ABSTRACT

ROBERTSON, SUZANNE LOUISE. A Global Proteomic Investigation of *Magnaporthe oryzae* During Nitrogen Starvation Conditions. (Under the direction of Dr. David C. Muddiman).

For a variety of crops the fungal pathogen, *Magnaporthe oryzae,* can lead to serious losses. *M. oryzae* can infect wheat, millet, and barley but it has proved most devastating in rice crops. Rice is predominantly grown and consumed in Asia, where *M. oryzae* is responsible for the greatest amount of crop loss, routinely destroying 10-30% of harvests. As a result, this filamentous ascomycete fungus has been the focus of a significant body of research that spans decades.

In the post-genome sequencing era, the field of proteomics has become highly prolific, contributing novel insights across a plethora of biological fields. Proteomics seeks to identify all proteins within a system and to understand the structures, quantities, locations, functions, and to learn how the proteome changes in physiological states which has only recently become possible to realize due to advances in mass spectrometry instrumentation, software, and bioinformatics tools. Indeed, global proteomic experiments can now routinely identify thousands of proteins in an experiment, yielding large amounts of information.

The focus of this current investigation is the response of *M. oryzae* to nitrogen starvation. A global proteomic experiment was performed using a protein lysate from mycelium that was cultured in a nitrogen-scarce environment and analyzed with the Q-Exactive HF mass spectrometer. Insight was gained about the infection process that *M. oryzae* could encounter *in planta* during nitrogen-limited conditions. It appears that *M. oryzae* initiates the infection process rather than dedicating resources to growth and proliferation when faced with a nitrogen-limited environment. When the proteomic results

were compared to a previous transcriptome study, similar results were observed between gene and protein expressions. This work will greatly contribute to the understanding of the infection process of *M. oryzae* when it encounters an *in planta* nitrogen-limited environment or when nitrogen is highly abundant as is the case when excessive nitrogenous fertilizers are applied in the field.

A Global Proteomic Investigation of *Magnaporthe oryzae* During Nitrogen Starvation Conditions

by
Suzanne Louise Robertson

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Masters of Science

Chemistry

Raleigh, North Carolina

2016

APPROVED BY:

_____	_____
David C. Muddiman	James N. Petitte
Committee Chair	
_____	
Edmond F. Bowden	

DEDICATION

This work is dedicated to a number of people who have helped me along my path. To my parents, for my Mom who has always been there to love and encourage me to pursue my dreams. My Dad who passed away far earlier than I would have liked, but who was my first mentor in nurturing an interest in science beginning from a very young age. As well as to my sister, even if she gets grossed out by harvesting mice organs. To Jeff and Tsila who are wonderfully supportive and loving in-laws. And of course, to my husband who is a constant source of love and comfort, this would have been so much harder without you by my side.

In addition to my family, I would like to also dedicate this work to some wonderful mentors at the University of Florida. To Dr. Mike Clare-Salzler, who gave me my first taste of research and inspired an appreciation and awe for the complexity of biological sciences, and continues to be a source of support. To Dr. Hatch, from who I gained insight into the fundamental ways of approaching research, as well as her continued guidance and support. To Dr. James Deyrup, Dr. Malgorzata Deyrup, and Dr. Harrison, for all of their valuable input and guidance during my undergraduate studies and in the years after graduation.

Thank you to all of you for the many years that you have contributed to the steps in my journey.

BIOGRAPHY

Suzanne Louise Robertson began her studies at the University of Florida where she graduated *Cum Laude* with a bachelor's of science in chemistry (focus biochemistry) and a bachelor's of arts in Spanish. She wanted to pursue her graduate studies in a combination of protein work and mass spectrometry, and joined the group of Dr. David C. Muddiman at North Carolina State University in August of 2013. Since then, she has gained experience in the field of proteomics.

ACKNOWLEDGMENTS

I thank a number of people for their help in completion of this thesis. First, thank you to Dr. David C. Muddiman for his guidance as my graduate mentor and to my committee members, Dr. Bowden and Dr. Petitte. Thanks to the entire Muddiman group, your support has been invaluable. I would like to thank our collaborators, Dr. Ralph Dean and Dr. Yeonyee Oh, you both have been great to work with. And of course, I greatly appreciate the USDA as my funding source in this investigation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. GENERAL BACKGROUND

1.1 Importance of Filamentous Fungi

Comprising their own kingdom, fungi are eukaryotic organisms that include yeasts, molds and mushrooms.[1] Filamentous fungi play an important role in the biotechnology industry as they are used in the production of human therapeutics and specialty chemicals.[1] For example, they manufacture statins, the cholesterol-lowering drugs which comprises a market of nearly $15 billion dollars per year in the USA alone.[1] In addition, filamentous fungi are well-known human, animal and plant pathogens. The focus of this particular study will be on the rice blast fungus, *Magnaporthe oryzae,* which has the ability to infect various grass species including rice, wheat, millet and barley.[2,3]

Rice is a major food staple in many Asian countries where approximately 90% of rice is grown and consumed, and it is most commonly threatened by M. oryzae which routinely causes a crop loss of 10-30%[3].[4] In China, rice blast epidemics destroyed 5.7 million hectares of rice from 2001-2005, and it causes an annual crop loss of 25% (275,000 tons) in Japan.[4,5] Given that *M. oryzae's* is the most substantial cause of rice crop loss, gaining a better understanding of its infection process is crucial in learning how to combat crop infection and create greater stability in the world's food supply.

In addition to social and economic impacts of rice blast, it is an ideal model organism for the study of plant pathogens.[6] Both *M. oryzae* and its primary host rice (*Oryzae sativa)* have had their complete genomes sequenced which facilitates genetic and proteomic studies of the pathogen and host-pathogen interactions.[7,8] DNA transformations are difficult to achieve in filamentous fungi, which was overcome by insertional mutagenesis which generated a collection of 20,000 mutants.[9] In addition, *M. oryzae* shares key characteristics

like appressorium formation and intracellular tissue invasion with similar plant pathogens that infect grains. As a result, its study could result in the discovery of conserved mechanisms and lead to treatments that target a wider variety of crop pathogens.[4]

1.2 Biology of Magnaporthe Oryzae

Magnaporthe oryzae is a filamentous ascomycete fungus. It can reproduce sexually and creates an ascus (sack) that contains 4 or 8 ascospores (spores formed from meiosis contained in the ascus).[4] It is classified as a hemibiotrophic fungus because it initiates its life with a biotrophic phase as it lives and grows as a part of the host in a symbiotic relation.[10] During this time, *M. oryzae* suppresses the host's immune system and programmed cell death, while invasive hyphae spread and absorb nutrients from its host. This is followed by a necrotrophic phase in which *M. oryzae* secretes toxins to kill its host cells and consumes the nutrients released from the dead cells while undergoing sporulation to spread its infection .[10]

Conidia are asexual spores released during the necrotic phase of *M. oryzae's* lifespan, and the infection process begins when a conidium adheres to the leaf surface in the presence of water, and rapidly geminates. A single, polarized germ tube extends across the leaf surface and differentiates into the appressorium, a specialized dome-shaped cell used by plant pathogenic fungi to penetrate through the plant's cuticle and initial layer of epidermal cells.[4,5] A hard, hydrophobic surface is required to trigger appressoria differentiation via the cyclic adenosine monophosphate (cAMP) and mitogen-activated protein kinase (MAPK) signaling pathways.[2,3]

The appressorium formation is the first critical step of a successful infection. Enormous turgor pressure of up to 8 MPa is generated to mechanically pierce through plant tissues and it is achieved through a high concentrations of glycerol present in the

appressorium.[2] Once enough pressures has been generated, the penetration peg extends from the bulbous appressorium, penetrating the plant's cuticle and epidermal layers.[2,4,5] The appressorium's cell wall is highly differentiated with an outer layer composed of chitin and an inner layer of melanin, which is a key virulence factor.[2,4] Mutants that lacked this melanin layer were unable to generate sufficient turgor pressure to penetrate into plant cells to initiate an infection.[2] For this reason, it's believed that melanin's role in the appressorium is to inhibit solute efflux, giving *M. oryzae* the ability to produce its massive turgor pressure for access to the host cell.[2] In addition to the appressorium's penetration peg, additional enzymes such as cutinases also aid in infiltration of the plant's epidermal cells.[5]

Once the host has been breached, primary infection hyphae begin to quickly fill this first infected cell, where a significant amount of endo-exocytosis occurs between the fungus and plant, and this process is well summarized in the 2007 review by Ribot *et al.*[5] The invasive hyphae then infect neighboring cells through the plant's plasmodesmata (tiny channels that connect the cytoplasm of adjacent plant cells). By the time *M. oryzae* has moved onto a new cell, the invaded cell shows reduced viability but is not yet killed. The sum of the invasive hyphae comprise the fungus's mycelia. In 4-5 days after the initial penetration, *M. oryzae* switches to its necrotic phase thereby killing the colonized tissues, allowing for sporulation to occur which further spreads the infection, leaving lesions in the damaged area of the host.[5]

M. oryzae can infect any exposed aerial rice tissues, as well its roots and xylem tissues.[5] Plant susceptibility depends on various environmental cues and the plant's developmental stage. The use of nitrogen fertilizers greatly increases rice's susceptibility to *M. oryzae* infections as a greater nitrogen content supports the growth of the rice blast for a

longer duration, giving rise to a greater disease severity.[5,11] Likewise, younger, developing plant leaves are more vulnerable to infection than adult plants.[12] Applications of highly nitrogenous fertilizers cause rice to mature slower, creating a greater window of time for *M. oryzae* to infect younger, more susceptible plants.[5,12]

Nitrogen is one of the primary macronutrients for plant growth and it is frequently the limiting factor in growth as plants are only capable of utilizing sources of nitrate or ammonium. Pathogenic fungi on the other hand, have the ability to use a plethora of alternate nitrogen sources when their primary sources of ammonium or glutamine are lacking from their environment. Indeed, a fungus can encounter drastically different environments throughout its life cycle. Its greater flexibility in nitrogen scavenging aids in fungal survival and infection.

The topic of nitrogen starvation in *M. oryzae* and its relation to pathogenicity and the infection process has been the focus of multiple studies.[11,13-21] Indeed, the idea that nitrogen starvation could be an environmental cue for the activation of pathogenic genes began to appear in the literature in the 1990's and has been eloquently summarized in the 2009 review by Donofrio and co-workers.[12] In 1993, Talbot *et al.* demonstrated that the MPG1 gene was highly expressed in the appressorium, as well as during lesion development in response to the stress of nitrogen or carbon starvation.[13] A few years later, Lau *et al.* identified two genes involved in the regulation of starvation-induced gene expression (*NPR1* and *NPR2).*[15] These studies launched a series of investigations probing into gene regulation under starvation conditions. A comparison of nitrogen and carbon-scarce environments were compared at the transcription level, and nitrogen starvation led to the largest expression of gene transcripts when compared to carbon-starved conditions.[12] Moreover, culture filtrates were taken from

nitrogen-limited mycelia and applied to plant leaves. The filtrate induced senescence (plant aging), however this effect was greatly diminished if the filtrate was taken from *M. oryzae* that lacked the nitrogen regulatory gene *NUT1*, which activates nitrogen scavenging pathways.[11,14] As a whole, these experiments have implied that nitrogen starvation both simulates conditions encountered *in planta* and activates genes necessary to achieve an efficacious infection of its host.[12]

Many *M. oryzae* studies have focused on the genetic components of conidiation and development of the appressorium. The question of the current investigation is to determine the effects of nitrogen starvation on the global proteome of mycelia tissues in *M. oryzae*. The remainder of this background describes the technology and techniques employed in a shotgun proteomics used to elucidate the effects of nitrogen starvation on the total proteome of *M. oryzae* mycelia.

1.3 Proteomics Defined

In recent years, mass spectrometry has made unprecedented contributions to fields of biology due to advances in instrumentation, software and bioinformatics. Within a single global proteomic experiment, thousands of proteins can be identified, yielding insightful contributions to disease mechanisms, identification of drug targets, determination of gene functions and environmental effects on a system. Proteomics seeks not only to identify all proteins in a system, but to ascertain their structures, quantities, locations, functions, and to learn how these proteins change over time and in physiological states.[22] This field has blossomed with the sequencing of entire genomes which has enabled rapid, large-scale protein identifications, and now generates more data about a biological system than can currently be analyzed and comprehended.

1.4 Shotgun Proteomics

The advent of soft ionization techniques such as *Matrix Assisted Laser Desorption Ionization* (MALDI) and Electrospray Ionization (ESI), advances in mass spectrometers, and the development of bioinformatics tools has led to proteomics becoming a prolific field. The typical proteomic workflow for bottom-up or shotgun proteomics is summarized in Figure 1.1. First, protein lysates are digested into peptides and are identified using a mass spectrometer via a Data Dependent Acquisition (DDA). Finally, search algorithms are used to assign peptide and protein identifications from the raw data acquired.[23]

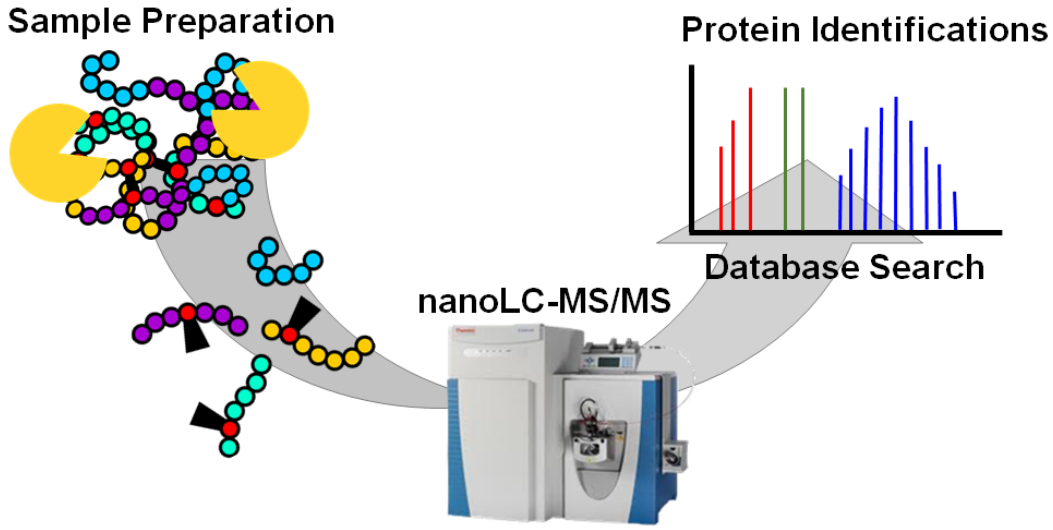


Figure 1.1. An overview of protein identifications in shotgun proteomics. A complex lysate is digested into peptides, and separated by reverse-phase chromatography. In a DDA acquisition, the precursor mass is isolated and fragmented to gain sequencing information. Bioinformatics are used to identify peptides and map them back to protein identifications.

1.5 Current Sample Preparation Methods

The heart of sample preparation for shotgun proteomics centers around the reduction, alkylation and complete digestion of proteins into peptides, and it's summarized in Figure 1.2 Initially, proteins are denatured to remove higher-ordered structures so that protease cleavage sites are accessible during digestion. The denaturation of proteins can be achieved with

various chaotropic reagents, detergents and/or heat. Disulfide bonds are reduced and alkylated to prevent their reformation. A protease which cleaves at specific sequences is typically used to digest proteins into peptides. Trypsin is a commonly employed protease because it specifically cleaves C-terminal to lysine and arginine residues.[24] Due to this cleavage site, tryptic peptides generally maintain at least a 2+ charge state because the N-terminus and basic residue of the cleaved lysine or arginine can each hold a +1 charge state during positive-mode electrospray ionization.

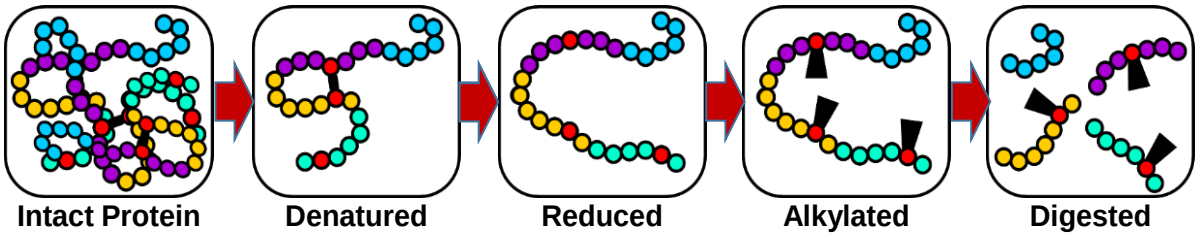


Figure 1.2. Bottom-up proteomics identifies proteins based on peptides. Thus, intact proteins must be denatured to remove secondary, tertiary or quaternary structures, and disulfide bonds reduced and alkylated prior to protein digestion into peptides

Within bottom-up proteomics there are various sample preparation methods; Figure 1.3. summarizes 3 of the most common methods. For a Filter Aided Sample Preparation (FASP), proteins are digested on a filter.[25] This allows detergents and other contaminating small molecules to be washed away with the aid of urea prior to their digestions.[25] An in-gel digestion occurs with the aid of a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) used to separate proteins based on size. Once the gel has run, a lane can be fractionated into multiple pieces, digested and analyzed separately, thereby decreasing the sample complexity.[26] In-solution digestions are a "one-pot" reaction in which the protein is reduced, alkylated and digested in the same vessel.

Recently, an enhanced FASP or eFASP protocol was developed with two key differences from the typical FASP approach.[27] First, the filter was prepassivated with Tween-20 to reduce peptide adsorption to the filter. Secondly, 2M urea was substituted with 0.2% sodium deoxycholic acid during the protein digestion. Enhanced FASP was shown to improve peptide recovery and coverage, increase the number of transmembrane helices identified and reduce peptide losses from filter adsorption.[27]

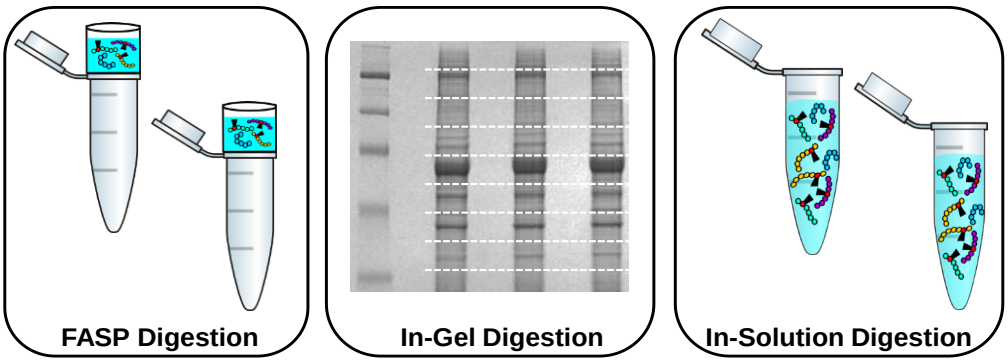


Figure 1.3. Three of the most commonly employed sample preparation techniques used in bottom-up proteomics are highlighted. With FASP, proteins are digested on-filter which has the advantage of removing contaminating salts or detergents. SDS-PAGE gels can be run and the bands excised with all reactions occurring within the polyacrylamide gel matrix. In-solution digestions are "one-pot" reactions where the reduction, alkylation and digestion occur in one vessel.

Surfactants are frequently employed in lysate buffers to aid in protein solubility. Many surfactants such as *SDS*, triton-X 100, or NP-40 that are commonly used for protein extractions are incompatible with down-stream analysis via mass spectrometry. In recent years, companies have produced mass spectrometry-compatible surfactants that can aid in protein solubility and denaturation without negatively impacting the chromatography or mass spectra.[28-31] Water's RapiGest SF and Agilent's PPS Silent Surfactant are acid-labile and can be removed upon acidification and centrifugation.[32] Invitrogen's Invitrosol elutes in three peaks that are distanced from the typical elution times for peptides.[32]

While commercially available surfactants may be effective, they are extremely expensive (several hundred dollars for milligrams of surfactant). Sodium deoxycholate/deoxycholic acid (SDC) is an inexpensive, naturally occurring bile acid detergent that has shown to increase detection of membrane proteins.[33-35] Masuda *et al.* demonstrated that using an in-solution digestion of 1% sodium deoxycholate with 9 µg of membrane-enriched pellet of HeLa cells achieved a higher amount of detected membrane proteins (53% of 1450 identified proteins).[33] Indeed, a recent study revealed that SDC outperformed the commercially available RapiGest and 8M urea with in-solution digestions yielding the greatest sequence coverage and number of protein identifications from mitochondrial protein fractions from rat livers.[36] Likewise, this study also demonstrated that SDC FASP protocols gave the highest recovery, lowest variation and the least bias towards peptide and protein abundance.[36]

1.6 Sample Preparation Complications

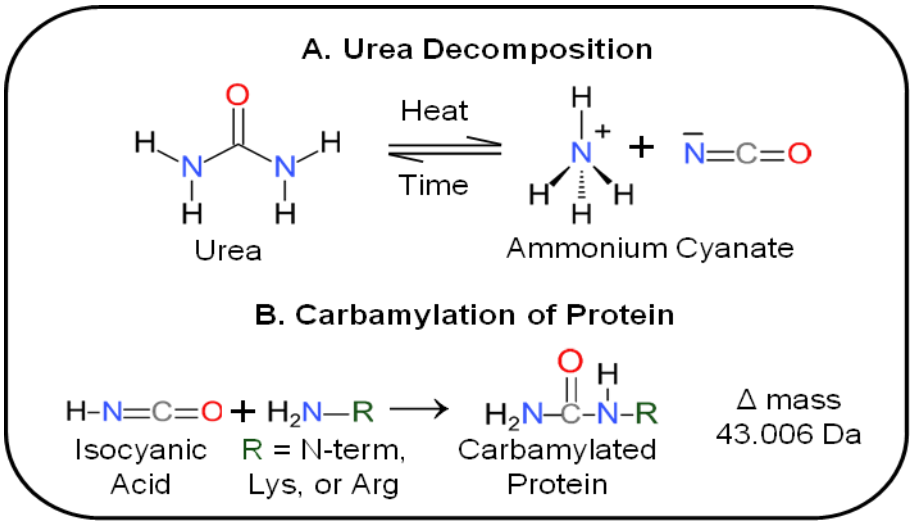


Figure 1.4. Mechanism of protein carbamylation which increases with exposure to heat or in highly concentrated solutions of urea for an extended period of time.

Various reagents frequently employed in a proteomic sample preparation can have adverse effects for peptide identification. Urea is a commonly used chaotropic agent that aids in the solubilization and denaturation of a protein during sample preparation. It is also used during FASP digestions to remove detergents from the sample buffer.[25] However, its use at elevated temperature for extended periods of time causes the carbamylation of proteins and its mechanism is illustrated in Figure 1.4.[37,38] Initially, urea decomposes into ammonium cyanate, and the concentration of ammonium cyanate increases until an equilibrium is reached.[37] Upon protonation, ammonium cyanate becomes isocyanic acid which can react with the N-terminus, lysine or arginine residues, resulting in the carbamylation of proteins with a mass shift of 43 Da.[37] Higher temperatures expedite the conversion of urea into ammonium cyanate.[37]

The chief problem with the carbamylation of proteins/peptides is that it increases the sample complexity as the mass of the peptide has then been shifted, reducing the over-all signal of the identified peptide.[39] Likewise, it can also lead to shifts in retention time during chromatographic separations, reduce the labeling ability of *iTRAQ* tags from N-terminal blockage and the study of *in vivo* carbamylation is also adversely affected.[39]

For the alkylation of proteins, iodoacetamide is a commonly used reagent that reacts with the reduced cysteine residues as illustrated in Figure 1.5.[40] The carbamiodomethyl modification results in a mass shift of 57 Da. However, non-specific or over-alkylation can also occur at the N-terminus, lysine, histidine and tyrosine residues.[41,42] Like carbamylation of proteins, the mass shift from over-alkylation increases the sample complexity and reduces the signal of the identified peptide.

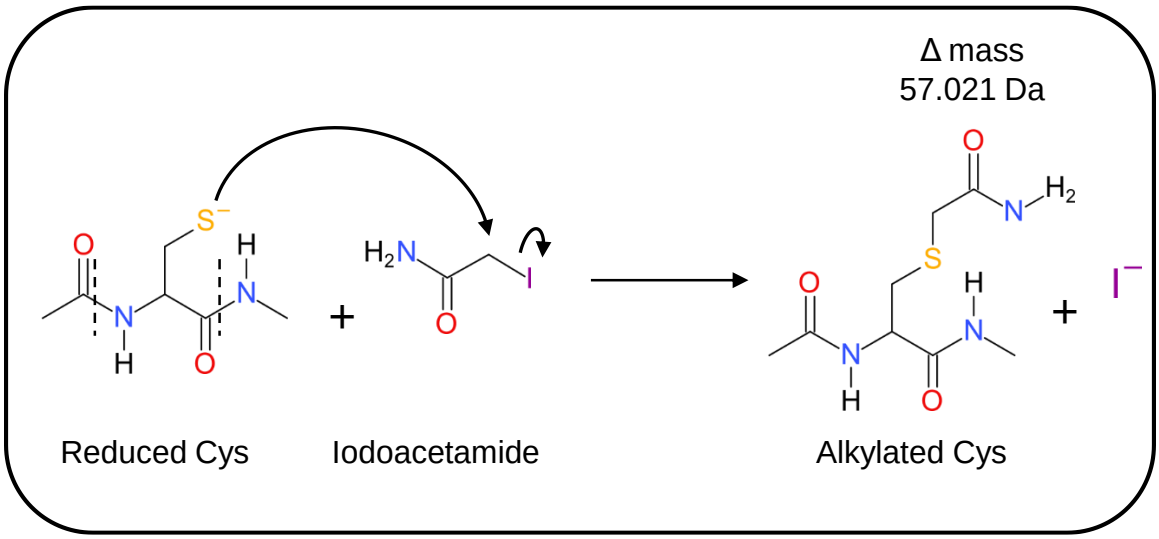


Figure 1.5. Mechanism of protein alkylation with iodoacetamide or chloroacetamide. Over alkylation can occur with all alkylation reagents.

1.7 Electrospray Ionization

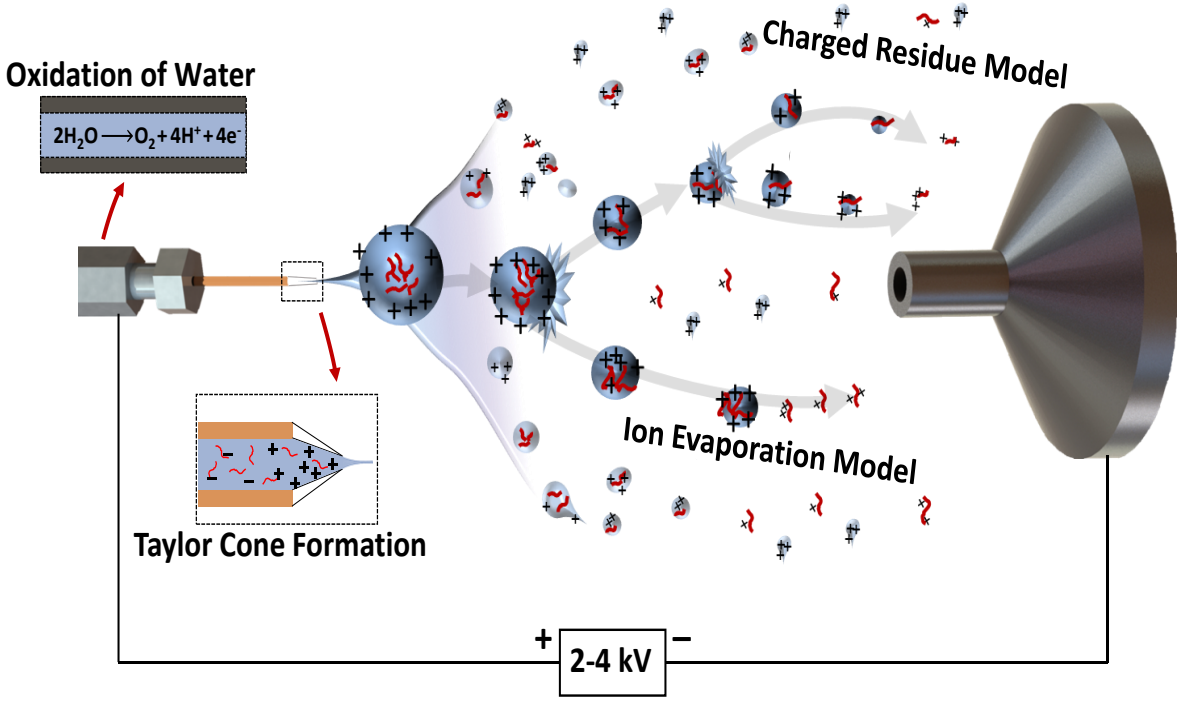


Figure 1.6. With electrospray ionization, peptides are converted into ionic, gas-phase molecules. There are 2 putative ESI mechanisms. In the charged residue model, coulombic fission results in smaller and smaller droplets until the droplets contains a single molecule. The charge from the droplet is imparted on that molecule as the solvent evaporates. On the other hand, the ion evaporation model hypothesizes that droplets at the Rayleigh limit eject molecules into the gas phase, alleviating coulombic repulsion at the droplet's surface.

Electrospray ionization (ESI) is a technique used to convert liquid phase, non-volatile species into charged gas phase molecules as depicted in Figure 1.6.[43,44] The conception of ESI enabled the measurement of large biomolecules such as proteins because it deposits very little internal energy into the molecule, even allowing the ionization of non-covalent interactions.[45] In addition, ESI's ability to impart multiple charges increases the mass range that can be measured by a mass spectrometer as an increase in charge state or *z*, lowers the *m/z* value of a large molecule, brining into the range that can be measured by a mass spectrometer.[46] Principal disadvantages of ESI are low salt tolerances, a large loss of ions upon transfer into a vacuum (common to any atmospheric pressure source), and biases seen from hydrophobicity/hydrophilicity or the molecule's conformation/geometry.[46]

Solvent is flowed through a capillary, and a voltage of 2-5 kV is applied to an emitter tip. The onset voltage is the voltage needed to initiate the ejection of charged droplets. Once the onset voltage has been reached, a Taylor Cone is formed and charged droplets are emitted in a jet.[47] The size of the droplets decrease due to solvent evaporation or through collisions with other gas molecules. As the surface area of the droplet decreases, the charge density increases until the Rayleigh Limit is reached as defined in Equation 1 where q_r represents the charge on the droplet, r is the radius of the droplet, ε_0 is the permittivity of free space and γ is the surface tension.[48]

$$q_r = \sqrt{64\pi^2\varepsilon_0\gamma r^3} \qquad \text{Equation 1}$$

At the Rayleigh limit, the charge density exceeds the surface tension of the droplet which results in "Coulombic explosions" or fission of a parent droplet into smaller daughter droplets. The total surface area of the daughter droplets is greater than the parent droplet, thereby relieving some the Coulombic repulsions on the surface of the parent droplet.[46]

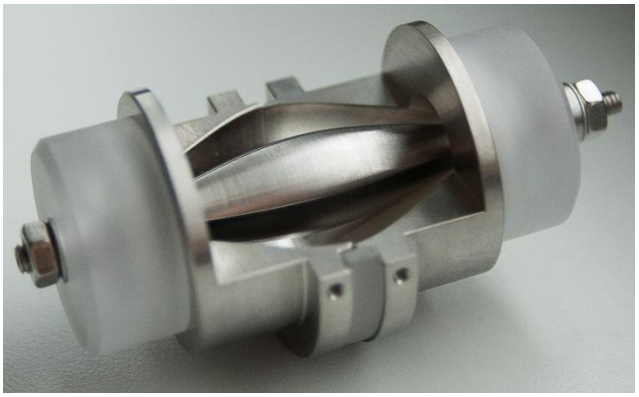
There are two predominant gas-phase ion production modes: Dole's Charged Residue Model (CRM) and Iribarne and Thomson's Ion Evaporation (IEM) model.[49,50] The CRM proposes that droplets continually undergo Coulombic fission forming smaller and smaller droplets until a droplet encapsulates one single molecule; upon solvent evaporation, the remaining charge is imparted on that sole molecule.[50] On the other hand, the IEM hypothesizes that once the Rayleigh limit has been reached, molecules at the surface of the droplet are ejected into the gas phase with the aid of the electric field, removing charge from the droplet and thereby relieving the Coulombic repulsion on the surface of the droplet.[49]

1.8 The Orbitrap: A Novel Mass Analyzer

The idea of employing an orbital trapping of ions was first introduced by Kingdon in 1923. The Kingdon trap simply consists of a central wire, an outer cylinder and two endcaps.[51] Once a voltage was applied between the wire and cylinder, this electric field attracted ions to the wire. Ions that possessed enough tangential velocity to miss the wire then began rotating around it.[51] Thus it was the idea of trapping ions in an electrostatic field from 1923 that gave rise to a novel mass analyzer, the Orbitrap which was initially published by Dr. Alexander Makarov in 2000.[52]

Other ion trapping mass analyzers have been used such as a quadrupole ion trap which uses dynamic electric fields of dc and rf voltages for ion storage and detection.[53,54] Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers also employ the use of a Penning trap, trapping ions within a magnetic field with electric trapping plates, prior to their excitation with resonant cyclotron frequencies.[55,56] However, ion traps suffers from lower mass resolving power and mass accuracy, and FT ICR mass spectrometers are costly and complex. Both mass spectrometers suffer from space charge effects.[52] The Orbitrap was

conceived as a mass analyzer with the advantages of high mass resolving power, mass accuracy, sensitivity, and speed that is significantly less expensive and easier to operate than a FT ICR.[52] It consists of a spindle-like central electrode and two electrically separated outer cup-shaped electrodes as illustrated in Figure 1.7.



$$Equation\ 2 \quad \omega_{\varphi} = \sqrt{\frac{\left(\frac{R_m}{R}\right)^2 - 1}{2}}$$

$$Equation\ 3 \quad \omega_{r} = \sqrt{\left(\frac{R_m}{R}\right)^2 - 2}$$

$$Equation\ 4 \quad \omega_{z} = \sqrt{\frac{e}{(m/z)} \cdot k}$$

Figure 1.7. A partial cut-out view of the Orbitrap with its three key frequencies of motion defined. The axial frequency is the only frequency that is independent of its initial velocity or spatial spread, so it the only frequency suitable to record its image current and convert its frequency into *m/z* values via an enhanced Fourier Transfer. Image taken from Wikimedia commons.

The Orbitrap is a pulsed mass analyzer, and it was a challenge to couple it with a continuous ionization source like ESI.[57] In order to interface the Orbitrap with ESI, ions are stored in the curved linear ion trap or C-trap prior to their pulsed injection into the Orbitrap.[57-59] The C-trap is a set of RF-only rods filled with a nitrogen bath-gas.[57,58] Once ions are injected into the C-trap, they lose their kinetic energy from collisions with the nitrogen gas and are gathered within a small axial potential well within the C-trap.[57,58] This axial potential well is generated by placing a ring over the end of the storage rods which is biased with a DC offset.[57] After accumulating sufficient ions, they are extracted by the application of a negative voltage pulse (for positive ions) through the exit lens.[57,58] Ions are

accelerated and spatially focused through deflection lenses in a Z-like pattern, delivering ion packets to the deflector at the entrance of the Orbitrap.[58] The deflector is a machined slot in the Orbitrap with a compensation electrode to allow for injections of ions with minimal losses as well as maintaining a stable voltage during ion detection to compensate for the electric field imperfections from the injection slot.[59]

Ions are injected into the Orbitrap offset from its equator and begin coherent axial oscillations without any additional excitation.[58] Ions of each mass to charge ratio are injected as short packets a few millimeters long which corresponds to a spread of flight times of a few hundred nanoseconds for mass to charge ratios of a few hundred Daltons per charge; thus ions are quickly delivered to the Orbitrap within a narrow temporal and spatial distribution.[57-59]

Ions are captured by electrodynamic squeezing: as ions are injected, the voltage is steadily decreased on the central electrode (while outer electrodes are maintained at a fixed potential), thereby increasing the field strength and causing the ions to rotate in a smaller radius around the central spindle.[57,58] The voltage of the central electrode continues to decrease to allow for newly arriving ions (typically with a higher *m/z*) to enter the Orbitrap without colliding with the outer electrodes.[57] Once all ions have been injected into the Orbitrap and are equidistant from the central and outer electrodes, the voltage on the central electrode is stabilized for image current detection.[57]

The ion motions within the Orbitrap can be described in three cyclic motions: rotational, radial and axial which are defined in Figure 1.7. The rotational motion or angular frequency (ω_φ) around the central electrode is defined by equation 2.[52,58,59] The radial motion with a frequency of ω_r is shown in equation 3 and the axial oscillations along the central

electrode composed of ω_z is demonstrated in equation 4.[52,58,59] In equations 2 and 3, R is the initial radius and R_m refers to the characteristic radius. For equation 4, k is a constant related to the field curvature and e is the elementary charge of an electron (1.602×10^{-19} C). The axial frequency is entirely independent of any initial velocity (R) and position/spatial spread which is why it is the only frequency used to determine the *m/z* value of the ions present.[52,58,59] The two outer electrodes detect an image current of the axial rotation of the ion packets in a time domain.[59] This complex image current is decomposed into individual sinusoidal waves and converted from the axial frequency into the *m/z* of a given ion using an enhanced Fourier transform (eFT) algorithm which is depicted in Figure 1.8.[57,60]

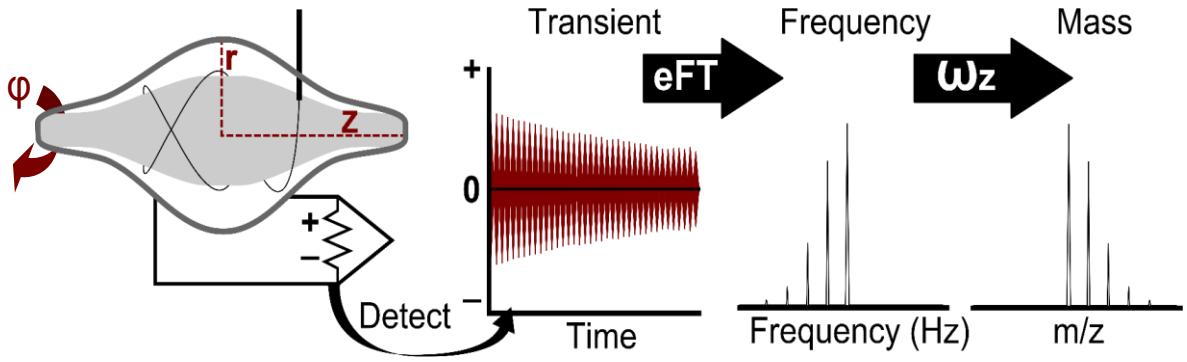


Figure 1.8. The Orbitrap collects an image current based the axial frequency of ion motion. An eFT transforms individual sine waves frequencies which can be converted into a *m/z* value. *m/z* value.

1.9 Q-Exactive HF Mass Spectrometer

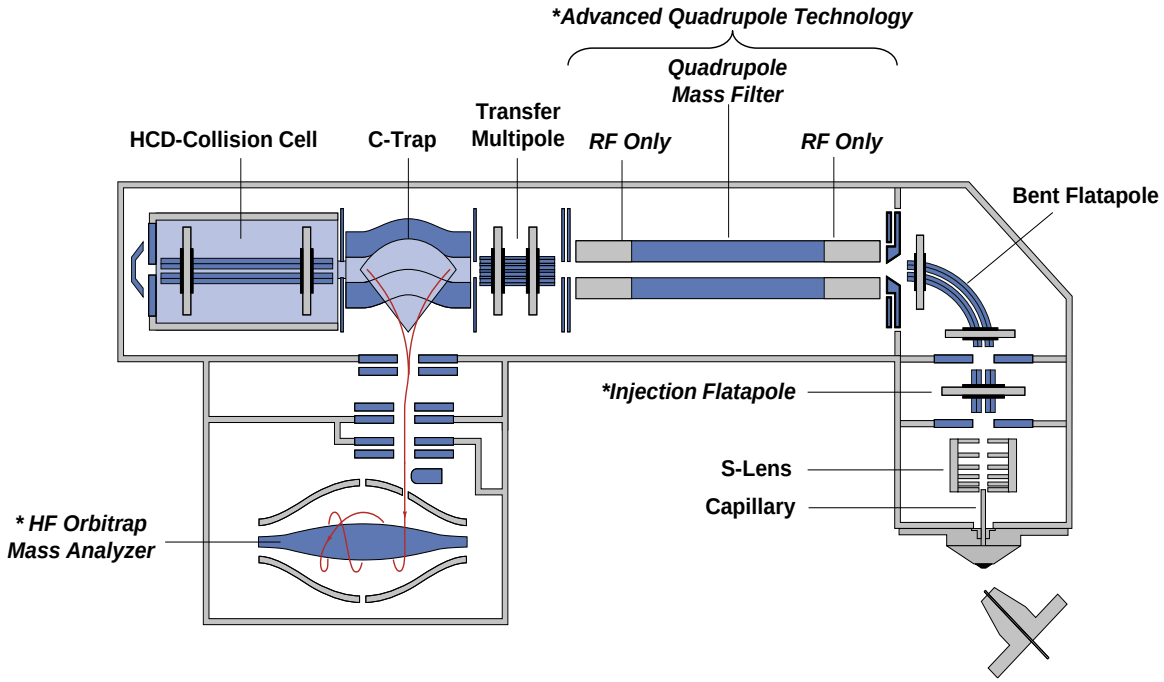


Figure 1.9. A schematic of the Q-Exactive High Field mass spectrometer. *Denotes improvements in performance compared to the original Q-Exactive mass spectrometer.

The Q-Exactive mass spectrometer is a hybrid quadrupole-Orbitrap instrument which has rapidly gained popularity among the proteomics field due to its fast speed of acquisition, high mass resolving power, large dynamic range, high mass accuracy, is simple to operate and possess a relatively small footprint.[60] As shown in Figure 1.9, ions enter through a heated capillary where they are focused through the stacked-ring ion guide or S-lens. The S-lens is composed of a series of ring electrodes with opposite RF voltages applied to alternating rings to generate a RF electric field that focuses the ion beam through the S-lens.[61] The diameters of the rings are larger at the entrance to capture as many ions as possible and they become concentrically smaller to focus the ion beam through the exit lens.

The beam then passes through an additional series of lenses and a bent flatapole. By forcing the ions to take a 90° turn, this removes any unwanted neutral species that cannot complete the bend, and the ion beam is then focused into a quadrupole.[60,62] The quadruple can be set to allow all ions to pass or select for a specific *m/z* range by applying various DC and RF voltages, and it has the ability to change its ion selectivity near instantly.[60] Ions are transferred from the quadrupole into the C-Trap where they are collisionally cooled by nitrogen bath gas. Once collected in the C-trap, the ions and can be injected into the Higher energy Collisional Dissociation cell (HCD) cell for fragmentation or into the Orbitrap for detection.[60] The HCD cell is a multipole filled with nitrogen as the collision gas.[60] Parent ions from the C-trap are accelerated by a voltage offset into the HCD cell where they are fragmented.[60] The fragmentation patterns obtained are comparable to those obtained from a triple-quadrupole instrument. The ions are then transferred back into the C-trap prior to their injection and measurement within the Orbitrap.

There were several improvements to the instrumentation of the Q-Exactive high field compared to its initial generation. First, the injection flatapole acts as a low-resolving power filter to remove unwanted ions before they can penetrate deeper within the instrument, increasing the robustness of the instrument.[62] Secondly, a segmented quadrupole mass filter was implemented with improved isolation efficiencies that resulted in nearly a 2-fold improvement in ion transmissions.[62]

Finally, the high field (HF) Orbitrap possess a higher electric field strength and frequency of ion motion. As a result, the resolution is doubled for the same transient times or the scan speed is doubled if the resolution is kept the same.[62,63] Depending on the

instrumental settings, it is possible to achieve a top 20 data dependent acquisition in approximately one second due to the fast speed of the Q-Exactive HF.

1.10 Data Dependent Acquisition and Protein Identifications

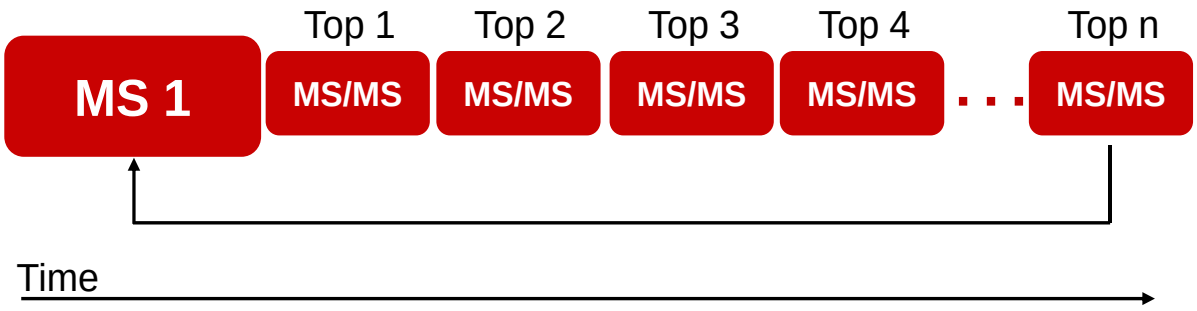


Figure 1.10. A summary of the DDA process. A MS 1 precursor scan surveys all m/z present and then selects abundant parent ions for MS/MS fragmentation to gain sequential information. DDAs are performed in a top N fashion, where N represents the number of MS/MS cycles following the survey scan that can be performed.

Data dependent acquisitions (DDA) has proven to be a powerful technique, routinely identifying thousands of proteins in shotgun proteomic experiments. First, peptides are separated chromatographically, typically on a reverse phase column. As they elute and enter the mass spectrometer, a full scan (MS1) of all ions and their isotopic distributions is collected.[62] In real time, the mass spectrometer selects which precursor mass is isolated and fragmented. The most abundant parent/precursor ions that have not yet been sequenced are selected for isolation and fragmentation.[62] In the HF Q-Exactive, parent ions are isolated by the quadrupole, sent to the HCD cell for fragmentation (MS/MS scan) and then analyzed by the HF-Orbitrap.[62]

While DDA has proven to be a highly efficacious technique, it is not without its limitations.[64-66] Because decisions for which ions are fragmented are made on-the-fly, DDA is stochastic in nature, and it may suffer in its reproducibility of identifying peptides.[64-66]

Moreover, DDA is biased towards identifying high-abundant species, thus low-abundant peptides present in a complex biological sample are less likely to be sampled.[64-66] Likewise, the number of sequencing events is limited by the instrument's scanning speed, which leads to an undersampling of the proteome.[64] Also, the co-isolation and fragmentation of 2 or more ions (chimeric species) is an additional limitation of DDA.[64]

The selection of parent ions for isolation and fragmentation is based on various experimental parameters such as: intensity threshold, charge state, isotopic distributions, and dynamic exclusion. The intensity threshold is the minimum intensity necessary for a precursor ion in the MS1 scan to be selected for MS/MS fragmentation and it's defined by Equation 5.

$$Intensity\ Threshold = \frac{AGC\ target\ value\ x\ Underfill\ Ratio}{Maximum\ Injection\ Time\ (msec)} \qquad \text{Equation 5}$$

The AGC target value is the amount of ions to be collected by the C-trap prior to injection into the Orbitrap. The maximum injection time is the longest amount of time that the C-trap is kept open to attempt to collect enough ions to reach the AGC target.[67] The underfill ratio is a minimum percentage of the AGC target value, and it's selected by the user to calculate the intensity threshold. Once a parent ion has triggered the intensity threshold and is selected for fragmentation, it is accelerated into the HCD cell, which fragments the parent ion into daughter ions as depicted by Figure 1.11. The masses of the fragmentation pattern can then yield sequential information of a peptide.

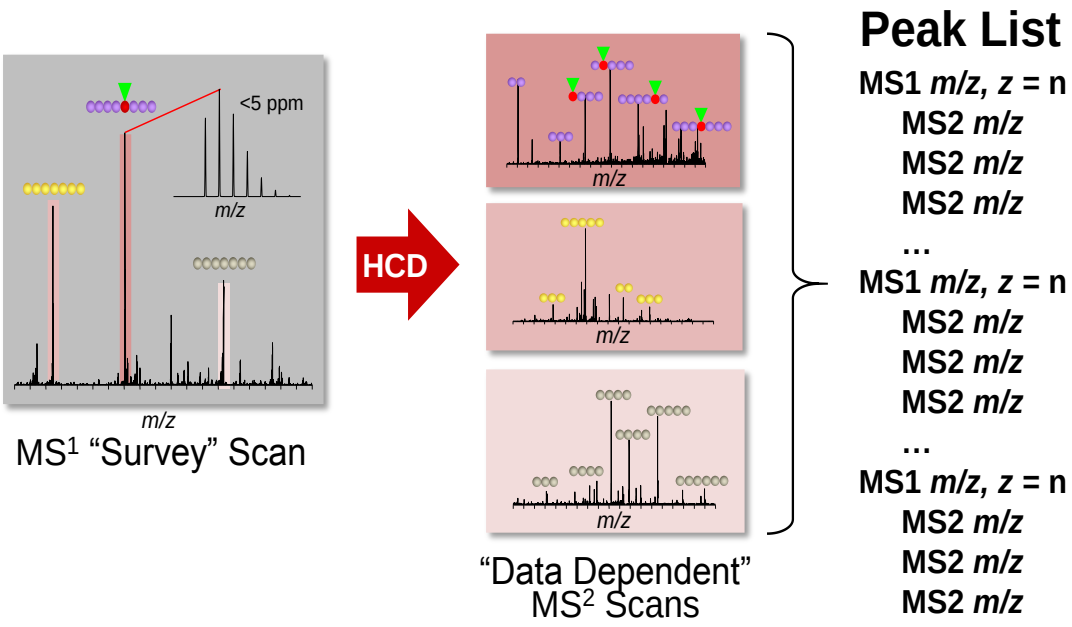


Figure 1.11. Example of the real-time decision making by the mass spectrometer during a DDA. The three most abundant peptide masses were selected for fragmentation which had not previously been sequenced. These masses are then transferred to a dynamic exclusion list so that they are not sequenced again, allowing new peptides to be selected for identification.

Charge state is an additional parameter to determine if an ion should be selected for sequencing. In proteomics, a charge of +1 or unassigned is generally excluded from fragmentation because a +1 state does not fragment as well as a higher charge state. Moreover, a +1 may be indicative of a contaminating polymer, or simply an ambient ion present from the environment. The 'peptide match' setting is used to prioritize ions with peptide-like isotopic distributions for fragmentation.[67] Likewise, the "exclude isotopes" setting is used to isolate and fragment the monoisotopic peak, thereby excluding the isolation and fragmentation of additional peaks in the same isotopic distribution. The dynamic exclusion setting is used so that once a peptide has triggered a MS/MS scan, it will not be sequenced again for the duration of time set by the user, ensuring that the instrument does not spend time re-sequencing the same peptides.[67]

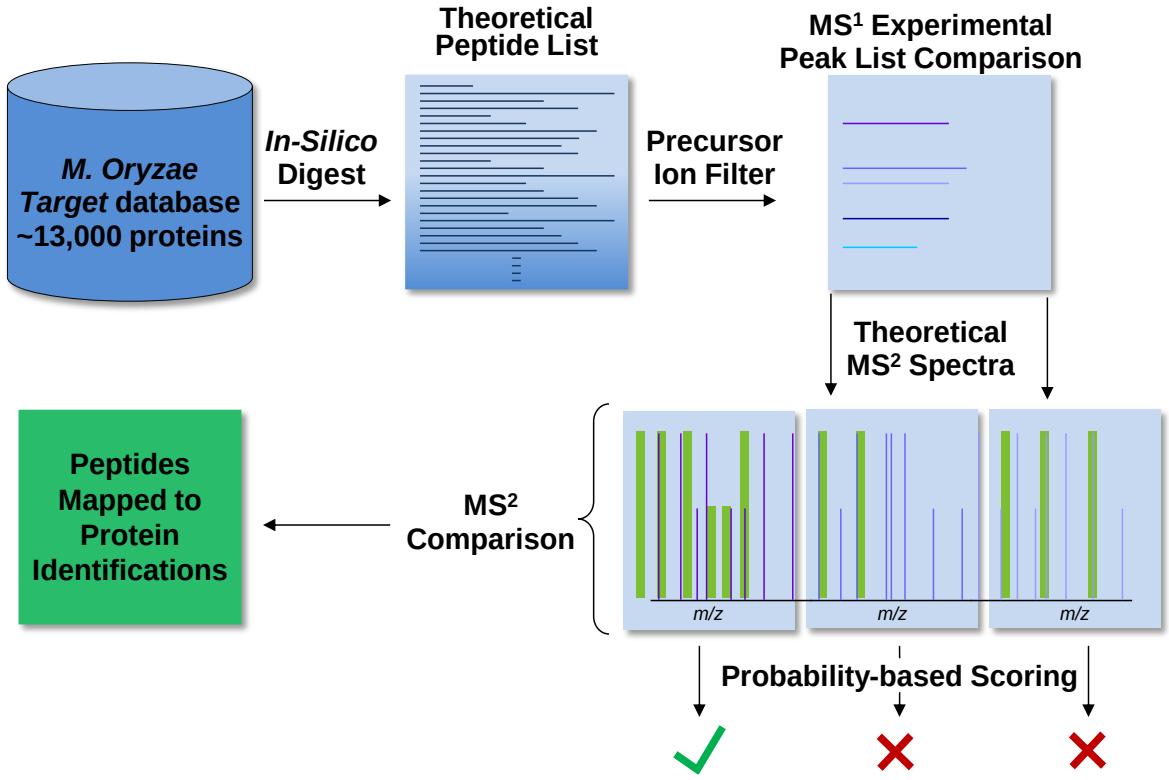


Figure 1.12. Workflow of bioinformatics employed to assign peptide and protein identifications. MS1 spectra are compared to theoretical masses generated from an *In-silico* digest of proteins in the data base. MS2 sequencing information is used to assign the peptide that most likely matches the MS1 mass. Identified peptides are then mapped back to proteins listed in the database.

Bioinformatics tools are used to assign peptide and protein identifications based on the precursor and MS/MS scans of peptides recorded during the DDA and their use is depicted above in Figure 1.12.[23,68-70] A protein database is provided and an *in silico* digest of the proteins is performed to generate the theoretical mass based on the protease used in digestion, mass tolerances of the instrument and any protein modifications.[68] The theoretical precursor mass is used to filter the experimental MS1 scans and the sequencing information from the MS/MS scan is used to find the best match to the precursor.[68,69] Various search algorithms are used for peptide identifications such as Sequest and Mascot. Sequest

functions based on a cross-correlation method in which identifications are based on how well the theoretical mass spectra correlate or match to the experimental mass spectra.[68,71] Mascot employs a probability-based assignment in which a statistical significance is calculated for the match between the theoretical and experimental spectra.[68,72] As with the instrumentation, several drawbacks occur within the bioinformatics aspect for DDAs. Dynamic modifications are modifications that may or may not be present, and the greater the number of dynamic modifications in the search parameters, the greater the search space and time.[73] Moreover, identifications are only made based on the specified search parameters which can bias novel discovery-based experiments.

In order for an identification to be assigned, the peptide identification must also undergo statistical validation, where the application of a false discovery rate (FDR) is frequently employed. One method a FDR is determined is by utilizing a search against a reverse database, generated by the reverse of the protein sequences in the database.[69] Any assignment to the reverse sequence would be incorrect, and this number is used to estimate the number of false positive identifications.[69] Similarly, the amino acid composition of the protein sequences in the data base can be randomized and also be used to determine false protein identification. A FDR threshold can then be applied to the dataset (typically at 1%). After validation, the peptide identification is then mapped back to a protein containing that sequence in the database. Figure 1.13 illustrates a peptide sequenced from a digested lysate of HeLa cells that was one of 17 peptides used to identify the α-enolase protein in a DDA. The greater the number of unique peptides and sequence coverage used to identify the protein, the greater the confidence in its identification.

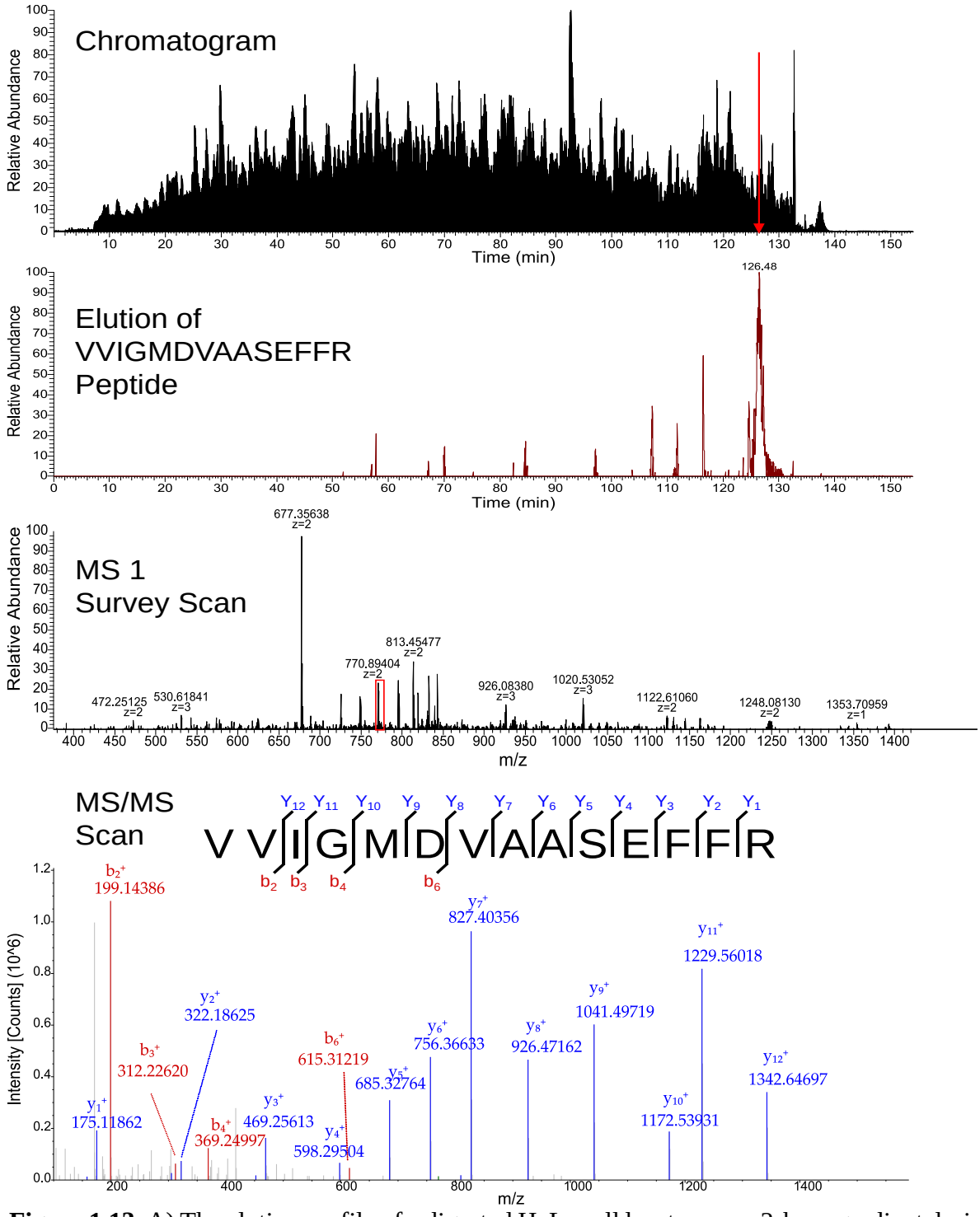


Figure 1.13. A) The elution profile of a digested HeLa cell lysate over a 2-hour gradient during a DDA. B) The elution of a peptide that belongs to α-Enolase. Note that peaks with the same *m/z* but were of a low abundance did not trigger the intensity threshold for their selection for sequencing. C) MS2 survey scan in which VVIGMDVAASEFFR peptide was selected for sequencing (it's highlighted in red). D) MS/MS sequencing information gained from the peptide's fragmentation pattern.

1.11 Relative Quantification

Relative quantification of bottom-up proteomics can be performed in a labeled or label-free manner. Chemical labeling such as Isobaric Tags for Relative and Absolute Quantification (iTRAQ), and Tandem Mass Tags (TMT) can be used to label samples and have the advantage of multiplexing (mixing multiple samples together).[74] However, these are very costly methods that may suffer from reproducibility issues during the derivatization process. Also, more MS/MS time is needed for iTRAQ because there is no peptide identification or quantification without its fragmentation.

The Stable Isotope Labeling by Amino Acids in Cell Culture or (SILAC) approach metabolically labels proteins with stably labeled isotopes that are supplemented in the culture media.[75] This sample is mixed together with an unlabeled sample, producing analytes that are chemically identical, and chromatographically coelute, but can be separated by the mass spectrometer due to the shift in mass from the "heavy" labeled amino acids.[75] This approach gained widespread use to due minimizing analytical variation, but it's limited to samples that can achieve an incorporation rate of the labeled isotope of at least 95%.[75] Also, the sample complexity is increased which results in fewer peptide identifications unless extensive fractionation is employed.

Label free quantification has gained popularity due to its simpler sample preparation and may be used with any type of sample.[76] Moreover, the reduced sample complexity enables a greater number of peptides to be identified, resulting in deeper proteome coverage with a larger dynamic range. However, because multiple experimental conditions cannot be combined into the same analytical run, label free quantification may have a greater variability associated with its measurement.[77] Label free quantification can be done using the area

under the curve of a peptide's MS1 scan, or with peptide spectral counts (the number of MS/MS spectra that identify a peptide).[77,78] It has been shown that these two techniques are comparable, with peak area yielding slightly greater accuracy, and spectral counting being more sensitive to changes in protein abundance.[78]

Various normalization strategies can be applied to label-free data to account for variations arising from sample preparation, chromatography and instrumentation, thereby reducing the variability between samples and replicates[76]. Total spectral count (TSpC) normalization is summarized in equation 6, where replicates or conditions with lower spectral counts are normalized to the replicate with the highest TSpC.[76] This approach assumes that the total number of spectral counts should be conserved across samples and is useful when comparing multiple conditions across various technical and biological replicates.[76]

$$TSpC = \frac{\Sigma\ Highest\ Total\ SpC}{\Sigma\ Total\ SpC\ Replicate\ N} \qquad \text{Equation 6}$$

The Normalized Spectral Abundance Factor *(NSAF*) accounts for the fact that longer proteins tend to yield a higher number of spectral counts compared to shorter proteins, and it accounts for this bias by dividing the spectral counts of a protein by its length, yielding the Spectral Abundance Factor (SAF). The SAF is the normalized to the total SAF to adjust for variations between runs and this calculates the NSAF as highlighted in Equation 7.[76,79] The NSAF approach assumes that the sum of all SAFs should be maintained between replicates in order to correct for differences in sampling rates.[76]

$$(NSAF)_x = \frac{(\frac{SpC}{L})_x}{\Sigma_{i=1}^{N} (\frac{SpC}{L})_i} \qquad \text{Equation 7}$$

Another approach that has been used is normalization to selected proteins (NSP). In this case it is assumed that the total SpCs for a standard protein should be conserved between replicates if it is found in equal concentrations, and any changes in spectral count of the standard protein should be due to variation between replicates or samples.[76] Changes in the standard protein are used to calculate a correction factor for all proteins. In this case, the standard protein may be an endogenous "house-keeping" protein or exogenously spiked into the sample.

When these three normalization methods were systematically compared, *NSP* did not effectively correct for variations within the data.[76] NSAF and TSpC both resulted in low variation across all data sets and it was also illustrated that accurate quantification was dependent upon the number of spectral counts. The lower the spectral counts, the greater the variation observed, thus it was suggested to apply less stringent fold-change cutoffs to proteins with higher spectral counts, and lower spectral count proteins receive higher fold-change cutoffs and more stringent significant testing.[76]

1.12 Conclusion

Infections by *M. oryzae* cause the greatest amount of crop loss in rice. By studying this filamentous fungi, we hope to make a contribution to combatting food shortages and world hunger, particularly in the Asian population which is heavily dependent upon rice as a staple food. By employing shotgun proteomics, this study investigates changes induced in the global proteome of mycelia in *M. oryzae* under nitrogen starvation conditions. Nitrogen-limited environments are frequently encountered *in planta* during *M. oryzae*'s infection process, thus this study seeks to ascertain novel insight into this fungus's pathogenicity and infection process.

CHAPTER 2. GLOBAL PROTEOMICS OF NITROGEN STARVATION IN *M. ORZYAE*

2.1 Introduction

The question of how nitrogen starvation affects *M. oryzae* has continued to be studied since the generation of its interest in the early 1990s. A Woronin body is a peroxisome-derived vesicle uniquely found in various filamentous fungi and its role is to close pores after cellular damage has occurred.[16] In 2003 Sundararajan *et al.* demonstrated that mutant *M. oryzae* lacking the hexagonal peroxisome protein that is necessary for the formation of Woronin bodies was unable to survive in nitrogen-starved environment.[16] This result suggested that Woronin bodies could play a role in fungal defense in nitrogen-limited environments encountered *in planta* in addition to their roles in pathogenicity.[16] The Dean lab performed a global transcriptome study of *M. oryzae* of the mycelia proteome under nitrogen starvation conditions.[17] Insight into global gene expression patterns, pathogenicity and the regulation of the nitrogen catabolic repression system was gained from this investigation. A 2008 review was the subjection nitrogen metabolism and regulation in plant pathogenic by Bolton and Thomma.[80]

In 2011, the secretome of nitrogen-limited *M. oryzae* was surveyed by two-dimension gel electrophoresis coupled to a mass spectrometry analysis, and identified 89 differentially expressed proteins from various gel spots.[18] Seven of randomly selected secreted proteins were probed with reverse transcriptase polymerase chain reaction and the genetic approach showed a good correlation between expressed protein and RNA levels.[18]

The Donofrio lab probed the effects of nitrogen and carbon starvation, oxidative stress and minimal media on small RNAs (sRNAs) which included small interfering and microRNAs.[19]

This was the first characterization of sRNAs in *M. oryzae* and they demonstrated that retrotransposons were induced by nitrogen starvation , potentially implicating their suppression through an epigenetic mechanism.[19] Overall, the stressors led to changes in the sRNA expression, and they proposed that sRNAs are transcriptional regulators.[19]

Quite recently, a proteomic investigation of *M. oryzae* mycelium during nitrogen starvation by Zhou and co-workers. Fungal cultures were grown for 3 days, mycelia harvested and cultured for another 3 days in nitrogen-starved or nitrogen-treated completed media. The samples were prepared using two-dimensional electrophoresis and spots were analyzed with a MALD-Time-of-flight (TOF)-TOF mass spectrometer which identified 975 protein spots from completed medium and 1,160 in nitrogen starved. Forty-nine protein spots showed at least a 2-fold up or down regulation and pathways involved in glycolysis, nitrogen metabolism and tricarboxylic acid were identified.[21]

While a study of mycelium from *M. oryzae* has recently been published, it was severely limited by antiquated techniques and instrumentation. First, protein spots were identified, which does not give an exact protein number or identification. Also, the fold-change was determined using differences in spot staining instead of using a label-free quantification approach. The maximum proteins identified was 1,160 protein spots which represents merely 20% of the 5,498 proteins identified in this study. Moreover, the Orbitrap technology possess far superior mass accuracy and mass resolving power than a TOF-TOF instrument.[62,81] The lower resolving power of the TOF-TOF results in less spectral accuracy and mass measurement accuracy when compared to the Q-Exactive HF's capabilities.[81-83] As a result, the proteomic study by Zhou and co-workers likely has less accurate results due to its dated methods and instrument.[81,82] The study we have undertaken regarding the global

effects of nitrogen starvation on the mycelia proteome resulted in far greater depth of proteome coverage and biological insights that this recent study, making a higher quality contribution to the study of *M. oryzae.*

2.2 Materials and Methods

Materials

HPLC grade water, and acetonitrile were procured from Burdick and Jackson (Muskegon, MI, USA). Formic acid, dithioreitol (DTT), iodoacetamide, urea, tris, sodium chloride, ethylenediaminetetraacetic acid (EDTA), phenylmethylsulfonyl fluoride (PMSF), calcium chloride, and TPCK treated porcine trypsin were purchased from Sigma Aldrich (St. Louis, MO, USA). Zwittergent 3-16 and 10 kDa concentration filters were purchased from EMD Millipore (Darmstadt, Germany). A protease inhibitor cocktail was purchased from Roche Diagnostics (Indianapolis, IN, USA). The analytical column was packed in-house with Kinetex C18 2.6 μm particles (Phenomenex, Torrance, CA, USA), and consisted of a 75 μm x 20 cm PicoFrit column (New Objective, Woburn, MA) packed.

Magnaporthe oryzae was grown in complete media culture for three days (strain 70-15). Mycelia were collected, washed, and inoculated into minima media with or without nitrogen sources and incubated for 12 hours. Samples were pelleted, flash-frozen with liquid nitrogen and lysed by grinding with mortar and pestle. Proteins were resuspended in 8 M urea, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA (pH 8.0), Roche protease cocktail inhibitor and 50 μM of PMSF, which were provided by the Dean lab. Three biological replicates were prepared and analyzed per condition.

To begin the FASP protein digestion, 250 μg of lysate were denatured in 50 mM DTT at 56°C for 30 minutes. Cysteine residues were alkylated with 200 mM iodoacetamide at

37°C for 60 minutes. Lysates were transferred to a 10 kDa molecular weight cutoff filter and concentrated by centrifugation at 14,000 g for 15 minutes at 20°C. All subsequent centrifugation steps were performed this way. Buffer exchange with 2 M Urea and 10 mM calcium chloride was done by applying excess buffer to the filters and concentrating the sample via centrifugation, ensuring the filter was properly conditioned for a trypsin digest. Samples were digested with a 1:50 ratio of enzyme: protein (w/w) and incubated overnight at 37°C. The digestion was quenched with 1% formic acid (v/v) and 0.001% Zwittergent 3-16, and peptides were eluted by centrifugation, frozen, and evaporated under vacuum.

Instrumental Settings

Peptides were reconstituted in Mobile Phase A (MPA, 98% water, 2% acetonitrile, 0.2% formic acid) to a concentration of 200 ng/µL based on an A_{280} reading using Thermo Fisher's Nanodrop spectrophotometer. Reverse-phase nano-LC was performed utilizing a Thermo Easy nano-LC 1000 (Thermo Fisher Scientific, Waltham, MA, USA). One microgram of digest was loaded onto an analytical column with a maximum pressure of 500 bar with a direct injection column configuration. Peptides were eluted over a 240 minute gradient from 5% to 30% Mobile Phase B (MPB, 98% acetonitrile, 2% water, 0.2% formic acid) at a flow rate of 300 nL/minute. Two blanks were run between each sample to eliminate carry-over between injections. Three technical replicates were injected for each biological replicate.

Data was collected on Thermo Fisher Scientific's Q-Exactive High Field Mass spectrometer with the electrospray voltage set to 2 kV in positive mode, a capillary temperature of 300°C, the S-Lens voltage at 55 V, and a scan range of 300-1600 *m/z*. A top 20 data dependent acquisition was performed using the following settings during a MS scan:

resolving power 120,000, AGC target of 3E6 and 50 millisecond injection time. For MS/MS scans parameters were: resolving power 15,000, AGC target 1E5, 30 millisecond maximum injection time, isolation window 2 *m/z*, and a scan range 200-2,000 *m/z*. The remaining data dependent settings included a normalized collision energy of 27, underfill ratio 2.0% (resulting in an intensity threshold of 6.7E4), a dynamic exclusion of 20 seconds, a charge exclusion of unassigned or 1, and the use of peptide match set to "preferred".

Data Analysis

Raw files were searched using Sequest HT and statistically validated using percolator in proteome discoverer (version 1.4) and a 1% false discovery rate was applied to peptides. The precursor mass range was 350-5,000 Da with a minimum S/N threshold of 1.5 and minimum peptide length of 5 amino acids. Data was searched using a concatenated target-reverse data base encompassing 12,991 *M. oryzae* proteins (genome version 8, Broad Institute of Harvard and MIT) and contaminating human proteins (i.e. keratin). The MS1 precursor mass tolerance was ± 5 ppm, and the fragment tolerance ± 0.02 Da. The enzyme was trypsin, and 4 missed cleaves were allowed to account for a possible higher rate of missed cleavages with ubiquitinated peptides. Dynamic modifications included deamidation (N, Q), oxidation (M), ubiquitination tag of GlyGly (K), and an additional ubiquitination marker of LeuArgGlyGly (K). The maximum amount of dynamic modifications per peptide was limited to 4. Carbamiodomethyl (C) was the only static modification. A label-free quantification was performed using spectral counts, and total spectral count normalization was applied to the technical replicates, biological replicates and across the two conditions of nitrogen starved vs. nitrogen treated.[76] Raw files were also evaluated using RawMeat

produced by Vast Scientific in collaboration with Thermo Fisher Scientific (Version 2.0b1007).

The differential protein expression was calculated by dividing the spectral counts of the nitrogen starved condition by the nitrogen treated condition (the normalized spectral count values across treatment conditions was used). The statistical significance of the protein fold-change was assessed using a pairwise Student's *t* test, and a strict cutoff of $p < 0.05$ was applied. For proteins with a significant fold-change going from a low spectral count number (i.e. <5) to a higher number of spectral counts, the upper number of spectral counts was filtered at a minimum of 15 spectral counts (averages to 5 spectral counts per biological replicate). This was done in order to minimize the chance of a false fold-change identification due to a lower number of spectral counts and to ensure a more accurate quantification.[77,84] If a condition had no spectral counts, it was assigned 1 spectral count for all biological replicates to enable the fold-change calculation.[85]

Proteins with a significant (p value < 0.05) and a 2-fold (or more) change in expression, or that were unique to one condition were used for protein enrichment analysis by the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics with its default settings (version 6.7).[86] The background for the DAVID analysis was comprised of the entire *M. oryzae* proteome. DAVID determines enrichment scores based on a modified Fisher exact probability test deemed the Expression Analysis Systematic Explorer or EASE score.[86,87] EASE is a more conservative measure that removes one element within a cluster and then calculates the Fisher exact probability, thereby removing false positives due to low number of genes/proteins in the cluster and favoring clusters with greater robustness.[87] Once enrichment clusters had been calculated, clusters with an enrichment score above 1.1

were analyzed for biological interpretation with a strict protein cutoff value of $p<0.05$. Moreover, the proteomic results were compared to a previous microarray study of the transcriptome to integrate a more comprehensive systems biology analysis.

2.3 Results and Discussion

In order to draw meaningful biological conclusions, fundamental parameters of the measurement must be recorded with the highest possible quality to ensure low technical variation and high reproducibility. The percent relative standard deviation (RSD), also known as the coefficient of variation, was calculated according to Equation 8 as it's widely used to assess the precision of measurements. A percent RSD value of 10% or less is typically desired. Thus, the initial portion of this analysis focuses on basic but crucial components of data to ensure that accurate biological observations can be made.

$$Percent\ RSD = \frac{\sigma}{\mu}\ x\ 100 = \frac{standard\ deviation}{mean}\ x\ 100 \qquad \text{Equation 8}$$

Technical Analysis

First, basic parameters of the data was analyzed per injection (n = 18) to demonstrate little variation occurred from injection to injection as illustrated in Figure 2.2. The average number of MS1 scans was 23,648 (0.5% RSD) and the average number of MS/MS scans was 152,900 (0.4% RSD). The average number of protein identifications was 3,638 (1.1% RSD) and the total spectral counts per injection was averaged at 55,792 (2.5% RSD). The average intensity per injection was 1.87 E8 (7.2% RSD), and the highest spectral count of a single protein per injection was 741 (9.8% RSD).

The percent RSD for each parameter is visualized in Figure 2.1 and it clearly demonstrates excellent precision with percent RSDs below 2% for the number of MS1 scans, MS2 scans, and number of protein identifications across all 18 injections. The precision for

total spectral counts is still quite good at 5% and within the expected amount of variability for spectral counting. The increased variability in MS1 intensity and highest spectral count are still within an acceptable limit of 10% and their increased percent RSDs may likely be explained due to variability in the amount of sample loaded on-column between each biological replicate.

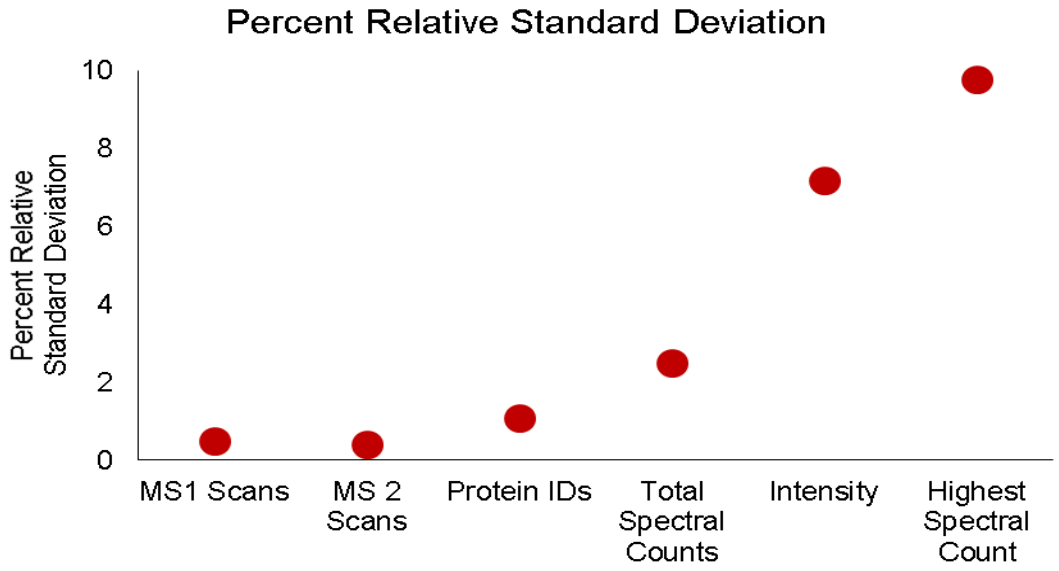


Figure 2.1. Comparison of variation across 6 fundamental parameters. Note that intensity refers to the MS1/precursor intensity. The low percent RSD values demonstrate good precision in the measurements with percent RSDs well below 10%. Higher variation observed with total spectral counts, MS1 Intensity and Highest Spectral count could likely be due to variation in the amount of sample loaded on-column.

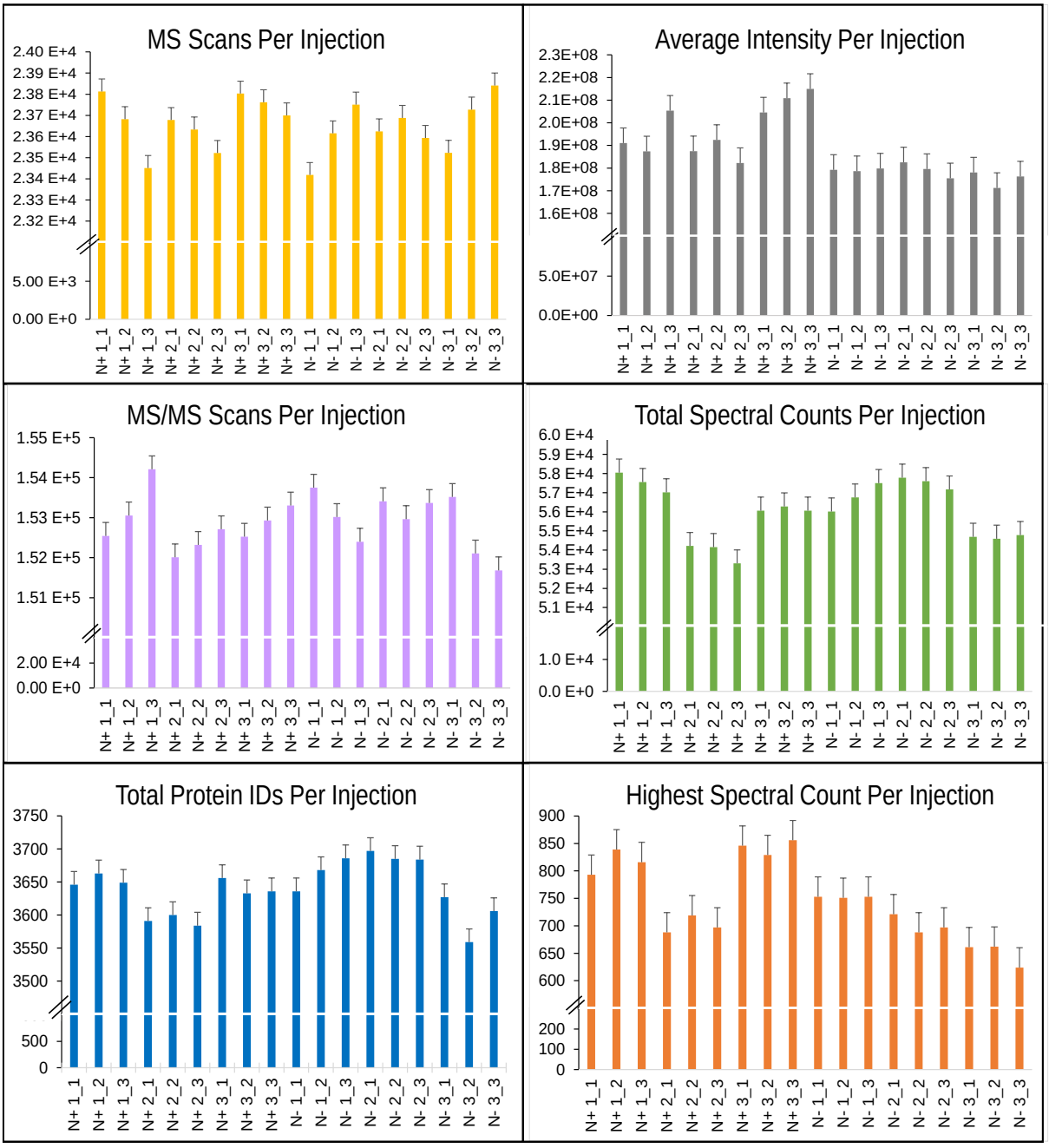


Figure 2.2. Comparison of fundamental parameters to show good reproduciblity in measuresments in a per-injection basis. Note that the y-axis is broken to allow a zoomed-in view of the data so that error bars could be observed. Error bars represent a 95% confidence interval. Good reproducibility indicates that instrumentation peformed well with little drift, resulting in highligh precise measurements from injection to injection.

In addition, the retention time and peak shape of 3 peptides were recorded for each injection at different retention times to ensure reasonable chromatographic reproducibility. Figure 2.3 depicts an example of the peak shape that was observed in each injection, while Figure 2.4 shows the distribution of retention times in half-minute bins. The average range of retention time was 2.24 minutes, and the average percent RSD was 0.41%, indicating that chromatography was also performing reliably.

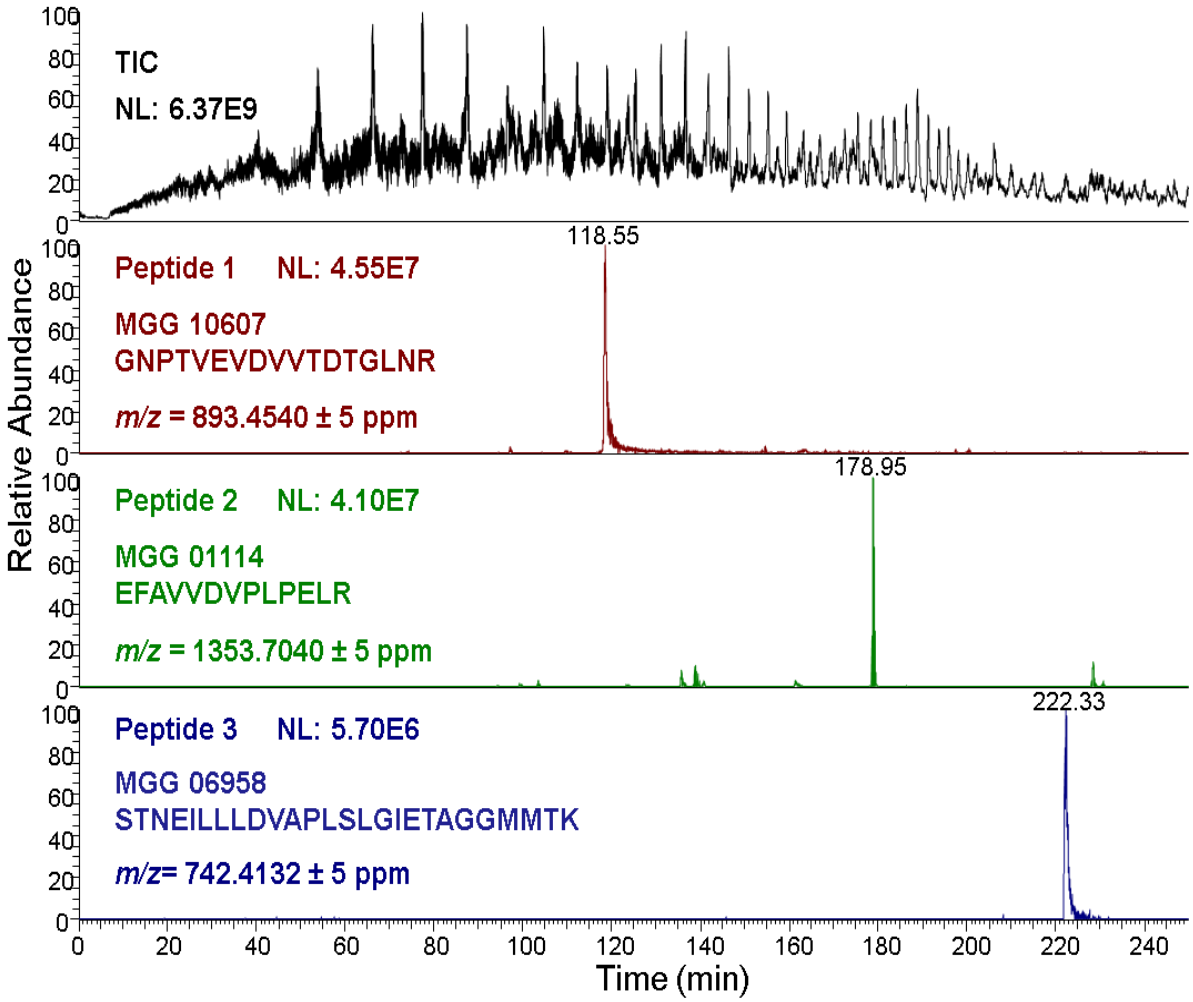


Figure 2.3. An example of the peptides that were used to track the retention time and peak shape of these 3 peptides for each injection to ensure reproducible chromatography. Example taken from Nitrogen Treated Sample 3, 3rd injection

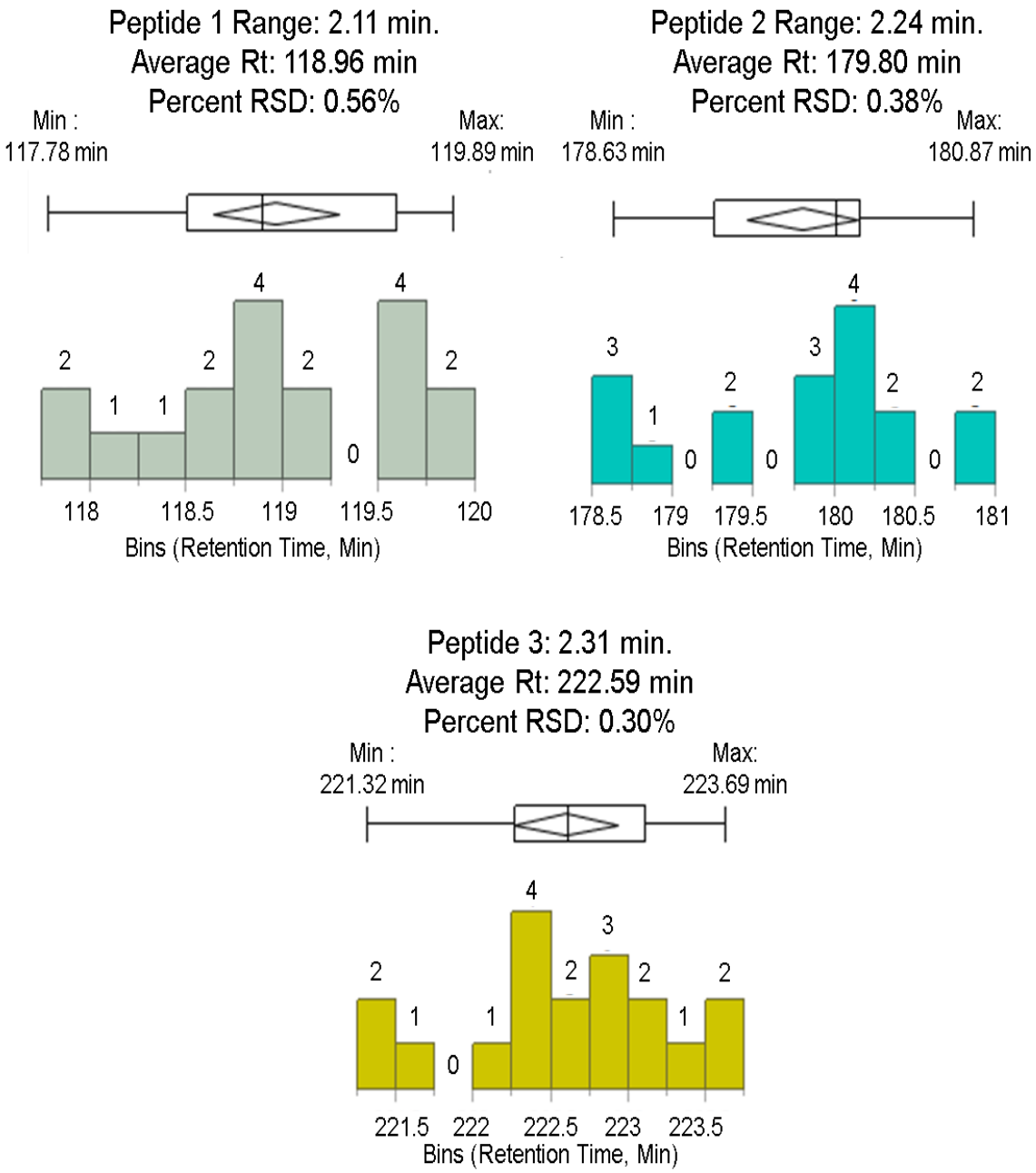


Figure 2.4. Summary of the retention time of each peptide tracked per injection throughout the experiment. Average retention time range was within 2.2 minutes. Overall, acceptable reproducibility of the chromatography was observed at three different time points in the gradient for each injection (n = 18).

An additional tool that can be used to assess data quality is RawMeat. This program was used to quickly visualize parameters for each injection. Figure 2.5 shows the visualization of the biological replicates for a nitrogen-starved sample as a representative sample for the experiment. The MS2 charge distribution visually demonstrates good precision in the charge state for each injection. Likewise, the Top N Spacing shows the number of MS/MS events after each MS1 survey scan. Given that the majority of the fragmentation events are occurring at a frequency less than 10 MS/MS scans per MS1 scan, this suggests that the proteome was adequately sampled. The rate of elution of the peptides during the 4-hour gradient was generally well-matched to the scan speed of the Q-Exactive High Field to minimize undersampling the proteome.

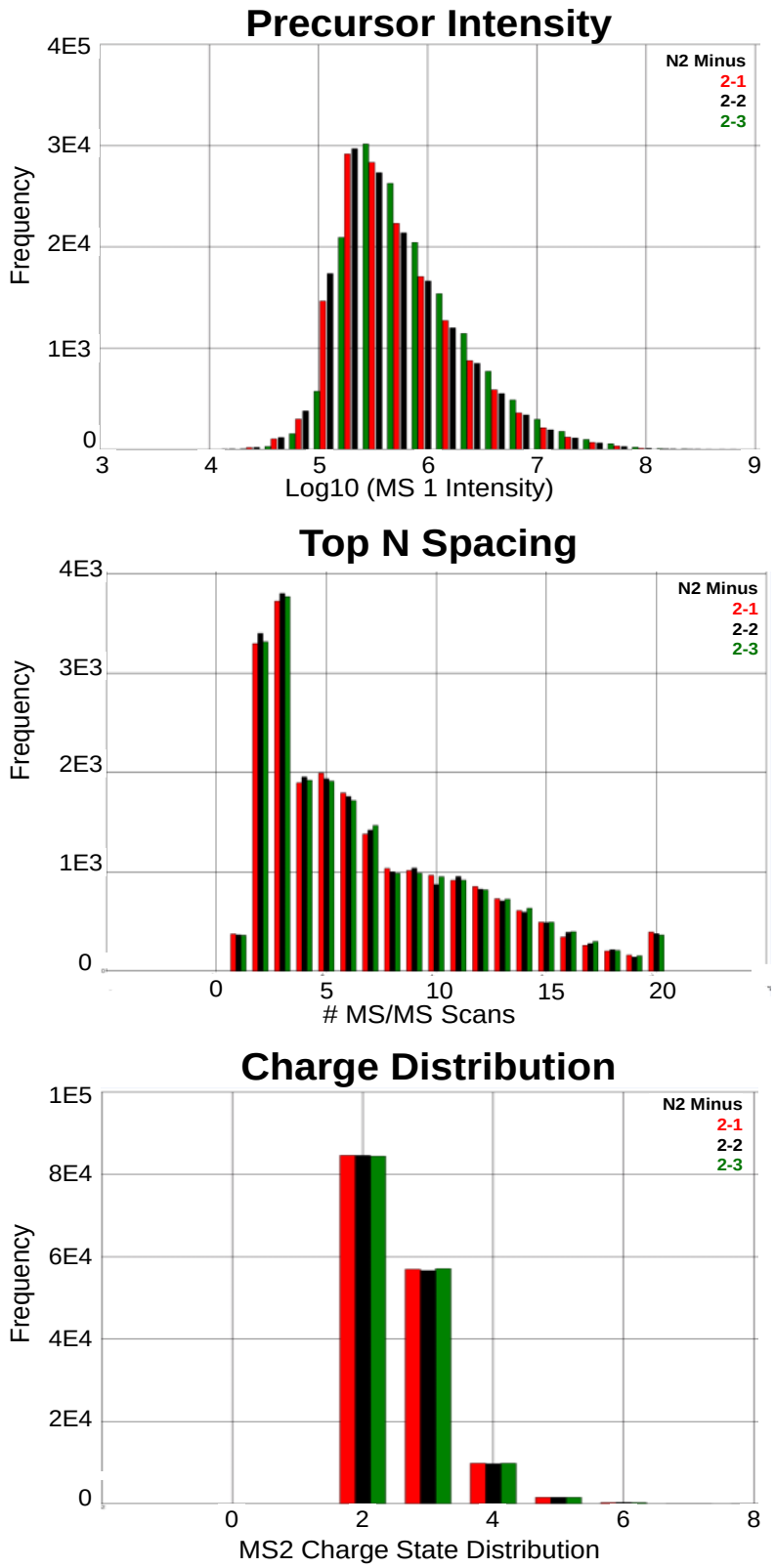


Figure 2.5. An example of raw data summarized by RawMeat used as a quality control tool to monitor the precision of each injection. Nitrogen starved sample (biological replicate 2) is used to illustrate this example.

Given this initial portion of the data, these fundamental parameters illustrate that the instrumentation was well calibrated and did not drift over the course of the experiment (~ 2 weeks). The high quality of these measurements yielded excellent precision and good accuracy, demonstrating that this data can be used to gain insight into the biology of *M. oryzae.* In order to proceed to an analysis of the biological data, the spectral counts was normalized. The total spectral count (TSpC) normalization was first applied to the three technical replicates of each biological replicate. The normalization process is represented in the scatter plots in Figures 2.6 and 2.7 where the average spectral count vs the spectral counts of each technical replicate is graphed for the unnormalized and normalized values. It is assumed that in the absence of biological variation, the slope of the regression 1ine should be unity since the plot consists of only technical replicates. Indeed, the technical replicates for both the nitrogen treated and nitrogen starved samples had unnormalized slopes nearly or at 1, with the average unnormalized slope at 0.9994 ± 0.0028 and the averaged normalized slope was 0.9996 ± 0.0022. Likewise, the average normalization factor was 1.0084 ± 0.0052. This data further supports the fact that each injection and technical replicate possessed a high degree of reproducibility as minimal normalization factors were applied, and unnormalized slopes were nearly 1.

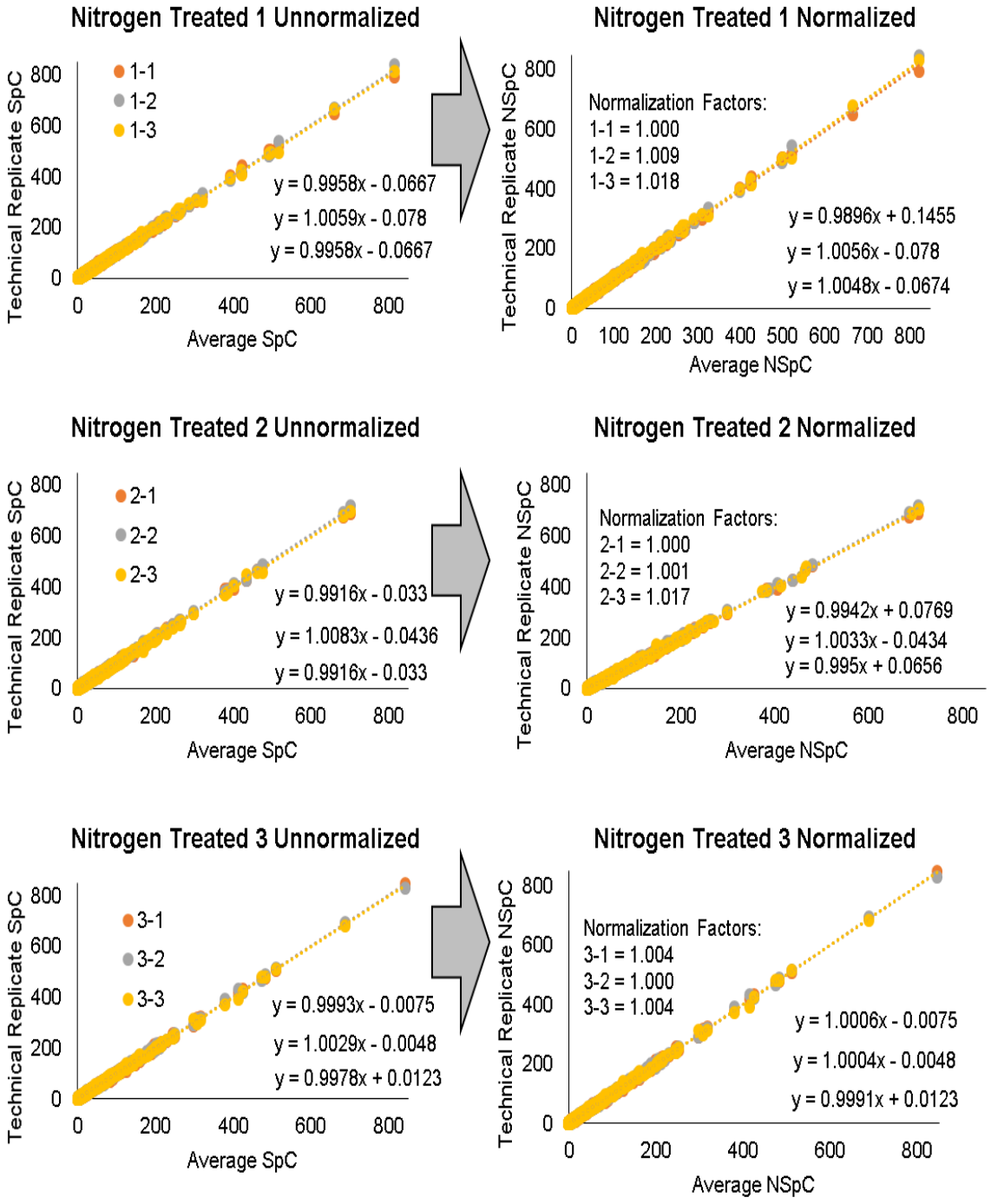


Figure 2.6. The normalization process using the total spectral count approach applied to each technical replicate for nitrogen treated samples. Slopes near unity indicate excellent precision within the technical replicates performed for each biological replicate.

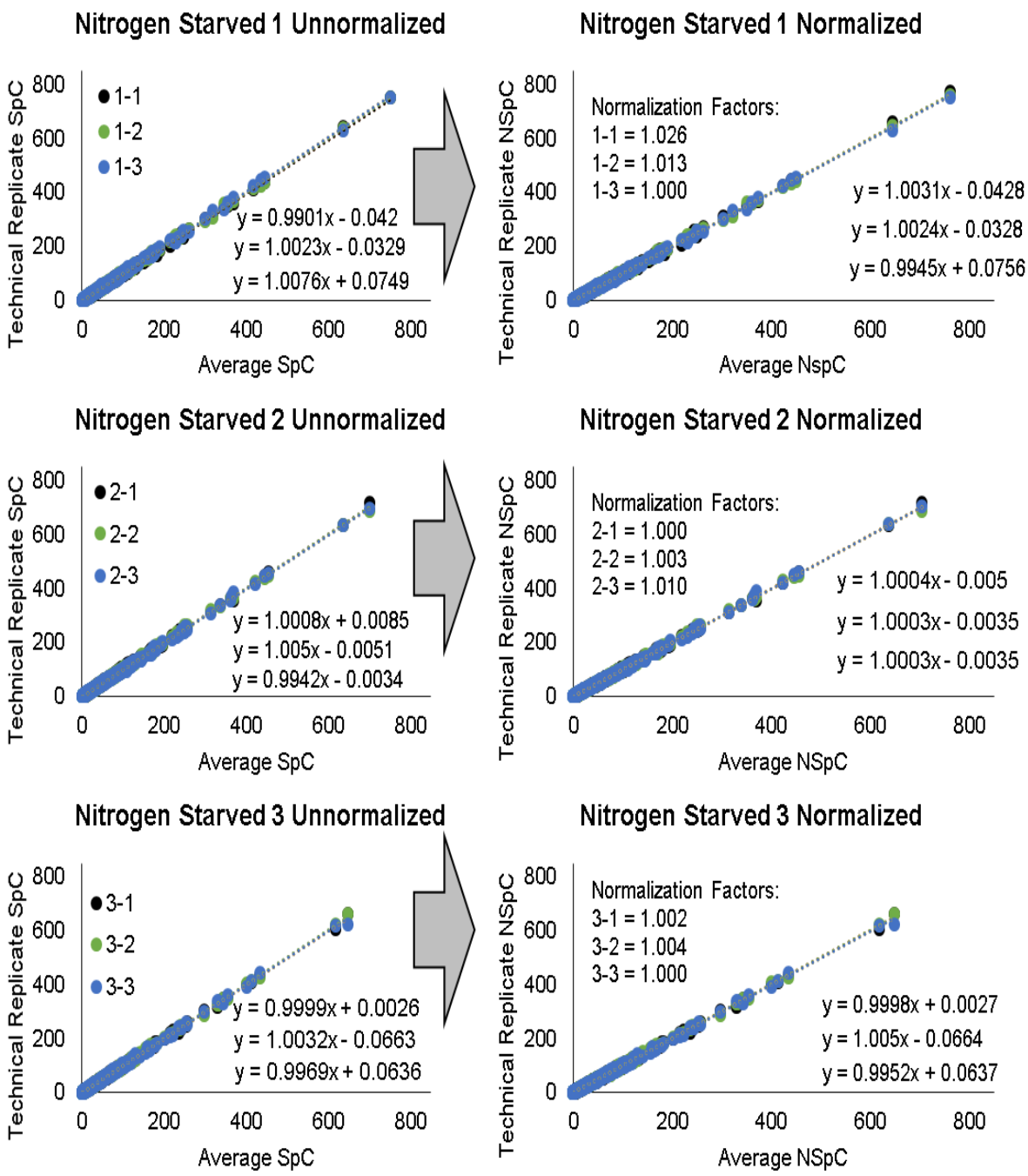


Figure 2.7. The normalization process using the total spectral count approach was applied to each technical replicate for nitrogen starved samples. Slopes near unity indicate excellent precision for the technical replicates performed for each biological replicate.

Following the technical replicate normalization, the TSpC normalization was applied to each biological replicate for the nitrogen treated and starved samples (respectively) as summarized in Figure 2.8. The biological replicates also had highly reproducible spectral counts as the average unnormalized slope for the nitrogen treated samples was 1.0133 ± 0.0450 and nitrogen starved was 1.0000_{33} ± 0.0769. The averaged normalized slope for both biological samples was 1 with an average normalization factor of 1.0280 ± 0.0496.

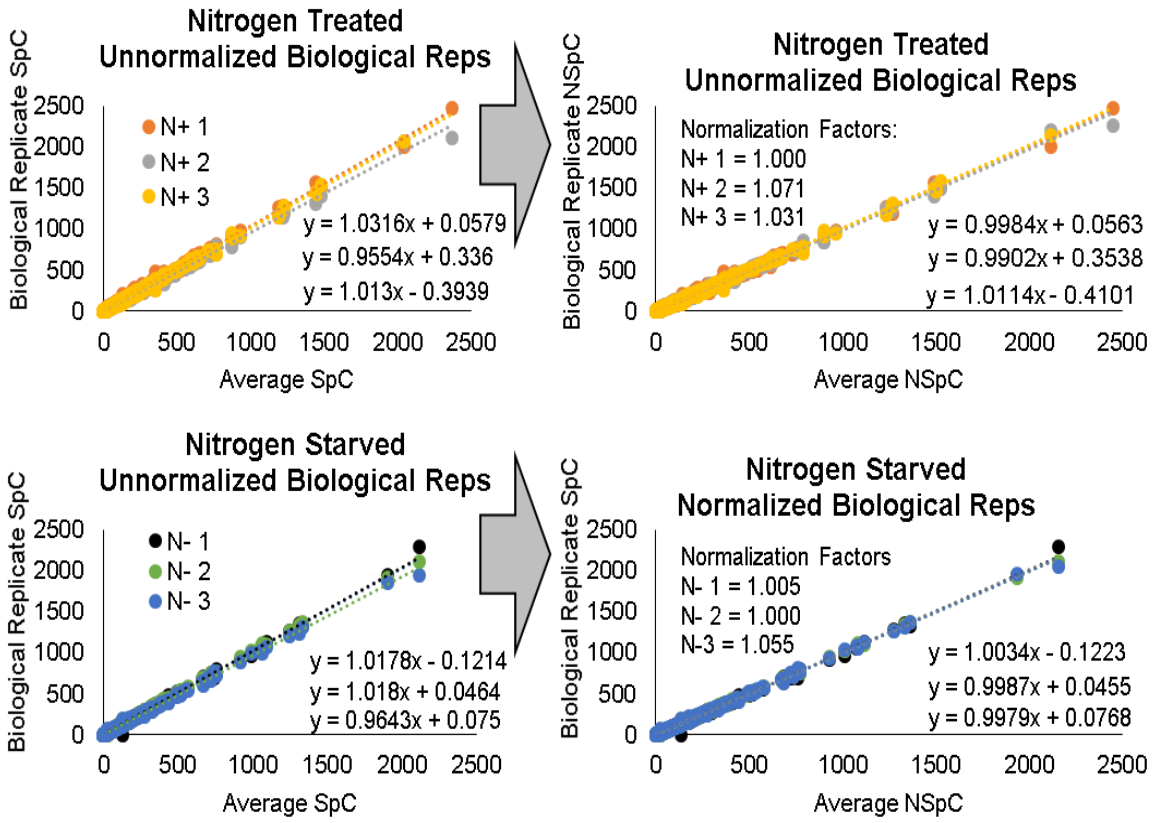


Figure 2.8. The second application of the total spectral count normalization, applied to the biological replicates of each experimental condition. The biological samples of each treatment condition were all highly reproducible.

The degree of homogeneity of the biological samples is described in Figure 2.9A and 2.9B**.** For the nitrogen treated samples, 74% of proteins were identified in all 3 biological replicates with an average of 5% of proteins uniquely identified in a single replicate. Similarly, in the nitrogen treated samples, 76% of proteins were observed in each biological

replicate with an average of 5% of proteins uniquely identified in 1 replicate. Given the high degree of overlapping identifications, this indicates that for each condition, the samples were biologically similar to one another. Given that these were cultures grown under a controlled environment, this high degree of homogeneity is both logical and expected.

In addition, good technical reproducibility was observed across all biological replicates, regardless of treatment as evidenced by Figure 2.9C and 2.9D. The average number of protein identifications was 4,164 with a percent RSD of 1.01%. The percent change of the number of protein identifications was calculated relative to its average, which was less than 2%. Likewise, the average amount of spectral counts per biological replicate was 169,299 and its percent RSD was 2.88%. The percent change was also calculated relative to its average and was less than 4%. This data continues to demonstrate a high degree of reproducibility with percent RSD's below 3% and percent changes below 4% even across biological treatments, further highlighting excellent precision.

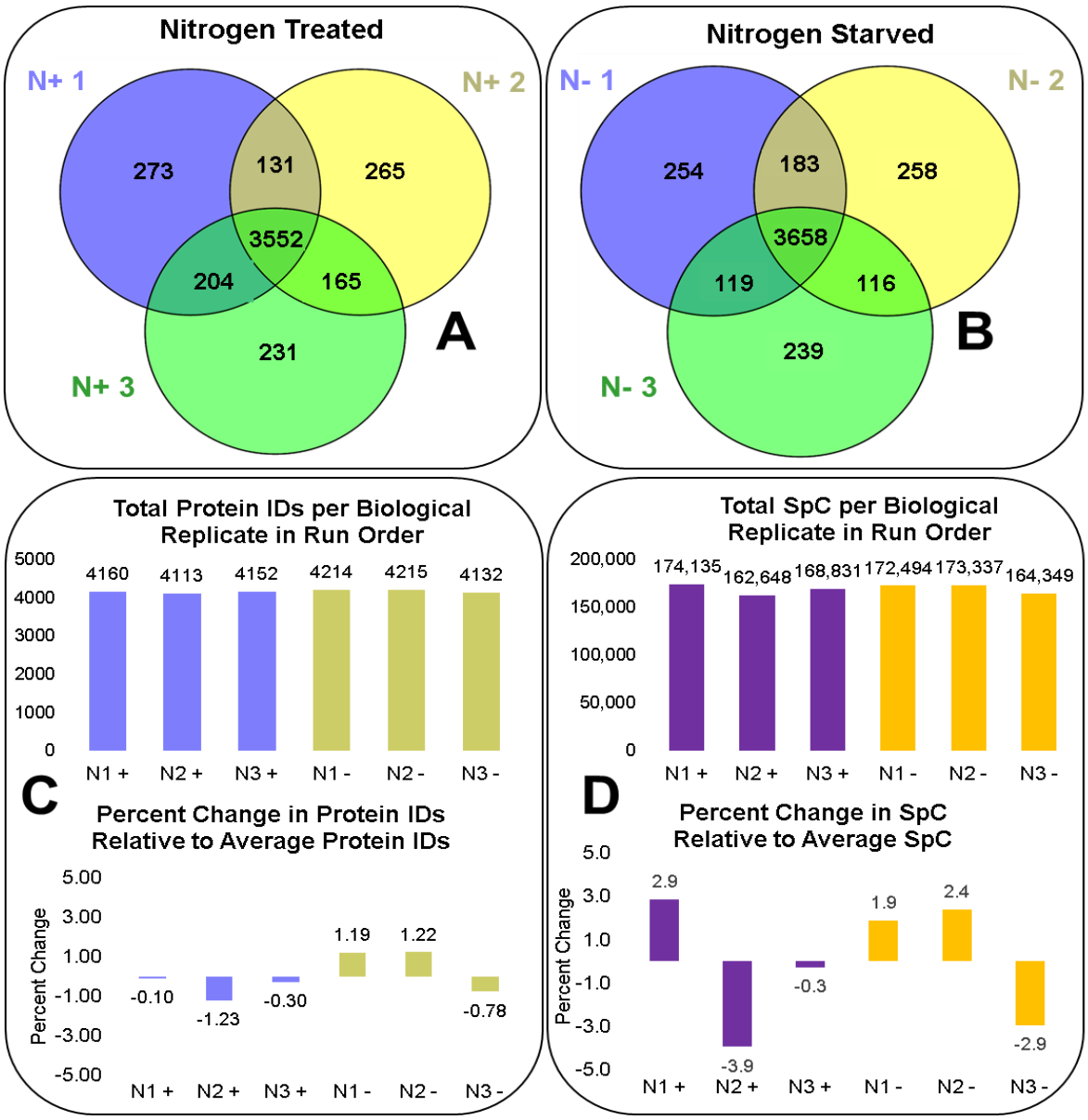


Figure 2.9. The homogeneity of the biological replicates is illustrated in parts A and B demonstrating good reproducibility among biological replicates of each sample treatment. C) Low variation across the total number of protein identifications per biological replicate was observed, regardless of the sample treatment with less than 2% change. D) Good precision for the total spectral count per biological replicate was also shown with less than a 4% change in the total number of spectral counts in each biological replicate.

Lastly, the TSpC normalization was applied across each condition to be able to directly compare between the nitrogen treated and nitrogen starved conditions. This was represented by graphing the unnormalized and normalized spectral count values of the

nitrogen treated vs nitrogen starved conditions in Figure 2.10A**.** The lower slope values shown of 0.9079 (unnormalized) and 0.9120 (normalized) indicate some differences between the two conditions. At this point greater difference is expected since the two experimental conditions were treated differently. Since the slopes for unnormalized and normalized values were near unity for the technical and biological replicates, this lower slope value is likely due to biological differences observed between the two different sample treatments.

In total, 5,498 proteins were identified across all samples with 75% of proteins being identified in both the nitrogen treated and nitrogen starved states. In the nitrogen starved samples, 704 proteins were uniquely identified, and 696 proteins uniquely identified in the nitrogen treated condition. However, as highlighted in the histograms in Figure 2.10B, ~84% of the unique identifications made in either condition have 5 or less spectral counts. Protein identifications with low spectral counts are problematic for several reasons. It has shown in the literature that a minimum of 5 spectral counts should be used to achieve a more accurate quantification.[77,84] Moreover, low spectral counts are identified from proteins of low abundance whose quality of the MS1 elution profile may be of a poorer quality causing the identification of the protein to be less certain. As such, these unique proteins with low (< 5) spectral counts were analyzed to contribute their biological story, but will require acquire additional validation prior to publication.

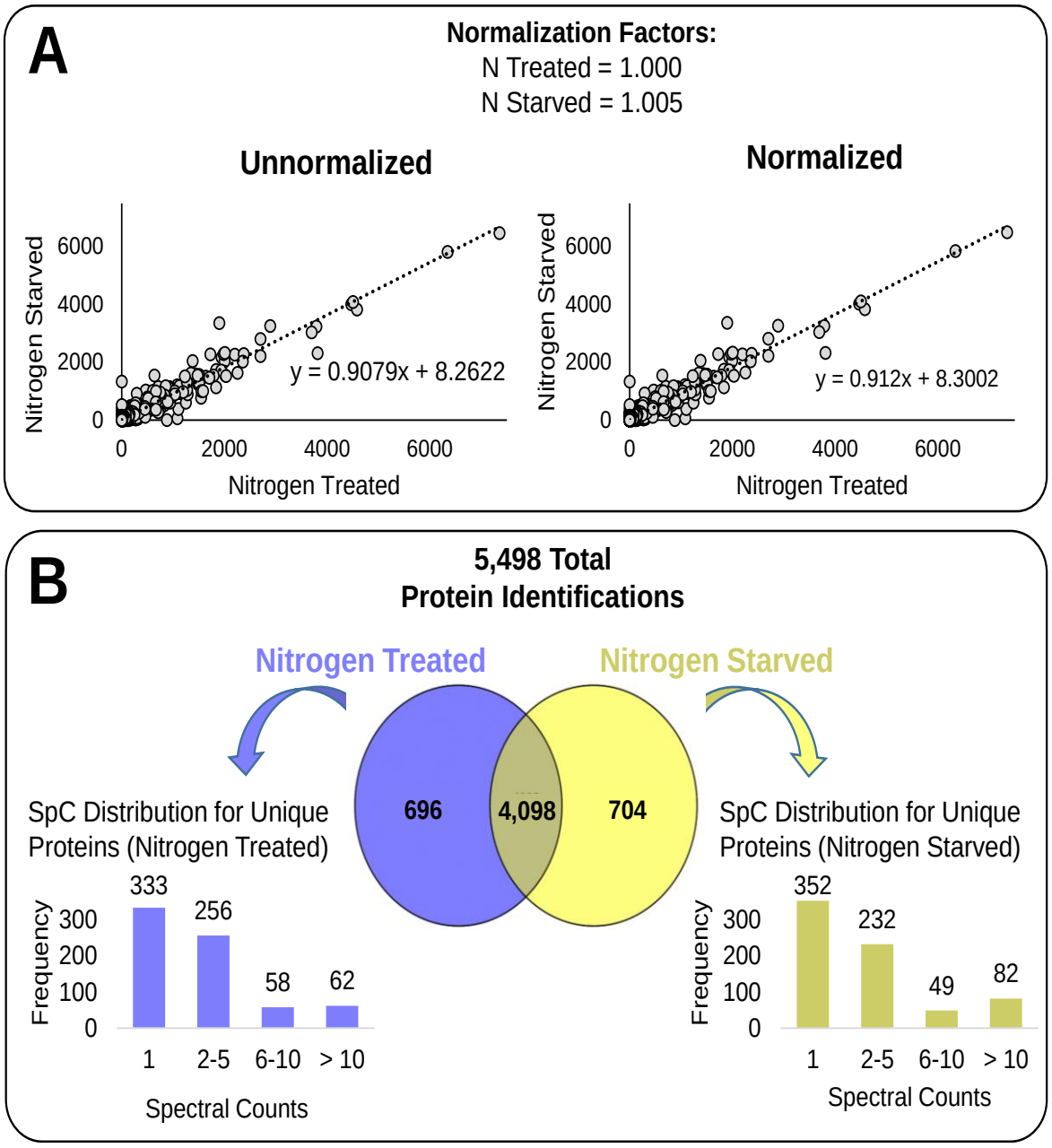


Figure 2.10. A) Normalization using total spectral counts to account for the variation between the two different sample treatments of nitrogen treated and nitrogen starved. A total of 5,498 proteins were identified across all samples B) Comparison of protein identifications between the two conditions and the spectral count distributions of uniquely identified proteins for each state is shown.

Biological Analysis

For plants, nitrogen is primarily the limiting nutrient for development and growth, and plants are restricted to nitrogen sources of ammonium or nitrate. While the preferred nitrogen source for many fungi are ammonium and glutamine, fungi have the ability to use alternate nitrogen sources to survive in a wide array of environments, including potential *in planta* conditions that are limited in nitrogen.[12,17] Thus this study examined the global effects of nitrogen starvation on the mycelia proteome of *M. oryzae* and its potential role in infection.

The protein ratios of nitrogen starved to nitrogen treated were calculated using the normalized spectral counts for each condition, and the nitrogen starved samples were divided by the spectral counts of the nitrogen treated samples; ratios were transformed to a $\log_2$ scale. Differential protein expression was determined using a pairwise student's *t* test and a strict cutoff value of $p<0.05$ was applied with a minimum of a 2-fold change. Figure 2.11A demonstrates that the fold-change for all proteins was a normal distribution, which is important to accurately apply the student's *t* test. Figure 2.11B depicts the volcano plot of the fold change vs the p-value which resulted in 185 proteins were downregulated in the nitrogen starved samples, and 271 proteins upregulated. Figure 2.11C also illustrates the same volcano plot, only zoomed-in on the fold-change (x) axis to illustrate that the expected shape of volcano plot was observed.

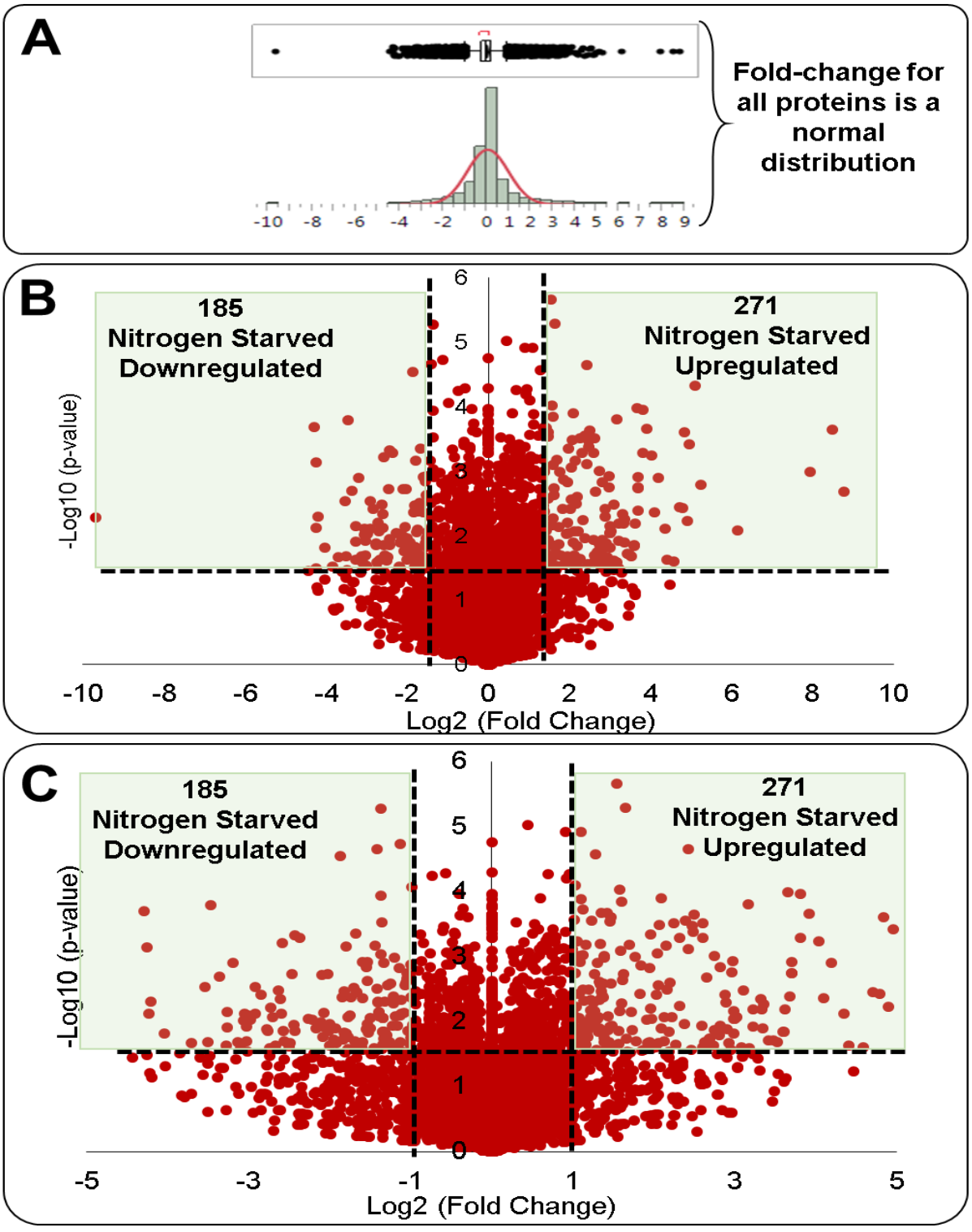


Figure 2.11. A) The protein fold-change fits a normal distribution, demonstrating that a student's t-test may be appropriately applied to the data. B) Scatter plot of the log2 fold change vs. the –log10 (p-value) which depicts 185 nitrogen-starved proteins downregulated at 271 proteins upregulated. C) A closer view of the axis from part B to show a typical distribution for up and downregulated proteins.

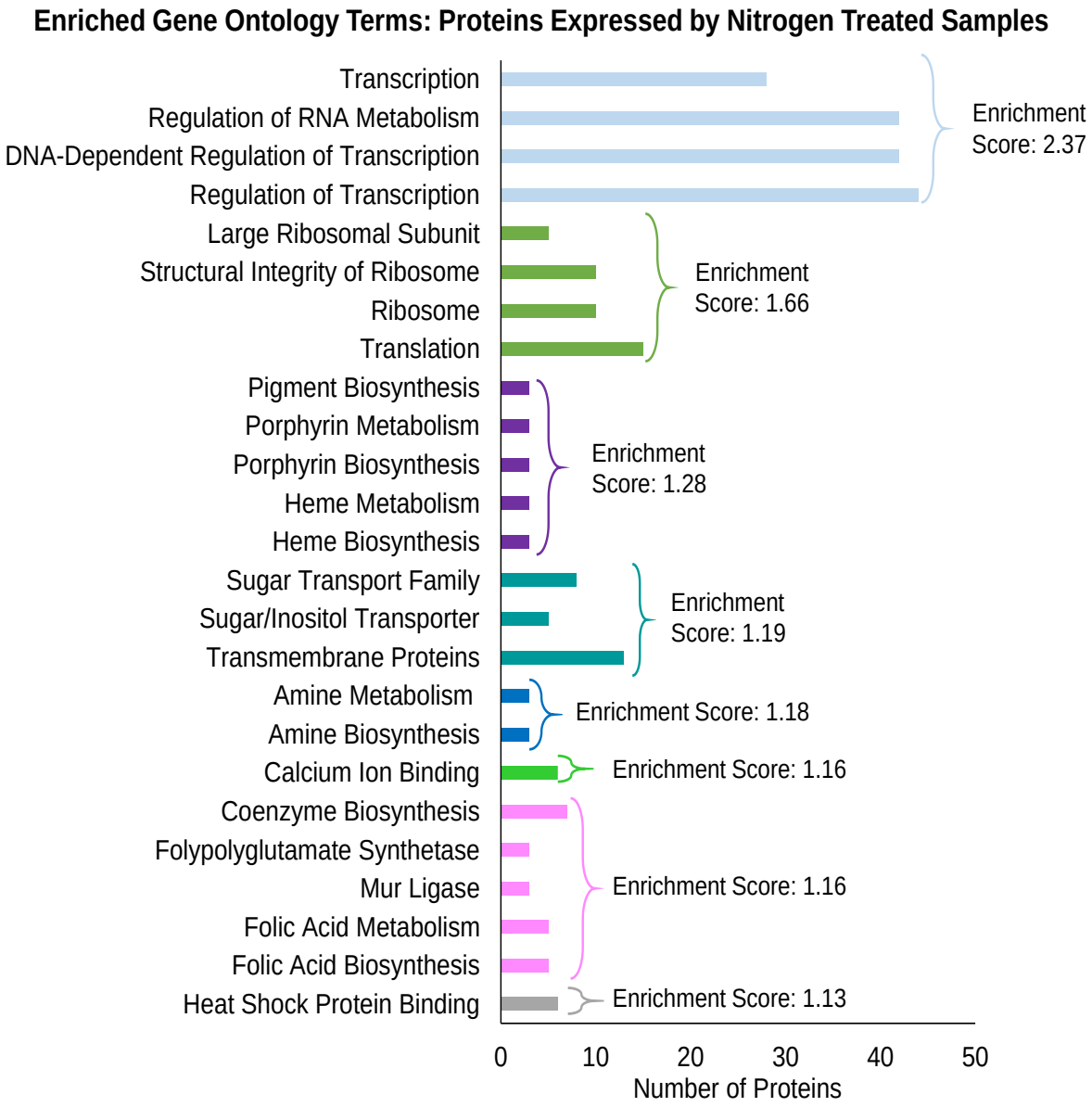


Figure 2.12. Protein enrichment clusters observed in the nitrogen treated condition.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics was used to perform an enrichment analysis where the higher the enrichment score, the better the enrichment for the cluster.[86] Figure 2.12 shows the enriched proteins that were expressed by the nitrogen treated control samples (proteins that were uniquely expressed by nitrogen treated samples or downregulated by the nitrogen deficient samples). The first cluster pertaining to the regulation of transcription was uniquely expressed by the

nitrogen treated samples with an enrichment score of 2.37 with 28-44 proteins expressed in the group. This is logical because once *M. oryzae* has infiltrated a host plant cell, it begins to transcriptionally reprogram its host in order to support a rapid generation of DNA, RNA and proteins for its infection.[5]

Likewise, a protein cluster relating to translation and ribosomal proteins had an enrichment score of 1.66, and was downregulated in the nitrogen starved samples, thereby being expressed in nitrogen-treated samples. A previous transcriptomic study demonstrated that in 17-day-old plants that were 5 days post infection with *M. oryzae*, 50% of the genes expressed were utilized for DNA, RNA, amino acid, protein synthesis and energy production.[5] In addition, a slight enrichment of 1.18 in amine metabolism and biosynthesis was noted which further supports the growth and propagation of the fungus. Thus under nitrogen-present conditions, a large portion of *M. oryzae's* resources are used in support of immediate proliferation to infect its host cells demonstrated by the enrichment of proteins related to transcription and translation.

Similarly, 6 proteins clustered in heat shock protein binding were uniquely expressed by the nitrogen-treated sample with a slight enrichment score of 1.13. Heat shock proteins (HSPs) are ubiquitously expressed across all organism, and are typically expressed in response to stressful conditions including changes in temperature, nutrient supply, pH, oxidative, osmotic or antifungal stresses.[88-90] These proteins play crucial roles as chaperone proteins by aiding in the proper folding of proteins, facilitating the correct refolding of misfolded proteins, and preventing protein aggregation.[88-90] They also assist in various biological process such as transcription, translation, post-translational modifications, cell cycle signaling, and the development and polymorphism of membrane lipids.[88-90]

The expression of HSPs could be in response to the fact that while there is a nitrogen source supplemented to the media, *M. oryzae* was still grown in a minimal media which could be a stressful environment compared to growth in a rich media. Regardless of stressors present, the function HSPs pairs well with the increased levels of protein transcription and translation. Moreover, within the fungal kingdom, HSPs have shown to hold crucial roles in morphogenetic changes, stress adaptation and antifungal resistance.[89] Given that heat shock proteins can play a role in the morphogenesis of the fungi, perhaps the observed HSPs could also be related to the remodeling of the fungal and/or plant wall which occurs during the first few days of infection.

Pigment biosynthesis as well as porphyrin and heme biosynthesis and metabolism were enriched with a score of 1.28 and were expressed by the nitrogen treated samples (downregulated in nitrogen starved condition). Pigment plays a critical role as a virulence factor for various pathogenic fungi, including *M. oryzae.*[2,4] Melanin is used in various capacities such as an antioxidant, protects from UV exposure or can be a toxic metabolite.[2] In *M. oryzae*, melanin is an essential part of the appressorium structure.[2,4] Turgor pressure can reach 8 MPa prior to the penetration peg formation in the appressorium.[2,4] Mutants lacking melanin synthesis are nonpathogenic as they cannot generate a high enough turgor pressure to penetrate the plant's cuticle.[2,4] Thus it has been hypothesized that melanin strengthens and acts as an impermeable layer to hinder leakage of osmolytes (such as glycol) needed for turgor generation to puncture through the plant tissue.

Note that appressorium differentiation and penetration requires 2-28 hours whereas hyphal invasion and mycelium development requires 2-6 days. These cultures were grown for three days and the material analyzed is mycelium. Given the crucial role of melanin as a

virulence factor in appressoria, which would have only recently differentiated and maturated, the role of the pigment synthesis and metabolism could continue to be a virulence factor to increase the success of the infection process, even after the differentiation of the appressoria.

The following cluster of enriched proteins relate to sugar transport with an enrichment score of 1.19; which were downregulated in the nitrogen starved samples. During the biotrophic phase of infection, an extensive amount of endocytosis and exocytosis occurs between *M. oryzae's* mycelia and the plant cell.[5] As previously mentioned, the host cell is transcriptionally reprogrammed to support *M. oryzae's* production of DNA, RNA and proteins.[5] Therefore, the expression of sugar transporters in the nitrogen treated samples may reflect the re-routing of plant resources in order to provide energy to support fungal expansion and reproduction. Likewise, the cell wall is primarily composed of glucans, chitin and glycoproteins. Given the need for glucan for cell wall synthesis, the enrichment of sugar transporters could also be in support of supplying necessary carbohydrates for *de novo* cell wall biosynthesis.

Calcium ion binding is a highly conserved mechanism in which Ca^{2+} ions can act as second messengers for various cell signaling pathways, and has been extensively characterized within animal systems.[91] Cytosolic Ca^{2+} concentrations are maintained at low levels by Ca^{2+} pumps and transporters.[92] Upon external cellular signals or environmental cues, the Ca^{2+} concentration is transiently increased, providing the second message which can activate various downstream signals through proteins like the protein kinase C and Ca^{2+}/calmodulin binding kinases.[92]

Six calcium ion binding proteins were expressed in the nitrogen treated samples and slightly enriched with a score of 1.16. In *M. oryzae,* the *ΔMoCMK1* gene was shown to

encode a putative calcium/calmodulin-dependent kinase.[93] The disruption of this gene affected: fungal morphogenesis, conidia development and appressorium formation which contributed to a significant reduction in pathogenicity.[93] Similar results were found in a study of the *M. oryzae* phospholipase C1 (*MoPLC1)* gene which was also shown to regulate intracellular Ca^{2+} levels.[93] The *ΔMoPLC1* deletion mutants clearly demonstrated that removal of the calcium flux greatly reduced the number of conidia produced, and impacted appressorium development.[93] When a rice cultivar of young seedlings was inoculated with *ΔMoPLCstrain1* , the fungus was nearly non-pathogenic and only induced a few minor lesions which did not further develop.[93]

Thus the same effects of fungal morphogenesis, conidia and appressorium development and reduced pathogenicity were observed in *ΔMoPLC1* and *ΔMoCMK1* mutant *M. oryzae* strains. While calcium signal transduction has been less-characterized in filamentous fungi, it has been demonstrated to play a vital role in the development and pathogenicity of *M. oryzae*, so the expression of calcium ion binding proteins in the nitrogen treated samples could be another developmental signal to achieve plant infection.

The following protein cluster is related to fungal cell wall biosynthesis and has an enrichment score of 1.16. Mur ligases are ATP-dependent enzymes that add key amino acid moieties to the cell wall.[94,95] Folypolyglutamate synthases are another member of mur ligase enzymes and they are used to add a polyglutamate tail to folate and its derivatives.[96] Pathways involving folate generally involve the transfer of a 1-carbon unit to amino acids, nucleotides and other biomolecules. As this is a cluster of mur ligase enzymes, coenzymes and folic acid related proteins used in fungal wall development and growth, their presence

continues to suggest that the nitrogen treated are actively growing, most likely in preparation of the fungal invasion of the host plant.

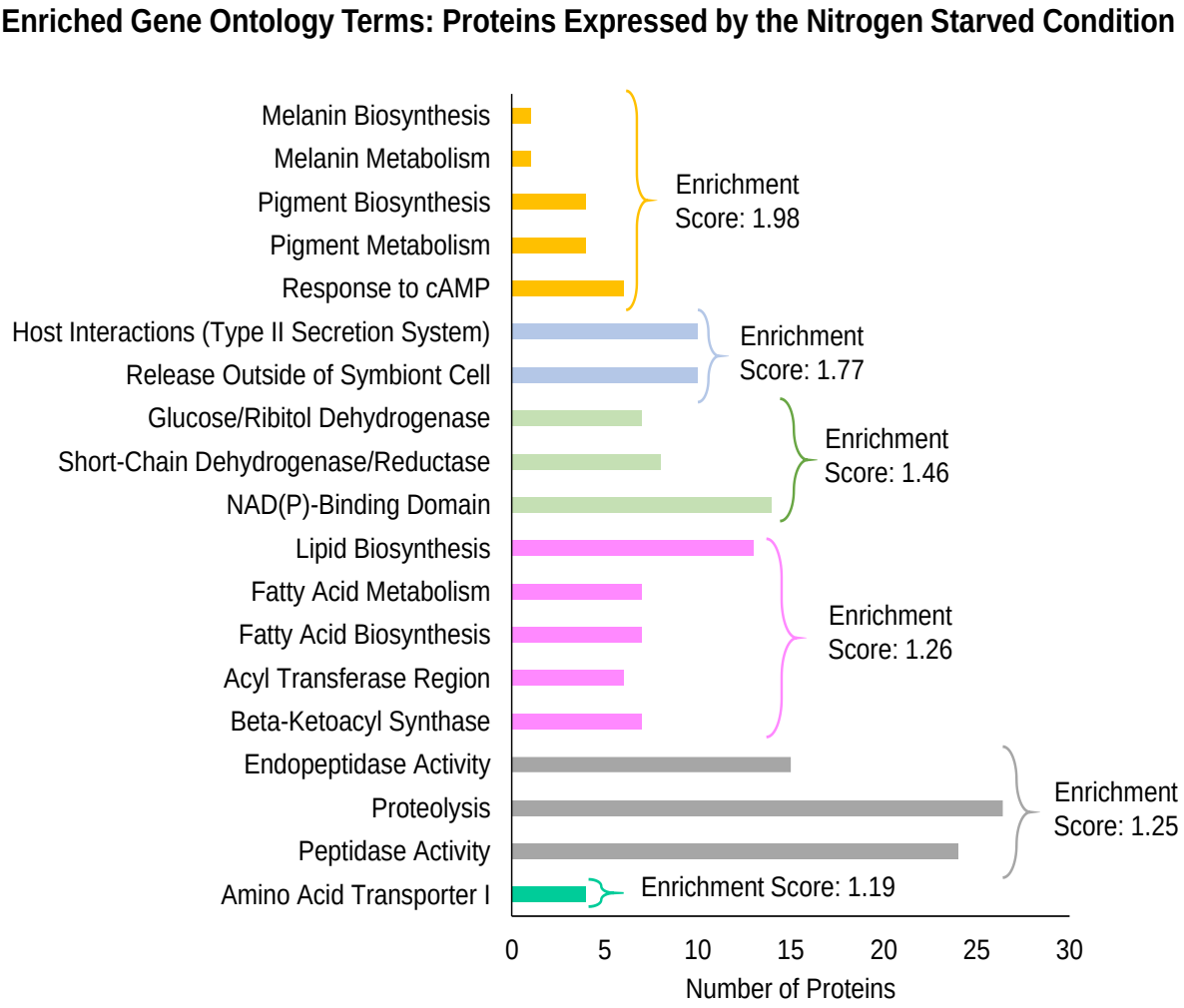


Figure 2.13 Protein enrichment clusters observed during nitrogen starvation.

Figure 2.13 highlights the enriched proteins for the Nitrogen starved samples which interestingly differ from their control. The protein cluster with the greatest enrichment is related to melanin development as well as a response to cAMP, and it was upregulated in the nitrogen starved samples. As mentioned previously, melanin synthesis and metabolism is a crucial virulence factor for the development of appressorium.[2,4] Similarly, cAMP signaling

regulates appressorium development.[3] Franck *et al*. showed that when the cAMP-dependent protein kinase A gene was deleted, *M. oryzae* was unable to form appressorium.[3] While an enrichment score of 1.3 for pigment-related proteins was observed in the nitrogen treated control group, there is a greater enrichment for pigment and melanin in the nitrogen starved samples with as its enrichment score is 2.0. Also, a response to cAMP was not observed in the control group. This could suggest that the nitrogen starvation condition has triggered a greater virulence response compared to nitrogen treated samples.

The second protein cluster is comprised of twenty proteins relating to host interactions with a score of 1.77, and it was uniquely expressed by the nitrogen starved samples. Host interactions by type II secretion system as well as releasing proteins outside of the symbiont cells were enriched. In hemibiotrophic fungi, a large diversity of effector proteins are secreted into the plant's cytoplasm or apoplast in order suppress plant immunity, adjust host metabolism and prevent recognition of the invading pathogen.[10,97] Given the intimate relationship between symbionts and their hosts, and the fact that extensive endo/exocytosis occurs at the mycelia, secretion of effector proteins to support the infection of the host plant is both likely and expected.[5]

The third enriched group has a score of 1.46 and consists of dehydrogenase and reductase proteins. Within the initial developing infection cells, an oxidative burst is detected as part of the cell wall differentiation of the appressorium.[4,5] It has been proposed that reactive oxygen species (ROS) may be used in the cross-linking of proteins into the cell wall during appressorium maturation.[4] The expression of oxidative/reductive proteins could serve a role in the generation of ROSs for cell wall biosynthesis, as well as serve a protective function from the generation of free radicles. Moreover, during germination of conidia,

glycerol can be produced from carbohydrates, which is a NADH-dependent reaction. Given the importance of glycerol for turgor pressure generation and appressoria maturation, the dehydrogenase/reductase protein expression could also be a sign of *M. oryzae* actively infecting its host through the generation of glycerol.

Polyketides are complex organic compounds that comprise a class of secondary metabolites produced by many fungi to confer a survival advantage, possessing a plethora of pharmacological properties such as antibacterial, antiparasitic and anticancer properties.[98,99] These sophisticated metabolites are generated from enzymatic complexes such as polyketide synthases (PKSs), non-ribosomal peptide synthases (NRPSs) and hybrid PKS-NRPS systems.[4,98]

The *M. oryzae* genome encodes for a large number of these enzymes with 22 PKSs, 8 NRPSs and 10 PKS-NRPSs, granting a considerable capacity for polyketide expression.[4] Six proteins relating to the acyl transferase (AT) function were uniquely expressed by the nitrogen starved samples. AT domains are a part of modular PKSs and are responsible for the specific selection of α-carboxyacl-CoA substrates to be added to the growing polyketide chain.[100] Likewise, 7 proteins for beta-ketoacyl synthase (BKS) which catalyze a condensation reaction between the selected extender unit and the growing polyketide chain were observed.[98] Interestingly, BKS is also a highly conserved enzyme as a domain in fatty acid synthases, which highlights a high degree of mechanistic similarities between fatty acid and polyketide biosynthesis.[99] Indeed, fatty acid synthesis and metabolism products can used in these secondary polyketide pathways as precursors or building blocks to be added to the growing product.[4] Moreover, it has been proposed that secondary metabolite production produced by PKS/NRP/PKS-NRSP enzymes in *M. oryzae* could perform functions similar to

protein effectors in plant pathogenesis such as suppressing plant defenses and altering the host metabolism.[4] Melanin, which is a crucial virulence factor is synthesized by a pentaketide pathway in *M. oryzae.* Thus this cluster of proteins could be key to adjusting plant response during the infection, thereby expressing additional proteins in support of *M. oryzae's* infection by the nitrogen starvation samples.

Finally, an enrichment of proteolytic enzymes was observed with a score of 1.25. The activity in protein degradation is likely linked to the recycling of nitrogen materials in the limited environment. Precious cellular resources are conserved and reused in the support of the infection process, saving *M. oryzae* an enormous amount of energy in a nitrogen-scarce environment. It has been shown that in the mycelium of the fungus, *Schizophyllum commune*, the proteolysis of proteins in older mycelium occurs during nitrogen starvation to release and recycle amino acids needed for its continued survival in a limited environment.[101] The fact that the amino acid transporter was also enriched could also be evidence of the recycling of cellular materials because these proteins were upregulated in the nitrogen starved condition. It is possible that the transporters are being used to re-route materials from the host cells which can then be further broken down by proteolysis to provide amino acids and other limited building blocks to *M. oryzae* that are scarce during nitrogen starvation.

Noteworthy differences appeared in the global proteome expressions of mycelia between the nitrogen starved and treated conditions. As a whole, it appears that the nitrogen-treated control group largely expressed proteins for the general growth and propagation of *M. oryzae* whereas the nitrogen starved samples was enriched for the recycling of cellular materials. The nitrogen limited samples also appeared to express a greater number of proteins related to virulence and pathogenicity compared to their control samples. While the

nitrogen treated samples did express some proteins relating to pigment biosynthesis and metabolism, a greater specificity in the role of pathogenesis was seen in the nitrogen starved samples by the larger enrichment score of melanin biosynthesis and metabolism, and a response to cAMP. Furthermore, the expression of enzymatic machinery related to PKS systems was uniquely observed in the nitrogen starved group. Likewise, an enrichment of host interactions were also uniquely expressed in nitrogen limited samples, potentially implying a greater release of effector proteins to aid the infection process.

Comparison to Transcriptome Study

A previous study by Donofrio *et al.* investigated the effects of nitrogen starvation on *M. oryzae's* global gene expression.[17] Similarly to this current proteomic investigation, *M. oryzae* was grown in complete medium for 48 hours.[17] Mycelia were harvested and grown in minimal media with or without nitrogen sources for 12 and 48 hours, and performed with three biological replicates.[17] Gene expression was probed by hybridizing extracted cRNA with Agilent's oligo microarray kit which included 13,666 genes comprised of the entire predicted gene set for *M. oryzae* and 7,137 rice defense-related genes.[17] This experiment gave insight into the regulation of nitrogen catabolite repression genes as well as the expression of genes related to pathogenicity during nitrogen expression; interesting parallels can be seen between this transcriptomic study and the proteomic approach in this investigation.

Initially, both experiments demonstrate similar directions of fold-change in the up and downregulation of the nitrogen-starved condition. The transcriptome study showed 525 genes upregulated and 345 genes downregulated in response to nitrogen starvation and the proteomic approach yielded 271 proteins upregulated and 185 downregulated, consequently

showing a comparable trend of having a greater number of proteins upregulated in response to nitrogen starvation between the two approaches.[17]

Fifty-five genes that had the greatest upregulated fold-change were highlighted in the transcriptome study. The protein expression levels of these 55 genes were compared in Table 1. From the proteomics data, 60% of these 55 genes had their associated proteins detected, and 38% of these proteins possessed sufficient spectral counts for quantification, which are the proteins outlined in Table 1. While the magnitude of the fold-change for the upregulated genes was generally greater than the proteomic fold-change, a similar trend was observed in that 62% of proteins in common with this gene list (13/21) showed a statically significant ($p < 0.05$) upregulation, the same trend observed in their genes, and this group is colored green. Five out of the 21 proteins (24%) showed no significant fold-change and are presented in black, and 14% of the proteins showed a down regulation while their genes showed an upregulation and are colored in red (downregulation denoted by the negative fold-change).

Most notably in this set of genes and proteins that were both upregulated, are similar expressions of key regulators of nitrogen metabolism in *M. oryzae.* The nitrogen catabolite repression system (NCR) maintains a tight control over the gene and protein expression of nitrogen scavengers.[12] Much of what is known about the NCR has come from the study of non-pathogenic fungi.[12] The NCR proteins are a specific class of zinc-finger proteins that are transcriptional activators.[12] They possess a "GATA" sequence and bind to promoter regions with "HGATAR" domains that are specific to nitrogen-utilization pathways when preferred nitrogen sources such as ammonium or glutamine are absent or severely limited.[12] These GATA factors bind to their HGATAR promoters, activating genes used in nitrogen

scavenging.[12] If preferred nitrogen sources are present, these promoters are inhibited from binding by the transcriptional repressor, NMR, thereupon achieving tight control of the expression of NCR genes.[12]

Interestingly a yeast homolog, L-serine dehydratase (MGG 06950) was found to have an upregulation with a 42.4 fold-change proteomic expression and a 37.3 fold-change in the genetic data. In yeast, this protein breaks down serine and threonine as an alternate nitrogen source. The fact that MGG 06950 was highly upregulated in the nitrogen starved condition for both the proteomic and genetic studies further supports that it plays a principal role in alternate nitrogen metabolism.[12] Additionally, this protein was unexpressed by the nitrogen treated samples in the proteomic experiment. MGG 06950 also contains 4 HGATAR binding motifs, further suggesting that MGG 06950 is regulated by the NCR system, and a key player in its pathway.[12]

Moreover, the NADP glutamate dehydrogenase (MGG 08074) was upregulated in response to nitrogen starvation. While a greater fold-change was observed in the genetic data, it was still significantly upregulated by a factor of 1.94 in the proteomic data. MGG 08074 generates ammonium through the conversation of 2-oxoglutarate to glutamate.[17] As a result, this protein could play a key role during nitrogen starvation by recycling ammonium from internal sources of *M. oryzae*.[17] Like MGG 06950, it also contains 3 HGATAR motifs, further supporting a role of this gene and protein during nitrogen catabolism and its potential regulation by the NCR system.[17]

Table 2.1. Comparison of expression levels of 55 most upregulated genes from transcriptome study with protein expression levels

Accession	Proteomic Fold-Change	P Value (Proteomic)	Genetic Fold-Change	Putative Identity
General Metabolism				
MGG 06950	42.4	0.020	37.3	L-Serine Dehydratase
MGG 07933	1.1	0.113	20.9	Dihydrodipicolinate Synthase
MGG 07935	-1.1	0.151	16.7	Galactonate Dehydratase
MGG 09297	2.0	0.007	15.4	Hypothetical Protein/COG0604
Nitrogen Metabolism				
MGG 08074	1.9	5.39E-5	25	NADP-Specific Glutamate Dehydrogenase
MGG 03494	1.3	6.07E-4	19.4	Aminotransferase
MGG 06888	1.2	0.800	16.4	Glutamine Synthetase
Carbohydrate Metabolism				
MGG 05828	2.0	0.058	50.2	Chitin Deacetylase
MGG 02544	2.8	0.012	21.9	D-Arabinitol Dehydrogenase
MGG 03287	-2.7	6.88E-4	11.4	α-Amylase
General Transport				
MGG 02072	-4.3	0.002	33.2	Amino Acid Permease
MGG 10800	5.6	3.38E-4	25.9	Sarcosine Oxidase
MGG 01925	1.8	0.004	14.2	Bifunctional P450:NADPH-P450 Reductase
MGG 09063	15.5	2.31E-4	12.5	Urea Active Transporter
Proteolysis				
MGG 08736	-30.1	0.028	14.7	Carboxypeptidase
Sporulation				
MGG 02612	1.6	4.97E-4	18.8	3-Oxoacyl-Reductase
Unknown Process				
MGG 08589	242.3	5.13E-5	25.0	Predicted Protein
MGG 08072	1.7	0.027	21.7	Cholesterol Oxidase
MGG 10751	-2.2	0.315	20.3	Peroxisomal Copper Amine Oxidase
MGG 02654	1.2	0.509	17.3	Predicted Protein
MGG 01569	1.5	1.26E-4	11.7	Minor Allergen Alt A

An additional comparison was made between gene and protein expression with potential involvement in nitrogen-regulator pathways. In the transcriptomic study, 23 genes were highlighted as differentially expressed that played a role in nitrogen-regulatory pathways, and 75% (15/20) of the proteins associated with these genes were detected. Fourteen proteins were able to be used in relative quantification with 57% of the proteins showing an upregulation in the nitrogen starved condition, 14% of proteins showing no significant change, and 29% of proteins being downregulated in the nitrogen starved group. This comparison is highlighted in Table 2. Proteins and genes that were both significantly upregulated are highlighted in green, no change shown in black and proteins that were downregulated but genes upregulated are shown in red (downregulated being shown as the negative fold-change).

The NUT1 protein is a principal GATA nitrogen regulator whose expression was upregulated by nitrogen starvation in both the transcriptome and proteomic experiments. In the genetic experiment, NUT1 was also detected *in planta* 5 days after infecting rice seedlings by reverse transcriptase polymerase chain reaction.[17] It has been shown that when the NUT1 gene is disrupted, *M. oryzae* cannot use nitrate, nitrite, formamide, histidine or uric acid as nitrogen sources, suggesting that these pathways are controlled by NUT1 through the NCR mechanism.[102] However, these mutants maintained wild-type growth when presented with proline, glutamate and alanine as nitrogen sources, which implies an additional key nitrogen-regulatory gene.[102] The gene that encodes the nitrogen-regulatory protein OTam possess a high degree of homology to the protein TamA in *Aspergillus nidulans*, which is known to influence the transcription of nitrogen metabolic genes.[17] In the genetic study, MGG 06492 was shown to be upregulated in the nitrogen starved samples, and it was also

expressed in the proteomic study. It is possible that OTam is another key regulator of the NCR system and would be an interesting target for future studies.[17]

Table 2.2. A comparison of differentially expressed genes in nitrogen-regulatory pathways with proteomic expressions

Accession	Proteomic Fold-Change	P Value (Proteomic)	Genetic Fold-Change	Putative Identity
Global Nitrogen Regulation				
MGG 06492	1.5	0.117	11	Nitrogen Regulatory Protein, OTam
MGG 02755	3.0	0.005	1.8	Nitrogen Regulatory Protein, NUT1
Alanine				
MGG 02525	-1.5	0.002	2.7	Alanine-Glyoxylate Aminotransferase
Allantoin				
MGG 10412	2.0	0.0124	11.1	Allantoinase
Histidine				
MGG 08802	-1.4	7.89E-4	3.6	GMP Synthase
Nitrate/Nitrite				
MGG 06062	-18.9	0.007	2.1	Nitrate Reductase
MGG 00634	-829.7	0.005	3.6	Nitrite Reductase
Proline				
MGG 04244	5.6	4.17E-4	8.2	Hypothetical Protein/ Proline Oxidase
MGG 02899	1.3	0.032	2.1	Proline-Specific Permease
Uric Acid				
MGG 09063	15.5	2.31E-4	12.5	Urea Active Transporter
MGG 01324	2.1	0.013	2.1	Urease
MGG 08056	7.8	0.002	1.9	Uric Acid-Xanthine Permease
Glutamate/Glutamine				
MGG 08074	1.9	5.39E-5	25.8	NADP-Specific Glutamate Dehydrogenase
MGG 06888	1.2	0.800	16.4	Glutamine Synthetase

Given that both gene and protein expressions were observed at two very different time points yields an additional level of confirmation in their putative roles of nitrogen regulation and catabolism. Indeed, 38% of the most up-regulated genes also exhibited protein expression and 70% of genes related to nitrogen regulatory pathways also expressed their corresponding proteins. Within this shared group, 60% of the genes and proteins were both upregulated and 40% of the genes and proteins either showed no significant fold-change or were downregulated. While 40% of genes and proteins expressed did not show the same direction of fold-change, several studies have highlighted the fact expression levels between gene and proteins are not always correlated.[103-105] Having 60% of both genes and proteins being both upregulated is an excellent correlation between two different systems biology approaches, with these global explorations setting a basis for a more in-depth analysis of the overlapping data. Thus the proteomic study presented is highly complementary to the transcriptome study of the effects of nitrogen starvation on mycelia development and the infection process of the fungus, *Magnaporthe oryzae.*

2.4 Conclusion

The filamentous fungus *M. oryzae* causes annual crop losses in rice of 10-30%, while epidemics can lead to even greater losses. By gaining a greater understanding of its infection process and pathogenicity, progress is made towards greater food stability for Asian countries where rice is a staple food. This investigation employed global proteomics to ascertain changes in protein expression levels in response to the stress of nitrogen starvation. *M. oryzae* was grown in a rich culture, the mycelium was harvested and then grown in nitrogen starved or nitrogen treated media for 12 hours. A FASP digestion was utilized and

samples were analyzed with a Q-Exactive HF mass spectrometer in a data dependent acquisition.

First, excellent precision in the measurements was demonstrated on a per-injection basis by showing percent RSDs at 10% or less in parameters of the number of scans, proteins identified and spectral counts in addition to good chromatographic reproducibility. The spectral counts were normalized with the total spectral count normalization applied to the technical replicates, biological replicates and across the two different experimental conditions of nitrogen treated vs nitrogen starved. The spectral counts were used to calculate the differential protein expression and in the nitrogen starved condition, 271 proteins were significantly ($p < 0.05$) upregulated and 185 proteins downregulated with a minimum of a 2-fold change.

Noteworthy differences were observed in protein enrichment between the nitrogen-scarce and nitrogen-treated conditions. In the nitrogen-rich samples, *M. oryzae* seemed to dedicate most of its resources to active growth and proliferation as the greatest enrichment was observed in protein transcription, translation and putative cell wall remodeling and biosynthesis. The nitrogen-limited samples appeared to be expressing greater numbers of proteins related to inducing infection such as its response to cAMP, host secretion interactions and polyketide machinery. Moreover, the nitrogen starved samples also expressed enrichment in proteolysis, implicating its role in energy recycling in a nutrient-limited environment.

In addition to protein enrichment, the proteomic data was compared to a previous global transcriptome study of nitrogen starvation in *M. oryzae*. Overall, similar trends were observed between the two studies. From 55 of the highest upregulated genes, 38% of their

correlating proteins could be quantified. For the genes that were investigated in nitrogen-regulated pathways, 60% of their proteins were also quantified. Within this shared group, 60% of genes and proteins were both upregulated. Proteins such as L-Serine dehydratase, NADP glutamate dehydrogenase, and NUT-1 were highlighted as they have been implicated in nitrogen scavenging and regulation, and they were both upregulated in the genetic and proteomic studies. By employing state-of-the-art mass spectrometry methods and technology, this study has made a novel contribution to understanding the role of nitrogen starvation in pathogenicity and nitrogen regulation in *M. oryzae*. This effort aids in the progression of learning about *M. oryzae*'s biological processes in order to better combat its infections, and gain a better stability in the staple crop of rice.

REFERENCES

1 Loginov, D. & Šebela, M. Proteomics of survival structures of fungal pathogens. *New Biotechnology*, doi:http://dx.doi.org/10.1016/j.nbt.2015.12.011.

2 Talbot, N. J. On the trail of a cereal killer: Exploring the biology of Magnaporthe grisea. *Annu Rev Microbiol* **57**, 177-202, doi:10.1146/annurev.micro.57.030502.090957 (2003).

3 Franck, W. L., Gokce, E., Oh, Y. Y., Muddiman, D. C. & Dean, R. A. Temporal Analysis of the Magnaporthe Oryzae Proteome During Conidial Germination and Cyclic AMP (cAMP)-mediated Appressorium Formation. *Mol Cell Proteomics* **12**, 2249-2265, doi:DOI 10.1074/mcp.M112.025874 (2013).

4 Wilson, R. A. & Talbot, N. J. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. *Nat Rev Micro* **7**, 185-195 (2009).

5 Ribot, C. *et al.* Susceptibility of rice to the blast fungus, Magnaporthe grisea. *J Plant Physiol* **165**, 114-124, doi:10.1016/j.jplph.2007.06.013 (2008).

6 Valent, B. Rice Blast as a Model System for Plant Pathology. *Phytopathology* **80**, 33-36, doi:Doi 10.1094/Phyto-80-33 (1990).

7 Dean, R. A. *et al.* The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature* **434**, 980-986, doi:http://www.nature.com/nature/journal/v434/n7036/suppinfo/nature03449_S1.html (2005).

8 Sequencing ProjectInternational Rice, G. The map-based sequence of the rice genome. *Nature* **436**, 793-800, doi:http://www.nature.com/nature/journal/v436/n7052/suppinfo/nature03895_S1.html (2005).

9 Jeon, J. *et al.* Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nat Genet* **39**, 561-565, doi:http://www.nature.com/ng/journal/v39/n4/suppinfo/ng2002_S1.html (2007).

10 Koeck, M., Hardham, A. R. & Dodds, P. N. The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cellular microbiology* **13**, 1849-1857, doi:10.1111/j.1462-5822.2011.01665.x (2011).

11 Talbot, N. J., McCafferty, H. R. K., Ma, M., Moore, K. & Hamer, J. E. Nitrogen starvation of the rice blast fungus Magnaporthe grisea may act as an environmental cue for disease symptom expression. *Physiol Mol Plant P* **50**, 179-195, doi:DOI 10.1006/pmpp.1997.0081 (1997).

12 Donofrio, N. M., Mitchell, T. K. & Dean, R. A. in *Advances in Genetics, Genomics and Control of Rice Blast Disease* (eds Guo-Liang Wang & Barbara Valent) 59-72 (Springer Netherlands, 2009).

13 Talbot, N. J., Ebbole, D. J. & Hamer, J. E. Identification and Characterization of Mpg1, a Gene Involved in Pathogenicity from the Rice Blast Fungus Magnaporthe-Grisea. *Plant Cell* **5**, 1575-1590, doi:DOI 10.1105/tpc.5.11.1575 (1993).

14 Froeliger, E. H. & Carpenter, B. E. NUT1, a major nitrogen regulatory gene in Magnaporthe grisea, is dispensable for pathogenicity. *Molecular & General Genetics* **251**, 647-656, doi:Doi 10.1007/Bf02174113 (1996).

15 Lau, G. K. K. & Hamer, J. E. Regulatory genes controlling MPG1 expression and pathogenicity in the rice blast fungus Magnaporthe grisea. *Plant Cell* **8**, 771-781 (1996).
16 Soundararajan, S. *et al.* Woronin Body Function in Magnaporthe grisea Is Essential for Efficient Pathogenesis and for Survival during Nitrogen Starvation Stress. *The Plant Cell* **16**, 1564-1574, doi:10.1105/tpc.020677 (2004).
17 Donofrio, N. M. *et al.* Global gene expression during nitrogen starvation in the rice blast fungus, Magnaporthe grisea. *Fungal Genet Biol* **43**, 605-617, doi:10.1016/j.fgb.2006.03.005 (2006).
18 Wang, Y. *et al.* Comparative Secretome Investigation of Magnaporthe oryzae Proteins Responsive to Nitrogen Starvation. *J Proteome Res* **10**, 3136-3148, doi:10.1021/pr200202m (2011).
19 Raman, V. *et al.* Physiological stressors and invasive plant infections alter the small RNA transcriptome of the rice blast fungus, Magnaporthe oryzae. *Bmc Genomics* **14**, 1-18, doi:10.1186/1471-2164-14-326 (2013).
20 Yang, J. *et al.* Expression of Magnaporthe oryzae genes encoding cysteine-rich proteins secreted during nitrogen starvation and interaction with its host, Oryza sativa. *Genetics and Molecular Research* **14**, 17099-17108, doi:10.4238/2015.December.16.10 (2015).
21 X. Zhou, P. Y., C. Dong, C.X. Yao, Y.-M. Ding, N. Tao, Z.W. Zhao. Proteomic Analysis of mycelial proteins from *Magnaporthe oryzae* under nitrogen starvation. *Genetics and Molecular Research* **15**, doi: 10.4238/gmr.15028637 (2016).
22 Kenyon, G. L. *et al.* Defining the Mandate of Proteomics in the Post-Genomics Era: Workshop Report: ©2002 National Academy of Sciences, Washington, D.C., USA. Reprinted with permission from the National Academies Press for the National Academy of Sciences. All rights reserved. The original report may be viewed online at http://www.nap.edu/catalog/10209.html. *Mol Cell Proteomics* **1**, 763-780 (2002).
23 Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **314**, 65-66, doi:10.1126/science.1133987 (2006).
24 Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol Cell Proteomics* **3**, 608-614, doi:10.1074/mcp.T400003-MCP200 (2004).
25 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-U360, doi:Doi 10.1038/Nmeth.1322 (2009).
26 Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856-2860, doi:10.1038/nprot.2006.468 (2006).
27 Erde, J., Loo, R. R. O. & Loo, J. A. Enhanced FASP (eFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments. *J Proteome Res* **13**, 1885-1895, doi:10.1021/pr4010019 (2014).
28 Waas, M. *et al.* Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins. *Anal. Chem.* **86**, 1551-1559, doi:10.1021/ac403185a (2014).

29 Klammer, A. A. & MacCoss, M. J. Effects of Modified Digestion Schemes on the Identification of Proteins from Complex Mixtures. *J Proteome Res* **5**, 695-700, doi:10.1021/pr050315j (2006).
30 Chen, E. I., Cociorva, D., Norris, J. L. & Yates, J. R. Optimization of Mass Spectrometry-Compatible Surfactants for Shotgun Proteomics. *J Proteome Res* **6**, 2529-2538, doi:10.1021/pr060682a (2007).
31 Wu, F., Sun, D., Wang, N., Gong, Y. & Li, L. Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. *Anal Chim Acta* **698**, 36-43, doi:http://dx.doi.org/10.1016/j.aca.2011.04.039 (2011).
32 Chen, E. I., McClatchy, D., Park, S. K. & Yates Iii, J. R. Comparisons of Mass Spectrometry Compatible Surfactants for Global Analysis of the Mammalian Brain Proteome. *Anal. Chem.* **80**, 8694-8701, doi:10.1021/ac800606w (2008).
33 Masuda, T., Tomita, M. & Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *J Proteome Res* **7**, 731-740, doi:10.1021/pr700658q (2008).
34 Lin, Y. *et al.* Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal Biochem* **377**, 259-266, doi:http://dx.doi.org/10.1016/j.ab.2008.03.009 (2008).
35 Zhou, J. *et al.* Evaluation of the Application of Sodium Deoxycholate to Proteomic Analysis of Rat Hippocampal Plasma Membrane. *J Proteome Res* **5**, 2547-2553, doi:10.1021/pr060112a (2006).
36 León, I. R., Schwämmle, V., Jensen, O. N. & Sprenger, R. R. Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Mol Cell Proteomics* **12**, 2992-3005, doi:10.1074/mcp.M112.025585 (2013).
37 McCarthy, J. *et al.* Carbamylation of proteins in 2-D electrophoresis - Myth or reality? *J Proteome Res* **2**, 239-242, doi:Doi 10.1021/Pr025564b (2003).
38 Stark, G. R. Reactions of Cyanate with Functional Groups of Proteins. III. Reactions with Amino and Carboxyl Groups*. *Biochemistry-Us* **4**, 1030-1036, doi:10.1021/bi00882a008 (1965).
39 Sun, S., Zhou, J.-Y., Yang, W. & Zhang, H. Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal Biochem* **446**, 76-81, doi:http://dx.doi.org/10.1016/j.ab.2013.10.024 (2014).
40 Dickens, F. Interaction of halogenacetates and sh compounds. The reaction of halogenacetic acids with glutathione and cysteine. The mechanism of iodoacetate poisoning of glyoxalase. *Biochem J* **27**, 1141-1151 (1933).
41 Korman, S. & Clarke, H. T. Carboxymethylamino Acids and Peptides. *J Biol Chem* **221**, 113-131 (1956).
42 Boja, E. S. & Fales, H. M. Overalkylation of a protein digest with iodoacetamide. *Anal. Chem.* **73**, 3576-3582, doi:Doi 10.1021/Ac0103423 (2001).
43 Yamashita, M. & Fenn, J. B. Electrospray Ion-Source - Another Variation on the Free-Jet Theme. *J Phys Chem-Us* **88**, 4451-4459, doi:Doi 10.1021/J150664a002 (1984).

44 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **246**, 64-71, doi:DOI 10.1126/science.2675315 (1989).
45 Veenstra, T. D. Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions. *Biophysical Chemistry* **79**, 63-79, doi:http://dx.doi.org/10.1016/S0301-4622(99)00037-X (1999).
46 Cech, N. B. & Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **20**, 362-387, doi:10.1002/mas.10008 (2001).
47 Taylor, G. Disintegration of Water Drops in an Electric Field. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **280**, 383-397 (1964).
48 Taflin, D. C., Ward, T. L. & Davis, E. J. Electrified Droplet Fission and the Rayleigh Limit. *Langmuir* **5**, 376-384, doi:Doi 10.1021/La00086a016 (1989).
49 Iribarne, J. V. & Thomson, B. A. Evaporation of Small Ions from Charged Droplets. *J Chem Phys* **64**, 2287-2294, doi:Doi 10.1063/1.432536 (1976).
50 Dole, M., Mack, L. L. & Hines, R. L. Molecular Beams of Macroions. *J Chem Phys* **49**, 2240-2249, doi:Doi 10.1063/1.1670391 (1968).
51 Kingdon, K. H. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys Rev* **21**, 408-418, doi:Doi 10.1103/Physrev.21.408 (1923).
52 Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **72**, 1156-1162, doi:10.1021/ac991131p (2000).
53 Todd, J. F. J. & March, R. E. A retrospective review of the development and application of the quadrupole ion trap prior to the appearance of commercial instruments. *Int J Mass Spectrom* **190–191**, 9-35, doi:http://dx.doi.org/10.1016/S1387-3806(99)00065-2 (1999).
54 March, R. E. Quadrupole ion trap mass spectrometry: a view at the turn of the century. *Int J Mass Spectrom* **200**, 285-312, doi:http://dx.doi.org/10.1016/S1387-3806(00)00345-6 (2000).
55 Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom Rev* **17**, 1-35, doi:10.1002/(sici)1098-2787(1998)17:1<1::aid-mas1>3.0.co;2-k (1998).
56 Marshall, A. G. & Hendrickson, C. L. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *Int J Mass Spectrom* **215**, 59-75, doi:http://dx.doi.org/10.1016/S1387-3806(01)00588-7 (2002).
57 Hardman, M. & Makarov, A. A. Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source. *Anal. Chem.* **75**, 1699-1705, doi:10.1021/ac0258047 (2003).
58 Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430-443, doi:10.1002/jms.856 (2005).
59 Scigelova, M. & Makarov, A. in *Encyclopedia of Analytical Chemistry* (John Wiley & Sons, Ltd, 2006).
60 Michalski, A. *et al.* Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics* **10**, doi:10.1074/mcp.M111.011015 (2011).

61 Olsen, J. V. *et al.* A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics : MCP* **8**, 2759-2769, doi:10.1074/mcp.M900375-MCP200 (2009).
62 Scheltema, R. A. *et al.* The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol Cell Proteomics* **13**, 3698-3708, doi:10.1074/mcp.M114.043489 (2014).
63 Makarov, A., Denisov, E. & Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **20**, 1391-1396, doi:10.1016/j.jasms.2009.01.005 (2009).
64 Law, K. P. & Lim, Y. P. Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Rev Proteomic* **10**, 551-566, doi:10.1586/14789450.2013.858022 (2013).
65 Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *PROTEOMICS* **9**, 1683-1695, doi:10.1002/pmic.200800562 (2009).
66 Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *P Natl Acad Sci USA* **104**, 5860-5865, doi:10.1073/pnas.0608638104 (2007).
67 Randall, S. M., Cardasis, H. L. & Muddiman, D. C. Factorial Experimental Designs Elucidate Significant Variables Affecting Data Acquisition on a Quadrupole Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **24**, 1501-1512, doi:DOI 10.1007/s13361-013-0693-y (2013).
68 Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
69 Resing, K. A. & Ahn, N. G. Proteomics strategies for protein identification. *Febs Lett* **579**, 885-889, doi:http://dx.doi.org/10.1016/j.febslet.2004.12.001 (2005).
70 Nesvizhskii, A. I. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics* **73**, 2092-2123, doi:http://dx.doi.org/10.1016/j.jprot.2010.08.009 (2010).
71 Eng, J. K., Mccormack, A. L. & Yates, J. R. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989, doi:Doi 10.1016/1044-0305(94)80016-2 (1994).
72 Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567, doi:Doi 10.1002/(Sici)1522-2683(19991201)20:18<3551::Aid-Elps3551>3.0.Co;2-2 (1999).
73 Hoffman, M. D., Sniatynski, M. J. & Kast, J. Current approaches for global post-translational modification discovery and mass spectrometric analysis. *Anal Chim Acta* **627**, 50-61, doi:http://dx.doi.org/10.1016/j.aca.2008.03.032 (2008).
74 Zhu, W., Smith, J. W. & Huang, C.-M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology* **2010**, 6, doi:10.1155/2010/840518 (2010).

75 Ong, S.-E. *et al.* Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* **1**, 376-386, doi:10.1074/mcp.M200025-MCP200 (2002).
76 Gokce, E., Shuford, C. M., Franck, W. L., Dean, R. A. & Muddiman, D. C. Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation during Proteomic Workflows. *J. Am. Soc. Mass Spectrom.* **22**, 2199-2208, doi:10.1007/s13361-011-0237-2 (2011).
77 Collier, T. S. *et al.* Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression. *Rapid Commun Mass Sp* **25**, 2524-2532, doi:Doi 10.1002/Rcm.5151 (2011).
78 Old, W. M. *et al.* Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **4**, 1487-1502, doi:DOI 10.1074/mcp.M500084-MCP200 (2005).
79 Zybailov, B. *et al.* Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res* **5**, 2339-2347, doi:10.1021/pr060161n (2006).
80 Bolton, M. D. & Thomma, B. P. H. J. The complexity of nitrogen metabolism and nitrogen-regulated gene expression in plant pathogenic fungi. *Physiol Mol Plant P* **72**, 104-110, doi:10.1016/j.pmpp.2008.07.001 (2008).
81 Russell, D. H. & Edmondson, R. D. High-resolution mass spectrometry and accurate mass measurements with emphasis on the characterization of peptides and proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Mass Spectrom* **32**, 263-276, doi:Doi 10.1002/(Sici)1096-9888(199703)32:3<263::Aid-Jms501>3.3.Co;2-T (1997).
82 Wang, Y. D. & Cu, M. The Concept of Spectral Accuracy for MS. *Anal. Chem.* **82**, 7055-7062, doi:Doi 10.1021/Ac100888b (2010).
83 Wenger, C. D., McAlister, G. C., Xia, Q. & Coon, J. J. Sub-part-per-million Precursor and Product Mass Accuracy for High-throughput Proteomics on an Electron Transfer Dissociation-enabled Orbitrap Mass Spectrometer. *Mol Cell Proteomics* **9**, 754-763, doi:10.1074/mcp.M900541-MCP200 (2010).
84 Collier, T. S. *et al.* Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for Quantitative Proteomics. *Anal. Chem.* **82**, 8696-8702, doi:10.1021/ac101978b (2010).
85 Franck, W. L., Gokce, E., Oh, Y., Muddiman, D. C. & Dean, R. A. Temporal Analysis of the Magnaporthe Oryzae Proteome During Conidial Germination and Cyclic AMP (cAMP)-mediated Appressorium Formation. *Mol Cell Proteomics* **12**, 2249-2265, doi:10.1074/mcp.M112.025874 (2013).
86 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).
87 Hosack, D. A., Dennis, G., Sherman, B. T., Lane, H. C. & Lempicki, R. A. Identifying biological themes within lists of genes with EASE. *Genome Biol* **4**, doi:Artn R70
Doi 10.1186/Gb-2003-4-10-R70 (2003).
88 Li, Z. & Srivastava, P. in *Current Protocols in Immunology* (John Wiley & Sons, Inc., 2001).

89 Tiwari, S., Thakur, R. & Shankar, J. Role of Heat-Shock Proteins in Cellular Function and in the Biology of Fungi. *Biotechnology Research International* **2015**, 11, doi:10.1155/2015/132635 (2015).
90 Wu, J., Wang, M., Zhou, L. & Yu, D. Small heat shock proteins, phylogeny in filamentous fungi and expression analyses in Aspergillus nidulans. *Gene* **575**, 675-679, doi:http://dx.doi.org/10.1016/j.gene.2015.09.044 (2016).
91 Yáñez, M., Gil-Longo, J. & Campos-Toimil, M. in *Calcium Signaling* (ed Shahidul Md Islam) 461-482 (Springer Netherlands, 2012).
92 Nguyen, Q. B. *et al.* Systematic functional analysis of calcium-signalling proteins in the genome of the rice-blast fungus, Magnaporthe oryzae, using a high-throughput RNA-silencing system. *Molecular Microbiology* **68**, 1348-1365, doi:10.1111/j.1365-2958.2008.06242.x (2008).
93 Liu, X.-H., Lu, J.-P., Dong, B., Gu, Y. & Lin, F.-C. Disruption of MoCMK1, encoding a putative calcium/calmodulin-dependent kinase, in Magnaporthe oryzae. *Microbiological Research* **165**, 402-410, doi:http://dx.doi.org/10.1016/j.micres.2009.08.007 (2010).
94 Smith, C. A. Structure, Function and Dynamics in the mur Family of Bacterial Cell Wall Ligases. *J Mol Biol* **362**, 640-655, doi:http://dx.doi.org/10.1016/j.jmb.2006.07.066 (2006).
95 Munshi, T. *et al.* Characterisation of ATP-Dependent Mur Ligases Involved in the Biogenesis of Cell Wall Peptidoglycan in Mycobacterium tuberculosis. *Plos One* **8**, e60143, doi:10.1371/journal.pone.0060143 (2013).
96 Cossins, E. A. & Chen, L. Folates and one-carbon metabolism in plants and fungi. *Phytochemistry* **45**, 437-452, doi:http://dx.doi.org/10.1016/S0031-9422(96)00833-3 (1997).
97 Giraldo, M. C. *et al.* Two distinct secretion systems facilitate tissue invasion by the rice blast fungus Magnaporthe oryzae. *Nat Commun* **4**, doi:10.1038/ncomms2996 (2013).
98 Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat Prod Rep* **29**, 1050-1073, doi:10.1039/c2np20019h (2012).
99 Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **26**, 90-114 (2009).
100 Nemes, P., Barton, A. A., Li, Y. & Vertes, A. Ambient molecular imaging and depth profiling of live tissue by infrared laser ablation electrospray ionization mass spectrometry. *Anal. Chem.* **80**, 4575-4582, doi:Doi 10.1021/Ac8004082 (2008).
101 Lilly, W. W., Wallweber, G. J. & Higgins, S. M. Proteolysis and amino acid recycling during nitrogen deprivation inSchizophyllum commune. *Current Microbiology* **23**, 27-32, doi:10.1007/bf02092305 (1991).
102 Froeliger, E. H., Carpenter, B. E. & Froeliger, E. NUT1, a major nitrogen regulatory gene inMagnaporthe grisea, is dispensable for pathogenicity. *Molecular and General Genetics MGG* **251**, 647-656, doi:10.1007/bf02174113 (1996).
103 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:http://www.nature.com/nature/journal/v473/n7347/abs/10.1038-nature10098-unlocked.html#supplementary-information (2011).

104 Vogel, C. *et al.* Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular Systems Biology* **6**, 400-400, doi:10.1038/msb.2010.59 (2010).
105 Lundberg, E. *et al.* Defining the transcriptome and proteome in three functionally different human cell lines. *Molecular Systems Biology* **6**, 450-450, doi:10.1038/msb.2010.106 (2010).