_13253</td>
<td>hypothetical protein</td>
<td>1.1  18.2  12.8  56.8*</td>
<td>2.0  6.1*</td>
</tr>
<tr>
<td>MGG_03436</td>
<td>hypothetical protein</td>
<td>1.1  42.2  75*   31.8*</td>
<td>2.0  35.6*</td>
</tr>
<tr>
<td>MGG_00659</td>
<td>glucan 1,3-beta-glucosidase</td>
<td>3.5  39.0*  77.7*  19.7</td>
<td>7.3  21.6*</td>
</tr>
<tr>
<td>MGG_05366</td>
<td>feruloyl esterase B</td>
<td>1.0  8.3  19.6  7.9*</td>
<td>1.9  9.5*</td>
</tr>
<tr>
<td>MGG_17528</td>
<td>hypothetical protein</td>
<td>2.9  4.6*  7.7   5.2*</td>
<td>5.6  9.2*</td>
</tr>
<tr>
<td>MGG_08975</td>
<td>3-hydroxy-3-methylglutaryl-coenzyme A reductase</td>
<td>1.7  4.1  4.1   4.2*</td>
<td>14.1 31.4*</td>
</tr>
<tr>
<td>MGG_07790</td>
<td>ligninase H2</td>
<td>1.1  7.4*  4.9   4.1</td>
<td>2.0  7.2*</td>
</tr>
<tr>
<td>MGG_13669</td>
<td>peptide transporter PTR2</td>
<td>0.4  1.3  19.1*  3.8</td>
<td>1.5  5*</td>
</tr>
<tr>
<td>MGG_00129</td>
<td>acetyl-hydrolase</td>
<td>1.6  0.7  1.4   3.8*</td>
<td>1.0  7.5*</td>
</tr>
<tr>
<td>MGG_09322</td>
<td>hypothetical protein</td>
<td>1.2  2.1  1.4   3.4*</td>
<td>1.4  3.8*</td>
</tr>
<tr>
<td>MGG_11475</td>
<td>hypothetical protein</td>
<td>1.5  1.5  1.7   2.9*</td>
<td>2.0  7.2*</td>
</tr>
<tr>
<td>MGG_15950</td>
<td>hypothetical protein</td>
<td>1.4  6.7  6.1*  2.7</td>
<td>1.7  25.7*</td>
</tr>
<tr>
<td>MGG_05433</td>
<td>solute carrier family 6 protein</td>
<td>1.8  1.5  3.9   2.6*</td>
<td>2.0  10.4*</td>
</tr>
<tr>
<td>MGG_00141</td>
<td>leptomycin B resistance protein pmd1</td>
<td>2.3* 3.3  5.0   2.5</td>
<td>1.7  10.3*</td>
</tr>
<tr>
<td>MGG_05989</td>
<td>seprase</td>
<td>1.9  2.0  3.0   2.5*</td>
<td>1.0  6.3*</td>
</tr>
<tr>
<td>MGG_00871</td>
<td>phosphatidylinositol transfer protein CSR1</td>
<td>0.7  2.3  2.1   2.4*</td>
<td>1.6  7.2*</td>
</tr>
<tr>
<td>MGG_01026</td>
<td>hydroxymethylglutary-CoA synthase</td>
<td>1.1  3.4*  2.4   2.3*</td>
<td>2.1  28.9*</td>
</tr>
<tr>
<td>MGG_07219</td>
<td>polyketide synthase ALB1</td>
<td>4.0* 3.9*  2.0   2.0</td>
<td>1.0  29.2*</td>
</tr>
<tr>
<td>MGG_00878</td>
<td>hypothetical protein</td>
<td>1.7  1.7*  1.3   2.0</td>
<td>1.8  3.7*</td>
</tr>
<tr>
<td>MGG_03554</td>
<td>60S ribosomal protein L36</td>
<td>1.1  1.9  2.2*  2.0*</td>
<td>2.0*  1.2</td>
</tr>
<tr>
<td>MGG_16203</td>
<td>SMc1A protein</td>
<td>3.6* 2.3  1.9   1.7</td>
<td>0.6  5.5*</td>
</tr>
<tr>
<td>MGG_01210</td>
<td>mitochondrial hypoxia responsive domain-</td>
<td>2.0  2.1*  1.0   1.6</td>
<td>2.3*  22.1*</td>
</tr>
<tr>
<td></td>
<td>containing protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGG_00819</td>
<td>niemann-Pick C1 protein</td>
<td>1.2  2.1  2.3*  1.6</td>
<td>2.4  11.7*</td>
</tr>
<tr>
<td>MGG_01557</td>
<td>phoshatidyglycerol/phosphatidyinositol transfer</td>
<td>1.1  1.3  1.1   1.6*</td>
<td>1.3  2.4*</td>
</tr>
<tr>
<td></td>
<td>protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGG_09272</td>
<td>beta-glucosidase 1</td>
<td>1.2  1.4  1.7*  1.5</td>
<td>1.9*  1.4</td>
</tr>
<tr>
<td>MGG_00755</td>
<td>indoleamine 2,3-dioxygenase</td>
<td>1.4  1.3  1.8*  1.4</td>
<td>1.2  2.4*</td>
</tr>
<tr>
<td>MGG_01674</td>
<td>multidrug resistance-associated protein 1</td>
<td>1.3  1.6*  1.4   1.4</td>
<td>0.5  5.3*</td>
</tr>
<tr>
<td>MGG_04378</td>
<td>integral membrane protein</td>
<td>2.5* 1.7  1.2   1.3</td>
<td>1.9  7.2*</td>
</tr>
<tr>
<td>MGG_13508</td>
<td>ubiquitin-protein ligase Sel1/Ubx2</td>
<td>0.9  3.6*  3.4   1.1</td>
<td>0.2  2.1*</td>
</tr>
<tr>
<td>MGG_13610</td>
<td>hydrolase</td>
<td>0.9  1.1  1.8*  1.0</td>
<td>1.0  2.5*</td>
</tr>
</tbody>
</table>

* Fold changes marked by an asterisk were determined to be significantly differentially expressed.
A total of 33 proteins were up-regulated (Table 5.2) and 26 down-regulated (Table 5.3) in the cAMP treated vs. untreated comparison and the wild-type vs. ΔcpkA mutant comparison. These comprise a subset of proteins regulated during appressorium formation that are both responsive to cAMP treatment and also require a functional CPKA for normal expression.

Table 5.3 Fold Changes of Proteins Down-Regulated in Response to cAMP Treatment and Down-Regulated in the Wild Type When Compared to the ΔcpkA Mutant.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Annotation</th>
<th>+cAMP vs. untreated</th>
<th>wild type vs. ΔcpkA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGG_00432</td>
<td>hypothetical protein</td>
<td>0.71 0.56 0.68 0.45*</td>
<td>0.46* 0.28</td>
</tr>
<tr>
<td>MGG_00522</td>
<td>hypothetical protein</td>
<td>0.57 0.40 0.28* 0.13*</td>
<td>0.20* 0.57</td>
</tr>
<tr>
<td>MGG_00719</td>
<td>aldehyde dehydrogenase 3H1</td>
<td>0.93 0.14* 0.30 2.68</td>
<td>0.37* 4.29</td>
</tr>
<tr>
<td>MGG_00968</td>
<td>mitochondrial outer membrane protein porin</td>
<td>1.04 0.95 0.62* 0.84</td>
<td>1.19 0.67*</td>
</tr>
<tr>
<td>MGG_01153</td>
<td>proteasome subunit alpha type-6</td>
<td>0.83 0.76 0.91 0.52*</td>
<td>0.65 0.41*</td>
</tr>
<tr>
<td>MGG_01622</td>
<td>uroporphyrinogen decarboxylase</td>
<td>0.82 0.61* 0.57 0.68</td>
<td>0.20* 0.65</td>
</tr>
<tr>
<td>MGG_02570</td>
<td>dienelactone hydrolase</td>
<td>0.52 0.38 0.29 0.14*</td>
<td>0.20* 1.08</td>
</tr>
<tr>
<td>MGG_02713</td>
<td>nascent polypeptide-associated complex subunit beta</td>
<td>1.07 0.65* 0.56 0.89</td>
<td>0.69 0.45*</td>
</tr>
<tr>
<td>MGG_02833</td>
<td>GTPase-activating protein GYP2</td>
<td>0.69 0.60 0.46* 0.76</td>
<td>0.27* 0.26</td>
</tr>
<tr>
<td>MGG_04311</td>
<td>3',5'-bisphosphate nucleotidase</td>
<td>1.19 0.46* 0.58 0.41</td>
<td>0.21* 0.54</td>
</tr>
<tr>
<td>MGG_04425</td>
<td>phenylalanyl-tRNA synthetase subunit alpha</td>
<td>0.66 0.56 0.62 0.59*</td>
<td>0.09* 1.09</td>
</tr>
<tr>
<td>MGG_04584</td>
<td>HNRNP arginine N-methyltransferase</td>
<td>1.06 0.60 0.33* 0.66</td>
<td>0.50* 1.16</td>
</tr>
<tr>
<td>MGG_05250</td>
<td>hypothetical protein</td>
<td>0.88 0.39 0.51 0.32*</td>
<td>0.38* 0.95</td>
</tr>
<tr>
<td>MGG_05956</td>
<td>glutamyl-tRNA synthetase</td>
<td>0.97 0.62* 0.94 0.85</td>
<td>0.61* 0.84</td>
</tr>
<tr>
<td>MGG_06751</td>
<td>hypothetical protein</td>
<td>0.84 0.56* 0.26* 0.30*</td>
<td>0.26* 0.24*</td>
</tr>
<tr>
<td>MGG_08622</td>
<td>nucleoside diphosphate kinase</td>
<td>0.70 0.77 0.64 0.61*</td>
<td>1.31 0.67*</td>
</tr>
<tr>
<td>MGG_09289</td>
<td>hexokinase</td>
<td>1.06 0.45 0.67 0.28*</td>
<td>2.07 0.13*</td>
</tr>
<tr>
<td>MGG_09355</td>
<td>NACHT domain-containing protein</td>
<td>0.46 0.25* 0.19* 0.12*</td>
<td>0.33* 0.30</td>
</tr>
<tr>
<td>MGG_09902</td>
<td>F-actin-capping protein subunit beta</td>
<td>0.63 0.48 0.45* 0.52</td>
<td>0.22* 0.51</td>
</tr>
<tr>
<td>MGG_10111</td>
<td>glucose and ribitol dehydrogenase</td>
<td>0.83 0.43* 0.40 0.13</td>
<td>0.36* 0.65</td>
</tr>
<tr>
<td>MGG_11916</td>
<td>CAP20</td>
<td>1.04 0.86 0.58* 0.50</td>
<td>1.40 0.56*</td>
</tr>
<tr>
<td>MGG_12822</td>
<td>glucose-6-phosphate isomerase</td>
<td>0.98 0.42* 0.36 0.43*</td>
<td>0.79 0.47*</td>
</tr>
<tr>
<td>MGG_13020</td>
<td>glucose-repressible alcohol dehydrogenase</td>
<td>0.91 0.59 0.28* 0.92</td>
<td>0.43* 2.91</td>
</tr>
<tr>
<td>MGG_13068</td>
<td>venom protein 2</td>
<td>0.72 0.52 0.27* 1.04</td>
<td>0.16* 1.62</td>
</tr>
<tr>
<td>MGG_15527</td>
<td>hypothetical protein</td>
<td>1.09 0.42* 0.54 0.71</td>
<td>0.26* 1.12</td>
</tr>
<tr>
<td>MGG_17278</td>
<td>argininosuccinate lyase</td>
<td>0.93 0.56* 0.56 0.67</td>
<td>0.22* 0.76</td>
</tr>
</tbody>
</table>

* Fold changes marked by an asterisk were determined to be significantly differentially expressed.
5.3.2 Gene Ontology Analysis of Regulated Proteins

The regulated proteins from each comparison were subjected to gene ontology (GO) analysis using the BLAST2GO program. Over and under-represented GO categories for each comparison were identified using Fisher’s exact test within the BLAST2GO program with the 3170 and 1651 proteins from experiments 1 and 2 as reference datasets. Two specific trends involving regulation of mitochondrial and transport proteins were identified based GO analysis. Proteins up-regulated during conidial germination were under-represented in the mitochondrial category (GO:0005739) and conversely, the down-regulated protein set was enriched for the mitochondrial category as well as the ion transport category (GO:0006811). Additionally, proteins associated with the transport activity category (GO:0005215) were over-represented in the 166 proteins up-regulated by cAMP treatment. Proteins from these GO categories will be addressed below in the context of germination and appressorium formation.

5.3.3 Extracellular Proteins Up-Regulated in Appressoria

Analysis of the subcellular localization of wild-type proteins up-regulated by cAMP treatment (Figure 5.3A and C) or when compared to the ΔcpkA mutant in the presence of cAMP (Figure 5.3B and D) identified a higher percentage of extracellular proteins relative to the whole set of proteins identified in each experiment. The 166 proteins up-regulated by cAMP treatment contain 39 proteins predicted to be extracellular by WolfPSORT. Likewise, another 32 predicted extracellular were more abundant in the wild-type when compared with the ΔcpkA mutant. Cumulatively, 60 unique putative extracellular proteins
were identified as up-regulated by the cAMP signaling pathway, either in response to cAMP treatment or by loss of regulation in the ΔcpkA mutant.

**Figure 5.3** Predicted subcellular localization of identified proteins in the wild-type time course study (3170 proteins, experiment 1) (A) or the wild type vs. ΔcpkA mutant study (1651 Proteins, experiment 2) (B). Predicted subcellular localization of proteins up-regulated in response to cAMP (166 Proteins) (C) and proteins up-regulated in the wild type when compared to the ΔcpkA mutant (148 Proteins) (D).
Several extracellular proteins likely involved in degradation of cell walls were induced during cAMP-mediated appressorium formation. A cutinase, CUT2 (MGG_09100), previously shown to up-regulated at the transcript level during appressorium formation and also demonstrated to be required for normal development of appressoria and complete virulence of *M. oryzae*[^41^], was strongly up-regulated by cAMP treatment at 18 hours. Interestingly, treatment of a Δcut2 mutant with cAMP or other inducers of the cAMP signaling pathway restored appressorium formation to wild-type levels indicating that CUT2 lies upstream of the cAMP signaling pathway[^41^]. A putative fungal lignin peroxidase, MGG_07790, was up-regulated 7.4 fold by cAMP treatment at 8 hours and up-regulated 7.2 fold at 18 hours when compared to the ΔcpkA mutant. Transcripts for MGG_07790 were up-regulated during appressorium formation in response to cAMP and on an inductive hydrophobic surface[^20^]. Two putative feruloyl esterases, MGG_05366 and MGG_09404, were induced during appressorium formation. MGG_05366 was up-regulated at 18 hours by cAMP treatment and was less abundant in the ΔcpkA mutant (Table 5.3). MGG_09404 was up-regulated at 8 hours following cAMP treatment. Both proteins contain a tannase and feruloyl esterase family domain (pfam07519) and show homology to the *Aspergillus niger* FAEB protein which is involved in the degradation of plant cell walls via release of ferulic acid or other aromatic compounds from cell wall polysaccharides[^42^]. The *A niger* FAEB protein shows a preference for plant pectins[^42^].

A multicopper oxidase, MGG_13764, with homology to fungal laccases and billirubin oxidases was 45.7 fold more abundant in the wild type than the ΔcpkA mutant. This protein was recently demonstrated to contain billirubin oxidase activity[^43^]. However, like billirubin
oxidases from other filamentous fungi, its biological substrate and function remains unclear. Related laccases from filamentous fungi are known to catalyze oxidation of phenolic compounds and play roles in developmental processes as well as degradation of plant lignins\textsuperscript{44,45}.

Finally, a putative acid phosphatase, MGG_00552, was more abundant in the wild type than the ΔcpkA mutant at 18 hours. No regulation of the protein was observed in the wild type during appressorium formation or germination. In an analysis of candidate secreted \textit{M. oryzae} effector proteins, transcripts of MGG_00552 were identified from cAMP treated germinating conidia as well as \textit{M. oryzae}-infected rice leaves\textsuperscript{46}. Pathogenicity of a MGG_00552 gene disruption mutant on barley was indistinguishable from the wild-type parent and no further characterization was reported\textsuperscript{46}. MGG_00552 is homologous to \textit{Aspergillus fumigatus} PhoAp (57%ID on 82% coverage), a cell wall associated GPI anchored phosphatase that is active on both phosphomonoesters and phosphodiesters and is inducible by low phosphate concentrations\textsuperscript{47}. Fungal acid phosphatases are involved in the acquisition of phosphorous via its release from organic compounds under phosphorous limiting conditions.

5.3.4 Cell Wall Biosynthesis and Remodeling

The process of conidial germination in \textit{M. oryzae} is characterized by the emergence of a germ tube which elongates in a polarized fashion and develops into an appressorium in presence of an appropriate stimulus. \textit{De novo} cell wall biosynthesis and remodeling of existing cell wall structures occurs in the extracellular space on the external surface of the
plasma membrane and plays a critical role in both the germination and appressorium formation processes. Cell walls of filamentous fungi are composed primarily of β- and α-glucans, chitin, and glycoproteins (mannan or galactomannan derivatives)\textsuperscript{48-50}. Fungal chitin is composed of repeating β-1,4-linked N-acetylglucosamine units which form chains that assemble into microfibrils which are then covalently attached to cell wall β-1,3-Glucans. Chitin is synthesized at the fungal plasma membrane in regions of active cell wall growth or remodeling and is extruded into the extracellular space where microfibril formation and crosslinking to cell wall glucans occurs. Fungal chitin synthases are grouped into seven distinct classes with three specific to filamentous fungi\textsuperscript{51} and are frequently found in multiple copies in a single organism\textsuperscript{48}. The \textit{M. oryzae} genome encodes seven chitin synthase proteins with a single representative of each of the seven classes of chitin synthases\textsuperscript{52}. Six of the seven \textit{M. oryzae} chitin synthase proteins were identified in conidia and all time points during germination. The most abundant in conidia, CHS1 (MGG_01802), was previously shown to play a role in pathogenesis. A \textit{chs1} mutant was reduced in conidial germination, appressorium formation and virulence\textsuperscript{52}. Two chitin synthase proteins, CHS5 (MGG_13014) and CHS7 (MGG_06064), were significantly up-regulated at the four hour germination time point. A \textit{chs7} mutant was demonstrated to be impaired in appressorium development and virulence\textsuperscript{52,53}. Interestingly, CHS7 and CHS5 were down-regulated by cAMP treatment at 12 and 18hrs respectively. Finally, CHS6 (MGG_13013), the only chitin synthase absolutely required for plant infection\textsuperscript{52}, was identified but not regulated. In addition to chitin synthases, a number of proteins involved in chitin metabolism were regulated during germination and appressorium development. Two of the most strongly up-
regulated proteins during conidial germination are a putative glucosamine-6-phosphate isomerase (MGG_00625) and a putative N-acetylglucosamine-6-phosphate deacetylase (MGG_00620). These two proteins are predicted to be involved in the catabolism of N-acetylglucosamine, catalyzing the conversion of N-acetylglucosamine-6-phosphate to fructose-6-phosphate. Homologous proteins in *C. albicans*, NAG1 and NAG2, are required for utilization of N-acetylglucosamine as a carbon source suggesting *M. oryzae* may be consuming intermediates of chitin biosynthesis as a source of carbon during conidial germination. A putative β-N-acetyl-D-glucosaminidase, MGG_09922, containing a glycosyl hydrolase family 20 domain was up-regulated during both germination and appressorium formation. β-N-acetyl-D-glucosaminidases catalyze the removal of N-acetylglucosamine residues from oligosaccharides. MGG_0992 is homologous to the EXC1Y protein of *Trichoderma asperellum*, a glucosamine inducible secreted protein that is up-regulated during the mycoparasitic interaction of *T. asperellum* with a host fungus.

Furthermore, a putative chitin deacetylase, (MGG_09159), catalyzing the conversion of chitin to chitosan was strongly up-regulated in response to cAMP treatment but was not regulated during germination. Chitosan is detectable on the surface of germ tubes, appressoria and infectious hyphae. Chitin is detectable on the surface of germ tubes and appressoria but not infectious hyphae colonizing the first invaded rice cell. During later stages of infection chitin again becomes detectable on the surface of infectious hyphae. Chitosan has been documented on the surface of infection structures in planta for a number of plant pathogenic fungi and it has been proposed that conversion of chitin to chitosan prevents degradation of the fungal cell wall by plant-derived chitinases. Finally, a
hypothetical protein (MGG_06771) with a carbohydrate-binding module belonging to the CBM18 family\textsuperscript{59} which is characterized by chitin binding activity was up-regulated at the 12 and 18 hour germination time points. Expression of MGG_06771 transcripts was documented in infected rice leaves but not mycelium cultured \textit{in vitro} however its role in the infection process is currently unknown\textsuperscript{57}.

β-1,3-Glucans, the most abundant component of fungal cell walls, are synthesized in a vectorial manner similar to chitin\textsuperscript{49}. The synthesis of β-1,3-Glucan chains is carried out by the glucan synthase complex located at the plasma membrane\textsuperscript{49}. In \textit{S. cerevisiae}, two of three \textit{fks} genes encoding β-1,3-Glucan catalytic subunits are involved in synthesis of β-1,3-Glucan from UDP-Glucose. The \textit{M. oryzae} genome encodes a single \textit{fks1} homolog, MGG_00865, which was detected in all biological conditions but was not differentially expressed. β-1,3-Glucan chains extruded into the periplasmic space by the glucan synthase complex require elongation and covalent linkage to other cell wall components during incorporation into the cell wall. Elongation of β-1,3-Glucans is catalyzed by β-1,3-glucanosyltransferases which cleave the β-1,3 linkage of an existing β-1,3-Glucan chain and transfer the new reducing end to the nonreducing end of another β-1,3-Glucan chain producing a new β-1,3 linkage. This enzymatic activity belongs to the GH72 family of glycoside hydrolases\textsuperscript{59}. The first member of the family was described in \textit{A. fumigatus} whose genome encodes seven members of the GH72 family\textsuperscript{60}. Four members of the GH72 family are present in \textit{M. oryzae} (MGG_03208, MGG_06722, MGG_07331, and MGG_08370) with a fifth protein, MGG_11861, not assigned to this family but having homology to the \textit{A. fumigatus GEL5} protein. One of the \textit{M. oryzae} β-1,3-glucanosyltransferases, MGG_06722, was up-regulated in the wild type when
compared to the ΔcpkA mutant and a second protein, MGG_11861, was up-regulated at 12 hours following cAMP treatment. Furthermore, a third β-1,3-glucanosyltransferase, MGG_08370, was up-regulated at 8, 12, and 18 hours during germination. Although no information exist regarding the catalytic nature of these proteins, their regulation patterns suggest specific roles during conidial germination and appressorium development.

A putative β-1,3-exoglucanase, MGG_00659, is one of the most strongly up-regulated proteins in response to cAMP treatment at 8 and 12 hours and is also down-regulated in the ΔcpkA mutant at 18 hours. The MGG_00659 transcript was induced by formation of appressoria on hydrophobic surfaces as well as by cAMP treatment\textsuperscript{20}. β-1,3-exoglucanases catalyze the release of single glucose residues from the nonreducing end of β-1,3-glucan chains. A PSORT prediction suggests the MGG_00659 protein is extracellular and the CAZY database describes it is as a member of the glycosyl hydrolase 55 protein family\textsuperscript{59}. The protein is highly homologous to Lam1.3 (59% identity) of \textit{Trichoderma harzianum}\textsuperscript{61,62} and EXG1 (46% identity) of \textit{Cochliobolus carbonum}\textsuperscript{63}, both of whose β-1,3-exoglucanase activity has been experimentally characterized. A \textit{M. oryzae} gene deletion mutant of MGG_00659 retained pathogenicity and the ability to form appressoria\textsuperscript{20}. A second putative β-1,3-exoglucanase, MGG_14087, with 50% identity to LAM1.3 and was significantly up-regulated at 4 and 18 hours during germination. In addition, a putative endo β-1,3-glucanase with homology to ENGL1 of \textit{A. fumigatus} was also up-regulated at 18 hours in the wild type when compared to the ΔcpkA mutant.

A host of additional proteins likely involved in cell wall modification were up-regulated by cAMP treatment or down-regulated in the ΔcpkA mutant. Included were a GH3 family
putative β-glucosidase (MGG_10038), up-regulated at 18 hours following cAMP treatment and previously demonstrated to be induced during appressorium formation\textsuperscript{20}. A second putative β-glucosidase, (MGG_01885) belonging to the GH3 family, was down-regulated in the \textit{ΔcpkA} mutant at 18 hours. A GH92 family member encoding a putative α-1,2-mannosidase (MGG_07146) was up-regulated at 12 hours in the cAMP treated sample. Two GH16 family proteins, MGG_00592 and MGG_01134, encoding putative glucosyltransferases were down-regulated in the \textit{ΔcpkA} mutant at 18 hours. Interestingly, MGG_00592, and to a lesser extent, MGG_01134, are homologous to the \textit{A. fumigatus} CRF1 protein which belongs to the Crh family of glycosylphosphatidylinositol (GPI)-anchored glucanosyltransferases\textsuperscript{64}. AfCRF1 is an orthologue of the \textit{S. cerevisiae} GPI-anchored CRH1 protein which is involved in the crosslinking of β-1,6-glucans to chitin\textsuperscript{65}. In addition to the GH16 domain, MGG_01134 also contains a CBM18 domain typically involved in the binding of chitin.

A critical role for melanin in the formation of functional appressoria has long been recognized\textsuperscript{4,66-68}. The hyphal and appressorial cell walls of \textit{M. oryzae} are subject to melanization. Three genes essential for biosynthesis of melanin are required for pigmentation of mycelium and development of functional appressoria. Generation of mutations in the hydroxynapthalene reductase \textit{BUF1} (MGG_02252), the polyketide synthase \textit{ALB1} (MGG_07219), or the scytalone dehydratase \textit{RSY1} (MGG_05059) all produce nonpathogenic strains with abnormal culture pigmentation\textsuperscript{20,66}. A fourth protein, 4HNR (MGG_07216), with homology to \textit{BUF1} catalyzes the reduction of tetrahydroxynaphthalene to scytalone but is not required for melanin production most likely as a result of a functional redundancy derived
from BUF1. All four of the aforementioned proteins involved in melanin biosynthesis were identified in a transcriptomics study as up-regulated during early appressorium formation on a hydrophobic surface and in response to cAMP treatment. We previously reported that BUF1, 4HN R and RSY1 are amongst the most abundant proteins in M. oryzae conidia. All three proteins remained highly abundant during germination and appressorium formation with no significant regulation. However, the polyketide synthase ALB1, that catalyzes the first step in melanin biosynthesis from acetyl-CoA and whose abundance is much lower in conidia was down-regulated during germination at 4 and 8 hours, but up-regulated by cAMP treatment at 4 and 8 hours. In addition, ALB1 was also strongly down-regulated in the cpkA mutant at 18 hours.

The emergence and elongation of conidial germ tubes followed by the development of an appressorium is a dynamic process that necessitates de novo cell wall biogenesis and substantial remodeling of the newly formed structures. The importance of this process is reflected in the both the abundance and regulation of a large number of proteins involved in cell wall metabolism. Included are proteins central to chitin, β-1,3-Glucan, and melanin metabolism as well proteins containing domains associated with carbohydrate modification whose specific catalytic activities remain unclear.

5.3.5 Regulation of Transport Functions in Response to Conidial Germination and Appressorium Development

Gene ontology analysis revealed an enrichment of proteins with transport functions in the set of proteins up-regulated by cAMP as well as those down-regulated during germination. This
suggests that regulation of transport functions is an important aspect of germination and the cAMP response leading to appressorium development. The transporter activity GO category (GO:0005215) was enriched in the set of 166 proteins up-regulated by cAMP with 21 of the regulated proteins assigned to this category. An additional 12 of the 166 cAMP up-regulated proteins are assigned to at least one of the following transport related GO categories; transport (GO:0006810), ion transport (GO:0006811), or protein transport (GO:0015031).

In addition, the regulation of 9 of these 33 proteins was dependent upon the presence of CPKA. Finally, the process of conidial germination resulted in the down-regulation of 77 proteins associated with at least one of the aforementioned GO categories and also an enrichment of the ion transport GO category (GO:0006811).

The Golgi-localized P-type ATPase, APT2 (MGG_02767), which is involved in exocytosis of an as yet undefined collection of extracellular proteins is amongst the transport proteins that respond most strongly to cAMP treatment. APT2 is required for normal pathogenicity and elicitation of the hypersensitive response during incompatible host-pathogen interaction. A putative plasma membrane localized H+-ATPase, PMA2 (MGG_04994), was also induced by cAMP treatment at 18 hours. The M. oryzae genome encodes a second H+-ATPase, PMA1 (MGG_07200), which is one of the most abundant proteins of the spore proteome and is not regulated during germination or appressorium formation. PMA1 is a homolog of yeast PMA1p and likely represents the housekeeping H+-ATPase responsible for generation of proton gradients across the plasma membrane. On the other hand, PMA2 is much less abundant and strongly up-regulated during appressorium formation. In addition, expression of PMA2 (referred to as PMA1 by Zhang et al. and PMA2 by Gilbert et al.) is
higher when using ammonia rather than nitrate as a nitrogen source and its regulation is partially dependent upon the activity of the glucokinase GLK1. However, its role in appressorium formation remains unclear.

The *M. oryzae* genome encodes 4 TRK family transporters which are involved in the uptake of K\(^+\) and Na\(^+\) in other organisms. One member of the family, MGG_09119, was up-regulated during appressorium formation. The other three members of the TRK family as well as the sole member of the HAK family of K\(^+\) and Na\(^+\) transporters in *M. oryzae* were not detected in this study. Two putative K\(^+\)-Na\(^+\) efflux P-type ATPases, MGG_02074 and MGG_10730, were up-regulated and a third, MGG_05078 was down-regulated by cAMP treatment. Interestingly all three were down-regulated by germination. RNA silencing of these K\(^+\)-Na\(^+\) efflux P-type ATPases resulted in strains with growth, conidiation and pathogenic defects including deficiencies in establishment of appressoria.

Two putative Ca\(^{2+}\)-transporting ATPases, MGG_02487 and MGG_07971, and a putative Ca\(^{2+}\)/H\(^+\) antiporter, MGG_01381, were down-regulated by germination. MGG_02487 and MGG_07971 are homologs of the yeast vacuolar calcium pump, PMC1p, which plays a role in cellular calcium homeostasis by transporting Ca\(^{2+}\) into the vacuole. RNA silencing mediated knock down of MGG_02487 and MGG_07971 results in strains that have severe growth defects and reduced conidiation. A third PMC1 like protein in *M. oryzae*, MGG_04890, was not detected. The putative Ca\(^{2+}\)/H\(^+\) antiporter, MGG_01381, is a homolog of the yeast vacuolar Ca\(^{2+}\)/H\(^+\) antiporter, VNX1p, which is also involved in cellular calcium homeostasis. The conidiation defects associated with these Ca\(^{2+}\) transporters and their down-
regulation during germination points towards an importance for calcium homeostasis during germination.

Amongst the most strongly down-regulated proteins during germination are two putative MFS family transporters (pfam domain PF07690), MGG_15745, and MGG_10131 as well as a putative sugar transporter (pfam domain PF00083), MGG_00045. Localization to the plasma membrane is predicted for MGG_15745; however, it is a homolog of *S. cerevisiae* YMR221C. YMR221C was identified in the yeast mitochondrial proteome and was demonstrated to interact with *S. cerevisiae* Atg27, a protein localized to multiple organelles, including mitochondria, and proposed to direct donor membranes to the pre-autophagosomal structure. A putative alpha glucoside permase, MGG_00045, was down-regulated during germination at 12 and 18 hours. MGG_00045 shows 32% identity to *S. cerevisiae* maltose transporters, MPH2 and MPH3. In addition, two ABC type transporter proteins, MGG_01674 and MGG_03736, were down-regulated at all four germination time points. MGG_01674 is homologous (49% identity) to YCF1, a *S. cerevisiae* involved in transport of glutathione-metal conjugates into the vacuole. MGG_03736 is homologous to the *S. cerevisiae* YBT1 protein isolated from vacuoles and demonstrated to be an ATP dependent bile acid transporter.

Previous studies examining gene expression during the formation of appressoria revealed the regulation of a large number of transport related genes. A similar trend was observed in the proteomics data presented here further supporting an important role for transport processes in environmental adaptation of the germinating conidium and infection related development. Conidia germinate in the absence of exogenous nutrients and rely solely on
storage reserves to support the earliest stages of germination. Interestingly, a large number of transport related proteins were down-regulated during the germination process supporting the concept of consumption of reserves stored in the conidia for germination. However, upon initiation of appressorium formation, a number of proteins involved in ion transport were up-regulated, several of which are known to be important for appressorium function and pathogenicity.

5.3.6 Regulation of Ribosomal Proteins During Conidial Germination

The 3200 proteins identified in this study include 114 ribosomal proteins of which 12 proteins are up-regulated and 16 proteins are down-regulated during germination. The cytoplasmic and mitochondrial ribosomes of *S. cerevisiae* have been extensively characterized and the closest yeast orthologs for each of the *M. oryzae* regulated ribosomal proteins were identified. Examination of the down-regulated ribosomal proteins revealed that 14 of the 16 proteins are homologs of yeast mitochondrial proteins. This observation is consistent with an enrichment of the mitochondrial GO category in the proteins down-regulated during germination. Conversely, all 12 of the ribosomal proteins up-regulated during germination are homologs of yeast cytoplasmic ribosomal proteins. In addition, the process of germination resulted in an increase in abundance of six proteins predicted to be associated with the nuclear pre-ribosome complex. MGG_02505, a C$_2$H$_2$ type zinc finger domain containing protein, is significantly up-regulated at three of the four germination time points and shares homology to the yeast cytoplasmic proteins REH1p and REI1p which are involved in stabilization of the cytoplasmic 60s ribosomal subunit and also

131
nuclear import of the 60s ribosomal subunit export factor, ARX. Four proteins, MGG_01561, MGG_00482, MGG_06242 and MGG_07258 were up-regulated during germination at 18 hours. MGG_01561, a putative nucleolar GTP-binding protein, is a homolog of yeast NOG2p which is a putative GTPase required for processing of the precursor 27S rRNA as well as export of the pre 60s ribosome from the nucleus. MGG_00482 is a homolog of the yeast NIP7 protein which is required for both pre-rRNA processing and biogenesis of the 60S ribosomal subunit. MGG_06242 contains a KRI1-like family C-terminal domain and shows weak homology to yeast KRI1, a protein required for biogenesis of the 40S ribosomal subunit. MGG_07258 is similar to the yeast MPP10 protein which is required for processing of pre-18s rRNAs in yeast. Finally, MGG_03080 contains two WD40 like domains and shows weak homology to the yeast SQT1 protein which is required for normal assembly of the 60S ribosomal subunit. The up-regulation of numerous cytoplasmic ribosomal proteins and putative ribosomal biogenesis factors indicates that conidial germination stimulates production of ribosomes.

5.3.7 Down-Regulation of Mitochondrial Proteins During Germination

As mentioned previously, the set of 383 proteins down-regulated during conidial germination were enriched in the mitochondrial cellular component GO category (GO:0005739). A total of 45 proteins from the 383 down-regulated proteins were assigned to this category. The set of down-regulated proteins also includes the aforementioned 14 putative mitochondrial ribosomal proteins. Furthermore, down-regulated during germination were an abundance of electron transport components. An examination of the oxidative phosphorylation pathways in
filamentous fungi identified 76 *M. oryzae* genes likely to encode components of the oxidative phosphorylation pathway. A total of 57 of these proteins were identified in the proteomics data, 17 were also down-regulated in at least one germination time point and none were up-regulated. Included in the down-regulated proteins were 11 putative NADH dehydrogenase subunits of electron transport complex I and two putative alternative NADH dehydrogenases. An additional 9 down-regulated proteins are probable components of the mitochondrial import complexes of the inner (TIM23 complex) and outer (TOM complex) mitochondrial membranes. These complexes are involved in the mitochondrial import of nuclear-encoded preproteins synthesized in the nucleus. Also down-regulated was a putative mitochondrial presequence protease, MGG_02440, whose yeast homolog CYM1 is involved in the degradation of mitochondrial presequence peptides after their removal from imported preproteins.

The extensive down-regulation of mitochondrial proteins during conidial germination indicates that mitochondrial proteins are generally more abundant in conidia prior to germination and suggests that mitochondria in fact may be more abundant in conidia as well. Conidia rely on mobilization of storage reserves to provide energy and precursor metabolites for biosynthetic processes. Rapid utilization of these reserves is required for timely germination when environmental conditions are favorable and an increased abundance of mitochondria in conidia could reflect need for a large pulse of energy generation during the initial stages of germination prior to germ tube emergence.
5.3.8 Protease Regulation

Previous reports have established a link between protein turnover and the processes of conidial germination and appressorium development\textsuperscript{20,93}. Two putative thiol proteases, MGG\textunderscore 08526 and MGG\textunderscore 14872, are both strongly up-regulated at 18 hours in the presence of cAMP, with MGG\textunderscore 08526 being the most strongly up-regulated protein in the dataset (Table 5.2). MGG\textunderscore 08526 is also regulated in a CPKA dependent fashion (Table 5.2). Transcripts for both proteins were previously reported to be up-regulated during appressorium formation\textsuperscript{20}. MGG\textunderscore 14872 contains a cysteine proteinase domain with similarity to domain II of the calpain family of calcium-dependent cysteine proteinases but lacks other protein domains typically associated with calpains. A total of four \textit{M. oryzae} proteins contain this cysteine proteinase domain and all are predicted to be localized to the nucleus. The three remaining cysteine proteinase domain containing proteins (MGG\textunderscore 06335, MGG\textunderscore 07573 and MGG\textunderscore 15810) were not detected in this study. MGG\textunderscore 08526 is also predicted to be nuclear in localization and contains a partial cysteine proteinase domain lacking the N-terminal portion of this domain harboring a conserved cysteine residue required for catalytic activity suggesting it may not have proteolytic activity. Analysis of a MGG\textunderscore 08526 gene deletion mutant revealed no gross defects in appressorium formation or pathogenicity\textsuperscript{20}. Calpain like cysteine proteinases have been identified in other fungi and include the \textit{Aspergillus nidulans} PalB and \textit{S. cerevisiae} YMR154c, both of which are involved in the adaptation to alkaline conditions and the latter also being required for efficient sporulation\textsuperscript{94-96}. In mammalian systems, nuclear calpains are known to mediate cleavage of calcium-dependent protein kinases and thereby influence Ca\textsuperscript{2+} signaling.
pathways. Based upon the roles similar signaling proteases in other organisms and their regulation during appressorium formation it suggests that these putative proteases have a role in signal transduction during the early stages of development.

5.3.9 Regulation of Secondary Metabolism

During the process of appressorium formation a number of enzymes central to the biosynthesis of secondary metabolites were induced including enzymes representing key entry points to secondary metabolism. A putative phenylalanine ammonia lyase (PAL), MGG_10036, was up-regulated by cAMP treatment in a CPKA dependent manner (Table 5.2). MGG_10036 transcripts were previously reported to be up-regulated during appressorium formation. PAL catalyzes the production of cinnamic acid and ammonia from phenylalanine and serves as the entry point to the phenylpropanoid pathway for biosynthesis of secondary metabolites. Although MGG_10036 is up-regulated at both the transcript and protein level during appressorium formation, deletion of this gene had no affect on appressorium development or pathogenicity.

The mevalonate pathway converts acetyl-CoA to isopentenyl diphosphate from which sterol and nonsterol isoprenoid biosynthesis is initiated. HMG-CoA synthase (MGG_01026) and HMG-CoA reductase (MGG_08975) are responsible for the production of mevalonate and represent the rate limiting steps of the pathway. Both proteins are up-regulated at 18 hours during appressorium formation. In addition, up-regulation of transcripts of both genes was observed in appressoria. Interestingly, the induction of both proteins by cAMP treatment is CPKA dependent as both proteins were more abundant in wild-type appressoria than the
**ΔcpkA** mutant. The three proteins downstream of HMG-CoA reductase responsible for the conversion of mevalonate to isopentenyl diphosphate including, MGG_016219 (a putative mevalonate kinase), MGG_05812 (a putative phosphomevalonate kinase) and MGG_09750 (a putative diphosphomevalonate decarboxylase) were all identified but not significantly regulated.

5.3.10 Validation of the Global Proteomics Data via Absolute Quantification

To further validate the quantification and differential expression of proteins within the global proteomics datasets an absolute quantification of a select group of regulated proteins was generated by selected reaction monitoring (SRM)\(^98\). The selected proteins included two proteins up-regulated by cAMP treatment (MGG_08526 and MGG_04994), a protein down-regulated by cAMP treatment (MGG_09355) and a protein up-regulated during germination (MGG_05580). Synthetic peptides containing a heavy isotope labeled arginine or lysine and corresponding to tryptic peptides of the target protein were designed as described in the material and methods. Triplicate protein samples were produced from conidia and conidia germinated in the presence or absence of 50mM cAMP at 4, 8, 12, 18 and 24 hours for both the wild type and **ΔcpkA** mutant. The synthetic peptides were spiked into protein samples at known concentrations prior to trypsin digestion and the resulting peptides analyzed on a TSQ Vantage triple stage quadrupole mass spectrometer. Identification of both the heavy isotope labeled synthetic peptide and nonlabeled target peptide by SRM facilitated the quantification of the target peptide, thereby giving an accurate measurement of the abundance of each target protein in a given sample.
The absolute quantification of four target proteins is presented in **Figure 5.4**. In each instance the regulation of the target protein as determined by absolute quantification is in agreement with the global proteomics data. A putative CAIB/BAIF family CoA transferase, MGG_05580, was up-regulated during germination (**Figure 5.4A**). However, the abundance of this protein is reduced by cAMP treatment. Interestingly, the abundance of this protein is greater in the ΔcpkA mutant than the wild type and the reduction in protein levels by cAMP treatment is also observed in the ΔcpkA mutant suggesting a suppression of MGG_05580 protein levels by cAMP signaling. A protein of unknown function, MGG_09355, containing a NACHT domain typically found in components of signaling pathways with scaffold functions was strongly repressed by cAMP treatment (**Figure 5.4C**). This protein is abundant in conidia and remains abundant during germination in the absence of cAMP. Similar to MGG_05580, the MGG_09355 protein is more abundant in the ΔcpkA mutant but still responds negatively to cAMP treatment. Conversely, PMA2, a putative plasma membrane localized H⁺-ATPase and the putative thiol protease, MGG_08526, are up-regulated by cAMP treatment in the global proteomics data and the SRM data (**Figures 5.4B and D**). PMA2 increases in abundance during germination but responds strongly to cAMP treatment in a CPKA dependent manner. MGG_08526 is the protein most strongly up-regulated by cAMP treatment and was undetected in the SRM analysis of germinated samples not treated with cAMP. The agreement of protein regulation observed in the global proteomics data with that observed in the SRM data indicates that the spectral counting based label free quantification utilized in this study provides an accurate means of identifying regulated proteins. Finally, the SRM data includes a 24 hour time point not included in the global
proteomics studies. It is apparent from this data that changes in protein abundance in response to cAMP treatment were stronger or at least equivalent for the analyzed proteins at 24 hours as compared to 18 hours suggesting that further analysis of mature appressoria at later time points is warranted.
Figure 5.4 Absolute quantification by SRM (A, C, E, G) and average spectral count data (B, D, F, H) for four regulated proteins including MGG_05580 (A, B), MGG_09355 (C,D), MGG_04994 (E, F) and MGG_08526 (G, H). Protein abundance in SRM data is reported as fmol of target protein per 10 µg of total protein with averages from three biological replicates and standard errors shown (A, C, E, G). Average normalized spectral counts from experiment 1 are plotted with standard errors shown (B, D, F, H). Wild-type (blue and red bars or lines) and ΔcpkA mutant (green and purple bars) conidia were germinated on a hydrophilic surface in the presence (red and purple bars, red lines) or absence (blue and green bars, blue lines) of 50mM cAMP and three biological replicates were harvested 0, 4, 8, 12, 18 and 24 hours.
5.3.11 Comparison of the cAMP Responsive Transcriptome and Proteome

Changes in protein abundance observed in proteomic studies often fail to correlate well with changes in gene expression observed in transcriptomic studies performed on similar biological samples\textsuperscript{99}. A DNA microarray based analysis of gene expression revealed the up- and down-regulation of 644 genes and 370 genes, respectively, in a comparison of cAMP treated and untreated \textit{M. oryzae} conidia germinated for 9 hours on a hydrophilic surface\textsuperscript{20}. No correlation between changes in mRNA and protein abundance is observed when fold changes of the 9 hour transcriptomes are compared to fold changes of the 8 hour proteomes when only those genes for which both transcript and protein information is available. A similar lack of correlation was observed when proteomes from the other time points were compared to the 9 hour transcriptomes (data not shown). However, 53 genes (29 up- and 24 down-regulated) demonstrating a change in protein abundance at 8, 12, or 18 hours were also significantly regulated in the same direction at the transcript level at the 9 hour time point.

The lack of correlation of between differentially expressed transcripts and proteins is consistent with observations made in other biological systems. This is lack of correlation is typically attributed, in part, to the independent analysis of different experimental samples at separate times and often in different laboratories. On the other hand, increasing evidence suggests that translational regulation may be the most important factor controlling cellular protein levels and therefore some discordance between protein and mRNA expression is expected\textsuperscript{99}. Finally, although changes in gene expression do not accurately reflect changes in protein abundance in the cAMP-induced appressorium formation data sets, the set of 53
proteins whose differential expression is in agreement at both the transcript and protein level likely represent proteins central to the development and function of appressoria.

5.4 Conclusions
Technological advances in the field of mass spectrometry based proteomics now facilitate proteome analysis at depths unobtainable just a few years ago. In the present study, we identified 3200 unique proteins present in conidia, germinating conidia, or cAMP-induced appressoria. Utilizing spectral counting-based label free quantification we were able to observe changes in protein abundance during both conidial germination and appressorium development. A number of biological processes including cell wall biosynthesis and remodeling, transport, and production of extracellular proteins were dynamically regulated and provide insight into the process of appressorium formation. Validation of the global proteomics data via the absolute quantification of four proteins indicates that regulation observed in the global proteomics data can be reproduced by SRM analysis which provides very accurate measurements of protein abundance. In addition, a comparison of the cAMP-responsive transcriptome and proteome indicates that changes in transcript abundance generally do not correlate well with changes in protein abundance with the exception of some of the most strongly regulated proteins. The data presented here provides the most extensive protein level analysis of appressorium formation performed to date and provides key insights into the development and physiology of the early stages of infection related development as well as a foundation for future research on post translation modification of the *M. oryzae* proteome.
References


CHAPTER 6
Polyubiquitin is Required for Growth, Development and Pathogenicity in the Rice Blast Fungus *Magnaporthe oryzae*

The following work was reprinted from: Oh Y., Franck W. L., Han S., Shows A., Gokce E., Muddiman D. C., Dean R. A., *PLoS ONE* 2012, 7 (8), e42868. No permission was required as the article has an open-access license. This work was part of a collaborative effort and only a portion constitutes the author’s personal dissertation research; however, all results are shown to allow for a complete interpretation of the findings.

6.1 Introduction

Ubiquitin mediated protein degradation is a highly conserved process and plays important roles in a variety of cellular processes, including transcriptional regulation, signal transduction, cell cycling, cellular differentiation and pathogenesis. Ubiquitin, a highly conserved 76-amino acid protein, is activated by the ubiquitin activating enzyme E1 using ATP, which is then transferred to a ubiquitin conjugating enzyme E2. The E2 enzyme and the protein substrate bind specifically to a particular ubiquitin-protein ligase E3 resulting in the carboxy-terminal glycine of ubiquitin becoming covalently attached to a lysine residue of the protein substrate through an iso-peptide bond. The specificity of targeted proteins is largely controlled by E3 ligases. Successive conjugation of ubiquitin generates a polyubiquitin chain that is recognized by regulatory particles in the proteasome for degradation or as a trigger for various signaling pathways.

Ubiquitin is one of the most abundant cellular proteins, representing 1-5% of total cellular protein. Ubiquitin levels are maintained by recycling of ubiquitin from ubiquitin substrate conjugates by deubiquitinating enzymes (DUBs) and de novo ubiquitin synthesis through transcriptional regulation. In yeast, ubiquitin is produced by cleavage from precursor
proteins, where ubiquitin is fused to unrelated peptide sequences (UBI1, UBI2 and UBI3) or to itself (UBI4). UBI4 encodes a polyubiquitin protein that contains five consecutive ubiquitin repeats and is highly induced under stress conditions. Ubiquitination appears to play an important role in host-pathogen interactions. In a number of plant pathogenic fungi, polyubiquitin transcript levels significantly increased during *in planta* colonization or under environmental stress. In the human pathogen *Candida albicans*, inactivation of polyubiquitin gene UBI4 affected fungal growth, stress resistance and virulence. However, direct evidence for the role of ubiquitination in plant pathogenic fungi is lacking.

Rice blast is the most important disease of rice worldwide, and is caused by the filamentous ascomycete fungus, *Magnaporthe oryzae*. Common to many other phytopathogenic fungi, *M. oryzae* elaborates a specialized infection cell, the appressorium to infect its host. Perception of environmental cues, starvation responses, cell signaling pathways, turgor pressure generation, recycling cellular contents (autophagy) and cell cycle checkpoints are known to orchestrate the development of this specialized cell. Through global gene expression and functional analyses, we previously identified a connection between protein turnover and the infection process. Genes including *SPM1*, a vacuolar protease and *MGD1*, a putative NAD(+) dependent glutamate dehydrogenase (required for recycling carbon and nitrogen from amino acids back into metabolism) were up regulated during appressorium formation. Deletion of either gene resulted in defective appressoria and greatly reduced ability to cause disease. We also observed the transcripts of polyubiquitin gene (MGG_01282) to be significantly enriched during appressorium development as well as under nitrogen starvation conditions. In this study, we conducted a comprehensive
investigation of the role of protein ubiquitination in *M. oryzae* using pharmacological, molecular and proteomic approaches. Our findings suggest that ubiquitin-mediated proteolysis, which is known to be highly selective, plays a key role in nutrient assimilation, fungal development and pathogenicity of *M. oryzae*.

6.2 Materials and Methods

6.2.1 Effect of a Proteasome Inhibitor, Bortezomib, on *M. oryzae*

Wild type *M. oryzae* strain 70-15 was used for all experiments unless otherwise indicated. Conidia were collected from solidified V8 medium after 8 days in water and adjusted to 5x10⁴ conidia/ml. Bortezomib (LC laboratories) stock was prepared in DMSO (4ug/ul) and added to the conidial suspension to bring the final concentration up to 50 μM. 50ul of conidial suspension was placed on the appressorium inducing hydrophobic surface of Gelbond film (Lonza). After 24 and 48 hr incubation, conidia germination and appressorium formation were assessed. For infection assays, 10 ul of conidial suspension in Bortezomib was spotted onto detached 6 day old barley (ROBUST) leaves, placed in a humid plastic container and disease progress monitored for 5 days. Experiments were repeated three times with 3 replicates.

6.2.2 Prediction of Ubiquitin Associated Proteins in *M. oryzae* and their Transcription Analysis

Ubiquitin associated (UA) InterPro domains were selected based on other studies²⁰. *M. oryzae* proteins with UA Interpro domains were extracted from European Bioinformatics
Institute (EBI) Interpro databases (http://www.ebi.ac.uk/interpro/) using the Biomart filtering tool and were categorized according to function in the ubiquitination pathway. Gene expression for UA proteins during fungal development and nitrogen starvation were extracted from the previously reported genome-wide *M. oryzae* microarray data (NCBI GenBank: GSE1945, GSE10173 and GSE 2716).

6.2.3 Targeted Gene Replacement and Complementation of MGG_01282

Gene replacement cassettes were constructed using adaptamer mediated PCR as previously described\(^\text{21}\). Briefly about 1 kb of upstream and downstream sequence of MGG_01282 gene was amplified with primers that contained adaptamer sequences. A 1.5 kb fragment containing the hygromycin B phosphotransferase gene (HPH) driven by the *trpC* promoter from *Aspergillus nidulans* was amplified from plasmid PCB1003 using the adaptamer sequence attached to the forward HPHF and reverse HPHR primer set. Using nested primers from inside of the 5’ upstream fragment and from inside of the 3’ end of the downstream fragment of the target gene, the individual fragments and hygromycin resistance gene fragment were combined and amplified together to construct a hygromycin cassette for gene replacement approximately ~ 3.1 kb in length. The hygromycin cassette was transformed into 70-15 protoplasts as previously described\(^\text{22}\). Gene replacement mutants were identified by PCR screening and further confirmed by Southern blot analysis using ECL system of oligolabeling and detection (Amersham Co.). For complementation, a 2.6 kb DNA fragment corresponding to the MGG_01282 gene and its promoter region was PCR amplified from 70-15 wild type genomic DNA, and using gateway cloning system was cloned into a modified
pDONR221 plasmid in which the Bialophos resistance gene had been inserted. The complementation construct was introduced into mutant protoplasts using standard protocols and transformants were screened on Bialophos at 200μg/ml. The complemented strains were identified by PCR amplification of the insert and further confirmed by the recovery of wild type phenotypes.

6.2.4 Mutant Phenotype Assays

A series of phenotype analyses were conducted on several knockout mutants (≥3) and ectopic (≥2) transformants. Germination and appressorium assays were conducted using conidia collected from 8 day old V8 agar plates and adjusted to 5x10⁴ conidia/ml. Conidia suspension was spotted on the hydrophobic and hydrophilic surface of GelBond film and rate of germination and appressorium formation was measured after 24 hr incubation at 25 °C in the dark. To test for pathogenicity, barley and rice seedlings were spray inoculated with M. oryzae conidia suspension (5x10⁴ conidia/ml, Tween 20 0.025%) and incubated in dark humid conditions at 25 °C. The number and size of lesions were evaluated 5 days post-inoculation. Lesions on three leaves were counted for each strain, and this experiment was repeated three times. Wound assays were performed by a pinprick with a sterile needle on detached barley leaves. 5 millimeter square agar blocks from 8 day old culture on V8 of each strain were directly placed onto the wound site, and incubated in humid chambers as described above. Measurements were taken from nine wound sites (three per leaf) per strain, and this experiment was repeated twice. Disease progress and symptom development of MGG_01282 mutant was compared to wild type, ectopic, complemented strain or control
treatment. Growth rate assays were conducted by placing 10 μl conidia suspension (5x10⁴ conidia/ ml) on agar plates with complete media, minimal media and minimal media without nitrogen source. Colony morphology and diameters were recorded periodically for 15 days. The total number of conidia on minimal media plates was counted after 15 days incubation. Conidia size and morphology were also assessed using a minimum of 100 conidia per replicate. All experiments were conducted in triplicate and performed at least 3 times.

6.2.5 Sexual Reproduction

*M. oryzae* 70-15 strains were crossed with *M. oryzae* strain 4091-5-8, a weeping lovegrass (*Eragrostis curvula*) pathogen on oatmeal media. Plates were incubated at 25 °C for 7 days and then incubated at 20 °C for 21 days under constant light. Plates were examined for fruiting body formation. Perithecia were excised and crushed to identify ascospores under a brightfield microscope.

6.2.6 Western Blotting

Mycelia samples from wild type, mgg_01282 mutant and ectopic strains were collected from 3 day old liquid minimal media cultures, filtered through miracloth and washed with water. Excess liquid was squeezed out and 500 mg of semi-dried mycelia was ground using liquid nitrogen and resuspended with 2 ml lysis buffer containing 50mM HEPES (pH 7.5), 0.5% Nonidet P-40, 250mM NaCl, 10% (v/v) glycerol, 2mM EDTA (pH 8.0), and a complete protease inhibitor cocktail (Roche). Protein quantification of each soluble lysate was performed by Bradford assay. 85 μg of each protein sample was separated on a 4-20%
gradient gel (Invitrogen) and transferred to a nitrocellulose membrane. Blots were blocked in 5% bovine serum albumin (BSA) and antibody incubations were carried out in 5% skim milk followed by washes. Signals were detected by SuperSignal West Pico Chemiluminescent Substrate (Pierce), following the manufacturer’s instructions. The following antibodies were used. P4D1 (1:1000, Cell Signaling Technology), K63 (1:1000, Cell Signaling Technology) and Anti-mouse IgG (1:3000, Cell Signaling Technology).

6.2.7 Ubiquitinated Protein Profiling by NanoLC-MS/MS

Wild type *M. oryzae* 70-15 was incubated in liquid complete media for 3 days. Mycelia were collected, washed thoroughly with distilled water and then inoculated into liquid minimal media without nitrogen source. After 12 hr incubation, 3 replicates were pooled, mycelia tissues were collected and proteins extracted. The putative polyubiquitin protein sample was enriched using Agarose-Tube2 (LifeSensors) following manufacturer’s protocol. To prepare the sample for NanoLC-MS/MS, dithiothreitol (Biorad, Hercules) was added to the protein sample to a final of 5 mM and was incubated for 30 min at 56 °C to reduce the protein disulfide bonds. The samples were then mixed with 200 μl of 8M Urea (Sigma Aldrich), loaded onto Vivacon 500 μL ultrafiltration spin columns with 30kDa MW cutoff (Sartorius Stedim Biotech) and centrifuged at 14,000 x g for 15 min. The columns were washed one more time before Iodoacetamide (Sigma Aldrich) was added to a final of 20 mM. The samples were incubated for 30 min in the dark at room temperature for alkylation of the free thiols. The filter units were then centrifuged at 14,000 x g for 10 min. Three 100 μl 8M Urea washes and 3 0.05 M ammonium bicarbonate (Sigma Aldrich) washes with 15 min
subsequent centrifugations were performed. The collection vials were changed and 0.4 μg/μl trypsin (Sigma Aldrich) was added at a 1:100 enzyme:protein ratio. Digestion was performed over night at 37 °C and peptides were eluted by adding 40 μl ammonium bicarbonate and centrifuging the filter units at 14,000 x g for 10 min. Each sample was injected 3 times to a LTQ FT Ultra Hybrid Mass Spectrometer and the data analyzed as previously described24.

6.3 Results

6.3.1 Inhibition of Proteasome Mediated Protein Degradation Blocks Conidial Germination, Appressorium Induction and Pathogenicity in M. oryzae

To investigate the role of ubiquitin mediated proteasomal protein degradation during the infection-related fungal development, we treated conidia with the proteasome inhibitor, Bortezomib. On an appressorium inductive hydrophobic surface, both conidial germination and appressorium formation were significantly delayed in a dose-dependent manner with 20 μM Bortezomib being completely inhibitory after 24 hr incubation. However, at 48 hr, most conidia germinated and successfully formed melanized appressoria (Figure 6.1 A, B). Bortezomib also inhibited pathogenicity. Compared to typical spreading necrotic lesions on control infected barley leaves, the addition of 1 μM Bortezomib resulted in smaller lesions. Symptom development was completely blocked when the conidia solution contained 50 μM Bortezomib (Figure 6.1C). 50 μM Bortezomib solution had no observable effect on barley leaves. These data are consistent with proteasome mediated protein turnover being required for infection related development and pathogenicity in M. oryzae.
Figure 6.1 Bortezomib blocks conidia germination, appressorium formation and pathogenicity. A) Conidia germination and appressorium formation (green and red bar respectively) were measured after 24 and 48 hr incubation on a hydrophobic surface. B) Pathogenicity assays were performed on 6 day old barley leaves. Disease progress was assessed after 5 day incubation.
6.3.2 Components of Ubiquitin Mediated Protein Modification are Highly Conserved and Regulated in Response to Developmental and Nutritional Stimuli in *M. oryzae*

**Table 6.1 Ubiquitin Pathway Associated Proteins in *M. oryzae***

<table>
<thead>
<tr>
<th>Class</th>
<th>InterPro domain</th>
<th># of proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA</td>
<td>IPR00626 Ubiquitin</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>IPR004854 Ubiquitin fusion degradation protein</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>IPR001012 UBX</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>IPR003892 Ubiquitin system component Cue</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>IPR000449 Ubiquitin-associated/translation elongation factor EF1B, N-terminal</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>IPR003903 Ubiquitin interacting motif</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td><strong>UA total</strong></td>
<td><strong>44</strong></td>
</tr>
<tr>
<td>E1</td>
<td>IPR000011 Ubiquitin-activating enzyme, E1-like</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>IPR000127 Ubiquitin-activating enzyme repeat</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>IPR000594 UBA/THIF-type NAD/FAD binding fold</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td><strong>E1 total</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td>E2</td>
<td>IPR00608 Ubiquitin-conjugating enzyme, E2</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td><strong>E2 total</strong></td>
<td><strong>21</strong></td>
</tr>
<tr>
<td>E3</td>
<td>IPR001232 SKP1 component</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>IPR003126 Zinc finger, N-recognition</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>IPR003613 U box domain</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>IPR001373 Cullin, N-terminal</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>IPR000569 HECT</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>IPR001810 F-box domain, cyclin-like</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>IPR001841 Zinc finger, RING-type</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td><strong>E3 total</strong></td>
<td><strong>94</strong></td>
</tr>
<tr>
<td>DUB</td>
<td>IPR001607 Zinc finger, UBP-type</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>IPR001578 Peptidase C12, ubiquitin carboxyl-terminal hydrolase 1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>IPR001394 Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td><strong>DUB total</strong></td>
<td><strong>20</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>183</strong></td>
</tr>
</tbody>
</table>

To assess the degree of conservation of proteins associated with ubiquitin mediated protein modification, we conducted an InterPro domain search. One hundred and eighty three
proteins putatively involved in ubiquitination pathway in *M. oryzae* were identified (Table 6.1). Forty four proteins directly related to ubiquitin (UA) were found including 12 containing ubiquitin and 13 with an ubiquitin interacting motif. Eight proteins were ubiquitin-activating enzymes, E1, and 21 proteins were ubiquitin-conjugating enzymes, E2. The most diverse and largest group was ubiquitin ligase, E3, which had 94 members and included 53 zinc finger, ring-type (IPR001841) domain containing proteins. Components of multi-protein E3 ubiquitin ligase Skp, Cullin and F-box containing complex (SCF complex) were found and included one Skp, four Cullins and 26 F-box containing proteins. A total of 20 proteins were identified as proteins for de-ubiquitination.

In previous work, we conducted extensive microarray analysis. Further inspection of the data revealed that 161 of the 183 ubiquitin pathway genes were expressed during conidia germination, appressorium formation or nitrogen starvation (Figure 6.2). The most dramatic changes in gene expression occurred during conidia germination, where 87.6 % of the genes (141 out of 161) showed at least a 50% increase in transcript levels and 39 % were significantly induced (>2 fold up, p<0.05). This is in contrast to the entire transcriptome, where only 21% (2,087 of 10,176) of genes were significantly (>2 fold up, p<0.05) induced. Sixty three percent (5 out of 8) of genes encoding ubiquitin-activating enzyme, E1, 42 % (8 out of 19) of ubiquitin-conjugating enzyme, E2, 29 % of (23 out of 79) ubiquitin ligase, E3 and 39 % (7 out of 18) of de-ubiquitinating enzyme showed increased gene expression. No ubiquitin pathway genes exhibited significant down regulation during conidia germination, except for MGG_01282. Transcripts of MGG_01282, predicted to encode a polyubiquitin protein, were significantly more abundant in intact conidia compared with germinating cells.
Contrary to the conidia germination, transcript levels of the majority of ubiquitin associated proteins did not change significantly during appressorium formation. Only 3.7 % (6) and 5.0 % (8) of genes were induced during appressorium induction in response to hydrophobic surface signal and cyclic AMP, respectively, and 0.6 % (1) and 2.5 % (4) genes were down regulated. Transcripts of MGG_01282, a polyubiquitin protein and MGG_07127, an autophagy-related E1 like protein were more abundant during appressorium formation compared to germinating conidia. No ubiquitin-conjugating enzyme, E2, genes were differentially expressed. Among 79 ubiquitin ligases, E3, 4% (3) and 8% (6) genes were induced by physical and chemical signals and only one gene, MGG_10932 was down regulated. Two F-box proteins, MGG_07785 and MGG_08019, were up regulated by both

Figure 6.2 Ubiquitin associated genes are differentially expressed during conidia germination, appressorium formation and under nitrogen starvation. Each column represents hierarchical clustering profile for spore germination (SG), appressorium maturation (AM), cAMP induced appressoria formation (CI) and nitrogen starvation for 12 (N12) and 48 (NS48) hr. Genes of same expected function were grouped together as ubiquitin associated (UA), ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2), ubiquitin protein ligases (E3) and de-ubiquitinating enzymes (DUB).
signals. MGG_08638, a ubiquitin C-terminal hydrolase was the sole de-ubiquitination gene that was induced by the hydrophobic physical cue.

Examination of previous data under nitrogen starvation conditions revealed transcriptional induction of about 9% (14) and 11% (18) of ubiquitin pathway genes, and 3 % (4) and 2 % (3) with reduced expression of 12hr and 48hr, respectively. Two F-box containing proteins (MGG_00768 and MGG_04395) and one zinc finger, ring-type E3 ligase (MGG_02837) were induced at both 12hr and 48hr. The expression of only 2 genes was up-regulated by appressorium inducing physical and chemical signals as well as by nitrogen starvation. Transcript levels of MGG_01282 were significantly elevated during appressorium maturation (3.7 fold) and cyclic AMP induction (3.0 fold) as well as with nitrogen starvation for 12 hr (8.2 fold) and 48 hr (2.4 fold). Expression of the F-box protein, MGG_07785 was induced during appressorium formation and at 12 hr nitrogen starvation. In sum, transcription levels of a large proportion (39%) of genes associated with the ubiquitination pathway were induced during conidia germination. However, only a small proportion (between 5-11%) was induced during appressorium formation or in response to nitrogen starvation. Very few genes were induced under both conditions. One of the notable exceptions was the gene encoding polyubiquitin, MGG_01282.
6.3.3 Highly Conserved but Structurally Diverse Ubiquitin Genes in *M. oryzae* Show Different Expression Patterns during Development and Starvation

In addition to MGG_01282, which encodes a protein with 4 ubiquitin monomers, we identified 11 other proteins in *M. oryzae* that contain single or partial ubiquitin motifs. Some of which contained other functional domains that are related to the ubiquitination pathway (Figure 6.3A). MGG_09887 encodes a single ubiquitin monomer, whereas other single ubiquitin proteins MGG_06044 and MGG_07928 contain large and small ribosomal subunits at their C terminals, respectively. Among the proteins that contain a partial ubiquitin moiety, MGG_01318 and MGG_01656 possess an ubiquitin associated domain and MGG_02970 has a ubiquitin carboxyl-terminal hydrolase domain at the C terminal region. Two proteins, MGG_05137 and MGG_05737 contain an ubiquitin-like SUMO domain at their C terminal and N terminal regions, respectively.

Except for MGG_01282, expression of all ubiquitin containing genes were induced during conidia germination, however, no significant change was observed during appressorium induction and nitrogen starvation in most cases (Figure 6.3). The exceptions were MGG_07928 and MGG_06044, which were down-regulated during both conditions. Gene expression of MGG_09887, encoding a single ubiquitin moiety, was induced under nitrogen starvation. Because MGG_01282 showed the most dynamic changes in gene expression, being highly induced during appressorium formation and nitrogen starvation, it was subjected to further examination.
Figure 6.3 Clustering and gene expression of 12 ubiquitin related proteins in *M. oryzae*. A) Proteins were grouped using ClustalW and functional domains indicated. Green oval structure indicates a ubiquitin moiety and others in green indicate partial or ubiquitin like domains. B) Gene expression during spore germination (SG), appressorium maturation (AM), cAMP induced appressoria formation (CI) and nitrogen starvation for 12 (NS12) and 48 (NS48) hr.
6.3.4 Nitrogen Starvation Results in a Dramatic Increase in Protein Ubiquitination which is Mediated by the Polyubiquitin Protein, MGG_01282

Protein ubiquitination is known to be directly linked to the cellular nutrient status\(^{25}\). To examine this relationship in \textit{M. oryzae}, we investigated the correlation between ubiquitination and one of the major developmental signals, nitrogen starvation. As shown by immunoblots with antibodies recognizing polyubiquitin and ubiquitin in \textbf{Figure 6.4A}, when the fungus was exposed to the nitrogen limiting conditions for 12 hr, protein ubiquitination dramatically increased compared to the condition without nitrogen stress.

\textbf{Figure 6.4} Protein ubiquitination is induced by nitrogen starvation and requires the polyubiquitin gene, MGG_01282. A) 70-15 \textit{M. oryzae} was grown in liquid minimal medium with (+N) and without (-N) nitrogen sources. Protein extracts from each sample were probed with an antibody recognizing both ubiquitin and polyubiquitin (P4D1). An anti-actin antibody was used to compare the relative amount of total proteins in each lane. B) and C) Wild type (WT), MGG_1282 deletion mutants (MT1, MT2) and ectopic (EC) strains were incubated in minimal medium. Protein extracts from each sample were probed with antibodies recognizing ubiquitin and polyubiquitin (P4D1) or specifically polyubiquitin (K63), respectively. Panel B were exposed to X-ray film for a longer compared to panel A.
To further investigate the function of MGG_01282, we generated deletion mutants using standard protocols. Immunoblot analyses of proteins extracted from growth on minimal medium or under nitrogen starvation showed mutant strains contained strikingly less ubiquitinated proteins compared to wild type and ectopic strains (Figure 6.4B). In addition, immunoblot analyses using the linkage specific antibody (K63) revealed mgg_01282 deletion mutants contained dramatically reduced K63 linked polyubiquitin targets. K63 ubiquitin targets are known to be involved in cellular development and signal transduction in eukaryotic cells. In summary, protein ubiquitination was significantly elevated under nitrogen starvation and was primarily mediated by polyubiquitin protein MGG_01282.

6.3.5 The Polyubiquitin Protein MGG_01282 is Essential for Fungal Growth, Development and Pathogenicity

To investigate the biological role of polyubiquitin in M. oryzae, we compared phenotypes of the knockout mutant with ectopic and wild type strains as well as with a strain complemented with MGG_01282. Deletion of MGG_01282 resulted in significant changes in fungal growth, morphology and development in M. oryzae. Radial growth of mutants on solid nutrient rich complete medium or minimal medium was significantly retarded (about 17%) and aerial hyphae was suppressed compared with the wild type. This growth reduction was more severe under nitrogen limiting conditions (Figure 6.5A and B).
Figure 6.5 A) 70-15 wild type (WT) and mutant strain (MT) were incubated for 7 days on complete medium (CM), minimal medium (MM) and minimal medium without nitrogen source (MM-N). B) Radial mycelial growths were measured in mutant stain MT1 and MT2 and were compared to those of ectopic (EC), WT and MGG_012982 complemented strain (CP).

MGG_01282 mutants also showed defects in conidiation. The number of conidia generated by the mutants after 7 day incubation on V8 medium was significantly reduced (more than 90%) compared to the wild type (Figure 6.6).
Figure 6.6 MGG_01282 is required for normal conidia production. 70-15 wild type (WT), MGG_01282 deletion mutants (MT1, MT2), ectopic (EC) and MGG_01282 complemented (CP) strains were incubated on V8 medium (A, B) and the average number of conidia produced 10 days after inoculation (C).

In addition, while most conidia (≥94%) from wild type consisted of three cells, divided by two septa, about half of mutant conidia contained single (35%) or no septa (13%) (Figure 6.7A). Mutant conidia were smaller and more rounded than those produced by the wild type strain (Figure 6.7B). Ectopic and complemented strains were indistinguishable from the wild type for growth, conidiation and conidia morphology (Figures 6.5-6.7).
Figure 6.7 MGG_01282 deletion mutants produced abnormal conidia. A) For each strain, conidia collected from V8 medium were grouped according to the number of septa, double (a); single (b) or no septa (arrow marked). The percentage of each group per strain is presented. B) Average number of septa, conidia length and width is presented.

<table>
<thead>
<tr>
<th>Strain</th>
<th># septum</th>
<th>conidia length (um)</th>
<th>conidia ore width (um)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>1.9 ± 0.3</td>
<td>24.9 ± 2.6</td>
<td>10.3 ± 0.9</td>
</tr>
<tr>
<td>MT</td>
<td>1.4 ± 0.7</td>
<td>23.8 ± 5.6</td>
<td>8.6 ± 1.3</td>
</tr>
<tr>
<td>EC</td>
<td>1.9 ± 0.2</td>
<td>24.9 ± 2.5</td>
<td>10.0 ± 1.1</td>
</tr>
<tr>
<td>CP</td>
<td>2.0 ± 0.2</td>
<td>25.0 ± 1.6</td>
<td>9.6 ± 0.9</td>
</tr>
</tbody>
</table>

Almost all wild type and ectopic conidia germinated on hydrophobic surface and about 90% of them successfully developed appressoria after 24 hr incubation. In mutant strains, 62% of conidia germinated but most failed to develop appressoria (Figure 6.8 A and B). In addition,
mutant appressoria were less pigmented and unable to cause disease. When sprayed on barley and rice leaves, mutant strains produced no disease symptoms (Figure 6.8 C).

Figure 6.8 MGG_01282 is required for appressorium formation and pathogenicity in rice and barley. A) Germination and appressorium formation of conidia from 70-15 wild type (WT), MGG_01282 deletion mutants (MT1, MT2) and ectopic (EC) strains on a hydrophobic surface after 24 hr incubation. B) Percentage of conidia germination and appressorium formation from at least 100 conidia per replicate with 3 replicates per strain. C) Disease development of each strain including a complemented (CP) strain inoculated onto barley (upper panel) and rice (lower panel) seedlings. Disease progress was evaluated compared to water treated control (C) 5 days after inoculation. D) Wound assay on detached barley leaves. Barley leaves were wounded by making a tiny pinprick with a sterile needle. Leaves were inoculated with 8 day old V8 agar block of each strain and incubated in a humid chamber. Disease progress was evaluated compared to control agar and photographed 5 days post-inoculation.

To investigate whether MGG_01282 mediated growth in planta, barley leaves were inoculated onto pinprick wounded sites. MGG_01282 deletion mutants showed no blast symptoms on wounded plants (Figure 6.8D). Taken together, these data indicate that the
polyubiquitin gene, MGG_01282 is essential for conidiation, appressorium formation and invasive growth in planta.

*M. oryzae* 70-15 is heterothallic and requires an opposite mating type strain for sexual reproduction. In this study, we crossed the wild type 70-15, MGG_01282 deletion mutant, ectopic and complemented strains with strain 4091-5-8. Four weeks after crossing, wild type, ectopic and complemented strains developed melanized perithecia, which contained ascoconidia within the asci (**Figure 6.9**). In these crosses, all strains were female fertile and two rows of perithecia were clearly visible.

**Figure 6.9** MGG_01282 deletion mutants are female sterile. 70-15 wild type (WT), MGG_01282 deletion mutant (MT), ectopic (EC) and MGG_01282 complemented (CP) strains were crossed with the opposite mating type, 4091-5-8 strain, on oat meal media. Fruiting bodies formed in each test strain and 4091-5-8 strain are marked with white and black arrows respectively. Below are shown corresponding perithecia and erupting ascospores. The melanized structure in MT is indicated by a gray arrow with corresponding enlargement below.
In crosses with MGG_01282 deletion mutant, the number of perithecia was reduced. Within 4091-5-8 tissue, typical asci were observed. However, within the deletion mutant tissue only highly condensed and melanized fungal structures were detected that failed to further develop into perithecia (Figure 6.9). These data indicate that MGG_01282 is required for normal sexual development.

6.3.6 Ubiquitinated Protein Profiling during Fungal Growth under Nitrogen Stress

We found that protein ubiquitination is highly increased during nitrogen starvation. This suggests that when nitrogen is limiting, M. oryzae actively recycles cellular materials and redirects biological pathways to cope with the nitrogen deficiency for survival, at least part of which is mediated by protein ubiquitination. To identify proteins targeted for ubiquitination under nitrogen stress, we incubated wild type M. oryzae in minimal media without a nitrogen source for 18 hr and enriched polyubiquitinated proteins using agarose TUBE2. TUBE2 contains a protein with very high affinity for polyubiquitinated proteins\textsuperscript{27}. As a negative control, we used agarose lacking the affinity protein. We found the enrichment to be effective as evidenced by protein staining and immunoblot analysis of the enriched protein fraction. Mass spectral analysis revealed 63 proteins to be unique or significantly enriched in the affinity purified sample compared to negative control. As expected, ubiquitin was found to be the most abundant protein based on spectral counts. Based on gene ontology using BLAST2GO\textsuperscript{28}, compared to the entire proteome, proteins involved in translation (30.2%), metabolic process (17.4%), transport (14.3%), and protein metabolic process (9.5%) including ubiquitin proteins and components of the proteasome were over-represented.
Several proteins associated with cytoskeleton and stress responses including actin, tubulins and heat shock proteins were also identified. We further identified proteins predicted to be involved repression of carbon catabolism such as homologs of a hulA E3 ligase (MGG_07255) and as arrestin containing protein (MGG_01045) as well as proteins involved in cell signaling including MGG_01588 and MGG_13806, both 14-3-3 proteins.

6.4 Discussion

The availability of the entire genome sequence and global gene expression profiles for *M. oryzae* has enabled new insight into infection related development and pathogenicity\(^{19,29}\). Our previous studies revealed a hitherto unknown link between protein degradation and infection structure development in *M. oryzae*\(^ {19}\). Genes required for the non-selective protein degradation process referred to as autophagy have been characterized and shown to be important for fungal pathogenicity in *M. oryzae*\(^ {30,31}\). On the other hand, prior to this study, little direct evidence was available linking the highly selective protein degradation process, mediated by ubiquitin, to fungal pathogenicity.

To establish a link between protein turnover through the ubiquitin-proteasome complex and fungal pathogenicity, we first demonstrated that treatment with the 26S proteasome inhibitor, Bortezomib, resulted in a significant delay in germination and appressorium formation. Moreover, addition of Bortezomib to inoculum blocked symptom development, even though melanized appressoria were formed after 48 hr. A recent report, which was published during the course of this research, also showed the proteasome inhibitors MG-132, proteasome
inhibitor I and proteome inhibitor II delayed conidia germination and appressorium formation in *M. oryzae* as well as suppressed infection of rice leaves\textsuperscript{32}. In order to elucidate the underlying mechanisms, we then explored the machinery associated with ubiquitin mediated protein turnover. Similar to other eukaryotic organisms\textsuperscript{20}, many of the components are highly conserved in *M. oryzae*. Examination of whole genome microarray gene expression data further revealed that most of the components for protein ubiquitination were induced upon conidia germination but did not dramatically change expression during appressorium formation. The elevated expression of genes associated with ubiquitination during germination in the absence of external nutrients likely reflects the cells preparation for recycling proteins and other storage components. A few genes were observed to be induced during appressorium induction and nitrogen starvation but none of them had been previously characterized. Interestingly, we discovered that a polyubiquitin encoding gene MGG_01282 was most highly expressed in intact conidia rather than in germinating conidia and was significantly induced during appressorium formation and nitrogen starvation.

In fungi, polyubiquitin does not appear to be an essential gene, although it likely provides the main supply of cellular ubiquitin protein in response to developmental and environmental stimuli. Here, we showed by Western blot analysis that protein ubiquitination is highly induced under nitrogen starvation and through examination of the knock-out mutant that protein ubiquitination is mainly mediated by MGG_01282 gene products. The gene deletion mutant in *M. oryzae*, although viable, exhibited numerous phenotypic detects, including defects in mycelia growth, conidia morphology, sexual reproduction, infection structure
development and pathogenicity. Similar types of defects have been observed in *Saccharomyces cerevisiae* and *C. albicans*. Loss of *UBI4* in *S. cerevisiae* resulted in increased sensitivity to starvation and to amino acid analogs as well as reduced growth at high temperatures\(^{33-35}\). *UBI4* deletion mutant of *C. albicans* grew relatively normally on rich media, but displayed morphological and cell cycle defects when exposed to a number of stresses including temperature, peroxide and several anti-fungal drugs that interfere with cell wall biosynthesis\(^9\). In *M. oryzae*, the polyubiquitin deletion mutant although growing relatively poorly on all media, was most affected under nitrogen starvation. These data are consistent with ubiquitination playing a major role in protein turnover required for normal growth and development under a variety of stress conditions.

Examination of the 63 proteins unique or significantly enriched following affinity purification for polyubiquitination revealed that the most abundant group of proteins was components of the ribosome. Other studies have also shown that proteins associated with ribosome are ubiquitinated\(^9,36\). As may be expected under nitrogen starvation, translation is curtailed and the machinery is recycled via ubiquitination and proteasome mediated degradation. We also identified other proteins including heat shock proteins and a succinate dehydrogenase, which in *C. albicans* showed increased ubiquitination in response to heat and oxidative stress\(^9\). In a global analysis of ubiquitination in the human cell line HEK293, 236 proteins were identified to be ubiquitinated\(^37\). Among these, we found 20 that matched *M. oryzae* proteins in our ubiquitinated protein data set and these included heat shock proteins, actin and tubulins and ribosomal proteins. These data suggest a certain level of conservation of proteins targeted for ubiquitination across kingdoms.
It was noteworthy that two of the proteins targeted by ubiquitin identified in our data set were a HECT domain containing E3 ligase, MGG_07255 and an arrestin domain containing protein MGG_01045, orthologs of Rsp5 and Rod1 in *S. cerevisiae* respectively. Rsp5 is known to be involved in a variety of cellular process including endocytosis, multivesicular body sorting and RNA stability\textsuperscript{38-40}. The arrestin-like adaptor, Rod1 binds to Rsp5 and mediates ubiquitination and endocytic internalization of membrane transporters, which are degraded in the vacuole. In *A. nidulans*, CreD, an ortholog of Rod1, has been shown to control carbon catabolite repression with possible interaction with HECT ligase, HulA\textsuperscript{41}. Self-ubiquitination of other HECT ligase ortholog NEDD4-1, NEDD4-2 is regulated through intramolecular interaction between the WW domains and PY motifs in HECT domains\textsuperscript{42,43}. This suggests that ubiquitination and following degradation of MGG_07255 is tightly regulated depending upon environmental conditions. During nitrogen starvation, ubiquitination of MGG_07255/MGG_01045 could be important for reprogramming fungal cells to cope with nutrient limiting conditions.

Over the past twenty years, the role of protein ubiquitination in eukaryotic cells has been emerging, however, relatively little is known related to fungal development, pathogenicity and disease control. Here, we showed through a combination of pharmacological, molecular and proteomic analysis that ubiquitin mediated posttranslational modification is a central regulator in fungal nutrition, development and pathogenicity. Future study will be focused on specific components of the ubiquitination processes including identification of specific ubiquitination target proteins, which may offer up novel strategies for plant disease control.
6.5 References


(38) Shcherbik, N.; Pestov, D. G.: The Ubiquitin Ligase Rsp5 is Required for Ribosome Stability in *Saccharomyces cerevisiae*. *RNA* 2011, 17, 1422-1428.


## GLOSSARY

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1D-PAGE</strong></td>
<td>One dimensional polyacrylamide gel electrophoresis, protein separation method based on molecular weight.</td>
</tr>
<tr>
<td><strong>2D-PAGE</strong></td>
<td>Two dimensional polyacrylamide gel electrophoresis, protein separation method based on molecular weight and isoelectric point.</td>
</tr>
<tr>
<td><strong>Conidium</strong></td>
<td>Asexually produced spore.</td>
</tr>
<tr>
<td><strong>Mycelium</strong></td>
<td>Vegetative part of the fungus.</td>
</tr>
<tr>
<td><strong>GelFree</strong></td>
<td>Fractionation technique based on molecular weight with the ability to recover the sample in liquid form.</td>
</tr>
<tr>
<td><strong>Glufosinate</strong></td>
<td>Herbicide that kills unwanted weed by inhibiting the glutamine synthetase.</td>
</tr>
<tr>
<td><strong>Glyphosate</strong></td>
<td>Herbicide that kills unwanted weed by inhibiting aromatic amino acid synthesis.</td>
</tr>
<tr>
<td><strong>Gradient Elution</strong></td>
<td>Steady changes of the mobile phase composition during the chromatographic run.</td>
</tr>
<tr>
<td><strong>Hemibiotrophic</strong></td>
<td>Organism that parasites the host for a period of time and continues its life cycle on dead tissues.</td>
</tr>
<tr>
<td><strong>Isocratic Elution</strong></td>
<td>Composition of the mobile phase stays unchanged during the chromatographic run.</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>MS/MS</strong></td>
<td>MS measurement of ions resulting from the fragmentation of a previously isolated precursor ion.</td>
</tr>
<tr>
<td><strong>NanoLC</strong></td>
<td>Liquid chromatography performed at nano liter flow rates.</td>
</tr>
<tr>
<td><strong>OFFGEL</strong></td>
<td>Fractionation technique based on isoelectric focusing with the ability to recover the sample in liquid form.</td>
</tr>
<tr>
<td><strong>Random-Database</strong></td>
<td>Masses of the original protein sequences are calculated and decoy proteins of the same mass but consisting of different amino acids are created.</td>
</tr>
<tr>
<td><strong>Reverse-Database</strong></td>
<td>Forward protein sequences are reversed.</td>
</tr>
<tr>
<td><strong>Shuffle-Database</strong></td>
<td>Amino acids of the target proteins are shuffled.</td>
</tr>
<tr>
<td><strong>Soft Ionization Technique</strong></td>
<td>Produces mass spectra with little or no fragment-ion content.</td>
</tr>
<tr>
<td><strong>Spectral Counting</strong></td>
<td>Counts the number of spectra identified for a protein.</td>
</tr>
<tr>
<td><strong>Turgor pressure</strong></td>
<td>Pressure generated on the plant cell wall through the mechanism of osmosis.</td>
</tr>
</tbody>
</table>