IDENTIFICATION OF PROTEIN VACCINE CANDIDATES USING COMPREHENSIVE PROTEOMIC ANALYSIS STRATEGIES

by

James G. Rohrbough

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF BIOCHEMISTRY & MOLECULAR BIOPHYSICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN BIOCHEMISTRY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
31 DEC 2007	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Identification Of Protein Vaccine Candidates Using Comprehensive Proteomic Analysis Strategies	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Arizona**	8. PERFORMING ORGANIZATION REPORT NUMBER **CIO8-0003**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **The Department of the Air Force AFIT/ENEL, Bldg 16 2275 D Street WPAFB, OH 45433**	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**188**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by JAMES G. ROHRBOUGH

entitled IDENTIFICATION OF PROTEIN VACCINE CANDIDATES USING COMPREHENSIVE PROTEOMIC ANALYSIS STRATEGIES

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of DOCTOR OF PHILOSOPHY

_____ Date: 10/15/07
Dr. Vicki H. Wysocki

_____ Date: 10/15/07
Dr. John Galgiani

_____ Date: 10/15/07
Dr. Roger Miesfeld

_____ Date: 10/15/07
Dr. Megan McEvoy

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____Date: 10/15/07

Dissertation Director: Dr. Vicki H. Wysocki

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation form or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _____

James G. Rohrbough

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. Paul Haynes for starting me on the path to my PhD, and Dr. Vicki Wysocki for leading me to its completion. Both have provided the guidance to initiate my work and the freedom to explore and truly make my research my own. I would also like to thank Dr. John Galgiani for his mentoring and support for my research, and the members of my dissertation committee Dr. Roger Miesfeld, Dr. Megan McEvoy and Dr. Elizabeth Vierling for their advice and guidance.

Additionally, I would like to thank the members of the Wysocki research group for encouragement, friendship, and helping me keep my sense of humor intact. Specifically, I thank Dr. Amy Hildebrand and Dr. Richard Beardsley for enlightening discussions and great memories; and Samanthi Wickramasekera, Forrest Helfrich, and Ashley Gucinski for their friendship and collaborative efforts.

I would also like to thank the members of the Valley Fever Center for Excellence including Lourdes Lewis, Kris Orsborn, Julie Simons, Lisa Shubitz, Tao Peng, and Bob Perrill for help and encouragement involving the study of that fascinating fungus known as *Coccidioides*. I also want to thank the UA Mass Spectrometry Facility and the UA Proteomics Consortium, especially Dr. Linda Breci, Dr. Arpad Somogyi, Dr. Todd Mize, and Dr. George Tsaprailis for helping to enlighten me in the ways of Mass Spectrometry and proteomics.

Finally, I would like to express my appreciation to the U.S. Air Force Academy and the Air Force Institute of Technology for providing me the support and the tools to succeed in the most difficult challenge I have ever undertaken.

DEDICATION

This work is dedicated to my sisters and “Col. Mom” whose associations to me are both biological and voluntary. Without your encouragement, love and understanding this entire process would not have been possible. Thanks for your unwavering support and not letting me give up on myself. I would also like to thank Dundee for reminding me daily that nothing in life is more important than a well-thrown Frisbee.

“Twenty years from now you will be more disappointed by the things that you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor. Catch the trade winds in your sails. Explore. Dream. Discover.”

- Mark Twain

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

LIST OF FIGURES

LIST OF FIGURES - *Continued*

LIST OF TABLES

ABSTRACT

Presented in this dissertation are proteomic analysis studies focused on identifying proteins to be used as vaccine candidates against Coccidioidomycosis, a potentially fatal human pulmonary disease caused by inhalation of a spore from the soil-dwelling pathogenic fungi *Coccidioides posadasii* and *C. immitis*. A method of tandem mass spectrometry data analysis using dual protein sequence search algorithms for increasing the total protein identifications from an analysis is described. This method was utilized in a comprehensive proteomic analysis of cell walls isolated from the dimorphic fungal pathogen *C. posadasii.* A strategy of tandem mass spectrometry-based protein identification coupled with bioinformatic sequence analysis was used to produce a list of protein vaccine candidates for further testing. A differential proteome analysis using stable isotope protein labeling was undertaken to identify vaccine candidate proteins that are more highly expressed in the spherule, or pathogenic phase, of *C. posadasii*. The results of these analyses are 9 previously undescribed protein vaccine candidates isolated from spherule cell walls that have sequence indications of extracellular association such as GPI anchors and N-terminal signal sequences and antigen potential based on homology to known antigenic or secreted proteins. An additional 14 proteins identified from spherule cell walls are potential vaccine candidates based on extracellular sequence predictions without any indications of antigenic potential. The stable isotope labeling study has identified 3 more proteins that are preferentially expressed in spherules and exhibit antigenic potential based on extracellular localization or homology to known antigenic proteins. Additionally, there were 5 unknown function proteins identified by

stable isotope labeling that are more highly expressed in spherules that may be good vaccine candidates but cannot be identified or localized by sequence analysis.

The dual algorithm protein identification method presented here is a new technique to address some common shortcomings associated with a proteomic analysis. The comprehensive proteomic analyses of *Coccidioides posadasii* presented here have provided new targets for Coccidioidomycosis vaccine development as well as insights into the proteome of this pathogen, such as the sequence comparison of *C.posadasii* proteins to human proteins, as well as a comprehensive analysis of predicted protein function in the *Coccidioides* proteome.

1 CHAPTER ONE: BACKGROUND AND SIGNIFICANCE

Partial content of this chapter has been published in:

Rohrbough, J.G., J. Galgiani, V.H. Wysocki; 2007. **The Application of Proteomic Techniques to Fungal Protein Identification and Quantification**. Annals of the New York Academy of Sciences 1111: 133-146.

1.1 Coccidioides and vaccine development

1.1.1 History of coccidioides spp. and coccidioidomycosis (Valley Fever)

Coccidioides immitis and *Coccidioides posadasii* (collectively referred to as *Coccidioides*) are dimorphic fungal pathogens of humans and other mammals. Both species are essentially morphologically identical and are not known to differ in pathogenicity. Infection with *Coccidioides* is normally manifested as a flu-like human pulmonary disease called coccidioidomycosis, or more commonly, Valley Fever. It is estimated that approximately 150,000 new human infections occur each year[1], of which 95% resolve with no or minimal medical intervention,[2] although approximately 5% of cases result in disseminated disease which can be fatal. Recovery from Valley Fever is associated with lifelong immunity to the disease, [1, 3] suggesting that creation of a human vaccine is biologically possible.

Infection by *Coccidioides* was first described by Argentinean physicians Alejandro Posada and Roberto Johann Wernicke in 1892 .[4, 5] The Posada-Wernicke Disease, characterized by a severe disseminated form of coccidioidomycosis termed coccidioidal

granuloma, was thought to be caused by a protozoan.[6] Initially, the only known form of the fungus was the pathogenic spherule form isolated from infected tissue. The soil dwelling form of the fungus, known as mycelia, was initially thought to be a contaminant of growth media. The correct identification of *Coccidioides immitis* (the only known species at the time) as a fungus was not made until 1900 by Ophüls and Moffitt.[7] Ophüls later identified *Coccidioides immitis* as a dimorphic pathogen when he made the link between mycelia and spherules.[8]

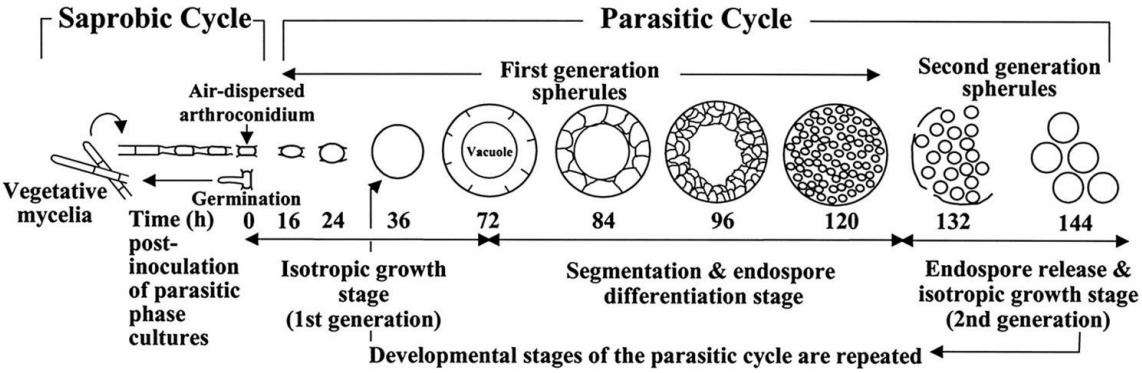
For many years, coccidioidal infection was believed to manifest itself as only the disseminated form of the disease. The link between coccidioidal granuloma and the relatively mild flu-like Valley Fever was not made until the accidental infection of a medical student working with a laboratory strain of *Coccidioides* in 1929.[9] A review of several cases by Drs Ernest Dickson and Myrnie Gifford led to the proposal of the term coccidioidomycosis to describe both the severe granuloma infection, as well as the relatively benign Valley Fever infection.[10, 11]

1.1.2 *Coccidioides* life cycle

The life cycle of *Coccidioides* consists of two phases, the saprobic soil-dwelling phase, and the parasitic phase into which the fungus differentiates upon entering a host mammalian lung, as shown in Figure 1.1. The saprobic or mycelial phase is characterized by growth of filamentous hyphae. The multinucleate hyphae elongate and separate, producing small arthroconidia spores with 1-3 nuclei[12, 13] which can become aerosolized and germinate to form more mycelia in the soil. Alternatively, if deposited in a suitable

host environment, they may begin differentiation into the parasitic cycle. Upon entering the host lung, a barrel-shaped arthroconidium remodels into a spherical cell, enlarges, undergoes multiple cycles of mitosis, undergoes repeated cell division and internal segmentation, resulting in the development of mature spherules containing scores of endospores.[14] Upon maturation, the mature spherules burst, releasing endospores that can regenerate spherules in the host. Upon the death of the host, fungal elements can return to the soil, and the saprobic phase can resume.[15] It is believed that spherule maturation is the same both *in vivo* and *in vitro.*[16]

Figure 1.1 The *Coccidioides* life cycle
(Figure reproduced with permission from [17] Delgado, N., J. Xue, J. J. Yu, et al. 2003. A recombinant beta-1,3-glucanosyltransferase homolog of Coccidioides posadasii protects mice against coccidioidomycosis. *Infect Immun* 71: 3010-3019.)



1.1.3 *Coccidioides* speciation

Based upon DNA sequence analysis, *Coccidioides* species are fungi in the phylum Ascomycota, class Euascomycetes, order Onygenales, and family Onygenaceae. Phylogenetic analysis has identified the non pathogenic fungus *Uncinocarpus reesii* as the

most closely related to *C. immitis/posadasii*, with the fungal pathogens *Histoplasma capsulatum* and *Blastomyces dermatitidis* also belonging to the family Onygenaceae.[18] Understanding these taxonomic relationships is helpful not only by providing sources of comparison to help understand the biology of *Coccidioides*, but also by identifying closely related species whose sequenced genomes can provide protein sequences for databases used in proteomic analyses (described later).

All strains of *Coccidioides* were designated *C. immitis* until 2002 when Fisher *et al.*[19] described a separate species (*C. posadasii*-named after Dr. Posada) based on sequence analysis of multiple strains. Today, it is believed that *C. immitis* is localized primarily to the San Joaquin Valley of California, and *C. posadasii* exists in the environment outside of California, including Arizona, New Mexico, Utah, Texas and Central/South America.

1.1.3.1 Coccidioides strains in common research use

There are multiple laboratory-maintained *Coccidioides* strains that are routinely used for research. These strains include Silveira, C735 and RS. Of these, Silveira has been used for the longest time, especially for vaccine development and other immunologic studies. Silveira was isolated by Friedman et al in 1951[20] from a human with non-disseminated disease who recovered within 4 years. Although originally isolated from a patient in California, Silveira is a strain of *C. posadasii*. Silveira is the strain used for all experiments performed and detailed in this dissertation.

Another common *C. posadasii* strain is C735. This strain was isolated by Yuan and Cole from a patient with disseminated coccidioidomycosis first reported in 1987.[21] C735 is

the *C. posadasii* strain used for genome sequencing by The Institute for Genomics Research (TIGR), now known as the J. Craig Venter Institute (http://www.tigr.org).

The other commonly used laboratory strain that will be discussed here is the *C. immitis* strain RS. RS was isolated by Henry and Ruth Walch , initially reported in 1967.[22] RS is the strain used by the Broad Institute for genome sequencing of *C. immitis* (http://www.broad.mit.edu). Predicted protein sequences from both sequenced genomes are used for proteomic analyses described in this dissertation.

1.1.4 Immunization against coccidioidomycosis

The understanding that coccidioidal infection produced immunity was first described in 1896 by Rixford after inoculating a dog from infected human tissue.[6] The link between human infection and subsequent immunity was cemented by Smith in 1940 with a comprehensive epidemiological study of 432 *Coccidioides*-infected patients from Kern and Tulare counties in California. In this study, only 2 of the 432 patients showed signs of a second coccidioidal infection, and these 2 cases were described as "clinical pictures which were not clear-cut".[23]

While coccidioidomycosis can be caused from infection by either species of *Coccidioides*, there has been some concern that immunity would be species or even strain specific.[24] However, Pappagianis has concluded that immunization with one strain of *Coccidioides* is effective in providing protection from infection by other strains and species.[25] This idea bodes well for the development of vaccine components by separate research groups utilizing different *Coccidioides* laboratory strains.

1.1.5 Vaccine development

Vaccine preparations produced from spherule cells have long been known to perform better than vaccines derived from mycelia cells. Studies of immunization in mice by Levine in 1960 identified spherule-based vaccinations as superior to mycelial vaccination. In that study, mice immunized with spherules exhibited a 3 percent mortality rate, compared to 28 percent in mice immunized with mycelia, and 75 percent mortality in the unvaccinated control animals.[26] This study and others indicate the importance of analyzing the pathogenic phase of the fungus in vaccine development research.

1.1.5.1 Whole cell vaccine

Studies to assess the protective effect of whole formalin-killed spherule (FKS) vaccine in humans was undertaken by Pappagianis, Levine and Smith in 1967.[27, 28] In these studies, sensitivity to coccioioidin (a response seen in individuals who have recovered from *Coccidioides* infection) was conferred to subjects that were initially negative, indicating an immune response to vaccination. In light of these results, field trials were undertaken in the early 1980s with a total of approximately 1400 humans that were inoculated with a killed spherule vaccine, versus approximately 1400 patients given a NaCl placebo. The results were a slight but statistically insignificant decrease in *Coccidioides* infection in the vaccinated patients with no difference in disease severity versus the control group.[29] It is believed that the amount of vaccine dose received by the human subjects was only 0.1% of the dose needed to immunize mice due to the local irritant effect of the vaccine injection.

Because of these findings, researchers have shifted the focus from whole cell to subcellular vaccine candidates.

1.1.5.2 Subcellular vaccines

A study by Kong et al showed the effectiveness of cell wall vaccines from formalin-killed spherules.[30] In this study, the spherules were disrupted and separated into cell-wall and soluble cellular fractions used to vaccinate mice. Mice vaccinated with the cell wall fraction displayed a 70% survival rate after 61 days, compared to a 10% survival rate of mice vaccinated with the soluble cellular fraction, and 0% survival of the unvaccinated control animals. This study, published in 1963, pushed the focus of *Coccidioides* vaccine research towards spherule cell-wall associated components.

1.1.5.2.1 Aqueous extraction

In an attempt to reduce the irritant effect of whole cell vaccines, Pappagianis and coworkers produced a phosphate buffered saline (PBS) extract of disrupted spherule cell walls that was tested in mice for protective effect.[31] When used as a vaccine in mice, the PBS extract produced an 80% survival rate versus a 10% survival of unvaccinated controls. While the results were encouraging, additional analysis of this subcellular vaccine has not been reported.

1.1.5.2.2 Alkali extraction

The Alkali-Soluble, Water-Soluble (ASWS) extract of mycelial cell walls was originally produced as a skin test for *Coccidioides* infection since it elicits an immune response in infected animals.[32] Further analysis of the extract indicated a potential use as a vaccine when it was discovered that mice immunized with ASWS had an 80% survival rate after 35 days of infection, versus a 10% survival rate in control mice.[33] Later work by Cox and Magee indicated that the ASWS derived from spherule cell walls was more protective than that from the mycelia.[34] Additional studies to identify the protective components of ASWS reported the presence of two immunogenic components,[35] one of which was Antigen 2 (Ag2), later identified as the Proline-Rich Antigen (PRA) as described below.

1.1.5.2.3 Spherule outer wall

When grown in vitro, cultured spherules produce a membranous spherule outer wall (SOW) component that is shed into the liquid media.[36] Observations in culture suggest that the antigenic components of the SOW become concentrated as the cells mature. This SOW component was shown to be immunoreactive against serum from infected human patients[37], and contains SOW glycoprotein (SOWgp), Antigen-2 (Ag2) and the Coccidioides-Specific Antigen (CSA) that are discussed in detail below. It has been suggested that the SOW may function as a protective barrier against host defense upon infection.

1.1.5.2.4 T27K

The irritant effects of the whole cell FKS vaccine lead researchers to look at a soluble, aqueous fraction called 27K in which the killed spherules are disrupted and centrifuged at 27,000 × g.[38] When analyzed for protection in mice, 27K proved as effective as the FKS vaccine. Attempts to fractionate 27K into individual protective components was unsuccessful,[39] so a new preparation was produced from thimerisol-killed (instead of formalin-killed) spherules to produce the T27K subcellular vaccine.[40] T27K is as protective as the original 27K, but is more amenable to fractionation attempts using gel filtration and anion exchange chromatography. Testing of some of these sub-fractions of T27K have shown some protective characteristics, however identification of individual protective components are ongoing. Two recent identifications by proteomic analyses of T27K, a superoxide dismutase (SOD) and an alpha mannosidase (Amn1) are described below. Another recent analysis of the T27K preparation using MS identification of proteins separated by 2-D gel electrophoresis found several of the vaccine antigens discussed below, including HSP-60, GEL1, ELI-Ag1 and Pmp1.[41]

1.1.5.3 Recombinant protein vaccine antigens

1.1.5.3.1 Ag2/PRA

Antigen 2 (Ag2), was first identified in 1978 by two-dimensional immunoelectrophoresis (IEP) from the crude antigen preparations coccidioidin and spherulin,[42] and also found in the alkali-soluble, water-soluble mycelial and spherule extracts.[43] Ag2 was not sequenced until 1996.[44] A parallel antigen discovery of a proline-

rich antigen (PRA) from a toluene spherule lysate was made in 1991.[45] PRA was also sequenced in 1996,[46] at which point it was discovered Ag2 and PRA were the same protein.

A recent study of truncated recombinant versions of Ag2/PRA[47] demonstrated the effectiveness of the first 106 residues of this 194 amino acid protein. The recombinant peptide consisting of residues 1-106 was as protective in vaccinated mice as the full-length protein. These findings allow for the substitution of the full length protein in antigen preparations which reduces the total amount of protein in a vaccine dose, and improves the likelihood that a recombinant chimeric protein consisting of antigenic portions of multiple proteins would provide an effective vaccine.

1.1.5.3.2 CSA

The *Coccidioides*-specific antigen (CSA) was first identified by Kaufman and Standard in 1978 as an exoantigen found in all tested *Coccidioides* strains, but not found in other fungal pathogens.[48] CSA has been isolated from both spherules,[49] and infectious arthroconidia[50] and was classified as a serine protease in 1987.[21] It was not until 1995 that the protein sequence for CSA was determined by sequencing of the N-terminal and proteolytically produced peptides, as well as sequence determination of isolated cDNA.[51] Further testing was not attempted until recently, when recombinant CSA combined with rAg2/PRA (the 1-106 fragment described above) produced in *Saccharomyces* was tested in mice[52]. This chimeric antigen preparation produced a 90% survival rate in mice after 60 days of infection, compared to a 30% survival rate with rCSA alone, and 60% with rAg2/PRA alone. All unvaccinated control animals were dead within 40 days (with 34 of

35 control mice gone within 20 days). This study also highlighted the effectiveness of a multivalent versus a univalent vaccine.

1.1.5.3.3 URE

The urease (URE) gene codes for a urea amidohydrolase protein that catalyzes urea hydrolysis. The protein was first isolated from *C. immitis* and characterized by Yu *et al.* from the Cole laboratory.[53] The gene was cloned in *E. coli*, and both the recombinant protein and the URE gene itself were tested for protection against coccidioidomycosis in mice. Vaccination with rURE protein produced an 80% survival rate in mice after 40 days of infection, compared to a 0% survival of control animals. The URE gene itself, given as an expression vector cDNA vaccine, was also somewhat effective in mice, producing a 40% survival rate after 40 days.[54] The high eukaryotic protein homology of rURE and the poor track record of DNA vaccines in humans,[15] has created some skepticism as to the effectiveness of URE protein or cDNA as vaccine candidates.[34] Recent work, however, suggests that URE contributes to the virulence of *Coccidioides*, perhaps by increasing the localized ammonia concentration and exacerbating the host inflammatory response[55]. This study also showed a 55% survival rate in mice after 60 days of infection with a URE knockout strain of *Coccidioides*. These results suggest that URE is perhaps more attractive as a candidate for gene suppression rather than direct use as a vaccine antigen.

1.1.5.3.4 GEL1

GEL1 was identified by the Cole laboratory from computational analysis of the *C. posadasii* genome.[17] Sequence analysis indicates that it is a β-1,3-glucanosyltransferase with a predicted N-terminal signal sequence and a predicted glycosylphosphatidylinositol (GPI) anchor site, all of which suggest a cell-wall association. This localization was confirmed by immunofluorescent staining which placed GEL1 on the exterior of the endospore cell wall. In the vaccination study, rGEL1 expressed in *E. coli* produced a 70% survival rate after 40 days in vaccinated mice. All unvaccinated control animals were dead by day 20. This study was followed by a more detailed analysis of the specific immune response of rGEL1-vaccinated mice.[56] The identification of GEL1 as a viable protein antigen was one of the first studies initiated using bioinformatic methods, pointing out the value of this technique for future vaccine candidate identifications.

1.1.5.3.5 ELI-Ag1

ELI-Ag1 is a protein identified from an expression library immunization (ELI) method[57]. In this method, the sequenced genome is divided into ten sub-libraries utilizing cDNA isolated from *C. posadasii* (strain Silveira), with each sub-library containing between 80-100 genes. Each sub-library was inserted into an expression vector and used to vaccinate mice. Two of the ten initial sub-libraries were shown to be protective in mice challenged with *Coccidioides* infection. The most protective initial sub-library was fractionated into 5 daughter libraries. This manner of sub-library isolation and fractionation continued until a single gene was isolated. The ELI-Ag1 cDNA that was

isolated produced a survival rate of 80% at 40 days when used to vaccinate mice. Analysis of the 224 amino acid gene product sequence indicates the presence of a GPI anchor and an N-terminal signal sequence. Work is currently underway to express ELI-Ag1 in a eukaryotic system for testing of the recombinant antigen. This study is another example of the value of bioinformatic techniques combined with biochemical isolation of potential vaccine antigens.

1.1.5.3.6 SOWgp

The Spherule Outer Wall glycoprotein (SOWgp) was isolated in 2000 from the SOW preparation described above (Section 1.1.5.2.3) and shown to be the major antigenic component of the SOW complex that is capable of eliciting both humoral (antibody) and cellular immune responses.[58] It was also shown that SOWgp was expressed only on the surface of immature spherules. Later, SOWgp was cloned and expressed in *E. coli,* and shown to contain proline- and aspartic acid-rich region repeats, as well as a GPI anchor and N-terminal signal sequence.[59] In the same study, SOWgp was putatively identified as an adhesin, based on its ability to bind host extracellular matrix proteins.

An analysis of recombinant SOWgp as a vaccine showed a "modest level of protection" in mice based on increased clearance of the fungus from the lung. The rSOWgp was used to create antiserum for western blots which were used to probe the expression levels of SOWgp in the parasitic phase of *Coccidioides*. The production of SOWgp was shown to be cyclic, with the highest expression levels occurring in presegmented spherules, and decreased levels during spherule maturation.[60] The control of

SOWgp levels was later shown to be modulated by activity of a metalloprotease known as MEP1,[61] which is capable of proteolytic digestion of SOWgp in vitro. Mice immunized with rSOWgp and then infected with a MEP1 knockout strain of *Coccidioides* had high survival rates (approx. 55%) compared to immunized mice infected with wild-type *Coccidioides* or with a MEP1 knockout strain with restored MEP1 activity. MEP1 expression levels are also shown to increase during the early stages of endospore formation. These results suggest that *Coccidioides* is able to modulate the host immune response by the presence of SOWgp on the surface of maturing spherules, which elicits an ineffective humoral immune response, instead of the more effective cellular immune response. This idea is supported by recent work that analyzed the evolution of SOWgp proteins in both *C. posadasii* and *C. immitis* strains which suggests that this protein is under selective evolutionary pressure to allow the fungus to more efficiently evade host immune defenses.[62]

Since endospores are the only cell morphology of the pathogenic phase of *Coccidioides* that are small enough to be phagocytized by host immune cells, the removal of SOWgp proteins from the cell surface by MEP1 allow the endospores to escape detection during the short period of time they are susceptible to phagocytosis.[63] This theory also helps to explain the survival of phagocytized endospores seen by Drutz and Huppert in 1983.[16] Despite some success with vaccination, SOWgp is not considered a prime vaccine antigen target,[15, 34] however the studies described above highlight the potential usefulness of SOWgp/MEP1 function in understanding host immune response and the virulence of *Coccidioides* infection.

1.1.5.3.7 TCRP

The T-Cell Reactive Protein (TCRP) was first isolated by the Kirkland *et al.*[64] from an arthroconidia preparation known as the soluble conidial wall fraction (SCWF) that had previously been shown to elicit a T-cell mediated (cellular) immune response.[65] The protein was cloned and expressed in *E. coli* with predicted homology to mammalian 4-hydroxyphenylpyruvate dioxygenase which degrades phenylalanine to tyrosine.[66] The recombinant TCRP was subsequently tested for immunogenicity in mice,[67] where is was deemed to have a "modest protective effect". This protein's effect as a stand-alone vaccine antigen is limited, but may prove useful as part of a multivalent vaccine for its T-cell directed immunogenicity.

1.1.5.3.8 HSP-60

The heat shock protein HSP-60 of *Coccidioides immitis* was identified as a possible antigen based on homology to known bacterial and lower eukaryotic HSPs that have been shown to be immunoprotective antigens. In the fungal pathogen *Histoplasma capsulatum*, the HSP-60 homolog is a glycoprotein isolated from both cell wall and membrane fractions. Based on these observations, the HSP-60 of *Coccidioides* was cloned and expressed in *E. coli* and shown to elicit a T-cell immune response.[68] Despite these encouraging findings, testing of rHSP-60 as a vaccine in mice resulted in a disappointing 16% survival rate of vaccinated mice at 40 days post-infection (versus loss of all control animals by day 22).[54] Based on these results, HSP-60 is no longer considered a viable vaccine candidate.[15]

1.1.5.4 Antigen identifications by proteomic methods

While there are numerous examples of protein antigen identifications in *Coccidioides,* more modern proteomic analyses utilizing mass spectrometry (MS) have only recently been reported. These studies have primarily focused on identification of antigenic proteins. The analysis of T-cell reactive antigens associated with the spherule cell wall by 1 and 2-D electrophoresis protein separations followed by peptide identification via tandem MS (MS/MS) identified a protective aspartyl protease (Pep1).[69] Another analysis of seroreactive spherule cell wall proteins separated by 2-D electrophoresis and analyzed by MS/MS identified two more protective protein antigens, Phospholipase B (Plb) and a 1,2 Alpha-mannosidase (Amn1), in addition to Pep1, all of which were shown to be protective in mice as a multivalent recombinant protein vaccine.[70]

A 2-D DIGE analysis of differential protein expression between the mycelial and spherule phases of *C. posadasii*, resulted in the identification of a new vaccine candidate protein, a peroxisomal matrix protein known as Pmp1, also shown to be protective in mice against coccidioidal infection.[71] Immunoblot analysis of a 2-D gel of the thimerisol-inactivated spherule vaccine (T27K) was analyzed by MS, resulting in the identification of a putative Cu, Zn superoxide dismutase (SOD),[41, 72] as well as Amn1.[73, 74] While the protective effects of SOD have yet to be determined, homology to similar dismutases in other pathogenic fungi suggest a possible role in virulence.

Another study purified N-glycan containing glycoproteins from the T27K subcellular preparation by lectin affinity chromatography followed by SDS-PAGE separation.[75] From this study, a 60 kDa protein component was identified by MS, with homology to a 1,3

glucanosyltransferase from *C. posadasii* and other fungi. Further analysis of this protein is ongoing.

1.1.6 Future vaccine development

The future of vaccine development for coccidioidomycosis likely rests in the engineering of a multivalent protein vaccine.[34] Recent work detailed above by Tarcha and coworkers has shown the efficacy of a mixture of multiple recombinant proteins[70] as well as a promising chimeric antigen expressed in the Galgiani laboratory containing the sequences of two separate protein antigens.[52] The production of a vaccine for this human pathogen probably does not depend on the discovery of a single, as-yet unknown antigen, but rather depends on the concerted development of multiple immunoreactive components that provide protective immunity when administered together. To this end, efforts to identify multiple new protein antigens are likely to pay great rewards. It is with this goal in mind that the *Coccidioides* proteomic analyses described in this dissertation were performed.

1.2 Proteomics

With the increasing number of sequenced genomes, proteomics is a field of study that has expanded rapidly in the last decade to encompass a wide variety of techniques and technologies. While protein analysis is not new, many recent advances in bioinformatics as well as mass spectrometry have increased the speed and breadth of samples that can be efficiently analyzed. Older methodologies such as protein sequence

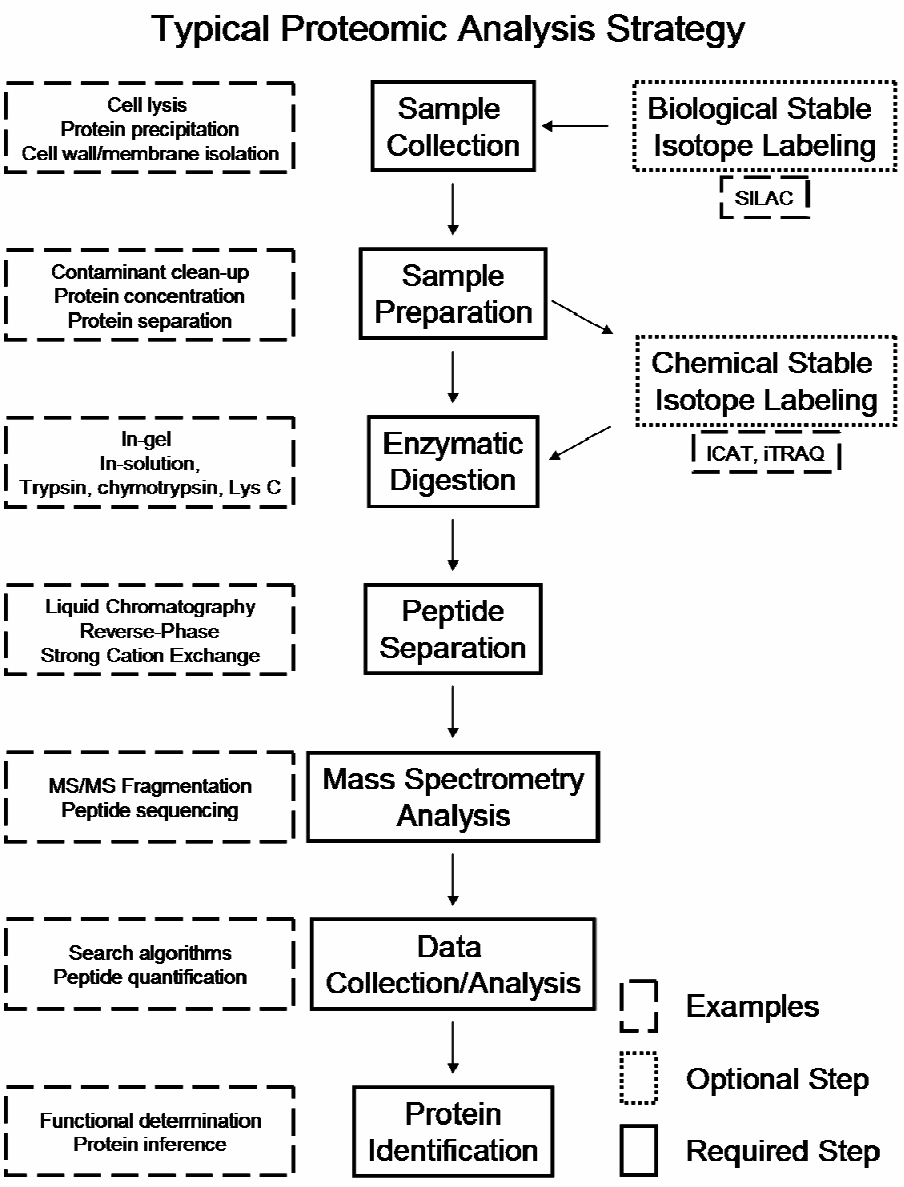
determination by Edman degradation, amino acid composition analysis or gel based analytical methods like western blots still have their place in specific research endeavors, but are somewhat less applicable to the high-throughput nature of modern proteomic analyses.

A proteomic analysis of a system involves the collection, separation, identification, and functional determination of the expressed proteins of a sample,[76] which can lead to a better understanding of protein function and regulation of a system. A detailed analysis of the proteome can then lead to protein targets for disease identification, treatment or vaccine development. The general steps of a proteomic analysis after the collection of a sample of interest are extraction of the proteins from the sample mixture, proteolytic digestion to produce peptides, peptide separation, peptide identification, and determination of identity and function of proteins present in the original sample. A brief overview of a typical proteomic analysis strategy is shown in Figure 1.2. The methods (sample preparation, ionization type, mass analyzer, etc) employed for a proteomic analysis may vary depending on the starting material and the goal of the analysis.

1.2.1 Protein analysis

There are many difficulties encountered in the analysis of proteins in a biological sample. The most obvious is the inherent complexity of many samples. Many proteins are present in locations that prohibit easy analysis, such as membrane-bound proteins that are difficult to solubilize. Very large and very small proteins can also be difficult to

Figure 1.2 Typical proteomic analysis strategy



analyze and detect. Not all proteins are present in equal abundance, a concept known as dynamic range.[77] This leads to a difficulty in detection of low-abundance proteins in the presence of highly abundant ones. Unlike RNA-based methods of transcript

amplification, sample protein levels cannot be increased to facilitate analysis of low abundance proteins. Any undertaking of a proteomic analysis will likely require addressing at least one of these difficulties.

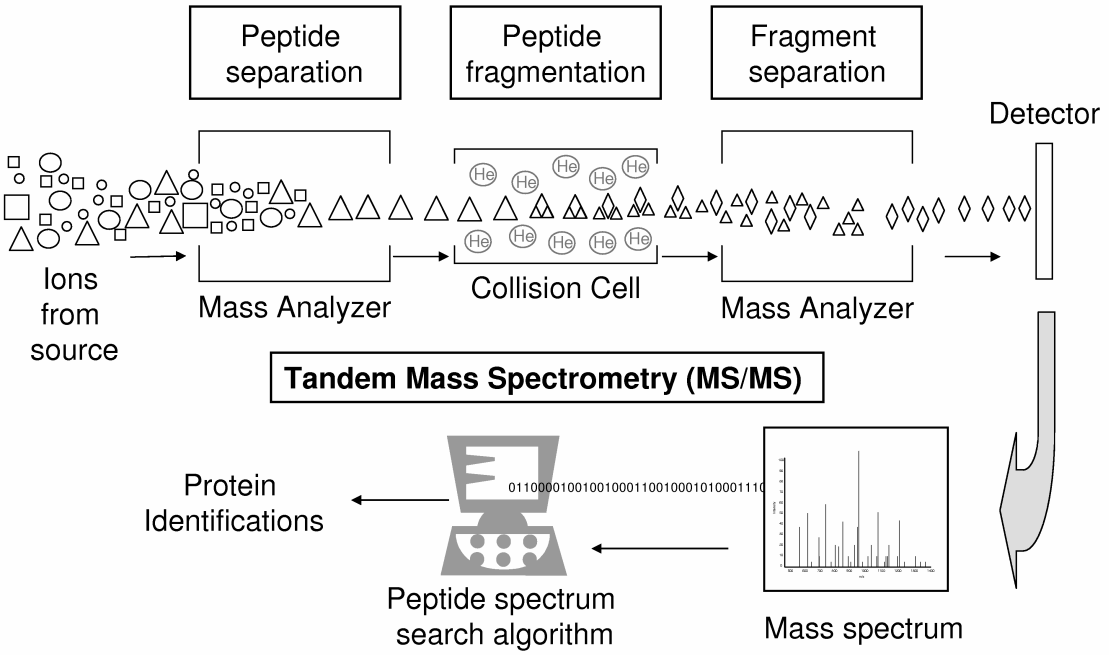
Recent advances in transcript identification may lead investigators to use mRNA analysis to infer protein presence. It is important to note, however, that while analysis of the mRNA levels of a system provides insights into gene expression, those levels may not correlate with protein abundance. Protein levels can vary as much as 30-fold relative to the mRNA levels coding for that protein.[78]

1.2.2 Mass spectrometry

The most common methods of analysis utilize mass spectrometry (MS) for protein and peptide identification. Sample proteins can be analyzed whole, in what is known as a top-down proteomic analysis,[79] or analyzed as peptides from protein digestion in a bottom-up approach. Top-down proteomics is less popular primarily due to the need for expensive and complex high-resolution mass spectrometers. Bottom-up methods can include peptide identification by high-resolution mass determination from a single round of MS, known as peptide mass fingerprinting,[80] or more commonly, peptides are identified by peptide fragmentation in a tandem mass spectrometry experiment (MS/MS) to facilitate amino acid sequence correlation.[81] In MS/MS, the peptides are separated by the mass analyzer and then subjected to fragmentation. The masses of the fragment ions are determined by a second mass analyzer (or in a second round of MS in a trapping-type instrument). Since different peptides will fragment differently, this

technique allows not only for the identification of peptides with different masses, but also those with the same or similar masses.[82, 83] An overview of the MS/MS process for identifying proteins from peptide fragmentation is shown in Figure 1.3. In the general

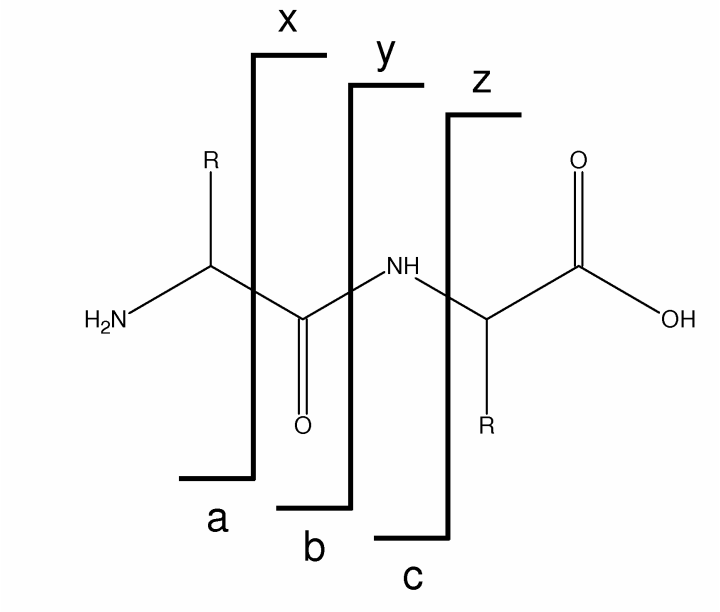
Figure 1.3 Overview of protein identification by tandem mass spectrometry



process, peptides are isolated from a mixture of ions in the first mass analyzer (or in the first round of MS in an ion-trapping type instrument), then transferred to a collision cell containing a background gas such as helium where the peptide ions collide with the gas molecules and fragment. The mixture of fragment ions is then transferred to a second mass analyzer (or the second round of MS in an ion trap) to separate the fragment ions prior to passing to a detector. The final protein identifications are made by computer search algorithm analysis of the tandem mass spectrum.

There are six common peptide fragment ion types produced by backbone cleavage as shown in Figure 1.4. If the fragmentation process results in the cleavage of the peptide bond, the N-terminal fragment ion is called a b ion, and the C-terminal ion is known as a y ion. Similarly, if the cleavage occurs between the alpha carbon and the carbonyl of the peptide backbone, a and x ions result from a charged N-terminal fragment and C-terminal

Figure 1.4 Peptide backbone fragmentation ions



fragment, respectively. Finally, cleavage of the peptide backbone between the alpha carbon and the amine results in the analogous c and z ions. The most common ions produced by collision induced dissociation with a background gas in an ion trap are the b and y ions.

1.2.3 Protein/peptide separation

Protein complexity can be reduced with some relatively simple methods of protein separation such as one (1-DE)[84] and two-dimensional (2-DE)[85] electrophoresis. These gels are run under denaturing conditions, including heat, detergent (such as SDS), and a reductant (such as Dithiothreitol or β-mercaptoethanol) for disulfide bond cleavage. In addition to reducing the sample complexity, gel electrophoresis can also be used for sample clean-up (removal of salts, detergents, etc). 1-DE involves the separation of denatured proteins based on size, while 2-DE starts with a separation of proteins by isoelectric point, followed by separation by size. A variation of the 2-DE method utilizing fluorescent dyes for protein quantification is known as Difference Gel Electrophoresis (DIGE), which will be discussed later (Section 1.2.7.4). After protein separation using electrophoresis, in-gel digestion is often used to produce and extract peptides prior to MS analysis.[86]

1.2.3.1 Liquid chromatography peptide separation

The complex peptide mixtures from a 1-DE gel or solution digest in a bottom-up experiment can be further separated by utilizing High Performance Liquid Chromatography (HPLC) on-line peptide separation methods. The most common peptide separation is known as reverse-phase (RP) LC. Using this method, the peptides are separated by hydrophobicity by eluting peptides bound to the RP packing material using an organic solvent (such as acetonitrile or methanol) gradient flow by HPLC.[87] Another LC separation method used is strong cation-exchange (SCX) which separates peptides by

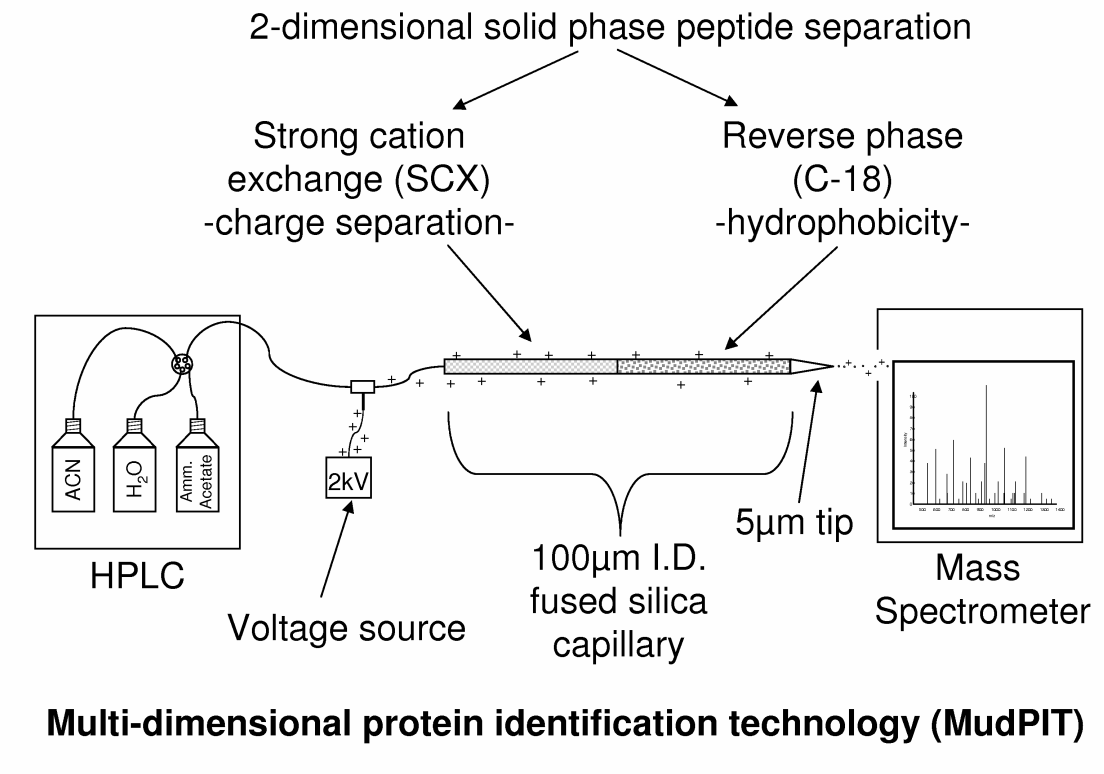
charge, where the peptides bound to the SCX material are eluted by a salt buffer. SCX is often used in conjunction with RP separation in a method known as MudPIT,[88] (Multi-dimensional Protein Identification Technology). MudPIT utilizes a column containing both RP and SCX chromatography phases described above, which allows for easier and automated analysis of biological mixtures, by reducing the complexity of peptides with a method that does not require protein separation by electrophoresis. A diagram of the MudPIT technique is shown in Figure 1.5. While MudPIT is often used to analyze solution-digested proteins, 1-DE has been used as a sample clean-up step, followed by MudPIT analysis.

1.2.4 Ionization methods

There are several types of ionization methods used for MS analysis, but there are two primary ones used in proteomics. Electrospray ionization (ESI),[89] and its closely-related small volume cousin nano-electrospray (nano-ESI),[90] involve injection of analyte (peptide or protein) molecules exiting the LC in solution into the mass spectrometer. A major advantage of ESI is the ease of coupling on-line separation methods such as RP and SCX prior to ionization and MS analysis. Also, ESI produces multiply-charged ions, allowing for identification of larger ions in an instrument with a low mass-to-charge (m/z) ratio limit.

The second major ionization method used in proteomics is matrix-assisted laser desorption/ionization (MALDI).[91, 92] The sample of interest is mixed with a matrix,

Figure 1.5 2-D Liquid chromatography peptide separation by MudPIT



spotted onto a sample plate, and then excited by a laser beam that ionizes and transfers analyte molecules into the gas phase for analysis. Advantages of MALDI ionization include a larger analyte mass range and higher tolerance of salts than ESI. Disadvantages include the cost of a laser-based system as well as interference from the background signal produced by the matrix.

1.2.5 Mass analyzers

After the protein or peptide molecules have been ionized and put into the gas phase, they enter the mass spectrometer for analysis. There are several methods of mass

analysis in mass spectrometry, but there are four major types utilized in proteomic analyses. All mass analyzers operate on the same basic principle of separation of ions by their mass to charge (m/z) ratio. The first type is the quadrupole mass analyzer, which utilizes four parallel metal rods that carry both a radio-frequency (RF) and a direct current (DC) voltage to produce a magnetic field to influence the path of ions. Manipulation of the settings for the RF and DC voltages allow for the selection of ions of a particular m/z ratio as they flow through the instrument. Ion identification is made by automated interpretation of the RF/DC settings of the quadrupole correlated to the time ions impact a detector. Quadrupole mass analyzers are some of the oldest and best defined and have been mainstays of MS analysis for decades. These reasons plus the relative simplicity of operation and more affordable cost are advantages associated with quadrupole mass analyzers. A major disadvantage associated with quadrupoles is the necessity of multiple mass analyzers to accomplish MS/MS, which increases both the complexity and cost of a system.

1.2.5.1 Ion traps

There are two common ion-trap mass analyzers based on the physics of the quadrupole mass analyzer. The first is the Quadrupole Ion Trap (QIT),[93] which utilizes a 3-dimensional ion trapping configuration that allows for trapping of all sample ions, followed by the selective release of ions for detection. The second common ion trap instrument is the newer Linear Ion Trap (LIT),[94] which is also based on the RF/DC combinations of the quadrupole, only in a 2-dimensional linear arrangement of the trap.

Like the QIT, the LIT traps all sample ions allowing for selective ion release to the detector. Another ion trap mass analyzer is the Fourier transform ion cyclotron resonance (FTICR, or just FT)[95] MS that utilizes a superconducting electromagnet for ion control. The benefits of ion-trapping mass analyzers include the ability to perform multiple rounds of MS and parent ion fragmentation within the same mass analyzer, as well as an increase in signal to noise ratio. Disadvantages of the QIT/LIT analyzers include limited resolution and mass accuracy. While the FT has excellent resolution and mass accuracy appropriate for top-down sequencing, it is relatively large and expensive and is more difficult to couple to LC-based sources.

1.2.5.2 Time of flight

The final common mass analyzer is the Time of Flight (TOF)[96] which is a much simpler system than either a quadrupole or FT-based mass analyzer. In the TOF, ions are separated by the time it takes them to travel the length of the analyzer, with the smaller m/z ratio ions impacting before larger ones. In addition to its simplicity of operation, the TOF mass analyzer is also valued for its enhanced m/z range, mass accuracy and resolution over quadrupole-based instruments. Disadvantages of TOF include the difficulty of coupling the pulsed analysis with continuous-ionization LC-based peptide separation, and the need for an additional mass analyzer to accomplish MS/MS. TOF instruments that are commonly used in proteomic analysis include the quadrupole coupled to a TOF (QTOF), or two TOF analyzers in sequence (TOF-TOF).

1.2.6 Data analysis

MS/MS spectra typically do not directly provide a peptide sequence. While it is possible to interpret MS/MS spectra for de novo sequence identification, the current level of understanding of peptide fragmentation is not advanced enough to make de novo sequencing as effective as spectrum matching. In spectrum matching, spectral information is matched to known peptide sequences and predicted fragment ion m/z values from protein sequence databases. Some of the more popular database collections include the National Center for Biotechnology Information (NCBI) which includes most of the public domain sequence databases, included in the non-redundant (NR) database. Another collection of sequences is the Swiss Prot database which has many sequences, but also has a large amount of functional annotation included with the sequences to allow for easier identification of functionality of listed proteins.[83] If the genome of the organism being analyzed has not been sequenced, the best strategy is to build a database of closely-related species, or search against the NR database, with the realization that the larger the database, the longer the search process will take, and the greater the rate of false positive (random) identifications. There are several different sources for sequenced fungal genomes. Among these are the Broad Institute (http://www.broad.mit.edu), the Sanger Institute (http://www.sanger.ac.uk), The Institute for Genomic Research (TIGR) (now known as the J. Craig Venter Institute) (http://www.tigr.org), and Génolevures (http://cbi.labri.fr/Genolevures).

1.2.6.1 Database search algorithms

Peptide sequences are matched to spectral information using a database search algorithm. Two of the most common licensed programs are SEQUEST[97] and Mascot.[98] A third, newer algorithm is the open-source XTandem.[99, 100] While each of these programs is in common use in proteomic research today, a recent evaluation[101] highlighted the strengths and weaknesses of these algorithms and also suggested the validity of using multiple search algorithms as a way of minimizing false positive identifications in a consensus approach. In this review of search algorithms, the authors found that the SEQUEST search algorithm was more sensitive than Mascot or XTandem which means it is better able to correctly identify spectra of poor quality. Mascot and XTandem on the other hand, were found to be more specific than SEQUEST, meaning that they do a better job of discriminating between correct and incorrect matches. The authors concluded that a consensus approach that pairs a more specific search algorithm (such as XTandem) with a more sensitive one (such as SEQUEST) is likely to reduce false positive protein identifications. An investigation of this consensus approach for identification of proteins from single peptide matches is presented in Chapter 2 of this dissertation. A similar method using SEQUEST and Mascot for validation of protein identifications is described by Resing *et al.*[102] In addition, there is a commercial software algorithm known as Scaffold (Proteome Software Inc.) that uses validation between three search algorithms: Mascot, SEQUEST and XTandem. It is important to note, however, is that neither of these methods focuses on single-peptide identifications, and both require additional software licenses.

1.2.6.1.1 Mascot

The Mascot search algorithm is a probability-based database search algorithm. It provides a calculation of the probability that a peptide sequence from the database being searched matches the experimental spectrum by chance. The algorithm analyzes each experimental spectrum using an iterative process to find the set of most intense fragment ion peaks that produce the highest Mascot score. The Mascot score is calculated from the random sequence match probability (P) using the equation: Mascot score = $-10\log_{10}P$. While the score is dependent on the protein length, a good score is typically 70 or greater.

1.2.6.1.2 SEQUEST

The SEQUEST search algorithm is what is known as a heuristic algorithm, which predicts the fragmentation spectrum of each peptide in the database (that matches the parent ion mass) and compares it with the experimental spectrum. SEQUEST analyzes the 200 most abundant ions in the experimental spectrum, divides the spectrum into 10 bins, and normalizes the relative intensity of the ions in each bin to 100. The program then compares the modified predicted and binned normalized experimental spectra and scores each predicted database spectrum based on criteria such as continuity of b and y ion series, presence of immonium ions for H, Y, W, M or F, as well as the total number of predicted ions found in the experimental spectrum. The fragment ions in the top 500 scoring theoretical spectra are then assigned an abundance of 50, 25 or 10. Ions corresponding to the b and y ion series are assigned abundances of 50, any ions within 1 mass unit of the b and y ions are assigned an abundance of 25, and ions corresponding to

water and ammonia losses of the b and y ions are assigned an abundance of 10. Each of the 500 theoretical spectra are then compared to the experimental spectrum and assigned a cross-correlation (XCorr) score. The peptide corresponding to the theoretical spectrum with the highest XCorr score is reported as the match. In the event that two different peptides match the spectrum almost as well, the result is considered ambiguous. An additional scoring parameter that compares the XCorr scores of the top two matching peptides is known as the ΔCn score. If the difference between the top two XCorr scores (the ΔCn) is below a user-set threshold (such as 0.1), the matches are too close and no identification is made because the result is considered ambiguous.

1.2.6.1.3 XTandem

The XTandem search algorithm differs from the previous programs by virtue of the fact that it is an open-source program that does not require purchase of a license for use. It is also a heuristic search algorithm like SEQUEST that predicts fragmentation spectra for database peptides, but only performs those predictions on peptides that have few internal missed enzymatic cleavage sites. This allows the search algorithm to operate much faster than either SEQUEST or Mascot on a given dataset. XTandem also incorporates some known variations in fragmentation based on peptide sequence, such as the trend for increased fragmentation on the N-terminal side of proline residues.[103]

1.2.6.2 Protein sequence analysis tools

After the identified peptides are matched to protein sequences, there may be a need to further analyze the protein sequence to elucidate function, modification or cellular location. One of the tools available for this is the basic local alignment search tool (BLAST).[104] BLAST can search both nucleotide and protein databases to identify protein homology, which is useful when the database used for peptide identification is insufficiently annotated with functional information. Additional information such as subcellular localization can be found using the TargetP[105] localization predictor. Identification of possible cell membrane or cell wall glycosylphosphatidylinositol (GPI) anchors can be performed using the big-PI Fungal Predictor,[106] or the automated detection of GPI-anchored proteins using the DGPI prediction algorithm.[107] A third online algorithm (called TMHMM) can be used to predict hydrophobic transmembrane regions within a protein sequence using both a Hidden Markov Model as well as Neural Network prediction schemes.[108, 109] Another useful program is the Gene Ontology Tool [110] to infer functional classification of proteins. More helpful sequence analysis programs can be found at the EXpert Protein Analysis System (ExPASy) proteomics server (http://www.expasy.ch). Additional tools for analysis of mass spectrometry data can be found at the Protein Prospector (http://prospector.ucsf.edu), as well as helpful proteomics software at the Proteome Commons (http://www.proteomecommons.org).

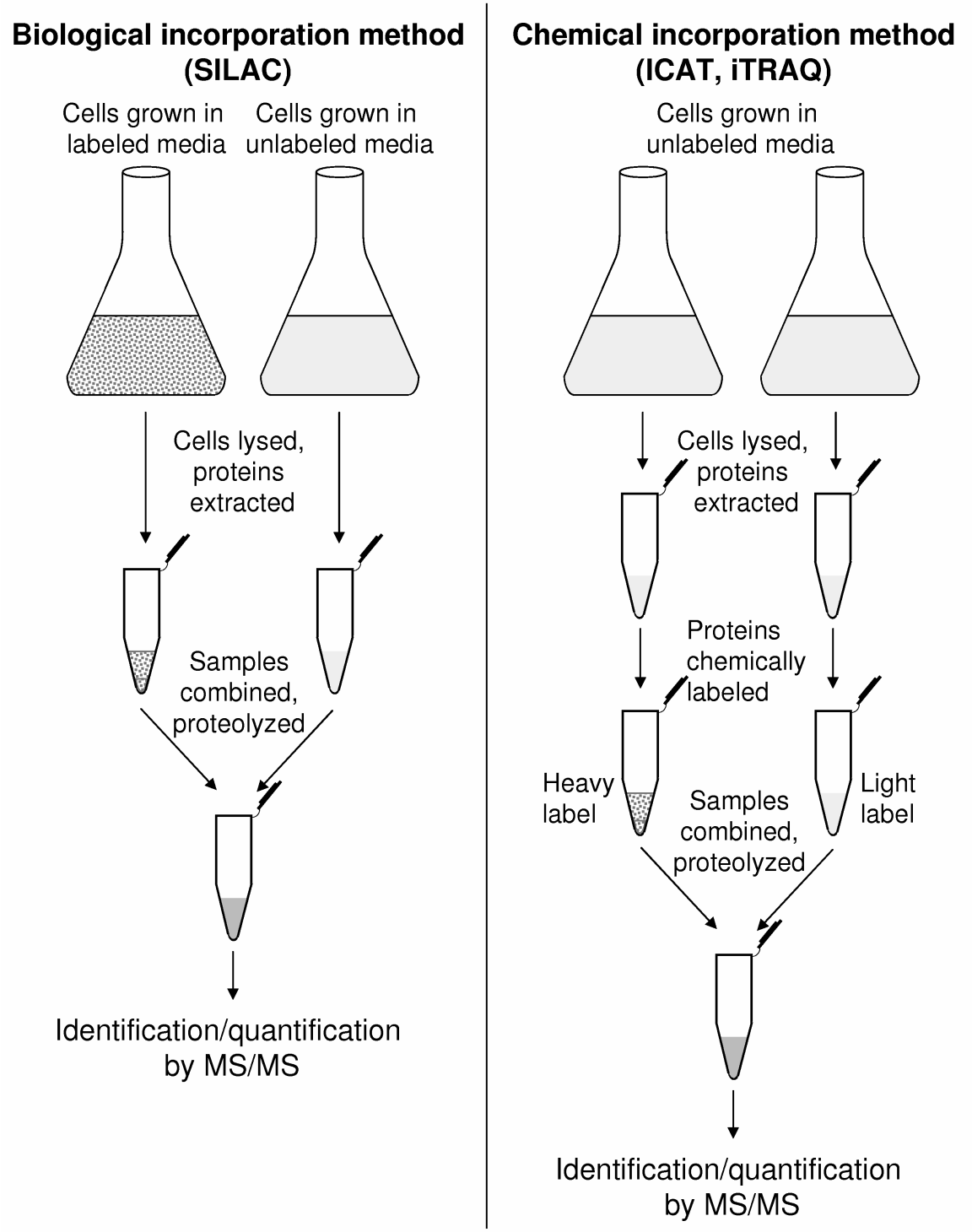
1.2.7 Protein quantification

The area of differential proteomics has seen several advances in recent years. Three of the newer techniques can be used for differential quantitative analysis using mass spectrometry (a technology that does not lend itself well to quantitative measurement unless internal standards are used). Each of the techniques involve differential labeling of proteins from different samples using stable isotopes, such as deuterium, ^{15}N, or ^{13}C. There are two ways to label proteins with these isotopes: biological incorporation, where cells are grown on media enriched with the isotope being used. The second method is known as chemical incorporation, where the isotopically-labeled tag is added to proteins after extraction (often reacting with primary amine groups or cysteine residues). Regardless of label incorporation method, corresponding labeled and unlabeled peptides will be detected in the mass spectrometer at the same time. A diagram of chemical and biological stable isotope labeling is shown in Figure 1.6. Quantitative data is derived by comparing the ratio of areas of the MS peaks for labeled and unlabeled peptides.[111]

1.2.7.1 Isotope coded affinity tag (ICAT)

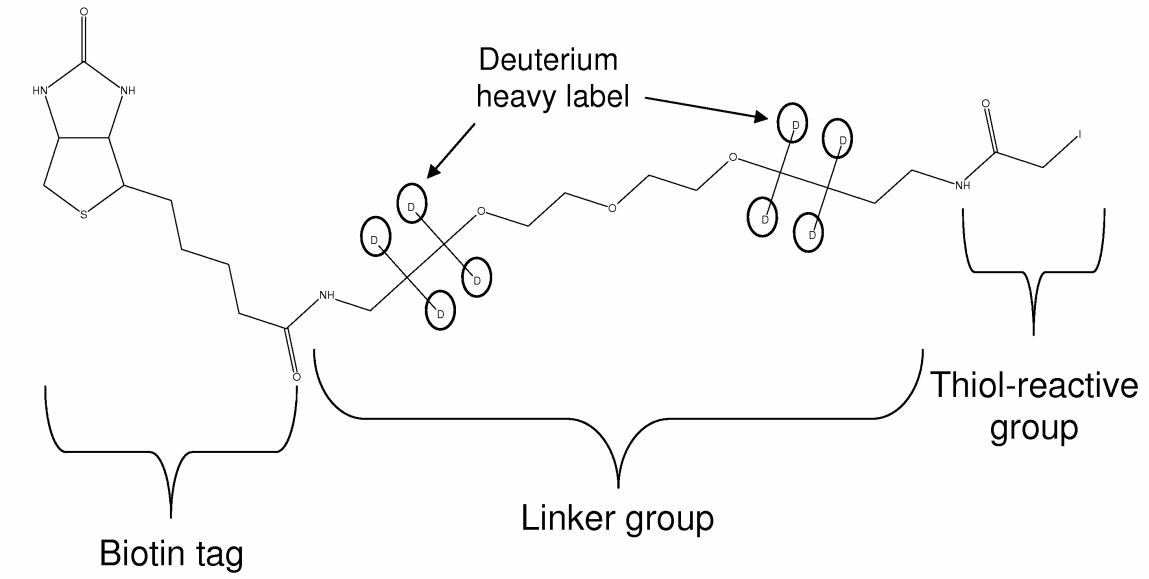
Of the three main methods, for differential protein analysis, ICAT (Isotope Coded Affinity Tag) was the first established. It involves a chemical incorporation of a deuterium-labeled reagent consisting of a thiol-specific reactive group (derived from iodoacetamide) that binds to cysteine residues in proteins. Attached to the reactive group is a poly-ether amide linker that can be labeled with deuterium atoms, or remain

Figure 1.6 Stable isotope label incorporation strategies for protein quantification by MS/MS



unlabeled with hydrogen, or other stable isotopes such as ^{13}C. This linker is also attached to a biotin moiety.[112] A diagram of a deuterated ICAT label is shown in Figure 1.7. Using the ICAT method, proteins are collected from two conditions to be analyzed (such as diseased and healthy cells). One of the protein mixtures is reacted with unlabeled (light) ICAT and the other mixture with labeled (heavy) ICAT. In both cases, the thiol-

Figure 1.7 Isotope Coded Affinity Tag (ICAT) stable isotope label

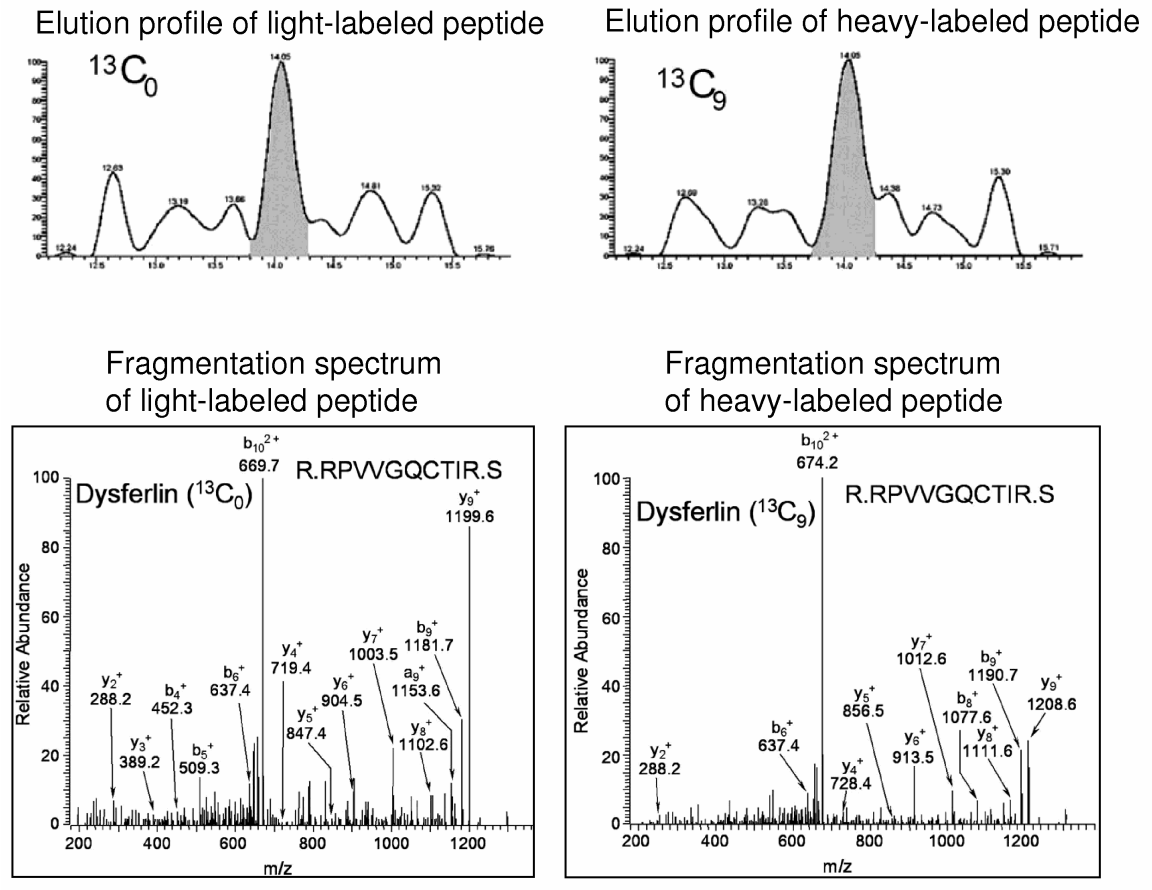


reactive group binds to cysteine residues of the proteins. The mixtures are then combined and subjected to proteolytic cleavage. Each of the labeled peptides now has a biotin affinity tag, which can be used to separate the heavy-labeled peptides from light-labeled peptides using a streptavidin column.[113] This leads to a less complex mixture of peptides, easing the separation and analysis by mass spectrometry. Quantification is then performed by comparing the ratio of areas of the MS peaks for labeled and unlabeled

peptides. An example of ^{13}C ICAT labeling for differential protein quantification is shown in Figure 1.8.[114] In this figure, the area under the curve of the elution profiles of both labeled and unlabeled peptides is shown, along with the MS/MS fragmentation spectra of each peptide for sequence identification.

The primary advantage of ICAT labeling is the ability to label proteins that cannot be labeled using biological incorporation techniques, such as human serum samples. ICAT also allows for decreasing the complexity of a complex biological mixture by the

Figure 1.8 Differential protein quantification using ICAT labeling
(Figure reproduced with permission from [114] Karsan, A., I. Pollet, L. R. Yu, et al. 2005. Quantitative proteomic analysis of sokotrasterol sulfate-stimulated primary human endothelial cells. *Mol Cell Proteomics* 4: 191-204.)



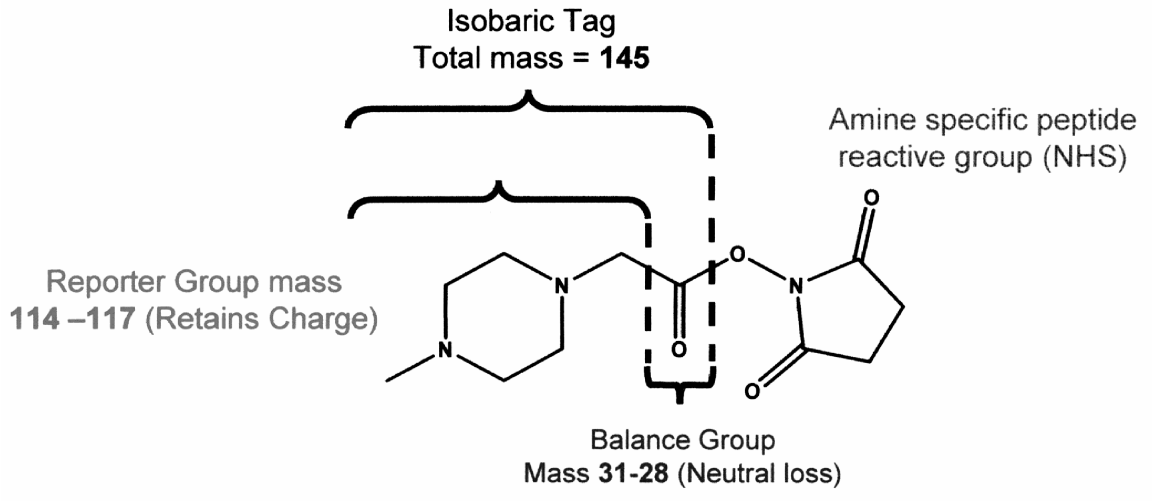
use of biotin affinity separation, as well as allowing for the identification of low abundance proteins by biotin-assisted concentration of the protein. The main disadvantage of ICAT compared to other labeling methods is the fact that only cysteine-containing proteins and peptides can be identified. This not only increases the ambiguity of some protein identifications,[115] it also prevents the identification of any proteins that do not contain a cysteine.[116]

1.2.7.2 Isotope tagging for relative and absolute protein quantitation (iTRAQ)

A variation of the ICAT technique has been recently introduced called iTRAQ (Isotope Tagging for Relative and Absolute protein Quantitation). The iTRAQ system (developed by Applied Biosystems) consists of a Peptide-Reactive Group (PRG) that reacts with primary amines of peptides (i.e. lysine side chains and amino-termini). The PRG is attached to a balancing group with a mass of 28, 29, 30, or 31 Da. This balancing group is also attached to a reporter group with a mass of 117, 116, 115 or 114 Da. The balance group and the reporter group are matched so that the entire tag is isobaric with the other three tags. A diagram of the iTRAQ tag is shown in Figure 1.9.[117] When subjected to fragmentation in MS/MS, the tag is cleaved between the PRG and balance groups and between the balance and reporter groups. The balance group is a neutral molecule that will not be detected by MS. The cationic reporter group can then be detected along with all of the normal peptide fragmentation ions. The quantification is performed by analyzing the differential abundances of the four product ions of the reporter groups.[118] An example of protein quantification using iTRAQ labeling is shown

Figure 1.9 Isotope Tagging for Relative and Absolute Protein Quantitation (iTRAQ) stable isotope label

(Reproduced from [117] Zieske, L. R. 2006. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* 57: 1501-1508. by permission of Oxford University Press [on behalf of the Society for Experimental Biology].)

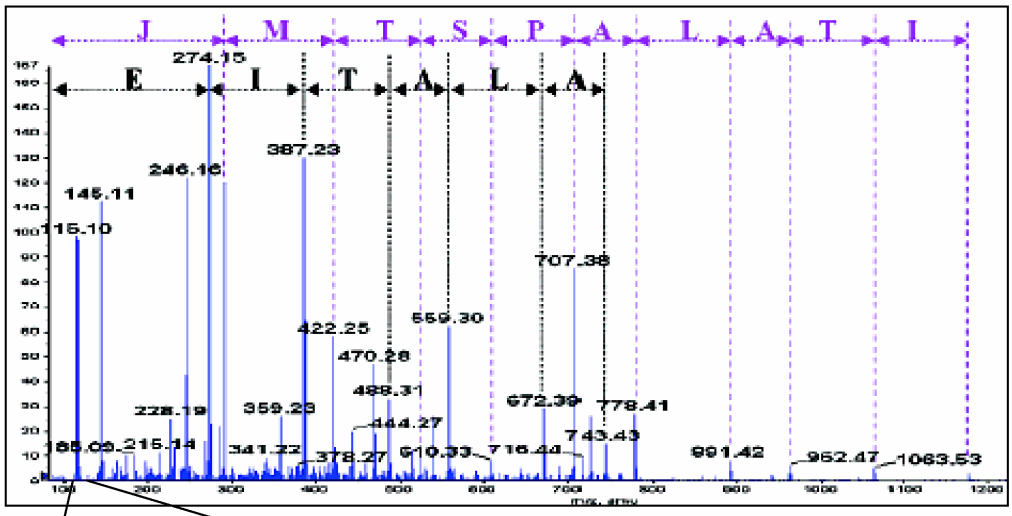


in Figure 1.10. This figure shows the typical peptide fragmentation spectrum in Panel A, including amino acid sequence from b and y ions. Panel B is a magnification of the m/z region corresponding to the location of the reporter groups.

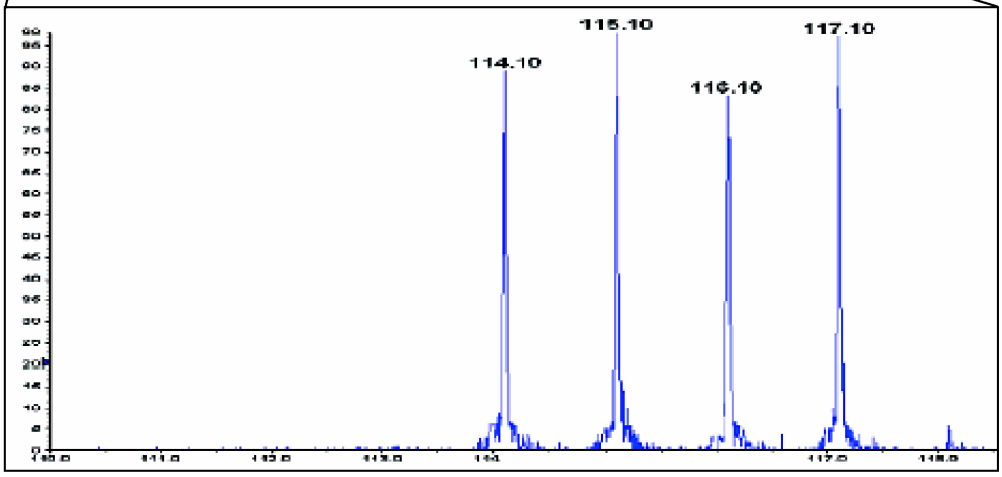
The primary advantage of the iTRAQ system is the same as that for standard ICAT: the chemical incorporation of label. It also has the advantage over ICAT in being able to label all peptides in a mixture, not just those containing cysteine. It also allows for the analysis of four different conditions at one time, rather than just two with ICAT. The biggest disadvantage of iTRAQ comes from the MS identification. The low masses of the reporter groups requires analysis by an instrument with no (or a low) mass cutoff that allows for identification of product ions of 114-117 Da. This means a standard quadrupole ion trap instrument cannot be utilized for iTRAQ identifications. Analysis

Figure 1.10 Differential protein quantification using iTRAQ labeling
(Figure reprinted with permission from [118] *J Proteome Res* **2005**, *4(2)*, 377-386. Copyright 2005 American Chemical Society.)

Panel A Full MS/MS spectrum of iTRAQ-labeled peptide



Panel B Magnification of 110-120 m/z region



would most likely be done with a QTOF or a MALDI-TOF instrument (with an associated collision cell), although recent advances in 2D linear ion trap operation can

allow for a low mass cutoff around 50 Da.

1.2.7.3 Stable isotope labeling by amino acids in cell culture (SILAC)

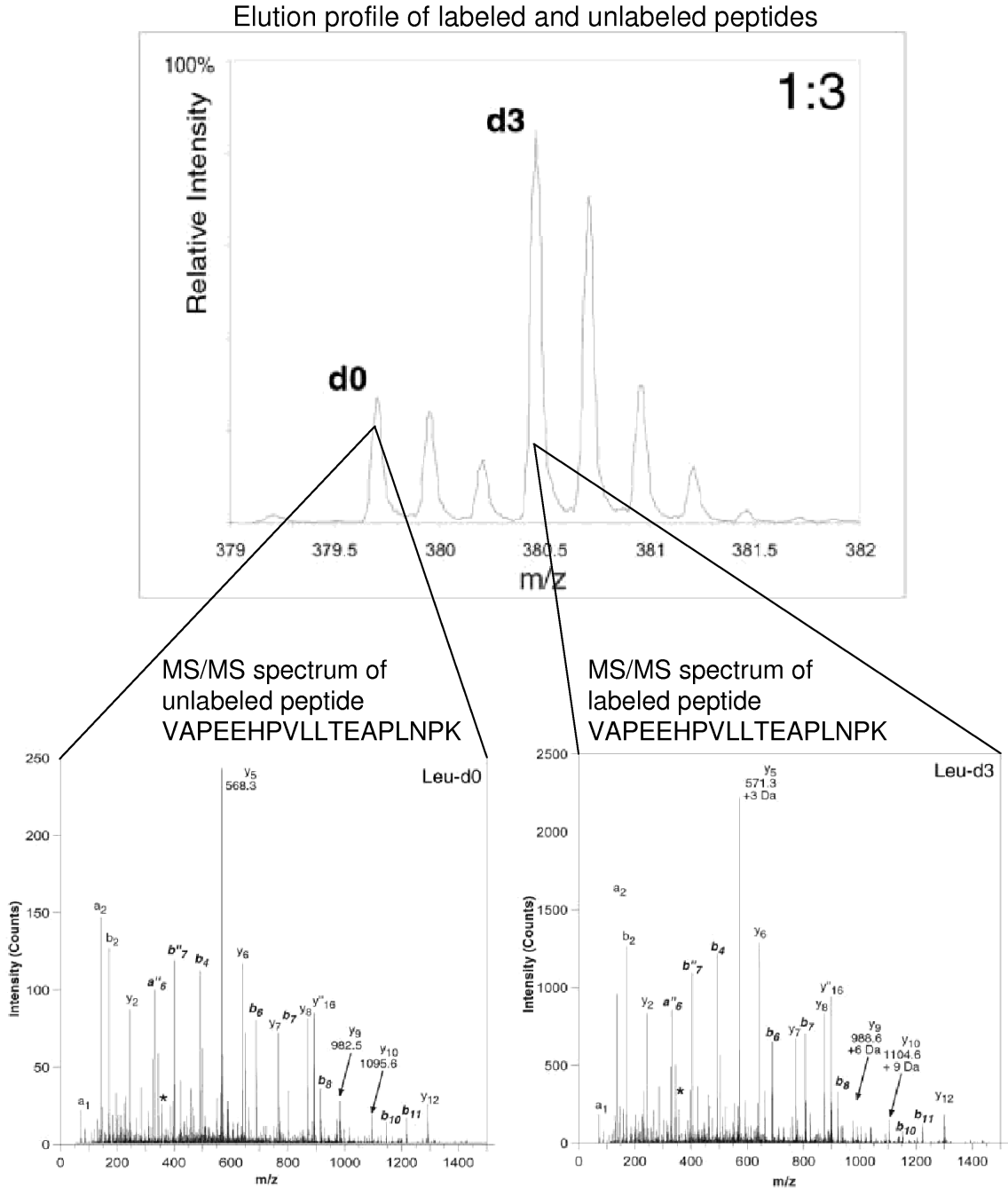
The last quantitative labeling technique is SILAC (Stable Isotope Labeling by Amino acids in Cell culture). SILAC is a biological incorporation method in which the cells of interest are grown in stable isotope-enriched media. The first introduction of SILAC[116] was done with deuterium-labeled leucine added to the media of mouse primary cells. An example of protein quantification using SILAC labeling is shown in Figure 1.11. While the acronym SILAC was coined relatively recently, it is a derivative of an earlier technique involving the growth of cell cultures on ^{15}N isotopically-enriched media resulting in the incorporation of the isotope into all amino acids.[119, 120] This method results in more complicated analysis of the mass spectra (due to each amino acid residue having at least one ^{15}N incorporated), but allows for analysis of changes of protein expression levels as low as 10%.[115] A disadvantage of this method is the increase in the number of isobaric amino acids with ^{15}N glutamic acid/^{15}N glutamine, and ^{15}N aspartic acid/^{15}N asparagine, in addition to the normal isobaric residues of leucine/isoleucine and glutamine/lysine. This increased number of indistinguishable amino acids leads to increased ambiguity in peptide identification by MS/MS of ^{15}N-labeled peptides.[121]

1.2.7.4 Two-dimensional difference gel electrophoresis (DIGE)

Another widely used quantification technique is the 2-Dimensional Difference Gel Electrophoresis (DIGE).[122] DIGE uses a system of fluorescent markers that bind to

Figure 1.11 Differential quantification of SILAC labeled peptides

(Figure reproduced with permission from [116] Ong, S. E., B. Blagoev, I. Kratchmarova, *et al.* 2002. 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.)



proteins in a sample, allowing for quantification of labeled proteins in a 2-D gel upon excitation of the marker by a laser. Samples of interest can then be in-gel digested with a protease and analyzed by MS/MS. Advantages of DIGE include the relative ease of use of the fluorescent markers compared to stable-isotope label incorporation. Disadvantages include the cost of markers and the requisite laser scanner as well as the fact that most spots on a 2-D protein gel contain more than one protein, leading to ambiguity in assignment of abundance.

1.2.8 Application of proteomic analysis to fungal systems

There are several examples of proteomic analyses of fungal organisms to be found in the current literature. Many of these utilize some of the techniques described above, such as an analysis of the obligate plant pathogen *Uromyces appendiculatus*.[123] In this analysis, proteins were extracted from uredospores, digested and then separated by MudPIT. The analysis identified over 400 proteins, many of which are associated with protein-production such as translation factors, ribosomal proteins, and amino acid synthetases. These results led the authors to hypothesize that the uredospores exist in a suspended state of translation that allows the spore to begin protein production rapidly upon germination.

An analysis of the human pathogen *Candida albicans* incorporated 2-DE separation of cell wall proteins prior to MS analysis.[124] In this study, the cell walls of the yeast and hyphae morphologies were subjected to protein extraction by SDS and Dithiothreitol (DTT) or cyanogen bromide (CNBr)/trypsin digestion. This study

produced a total of 82 SDS/DTT-extractable cell wall proteins from both yeast and hyphal samples. Seven of these proteins were shown to be upregulated in the yeast-hyphae transition, and 2 were down-regulated. There were an additional 29 proteins identified from the CNBr/trypsin digestion of both cell types, 12 of which are hyphae-specific, and 6 that are yeast-specific. These protein identifications have not only increased the understanding *C. albicans* biology, but also identified a heat-shock protein that is up-regulated in the yeast-hyphae transition, but not at the mRNA level. These results suggest that this protein is regulated at a post-translational level in the fungal cell wall.

Another analysis also focused on *C. albicans*, illustrating the applicability of proteomics to vaccine development.[125] In this study, an extract of yeast cell wall proteins was shown to be effective in protecting mice from infection. This study identified and characterized 20 proteins that reacted with antibodies from the serum of immunized animals. Many of the identified proteins were determined to play important roles in adhesion, cell surface hydrophobicity and immunogenic activity. These protein identifications have produced target antigens to be used in the development of a subcellular vaccine against *C. albicans* infection.

There are other examples of fungal proteomics such as the analysis of proteins secreted by the phytopathogen *Sclerotinia sclerotiorum.*[126] In this study, both mycelial and secreted proteins were separated by 2-DE. This analysis identified 18 secreted proteins, along with 95 mycelial proteins that provide insight into the fungal lifecycle and pathogenicity. One protein had not been previously identified in analysis of mRNA

levels, highlighting the value of direct protein identifications, rather than protein presence inferred from transcript analysis.

The quantitative technique SILAC was used in a study of the complete proteome of *Saccharomyces cerevisiae.*[127] In this analysis, yeast cells were grown in normal media or media containing labeled lysine. The proteins collected from the cells were digested and the resulting peptides were analyzed on a linear ion trap-FT mass spectrometer capable of extremely high peptide mass accuracy. Peptides were identified by MS/MS fragmentation resulting in identification of over 2000 *S. cerevisiae* cytoplasmic proteins. These identifications included low abundance proteins corresponding to about 100 protein copies per cell.

Another recent analysis used SILAC-like stable isotope labeling for protein quantification in *Schizosaccharomyces pombe*.[128] In this study, fungal cells were treated with Cd^{2+} and labeled with deuterated leucine to determine what effect the toxic metal had on protein production. This study identified 106 proteins that were up-regulated and 55 that were down-regulated in response to Cd^{2+} treatment. In addition, 28 of the up-regulated proteins were revealed to be proteins involved in detoxification of reactive oxygen species (ROS) or repair of damaged cellular components. This study serves to highlight the applicability of proteomics to analysis of environmental effects on cellular metabolism.

1.2.9 Proteomic analyses for extracellular protein identification

There are many cases in the literature of proteomic analysis applied to identification of vaccine targets, excreted proteins, and proteins associated with the cell membrane. These include an analysis of membrane proteins in the opportunistic human pathogen *Psuedomonas aeruginosa,*[129] as well as identification of proteins excreted by *Mycobacterium tuberculosis* as potential protein antigens.[130, 131] Proteomic analysis has also been utilized to identify specific pathogen-associated proteins from *M. tuberculosis* by comparison to the nonpathogenic relative *M. bovis.*[132] Additional studies have been performed on other organisms prevalent in human disease, such as the analysis of excreted proteins from the human parasitic liver fluke *Fasciola hepatica*, in the search for potential vaccine candidates.[133] More recent studies have employed MudPIT analysis for the identification of potential vaccine antigens from cell membranes of erythrocytes infected with the malaria parasite *Plasmodium falciparum*[134] as well as comparative proteomics between *Plasmodium spp.*.[135]

1.2.10 Fungal cell wall proteomes

Additional studies involving the specific identification of fungal cell wall proteins by comprehensive proteomic analysis of covalently attached proteins have been recently undertaken. A proteomic analysis of the human opportunistic fungal pathogen *Candida albicans* using HF-pyridine to specifically cleave GPI anchored proteins by hydrolysis of the phosphodiester linkage between the protein and the cell wall, as well as NaOH incubation to remove alkali-senstitive covalently attached cell wall proteins.[136] A similar

study by the same research group used a similar chemical fractionation strategy for identifying proteins from the cell wall of *Saccharomyces cerevisiae* leading to the identification of 19 GPI-linked and alkali-sensitive proteins using both HF and NaOH extraction as well as a direct cell wall digestion with proteases. While the chemical treatment steps using HF and NaOH removed proteins that were later identified by MS/MS analysis, all of the proteins were identified using the protease digestion.[137] This suggests that direct protease digestion of cell wall components is a valid strategy for identifying cell wall associated proteins.

1.2.11 Criteria to be used in *Coccidioides* proteomic analyses

In the cell wall proteome analysis described in Chapter 3 of this dissertation, the dual algorithm search technique detailed in Chapter 2 will be utilized for the identification of proteins from single-peptide matches from MS/MS data. Only those spectra that are identified by both SEQUEST and XTandem as the same peptide sequence, and match *C. posadasii* sequences will be included in the final list of identified proteins. Following the bioinformatic analysis approach detailed in Chapter 3, any potential vaccine antigen targets identified from single peptide matches from both search algorithms will be manually validated from the source spectrum. An example of this process is shown in Figures 3.6 and 3.7. Chapter 4 describes a differential proteomic analysis to search for high-abundance spherule proteins using ^{15}N stable isotope labeling. In this analysis, proteins more highly expressed in spherules (compared to mycelia) are analyzed using bioinformatic methods to identify potential protein vaccine candidates.

2 CHAPTER TWO: VERIFICATION OF SINGLE-PEPTIDE PROTEIN IDENTIFICATIONS BY THE APPLICATION OF COMPLEMENTARY DATABASE SEARCH ALGORITHMS

The content of this chapter has been published in:

Rohrbough, J. G., L. Breci, N. Merchant, S. Miller and P.A. Haynes; 2006. **Verification of Single-Peptide Protein Identifications by the Application of Complementary Database Search Algorithms**. Journal of Biomolecular Techniques 17(5)**:** 327-332

Data produced from MudPIT analysis of yeast (*S. cerevisiae*) and rice (*O. sativa*) were used to develop a technique to validate single-peptide protein identifications using complementary database search algorithms. This results in a considerable reduction of overall false-positive rates for protein identifications; the overall false discovery rates in yeast are reduced from near 25% to less than 1%, and the false discovery rate of yeast single-peptide protein identifications becomes negligible. This technique can be employed by laboratories utilizing a SEQUEST-based proteomic analysis platform, incorporating the XTandem algorithm as a complementary tool for verification of single-peptide protein identifications. We have achieved this using open-source software, including several data-manipulation software tools developed in our laboratory, which are freely available to download.

2.1 Introduction

Protein identification from complex biological mixtures often involves the application of tandem mass spectrometry techniques [138, 139] such as MudPIT [140, 141], which involves digestion of the protein mixture with a protease such as trypsin, followed by two stages of liquid chromatography separation using strong cation exchange (SCX) and reverse-phase (RP) separation. Peptides eluting after these separations are subjected to ionization and fragmentation in the mass spectrometer. Database search algorithms are then used to match the acquired spectra to peptide sequences from a protein database. Examples of such programs include SEQUEST [138, 142], Mascot [143], Spectrum Mill [144], ProteinLynx [145], XTandem [146-148], and OMSSA.[149] When a protein is identified from several unique peptide spectra, the inherent redundancy of identification improves the confidence in protein identification, even if the confidence of some of the peptide identifications is low. As the number of peptides assigned to each protein sequence decreases, the confidence of protein identification drops correspondingly.

There are many examples in current literature of proteomic analyses performed by application of the MudPIT technique.[150-154] However, there is no consensus on the search parameters used for the database search algorithms, or the treatment of proteins identified from single peptides. It is not correct to simply disregard single-peptide matches. Such peptides may be the only detectable peptide from an enzymatic digest, and therefore perfectly valid for identification purposes. It is equally incorrect to include all proteins identified from single peptides, because of the variability in protein identification from poor mass spectra, resulting in a high rate of false-positive identifications.[155-158]

There have been numerous attempts to validate protein identifications from current database search algorithms, including: linear discriminate analysis used to determine the accuracy of search algorithm assignments [159]; the Qscore algorithm using a probabilistic scoring system and analysis of false-positive identification rates using a reverse database [160]; the heuristic approach to assigning false discovery rates [161]; the normalization of peptide identification scoring systems based on the length of the peptide [162]; utilization of the tryptic status of peptides as an additional level of validation [140, 162-164]; the application of a support vector machine (SVM) to distinguish between correct and incorrect peptide identifications by SEQUEST [165]; and the inclusion of orthogonal parameters such as exact mass measurements of selected peptides.[166] One published report describes a proteomic analysis in which the final results were in the form of a consensus between the output from two different search algorithms.[167] However, neither this report, nor any of those mentioned above, specifically addresses the issue of improving the confidence rate of assignment for proteins identified from a single peptide. Several authors, however, have noted that consensus analysis of dual algorithm searching programs has considerable merit in terms of protein identification confidence levels.[144, 168]

Our aim in this study was to develop a basic set of software tools that would enable us to achieve 95%, or greater, confidence of assignment for both single- and multiple-peptide based protein identifications, using only freely available, open-source software in addition to our existing SEQUEST analysis platform. As a consequence, all software tools developed and used in this project are made freely available via our laboratory website.

2.2 Materials and Methods

The data used in the development and testing of this approach were acquired from triplicate MudPIT analyses of yeast (*S. cerevisiae*) mixed organelle lysate sample (designated Y1,Y2 and Y3), prepared and analyzed as described [150], and rice (*O. sativa*) leaf, root and seed organ lysate samples (designated R1seed, R2root and R3leaf), prepared [76] and analyzed [150] as described.

The entire set of tandem mass spectra collected from all 13 chromatographic steps in each experiment were searched using TurboSEQUEST (BioWorks version 3.1, Thermo Electron) [138, 142] run on a 16-processor IBM Beowulf cluster; with dta files generated from peptide spectra meeting the following criteria: Peptide MW Range = 400-3500 Daltons; Threshold = 1000; Precursor Mass = 1.40; Group Scan = 1; Minimum Group Count = 1; and Minimum Ion Count = 35.

All SEQUEST searches were performed with no enzyme specificity indicated. The search parameters used were default settings except for: peptide mass tolerance = 1.5; max number of modified amino acids per differential modification in a peptide = 4; static modification mass of +57.0 for acetylated cysteine; differential residue modification mass of +16.0 for oxidized methionine; a maximum of 2 internal cleavage sites; one allowed error in matching auto-detected peaks, and a mass tolerance of 1.0 for matching auto-detected peaks. SEQUEST search results were filtered using DTA-select v 1.9 [169] using our laboratory default cutoff parameters : Xcorr for a 1+ ion = 1.8, Xcorr for a 2+ ion = 2.5, Xcorr for a 3+ ion =3.5, deltaXcorr = 0.1.[150, 170-172]

The single-peptide matches from SEQUEST were re-searched against the same database by XTandem version 2005.10.01.5 (open source software, available from http://www.proteome.ca/opensource.html).[146-148] The default XTandem search parameters were used, except for the following: a maximum valid expectation value of 0.02; residue mass modification of +57.022 for carbamidomethylated cysteine; potential residue mass modification of +16.0 for oxidized methionine; enzyme specificity = none specified; spectrum parameters including a fragment monoisotopic mass error of 0.5 Daltons and a parent monoisotopic mass error of +/- 2.5 Daltons; spectrum conditioning parameters of 100 .0 spectrum dynamic range, total spectrum peaks 50, a minimum parent M+H of 400.0 and a minimum fragment m/z of 150.0.

Tandem MS spectra from rice organ samples were searched against a database of rice (*Oryza sativa japonica*) protein sequences (36318 sequences- April 2005 version), representing the complete rice genome, from NCBI (www.ncbi.nlm.nih.gov). The yeast samples were searched against a yeast genome protein sequence database (6882 sequences, March 2005) from the *Saccharomyces* Genome Database (www.yeastgenome.org). Both the rice and yeast databases were supplemented with common laboratory contaminants.[150] Manipulation of mass spectrometry data was assisted by the use of several Perl script programs designed in-house, all of which are freely available for download from our laboratory website as part of the Wildcat Toolbox (http://proteomics.arizona.edu/toolbox.html), and which are described in detail in a separate report.[173]

For the data analysis outlined in this report, six distinct sets of MudPIT data were acquired, and all six data sets were searched using SEQUEST against both a forward and reversed database.[160-162, 174] False discovery rates (FDR) were calculated by determining the number of matches against the reversed database as a percentage of the number of matches against the forward database, which gives an estimate of random sequence matches to the database, in accordance with recently published proteomics data guidelines.[156, 157] In numerical terms, FDR is FP/(TP + FP), where FP is false positives and TP is total positives.[161] It is important to note that we have not addressed false negative assignments in this report for two reasons: first, identification of false negative assignments from a biological sample where the "correct" answer is not known is problematic, and second, the method presented here is simply intended to limit the false discovery rate using available search algorithms.

2.3 Results and discussion

The number of proteins identified in each experiment, along with the false discovery rate in each experiment, is shown in Table 2.1. The salient features of this data are first that the largest contributor to the overall false-positive rate is very clearly those proteins identified from single peptides, and second that by using a two peptide minimum criteria our currently used SEQUEST cutoff parameters would give us a satisfactory confidence of protein assignment. When a minimum of two peptides per protein is imposed, our current SEQUEST parameter cutoff scores produce a false discovery rate

below the targeted 5% threshold. One data set out of six has a FDR of 5.7%, but the average for all six experiments is 3.1%.

Table 2.1 Protein identifications and false discovery rates in SEQUEST analysis of MudPIT data

Experiment	Total proteins identified[a]	Single peptide proteins identified[b]	FDR[c] Single peptides only	FDR overall	FDR 2 peptides minimum
Y1	532	248	50.4	23.9	1.1
Y2	604	295	51.2	25.5	2.9
Y3	517	262	47.7	25.5	5.7
R1seed	221	155	41.9	29.9	3.1
R2root	258	175	28.6	19.4	0
R3leaf	247	169	59.2	40.9	2.6

a) Number of proteins identified in Yeast and Rice MudPIT protein identifications using SEQUEST cutoff scores of: Xcorr for a 1+ ion = 1.8, Xcorr for a 2+ ion = 2.5, Xcorr for a 3+ ion =3.5, deltaXcorr = 0.1

b) Number of proteins identified from single peptides only using SEQUEST with cutoff parameters detailed in footnote a.

c) false discovery rates assessed by searching against a reversed sequence database, calculated using FDR is FP/(TP + FP), where FP is false positives and TP is total positives [24], expressed as a percentage.

The DTA_sorter.pl script was developed to extract those .dta files corresponding to SEQUEST single-peptide identifications. This script uses the DTASelect-filter.txt output file[169] and separates all .dta files from a MudPIT run into three newly created folders: singlexcel, which contains all .dta files that correspond to single-peptide identifications; inexcel, which contains all of the .dta files that correspond to multiple-peptide protein identifications; and notinexcel, which contains all of the remaining .dta

files. The script then creates a concatenated .dta file from all of the individual .dta files contained in each newly created subdirectory for use in further searching.

For data output comparison purposes, the CommonSingles.pl script was developed, which compares a DTASelect output file (DTASelect-filter.txt) to an XTandem Excel table output (obtained using the Global Proteome Machine xml input upview page at: http://ww.thegpm.org). The CommonSingles script produces a modified DTASelect output file that includes all of the single peptides found by XTandem that are also found by SEQUEST.

Spectra corresponding to the single peptide based protein identifications from all six experiments were sorted using DTA-sorter.pl, re-searched using XTandem, and the single peptide identifications common to both algorithms were combined with the multiple peptide based protein identifications using the Commonsingles.pl program. The same procedure was used for both forward reverse databases to allow calculation of FDR. Table 2-2 shows the revised numbers of proteins identified in each of the six MudPIT experiments. The false discovery rates of the overall data sets have dropped from approximately 25% in the initial SEQUEST searches to less than 1% in the dual algorithm search results, while the false discovery rates for the single peptides considered in isolation have dropped from around 50% to less than 1%, zero in some cases. This is a dramatic improvement in overall data quality, and has been obtained without increasing the number of false negative assignments by simply excluding all of the single peptide based matches.

Table 2.2 Protein identifications and false discovery rates observed using dual algorithm searching

Experiment	Total proteins identified in SEQUEST searches	Revised Total proteins identified using dual algorithm search	Overall FDR[a] using dual algorithm search	FDR of single peptides retained in dual algorithm approach
Y1	532	417	0.005	0
Y2	604	467	0.011	0.013
Y3	517	384	0.021	0.008
R1seed	221	141	0.71	0
R2root	258	174	0	0
R3leaf	247	153	0.65	0

a) False discovery rates, determined as explained in Table 2.1

Within the yeast samples, there is a high level of reproducibility in the results. When compared to samples prepared from rice organs, there is a clear difference in false discovery rates, as expected in samples from different biological sources.[162] The reanalysis of the yeast MudPIT datasets results in the retention of an average of 76.7% of all proteins identified by SEQUEST, which includes on average 52.1% percent of the single-peptide identifications. For the rice MudPIT datasets an average of 64.4% of the total proteins are retained, which includes an average of 48.3% of the single peptide identifications.

While none of the partially tryptic peptides contained in the SEQUEST analysis data sets were confirmed by XTandem searching, a large number of fully tryptic peptides were dropped from the final dataset as they were not confirmed using the second

algorithm. This confirms that we are not simply filtering the single peptide matches on the basis of tryptic status, which is essential as not all of our experiments involve solely trypsin digestion. When analyzing the common singles, none of the dual algorithm consensus matches are partially tryptic; all are fully tryptic. However, out of 115 single peptide matches dropped from Y1, 58 (50.4%) are partially tryptic, for Y2, 91 of 137 (66.4%) are partially tryptic, and for Y3, 83 of 133 (62.4%) are partially tryptic. Further analysis of the forward and reverse database search results (data not shown) demonstrates that imposing a fully tryptic constraint on the single peptide matches would improve the FDR compared to the original SEQUEST results, but would not bring it below our desired threshold rate of <5%.

In conclusion, we have presented a method for verifying proteins identified from a single unique peptide during nanoLC-MS/MS experiments such as MudPIT analysis of a complex biological mixture. For the analysis of yeast MudPIT datasets, we are able to produce a revised results output with an overall false positive assignment rate of less than 1%, which still retains over 75% of the proteins initially identified. Similarly, for analysis of the rice organ MudPIT datasets, we are able to retain over 60% of the proteins initially identified, with a revised overall false discovery rate less than 1%. This indicates that application of this technique is highly reproducible for the analysis of similar samples, and likely to yield comparable, yet distinctly different, results for samples prepared from different biological sources.

We have developed a technique that can be employed by laboratories utilizing a SEQUEST-based proteomic analysis platform, incorporating the XTandem algorithm as a

complementary tool for verification of single-peptide protein identifications. We have achieved this using open-source software, including several data-manipulation software tools developed in our laboratory, which are freely available for download. We make these programs available to other users in the spirit of open-source collaboration, and we hope and expect that users will modify them to fit their own needs. For example, it would be relatively simple to adapt these tools for use with Mascot rather than SEQUEST as the primary search engine, or Mascot rather than XTandem as the secondary search engine.

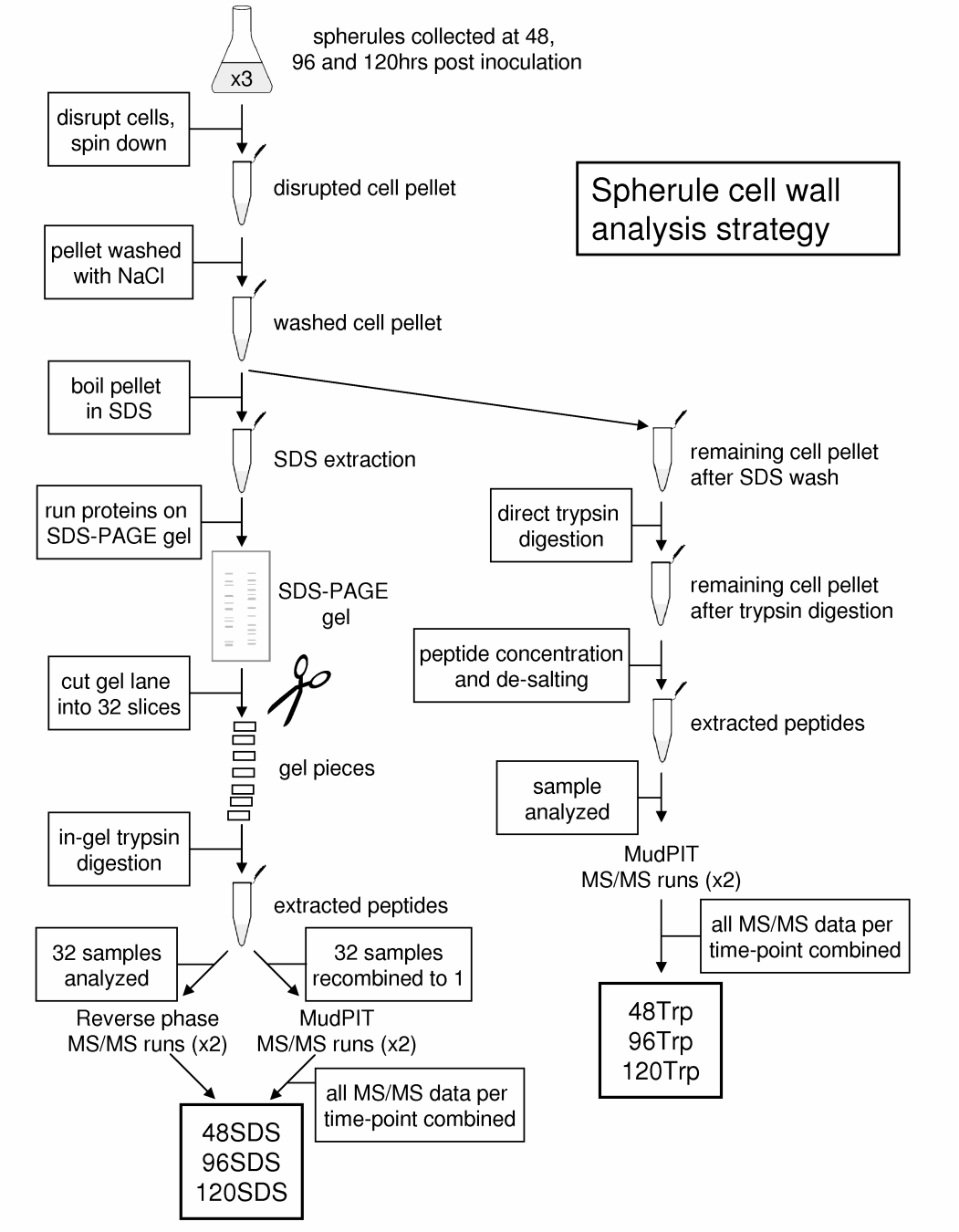
3 CHAPTER THREE: ISOLATION AND IDENTIFICATION OF PROTEINS ASSOCIATED WITH THE SPHERULE CELL WALL

3.1 Introduction

This chapter describes experiments that were performed with the goal of identifying protein vaccine candidates that are associated with the spherule cell wall of *Coccidioides posadasii*. As explained in Chapter 1 of this dissertation, vaccines derived from spherules are more effective than those derived from mycelia cells. In addition, most of the current protein vaccine candidates are associated with the spherule cell wall. Proteomic analysis of fungal cell walls has proven effective in identifying covalently-associated cell wall proteins in *Saccharomyces cerevisiae*[136, 137] as well as the opportunistic fungal pathogen *Candida albicans*.[136] An analysis of this type of *Coccidioides posadasii* is likely to identify previously uncharacterized protein antigen targets. The cell wall protein analysis described here is, to our knowledge, the most comprehensive analysis yet undertaken to describe the cell wall proteome of either *Coccidioides* spp. The results of this analysis will likely benefit areas of research from vaccine development to fungal biology.

The general approach applied to identify cell wall associated proteins by tandem mass spectrometry is shown in Figure 3.1. In brief, spherules from three separate time points (48, 96 and 120 hours post inoculation) were collected as representative samples of immature, mature and endosporulating spherules, respectively (see Figure 1.1). The cells were disrupted and spun down to isolate the cell wall and membranes. Proteins were

Figure 3.1 Strategy employed for identification of spherule cell wall associated proteins



extracted from this pellet using both SDS extraction and direct trypsin digestion of the pellet. Proteins were identified using both one-dimensional liquid chromatography

separation with reverse-phase packing material and two-dimensional separation known as MudPIT (both methods are described in detail in Section 1.2.3.1 of this dissertation).

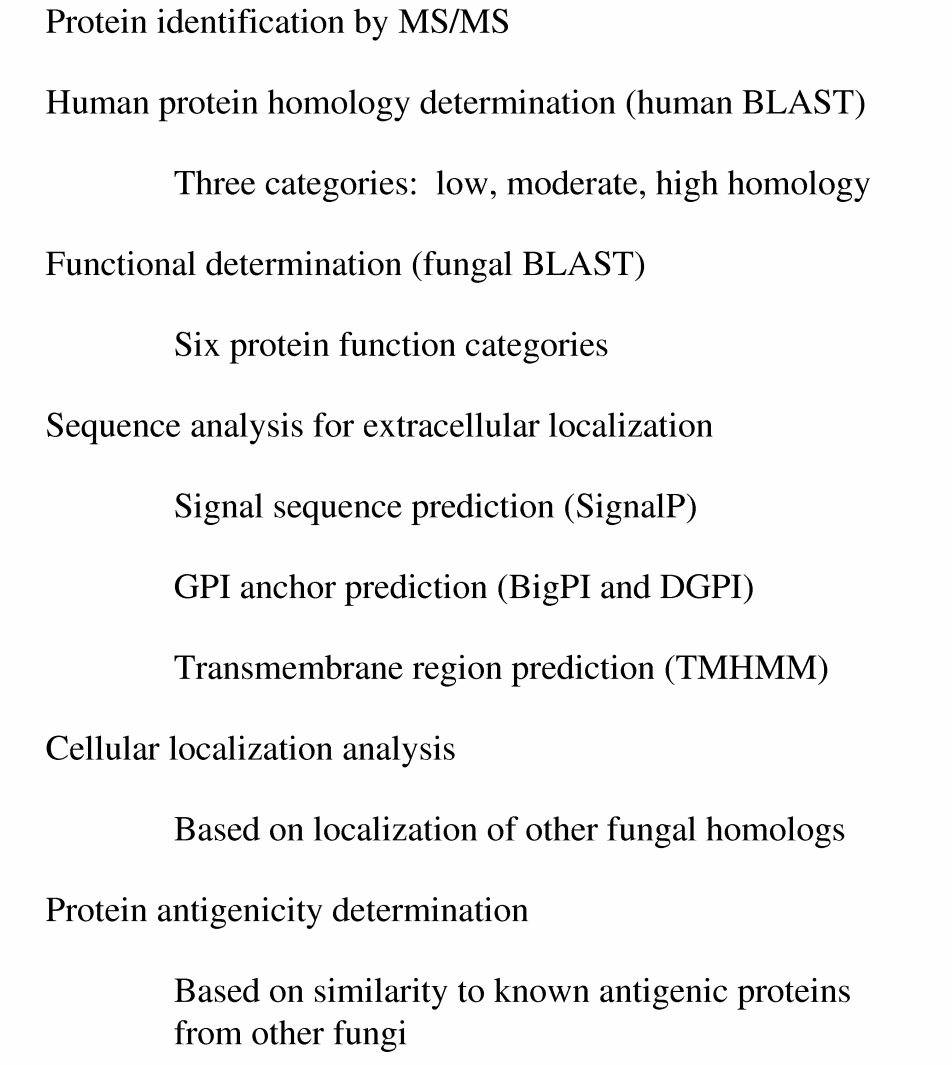
Proteins identified from the spherule cell walls were then analyzed for indicators of extracellular localization as well as homology to human and other fungal proteins in an effort to identify possible vaccine candidate proteins. The bioinformatic strategy used for this analysis is shown in Figure 3.2. Any *C. posadasii* protein that was identified by MS/MS analysis, had low or moderate human homology, contained sequence elements predicting extracellular localization and had not been previously analyzed as a vaccine candidate was considered a new vaccine target for further research analysis. Use of this approach on known vaccine antigens (see Section 1.1.5.3) that were found in the spherule cell wall analysis identified seven of nine as possible vaccine candidates. This suggests that the described strategy of antigen identification will be successful in identifying new protein vaccine candidates for further analysis.

3.2 Materials and methods

3.2.1 Protein separation for MS/MS analysis

In this study, we have employed a combined approach with regards to protein separation methods prior to MS analysis. Separation of proteins by gel electrophoresis is a good method for reducing complexity in a sample to be analyzed by MS, but is well known to be biased towards medium range molecular weights.[175] Proteins of very high and very low molecular weight as well as hydrophobic proteins and those with extreme isoelectric points are also not well separated by gel electrophoresis. Proteins with high

Figure 3.2 Bioinformatic analysis strategy of spherule cell wall proteins for the identification of vaccine candidate proteins



levels of glycosylation are known to spread out on a gel, lowering the concentration of the protein in a given section that can be identified by MS. An alternative to reducing the complexity of protein samples is the use of multi-dimensional liquid chromatography separation of peptides such as MudPIT. When analyzing complex mixtures such as cell

lysates, MudPIT peptide separation performs well, but identification of low abundance proteins may be masked by those of higher abundance. In an effort to capitalize on the strengths of both separation methods, we designed an analysis using the combined protein identifications of both gel-separated and MudPIT-separated MS/MS analyses.

3.2.2 Strains and growth conditions

Arthroconidia harvested from *Coccidioides posadasii* strain Silveira (isolated in 1951, a gift from H. B. Levine at the University of California, Berkeley) stock cultures were inoculated into 1L of modified Converse medium[176] at the following concentrations: 3×10^9 CFU for 48-hr spherules, 1.4×10^9 CFU for 96-hr spherules and 7×10^8 CFU for 120-hr spherules. Samples were incubated at 38°C with 20% CO_2 while shaking at 160 rpm for the appropriate time length. All manipulations of potentially viable cells were conducted in biosafety level 3 (BSL-3) conditions utilizing approved standard operating procedures in laboratories registered with the Centers for Disease Control for select agent possession.

3.2.3 Cell wall isolation

Spherules were harvested by centrifugation at 5100 rpm, 4°C for 30 min. and washed with sterile water. Pelleted cells were resuspended in equal volume cold lysis buffer (20mM Tris-HCl pH 7.9, 10mM $MgCl_2$, 1mM dithiothreitol (DTT), 200mM ammonium sulfate, 1mM PMSF, 5% v/v glycerol and 1x protease inhibitor cocktail (Calbiochem, San Diego, CA)) and an equal volume of glass beads. Cells were then

vortexed for 60 sec and placed on ice for 60 sec, alternately, 8 times. Disrupted cells were again centrifuged as above. The cell wall pellet was resuspended in 70% ethanol for 30 min to ensure complete sample sterilization per BSL-3 standard operating procedure prior to removal from the laboratory. After sterilization, the sample was again centrifuged, resuspended in 1mL lysis buffer and stored at -20°C.

3.2.4 Protein extraction

Cell wall samples were washed three times with 500µL 1M NaCl to remove contaminants and non-cell wall associated proteins prior to SDS extraction. The removal of loosely associated cell wall proteins with SDS extraction buffer (50mM Tris HCl pH 7.8, 2% w/v SDS, 100mM Na-EDTA and 20 mM DTT) was performed by boiling at 100°C for 5 minutes twice. SDS-extracted proteins (hereafter referred to as SDS-sample) were then dialyzed extensively against water with 0.1% formic acid. The remaining cell wall pellet was again washed with NaCl as above prior to direct trypsin digestion as described below.

3.2.5 Sample preparation for MS/MS analysis

SDS-sample proteins were separated by 1-D gel electrophoresis by adding approximately 100µg total protein (as determined using a bicinchoninic acid (BCA) assay) on a 12% linear SDS-PAGE gel (Bio-Rad, Hercules, CA). Proteins were visualized by silver-staining[86] and the entire gel lane was cut into 32 slices which were further cut into equal size cubes of approximately 2mm each before being placed into the

wells of a 96-well plate. Peptides from each of the 32 samples were extracted by automated in-gel trypsin digestion as previously described.[177] Briefly, gel pieces were destained using 30mM potassium ferricyanide (Sigma-Aldrich, St Louis, MO) with 100mM sodium thiosulfate (Spectrum, Gardena, CA) then dehydrated using 100% HPLC-grade acetonitrile. Proteins in the gel pieces were reduced using 10mM DTT (Fluka, Sigma-Aldrich) in 100mM ammonium bicarbonate (AmBic) then treated with 55mM iodoacetamide (Sigma-Aldrich, St Louis, MO) in 100mM AmBic for carbamidomethylation of cysteine residues. Proteins were then subjected to in-gel digestion using proteomics-grade trypsin (Sigma-Aldrich) suspended in 100mM AmBic and incubated at 37°C. Peptides were extracted from gel pieces using 5% acetonitrile with 2% formic acid.

Proteins remaining in the cell wall pellet after SDS extraction were manually reduced and alkylated using DTT and iodoacetamide then incubated in 2μg trypsin suspended in 300μL 100mM AmBic at 37°C for 3 hours followed by purification and concentration with a C-18 solid-phase extraction cartridge (3M, St Paul, MN). Purification and concentration were accomplished by passing peptide mixture through the cartridge to bind peptides. The cartridge was then washed repeatedly with water (5% acetonitrile) to remove all contaminants prior to elution of the concentrated peptides by an 80% acetonitrile wash. Peptides from the SDS-PAGE separated proteins (SDS-sample) and peptides from direct trypsin digestion of cell wall pellet (Trypsin-sample) were dried down to minimal volume (20 μL) and stored at -20°C until MS/MS analysis.

Peptides from SDS-samples intended for MudPIT analysis were recombined after in-gel digestion and also dried to minimal volume as illustrated in Figure 3.1.

3.2.6 HPLC

SDS-sample peptides were analyzed by Reverse-Phase (RP) LC MS/MS as previously described,[177] (further details are described below) using an HPLC elution coupled to the ESI source of a Thermo-Finnigan LTQ linear ion-trap mass spectrometer (Thermo Scientific, San Jose, CA). Recombined SDS-sample peptides and Trypsin-sample peptides were analyzed by 2-D LC MS/MS (MudPIT)[88] also using the LTQ instrument. Four stock buffer solutions were used for both RP and 2-D LC MS/MS analyses, consisting of water with 0.1% formic acid (Buffer A), acetonitrile with 0.1% formic acid (Buffer B), 250mM ammonium sulfate (Buffer C), and 1.5 M ammonium sulfate (Buffer D). Flow rates of 600 nL per minute were calibrated prior to each run.

3.2.6.1 RP LC MS/MS

Reverse-phase analysis of the 32-gel slice SDS-sample peptides was performed using a single-phase column consisting of 7cm of 5μm Zorbax Eclipse XDB C-18 resin (Agilent Technologies, Palo Alto, CA, USA) packed into a 100 μm I.D. fused silica capillary pulled to a 5 μm tip using a laser puller (Sutter Instrument, Novato, CA). Each sample from the gel-slice extraction was injected by direct bomb-loading of the capillary using 500 psi UHP helium gas or by HPLC injection using a Surveyor autosampler (Thermo Scientific) with Buffer A to deposit the sample on the C-18 column packing

material. Elution of the sample peptides from the C-18 column was done using a 30 minute gradient of 5 to 50% Buffer B (95 to 50% Buffer A) followed by a 5 minute gradient to 95% Buffer B (5% Buffer A), and then 5 minute at 95% B. After the Buffer B gradient, the column was re-equilibrated with a wash of 95% Buffer A for 15 minutes to prepare the C-18 material for the next run.

3.2.6.2 Two-D LC MS/MS (MudPIT)

MudPIT analysis of SDS-sample peptides (from recombined 32-gel slice samples) and peptides from trypsin-samples was performed by loading on a dual-phase column consisting of 5 cm of 5µm polysulfoethyl-A strong cation exchange (SCX) resin (PolyLC Inc., Columbia, MD) upstream of 7cm of 5µm Zorbax Eclipse XDB C-18 resin (Agilent) also packed in a 100 µm capillary as described above. Samples were injected onto the column by direct bomb-loading of the capillary using 500 psi UHP helium gas or by injection using a Surveyor autosampler (Thermo Scientific) with Buffer A to deposit the sample on the SCX phase of the column. Peptides that did not deposit on the SCX were eluted off the RP material by a 5-50% gradient of Buffer B (95-50% Buffer A) over 90 minutes followed by a 50-98% gradient (50-2% Buffer A) over 5 min and a 5 min wash of 95% Buffer B (5% Buffer A), followed by a 20-min re-equilibration using 95% Buffer A (5% Buffer B). Peptides that were deposited on the SCX were eluted in a series of 11 salt steps (from 10-100% Buffer C and 50% Buffer D) consisting of a 5 min pulse of the salt followed by a 7 min wash of 95% Buffer A (5% Buffer B) prior to a 60 min gradient from 5-50% Buffer B (95-50% Buffer A), followed by 50-98% Buffer B (50-2% Buffer

A) over 5 min and a 5 min wash of 95% Buffer B. After the Buffer B gradient, the column was re-equilibrated with a wash of 95% Buffer A for 20 min to prepare the C-18 material for the next salt elution step.

3.2.7 MS/MS analysis

Peptide samples separated as described above were ionized by electrospray voltage of 1.6-2.1 kV applied using a gold or platinum electrode attached to a liquid junction upstream of the packing material. Peptides introduced into the mass spectrometer were scanned over the mass-to-charge ratio (m/z) range from 400 to 2000. This m/z range allows for the identification of peptides up to a mass of 6000 daltons if the peptide carries a +3 charge, or 4000 daltons for a +2 ion. Utilizing data-dependent data acquisition, the seven most abundant peaks were automatically selected for fragmentation in the second round of MS using automatic peak recognition and a 30-second dynamic exclusion window after a maximum of 5 selections of the same parent ion. This dynamic exclusion window prevents the mass spectrometer from repeatedly selecting the same high abundant parent ions and allows for selection and fragmentation of lower intensity ions. Using these settings, the mass spectrometer runs one MS scan, followed by seven MS/MS scans (as long as there are seven ions of high enough intensity and not being excluded) before repeating the process with another MS scan. Parent ions were fragmented by RF excitation and collision induced dissociation with helium background gas at approximately 0.6 to 0.8 $\times 10^{-5}$ torr pressure in the ion trap. Data were

continually collected by Xcalibur instrument software version 1.4 SR1 (Thermo Scientific).

3.2.8 Protein database search algorithms

MS/MS data produced as described were analyzed using the SEQUEST database search algorithm[82, 178] against a FASTA database consisting of common contaminants (trypsin, human keratin, protein standards for MS calibration such as bovine serum albumin and angiotensin, etc) followed by the *C. posadasii* and *C. immitis* sequences with protein sequences of 18 more fungi as shown in Table 3.1. Xcalibur .raw files were searched using TurboSEQUEST (BioWorks v 3.1) on a 16-processor IBM Beowulf cluster. DTA files were generated by SEQUEST according to the following criteria: Peptide MW Range = 400-3500 Da; Threshold = 100;(the minimum abundance of the parent ion required to generate a file) Precursor Mass = 1.50 (search for all peptides in the database that have a mass +/- 1.5 daltons of the detected ion); Group Scan = 42 (a window of MS/MS scans where multiple spectra are averaged for the same parent ion appearing multiple times); Minimum Group Count = 2;(the minimum number of spectra to be averaged) and Minimum Ion Count = 10 (the minimum number of ions in the MS/MS scan of a parent ion required to generate a file). SEQUEST searches were performed with no enzyme specified utilizing the default search parameters except: peptide mass tolerance = 1.5 Da; max number of modified amino acids per differential modification in a peptide = 4; static modification of +57.0 Da for carbamidomethylated cysteine; a differential residue modification of +16.0 Da for oxidized methionine;

Table 3.1 Cocci protein sequence database

Organism	Strain	Predicted protein sequences	Source[a]	Version date
Coccidioides posadasii	C735	7202	TIGR	9/1/2005
Coccidioides immitis	RS	10457	Broad	4/26/2006
Uncinocarpus reesii	Unknown	7798	Broad	5/12/2006
Botrytis cinerea	Unknown	16448	Broad	4/26/2006
Sclerotinia sclerotiorum	Unknown	14522	Broad	4/21/2006
Stagonospora nodorum	SN15	16597	Broad	3/14/2006
Neurospora crassa	Unknown	10620	Broad	5/18/2006
Magnaporthe grisea	Unknown	11109	Broad	10/27/2003
Fusarium graminearum	PH-1	11640	Broad	4/24/2006
Chaetomium globasum	CBS 148.51	11124	Broad	4/26/2006
Aspergillus nidulans	FGSC A4	9541	Broad	10/27/2003
Aspergillus fumigatus	AF293	9926	TIGR	7/25/2003
Aspergillus terreus	NIH 2624	10406	Broad	5/11/2006
Candida lusitaniae	ATCC 42720	5940	Broad	4/26/2006
Saccharomyces cerevisiae	Unknown	6714	SGD	5/12/2006
Kluyveromyces lactis	CLIB210	5327	Geno	5/22/2006
Yarrowia lipolytica	CLIB99	6463	Geno	5/22/2006
Schizosaccharomyces pombe	Unknown	4992	Sanger	5/2/2006
Cryptococcus neoformans	H99	7302	Broad	4/26/2006
Rhizopus oryzae	RA 99-880	17467	Broad	4/21/2006

a) Protein sequence sources: TIGR: The Institute for Genomic Research (www.tigr.org); Broad: The Broad Institute (www.broad.mit.edu); SGD: The *Saccharomyce* Genome Database (www.yeastgenome.org); Geno: The Consortium Génolevures (cbi.labri.fr/genolevures); Sanger: The Wellcome Trust Sanger Institute (www.sanger.ac.uk)

maximum of 2 internal cleavage sites; one allowed error in matching auto-detected peaks; and a mass tolerance of +/- 1.0 Da for matching auto detected peaks. Search results from SEQUEST were filtered using DTASelect and Contrast (v 1.9)[179] with the default cutoff parameters (+1 ≥ 1.8, +2 ≥ 2.5, +3 ≥ 3.5, ΔCn ≥ 0.08), specification of at least half-tryptic peptides (meaning each peptide either contains a lysine or arginine residue on the C-

terminal end or the protein sequence contains K or R one residue removed from the N-terminus of the peptide), minimum of one peptide identification per protein and automated removal of all common contaminant identifications.

Data from multiple sample runs (such as MudPIT and RP LC MS/MS) for the same spherule time point were combined using the Contrast function of DTASelect and Contrast to create a combined dataset of all peptide identifications. The spectra corresponding to single-peptide protein identifications were re-searched against the same protein sequence database using the XTandem database search algorithm as previously described,[180] and as presented in Chapter 2 of this dissertation. The combined datasets, including the validated single-peptide identifications, were then analyzed for function prediction as well as vaccine target candidacy as described below. All non-*C. posadasii* peptides were scrutinized for any logical sequence errors (such as D-N substitutions that are isobaric with respect to the MS instrumental mass accuracy) that would allow a *C. posadasii* match. In addition, any DNA point mutations that could result in a changed amino acid sequence (such as a GUU to GUC codon change resulting in a V-A conversion) that matched a *C. posadasii* peptide sequence was kept. Any peptides that could not be justifiably matched to *C. posadasii* were excluded.

3.2.9 Bioinformatics

After compilation of all identified and validated protein identifications, each *C. posadasii* protein sequence (obtained from the FASTA database used for searching) was aligned[104] against all fungal sequences on NCBI to determine putative protein identity

based on homology to identified fungal proteins. Next, protein sequences were aligned against human protein sequences contained in NCBI to determine the level of human homology. Proteins with 50% or greater identity to a human protein sequence (with an E-value of 10^{-4} or less) were placed in the non-candidate category. Proteins with less than 50% but greater than 30% human protein identity were placed in the moderate candidate category. Finally, proteins with less than 30% human protein identity or greater than 10^{-4} expectation value were considered good vaccine candidates based on low human homology. Sequences of proteins in the good and moderate vaccine candidate categories were then analyzed for indicators of cell exterior or surface localization using several analysis tools found on the Expert Protein Analysis System (ExPASy) proteomics server tools page (http://us.expasy.org). Glycosylphosphatidyl-inositol (GPI) linkages were predicted using the Big PI Fungal Predictor (http://mendel.imp.ac.at/sat/gpi/fungi_server.html)[106], or the DGPI algorithm (http://129.194.185.165).[107] N-terminal signal sequences indicating possible extracellular localization were predicted using TargetP[105] (http://www.cbs.dtu.dk/services/TargetP). Protein sequences were analyzed for transmembrane region prediction using the Transmembrane Hidden Markov Model (TMHMM) server v 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

In the search for protein antigen targets to be used for further analysis as vaccine candidates we have used a structured bioinformatic method of estimating protein function and cellular localization. Proteins identified by MS/MS analysis are scrutinized for homology to human proteins followed by analysis of known fungal protein homology for

function determination. Proteins with moderate to low human homology (<50% sequence identity) are then analyzed for sequence cues for extracellular localization.

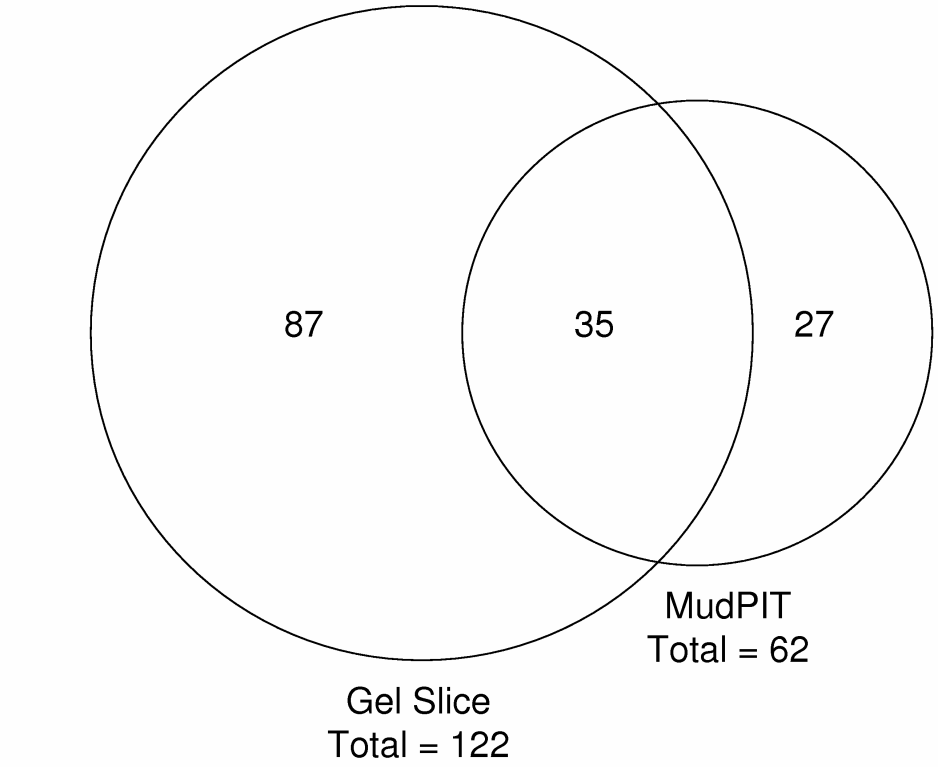
Proteins that fall into the low and moderate human homology categories are subsequently analyzed using web-based prediction algorithms to identify 3 indicators of extracellular localization: N-terminal signal sequence, GPI anchor, and transmembrane helices. Signal sequences indicate protein transport across a membrane after translation, which could correspond to the cell plasma membrane. GPI anchors indicate cell membrane or cell wall association, when combined with the requisite N-terminal signal sequence. Finally, a predicted transmembrane protein may be associated with the cell membrane and contain extracellular regions that may interact with host immune defense mechanisms.

3.3 Results

3.3.1 Comparison of gel-separated 1-D MS/MS and 2-D MS/MS (MudPIT) protein identifications

Both gel-separated 1-D LC and MudPIT methods of protein separation are commonly used in proteomic analyses. While there is certainly overlap between the two methods, there are numerous proteins that were found by only one method. Figure 3.3 shows an example for 120 hour spherules. When combined, these methods provide complementary results. This combined approach was used for all of the SDS-extracted proteins from each spherule time-point analyzed.

Figure 3.3. Comparison of proteins identified from 120 hour spherule cell wall SDS wash using Gel-slice (gel separation with 1-D LC MS/MS) or MudPIT (recombined gel separation with 2-D LC MS/MS) separations. (Data analyzed by SEQUEST using specified parameters and filtered using DTASelect with specified parameters except minimum of 2 peptides per protein identified.)



3.3.2 Spherule cell wall fraction protein identifications

Using a method of detergent extraction of loosely associated cell wall fraction proteins followed by direct trypsin digestion to identify covalently associated proteins we have produced the most comprehensive proteomic analysis of *Coccidioides posadasii* spherule cell walls to date. A total of 645 proteins were identified from three time-points analyzed (48, 96 and 120 hours post-inoculation) from the SDS-sample and trypsin-sample extraction of each. The total list of identified proteins is shown in Table 3.2.

Table 3.2 Total list of proteins identified from comprehensive proteomic analysis of spherule cell wall preparations. Category designations are detailed in Figure 3.5. Human homology designations are described in Section 3.2.9

TIGR locus	Description	Category	Human Homology
10.m00599	Isocitrate dehydrogenase	1	High
10.m00607	Fumarylacetoacetate hydrolase	1	Moderate
10.m00619	Phosphomannomutase	1	High
10.m00701	Enolase	1	High
12.m07458	UQCR subunit	1	Moderate
12.m07607	BGL2	1	Low
12.m07673	Pyruvate decarboxylase	1	Low
12.m07750	Methylglutaconyl-CoA hydratase	1	Moderate
12.m07770	Aldolase	1	Low
12.m07863	Co-A transferase family	1	Moderate
12.m07877	Vacuolar ATP synthase subunit	1	Moderate
12.m07934	Allantoicase	1	Low
12.m08204	ATP synthase subunit	1	Low
13.m01718	Cytochrome C1	1	High
13.m01737	Glutamate decarboxylase	1	Low
13.m01794	Malate synthase	1	Low
13.m01800	2-Me citrate synthase	1	High
13.m01811	2-Me citrate dehydratase	1	Low
13.m01819	ATP synthase, D chain	1	Low
13.m01907	Aldehyde reductase	1	Moderate
13.m01957	Phosphoglycerate kinase	1	High
14.m03050	3-hydroxyisobutyryl-CoA hydrolase	1	Moderate
14.m03111	TIM	1	High
14.m03166	ACR1	1	Moderate
14.m03285	Pyruvate carboxylase	1	High
45.m00866	Kynurenine 3-monooxygenase	1	Moderate
45.m00877	Isocitrate dehydrogenase.	1	Moderate
51.m00579	Cytochrome C peroxidase	1	Low
51.m00597	Lyophospholipase	1	Low
52.m06469	Transketolase	1	Low
52.m06498	Glucokinase	1	Moderate
52.m06581	Alcohol dehydrogenase.	1	Low
52.m06590	Electron transfer flavoprotein	1	High
52.m06668	ATP synthase beta chain	1	High
52.m06707	GAPDH	1	High
52.m06730	ATP synthase subunit	1	High
52.m06735	COX6	1	Moderate
52.m06796	Aldehyde dehydrogenase	1	High

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
52.m06868	Transaldolase	1	High
52.m06950	Ubiquinone-cytochrome C reductase precursor	1	Moderate
52.m06954	Succinate dehydrogenase Fe-S protein	1	High
52.m07021	Glycine cleavage protein	1	Moderate
52.m07023	Vacuolar ATP synthase subunit	1	Moderate
52.m07049	Glycogen phosphorylase	1	High
52.m07105	NADH-ubiquinone oxidoreductase subunit	1	High
52.m07137	Acyl-CoA dehydrogenase family protein	1	High
52.m07142	Succinyl-CoA synthetase	1	High
52.m07286	Plasma membrane ATPase	1	Low
52.m07288	Citrate synthase	1	High
52.m07290	Ubiquinone-cytochrome c reductase	1	Low
52.m07296	NADH-ubiquinone oxidoreductase	1	High
52.m07392	NDPK	1	High
52.m07505	Isocitrate dehydrogenase	1	High
52.m07552	BGL2	1	Low
52.m07616	CUE domain protein	1	Low
60.m01335	Carnitine shuttle protein	1	Moderate
60.m01345	Acetyl-CoA acyltransferase	1	Moderate
60.m01383	Malate dehydrogenase	1	High
60.m01430	Adenylate kinase	1	High
60.m01455	Formate dehydrogenase	1	Moderate
60.m01518	NADH-ubiquinone oxidoreductase subunit	1	Moderate
61.m01556	ACAT	1	Moderate
61.m01632	Succinate dehydrogenase	1	High
61.m01655	PDH E1	1	Low
61.m01710	Cytochrome b2	1	Moderate
65.m01749	G6P isomerase	1	High
65.m01809	Lactate dehydrogenase	1	Moderate
65.m01851	Mitochondrial phosphate carrier protein	1	High
65.m01908	Inorganic pyrophosphatase	1	Low
65.m01929	Isocitrate dehydrogenase	1	High
67.m08291	Aconitase	1	High
67.m08391	Alphaketoglutarate dehydrogenase	1	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
67.m08523	Aspartate aminotransferase	1	High
67.m08575	6-Hydroxy-D-nicotine oxidase	1	Low
67.m08592	NADH-ubiquinone oxidoreductase	1	High
67.m08638	PEPCase	1	Low
67.m08821	Hexokinase	1	Moderate
67.m08849	Altenative oxidase	1	Low
67.m08864	NADH-ubiquinone dehydrogenase subunit	1	High
67.m09060	ATP synthase f chain	1	Low
67.m09061	NADH-ubiquinone oxidoreductase subunit	1	Moderate
67.m09097	Ubiquinone-cytochrome c reductase	1	High
67.m09241	Cytochrome C oxidase subunit	1	Moderate
68.m01805	Enoyl reductase	1	Moderate
68.m01809	Pyruvate dehydrogenase subunit	1	High
68.m01845	NADH-ubiquinone oxidoreductase subunit	1	Low
68.m01886	Succinyl-CoA ligase	1	High
68.m01887	3HB-CoA dehydrogenase.	1	Moderate
68.m01947	Peroxisomal multifunctional beta-oxidation protein	1	Moderate
68.m01964	Probable fumarate reductase	1	Moderate
68.m02031	Aldehyde dehydrogenase	1	High
68.m02054	6PFK alpha subunit	1	Moderate
68.m02071	Cytochrome B2	1	Moderate
72.m01811	ATP sythase subunit	1	Moderate
72.m01865	ETF	1	High
72.m01880	Cytochrome B5 reductase	1	Moderate
72.m01909	Malate dehydrogenase	1	High
72.m01976	Dihydrolipoamide acetyltransferase	1	Moderate
72.m02021	Dihydrolipoamide dehydrogenase	1	High
73.m03409	Alcohol dehydrogenase	1	Low
73.m03439	Cytochrome C	1	High
73.m03469	ATP synthase subunit	1	High
73.m03591	NADH-cytochrome B5 reductase	1	Moderate
73.m03633	NADH-ubiquinone oxidoreductase subunit	1	High
73.m03649	ATP sythase gamma chain	1	Moderate
73.m03658	Aldo-keto reductase	1	Moderate
73.m03664	COX5	1	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
73.m03683	ATP synthase delta chain	1	Moderate
73.m03855	POX	1	Low
73.m03872	ATPase inhibitor	1	Low
73.m03905	Acetamidase	1	Moderate
73.m03957	ATP synthase	1	Low
73.m03967	ATP synthase alpha chain	1	High
10.m00610	ARP2/3	2	Moderate
10.m00647	Importin	2	High
12.m07407	Lipid transfer protein	2	High
12.m07447	Porin protein	2	Moderate
12.m07508	RhoA	2	High
12.m07524	Importin beta-3	2	Moderate
12.m07528	TCP1 eta	2	High
12.m07571	ATP/ADP carrier	2	High
12.m07702	Actin related protein 2/3	2	Low
12.m07706	EB1	2	Moderate
12.m07759	Nhp6	2	Moderate
12.m08095	Mitochondrial carrier protein	2	Moderate
12.m08164	Importin subunit	2	Moderate
12.m08180	Coatomer subunit delta	2	Moderate
13.m01895	GSP1/Ran	2	High
13.m01900	2ODC carrier	2	Moderate
13.m01917	Translocase of inner mitochondrial membrane	2	Low
14.m02878	Tom22	2	Low
14.m02913	GNBP	2	High
14.m02955	ARP3	2	High
14.m03284	Clathrin heavy chain	2	High
14.m03336	Rho-gdp dissociation inhibitor	2	Moderate
45.m00830	GDP dissociation inhibitor	2	High
45.m00843	ER vesicle protein	2	Moderate
45.m00923	SNARE protein	2	Low
51.m00553	Nucleoporin	2	Low
51.m00580	PI transfer protein	2	Low
52.m06433	ATP-binding transport protein	2	Low
52.m06664	Rab11	2	High
52.m06671	Cofilin	2	Moderate
52.m06675	Tubulin subunit	2	High
52.m06684	YPT1	2	High

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
52.m07029	Coatomer subunit beta	2	Moderate
52.m07037	Arp2	2	High
52.m07537	Tubulin subunit	2	High
60.m01419	Actin	2	High
60.m01501	Mitochondrial membrane transport	2	Low
61.m01451	Histone H2A	2	High
61.m01452	Histone H2B	2	High
61.m01685	Clathrin light chain	2	Moderate
65.m01831	Coatomer subunit alpha	2	High
67.m08201	Coatomer subunit gamma	2	Moderate
67.m08284	Sec24 protein	2	Moderate
67.m08515	Importin	2	Low
67.m08516	Sphingolipid transporter	2	Moderate
67.m08537	Tubulin	2	High
67.m08648	Mitochondrial DC carrier protein	2	Moderate
67.m08732	Cell wall glucanase	2	Moderate
67.m08858	Coatomer zeta	2	Moderate
67.m08916	Fimbrin	2	Moderate
67.m09132	Na transfer ATPase	2	Moderate
67.m09175	Histone H1	2	Moderate
67.m09237	Sla2	2	Low
67.m09469	Mitochondrial carrier protein	2	Moderate
67.m09597	SCP-2	2	Moderate
67.m09691	Rab-6	2	High
68.m01839	SAS1	2	High
68.m01855	Nuclear pore protein	2	Low
68.m02055	Profilin	2	Moderate
72.m01795	Carnitine acyltransferase	2	Moderate
72.m01900	Histone H4	2	High
72.m01999	Actin binding protein	2	Moderate
72.m02082	Mitochondrial outer membrane translocase	2	Moderate
73.m03428	SAR1	2	High
73.m03446	MAS5	2	Moderate
73.m03554	Chitobiase	2	Moderate
73.m03557	Actin related protein 2/3 subunit	2	Moderate
73.m03680	Sec23 protein	2	High
73.m03698	Actin-capping protein	2	High
10.m00595	Aminotransferase	3	High

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
10.m00596	DMRL Synthase	3	Low
12.m07356	Iron-sulfur cofactor synthesis protein	3	High
12.m07357	Gamma glutamyl phosphate reductase	3	Moderate
12.m07478	Acetolactase synthase precursor	3	Low
12.m07493	Homocysteine methyltransferase	3	Low
12.m07655	CP2 transcription factor	3	Low
12.m07684	Glutamyl tRNA synthase	3	Moderate
12.m07765	Homoserine dehydrogenase.	3	Low
12.m07776	Oxysterol binding protein	3	Moderate
12.m07785	HMG CoA synthase	3	High
12.m07812	Dolichol-phosphate mannose synthase	3	High
12.m07947	Arginosuccinate Sythase	3	High
12.m07956	Isoleucyl tRNA synthetase	3	High
12.m08029	Isopropylmalate synthase	3	Low
12.m08135	OAR	3	Moderate
12.m08258	Adenylsuccinate lyase	3	High
12.m08268	SPS2	3	Low
12.m08367	Beta 1,3 Glucanase	3	Moderate
13.m01721	DAHP synthase	3	Low
13.m01777	Arg-6 protein	3	Low
13.m01870	Cse1	3	Moderate
13.m01984	Adenosylhomocysteinase	3	Low
14.m02893	Homocitrate synthase	3	Low
14.m02922	Glutamine synthetase	3	High
14.m03150	CDC48	3	High
14.m03170	Saccharopine dehydrogenase	3	Low
14.m03283	Purine biosynthesis protein	3	High
14.m03364	TIP120	3	Low
45.m00893	LSP homolog	3	Low
52.m06442	Thiamine biosynthesis protein	3	Low
52.m06454	Oxygenase	3	Low
52.m06627	Pyridoxine biosynthesis protein	3	Low
52.m06641	Glycyl-tRNA sythetase	3	High
52.m06677	homocysteine synthase	3	Moderate
52.m06691	D3PG dehydrogenase	3	Moderate
52.m06696	Nap1	3	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
52.m06710	DPCK	3	Moderate
52.m06756	Septin	3	Moderate
52.m06882	ATP citrate lyase	3	Moderate
52.m06883	ATP citrate synthase	3	High
52.m06892	R5P isomerase	3	Moderate
52.m06947	Spermidine synthase	3	High
52.m06958	Anthranilate synthase	3	Low
52.m06964	CAP protein	3	Moderate
52.m07022	Adenylsulfate kinase	3	High
52.m07026	P5C dehydrogenase	3	High
52.m07050	GPA3	3	Moderate
52.m07097	GTP binding protein	3	High
52.m07214	Septin	3	High
52.m07224	Imidazole glycerol phosphate synthase	3	Low
52.m07240	SSBP	3	Low
52.m07275	Rvb2	3	Low
52.m07285	Glycosyltransferase	3	Low
52.m07353	Vps54	3	Low
52.m07365	Phosphoribosylformylglycinamidine synthase	3	Moderate
52.m07556	Choline kinase	3	Moderate
52.m07600	Aminotransferase	3	Low
52.m07634	Methionyl-tRNA synthetase	3	High
60.m01291	CF Antigen	3	Moderate
61.m01474	Valyl tRNA synthetase	3	Moderate
61.m01528	Cystathionine gamma-lyase	3	High
61.m01539	Alanyl tRNA synthetase	3	High
61.m01616	RNA1 protein	3	Moderate
65.m01772	Aspartate transaminase	3	High
65.m01793	UDP-N-acetylglucosamine pyrophosphorylase	3	Moderate
65.m01857	G6P dehydrogenase	3	High
65.m01927	MBF1	3	Moderate
65.m02039	Lanosterol 14-alpha-demethylase	3	Moderate
67.m08147	Serine hydroxymethyl transferase	3	High
67.m08206	Aspartyl tRNA synthetase	3	Moderate
67.m08218	Threonyl tRNA synthetase	3	High
67.m08287	Alpha-trehalose-phosphate synthase	3	Low
67.m08320	Anthranilate synthase component	3	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
67.m08358	Anthranilate phosphoribosyltransferase	3	Low
67.m08411	Saccharopine dehydrogenase	3	Moderate
67.m08453	Alpha-aminoadipate reductase	3	Low
67.m08487	RuvB-like helicase	3	High
67.m08624	GDP-mannose pyrophosphorylase	3	Moderate
67.m08724	PIL homolog	3	Low
67.m08726	Thiazole biosynthesis enzyme	3	Low
67.m08761	Phosphogluconate dehydrogenase	3	High
67.m08999	ACAC	3	Moderate
67.m09085	Citrate lyase subunit	3	Moderate
67.m09095	PNPO	3	Low
67.m09210	Transcription elongation factor spt6	3	Low
67.m09233	Phosphoribosylaminoimidazole-succinocarboxamide synthase	3	Low
67.m09268	ATP sulfurylase	3	Moderate
67.m09366	Prolyl tRNA synthetase	3	High
67.m09395	tRNA synthetase cofactor	3	Moderate
67.m09440	DHDP sythase	3	Low
67.m09522	HET C2	3	Moderate
68.m01780	Adenosylmethionine methyltransferase	3	Low
68.m01786	Choline sulfatase	3	Low
68.m01822	Fatty acid synthase subunit	3	Low
68.m01824	Fatty acid synthase beta subunit	3	Low
68.m01995	Threonine dehydratase	3	Moderate
68.m02122	Uridine/cytidine kinase	3	Low
72.m01800	Acetylornithine aminotransferase	3	Moderate
72.m01847	Arginyl tRNA synthetase	3	Moderate
72.m02091	K-A reductoisomerase	3	Low
73.m03359	Tryptophan synthase	3	Low
73.m03363	Mitochondria fission protein	3	Moderate
73.m03531	NPL4 protein	3	Moderate
73.m03695	Threonine synthase	3	Moderate
73.m03726	IMP dehydrogenase	3	High
73.m03772	Cystathionine gamma-synthase	3	Low
73.m03878	Uricase	3	Low

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
73.m03926	Phospho-2-dehydro-3-deoxyheptonate aldolase	3	Low
10.m00641	Proteasome subunit	4	High
12.m07360	SRP1	4	Moderate
12.m07431	Aspartyl protease 4	4	Low
12.m07453	Mannan polymerase II Anp1	4	Low
12.m07472	Homoserine kinase	4	Low
12.m07526	Ribosomal S26E	4	High
12.m07579	Ribosomal S12	4	Low
12.m07590	PP2A	4	High
12.m07592	Ribosomal S3	4	High
12.m07693	eIF3 subunit	4	Low
12.m07699	Ribosomal S7	4	High
12.m07727	Ef 1 alpha	4	High
12.m07738	eIF3	4	Moderate
12.m07741	26S proteasome subunit	4	High
12.m07742	PEP1	4	Moderate
12.m07854	26S proteasome subunit	4	Low
12.m07857	Arginine methyltransferase	4	High
12.m07871	Proteasome subunit	4	High
12.m07876	nop5	4	High
12.m07883	Ribosomal S0	4	High
12.m07994	Bfr1	4	Low
12.m08020	Amn1	4	Moderate
12.m08026	Ribosomal S5	4	High
12.m08103	Ribosomal L17	4	High
12.m08147	Fibrillarin	4	High
12.m08149	MPP	4	High
12.m08165	Ribosomal S22	4	High
12.m08213	Ribosomal S25	4	High
12.m08248	Ribosomal L18	4	High
12.m08443	T-complex protein 1 subunit	4	High
13.m01689	Tcp1 subunit	4	High
13.m01753	BTF3	4	High
13.m01763	RNA helicase	4	High
13.m01820	Nascent polypeptide complex	4	High
13.m01821	Ribosomal L14	4	Moderate
13.m01927	Subtilisin-like protease	4	Low
13.m01936	eIF3 subunit	4	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
14.m02801	Ribosomal L35	4	High
14.m02818	IWS1 transcription factor	4	Low
14.m02833	Ef Tu	4	High
14.m02926	Thioredoxin	4	High
14.m02971	Proteosome subunit	4	High
14.m02974	ATP-dependent protease	4	Moderate
14.m03017	Ran GTP-ase	4	High
14.m03028	PDI	4	Moderate
14.m03171	Ribosomal L43	4	High
14.m03222	Translation initiation factor	4	High
14.m03231	Ribosomal L32	4	High
14.m03249	Ribosomal L36	4	High
14.m03251	Proteasome component	4	High
14.m03267	eIF 2 gamma	4	High
14.m03301	Ribosomal L38	4	High
45.m00822	RNA pol II accessory	4	Low
45.m00895	Ribosomal S11	4	High
45.m00927	Ribosomal S2	4	High
52.m06507	Ribosomal L24	4	High
52.m06670	cytochrome c oxidase	4	Low
52.m06508	Ribosomal S30	4	High
52.m06556	Psi protein	4	Moderate
52.m06573	Ribosomal L2	4	High
52.m06600	Proteasome subunit p45	4	High
52.m06729	VpsA	4	Moderate
52.m06765	Ubiquitin carboxyl-terminal hydrolase	4	Moderate
52.m06832	Ribosomal L26	4	High
52.m06843	Peptide synthase	4	Low
52.m06866	Subtilisin-like protease	4	Low
52.m06871	Snd1 protein	4	Moderate
52.m06922	Dipeptidyl-peptidase 5 precursor	4	Low
52.m06938	Ribosomal L22	4	High
52.m07032	Ubiquitin carboxyl-terminal hydrolase	4	Moderate
52.m07045	Ribosomal L5	4	High
52.m07063	EF 1 beta	4	High
52.m07072	ARF	4	High
52.m07076	SURF4	4	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
52.m07100	PPT1	4	High
52.m07165	N-acetyl transferase subunit	4	High
52.m07274	UFD1	4	Moderate
52.m07309	Ribosomal S3aE	4	High
52.m07317	Ribosomal L23	4	High
52.m07319	PPRF	4	High
52.m07340	CPSF5	4	High
52.m07369	Ribosomal L4	4	High
52.m07545	Ribosomal L6	4	Moderate
52.m07559	Ribosomal L1	4	High
52.m07603	Ribosomal S23	4	High
52.m07610	eIF3	4	Moderate
52.m07630	Ribosomal L9	4	High
52.m07651	Protein disulfide isomerase	4	Moderate
60.m01366	Cap binding protein	4	Low
60.m01381	Ribosomal L28	4	Moderate
60.m01386	Carboxypeptidase Y	4	Moderate
60.m01388	Ribosomal S18	4	High
60.m01397	Mannosyltransferase	4	Moderate
60.m01428	Proteasome subunit	4	High
60.m01438	EF1 gamma	4	Moderate
60.m01475	TOM 40 protein	4	Low
61.m01462	ER oxidoreductin 1	4	Moderate
61.m01472	Casein kinase II	4	High
61.m01540	Ribosomal s15	4	High
61.m01585	Proteasome subunit	4	Moderate
61.m01628	TCP1 delta	4	High
61.m01630	Proteasome subunit	4	Low
61.m01653	TCP1	4	Moderate
61.m01677	RfeF	4	Moderate
65.m01732	DEAD box protein	4	High
65.m01736	Ribosomal L7	4	High
65.m01738	eIF3	4	Moderate
65.m01745	eIF4F	4	Moderate
65.m01791	Ribosomal s24	4	High
65.m01845	eIF3	4	High
65.m01902	Leucine aminopeptidase	4	Low
65.m01907	Proteasome subunit	4	Moderate
65.m01919	eIF2B	4	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
65.m01956	eIF4A	4	High
65.m01975	Ribosomal S28	4	High
65.m02094	Ribosomal S29	4	High
67.m08017	Ribosomal L27	4	High
67.m08092	Mitogen activated protein kinase	4	Moderate
67.m08125	SIK1	4	High
67.m08133	CaMK1	4	Moderate
67.m08170	Hsp 70	4	Moderate
67.m08171	UDP-glucose glycoprotein glucosyltransferase	4	Moderate
67.m08216	Mitochondrial protease	4	High
67.m08219	GAR1	4	High
67.m08299	Ribosomal S10	4	Moderate
67.m08322	Ribosomal L8	4	High
67.m08346	Ribosomal L31	4	High
67.m08456	Ribosomal S6	4	High
67.m08517	Calcineurin	4	Moderate
67.m08518	bZIP	4	Moderate
67.m08661	BipA	4	High
67.m08809	Ribosomal L10	4	High
67.m08947	Ribosomal L19	4	Moderate
67.m09088	Kex protease	4	Moderate
67.m09134	Polyubiquitin	4	High
67.m09156	Calnexin	4	Moderate
67.m09158	Nexin 3	4	Moderate
67.m09183	Proteasome subunit	4	Moderate
67.m09230	SUI1	4	High
67.m09252	eIF2A	4	Moderate
67.m09260	Ribosomal S27	4	High
67.m09261	Proteasome subunit	4	Moderate
67.m09272	Ribosomal L21	4	High
67.m09273	Ribosomal s9	4	High
67.m09340	Ubiquitin activating enzyme	4	High
67.m09348	Ribosomal S4	4	High
67.m09641	Sm D1	4	High
67.m09656	Peptidyl-prolyl cis-trans isomerase	4	High
68.m01784	DnaJ protein	4	Moderate
68.m01865	Proteasome subunit	4	High
68.m01881	Ribosomal L13	4	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
68.m01938	MEP1	4	Moderate
68.m01960	PP2A	4	High
68.m02024	Proteasome subunit	4	Moderate
68.m02038	mRNA capping enzyme	4	Low
72.m01869	cAMP dep. Protein kinase	4	High
72.m01899	Ribosomal L3	4	High
72.m01926	EF-3	4	Moderate
72.m01939	Proteasome component	4	High
72.m01960	Ribosomal S20	4	High
72.m02028	PPI	4	Moderate
73.m03394	PABC	4	Moderate
73.m03452	Ribosomal P0	4	High
73.m03466	eIF3	4	Moderate
73.m03498	PDI related protein	4	Moderate
73.m03507	EF-2	4	High
73.m03510	Ribosomal L12	4	Low
73.m03528	eIF5B	4	High
73.m03534	eIF5A	4	High
73.m03550	Ribosomal S8	4	High
73.m03579	Ribosomal L15	4	High
73.m03592	Ribosomal L18	4	High
73.m03654	DAD1	4	Moderate
73.m03667	Pre-mRNA splicing factor	4	High
73.m03675	Alkaline phosphatase	4	Moderate
73.m03706	26S proteasome subunit	4	High
73.m03724	Ribosomal S21	4	High
73.m03737	NTF2 domain protein	4	Low
73.m03758	P-P cis trans isomerase	4	High
73.m03781	Ribosomal S13	4	High
73.m03789	Pre mRNA splicing factor	4	Moderate
73.m03849	Ribosomal S14	4	High
73.m03851	Ribosomal S16	4	High
73.m03854	Ribosomal S19	4	High
73.m03950	eIF3 Subunit	4	High
12.m07720	Hsp 20	5	Low
12.m08080	Hsp 98	5	Moderate
13.m01705	Dienelactone hydrolase protein	5	Low
13.m01730	Dienelactone hydrolase	5	Low
13.m01916	Hsp 70 co chaperone	5	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
45.m00880	Fe-SOD	5	Low
45.m00953	Thiosulfate sulfurtransferase	5	Moderate
52.m06424	Hsp 90 co-chaperone	5	Moderate
52.m06460	Hsp 10	5	Moderate
52.m06672	Hsp 60	5	High
52.m06880	Woronin body major protein	5	Low
52.m07000	PMP1	5	Moderate
52.m07044	Epoxide hydrolase	5	Moderate
52.m07548	Hsp 78	5	Moderate
61.m01570	Hsp 70	5	Low
61.m01694	Hsp 70	5	High
65.m01782	MGM 101	5	Low
67.m08258	Hsp 20	5	Low
67.m08426	Hsp 90 co-chaperone	5	Low
67.m08567	3',5'-bisphosphate nucleotidase	5	Low
67.m08616	TCP1 epsilon	5	High
67.m09087	Thioredoxin reductase	5	Low
67.m09327	TCP1 alpha	5	High
67.m09426	Formaldehyde dehydrogenase	5	High
68.m02027	Peroxiredoxin	5	Moderate
73.m03535	CPY20 protein	5	Low
73.m03734	Hsp 90	5	High
73.m03913	Hsp 70	5	High
10.m00638	14-3-3 like protein	6	High
10.m00644	Unknown function	6	Low
12.m07479	KH domain protein	6	Moderate
12.m07497	Short chain dehydrogenase.	6	Moderate
12.m07521	Unknown function	6	Low
12.m07533	CBS domain protein	6	Low
12.m07658	Unknown function	6	Low
12.m07661	Unknown function	6	Moderate
12.m07682	Possible Stm1	6	Low
12.m07728	Unknown function	6	Low
12.m07774	Unknown function	6	Low
12.m07800	Possible stress response protein	6	Low
12.m07866	Possible virulence-related protein	6	Low
12.m07980	Unknown function	6	Low
12.m07997	NOL1/NOP2	6	Moderate
12.m08151	Zinc-binding dehydrogenase	6	Moderate

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
12.m08176	RNA splicing factor	6	Moderate
12.m08271	Unknown function	6	Moderate
12.m08273	Unknown function	6	Low
13.m01696	Unknown function	6	Low
13.m01797	Unknown function	6	High
13.m01884	Unknown function	6	Low
13.m01892	Unknown function	6	Low
13.m01911	SPFH domain protein	6	High
13.m01913	GTP binding protein	6	High
13.m01939	HMG box protein	6	Low
13.m01961	ELI Ag1	6	Low
14.m02867	Unknown function	6	Low
14.m02880	Unknown function	6	Low
14.m02927	Dipeptidase	6	Moderate
14.m03177	Unknown function	6	Moderate
14.m03213	Phospholipase D	6	Moderate
14.m03371	Unknown function	6	Low
45.m00878	Unknown function	6	Low
45.m00912	FAD Dependent oxidoreductase	6	Moderate
45.m00934	DUF 52 protein	6	Moderate
45.m00935	Opsin	6	Low
51.m00559	BAR protein	6	Low
51.m00603	Unknown function	6	Moderate
52.m06473	Prohibitin	6	High
52.m06501	Fasciclin domain protein	6	Low
52.m06597	ABC Transporter	6	High
52.m06741	Unknown function	6	Low
52.m06745	Unknown function	6	Low
52.m06817	Unknown function	6	Low
52.m06855	Ag2/PRA	6	Low
52.m06876	Unknown function	6	Low
52.m06891	Unknown function	6	Low
52.m06969	Acetylase	6	Low
52.m07114	Unknown function	6	Low
52.m07155	Glycine rich protein	6	Moderate
52.m07327	Unknown function	6	Low
52.m07368	Possible endoglucanase	6	Low
52.m07424	PQ loop protein	6	Moderate
52.m07484	Unknown function	6	Low

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
60.m01469	DUF410 protein	6	Low
61.m01563	Possible actin binding protein	6	High
61.m01576	Unknown function	6	Low
61.m01622	Possible nucleolin protein	6	Moderate
61.m01714	Unknown function	6	Low
65.m01757	NDP	6	Low
65.m01816	Short chain dehydrogenase	6	Moderate
65.m01827	Unknown function	6	Low
65.m01846	Zinc knuckle protein	6	Low
65.m01850	RNA recognition motif protein	6	Moderate
65.m01966	Tyrosinase-family	6	Low
65.m01987	PCI domain protein	6	Low
65.m02093	Unknown function	6	Low
67.m08055	Unknown function	6	Low
67.m08080	Short chain dehydrogenase	6	Low
67.m08087	KH domain protein	6	Moderate
67.m08111	Possible sexual development protein	6	Moderate
67.m08112	Unknown function	6	Low
67.m08141	Unknown function	6	Low
67.m08239	RNP domain protein	6	Moderate
67.m08372	Unknown function	6	Low
67.m08425	Unknown function	6	Low
67.m08560	Possible NAD dependent epimerase	6	Low
67.m08609	BYS1	6	Low
67.m08637	Possible proteasome subunit	6	Moderate
67.m08694	Mitochondrial GTPase	6	Moderate
67.m08739	Possible mitochondrial protein	6	Moderate
67.m08775	Unknown function	6	Low
67.m08795	KH domain protein	6	Moderate
67.m08867	Prohibitin 2	6	High
67.m08944	5-oxoprolinase	6	High
67.m09080	Oxidoreductase	6	Moderate
67.m09298	Unknown function	6	Low
67.m09306	CHCH domain protein	6	Moderate
67.m09353	Possible RRM domain protein	6	Low
67.m09356	Unknown function	6	Low
67.m09259	NipSnap protein	6	Moderate
67.m09383	CPR	6	Moderate
67.m09402	Possible extracellular matrix protein	6	Low

Table 3.2 Continued

TIGR locus	Description	Category	Human Homology
67.m09585	HIT protein	6	Moderate
67.m09598	Unknown function	6	Low
68.m01796	FAD dependent oxidoreductase family protein	6	Low
68.m01869	Outer membrane protein	6	Moderate
68.m01997	Hypoxia induced protein	6	Moderate
68.m02032	Possible MPD protein	6	Moderate
68.m02050	Possible Nmr-A	6	Moderate
68.m02052	Unknown function	6	Low
72.m01842	Unknown function	6	Low
72.m01860	Annexin	6	Moderate
72.m01879	YagE protein	6	Low
72.m01904	Lectin family protein	6	Moderate
72.m01921	Unknown function	6	Low
72.m01933	Unknown function	6	Low
72.m02032	Unknown function	6	Moderate
72.m02141	Possible pyruvate dehydrogenase e2	6	Low
73.m03525	Possible signal recognition particle	6	Low
73.m03585	NifU-like protein	6	Moderate
73.m03589	HMG box protein	6	Low
73.m03639	Unknown function	6	Low
73.m03646	Unknown function	6	Low
73.m03696	Unknown function	6	Low
73.m03719	Possible monoamine oxidase	6	Low
73.m03727	Unknown function	6	Low
73.m03765	C2 domain protein	6	High
73.m03826	Unknown function	6	Low
73.m03828	Possible vip1 protein	6	Moderate
73.m03871	14-3-3 like protein	6	High
73.m03891	Nucleic acid binding protein	6	Moderate
73.m03901	Unknown function	6	Moderate
73.m03922	GMC oxidoreductase	6	Low
73.m03944	SH3 domain protein	6	Moderate
73.m03949	Possible HHE cation domain protein	6	Low
73.m03954	ATG15	6	Low
73.m03980	Unknown function	6	Low
73.m03981	Possible Sec31 protein	6	Low
9.m00307	TCTP protein	6	Moderate

Each protein on this list was analyzed for human protein homology and separated into three categories: low human homology (<30% sequence identity), moderate human homology (between 30-50% sequence identity), and high homology (>50% sequence identity). The total numbers in each category of human homology are illustrated in Figure 3.4. The function of each identified protein, as determined by known fungal protein homology, was divided into six categories. Descriptions of these categories are shown in the figure legend of Figure 3.5. The separation of all identified cell wall associated proteins into these categories is shown in Figure 3.5.

3.3.4 Protein antigen target evaluation

Using the bioinformatic analysis described in Section 3.2.9, we produced a list of 74 potential vaccine antigen targets. These proteins were scrutinized further and the identifications were divided into three categories: Good target candidates (Table 3.4) Potential target candidates (Table 3.5) and Poor candidates that were not worth further analysis (data not shown). Proteins were divided into these categories by investigation of sequence homology with known fungal proteins for additional information regarding cell localization, as well as for identification of homology with known fungal proteins exhibiting antigenic activity or verified extracellular localization. Good target candidates are those proteins that contain predicted GPI anchor locations, or are homologous to other known secreted or antigenic proteins. Potential candidates are those proteins that contain N-terminal signal sequences or transmembrane regions but cannot be definitively localized. Poor candidates are the proteins that are determined to be localized internally to the cell by homology to other fungal proteins or by functional localization, as well as

Figure 3.4 Human homology of all proteins identified from spherule cell wall preparation. Each category includes total number of proteins in the category, along with percentage of the total proteins. Black indicates high homology, gray is moderate and white is low (as described in text)

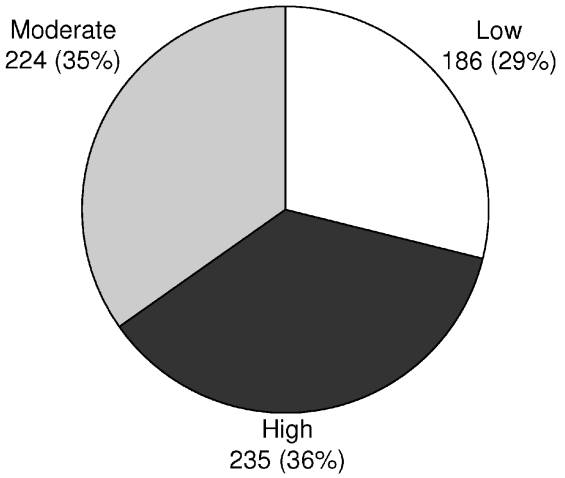


Figure 3.5 Functional categories of proteins identified from cell wall preparation. (Numbers represent the number of proteins in each category, number in parenthesis is percentage of total proteins (645))

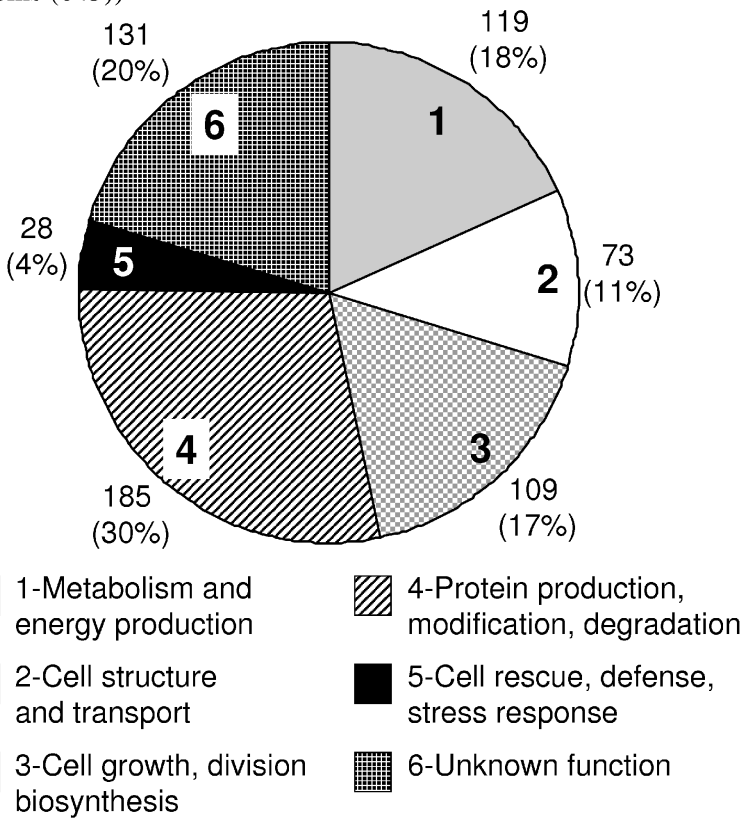


Table 3.3 Identification of good protein antigen targets in spherule cell wall proteomic analysis

Protein Name	Description	TIGR Locus	Broad Locus	Sequence Coverage	Identified Peptides	Sample identified from[a]	Sequence Predictions[b]	Antigen Potential[c]
PepC	Vacuolar Serine Protease	52.m06866	CIMG 06672	25.10%	Total of 14 peptides	96Trp, 96SDS,	N-Terminal Signal Sequence	Homolog is major allergen in *Aspergillus fumigatus* and *Penicillium chrysogenum*
					AFKGSAANMSLGGGK	96Trp		
					DEPVTENNAPWGLAR	96SDS, 120Trp, 120SDS		
					DSLTMGTFNK	96SDS, 120SDS		
					AVTVGASTLADER	96SDS		
					CTDIFAPGLNILSTWIGSK	96SDS, 120SDS		
					GSAANMSLGGGK	96SDS		
					VIDTGTNIEHVDFEGR	120Trp, 120SDS		
					IDTGTNIEHVDFEGR	120Trp		
					DTGTNIEHVDFEGR	120Trp		
					YGVAKKANVYAVK	120Trp		
					KANVYAVK	120Trp, 120SDS		
					KANVYAVKVLR	120Trp		
					PVTENNAPWGLAR	120SDS		
					PTGDDDVDGNGHGTHCSGTVAGK	120SDS		
Lap1	Leucine Amino-peptidase	65.m01902	CIMG 07629	13.10%	Total of 6 peptides	48SDS, 96SDS	N-Terminal Signal Sequence	Homolog in *Trichophyton rubrum* is secreted into extracellular space
					GTLGGVYDNTVPVVGISR	48SDS		
					SDYDAFIQK	48SDS, 96SDS		
					GAAYAAAEYAR	48SDS, 96SDS		
					LQSFADANGK	96SDS		
					TGYYDVTK	96SDS		
					TTWNVLAQTK	96SDS		
Unk1	Unknown function protein	67.m08112	CIMG 01935	32.60%	Total of 4 peptides	96SDS, 120Trp	N-terminal Signal Sequence, Trans-membrane region	Contains predicted RGD cell-attachment sequence suggesting association with extracellular components
					KFEEDASR	96SDS		
					DAEDIGAQAGAR	96SDS		
					VDEGVDKARDTAK	120Trp		
					SGVSSWFGGGKK	120Trp		

[a] Summary of which proteomic analyses the peptides were identified from, numbers indicate the hours post-inoculation, SDS indicates cell pellet wash with SDS, and Trp indicates direct digestion of cell pellet with trypsin.

[b] Summary of predicted extracellular localization clues from protein sequence analysis.

[c] Summary of evidence for antigenic activity or extracellular localization of identified protein.

Table 3.3 Identification of good protein antigen targets in spherule cell wall proteomic analysis (cont.)

Protein Name	Description	TIGR Locus	Broad Locus	Sequence Coverage	Identified Peptides	Sample identified from[a]	Sequence Predictions[b]	Antigen Potential[c]
Unk2	Unknown function protein	65.m02093	CIMG 07502	28.30%	3 peptides	96SDS, 120SDS	N-terminal Signal Sequence, GPI anchor	Contains predicted GPI anchor with requisite N-terminal signal sequence
					ASWCDAQSSTCPK	96SDS, 120SDS		
					LTFECVCGDGTVPK	96SDS, 120SDS		
					NTLPYFVCLANFGQCIDNHPNDLEGQR	96SDS		
Sub	Subtilase serine peptidase	13.m01927	CIMG 10287	8.50%	Total of 3 peptides	48SDS	N-terminal Signal Sequence	High sequence similarity to several secreted proteases from *Trichophyton rubrum*
					LGHGTHVAGTIASK	48SDS		
					SVVNMSLGGGR	48SDS		
					IPGNEVCNR	48SDS		
Unk3	Unknown function protein	73.m03696	CIMG 04772	21.10%	Total of 3 peptides	96SDS	N-terminal Signal Sequence, GPI anchor	Contains predicted GPI anchor with requisite N-terminal signal sequence
					AACVGVPCPDDAMANK	96SDS		
					TVQCAAQCDQGNGSPAAIEAYAK	96SDS		
					AQCDQGNGSPAAIEAYAK	96SDS		
Fdc	Fasciclin domain containing protein	52.m06501	CIMG 05852	5.00%	Total of 2 peptides	48SDS	N-terminal Signal Sequence	Contains Fasciclin domain predicted to function in cell-cell and cell-extracellular matrix interactions
					TLLEQSELGDHDQR	48SDS		
					VNAFLFSR	48SDS		
DppV	Dipeptidyl dipeptidase 5	52.m06922	CIMG 06782	1.80%	Total of 1 peptide	48SDS	N-terminal Signal Sequence	Homolog in *Aspergillus fumigatus* identified as antigen
					YSGIGESNPNAVK	48SDS		
Gh16	Glycosyl hydrolase family 16 protein	67.m08732	CIMG 01102	3.40%	Total of 1 peptide	48Trp	N-terminal Signal Sequence, GPI anchor	Contains predicted GPI anchor with requisite N-terminal signal sequence
					NIVITDYSTGKEYR	48Trp		

[a] Summary of which proteomic analyses the peptides were identified from, numbers indicate the hours post-inoculation, SDS indicates cell pellet wash with SDS, and Trp indicates direct digestion of cell pellet with trypsin.

[b] Summary of predicted extracellular localization clues from protein sequence analysis.

[c] Summary of evidence for antigenic activity or extracellular localization of identified protein.

Table 3.4 Identification of potential protein antigen targets in spherule cell wall proteomic analysis

Protein Name	Description	TIGR Locus	Sequence Coverage	Identified Peptides	Sample identified from[a]	Sequence Predictions[b]
Crp	Cysteine-Rich Protein	67.m08055	11.80%	Total of 3 peptides	48SDS, 96SDS	N-Terminal Signal Sequence
				ICAQWCADNFENPGR	48SDS	
				QSESQKLCDEQCR	48SDS, 96SDS	
				LSLLHILR	48SDS, 96SDS	
Unk 4	Unknown function protein	52.m07368	13.00%	Total of 6 peptides	48SDS, 96Trp, 96SDS, 120SDS	N-Terminal Signal Sequence
				TPYPYGPDTLTNGPLMNDLADFPCK	48SDS, 96SDS, 120SDS	
				VIHSIEGGCPANVDGNLPAD	48SDS, 96SDS	
				VPSSIAPGEYTLAWTWINR	48SDS	
				EFYMNCAPITVTASK	48SDS, 96SDS	
				QRPGVYDPPAQENIFAIGEEQTLSFR	96Trp, 96SDS	
				VIHSIEGGCPANVDGNLPADPNGDGASK	48SDS, 96SDS, 120SDS	
Unk 5	Unknown function protein	67.m08141	29.00%	Total of 4 peptides	96SDS	N-Terminal Signal Sequence, Transmembrane region
				LNQNGGGISIK	96SDS	
				DISQVEYTLDGQK	96SDS	
				VWYDLSNIDGYPFK	96SDS	
				DGGVSIAPSDSSCPK	96SDS	
Unk 6	Unknown function protein	73.m03826	10.50%	Total of 3 peptides	48SDS, 120Trp	N-Terminal Signal Sequence, Transmembrane region
				SAETEGLVAQK	48SDS	
				DEAEVQSWLK	48SDS	
				AKSAETEGLVAQK	120Trp	
Unk 7	Unknown function protein	67.m09298	20.40%	Total of 2 peptides	48SDS, 96SDS, 120SDS	N-Terminal Signal Anchor, Transmembrane region
				SPTTVTSESNIR	48SDS, 96SDS, 120SDS	
				SAPSALNEVVIPNVTLPK	96SDS, 120SDS	

[a] Summary of which proteomic analyses the peptides were identified from, numbers indicate the hours post-inoculation, SDS indicates cell pellet wash with SDS, and Trp indicates direct digestion of cell pellet with trypsin.

[b] Summary of predicted extracellular localization clues from protein sequence analysis.

Table 3.4 Identification of potential protein antigen targets in spherule cell wall proteomic analysis (cont.)

Protein Name	Description	TIGR Locus	Sequence Coverage	Identified Peptides	Sample identified from[a]	Sequence Predictions[b]
Unk 8	Unknown function protein	67.m08775	15.70%	Total of 2 peptides	48SDS	N-Terminal Signal Sequence
				GLAYNNPDALRPFK	48SDS	
				SILGFNEPDHGDQASMSPESAVDAFK	48SDS	
Unk 9	Unknown function protein	72.m01921	13.20%	Total of 2 peptides	96SDS	Transmembrane region
				GLGSTLLGGAAGGFAGHK	96SDS	
				GANLVGHAAK	96SDS	
Cah	Carbonic Anhydrase	72.m02032	5.80%	Total of 2 peptides	48SDS, 120SDS	N-Terminal Signal Sequence
				AAGGIGELPNFDYGPVR	48SDS, 120SDS	
				IPIEPVGFVR	120SDS	
Unk 10	Unknown function protein	12.m07728	4.00%	Total of 2 peptides	48SDS, 96Trp,	Transmembrane region
				ALLDPEQGEEVWNQILK	48SDS, 96Trp,	
				GVEQVSTGNSTIPHFPK	96Trp	
Unk 11	Unknown function protein	67.m08425	3.60%	Total of 2 peptides	48SDS, 96Trp	Transmembrane region
				DTFSLQDVPK	48SDS	
				HGVADKVQALR	96Trp	
Unk 12	Unknown function protein	67.m09598	11.60%	Total of 1 peptide	48SDS	N-Terminal Signal Anchor, Transmembrane region
				GQTDNPIAPR	48SDS	
Unk 13	Unknown function protein	68.m02052	11.00%	Total of 1 peptide	48Trp	N-Terminal Signal Sequence, Transmembrane region
				ELHRESADTAVTGNGGQAV	48Trp	
Unk 14	Unknown function protein	73.m03639	6.80%	Total of 1 peptide	48SDS, 96Trp	N-Terminal Signal Anchor, Transmembrane region
				LVGAHNSPFVGYLPQHNQEISVTK	48SDS, 96Trp	
Unk 15	Unknown function protein	52.m06817	4.50%	Total of 1 peptide	48SDS	N-Terminal Signal Sequence
				EGVALLNVDDLDAVK	48SDS	

[a] Summary of which proteomic analyses the peptides were identified from, numbers indicate the hours post-inoculation, SDS indicates cell pellet wash with SDS, and Trp indicates direct digestion of cell pellet with trypsin.

[b] Summary of predicted extracellular localization clues from protein sequence analysis.

those proteins that have been previously analyzed as vaccine candidates. Examples of this include membrane proteins that localized to subcellular organelle membranes, or proteins with predicted N-terminal signal sequences that stay within the endoplasmic reticulum, and are not transported across the plasma membrane. A summary of the supporting information for each good protein vaccine candidate identified is presented below.

3.3.4.1 Vacuolar serine protease (52.m06866) PepC

This protein was identified in 96Trp, 96SDS, 120Trp, 120SDS from 14 peptides comprising 25.1% sequence coverage (125/498 residues). It contains a predicted N-terminal signal sequence and two possible N-glycosylation sites, both of which suggest extracellular association. PepC has 27% sequence identity to human subtilisin/kexin type 9, and 72% sequence identity (84% sequence similarity) to the opportunistic fungal pathogen *Aspergillus fumigatus* serine protease Alp2 (genbank Y13338).[181] PepC also has 71% sequence identity (83% similarity) to the vacuolar serine protease Pen ch 18 (genbank AF263454) from the fungal allergen *Penicillium chrysogenum*.[182] Both the *A. fumigatus* and *P. chrysogenum* homologs are known to be allergens, which suggests that *C. posadasii* PepC may also interact with host immune defenses and thus may function well as a vaccine component.

3.3.4.2 Leucine aminopeptidase (65.m01902) Lap1

This protein was identified in 48SDS and 96SDS samples from 6 peptides comprising 13.1% sequence coverage (66/503 residues), and contains a predicted N-terminal signal sequence and 3 predicted N-glycosylation sites. Lap1 has 27% sequence identity to human folate hydrolase and 61% sequence identity (76% similarity) to the dermatophyte *Trichophyton rubrum* Lap1/Lap2 (there is some confusion as to the correct name of this protein. It is known as Lap2 in the paper,[183] but the sequence is designated Lap1 in genbank AY496929). *T. rubrum* Lap2 (as identified in the paper) is identified in cell culture supernatant by western blot using antisera specific to Lap2, indicating that the protein is secreted from the cell. This finding, along with the predicted N-terminal signal sequence and putative glycosylation sites suggest this protein may be secreted from the cell and potentially interact with host immune defenses.

3.3.4.3 Unknown 1 (67.m08112) Unk1

This protein was identified in 96SDS and 120Trp from 4 peptides comprising 32.6% sequence coverage (35/138residues), and exhibits 32% sequence identity (32/99 of 138 residue protein expectation score = .016) to human myosin-4. This protein matches only hypothetical and predicted proteins of unknown function from fungal BLAST search. Unk1 contains a predicted transmembrane helix, a predicted N-terminal signal sequence, as well as a RGD cell attachment sequence indicating the possibility that Unk1 is an external protein involved in cell-cell or cell-extracellular matrix interactions.[184]

3.3.4.4 Unknown 2 (65.m02093) Unk2

This protein was identified in 96SDS and 120SDS from 4 peptides comprising 28.3% sequence coverage (54/191 residues) and has 31% sequence identity (18/57 of 191 residue protein; expectation score = 1.8) to human muscle alpha kinase. This protein matches only hypothetical and predicted proteins of unknown function from fungal BLAST search. Unk2 contains a predicted N-terminal signal sequence, and GPI anchor indicating cell membrane or cell wall anchoring.

3.3.4.5 Subtilisin-like serine protease (13.m01927) Sub

This protein was identified in only 48 hour spherule SDS wash from 3 peptides comprising 8.5% sequence coverage (34/400 residues), and contains a predicted N-terminal signal sequence and 3 predicted N-glycosylation sites. Sub exhibits 28% sequence identity to human convertase subtilisin/kexin type 9, and has significant homology to a subtilisin gene family (Sub1 through-Sub7) identified in the dermatophyte *Trichophyton rubrum.*[185] *C. posadasii* Sub has 47% sequence identity (62% similarity) to Sub3,and 45% identity(64% similarity) to Sub4, both of which were detected in supernatant of cultured *T. rubrum* cells indicating secretion. These findings, with the predicted signal sequence and glycosylation suggest Sub may be a secreted serine protease, and may function as a vaccine component.

3.3.4.6 Unknown 3 (73.m03696) Unk3

This protein was identified in 96SDS from 3 peptides comprising 21.1% sequence coverage (39/185 residues) and has 41% sequence identity (13/31 of 187 residue protein expectation score = 0.70) to the human rootletin protein. Unk3 matches only hypothetical and predicted proteins of unknown function from fungal BLAST search. This protein contains a predicted N-terminal signal sequence, and GPI anchor indicating cell membrane or cell wall anchoring.

3.3.4.7 Fasciclin domain containing protein (52.m06501) Fdc

This protein was identified in 48SDS from 2 peptides comprising 5% sequence coverage (22/443 residues) and has 27% sequence identity to human transforming growth factor protein. Fdc has 43% identity (61% similarity) to fasciclin domain family protein in *Aspergillus fumigatus,* and contains a predicted N-terminal signal sequence, as well as four predicted N-glycosylation sites. Fdc contains a fascicilin domain, which is predicted to function in cell-cell and cell-extracellular matrix interactions.[186] This domain, combined with the predicted signal sequence and glycosylation sites, suggests *C. posadasii* Fdc may be involved in extracellular interactions.

3.3.4.8 Dipeptidyl-peptidase V (52.m06922) DppV

This protein was identified from only one peptide predicted from two spectra from 48hr spherule SDS wash. This protein has 21% sequence identity to human peptide hydrolase, and 64% sequence identity (78% similarity) to *A. fumigatus* Dipeptidyl

dipeptidase V (DppV). The *A. fumigatus* DppV is a protein recognized by sera from human and mouse aspergillosis patients,[187] suggesting the *C. posadasii* homolog may also elicit an immune response.

Due to the single-peptide identification, this MS/MS spectrum was re-analyzed both manually (shown in Figure 3.6) and using the search algorithm Mascot (searching against the NCBI non-redundant database). All three database search algorithms employed agree on the same peptide identification. The results of these searches were: SEQUEST (+2 ion XCorr = 3.916), XTandem (log(e)= -7.1) and Mascot (score = 37 vs nr database). Manual analysis of the spectrum shows a good b and y-ion series coverage. Based on this analysis, we have confidence in this peptide identification.

3.3.4.9 Glycosyl hydrolase family 16 (67.m08732) Gh16

This protein was identified from one peptide predicted from two spectra from 48hr spherule cell wall trypsin digestion, with 33% sequence identity (39/115 of 417 residue protein expectation score = .040) to human RNA polymerase II. Gh16 exhibits 49% sequence identity (64% similarity) to putative cell wall glucanase of *Aspergillus clavatus* Gh16 contains a predicted N-terminal signal sequence as well as a predicted GPI-anchor. These predictions, combined with the homology to a predicted cell wall associated protein of *A. clavatus* suggest Gh16 is extracellularly associated. It is worth mentioning that this protein does not match any of the 24 glycosyl hydrolase family proteins predicted from the *C. immitis* genome analysis undertaken by Cole and Hung in 2001.[188]

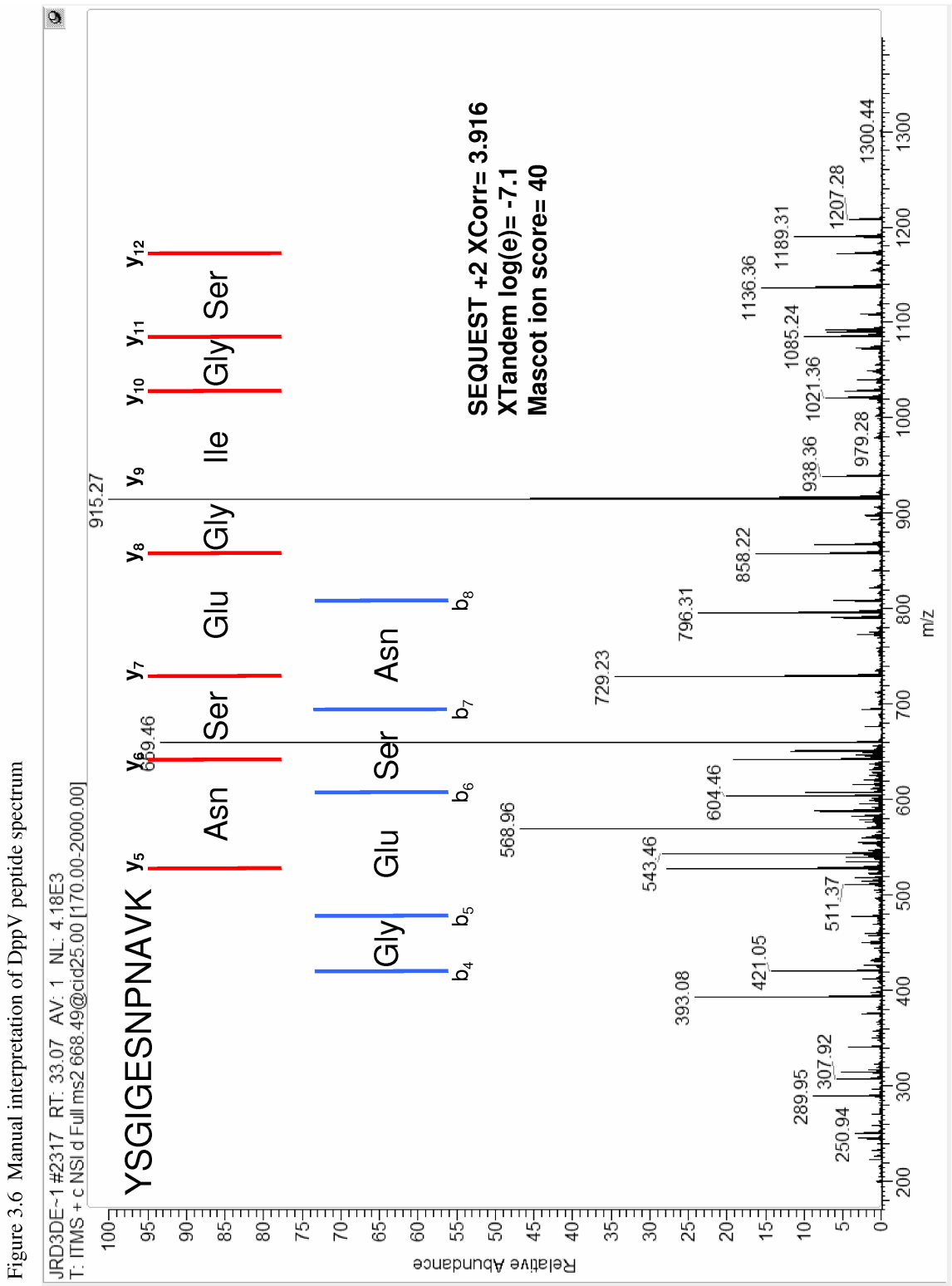


Figure 3.6 Manual interpretation of DppV peptide spectrum

Due to the single-peptide identification, this MS/MS spectrum was re-analyzed both manually (shown in Figure 3.7) and using the search algorithm Mascot (searching against the NCBI non-redundant database). All three database search algorithms employed agree on the same peptide identification for this protein. The results of these searches were: SEQUEST (+2 ion XCorr = 4.1794), XTandem (log(e)= -5.2) and Mascot (score = 44 vs nr database) all agree on identification. Manual analysis of the spectrum shows a good b and y-ion series coverage. Based on the manual interpretation of the spectrum, we have confidence in this peptide identification.

3.3.5 Known antigen identification

Several of the previously identified protein antigens from *Coccidioides* (as described in section 1.1.5.3) were found in our comprehensive spherule cell wall proteomic analysis. A summary of these antigens, along with what peptides and which dataset they were found in is shown in Table 3.5. Some of the previously identified antigens were not found, including the Coccidioides Specific Antigen (CSA), the glucanosyltransferase GEL1, the Cu, Zn, Superoxide Dismutase (SOD), the Spherule Outer Wall glycoprotein (SOWgp), the T-Cell Reactive Protein (TCRP), and Urease (URE).

It is likely that CSA is not identified in our experiments because it is an exoantigen isolated from extracellular extracts of *Coccidioides*. A similar reason may explain the absence of SOWgp in our data as well as URE. In addition, URE is believed to be glycosylated[53], which can preclude identification by mass spectrometry, using the

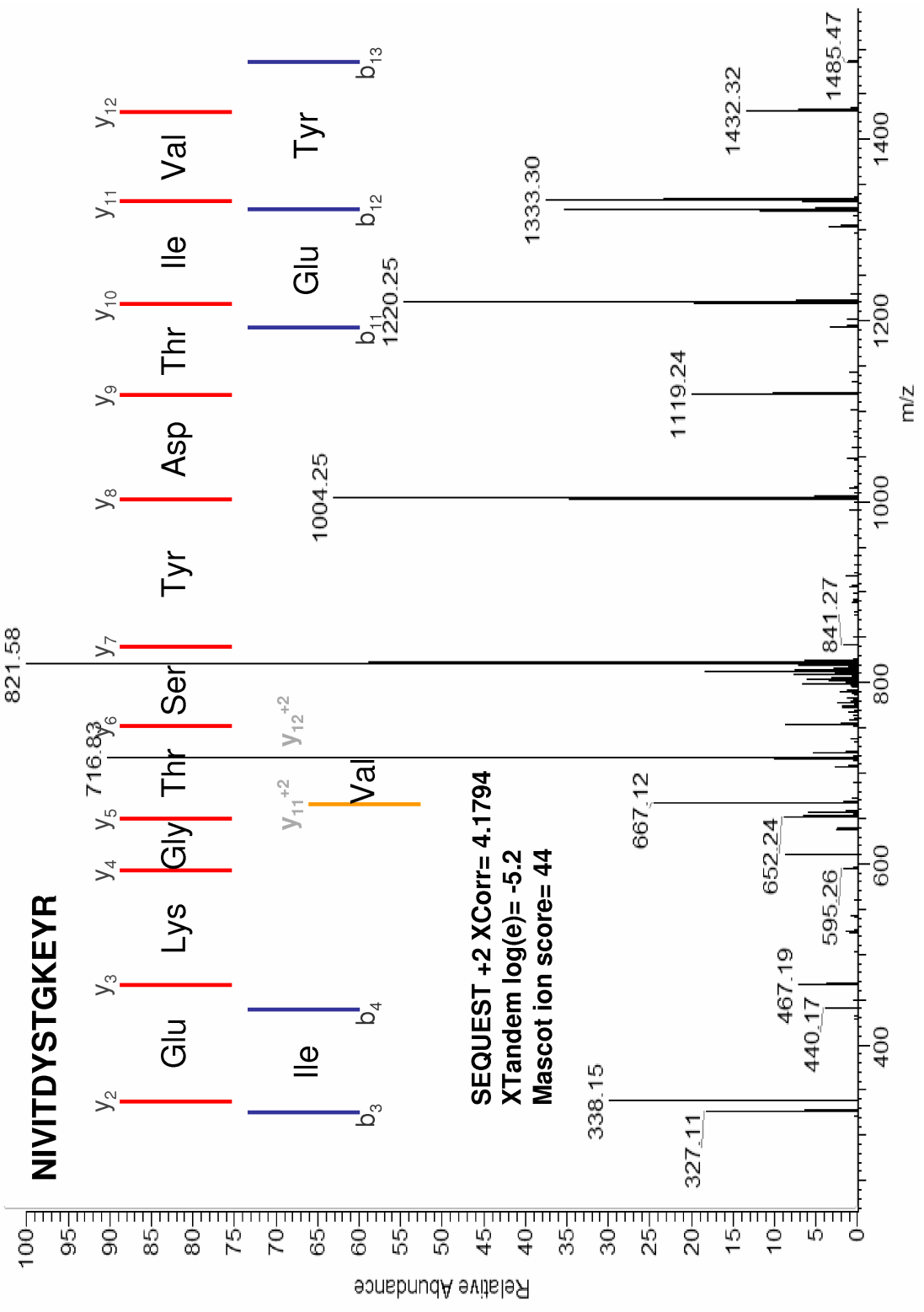


Figure 3.7 Manual interpretation of Gh16 peptide spectrum

Table 3.5 Summary of known antigens identified in cell wall preparation analysis.

Antigen	Description	TIGR locus	Broad locus	Sequence Coverage	Identified Peptides	Identified in
Ag2/PRA	Antigen 2/ Proline Rich Antigen	52.m06855	CIMG 09696	19.6%	CFVEALGNDGCTR	120SDS
					CHCSKPELPGQITPCVEEACPLDAR	120SDS
Amn1	1,2 Alpha Mannosidase	12.m08020	CIMG 03314	8.5%	IGPEGFGWDATK	48SDS
					LSDITGDPEYGR	48SDS
					TIDIETGLFR	48SDS
					VPEAQAEFYK	48SDS
BGL2/TP	β Glucosidase 2/ Tube Precipitin antigen	12.m07607	CIMG 03888	1.0%	HFIVNEQER	48SDS
ELI Ag1	Expression Library Immunization Antigen 1	13.m01961	CIMG 10032	16.5%	FPCGGQSMSK	96SDS
					GFSEDMLTR	48SDS, 96SDS, 120SDS
					QVGLGDFCLPSVSLDEQR	48SDS, 96SDS, 120SDS
Hsp60	Heat Shock Protein 60	52.m06672	CIMG 06278	43.1%	AAVEEGILPGGGTALLK	48SDS, 96SDS, 120SDS
					AALLKGVDTLAK	120Trp
					AHKELKFGVEGR	120Trp
					AISLQDKFENLGAR	48SDS, 96SDS, 120SDS, 120Trp
					AKAVTTTLGPKGR	120Trp
					ASANGLKDVKPANFDQQLGVSIVK	120Trp
					AVTTTLGPKGR	120Trp
					AVTTTLGPKGRNVLIESSYGSPKITK	120Trp
					DVKPANFDQQLGVSIVK	48SDS
					GFDSAKGEYVDMIGAGIVDPLKVVR	120Trp
					GIQAAVDSVVEYLQANK	48SDS, 120SDS
					GIQAAVDSVVEYLQANKR	48SDS, 96SDS, 120SDS
					GQLQVAAVK	48SDS, 96SDS
					GRNVLIESSYGSPK	120Trp
					GYVSPYFITDTK	48SDS, 96SDS, 120SDS
					HLGGLAIITR	96Trp
					ISAVQDIIPALEASTTLR	48SDS, 96SDS, 120SDS
					KAISLQDKFENLGAR	120Trp
					LISNAMER	48SDS, 96SDS
					LLQDVASK	48SDS, 96SDS
					LSGGVAVIK	96SDS
					LTDEFAGDFNR	48SDS, 96SDS
					NVAAGCNPMDLR	48SDS, 96SDS
					NVLIESSYGSPK	48SDS, 96SDS, 120SDS
					SIIADPATSEYEKEK	48SDS, 96SDS, 120SDS
					TNEIAGDGTTTATVLAR	48SDS, 96SDS, 120SDS
					TQKVEFEKPLILLSEK	48SDS
					VEFEKPLILLSEK	48SDS, 96SDS, 120SDS
					VGGASEVEVGEK	48SDS, 96SDS, 120SDS
					VGGASEVEVGEKK	96SDS, 120SDS
					VGKEGVITVKDGK	120Trp
					VVDALNATR	96SDS
					WVVIGDWNYGEGSSR	96Trp
MEP1	Metalloprotease 1	68.m01938	CIMG 08674	17.4%	DGCSVLSSSVPGGSGAPYDLGK	48SDS
					SPSSGCPVGR	48SDS
					TINPSWASDGNEIAMK	48SDS
PEP1	Aspartyl Protease 1	12.m07742	CIMG 03687	40.4%	EGDDSYATFGGVDSSLFSGEMIK	48SDS, 96SDS
					FDGILGLGYDTISVNK	48SDS, 96SDS, 120SDS
					FYTMYDLGNNLVGLAK	48SDS, 96SDS, 120SDS
					IGDLTIEGQDFAEATNEPGLAFAFGR	96SDS, 120SDS
					SWNGQYTVDCNK	48SDS, 96SDS, 120SDS
					VVLDTGSSNLWVPSSECGSIACYLHNK	48SDS, 96SDS
					YDSSASSTYK	96SDS, 120SDS
					YFSEISIGNPPQNFK	96Trp, 120SDS
					YGSGSLSGFVSQDTLR	48SDS, 96SDS, 96Trp, 120SDS
Plb	Phospholipase B	51.m00597	CIMG 08300	6.7%	ALSYQFINAK	48SDS, 96SDS
					FTGSTFLDVLR	48SDS, 96SDS
					VSITDYWGR	48SDS
					YYAQLQSAVAGK	48SDS, 96SDS
PMP1	Peroxisomal Matrix Protein 1	52.m07000	CIMG 05828	56.6%	ACGMPQNYEASK	120SDS
					AGDSFPSDVVFSYIPWTPDNK	96SDS, 96Trp, 120SDS
					AGDSFPSDVVFSYIPWTPDNKDIK	96SDS
					ANGVTGDDILFLSDPEAK	48SDS, 96SDS, 120SDS
					ANGVTGDDILFLSDPEAKFSK	120Trp
					GKANGVTGDDILFLSDPEAK	120Trp
					SIGWNAGER	96SDS, 120SDS
					SYIPWTPDNKDIK	120Trp
					VVLFSLPGAFTPT	120SDS
					YAMIIDHGQVTYA	120Trp

techniques described here. The absence of GEL1 may be explained by its association with endospores and the discovery that its highest levels of expression occur during the late endosporulation stage after 120hrs post inoculation.[17] The lack of identification of TCRP may be explained by its original identification from arthroconidia[64] and thus the possibility that it is not expressed in the pathogenic phase of *Coccidioides*. The missing SOD identification is likely due to the fact that it was identified from a largely cytosolic protein spherule extract (T27K)[41] and is believed to be a cytoplasmic protein (J.M. Lunetta personal communication).

3.3.6 Validation of antigen identification strategy

Each of the known antigens found in our cell wall proteome analysis was subjected to the same bioinformatic analysis used in our vaccine target search. Table 3.6 illustrates the results of this analysis. Of the nine known antigens found in our analysis, only two of them, Hsp60 and PMP1 would not have been identified as potential vaccine antigen targets.

3.4 Discussion

We have described here the first comprehensive proteomic analysis of *Coccidioides posadasii* spherule cell walls in an effort to identify proteins for future testing as coccidioidomycosis vaccine candidates. Utilizing a strategy of mass spectrometry-based proteomic analysis we have identified a total of 645 proteins from multiple spherule samples using a combination of protein extraction and peptide

Table 3.6 Analysis of antigen identification strategy applied to known antigens identified in spherule cell wall analysis.

Antigen	Human Homology	Sequence Predictions	Found in Cell Wall Analysis?	Found by Antigen Identification Strategy?
Ag2/PRA	Low	N-terminal Signal Sequence, GPI anchor	yes	yes
Amn1	Moderate	N-terminal Signal Sequence, Trans-membrane region	yes	yes
BGL2/TP	Low	N-Terminal Signal Sequence	yes	yes
ELI-Ag1	Low	N-Terminal Signal Sequence	yes	yes
HSP60	**High**	N-Terminal Signal Sequence	yes	**no (high human homology)**
MEP1	Moderate	N-Terminal Signal Sequence	yes	yes
PEP1	Moderate	N-Terminal Signal Sequence	yes	yes
PLB	Low	N-terminal Signal Sequence, Trans-membrane region	yes	yes
PMP	Moderate	**None**	yes	**no (no sequence predictions)**

separation techniques. By employing a bioinformatic approach of human protein homology determination and prediction of extracellular localization sequence markers we have narrowed this expansive list down to 23 proteins that have not, to our knowledge, been previously identified as vaccine candidates. These 23 proteins contain sequence cues along with low human protein homology that suggest they are viable protein targets for further testing as vaccine candidates.

We have found nine previously identified *Coccidioides* protein antigens, which indicates that our strategy of protein identification by proteomic analysis is sound. We

have also shown that our bioinformatic analysis strategy of vaccine candidate prediction is relevant and suggests our identified vaccine candidates are worthy of further testing.

We have identified nine good protein vaccine targets based on extracellular localization sequence predictions and/or homology to proteins known to be antigenically active or extracellularly associated in other fungal species. We have identified an additional 14 potential target proteins that have not been found to share homology with known antigenic proteins in other fungi but have somewhat ambiguous localization clues that suggest they may be extracellularly associated. The nine proteins from the list of good targets are ready for immediate analysis as vaccine candidates. The remaining 14 proteins are of considerably higher risk for testing due primarily to their lack of definitive functional identification or localization, but may prove fruitful regardless. We feel confident that the vaccine candidate targets described in this study will prove valuable in advancing the development of a vaccine for coccidioidomycosis.

4 CHAPTER FOUR: DIFFERENTIAL PROTEIN EXPRESSION ANALYSIS OF *COCCIDIOIDES POSADASII* USING STABLE ISOTOPLE LABELING AND TANDEM MASS SPECTROMETRY

4.1 Introduction

The search for a vaccine against coccidioidomycosis is one that has been on going for more than eighty years. This human respiratory infection is caused by the dimorphic fungal pathogens *Coccidioides posadasii* and *C. immitis,* and is endemic primarily to the southwestern United States. These soil-dwelling organisms have been cultured in the laboratory for decades in the search for vaccine components. It is known that vaccines derived from the pathogenic spherule phase of the fungus are more effective in inducing immunity than the saprobic mycelia phase found in the soil.[26, 60] Here we describe a proteomic analysis utilizing stable isotope labeling of proteins for differential quantification of mycelial and spherule proteins. Proteins that are more highly or exclusively expressed in spherule cells compared to mycelial cells are likely candidates for vaccine development.

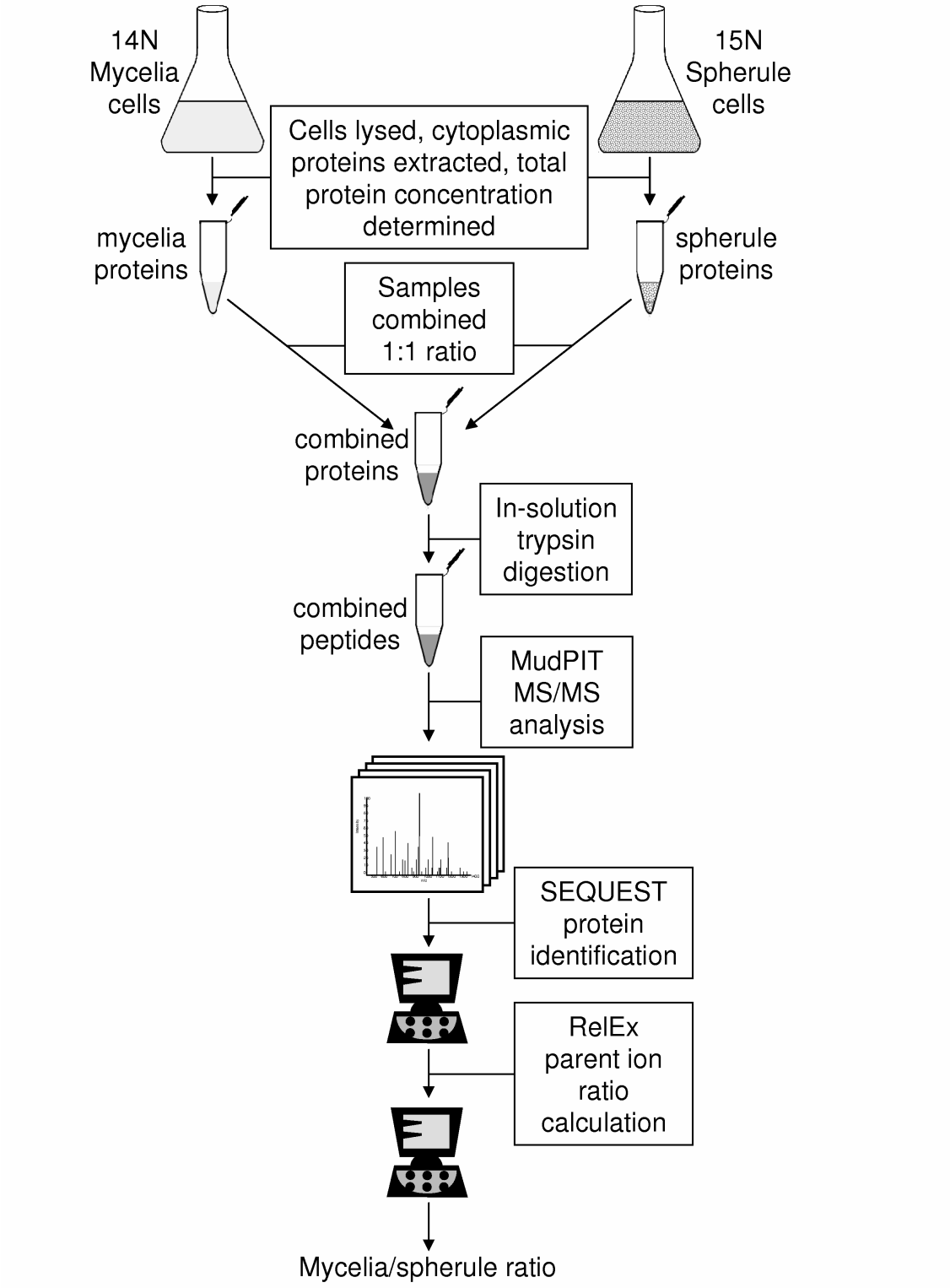
The area of differential proteomics has seen several advances in recent years. Several techniques have been used for differential quantitative analysis using mass spectrometry (a technology that does not alone lend itself well to quantitative measurement). Many of the techniques involve differential labeling of proteins from different cell states using stable isotopes, such as deuterium, ^{15}N, or even ^{13}C. There are two basic methods to label proteins with these isotopes. The first method is known as

biological incorporation, where one cell state is grown on normal media, while the other cell state is grown on media enriched with the isotope being used. The second method is known as chemical incorporation, where an isotopically-labeled tag is added to the protein mixture after cell growth is completed. Each of these methods derive quantitative data from the relative ion intensities of the labeled and unlabelled proteins or peptides or the tags associated with those peptides.[111]

Biological incorporation of ^{15}N is performed by culturing cells in ^{15}N isotopically-enriched media which allows for the incorporation of the isotope into all amino acids produced in the cells.[119, 120] This method produces more complex mass spectra (due to incorporation of at least one ^{15}N incorporated into each amino acid), but allows for detection of protein expression level differences as low as 10%.[115]

Established methods for culturing the two cell morphologies of *Coccidioides posadasii* allow for the isolation of the nitrogen source for cultured spherule cells to allow for ^{15}N incorporation into all proteins produced by the cell. Mycelia cells are cultured in a less defined medium that does not allow for the easy isolation of carbon or nitrogen sources. Due to these limitations, we have performed a differential protein quantification of spherules collected at three life-cycle time points (48, 96 and 120 hrs post-inoculation). These time points represent immature, mature and endosporulating spherules,respectively (See Figure 1.1). The spherule cells are labeled with ^{15}N followed by combination with normal ^{14}N-containing mycelial cells, and analyzing by tandem mass spectrometry. The overall experimental strategy for this study is shown in Figure 4.1.

Figure 4.1 Strategy employed for spherule protein quantification by stable isotope incorporation



4.2 Materials and Methods

4.2.1 Strains and growth conditions-Spherules

Arthroconidia harvested from *Coccidioides posadasii* strain Silveira (isolated in 1951, a gift from H. B. Levine at the University of California, Berkeley) stock cultures were inoculated into 1L of modified Converse medium[176] containing 99% ^{15}N-labeled Ammonium Acetate (Sigma-Aldrich, St Louis, MO), at the following concentrations: 2.5×10^9 CFU for 48-hr spherules, 7.0×10^8 CFU for 96-hr spherules and 120-hr spherules. Samples were incubated at 38°C in 20% CO_2 while shaking at 160 rpm for the appropriate time length. All manipulations of potentially viable cells were conducted in biosafety level 3 (BSL-3) conditions utilizing approved standard operating procedures in laboratories registered with the Centers for Disease Control for select agent possession.

4.2.2 Strains and growth conditions-Mycelia

Arthroconidia (as described above) were inoculated into 1L of liquid 2X GYE broth at a concentration of 1.0×10^8 CFU, at 37°C while shaking at 180 rpm in room air for 48 hours.

4.2.3 Cell disruption and cytoplasmic protein extraction

Spherules were harvested by centrifugation at 5100 rpm, 4°C for 30 min. and washed with sterile water. Mycelia cells were collected by filtering through a sterile paper filter followed by drying and physical scraping to remove cells. Pelleted spherules and collected mycelia cells were resuspended in equal volume cold lysis buffer (20mM

Tris-HCl pH 7.9, 10mM $MgCl_2$, 1mM DTT, 200mM ammonium sulfate, 1mM PMSF, 5% v/v glycerol and 1x protease inhibitor cocktail (Calbiochem, San Diego, CA)) and an equal volume glass beads. Cells were then vortexed for 60 sec and placed on ice for 60 sec, alternately, 8 times. Disrupted cells were again centrifuged as above. The supernatant was removed and proteins were precipitated in freshly prepared 10% cold Trichloroacetic Acid (Sigma-Aldrich, St Louis, MO) and allowed to sit on ice for 45 minutes. Precipitated proteins were centrifuged and washed with cold acetone prior to resuspension in 70% ethanol for 30 min to ensure complete sample sterilization per BSL-3 standard operating procedure prior to removal from the laboratory. After sterilization, sample was again centrifuged, resuspended in 1mL water and stored at -20°C.

4.2.4 Protein quantification and digestion

Precipitated protein samples were quantified by Bicinchoninic Acid (BCA) assay (Pierce, Rockford, IL) against a bovine serum albumin (BSA) standard after suspension in a BCA-compatible solubilization buffer (25mM Tris base, 192mM glycine, 0.1% SDS, pH=8.2). Mycelia and spherule extracts were combined in a 1:1 ratio based on total protein concentration for a total concentration of 150 µg protein (75 µg each sample). Three samples were prepared for analysis: Unlabeled (^{14}N) Mycelia with labeled (^{15}N) 48 hour spherules, 96 hour spherules or 120 hour spherules. The combined protein samples were subjected to disulfide bond reduction in 10mM Dithiothreitol (Fluka, Sigma-Aldrich, St Louis, MO) in 100 µL of 100mM ammonium bicarbonate (AmBic) at 50°C for 1 hour, followed by cysteine carbamidomethylation with 55mM iodoacetamide

(Sigma-Aldrich, St Louis, MO) in 100 µL of 100mM AmBic at room temp for 45 minutes in the dark. Samples were then digested by incubation in 2µg proteomics-grade trypsin (Sigma-Aldrich, St Louis, MO) suspended in 300µL 100mM AmBic at 37°C for 3 hrs.

4.2.5 Sample preparation for MS/MS analysis

Trypsin-digested peptides were purified and concentrated with a C-18 solid-phase extraction cartridge (3M, St Paul, MN). All samples were dried down to minimal volume (20 µL) and stored at -20°C until MS/MS analysis.

4.2.6 HPLC

Peptide samples were analyzed by 2-D LC MS/MS (MuDPIT)[88] with HPLC elution directed into an ESI source of a Thermo-Finnigan LTQ linear ion-trap mass spectrometer (Thermo Scientific, San Jose, CA). Peptides from were loaded on a dual-phase column consisting of 5 cm of 5µm polysulfoethyl-A strong cation exchange (SCX) resin (PolyLC Inc., Columbia, MD) upstream of 7cm of 5µm Zorbax Eclipse XDB C-18 resin (Agilent Technologies, Palo Alto, CA, USA) packed into a 100 µm I.D. fused silica capillary pulled to a 5 µm tip using a laser puller (Sutter Instrument, Novato, CA). Electrospray voltage of 1.6-2.1 kV was applied using a gold or platinum electrode attached to a liquid junction upstream of the packing material. Peptides were eluted using a 4-buffer system consisting of water with 0.1% formic acid (Buffer A), acetonitrile with 0.1% formic acid (Buffer B), 250mM ammonium sulfate (Buffer C), and 1.5 M

ammonium sulfate (Buffer D). Samples were injected onto the column by direct bomb-loading of the capillary using 500 psi UHP helium gas or by injection using a Surveyor autosampler (Thermo Scientific) with Buffer A to deposit the sample on the SCX phase of the column. Peptides that did not deposit on the SCX were eluted off the RP material by a 5-50% gradient of Buffer B over 90 minutes followed by a 50-98% gradient over 5 min and a 5 min wash of 95% Buffer B (5% Buffer A), followed by a 20-min re-equilibration using 95% Buffer A (5% Buffer B). Peptides that were deposited on the SCX were eluted in a series of 11 salt steps (from 10-100% Buffer C with a final step of 50% Buffer D) consisting of a 5 min pulse of the salt followed by a 7 min wash of 95% Buffer A (5% Buffer B) prior to a 60 min gradient from 5-50% Buffer B, followed by 50-98% Buffer B over 5 min and a 5 min wash of 95% Buffer B. After the Buffer B gradient, the column is re-equilibrated with a wash of 95% Buffer A for 20 min to prepare the C-18 material for the next salt elution step. Flow rates of 600 nL per min were calibrated prior to each run. An overview of the MudPIT strategy is described in Section 1.2.3.1 of this dissertation, and the experimental details are described in Section 3.2.6.2.

4.2.7 MS/MS analysis

Peptides introduced into the mass spectrometer were scanned over the m/z range from 400 to 2000. The seven most abundant peaks were selected for fragmentation in MS/MS using automatic peak recognition and a 30-second dynamic exclusion window after a maximum of 5 selections of the same parent ion. Parent ions were fragmented by

RF excitation leading to collision induced dissociation with helium background gas at approximately 0.6 to 0.8 × 10^{-5} torr pressure in the ion trap. Data were continually collected by Xcalibur instrument software version 1.4 SR1 (Thermo Scientific).

4.2.8 Bioinformatics

MS/MS data produced as described were analyzed using the SEQUEST database search algorithm[82, 178] against a FASTA database consisting of common contaminants (trypsin, human keratin, protein standards for MS calibration, etc) followed by the *C. posadasii* and *C. immitis* sequences with protein sequences from 18 additional fungi as shown in Table 3.1. Xcalibur .raw files were searched using TurboSEQUEST (BioWorks v 3.1) on a 16-processor IBM Beowulf cluster. DTA files were generated by SEQUEST according to the following criteria (for explanation of these criteria see Section 3.2.7): Peptide MW Range = 400-3500 Da; Threshold = 100; Precursor Mass = 1.50; Group Scan = 42; Minimum Group Count = 2; and Minimum Ion Count = 10. SEQUEST searches were performed with no enzyme specified utilizing the default search parameters except: peptide mass tolerance = 1.5 Da; max number of modified amino acids per differential modification in a peptide = 4; static modification of +57.0 Da for carbamidomethylated cysteine; a differential residue modification of +16.0 Da for oxidized methionine; maximum of 2 internal cleavage sites; one allowed error in matching auto-detected peaks; and a mass tolerance of +/- 1.0 Da for matching auto detected peaks. Each set of .raw files was searched with SEQUEST twice using a standard sequest.params file and again with a sequest.params file altered to search for

^{15}N-labeled peptides using static mass additions corresponding to the correct number of nitrogen atoms in each amino acid.

Search results from SEQUEST were filtered using DTASelect and Contrast (v 1.9)[179] with the default cutoff parameters (+1 ≥ 1.8, +2 ≥ 2.5, +3 ≥ 3.5, ΔCn ≥ 0.08), specification of at least half-tryptic peptides, minimum of two peptide identification per protein and automated removal of all common contaminant identifications.

4.2.9 Differential quantification

The relative protein expression levels between the ^{14}N mycelia and the ^{15}N spherules were calculated by the RelEx algorithm.[189] Use of RelEx provides an automated method of calculating peptide ratios from stable isotope proteomic experiments by calculating the ion current ratios from the ion chromatogram. The algorithm was used with the default parameters, except for the following. The EXTRACT-CHRO utility was run with a 99.0% ^{15}N atomic enrichment setting (based on the ^{15}N percentage of the labeled ammonium acetate used for cell media). After extraction of the ion chromatograms, a minimum signal-to-noise cutoff of 5, and regression filter cutoffs of 0.5 were used to filter the data as per Kolkman *et al.*[190] After peptide ratio calculation by RelEx on each of the three repeated experiments per spherule time point, the protein identifications were combined and filtered by a minimum of two peptides identified per protein, and coefficient of variation (the average peptide ratio divided by the standard deviation) of less than 25%. In addition to these criteria, all quantified proteins had to be identified in all three experiments to be included. Protein

identifications found in all three of the ^{15}N SEQUEST searches per time point, but not found in any ^{14}N SEQUEST searches were included as proteins exclusively produced in spherules.

4.3 Results

In an effort to identify new protein antigen targets against coccidioidomycosis, we utilized a method of stable isotope labeling to identify those proteins that are more highly expressed in the pathogenic spherule phase of *C. posadasii* than in the saprobic mycelial phase. To do this, we cultured mycelia using typical methods,[71] using standard ^{14}N-labeled media, and cultured spherule cells in modified Converse medium[176] using 99% ^{15}N ammonium acetate as the sole nitrogen source. Three time points of spherule cells were collected (48, 96 and 120 hours post-inoculation), along with mycelia cells cultured for 48 hours. The cell were disrupted, and the cytoplasmic proteins from both cell types were precipitated and combined in a 1:1 ratio based on total protein concentration. These combined protein samples were then analyzed by MudPIT tandem mass spectrometry for protein identification and quantification.

4.3.1 Protein quantification

We utilized the quantitative analysis algorithm RelEx[189] to analyze our MudPIT data for the determination of relative protein quantities between mycelia and our three spherule time points. This algorithm functions by determining the liquid chromatography elution profile of peptides identified by the search algorithm SEQUEST from the MS/MS data. Figure 4.2 illustrates a typical ion chromatogram from a MudPIT analysis,

including an MS spectrum with two parent ions that were identified as a ^{14}N and ^{15}N-labeled peptide pair from the antigen target protein PepC identified in the spherule cell wall proteome analysis described in Chapter 3 of this dissertation (See section 3.3.4.1). The parent ions 645.49 m/z and 653.45 m/z were subsequently fragmented and analyzed in a second round of MS as shown in Figure 4.3. The RelEx program uses the SEQUEST output file that identified the 645.49 ion as a PepC peptide to calculate the mass of the corresponding ^{15}N labeled peptide. RelEx then searches the ion chromatogram for both parent ions that elute within a time period corresponding to 50 MS scans above and below the MS/MS spectrum used by SEQUEST to identify the ^{14}N peptide. The relative protein concentration ratio is subsequently calculated as the total ion intensities (over the 100 scan window) for the unlabeled ^{14}N peptide divided by the total ion intensities for the ^{15}N-labeled version of the peptide. As can be seen in Figure 4.3, both peptides exhibit essentially identical fragmentation spectra, with the exception of the m/z ratios of the peaks corresponding to the ^{15}N incorporation.

4.3.2 SEQUEST analysis of ^{15}N-labeled peptides

Incorporation of ^{15}N in peptides identified by mass spectrometry results in an inflation in ambiguous identifications due to the ^{15}N-induced increase in nearly isobaric amino acids.[121] In normal ^{14}N-labeled proteins, the amino acid pairs of leucine/isoleucine and glutamine/lysine are isobaric in all but the highest resolution mass spectrometers. When ^{15}N labeling of all amino acids is performed, the additional amino acid pairs of

Figure 4.2 Ion chromatogram and MS spectrum of ^{14}N-^{15}N peptide pair

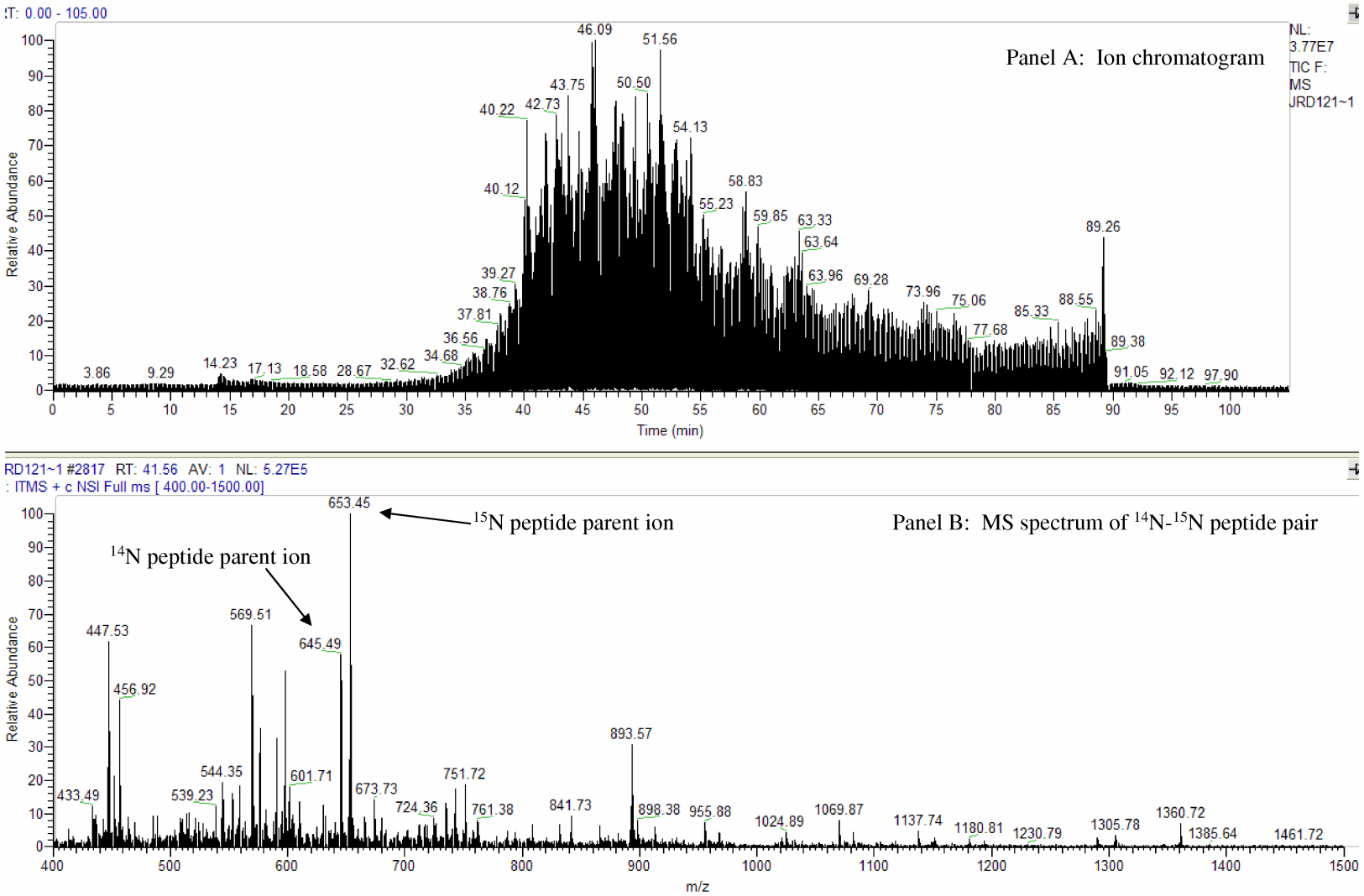
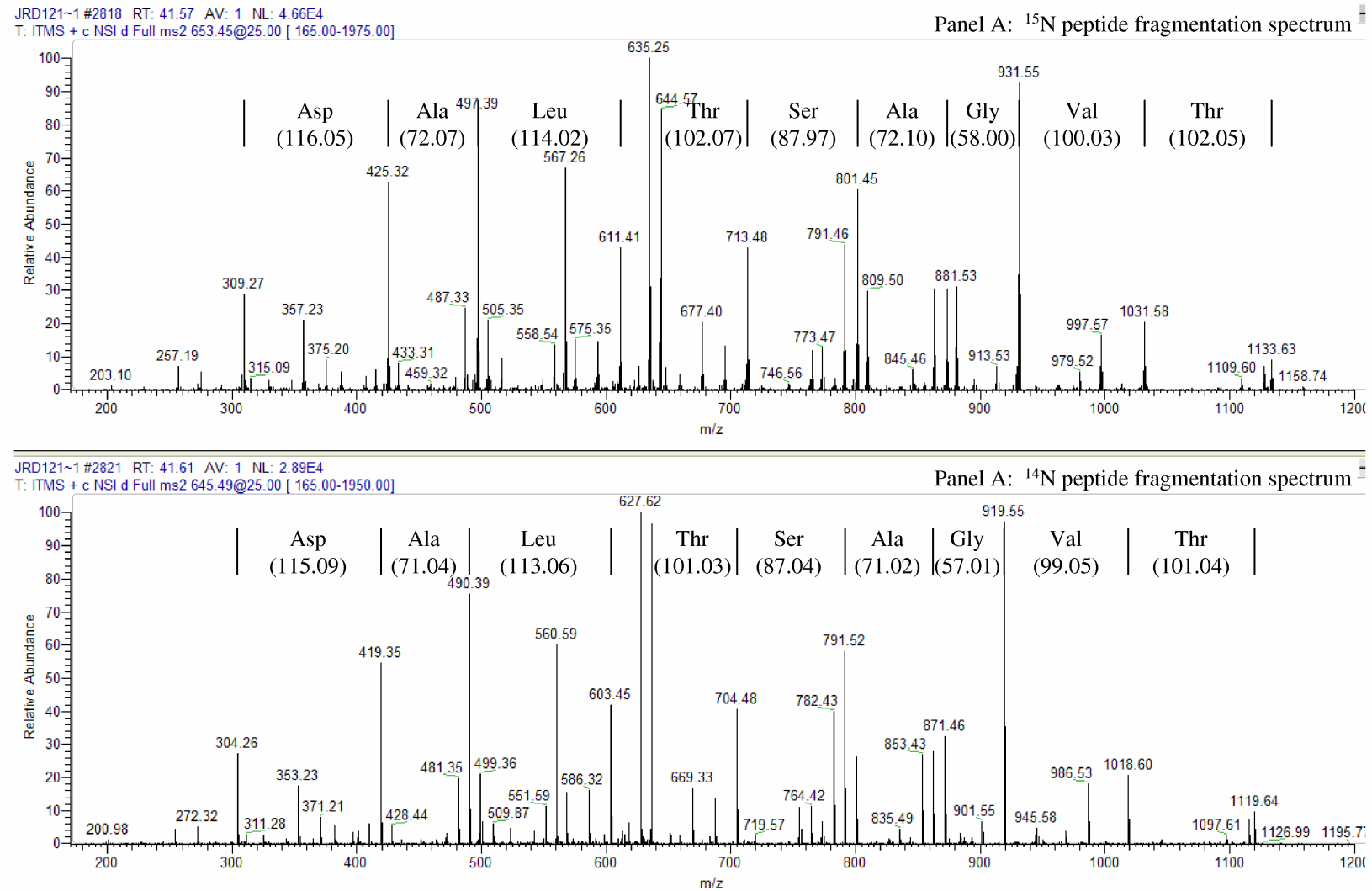


Figure 4.3 MS/MS fragmentation spectra of ^{14}N-^{15}N peptide pair AVTVGASTLADER



asparagine/aspartic acid and glutamine/glutamic acid become isobaric in addition to the previously mentioned pairs. This leads to ambiguity in peptide sequence identification via mass spectrometry, and an overall decrease in the number of proteins identified by common protein sequence database search algorithms. For example, the search algorithm SEQUEST calculates a cross correlation (XCorr) score for each peptide identified as a measure of sequence identification confidence. When the XCorr of the two highest scoring peptides for a spectrum are closer than a pre-determined amount (known as the ΔCn cutoff), the peptide assignment for that spectrum is not reported.

An example of this phenomenon of decreased protein identifications in ^{15}N-labeled samples is illustrated in Table 4.1. Here we show the number of proteins identified by SEQUEST in three experimental samples consisting of mycelial and 96-hour spherule proteins. In experiment number three, the total number of ^{15}N-labeled proteins identified with either a minimum of one or two peptides per protein is significantly lower than the total number of ^{14}N proteins identified from the same experiment. When we consider that both the mycelia and spherule samples consist of soluble cytoplasmic proteins, we would expect the total number of proteins identified to be roughly the same.

Using a method described by Nelson *et al,*[121] we have increased the number of total proteins identified in ^{15}N searches by lowering the ΔCn cutoff score from 0.08 to 0.04, without increasing the protein false discovery rate (FDR) above 1%. While this change only allowed an increase from 142 to 148 proteins identified by two peptides or more, further reduction in the ΔCn cutoff was not undertaken for fear of passing a 1%

Table 4.1 Changes of SEQUEST parameters to increase protein identification of ^{15}N labeled proteins from 96 hour spherules

Database search [a]	Minumum # peptides per protein	ΔCn cutoff	Experiment 1		Experiment 2		Experiment 3		Average	
			Number of Proteins Identified	FDR [b]	Number of Proteins Identified	FDR [b]	Number of Proteins Identified	FDR [b]	Number of Proteins Identified [c]	FDR [b]
^{14}N Forward	1	0.08	458	0.260	384	0.240	554	0.267	465 (85.2)	0.257
^{14}N Reverse	1	0.08	119		92		148		120 (28.0)	
^{14}N Forward	2	0.08	185	0.005	132	0.000	237	0.000	185 (52.5)	0.002
^{14}N Reverse	2	0.08	1		0		0		0.33 (.57)	
^{15}N Forward	1	0.08	415	0.301	336	0.345	393	0.321	381 (40.8)	0.321
^{15}N Reverse	1	0.08	125		116		126		122 (5.5)	
^{15}N Forward	2	0.08	144	0.007	120	0.000	161	0.000	142 (20.6)	0.002
^{15}N Reverse	2	0.08	1		0		0		0.33 (.57)	
^{15}N Forward	1	0.04	505	0.556	394	0.543	475	0.535	458 (57.4)	0.545
^{15}N Reverse	1	0.04	281		214		254		250 (33.7)	
^{15}N Forward	2	0.04	151	0.013	127	0.008	167	0.000	148 (20.1)	0.007
^{15}N Reverse	2	0.04	2		1		0		1 (1)	

a) SEQUEST search results filtered by DTASelect default cutoff parameters as described in materials and methods utilizing ^{14}N or ^{15}N labeled amino acids as indicated and searched against the forward or reversed protein sequence database

b) False Discovery Rate calculated as number of proteins identified against reversed database divided by number of proteins identified against forward database

c) Average number of proteins identified over the three experiments with standard deviation in parenthesis

FDR. An analogous strategy was used for the data analysis of the mycelia-48 hour spherule samples, resulting in a ΔCn cutoff decrease from 0.08 to 0.02 (data not shown). A corresponding change to ΔCn cutoff scores in mycelia-120 hour spherule samples was not done, due to ^{15}N protein identification levels that were already comparable to ^{14}N identifications.

4.3.3 Protein identifications in ^{15}N labeled spherule samples

Using the quantification algorithm RelEx, we have identified a substantial number of proteins produced in both mycelial and spherule cells. Table 4.2 lists all proteins identified by RelEx from MudPIT data sorted by spherule time point and $^{14}N/^{15}N$ ratio (mycelia/spherule ratio). These data were generated as described, and the proteins listed consist of only those proteins that were found in all three experiments for each time point, with a minimum of two peptides per protein identified. The $^{14}N/^{15}N$ ratio that is reported was calculated as the average ratio of all identified peptides between the three experiments. In addition, the standard deviation is reported as well as the coefficient of variation (CV) which is the standard deviation expressed as a percent of the average ratio (calculated as: SD/ratio * 100). A total of 73 proteins were identified from mycelia with 48 hour spherules, 59 from mycelia-96 hour spherules, and 56 from mycelia-120sph.

The total list of proteins identified was further reduced by limiting the possible target candidates to only those proteins that were expressed in higher abundance in spherules than the mycelia. In addition, any protein identified with a Mycelia/spherule ratio coefficient of variation (CV) greater than 25% was excluded, from a method

Table 4.2 Proteins identified in differential proteomic analysis of mycelia cells with 48hr, 96hr and 120hr spherules

Mycelia (^{14}N) and 48 hour spherule (^{15}N) proteins								
Description	**TIGR locus**	**Broad locus**	**Found in Cell Wall Analysis**	**Protein Category**	**Human Homology[a]**	**^{14}N/^{15}N ratio**	**Standard Deviation**	**CV[b]**
Malate dehydrogenase	72.m01909	CIMG_05466	Y	1	High	0.30	0.02	5.3%
CPY20 protein	73.m03535	CIMG_04756	Y	5	Low	0.33	0.02	5.1%
Aconitase	67.m08291	CIMG_01729	Y	1	High	0.36	0.03	8.9%
Cytochrome C peroxidase	51.m00579	CIMG_08209	Y	1	Low	0.37	0.05	12.5%
2-Methyl citrate synthase	13.m01800	CIMG_10114	Y	1	High	0.42	0.01	3.6%
Isocitrate lyase	13.m01812	CIMG_10137	**N**	1	Low	0.51	0.06	11.9%
Phosphoenol pyruvate carboxykinase	67.m08638	CIMG_01264	Y	1	Low	0.56	0.01	2.2%
2-Methyl citrate dehydratase	13.m01811	CIMG_10136	Y	1	Low	0.58	0.07	11.9%
Ketol-acid reductoisomerase	72.m02091	CIMG_05641	Y	3	Low	0.61	0.06	9.9%
Enolase	10.m00701	CIMG_07322	Y	1	High	0.69	0.06	9.1%
Transaldolase	52.m06868	CIMG_06675	Y	1	High	0.72	0.03	4.5%
Aldolase	12.m07770	CIMG_03654	Y	1	Low	0.87	0.18	20.7%
Triosephosphate isomerase	14.m03111	CIMG_09361	Y	1	High	0.89	0.05	5.8%
Aminotransferase	52.m07600	CIMG_07122	Y	3	Low	0.91	0.05	5.9%
Alcohol dehydrogenase	73.m03409	CIMG_04945	Y	1	Low	0.92	0.04	4.0%
Hsp70	73.m03913	CIMG_04436	Y	5	High	0.95	0.08	8.7%
Inorganic pyrophosphatase	65.m01908	CIMG_07626	Y	1	High	0.96	0.11	11.3%
Aspartate aminotransferase	67.m08523	CIMG_01452	Y	1	High	0.96	0.05	5.5%
Alphaketoglutarate dehydrogenase	67.m08391	CIMG_01597	Y	1	Moderate	1.00	0.44	43.6%
Malate dehydrogenase	60.m01383	CIMG_02580	Y	1	High	1.05	0.17	16.3%
Probable fumarate reductase	68.m01964	CIMG_08720	Y	1	Moderate	1.10	0.28	25.8%
Homocysteine methyltransferase	12.m07493	CIMG_04062	Y	3	Low	1.12	0.13	11.3%
Hsp70	61.m01694	CIMG_02494	Y	5	High	1.15	0.04	3.7%
Peptidyl-prolyl cis-trans isomerase	73.m03758	CIMG_04486	Y	4	High	1.17	0.05	4.2%
Glucose-6-phosphate isomerase	65.m01749	CIMG_07844	Y	1	High	1.28	0.15	11.4%
Fumarate hydratase	52.m06510	CIMG_05872	**N**	1	High	1.30	0.38	29.3%
ATP synthase beta chain	52.m06668	CIMG_06274	Y	1	High	1.32	0.17	12.7%
ATP synthase alpha chain	73.m03967	CIMG_04309	Y	1	High	1.33	0.13	9.8%
Adenosylhomocysteinase	13.m01984	CIMG_10311	Y	3	High	1.35	0.17	12.8%
Pyruvate dehydrogenase E1	61.m01655	CIMG_02447	Y	1	High	1.50	0.85	56.6%
Glyceraldehyde phosphate dehydrogenase	52.m06707	CIMG_06404	Y	1	High	1.50	0.08	5.6%
Methylmalonate semialdehyde dehydrogenase	67.m09155	CIMG_00614	**N**	1	High	1.54	0.38	24.6%
Acetyl-CoA acetyltransferase	61.m01556	CIMG_02262	Y	1	Moderate	1.54	0.23	14.7%
Zinc-binding dehydrogenase	12.m08151	CIMG_03135	Y	6	Moderate	1.61	0.65	40.4%
Phosphogluconate dehydrogenase	67.m08761	CIMG_01153	Y	3	High	1.61	0.08	5.2%
Delta-1-pyrroline-5-carboxylate dehydrogenase	52.m07026	CIMG_06927	Y	3	High	1.81	0.31	17.0%
3-Hydroxyisobutyrate dehydrogenase	67.m09538	CIMG_00035	**N**	1	Moderate	2.00	0.61	30.6%
Pyridoxine biosynthesis protein	52.m06627	CIMG_06181	Y	3	Low	2.00	0.62	30.8%
Peptidyl-prolyl cis-trans isomerase	72.m02028	CIMG_05579	Y	4	Moderate	2.08	0.51	24.3%
Hsp88	52.m06977	CIMG_06861	**N**	5	Moderate	2.24	0.62	27.7%
Ribosomal L22	52.m06938	CIMG_09618	Y	4	High	2.29	1.46	63.8%
Ribosomal S5	12.m08026	CIMG_03305	Y	4	High	2.46	0.16	6.3%
Ribosomal L10	67.m08809	CIMG_01033	Y	4	High	2.61	0.15	5.9%
Ribosomal L1	52.m07559	CIMG_07001	Y	4	High	2.65	0.10	3.9%
Ribosomal S0	12.m07883	CIMG_03501	Y	4	High	2.66	0.69	26.0%

a) Protein sequence identity to human proteins (Low = <30%, Moderate = 30%-50%, High = >50%)

b) Coefficient of Variation; calculated as: standard deviation/ratio * 100

Table 4.2 Proteins identified in differential proteomic analysis of mycelia cells with 48hr, 96hr and 120hr spherules (continued)

Mycelia (^{14}N) and 48 hour spherule (^{15}N) proteins (continued)

Description	TIGR locus	Broad locus	Found in Cell Wall Analysis	Protein Category	Human Homology[a]	$^{14}N/^{15}N$ ratio	Standard Deviation	CV[b]
Glutamate carboxypeptidase-like protein	67.m08009	CIMG_02104	N	4	High	2.71	0.69	25.3%
Cu,Zn superoxide dismutase	52.m06870	CIMG_06677	N	5	High	2.72	0.70	25.8%
Ribosomal S18	60.m01388	CIMG_02814	Y	4	High	2.74	0.91	33.1%
Ribosomal L17	12.m08103	CIMG_03194	Y	4	High	2.81	0.53	19.0%
Ribosomal S2	45.m00927	CIMG_08094	Y	4	High	3.01	0.76	25.2%
Ribosomal L4	52.m07369	CIMG_06503	Y	4	High	3.02	0.50	16.5%
Ribosomal S3	12.m07592	CIMG_03903	Y	4	High	3.16	0.45	14.2%
Ribosomal S7	12.m07699	CIMG_03754	Y	4	High	3.32	0.78	23.5%
Ribosomal L7	65.m01736	CIMG_07888	Y	4	High	3.55	0.82	23.0%
Aldehyde dehydrogenase	52.m06796	CIMG_09805	Y	1	High	3.56	0.71	20.1%
Ribosomal S11	45.m00895	CIMG_08046	Y	4	High	3.57	1.33	37.2%
Arginosuccinate synthase	12.m07947	CIMG_03406	Y	3	High	3.60	1.20	33.3%
Elongation factor 2	73.m03507	CIMG_05034	Y	4	High	3.74	0.46	12.2%
Hsp60	52.m06672	CIMG_06278	Y	5	High	3.77	0.68	17.9%
Ribosomal S4	67.m09348	CIMG_00391	Y	4	High	3.91	1.81	46.3%
Ribosomal L8	67.m08322	CIMG_01685	Y	4	High	3.91	1.24	31.8%
Aminopeptidase	52.m06685	CIMG_06320	N	4	Moderate	4.00	0.42	10.5%
Ribosomal L6	52.m07545	CIMG_06963	Y	4	Moderate	4.11	0.55	13.4%
Ribosomal L14	13.m01821	CIMG_10151	Y	4	Moderate	4.23	0.78	18.5%
Ribosomal L13	68.m01881	CIMG_08619	Y	4	Moderate	4.26	0.38	9.0%
Ribosomal L15	73.m03579	CIMG_04962	Y	4	High	4.33	0.50	11.5%
Ribosomal S22	12.m08165	CIMG_03113	Y	4	High	4.58	1.55	33.9%
Elongation factor 1 beta	52.m07063	CIMG_06970	Y	4	High	4.82	0.49	10.1%
Protein disulfide isomerase	52.m07651	CIMG_07225	Y	4	Moderate	4.97	1.15	23.1%
Hsp90	73.m03734	CIMG_04729	Y	5	High	5.27	0.81	15.5%
Peroxisomal membrane protein	52.m07000	CIMG_05828	Y	5	Moderate	6.02	1.44	23.9%
Endoribonuclease	65.m01864	CIMG_07737	N	3	Moderate	6.56	1.33	20.2%
Elongation factor 1 alpha	12.m07727	CIMG_03708	Y	4	High	13.63	1.72	12.7%

Mycelia (^{14}N) and 96 hour spherule (^{15}N) proteins

Description	TIGR locus	Broad locus	Found in Cell Wall Analysis	Protein Category	Human Homology[a]	$^{14}N/^{15}N$ ratio	Standard Deviation	CV[b]
Malate dehydrogenase	72.m01909	CIMG_05466	Y	1	High	0.26	0.03	10.1%
Cu,Zn superoxide dismutase	52.m06870	CIMG_06677	N	5	High	0.32	0.06	19.0%
Ketol-acid reductoisomerase	72.m02091	CIMG_05641	Y	3	Low	0.35	0.04	10.6%
Enolase	10.m00701	CIMG_07322	Y	1	High	0.37	0.04	10.7%
Alcohol dehydrogenase	73.m03409	CIMG_04945	Y	1	Low	0.45	0.10	23.0%
2-Methyl citrate dehydratase	13.m01811	CIMG_10136	Y	1	Low	0.48	0.03	6.5%
Peptidyl-prolyl cis-trans isomerase	73.m03758	CIMG_04486	Y	4	High	0.50	0.06	12.2%
Hsp10	52.m06460	CIMG_09671	Y	5	Moderate	0.51	0.08	16.6%
Aldolase	12.m07770	CIMG_03654	Y	1	Low	0.51	0.05	10.1%
Hsp70	73.m03913	CIMG_04436	Y	5	High	0.58	0.08	13.7%
ATP synthase alpha chain	73.m03967	CIMG_04309	Y	1	High	0.58	0.17	28.9%
Glyceraldehyde phosphate dehydrogenase	52.m06707	CIMG_06404	Y	1	High	0.61	0.09	14.8%
ATP synthase beta chain	52.m06668	CIMG_06274	Y	1	High	0.64	0.09	13.5%
Transaldolase	52.m06868	CIMG_06675	Y	1	High	0.67	0.09	14.0%
Malate dehydrogenase	60.m01383	CIMG_02580	Y	1	High	0.72	0.09	12.9%
Subtilisin-like protease	52.m06866	CIMG_06672	Y	4	Low	0.72	0.09	13.1%
Homocysteine methyltransferase	12.m07493	CIMG_04062	Y	3	Low	0.79	0.52	65.7%
Inorganic pyrophosphatase	65.m01908	CIMG_07626	Y	1	High	0.80	0.10	12.3%
Hsp70	61.m01694	CIMG_02494	Y	5	High	0.80	0.09	11.5%
Cytochrome C	73.m03439	CIMG_05096	Y	1	High	0.83	0.11	13.0%
Peroxisomal membrane protein	52.m07000	CIMG_05828	Y	5	Moderate	0.87	0.06	7.4%

a) Protein sequence identity to human proteins (Low = <30%, Moderate = 30%-50%, High = >50%)

b) Coefficient of Variation; calculated as: standard deviation/ratio * 100

Table 4.2 Proteins identified in differential proteomic analysis of mycelia cells with 48hr, 96hr and 120hr spherules (continued)

Mycelia (^{14}N) and 96 hour spherule (^{15}N) proteins (continued)

Description	TIGR locus	Broad locus	Found in Cell Wall Analysis	Protein Category	Human Homology[a]	$^{14}N/^{15}N$ ratio	Standard Deviation	CV[b]
Phosphoenol pyruvate carboxykinase	67.m08638	CIMG_01264	Y	1	Low	0.89	0.21	23.8%
Delta-1-pyrroline-5-carboxylate dehydrogenase	52.m07026	CIMG_06927	Y	3	High	1.00	0.20	20.0%
BipA chaperone protein	67.m08661	CIMG_01229	Y	4	High	1.20	0.25	21.2%
Ribosomal L19	67.m08947	CIMG_00911	Y	4	Moderate	1.20	0.41	34.4%
Hsp60	52.m06672	CIMG_06278	Y	5	High	1.23	0.18	14.8%
Ribosomal L1	52.m07559	CIMG_07001	Y	4	High	1.24	0.08	6.5%
Ribosomal L22	52.m06938	CIMG_09618	Y	4	High	1.25	0.60	48.3%
Acetyl-CoA acetyltransferase	61.m01556	CIMG_02262	Y	1	Moderate	1.27	0.17	13.0%
Ribosomal L13	68.m01881	CIMG_08619	Y	4	Moderate	1.31	0.22	16.7%
Ribosomal L8	67.m08322	CIMG_01685	Y	4	High	1.39	0.03	2.3%
Ribosomal S14	73.m03849	CIMG_04348	Y	4	High	1.41	0.67	47.8%
Ribosomal S6	67.m08456	CIMG_01482	Y	4	High	1.43	0.11	7.9%
Ribosomal L12	73.m03510	CIMG_04811	Y	4	High	1.54	0.25	16.2%
Endoribonuclease	65.m01864	CIMG_07737	N	3	Moderate	1.59	0.38	24.0%
Translationally-controlled tumor protein	9.m00307	CIMG_02984	Y	6	Moderate	1.59	1.12	70.7%
Ribosomal S7	12.m07699	CIMG_03754	Y	4	High	1.61	0.31	19.1%
Ribosomal L26	52.m06832	CIMG_09792	Y	4	High	1.64	0.18	10.9%
Ribosomal L18	12.m08248	CIMG_04241	Y	4	High	1.66	0.79	47.5%
Ribosomal L18	73.m03592	CIMG_04931	Y	4	High	1.71	0.27	15.7%
Ribosomal L6	52.m07545	CIMG_06963	Y	4	Moderate	1.73	0.26	15.1%
Ribosomal S5	12.m08026	CIMG_03305	Y	4	High	1.81	0.20	10.9%
Ribosomal L7	65.m01736	CIMG_07888	Y	4	High	1.82	0.53	29.1%
Ribosomal L2	52.m06573	CIMG_06034	Y	4	High	1.94	0.22	11.2%
Dihydrolipoamide dehydrogenase	72.m02021	unknown	Y	1	High	1.95	0.10	5.2%
Ribosomal L4	52.m07369	CIMG_06503	Y	4	High	1.97	0.40	20.1%
Ribosomal S2	45.m00927	CIMG_08094	Y	4	High	1.99	0.88	44.1%
Ribosomal S11	45.m00895	CIMG_08046	Y	4	High	2.07	0.54	25.9%
Ribosomal L17	12.m08103	CIMG_03194	Y	4	High	2.21	0.35	15.8%
Protein disulfide isomerase	52.m07651	CIMG_07225	Y	4	Moderate	2.71	0.44	16.3%
Aldehyde dehydrogenase	52.m06796	CIMG_09805	Y	1	High	2.78	0.67	24.3%
Ribosomal S22	12.m08165	CIMG_03113	Y	4	High	2.95	1.04	35.3%
Ribosomal S18	60.m01388	CIMG_02814	Y	4	High	3.12	1.36	43.6%
Glycine rich protein	52.m07155	CIMG_06083	Y	6	Moderate	3.17	0.90	28.4%
Arginosuccinate synthase	12.m07947	CIMG_03406	Y	3	High	3.36	0.84	25.1%
Elongation factor 1 alpha	12.m07727	CIMG_03708	Y	4	High	4.13	0.65	15.7%
Aminopeptidase	52.m06685	CIMG_06320	N	4	Moderate	5.00	1.68	33.6%
Elongation factor 2	73.m03507	CIMG_05034	Y	4	High	5.20	1.89	36.3%
Hsp90	73.m03734	CIMG_04729	Y	5	High	5.68	2.12	37.3%

Mycelia (^{14}N) and 120 hour spherule (^{15}N) proteins

Description	TIGR locus	Broad locus	Found in Cell Wall Analysis	Protein Category	Human Homology[a]	$^{14}N/^{15}N$ ratio	Standard Deviation	CV[b]
Cu,Zn superoxide dismutase	52.m06870	CIMG_06677	N	5	High	0.13	0.01	10.8%
Malate dehydrogenase	72.m01909	CIMG_05466	Y	1	High	0.23	0.02	8.9%
Alcohol dehydrogenase	73.m03409	CIMG_04945	Y	1	Low	0.25	0.04	17.8%
Enolase	10.m00701	CIMG_07322	Y	1	High	0.26	0.02	6.4%
Cytochrome C peroxidase	51.m00579	CIMG_08209	Y	1	Low	0.30	0.07	23.6%
Peptidyl-prolyl cis-trans isomerase	73.m03758	CIMG_04486	Y	4	High	0.38	0.08	21.1%
Glyceraldehyde phosphate dehydrogenase	52.m06707	CIMG_06404	Y	1	High	0.39	0.03	8.2%
Hsp70	73.m03913	CIMG_04436	Y	5	High	0.41	0.07	17.1%

a) Protein sequence identity to human proteins (Low = <30%, Moderate = 30%-50%, High = >50%)

b) Coefficient of Variation; calculated as: standard deviation/ratio * 100

Table 4.2 Proteins identified in differential proteomic analysis of mycelia cells with 48hr, 96hr and 120hr spherules (continued)

Mycelia (^{14}N) and 120 hour spherule (^{15}N) proteins (continued)								
Description	**TIGR locus**	**Broad locus**	**Found in Cell Wall Analysis**	**Protein Category**	**Human Homology[a]**	**$^{14}N/^{15}N$ ratio**	**Standard Deviation**	**CV[b]**
2-Methyl citrate dehydratase	13.m01811	CIMG_10136	Y	1	Low	0.45	0.04	9.1%
Hsp70	61.m01694	CIMG_02494	Y	5	High	0.45	0.07	15.5%
ATP synthase beta chain	52.m06668	CIMG_06274	Y	1	High	0.49	0.06	11.5%
ATP synthase alpha chain	73.m03967	CIMG_04309	Y	1	High	0.51	0.12	23.0%
Triosephosphate isomerase	14.m03111	CIMG_09361	Y	1	High	0.54	0.10	18.9%
Malate dehydrogenase	60.m01383	CIMG_02580	Y	1	High	0.55	0.07	13.7%
Transaldolase	52.m06868	CIMG_06675	Y	1	High	0.56	0.10	17.1%
Peroxisomal membrane protein	52.m07000	CIMG_05828	Y	5	Moderate	0.59	0.08	13.4%
Phosphoenol pyruvate carboxykinase	67.m08638	CIMG_01264	Y	1	Low	0.63	0.07	11.5%
Homocysteine methyltransferase	12.m07493	CIMG_04062	Y	3	Low	0.65	0.07	10.5%
Subtilisin-like protease	52.m06866	CIMG_06672	Y	4	Low	0.65	0.14	22.0%
Inorganic pyrophosphatase	65.m01908	CIMG_07626	Y	1	High	0.71	0.06	8.7%
Delta-1-pyrroline-5-carboxylate dehydrogenase	52.m07026	CIMG_06927	Y	3	High	0.73	0.14	19.3%
Ribosomal L31	67.m08346	CIMG_01655	Y	4	High	0.94	0.06	6.8%
Ribosomal L26	52.m06832	CIMG_09792	Y	4	High	0.95	0.07	7.7%
Ribosomal L13	68.m01881	CIMG_08619	Y	4	Moderate	0.95	0.09	9.6%
Ribosomal L28	60.m01381	CIMG_02581	Y	4	Moderate	0.97	0.03	3.4%
Ribosomal L22	52.m06938	CIMG_09618	Y	4	High	1.01	0.15	14.6%
Hsp60	52.m06672	CIMG_06278	Y	5	High	1.04	0.03	2.7%
Cytochrome C	73.m03439	CIMG_05096	Y	1	High	1.04	0.09	8.2%
Acetyl-CoA acetyltransferase	61.m01556	CIMG_02262	Y	1	Moderate	1.06	0.21	20.3%
Ribosomal L2	52.m06573	CIMG_06034	Y	4	High	1.08	0.06	5.2%
Ribosomal L8	67.m08322	CIMG_01685	Y	4	High	1.09	0.22	20.1%
Ribosomal S21	73.m03724	CIMG_04499	Y	4	High	1.13	0.19	16.6%
Ribosomal L6	52.m07545	CIMG_06963	Y	4	Moderate	1.29	0.17	12.9%
Elongation factor 1 beta	52.m07063	CIMG_06970	Y	4	High	1.37	0.19	13.8%
Ribosomal L4	52.m07369	CIMG_06503	Y	4	High	1.43	0.17	12.2%
Ubiquitin	12.m07650	CIMG_03821	**N**	4	High	1.45	0.24	16.4%
Ribosomal S11	45.m00895	CIMG_08046	Y	4	High	1.47	0.18	12.5%
Ribosomal S7	12.m07699	CIMG_03754	Y	4	High	1.47	0.33	22.5%
Protein disulfide isomerase	52.m07651	CIMG_07225	Y	4	Moderate	1.53	0.32	20.9%
Ribosomal S15	61.m01540	CIMG_02223	Y	4	High	1.55	0.35	22.7%
Ribosomal S5	12.m08026	CIMG_03305	Y	4	High	1.61	0.08	4.7%
Hsp88	52.m06977	CIMG_06861	**N**	5	Moderate	1.63	0.23	14.1%
Endoribonuclease	65.m01864	CIMG_07737	**N**	3	Moderate	1.63	0.50	30.5%
Ribosomal S3	12.m07592	CIMG_03903	Y	4	High	1.65	0.28	17.3%
Ribosomal L27	67.m08017	CIMG_02079	Y	4	High	1.67	0.13	8.0%
Ribosomal L12	73.m03510	CIMG_04811	Y	4	High	1.74	0.41	23.5%
Ribosomal L14	13.m01821	CIMG_10151	Y	4	Moderate	1.74	0.37	21.3%
Ribosomal S8	73.m03550	CIMG_04980	Y	4	High	1.77	0.13	7.2%
Ribosomal S4	67.m09348	CIMG_00391	Y	4	High	2.00	0.28	13.8%
Elongation factor 2	73.m03507	CIMG_05034	Y	4	High	2.24	0.45	20.3%
Hsp90	73.m03734	CIMG_04729	Y	5	High	2.28	0.31	13.4%
Aminopeptidase	52.m06685	CIMG_06320	**N**	4	Moderate	2.77	0.90	32.3%
Aldehyde dehydrogenase	52.m06796	CIMG_09805	Y	1	High	2.84	0.48	16.7%
Arginosuccinate synthase	12.m07947	CIMG_03406	Y	3	High	3.14	0.19	5.9%
Ribosomal S18	60.m01388	CIMG_02814	Y	4	High	3.16	1.91	60.4%
Elongation factor 1 alpha	12.m07727	CIMG_03708	Y	4	High	11.06	5.55	50.2%

a) Protein sequence identity to human proteins (Low = <30%, Moderate = 30%-50%, High = >50%)

b) Coefficient of Variation; calculated as: standard deviation/ratio * 100

adapted from Kolkman *et al.*[190] The final list of high abundance spherule proteins identified from all spherule time points with sequence prediction and human homology analysis is shown in Table 4.3. There are a total of 4 proteins that are considered good protein antigen targets, one of which (the Subtilisin-like protease, PepC) has been previously described in Chapter 3 of this dissertation. The remaining three proteins are described below.

4.3.3.1 Isochorismatase-family protein (67.m09017) IFP

This protein is predicted to be a potential vaccine antigen target based on identification only in spherule cells and the prediction of an N-terminal signal sequence. This protein was identified in 96 and 120 hour spherules, but not in mycelia. This protein has 36% sequence identity to the human Isochorismatase-domain protein ISOC1.

4.3.3.2 Ketol-acid reductoisomerase (72.m02091) KAR

This protein was found to be more highly expressed in 48 and 96 hour spherules, and contains a predicted N-terminal signal sequence suggesting possible extracellular transport.

4.3.3.3 Flavodomain-containing protein (73.m03535) CpY20

This protein exhibits 33% sequence identity (22/51 of 203 residue protein expectation score = 0.34) to an unknown function human protein. This protein was identified in the spherule cell wall analysis presented in Chapter 3, but was not selected

Table 4.3 Antigen target evaluation of highly expressed proteins from ^{15}N labeled spherule cells

Description	TIGR locus	Broad locus	Found in Cell Wall Analysis	Sequence Predictions[a]	Human Homology[b]	$^{14}N/^{15}N$ ratio[c]			Antigen Target
						48 hr Spherule	96 hr Spherule	120 hr Spherule	
Isochorismatase family protein	67.m09017	CIMG_00799	N	signal	moderate	-	15N only	15N only	yes
Subtilisin-like protease	52.m06866	CIMG_06672	Y	signal	low	-	0.72 (0.09, 13.1%)	0.65 (0.14, 22.0%)	yes
Ketol-acid reductoisomerase	72.m02091	CIMG_05641	Y	signal	low	0.61 (0.06, 9.9%)	0.35 (0.04, 10.6%)	-	yes
CPY20 protein	73.m03535	CIMG_04756	Y	none	low	0.33 (0.02, 5.1%)	-	-	yes
Myocardin-related transcription factor	45.m00937	unknown	N	none	moderate	-	15N only	-	unknown
Unknown function protein	61.m01706	CIMG_02469	N	none	low	-	-	15N only	unknown
Unknown function protein	67.m08559	CIMG_01333	N	none	low	-	15N only	15N only	unknown
Unknown function protein	13.m01892	CIMG_10240	Y	none	low	-	15N only	15N only	unknown
Unknown function protein	61.m01576	CIMG_02272	Y	none	low	15N only	15N only	15N only	unknown
Alcohol dehydrogenase	73.m03409	CIMG_04945	Y	none	low	0.92 (0.04, 4.0%)	0.45 (0.10, 23.0%)	0.25 (0.04, 17.8%)	no
Aminotransferase	52.m07600	CIMG_07122	Y	none	low	0.91 (0.05, 5.9%)	-	-	no
Aldolase	12.m07770	CIMG_03654	Y	none	low	0.87 (0.18, 20.7%)	0.51 (0.05, 10.1%)	-	no
2-Methyl citrate dehydratase	13.m01811	CIMG_10136	Y	none	low	0.58 (0.07, 11.9%)	0.48 (0.03, 6.5%)	0.45 (0.04, 9.1%)	no
Phosphoenol pyruvate carboxykinase	67.m08638	CIMG_01264	Y	none	low	0.56 (0.01, 2.2%)	0.89 (0.21, 23.8%)	0.63 (0.07, 11.5%)	no
Isocitrate lyase	13.m01812	CIMG_10137	N	none	low	0.51 (0.06, 11.9%)	-	-	no
Cytochrome C peroxidase	51.m00579	CIMG_08209	Y	membrane	low	0.37 (0.05, 12.5%)	-	0.30 (0.07, 23.6%)	no
Homocysteine methyltransferase	12.m07493	CIMG_04062	Y	none	low	-	-	0.65 (0.07, 10.5%)	no
Hsp10	52.m06460	CIMG_09671	Y	none	moderate	-	0.51 (0.08, 16.6%)	-	no
Peroxisomal membrane protein	52.m07000	CIMG_05828	Y	none	moderate	-	0.87 (0.06, 7.4%)	0.59 (0.08, 13.4%)	no
Ribosomal L13	68.m01881	CIMG_08619	Y	none	moderate	-	-	0.95 (0.09, 9.6%)	no
Polyketide synthase	9.m00282	CIMG_03014	N	none	moderate	15N only	-	-	no
Glutathione reductase	12.m08188	CIMG_03087	N	none	moderate	15N only	-	-	no
D-lactate dehydrogenase	52.m07346	CIMG_06424	N	none	high	15N only	-	-	no
3-Hydroxyacyl-CoA dehydrogenase	67.m09623	CIMG_01176	N	none	moderate	15N only	-	-	no
Tyrosinase-family	65.m01966	CIMG_07583	Y	signal	low	15N only	-	-	no
Glutamate dehydrogenase	12.m07468	CIMG_04095	N	none	low	15N only	15N only	-	no

a) Predicted extracellular localization cues derived from protein sequence analysis. signal = N-terminal signal sequence, membrane = transmembrane helix

b) Protein sequence identity to human proteins (Low = <30%, Moderate = 30%-50%, High = >50%)

c) Ratio of protein abundance in 14N-labeled mycelia/15N-labeled spherule at the time point indicated. Numbers in parenthesis indicate the ratio standard deviation and the coefficient of variation as explained in Table 4.2

Table 4.3 Antigen target evaluation of highly expressed proteins from ^{15}N labeled spherule cells (continued)

Description	TIGR locus	Broad locus	Found in Cell Wall Analysis	Sequence Predictions[a]	Human Homology[b]	$^{14}N/^{15}N$ ratio[c]			Antigen Target
						48 hr Spherule	96 hr Spherule	120 hr Spherule	
Acetyl-CoA hydrolase	68.m01968	CIMG_08724	N	none	low	15N only	15N only	15N only	no
Pyruvate decarboxylase	12.m07673	CIMG_03789	Y	none	low	15N only	15N only	15N only	no
Co-A transferase family	12.m07863	CIMG_03524	Y	signal	moderate	15N only	15N only	15N only	no
Malate synthase	13.m01794	CIMG_10106	Y	none	low	15N only	15N only	15N only	no
Uricase	73.m03878	CIMG_04344	Y	none	low	15N only	15N only	15N only	no
Carbohydrate kinase domain protein	61.m01486	CIMG_02220	N	none	moderate	-	-	15N only	no
Oxidoreductase	67.m08599	CIMG_01315	N	none	low	-	-	15N only	no
Ubiquitin-like protein	67.m08921	CIMG_00887	N	none	high	-	-	15N only	no
Dihydroxy-acid dehydratase	72.m01986	CIMG_05492	N	none	low	-	-	15N only	no
Aldehyde reductase	13.m01907	CIMG_10264	Y	none	moderate	-	-	15N only	no
Adenylsulfate kinase	52.m07022	CIMG_06918	Y	signal	high	-	-	15N only	no
Clathrin light chain	61.m01685	CIMG_02417	Y	none	moderate	-	-	15N only	no
Hsp70	61.m01694	CIMG_02494	Y	none	high	-	-	15N only	no
Short chain dehydrogenase	13.m01774	CIMG_10076	N	none	moderate	-	15N only	-	no
Alpha-methylacyl-CoA racemase	52.m06809	CIMG_06590	N	none	high	-	15N only	-	no
Acyl-CoA oxidase	52.m07157	CIMG_07097	N	none	moderate	-	15N only	-	no
Cystathionine beta-synthase	52.m07184	CIMG_06088	N	none	high	-	15N only	-	no
Formate dehydrogenase	60.m01455	CIMG_02650	Y	none	moderate	-	15N only	-	no
Unknown function protein	12.m07752	CIMG_03674	N	none	low	-	15N only	15N only	no
Unknown function protein	12.m07802	CIMG_03612	N	none	low	-	15N only	15N only	no
UV excision repair protein	52.m07256	CIMG_06202	N	none	moderate	-	15N only	15N only	no
Ribosomal L12	61.m01651	CIMG_02445	N	none	moderate	-	15N only	15N only	no
Vacuolar ATPase	65.m01840	CIMG_07768	N	none	moderate	-	15N only	15N only	no
GDSL-like Lipase	67.m09015	CIMG_00801	N	none	moderate	-	15N only	15N only	no
3-Ketoacyl-CoA thiolase	67.m08439	CIMG_01533	Y	none	high	-	15N only	15N only	no
Protein translation factor SUI1	67.m09230	CIMG_00415	Y	none	high	-	15N only	15N only	no
Aldo-keto reductase	73.m03658	CIMG_04588	Y	none	moderate	-	15N only	15N only	no

a) Predicted extracellular localization cues derived from protein sequence analysis. signal = N-terminal signal sequence, membrane = transmembrane helix

b) Protein sequence identity to human proteins (Low = <30%, Moderate = 30%-50%, High = >50%)

c) Ratio of protein abundance in 14N-labeled mycelia/15N-labeled spherule at the time point indicated. Numbers in parenthesis indicate the ratio standard deviation and the coefficient of variation as explained in Table 4.2

as a vaccine antigen candidate due to the lack of any predicted extracellular localization clue. Since it is approximately three times more abundant in 48 hour spherule cells, a closer inspection was made, and it was discovered that this protein is a likely homolog of the fungal pathogen *Paracoccidioides brasiliensis* Pby20 protein.[191] CpY20 and PbY20 share 78% sequence identity and 86% sequence similarity. PbY20 was also identified as a protein more highly expressed in the yeast form of the dimorphic pathogen, which is analogous to the spherule form in *Coccidioides*. In the same analysis, PbY20 was localized to the cytoplasm as well as to the cell wall.

In addition to the homology to the PbY20 protein, CpY20 also exhibits high sequence identity with two identified allergen proteins Cla h5 and Alt a7 produced by the allergenic molds *Cladosporium herbarum* and *Alternaria alternata*, respectively. CpY0 shares 63% sequence identity (74% similarity) with Cla h5 (also reported as Cla h7 in GenBank), and 67% identity (80% similarity) with Alt a7. Interestingly, CpY20 was also identified in a recent differential proteomic analysis of *C. posadasii*[71] where it was identified as a benzoquinone reductase and found to be approximately twice as abundant in 96 hour spherules as in mycelia. Due to the high level of spherule expression and homology to antigenically active allergen proteins, CpY20 is a strong candidate for further testing as a vaccine candidate for coccidioidomycosis.

4.3.3.4 Unknown function proteins

As presented in Table 4.3, there are five unknown function proteins that were found in spherules but not in mycelia. These proteins all exhibit low or moderate human

homology, but give no indication of cellular function, except one that shows some similarity to a myocardin-related transcription factor. These proteins do not have any predicted sequence clues to indicate extracellular association, but their identification in only spherule cells suggests they may be spherule specific. It is for this reason alone that they are included here.

4.3.3.5 Non-antigen target proteins with predicted signal sequences

As shown in Table 4.3, there are four other proteins identified as more highly abundant in spherules that have sequence predictions, but are not predicted to be antigen targets. One of these proteins (51.m00579) is predicted to have a membrane-spanning region, but is likely associated with the mitochondrial inner membrane, and not extracellularly associated. Another protein (65.m01966) is predicted to contain an N-terminal signal sequence, but is predicted by homology to be vacuolar associated. The protein 12.m07863 is also predicted to have an N-terminal signal sequence, but this protein is likely mitochondria-associated, and has a questionable signal sequence prediction probability of only 27%. The final protein with a predicted signal sequence is 52.m07022. This protein has a 62% N-terminal signal sequence probability, but is not considered an antigen target due to high human homology with 55% identity to the human phosphoadenosine phosphosulfate synthetase protein.

4.3.4 Comparison of relative protein quantification and relative mRNA expression levels

Using data produced from a serial analysis of gene expression (SAGE) study (performed by the Orbach research group in the Department of Plant Sciences at the University of Arizona)[192] comparing *C. posadasii* mycelia and spherule gene expression, we compared the protein abundance predictions from ^{15}N labeling experiments to the mRNA levels from SAGE analysis. Included were only those proteins that were identified in both the SAGE and MS/MS experiments and present in both cell types (i.e. no mycelia-only or spherule-only proteins). The results of this comparison are shown in Figures 4.3, 4.4 and 4.5. Each figure contains mycelia/spherule expression ratios for those genes with data from both experiments, including error bars indicating standard deviation of $^{14}N/^{15}N$ ratio averages based on three independent MS/MS experiments

Figure 4.4 Comparison of mRNA and protein expression levels for 48 hour spherules (see text for explanation of error bars)

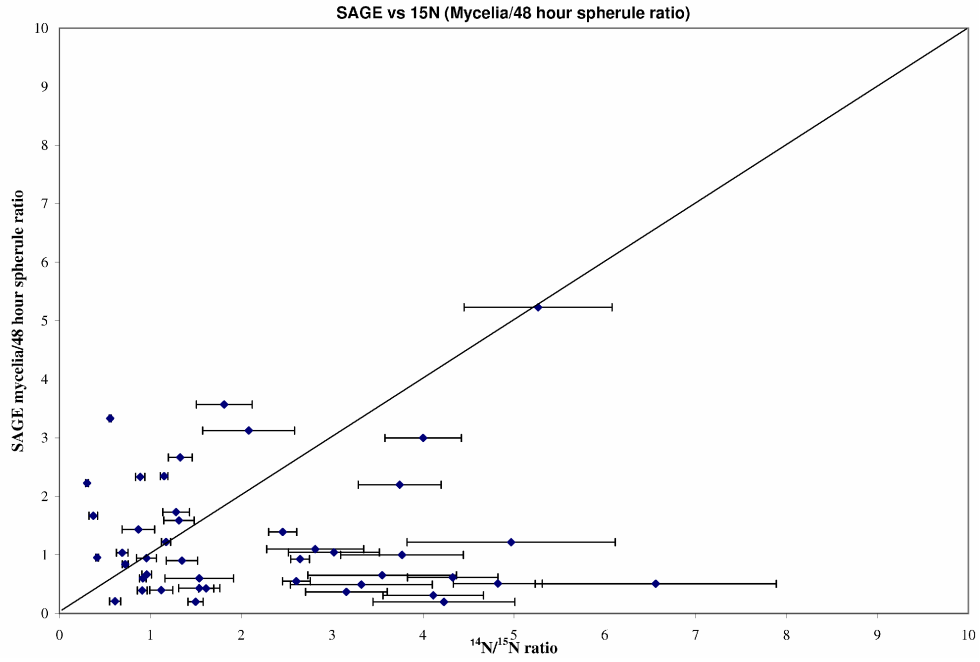


Figure 4.5 Comparison of mRNA and protein expression levels for 96 hour spherules (see text for explanation of error bars)

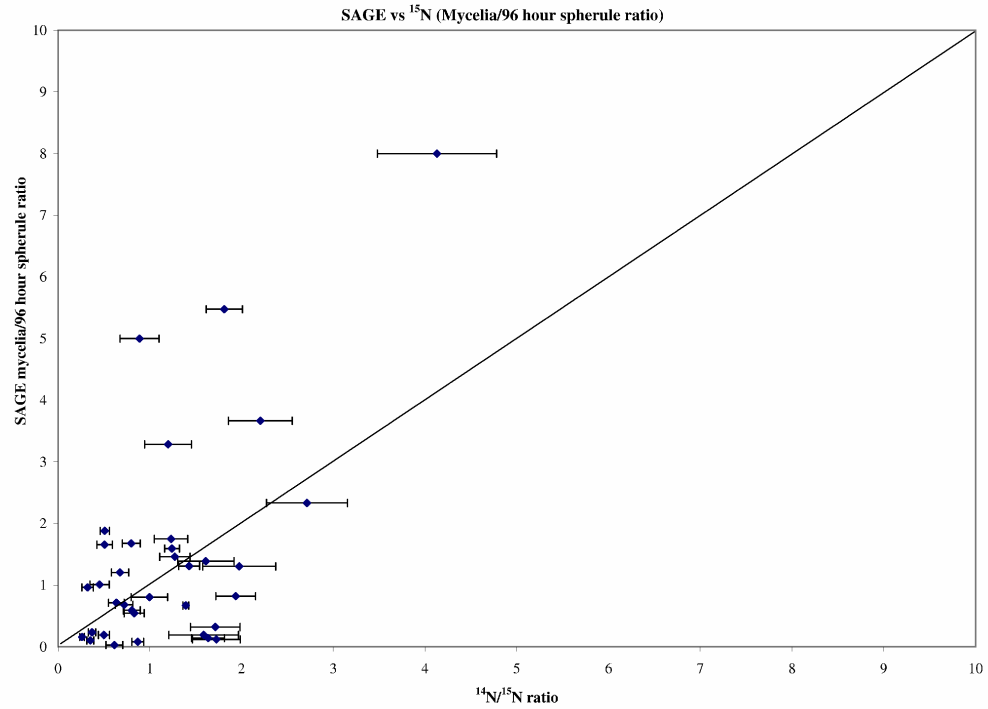
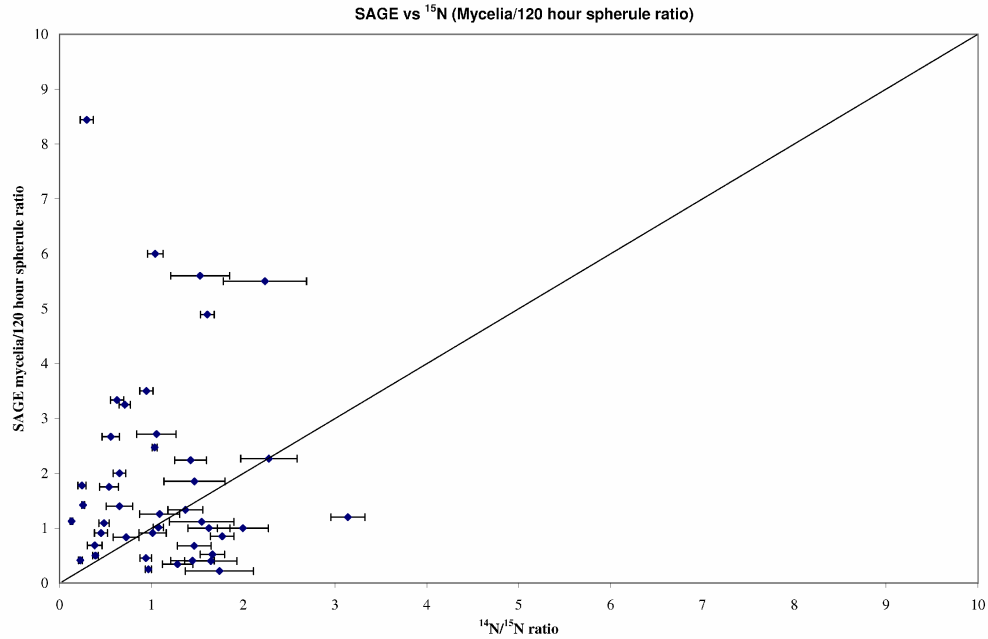


Figure 4.6 Comparison of mRNA and protein expression levels for 120 hour spherules (see text for explanation of error bars)



performed on the same mycelia/spherule sample. Each chart also contains a 1:1 correlation line for reference.

4.4 Discussion

We have described here a proteomic analysis of the *C. posadasii* cytoplasmic proteome using stable isotope labeling combined with 2-dimensional liquid chromatography tandem mass spectrometry in the search for protein vaccine antigen targets that are more highly expressed in the pathogenic spherule phase of the organism. With the help of a protein database search algorithm (SEQUEST) for protein identification and an algorithm for differential protein quantification (RelEx), we have identified a list of three new vaccine candidates along with one previously identified from the spherule cell wall proteome analysis described in Chapter 3. In addition to these four proteins, we have also identified five unknown function proteins that appear to be expressed only in spherules but cannot be localized by homology to known proteins to determine if they are extracellularly associated or similar to known antigenic proteins. These proteins may be worth further analysis, but with a possible risk of analyzing ineffective vaccine candidates.

Within our results, we have identified a known *C. posadasii* vaccine candidate, the Peroxisomal Membrane Protein (PMP1) that was identified in a previous differential protein expression analysis[71] and is known to be preferentially expressed in spherules. In addition, we have identified another known antigen, the recently reported[41] Cu, Zn superoxide dismutase that was not identified in the cell wall analysis from Chapter 3.

This protein is expected to be highly abundant, and is also predicted to be a cytoplasmic protein by homology (J. Lunetta personal communication). Both of these identifications serve to validate our technique for vaccine discovery.

During the data analysis of this work, we discovered an interesting trend in protein identification in ^{15}N labeled spherules from different time points. In order to increase the total number of proteins identified using ^{15}N-specified search parameters compared to ^{14}N-specific searches of the same data we reduced the SEQUEST ΔCn cutoff score down from 0.08 (default setting) to 0.02 in 48 hour spherules, 0.04 in 96 hour spherules, and no decrease in 120 hour spherules. The reason for this trend of greater protein ambiguity in younger ^{15}N labeled spherules is unknown. Perhaps there is some residual effect of ^{14}N-labeling in the arthroconidia spores used as the source cells for spherule production.

In the course of this study, we also compared the protein expression levels as produced by ^{15}N labeling experiments with mRNA expression levels from a serial analysis of gene expression (SAGE) experiment from a collaborating group. We were not able to find any discernible trend in these two expression profiles, much in line with previous studies. One study came to the conclusion that mRNA levels "provide little predictive value" when compared to protein expression.[193] Other studies have supported this conclusion,[194] even suggesting that changes in mRNA levels are only responsible for up to 40% of the variation in protein expression.[195] Thus, the lack of correlation presented here is not without precedent, and in fact supports the strategy of protein expression evaluation by proteomic analysis or other methods.

It is important to note, however that the spherule cell culture conditions between the SAGE study and the proteomic analysis were slightly different. While both experiments were performed using *C. posadasii* strain Silveira using modified Converse media, the ^{15}N-labeling was done without the addition of NZ amine in order to isolate the nitrogen source for labeling. NZ Amine is an enzymatic digest of the milk phosphoprotein casein which is used as a source of raw amino acids and small peptides in the cell culture medium, but is not absolutely necessary for the growth of spherules in culture. This additional carbon and nitrogen source does effect the growth of spherules in culture, and may be partly responsible for the lack of correlation between mRNA and protein levels.

By utilizing cutting edge research techniques like stable isotope labeling and tandem mass spectrometry-based proteomic analysis, we have advanced the search for protein vaccine candidates for coccidioidomycosis. From this analysis, we have identified spherule-dominant protein targets for further testing of vaccine components. We have also shown that analysis of protein expression levels is something that requires direct measurement of protein, rather than inference from mRNA expression levels.

5 CHAPTER FIVE: SUMMARY AND FUTURE DIRECTIONS

5.1 Summary of described work

The work presented in this dissertation represents a coalescing of multiple disciplines and areas of study such as analytical chemistry, bioinformatics, cell biology, and immunology. It is this bridging of scientific disciplines that helps to drive today's rapid flow of advancement in disease research in the biological sciences. In Chapter 2, a bioinformatic-based strategy of identifying proteins from single peptides in biological samples is described. Using a combination of protein database search algorithms we demonstrate an easily adopted method of increasing protein detection sensitivity derived from empirical analysis of protein identifications. In Chapter 3, this method is employed in the most comprehensive study of *Coccidioides posadasii* spherule cell wall proteins to date. This study utilized a combination of proteomics and bioinformatic approaches to identify several spherule protein antigen candidates for further immunologic testing as coccidioidomycosis vaccines. Chapter 4 describes the first application of stable isotope labeling for protein quantification in either *Coccidioides* species for the discovery of high abundance spherule antigens. As a result of these analyses, we have identified 10 proteins that are excellent candidates for immunologic testing as protein-based vaccines, including 17 additional unknown function or localization proteins that may be immunogenic but with substantially higher risk for analysis as vaccines.

5.2 Mass spectrometry and proteomics

As detailed in Chapters 1 and 2, there are a multitude of protein database search algorithms in use for peptide identification from mass spectrometric analyses. As described, these algorithms are far from perfect when it comes to complete automated identification of proteins in complex mixtures. The most pressing need at this time is for further understanding of peptide fragmentation in tandem mass spectrometry experiments. Current algorithms do very little, if any, matching of ion intensities from MS/MS spectra. Simple incorporation of knowledge like the tendency for increased cleavage N-terminal to proline residues[103] as well as changes in cleavage patterns based on the number and location of basic residues in a peptide[196] would likely improve the accuracy of current peptide matching algorithms. While some of this has already occurred, (see the description of XTandem in Section 1.2.6.1.3) more integration of this knowledge into search algorithms is necessary

A concept as simple as the application of liquid chromatography retention times to peptide identification would be useful in MS/MS experiments. Currently, LC is primarily used for separation of peptide mixtures prior to MS analysis. The LC process is capable of providing more information regarding peptide identification than is currently utilized.[197] There has been considerable work done regarding the use of LC retention times (as reviewed in [198]) in proteomic analyses, however this technique primarily involves the coupling of LC retention times with accurate mass tags (AMTs) generated by high mass accuracy instruments such as FTICR or TOF mass spectrometers. While beneficial, this process does not allow the use of more common quadrupole and linear ion

traps so often used in proteomics. However, the use of LC retention times could still be used in these instruments. For example: a spectrum that matches two peptide sequences well is likely to be thrown out as too ambiguous by current search algorithms. If the two peptides are sufficiently different in terms of hydrophobicity, the LC gradient conditions at the time of fragmentation could be used to rule out one of the matches, resulting in a decrease in overall false negative identifications. This is another aspect of peptide search algorithm development that could provide beneficial capabilities.

Identification of proteins from single peptides found in MS/MS analysis is an area that requires further study. While a technique for improving the protein identifications of proteomic analyses is detailed in Chapter 2, this method is merely a stopgap measure of increasing the data from a proteomic analysis. It can be argued that the SEQUEST and XTandem search algorithms are similar enough to each other (as described in section 1.2.6.1) such that they make the same incorrect assignment for a spectrum. While this argument certainly has merit, in practice the dual algorithm technique produces very different results between the search programs. A good example of this is the fact that in all of the MS/MS experiments performed on spherule cell walls (See Chapter 3), the majority (>90%) of the validated single peptide identifications were from *Coccidioides posadasii* proteins. It is important to realize that an incorrect peptide match is a random event, and therefore the incorrect peptides should be found from a random sampling of organisms present in the search database used (as shown in Table 3.1). Given that *C. posadasii* proteins comprise less than 4% (7202/201596) of the total number of protein sequences in the database, random *C. posadasii* matches should also be correspondingly

low. The fact that most of the protein identifications come from *C. posadasii* suggest single peptide identifications, as selected by both SEQUEST and XTandem, are not comprised of random matches.

5.3 Vaccine development efforts for *Coccidioides spp.*

Given that no single vaccine antigen identified to date provides protective immunity to the same degree as killed-cell vaccines, it is reasonable to predict that a vaccine fielded for coccidioidomycosis will be multivalent[70], or perhaps chimeric.[52] While we have not finished the task of identifying antigenic proteins for vaccine development, more comprehensive analyses of multiple-protein vaccines need to be undertaken.

It is important to understand, however, that the efficiency of a protein vaccine is based on more than just the amino acid sequence of the proteins included. There are additional issues such as the degree and type of protein post-translational modifications (PTMs) such as glycosylation that are present in *Coccidioides*-produced proteins that are not duplicated in standard expression systems such as *E. coli* or *S. cerevisiae*. If *Coccidioides*-produced PTMs are important in the host immune response, those PTMs are unlikely to be duplicated in an expression system in another organism. In addition, adjuvant selection is important in initiating the proper immune response. It is known that a Th1 (cell-mediated) immune response is more indicative of a good clinical prognosis, while a Th2 (antibody) mediated response is considered a poor protective response.[3] The combination of adjuvant and the antigen(s) present in the vaccine help determine which

immune response pathway is initiated. An understanding of the effects of glycosylation, adjuvant selection, as well as the type of immune response desired is also very important in the vaccine development effort.

Additional studies to evaluate the vaccine effectiveness of the protein vaccine candidates identified by the proteomic methods described here will be undertaken in future work. These candidates will ultimately be tested in mouse survival studies typically used for evaluation of recombinant protein antigens for effectivene immunity against coccidioidomycosis. Prior to this expensive, labor intensive process, however, it would be beneficial to evaluate these proteins using antibody-based methods. Two such methods have been discussed, including a microarray system in which proteins of interest are added to the array by *in vivo* cloning of the gene followed by *in vitro* transcription and translation.[199] The native protein is then attached to the array plate and evaluated for antigenic properties by passing sera from infected human patients over and looking for antibody recognition of the arrayed proteins. This method could prove useful in evaluating these vaccine candidates, however there are some drawbacks to be familiar with. First, since the *in vivo* cloning cannot easily be performed in *Coccidioides*, any protein produced by *in-vitro* methods would not have the same post translational modifications as the true native protein. Second, antibody recognition of three-dimensional structure is not as likely to occur as primary sequence recognition. This is due to the fact that antigen presenting cells in the immune system will proteolyze the antigen and only present peptide antigens for antibody recognition, rather than native structure. An alternative testing strategy is being proposed by the Biodesign Institute at

Arizona State University in which 20-residue peptides encompassing the entire sequence of the candidate protein to be tested are attached to a microarray plate and evaluated for antibody recognition in the same manner as described above. This method would evaluate antibody recognition of primary amino acid sequence. This method is currently limited by a need to eliminate cysteine residues from the oligopeptides attached to the array plate. This would prevent analysis of protein sequence regions containing cysteine residues. Despite the drawbacks described, either of these microarray systems would likely be beneficial in the testing of proposed protein vaccine candidates.

5.4 Stable isotope labeling and protein expression quantification

As explained in Chapter 4, stable isotope labeling is a valuable technique for differential protein expression analysis, but it certainly comes with limitations. While the method of ^{15}N labeling described here carries the advantages of biological incorporation as well as labeling of all peptides (as opposed to a residue-specific modification), the increased complexity of labeled peptides introduces additional ambiguity in protein identification. One way to help minimize this effect is to use a dual-labeling experimental approach in which each of the cell types is separately ^{15}N labeled and analyzed with the unlabeled cell type. Proteins quantified in this manner can be correlated between experimental runs for verification purposes. Unfortunately, this is not easily performed with the spherule-mycelia comparison as described, since the mycelia cell culture medium is significantly less defined than the spherule media, which makes isolation of the mycelia nitrogen source problematic. One possible solution to this would

be ^{15}N labeling of *S. cerevisiae*[200] to produce a labeled yeast extract for use in mycelia growth media. Since protein identification of unlabeled samples is less problematic, more spherule proteins would likely be identified in a ^{14}N spherule-^{15}N mycelia study and correlated to expression levels in mycelia.

Although the comparisons between mRNA and protein expression levels do not correlate well in the experiments described in Chapter 4, further use of mRNA expression data is not without benefit. Amplification of mRNA allows for the identification of gene products that are in low abundance, which are often missed by MS/MS analysis. Recent analyses of specific mRNA to protein correlations suggest that only 20%-30% of the difference between protein concentrations is attributable to mRNA levels alone.[201] One of the reasons for poor mRNA to protein correlation include the translational activity (TA) of a gene[202], which is calculated from the mRNA abundance and number of ribosomes per mRNA. Others reasons include tRNA concentrations that can slow translation of proteins with sub-optimal codon usage,[203] the blocking of mRNA translation,[202] or changes in protein degredation.[204] A gene-specific analysis of any or all of these translational control mechanisms is necessary to characterize protein abundance from mRNA levels.

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the Air Force, Department of Defense or the United States Government.

REFERENCES

1. Galgiani, J. N. 1993. Coccidioidomycosis. West J Med **159:** 153-171.

2. Kirkland, T. & G. Cole. 2002. Gene-finding in *Coccidioides immitis*. *In* Pathogen genomics : impact on human health. Shaw, K. J., Ed.: 247-254. Humana Press. Totowa, N.J.

3. Cox, R. A. & D. M. Magee. 1998. Protective immunity in coccidioidomycosis. Res Immunol **149:** 417-428.

4. Posada, A. 1892. Uno nuevo caso de micosis fungoidea con psorospermias. Ann Circulo Medico Argentino **15:** 585-596.

5. Wernicke, R. 1892. Ueber einen protozoenbefund bei mycosis fungoides. Zentralbl Bakteriol **12:** 859-861.

6. Rixford, E. & T. C. Gilchrist. 1896. Two cases of protozoan (coccidioidal) infection of the skin and other organs. Johns Hopkins Hosp Rev **1:** 209-268.

7. Ophüls, W. & H. C. Moffitt. 1900. A new pathogenic mould (formerly described as a protozoan: *Coccidioides immitis pyogenes*). Philadelphia Med. J. **5:** 1471-1472.

8. Ophüls, W. 1905. Further observations on a pathogenic mould formerly described as a protozoon (*Coccidioides immitis*, *Coccidioides pyogenes*). J Exp Med **6:** 443-U414.

9. Fiese, M. J. 1958. Coccidioidomycosis. Thomas. Springfield, Ill.,.

10. Dickson, E. C. 1937. "Valley Fever" of the san joaquin valley and fungus *Coccidioides.* Calif West Med **47:** 151-155.

11. Gifford, M. A., W. C. Buss, R. J. Douds*, et al.* 1936. Data on *Coccidioides* fungus infection. Kern County Dept Pub Health Ann Rep39-54.

12. Kwonchun, K. 1969. *Coccidioides immitis* - Cytological study on formation of arthrospores. Can J Gen Cyt **11:** 43-53.

13. Sun, S. H., S. S. Sekhon & M. Huppert. 1979. Electron-microscopic studies of saprobic and parasitic forms of *Coccidioides immitis*. Sabouraudia **17:** 265-273.

14. Sun, S. H. & M. Huppert. 1976. A cytological study of morphogenesis in *Coccidioides immitis*. Sabouraudia **14:** 185-198.

15. Cole, G. T., J. M. Xue, C. N. Okeke*, et al.* 2004. A vaccine against coccidioidomycosis is justified and attainable. Med Mycol **42:** 189-216.

16. Drutz, D. J. & M. Huppert. 1983. Coccidioidomycosis - factors affecting the host-parasite interaction. J Infect Dis **147:** 372-390.

17. Delgado, N., J. Xue, J. J. Yu*, et al.* 2003. A recombinant beta-1,3-glucanosyltransferase homolog of *Coccidioides posadasii* protects mice against coccidioidomycosis. Infect Immun **71:** 3010-3019.

18. Bowman, B. H., T. J. White & J. W. Taylor. 1996. Human pathogenic fungi and their close nonpathogenic relatives. Mol Phy Evo **6:** 89-96.

19. Fisher, M., G. Koenig, T. White*, et al.* 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-california population of *Coccidioides immitis*. Mycologia **91:** 73-84.

20. Friedman, L., C. E. Smith & L. E. Gordon. 1955. The assay of virulence of *Coccidioides* in white mice. J Infect Dis **97:** 311-316.

21. Yuan, L. & G. T. Cole. 1987. Isolation and characterization of an extracellular proteinase of *Coccidioides immitis*. Infect Immun **55:** 1970-1978.

22. Walch, H. A. & R. K. Walch. 1967. Studies with induced mutants of *Coccidioides immitis*. *In* Coccidioidomycosis: papers. Ajello, L., Ed.: 339-347. University of Arizona Press. Tucson.

23. Smith, C. E. 1940. Epidemiology of acute coccidioidomycosis with erythema nodosum ("san joaquin" or "valley fever"). Am J Pub Health **30:** 600-611.

24. Koufopanou, V., A. Burt & J. W. Taylor. 1997. Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. Proc Natl Acad Sci U S A **94:** 5478-5482.

25. Pappagianis, D. 2001. Seeking a vaccine against *Coccidioides immitis* and serologic studies: expectations and realities. Fungal Genet Biol **32:** 1-9.

26. Levine, H. B., J. M. Cobb & C. E. Smith. 1960. Immunity to coccidioidomycosis induced in mice by purified spherule, arthrospore, and mycelial vaccines. Trans N Y Acad Sci **22:** 436-449.

27. Levine, H. B. & C. E. Smith. 1967. The reactions of eight volunteers injected with *Coccidioides immitis* spherule vaccine: First human trials. *In* Coccidioidomycosis. Ajello, L., Ed.: 197-200. Univ. of Arizona Press. Tucson.

28. Pappagianis, D. 1967. Further studies on vaccination of human volunteers with killed *Coccidioides immitis*. *In* Coccidioidomycosis. Ajello, L., Ed.: 201-210. Univ. of Arizona Press. Tucson.

29. Pappagianis, D. 1993. Evaluation of the protective efficacy of the killed *Coccidioides immitis* spherule vaccine in humans. The Valley Fever Vaccine Study Group. Am Rev Respir Dis **148:** 656-660.

30. Kong, Y. C., H. B. Levine & C. E. Smith. 1963. Immunogenic properties of nondisrupted and disrupted spherules of *Coccidioides immitis* in mice. Sabouraudia **2:** 131-142.

31. Pappagianis, D., R. Hector, H. B. Levine*, et al.* 1979. Immunization of mice against coccidioidomycosis with a subcellular vaccine. Infect Immun **25:** 440-445.

32. Ward, E. R., Jr., R. A. Cox, J. A. Schmitt, Jr.*, et al.* 1975. Delayed-type hypersensitivity responses to a cell wall fraction of the mycelial phase of *Coccidioides immitis*. Infect Immun **12:** 1093-1097.

33. Lecara, G., R. A. Cox & R. B. Simpson. 1983. *Coccidioides immitis* vaccine: potential of an alkali-soluble, water-soluble cell wall antigen. Infect Immun **39:** 473-475.

34. Cox, R. A. & D. M. Magee. 2004. Coccidioidomycosis: host response and vaccine development. Clin Microbiol Rev **17:** 804-839.

35. Cox, R. A. & L. A. Britt. 1985. Antigenic heterogeneity of an alkali-soluble, water-soluble cell wall extract of *Coccidioides immitis*. Infect Immun **50:** 365-369.

36. Cole, G. T., K. R. Seshan, M. Franco*, et al.* 1988. Isolation and morphology of an immunoreactive outer wall fraction produced by spherules of *Coccidioides immitis*. Infect Immun **56:** 2686-2694.

37. Cole, G. T., T. N. Kirkland, M. Franco*, et al.* 1988. Immunoreactivity of a surface wall fraction produced by spherules of *Coccidioides immitis*. Infect Immun **56:** 2695-2701.

38. Zimmermann, C. R., S. M. Johnson, G. W. Martens*, et al.* 1998. Protection against lethal murine coccidioidomycosis by a soluble vaccine from spherules. Infect Immun **66:** 2342-2345.

39. Zimmermann, C. R., S. M. Johnson, K. M. Kerekes*, et al.* 1999. Development and characterization of a new soluble, subcellular vaccine against infection in mice by *Coccidioides immitis*. Abstr Gen Meet Am Soc Microbiol. Chicago, Illinois,

40. Johnson, S. M., K. M. Kerekes, J. M. Lunetta*, et al.* 2007. Characteristics of the Protective Subcellular Coccidioidal T27K Vaccine. Ann N Y Acad Sci **1111:** 275-289.

41. Lunetta, J. M., K. A. Simmons, S. M. Johnson*, et al.* 2007. Molecular cloning and expression of a cDNA encoding a *Coccidioides posadasii* Cu,Zn superoxide dismutase protein identified by proteomic analysis of the coccidioidal T27K vaccine. Ann N Y Acad Sci **1111:** 164-180.

42. Huppert, M., N. S. Spratt, K. R. Vukovich*, et al.* 1978. Antigenic analysis of coccidioidin and spherulin determined by two-dimensional immunoelectrophoresis. Infect Immun **20:** 541-551.

43. Cox, R. A., M. Huppert, P. Starr*, et al.* 1984. Reactivity of alkali-soluble, water-soluble cell wall antigen of *Coccidioides immitis* with anti-*Coccidioides* immunoglobulin M precipitin antibody. Infect Immun **43:** 502-507.

44. Zhu, Y., C. Yang, D. M. Magee*, et al.* 1996. Molecular cloning and characterization of *Coccidioides immitis* antigen 2 cDNA. Infect Immun **64:** 2695-2699.

45. Dugger, K. O., J. N. Galgiani, N. M. Ampel*, et al.* 1991. An immunoreactive apoglycoprotein purified from *Coccidioides immitis*. Infect Immun **59:** 2245-2251.

46. Dugger, K. O., K. M. Villareal, A. Ngyuen*, et al.* 1996. Cloning and sequence analysis of the cDNA for a protein from *Coccidioides immitis* with immunogenic potential. Biochem Biophys Res Commun **218:** 485-489.

47. Peng, T., L. Shubitz, J. Simons*, et al.* 2002. Localization within a proline-rich antigen (Ag2/PRA) of protective antigenicity against infection with *Coccidioides immitis* in mice. Infect Immun **70:** 3330-3335.

48. Kaufman, L. & P. Standard. 1978. Immuno-identification of cultures of fungi pathogenic to man. Curr Microbiol **1:** 135-140.

49. Cox, R. A. & L. A. Britt. 1986. Isolation and identification of an exoantigen specific for *Coccidioides immitis*. Infect Immun **52:** 138-143.

50. Cole, G. T., S. W. Zhu, S. C. Pan*, et al.* 1989. Isolation of antigens with proteolytic activity from *Coccidioides immitis*. Infect Immun **57:** 1524-1534.

51. Pan, S. & G. T. Cole. 1995. Molecular and biochemical characterization of a *Coccidioides immitis*-specific antigen. Infect Immun **63:** 3994-4002.

52. Shubitz, L. F., J. J. Yu, C. Y. Hung*, et al.* 2006. Improved protection of mice against lethal respiratory infection with *Coccidioides posadasii* using two recombinant antigens expressed as a single protein. Vaccine **24:** 5904-5911.

53. Yu, J. J., S. L. Smithson, P. W. Thomas*, et al.* 1997. Isolation and characterization of the urease gene (URE) from the pathogenic fungus *Coccidioides immitis*. Gene **198:** 387-391.

54. Li, K., J. J. Yu, C. Y. Hung*, et al.* 2001. Recombinant urease and urease DNA of *Coccidioides immitis* elicit an immunoprotective response against coccidioidomycosis in mice. Infect Immun **69:** 2878-2887.

55. Mirbod-Donovan, F., R. Schaller, C. Y. Hung*, et al.* 2006. Urease produced by *Coccidioides posadasii* contributes to the virulence of this respiratory pathogen. Infect Immun **74:** 504-515.

56. Xue, J., C. Y. Hung, J. J. Yu*, et al.* 2005. Immune response of vaccinated and non-vaccinated mice to *Coccidioides posadasii* infection. Vaccine **23:** 3535-3544.

57. Ivey, F. D., D. M. Magee, M. D. Woitaske*, et al.* 2003. Identification of a protective antigen of *Coccidioides immitis* by expression library immunization. Vaccine **21:** 4359-4367.

58. Hung, C. Y., N. M. Ampel, L. Christian*, et al.* 2000. A major cell surface antigen of *Coccidioides immitis* which elicits both humoral and cellular immune responses. Infect Immun **68:** 584-593.

59. Hung, C. Y., J. J. Yu, K. R. Seshan*, et al.* 2002. A parasitic phase-specific adhesin of *Coccidioides immitis* contributes to the virulence of this respiratory fungal pathogen. Infect Immun **70:** 3443-3456.

60. Kirkland, T. N. & G. T. Cole. 2002. Coccidioidomycosis: Pathogenesis, Immune Response, and Vaccine Development. *In* Fungal pathogenesis : principles and clinical applications. Calderone, R. A. & R. L. Cihlar, Eds.: 365-399. Marcel Dekker. New York.

61. Hung, C. Y., K. R. Seshan, J. J. Yu*, et al.* 2005. A metalloproteinase of *Coccidioides posadasii* contributes to evasion of host detection. Infect Immun **73:** 6689-6703.

62. Johannesson, H., J. P. Townsend, C. Y. Hung*, et al.* 2005. Concerted evolution in the repeats of an immunomodulating cell surface protein, SOWgp, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*. Genetics **171:** 109-117.

63. Hung, C. Y., J. Xue & G. T. Cole. 2007. Virulence Mechanisms of *Coccidioides*. Ann N Y Acad Sci **1111:** 225-235.

64. Kirkland, T. N., S. W. Zhu, D. Kruse*, et al.* 1991. *Coccidioides immitis* fractions which are antigenic for immune T lymphocytes. Infect Immun **59:** 3952-3961.

65. Cole, G. T., T. N. Kirkland & S. H. Sun. 1987. An immunoreactive, water-soluble conidial wall fraction of *Coccidioides immitis*. Infect Immun **55:** 657-667.

66. Wyckoff, E. E., E. J. Pishko, T. N. Kirkland*, et al.* 1995. Cloning and expression of a gene encoding a T-cell reactive protein from *Coccidioides immitis*: homology to 4-hydroxyphenylpyruvate dioxygenase and the mammalian F antigen. Gene **161:** 107-111.

67. Kirkland, T. N., P. W. Thomas, F. Finley*, et al.* 1998. Immunogenicity of a 48-kilodalton recombinant T-cell-reactive protein of *Coccidioides immitis*. Infect Immun **66:** 424-431.

68. Thomas, P. W., E. E. Wyckoff, E. J. Pishko*, et al.* 1997. The hsp60 gene of the human pathogenic fungus *Coccidioides immitis* encodes a T-cell reactive protein. Gene **199:** 83-91.

69. Tarcha, E. J., V. Basrur, C. Y. Hung*, et al.* 2006. A recombinant aspartyl protease of *Coccidioides posadasii* induces protection against pulmonary coccidioidomycosis in mice. Infect Immun **74:** 516-527.

70. Tarcha, E. J., V. Basrur, C. Y. Hung*, et al.* 2006. Multivalent recombinant protein vaccine against coccidioidomycosis. Infect Immun **74:** 5802-5813.

71. Orsborn, K. I., L. F. Shubitz, T. Peng*, et al.* 2006. Protein expression profiling of *Coccidioides posadasii* by two-dimensional differential in-gel electrophoresis and evaluation of a newly recognized peroxisomal matrix protein as a recombinant vaccine candidate. Infect Immun **74:** 1865-1872.

72. Lunetta, J., S. Johnson & D. Pappagianis. 2006. Identification of a Cu,Zn-superoxide dismutase in the coccidioidal T27K vaccine using proteomic methods. Presented at the 6th International Symposium on Coccidioidomycosis. Stanford University, CA, Aug 25.

73. Lunetta, J., S. Johnson & D. Pappagianis. 2006. Identification of a class I 1,2 alpha-mannosidase protein in the coccidioidal T27K vaccine using immunoproteomic methods. Presented at the 6th International Symposium on Coccidioidomycosis. Stanford University, CA, Aug 25.

74. Lunetta, J. M., K. A. Simmons, S. M. Johnson*, et al.* 2007. Molecular cloning and expression of a cDNA encoding a *Coccidioides posadasii* 1,2-{alpha}-mannosidase

identified in the coccidioidal T27K vaccine by immunoproteomic methods. Ann N Y Acad Sci **1111:** 164-180.

75. Johnson, S. & D. Pappagianis. 2006. Identification and molecular cloning of a pH sensitive protein-like protein present in the coccidioidal vaccine T27K. Presented at the 6th International Symposium on Coccidioidomycosis. Stanford University, CA, Aug 25.

76. Hunter, T. C., N. L. Andon, A. Koller*, et al.* 2002. The functional proteomics toolbox: methods and applications. J Chromatogr B Analyt Technol Biomed Life Sci **782:** 165-181.

77. Corthals, G. L., V. C. Wasinger, D. F. Hochstrasser*, et al.* 2000. The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis **21:** 1104-1115.

78. Pradet-Balade, B., F. Boulme, H. Beug*, et al.* 2001. Translation control: bridging the gap between genomics and proteomics? Trends Biochem. Sci. **26:** 225-229.

79. Kelleher, N. L., H. Y. Lin, G. A. Valaskovic*, et al.* 1999. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. J Am Chem Soc **121:** 806-812.

80. Thiede, B., W. Hohenwarter, A. Krah*, et al.* 2005. Peptide mass fingerprinting. Methods **35:** 237-247.

81. Yates, J. R., 3rd. 1998. Database searching using mass spectrometry data. Electrophoresis **19:** 893-900.

82. Yates, J., J. Eng, A. McCormack*, et al.* 1995. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem **67:** 1426-1436.

83. Siuzdak, G. 2003. The Expanding Role of Mass Spectrometry in Biotechnology. MCC Press. San Diego, CA.

84. Hames, B. D. 1998. Gel electrophoresis of proteins : a practical approach. Oxford University Press. Oxford ; New York.

85. Wittmann-Liebold, B., H. R. Graack & T. Pohl. 2006. Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. Proteomics **6:** 4688-4703.

86. Shevchenko, A., M. Wilm, O. Vorm*, et al.* 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem **68:** 850-858.

87. Yates, J. r., A. McCormack, A. Link*, et al.* 1996. Future prospects for the analysis of complex biological systems using micro-column liquid chromatography-electrospray tandem mass spectrometry. Analyst **121:** 65R-76R.

88. Washburn, M., D. Wolters & J. R. Yates, 3rd. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol **19:** 242-247.

89. Mann, M. & M. Wilm. 1995. Electrospray mass spectrometry for protein characterization. Trends Biochem Sci **20:** 219-224.

90. Wilm, M. & M. Mann. 1996. Analytical properties of the nanoelectrospray ion source. Anal Chem **68:** 1-8.

91. Karas, M., D. Bachmann, U. Bahr*, et al.* 1987. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. Int J Mass Spectrom Ion Process **78:** 53-68.

92. Karas, M. & F. Hillenkamp. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem **60:** 2299-2301.

93. Jonscher, K. R. & J. R. Yates, 3rd. 1997. The quadrupole ion trap mass spectrometer--a small solution to a big challenge. Anal Biochem **244:** 1-15.

94. Schwartz, J. C., M. W. Senko & J. E. Syka. 2002. A two-dimensional quadrupole ion trap mass spectrometer. J Am Soc Mass Spectrom **13:** 659-669.

95. Williams, E. R. 1998. Tandem FTMS of large biomolecules. Anal Chem **70:** 179A-185A.

96. Guilhaus, M., D. Selby & V. Mlynski. 2000. Orthogonal acceleration time-of-flight mass spectrometry. Mass Spectrom Rev **19:** 65-107.

97. Eng, J., A. McCormack & J. R. Yates, III. 1994. 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.

98. Perkins, D. N., D. J. Pappin, D. M. Creasy*, et al.* 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20:** 3551-3567.

99. Craig, R. & R. C. Beavis. 2003. A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid Commun Mass Spectrom **17:** 2310-2316.

100. Craig, R., J. P. Cortens & R. C. Beavis. 2004. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res **3:** 1234-1242.

101. Kapp, E. A., F. Schutz, L. M. Connolly*, et al.* 2005. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. Proteomics **5:** 3475-3490.

102. Resing, K. A., K. Meyer-Arendt, A. M. Mendoza*, et al.* 2004. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. Anal Chem **76:** 3556-3568.

103. Breci, L. A., D. L. Tabb, J. R. Yates, 3rd*, et al.* 2003. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. Anal Chem **75:** 1963-1971.

104. Altschul, S., W. Gish, W. Miller*, et al.* 1990. Basic local alignment search tool. J Mol Biol **215:** 403-410.

105. Emanuelsson, O., H. Nielsen, S. Brunak*, et al.* 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300:** 1005-1016.

106. Eisenhaber, B., G. Schneider, M. Wildpaner*, et al.* 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for A*spergillus nidulans, Candida albicans, Neurospora crassa, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. J Mol Biol **337:** 243-253.

107. Kronegg, J. & D. Buloz. 1999. Detection/prediction of GPI cleavage site (GPI-anchor) in a protein (DGPI), Retrieved 31 July 2007 from http://129.194.185.165/dgpi/

108. Krogh, A., B. Larsson, G. von Heijne*, et al.* 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol **305:** 567-580.

109. Sonnhammer, E. L., G. von Heijne & A. Krogh. 1998. A hidden markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol **6:** 175-182.

110. Ashburner, M., C. A. Ball, J. A. Blake*, et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25:** 25-29.

111. Ong, S. E., L. J. Foster & M. Mann. 2003. Mass spectrometric-based approaches in quantitative proteomics. Methods **29:** 124-130.

112. Gygi, S. P., B. Rist, S. A. Gerber*, et al.* 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol **17:** 994-999.

113. Turecek, F. 2002. Mass spectrometry in coupling with affinity capture-release and isotope-coded affinity tags for quantitative protein analysis. J Mass Spectrom **37:** 1-14.

114. Karsan, A., I. Pollet, L. R. Yu*, et al.* 2005. Quantitative proteomic analysis of sokotrasterol sulfate-stimulated primary human endothelial cells. Mol Cell Proteomics **4:** 191-204.

115. Lahm, H. W. & H. Langen. 2000. Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis **21:** 2105-2114.

116. Ong, S. E., B. Blagoev, I. Kratchmarova*, et al.* 2002. 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.

117. Zieske, L. R. 2006. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. J Exp Bot **57:** 1501-1508.

118. DeSouza, L., G. Diehl, M. J. Rodrigues*, et al.* 2005. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. J Proteome Res **4:** 377-386.

119. Oda, Y., K. Huang, F. R. Cross*, et al.* 1999. Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A **96:** 6591-6596.

120. Washburn, M. P., R. Ulaszek, C. Deciu*, et al.* 2002. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. Anal Chem **74:** 1650-1657.

121. Nelson, C. J., E. L. Huttlin, A. D. Hegeman*, et al.* 2007. Implications of 15N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. Proteomics **7:** 1279-1292.

122. Lilley, K. S. & D. B. Friedman. 2004. All about DIGE: quantification technology for differential-display 2D-gel proteomics. Expert Rev Proteomics **1:** 401-409.

123. Cooper, B., W. M. Garrett & K. B. Campbell. 2006. Shotgun identification of proteins from uredospores of the bean rust *Uromyces appendiculatus*. Proteomics **6:** 2477-2484.

124. Ebanks, R. O., K. Chisholm, S. McKinnon*, et al.* 2006. Proteomic analysis of *Candida albicans* yeast and hyphal cell wall and associated proteins. Proteomics **6:** 2147-2156.

125. Thomas, D. P., A. Viudes, C. Monteagudo*, et al.* 2006. A proteomic-based approach for the identification of *Candida albicans* protein components present in a subunit vaccine that protects against disseminated candidiasis. Proteomics **6:** 6033-6041.

126. Yajima, W. & N. N. Kav. 2006. The proteome of the phytopathogenic fungus *Sclerotinia sclerotiorum*. Proteomics **6:** 5995-6007.

127. de Godoy, L. M., J. V. Olsen, G. A. de Souza*, et al.* 2006. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. Genome Biol **7:** R50.01-R50.15.

128. Bae, W. & X. Chen. 2004. Proteomic study for the cellular responses to Cd2+ in *Schizosaccharomyces pombe* through amino acid-coded mass tagging and liquid chromatography tandem mass spectrometry. Mol Cell Proteomics **3:** 596-607.

129. Nouwens, A. S., S. J. Cordwell, M. R. Larsen*, et al.* 2000. Complementing genomics with proteomics: the membrane subproteome of *Pseudomonas aeruginosa* PAO1. Electrophoresis **21:** 3797-3809.

130. Sonnenberg, M. G. & J. T. Belisle. 1997. Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. Infect Immun **65:** 4515-4524.

131. Bhaskar, S., S. P. Khanna & R. Mukherjee. 2000. Isolation, purification and immunological characterization of novel low molecular weight protein antigen CFP 6 from culture filtrate of *M. tuberculosis*. Vaccine **18:** 2856-2866.

132. Jungblut, P. R., U. E. Schaible, H. J. Mollenkopf*, et al.* 1999. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. Mol Microbiol **33:** 1103-1117.

133. Jefferies, J. R., A. M. Campbell, A. J. van Rossum*, et al.* 2001. Proteomic analysis of *Fasciola hepatica* excretory-secretory products. Proteomics **1:** 1128-1132.

134. Carucci, D. J. 2002. Technologies for the study of gene and protein expression in *Plasmodium*. Philos Trans R Soc Lond B Biol Sci **357:** 13-16.

135. Carucci, D. J., J. R. Yates, 3rd & L. Florens. 2002. Exploring the proteome of *Plasmodium*. Int J Parasitol **32:** 1539-1542.

136. de Groot, P. W., A. D. de Boer, J. Cunningham*, et al.* 2004. Proteomic analysis of *Candida albicans* cell walls reveals covalently bound carbohydrate-active enzymes and adhesins. Eukaryot Cell **3:** 955-965.

137. Yin, Q. Y., P. W. de Groot, H. L. Dekker*, et al.* 2005. Comprehensive proteomic analysis of *Saccharomyces cerevisiae* cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. J Biol Chem **280:** 20894-20901.

138. Eng, J., A. L. McCormack & J. R. Yates, III. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Amer Soc Mass Spectrom **5:** 976-989

139. Aebersold, R. & M. Mann. 2003. Mass spectrometry-based proteomics. Nature **422:**198-207.

140. Washburn, M. P., D. Wolters & J. R. Yates, III. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol **19:** 242-247.

141. Wolters, D. A., M. P. Washburn & J. R. Yates, III. 2001. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem **73:** 5683-5690.

142. Yates, J. R., III, J. K. Eng, A. L. McCormack*, et al.* 1995. A method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem **67:** 1426-1436.

143. Perkins, D. N., D. J. Pappin, D. M. Creasy*, et al.* 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20:** 3551-3567.

144. Kapp, E. A., F. Schutz, L. M. Connolly*, et al.* 2005. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. Proteomics **5:** 3475-3490.

145. Skipp, P., J. Robinson, C. D. O'Connor*, et al.* 2005. Shotgun proteomic analysis of Chlamydia trachomatis. Proteomics **5:** 1558-1573.

146. Craig, R. & R. C. Beavis. 2003. A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid Commun Mass Spectrom **17:** 2310-2316.

147. Craig, R. & R. C. Beavis. 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics **20:** 1466-1467.

148. Craig, R., J. P. Cortens & R. C. Beavis. 2004. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res **3:** 1234-1242.

149. Geer, L. Y., S. P. Markey, J. A. Kowalak*, et al.* 2004. Open mass spectrometry search algorithm. J Proteome Res **3:** 958-964.

150. Breci, L., E. Hattrup, M. Keeler*, et al.* 2005. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. Proteomics **5:** 2018-2028.

151. Benzinger, A., N. Muster, H. B. Koch*, et al.* 2005. Targeted proteomic analysis of 14-3-3 sigma, a p53 effector commonly silenced in cancer. Mol Cell Proteomics **4:** 785-795.

152. Durr, E., J. Yu, K. M. Krasinska*, et al.* 2004. Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. Nat Biotechnol **22:** 985-992.

153. Phillips, G. R., T. R. Anderson, L. Florens*, et al.* 2004. Actin-binding proteins in a postsynaptic preparation: Lasp-1 is a component of central nervous system synapses and dendritic spines. J Neurosci Res **78:** 38-48.

154. Washburn, M. P., R. Ulaszek, C. Deciu*, et al.* 2002. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. Anal Chem **74:** 1650-1657.

155. Swanson, S. K. & M. P. Washburn. 2005. The continuing evolution of shotgun proteomics. Drug Discov Today **10:** 719-725.

156. Taylor, G. K. & D. R. Goodlett. 2005. Rules governing protein identification by mass spectrometry. Rapid Commun Mass Spectrom **19:** 3420.

157. Wilkins, M. R., R. D. Appel, J. E. Van Eyk*, et al.* 2006. Guidelines for the next 10 years of proteomics. Proteomics **6:** 4-8.

158. Klammer, A. A. & M. J. MacCoss. 2006. Effects of modified digestion schemes on the identification of proteins from complex mixtures. J Proteome Res **5:** 695-700.

159. Keller, A., A. I. Nesvizhskii, E. Kolker*, et al.* 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem **74:** 5383-5392.

160. Moore, R. E., M. K. Young & T. D. Lee. 2002. Qscore: an algorithm for evaluating SEQUEST database search results. J Am Soc Mass Spectrom **13:** 378-386.

161. Weatherly, D. B., J. A. Astwood, 3rd, T. A. Minning*, et al.* 2005. A Heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. Mol Cell Proteomics **4:** 762-772.

162. Qian, W. J., T. Liu, M. E. Monroe*, et al.* 2005. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res **4:** 53-62.

163. Link, A. J., J. Eng, D. M. Schieltz*, et al.* 1999. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol **17:** 676-682.

164. Han, D. K., J. Eng, H. Zhou*, et al.* 2001. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat Biotechnol **19:** 946-951.

165. Anderson, D. C., W. Li, D. G. Payan*, et al.* 2003. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. J Proteome Res **2:** 137-146.

166. Smith, R. D., G. A. Anderson, M. S. Lipton*, et al.* 2002. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. Proteomics **2:** 513-523.

167. Resing, K. A., K. Meyer-Arendt, A. M. Mendoza*, et al.* 2004. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. Anal Chem

168. Pedrioli, P. G., J. K. Eng, R. Hubley*, et al.* 2004. A common open representation of mass spectrometry data and its application to proteomics research. Nat BiotechnolNov.

169. Tabb, D. L., W. H. McDonald & J. R. Yates, III, inventors. 2002. Date of application: Jan-Feb.

170. Frigeri, L. G., T. R. Radabaugh, P. A. Haynes*, et al.* 2006. Identification of proteins from a cell wall fraction of the diatom Thalassiosira pseudonana: insights into silica structure formation. Mol Cell Proteomics **5:** 182-193.

171. Medina, M. L., P. A. Haynes, L. Breci*, et al.* 2005. Analysis of secreted proteins from Aspergillus flavus. Proteomics **41:** 327-335.

172. Orsborn, K. I., L. F. Shubitz, T. Peng*, et al.* 2006. Protein expression profiling of Coccidioides posadasii by two-dimensional differential in-gel electrophoresis and

evaluation of a newly recognized peroxisomal matrix protein as a recombinant vaccine candidate. Infect Immun **74:** 1865-1872.

173. Haynes, P. A., S. J. Miller, T. R. Radabaugh*, et al.* 2006. The Wildcat Toolbox: A set of perl script utilities for use in peptide mass spectral database searching and proteomics experiments. Journal of Biomolecular Techniques **17:** 97-102.

174. Peng, J., J. E. Elias, C. C. Thoreen*, et al.* 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res **2:** 43-50.

175. Hart, S. & S. Gaskell. 2005. LC-tandem MS in proteome characterization. Trends Analyt Chem **24:** 566-575.

176. Converse, J. L. & A. R. Besemer. 1959. Nutrition of the Parasitic Phase of *Coccidioides immitis* in a Chemically Defined Liquid Medium. J Bacteriol **78:** 231-239.

177. Breci, L., E. Hattrup, M. Keeler*, et al.* 2005. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. Proteomics **5:** 2018-2028.

178. Yates, J. r., J. Eng & A. McCormack. 1995. Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem **67:** 3202-3210.

179. Tabb, D., W. McDonald & J. r. Yates. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res **1:** 21-26.

180. Rohrbough, J. G., L. Breci, N. Merchant*, et al.* 2006. Verification of single-Peptide protein identifications by the application of complementary database search algorithms. J Biomol Tech **17:** 327-332.

181. Shen, H. D., W. L. Lin, M. F. Tam*, et al.* 2001. Identification of vacuolar serine proteinase as a major allergen of *Aspergillus fumigatus* by immunoblotting and N-terminal amino acid sequence analysis. Clin Exp Allergy **31:** 295-302.

182. Shen, H. D., H. Chou, M. F. Tam*, et al.* 2003. Molecular and immunological characterization of Pen ch 18, the vacuolar serine protease major allergen of *Penicillium chrysogenum*. Allergy **58:** 993-1002.

183. Monod, M., B. Lechenne, O. Jousson*, et al.* 2005. Aminopeptidases and dipeptidyl-peptidases secreted by the dermatophyte *Trichophyton rubrum*. Microbiology **151:** 145-155.

184. Hostetter, M. K. 2000. RGD-mediated adhesion in fungal pathogens of humans, plants and insects. Curr Opin Microbiol **3:** 344-348.

185. Jousson, O., B. Lechenne, O. Bontems*, et al.* 2004. Secreted subtilisin gene family in *Trichophyton rubrum*. Gene **339:** 79-88.

186. Elkins, T., K. Zinn, L. McAllister*, et al.* 1990. Genetic analysis of a *Drosophila* neural cell adhesion molecule: interaction of fasciclin I and abelson tyrosine kinase mutations. Cell **60:** 565-575.

187. Beauvais, A., M. Monod, J. P. Debeaupuis*, et al.* 1997. Biochemical and antigenic characterization of a new dipeptidyl-peptidase isolated from *Aspergillus fumigatus*. J Biol Chem **272:** 6238-6244.

188. Cole, G. T. & C. Y. Hung. 2001. The parasitic cell wall of *Coccidioides immitis*. Med Mycol **39 Suppl 1:** 31-40.

189. MacCoss, M. J., C. C. Wu, H. Liu*, et al.* 2003. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. Anal Chem **75:** 6912-6921.

190. Kolkman, A., P. Daran-Lapujade, A. Fullaondo*, et al.* 2006. Proteome analysis of yeast response to various nutrient limitations. Mol Syst Biol **2:** 2006 0026.

191. Daher, B. S., E. J. Venancio, S. M. de Freitas*, et al.* 2005. The highly expressed yeast gene *pby20* from *Paracoccidioides brasiliensis* encodes a flavodoxin-like protein. Fungal Genet Biol **42:** 434-443.

192. Kellner, E. M., M. A. Mandel, J. N. Galgiani*, et al.* 2005. Serial analysis of gene expression in *Coccidioides posadasii*. Presented at the 23rd Fungal Genetics Conference at Asilomar

193. Gygi, S. P., Y. Rochon, B. R. Franza*, et al.* 1999. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol **19:** 1720-1730.

194. de Nobel, H., L. Lawrie, S. Brul*, et al.* 2001. Parallel and comparative analysis of the proteome and transcriptome of sorbic acid-stressed *Saccharomyces cerevisiae*. Yeast **18:** 1413-1428.

195. Tian, Q., S. B. Stepaniants, M. Mao*, et al.* 2004. Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Mol Cell Proteomics **3:** 960-969.

196. Wysocki, V. H., K. A. Resing, Q. Zhang*, et al.* 2005. Mass spectrometry of peptides and proteins. Methods **35:** 211-222.

197. Listgarten, J. & A. Emili. 2005. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics **4:** 419-434.

198. Zimmer, J. S., M. E. Monroe, W. J. Qian*, et al.* 2006. Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom Rev **25:** 450-482.

199. Eyles, J. E., B. Unal, M. G. Hartley*, et al.* 2007. Immunodominant *Francisella tularensis* antigens identified using proteome microarray. Proteomics **7:** 2172-2183.

200. Wutzke, K. D. & I. Oetjens. 2005. 13C- and 15N-incorporation of doubly stable isotope labelled *Lactobacillus johnsonii* in humans. Eur J Clin Nutr **59:** 1167-1172.

201. Nie, L., G. Wu & W. Zhang. 2006. Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. Biochem Biophys Res Commun **339:** 603-610.

202. Beyer, A., J. Hollunder, H. P. Nasheuer*, et al.* 2004. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. Mol Cell Proteomics **3:** 1083-1092.

203. Brockmann, R., A. Beyer, J. J. Heinisch*, et al.* 2007. Posttranscriptional expression regulation: what determines translation rates? PLoS Comput Biol **3:** e57.

204. Irniger, S. & G. H. Braus. 2003. Controlling transcription by destruction: the regulation of yeast Gcn4p stability. Curr Genet **44:** 8-18.