

V. YILDIRIM

**FIRST REFERENCE MAP FOR PHANEROCHAETE CHRYSOSPORIUM  
PROTEOME**

**VOLKAN YILDIRIM**

**JANUARY 2006**

METU

**FIRST REFERENCE MAP FOR *PHANEROCHAETE CHRYSOSPORIUM*  
PROTEOME**

**A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES  
OF  
MIDDLE EAST TECHNICAL UNIVERSITY**

**BY**

**VOLKAN YILDIRIM**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF MASTER OF SCIENCE  
IN  
BIOLOGY**

**JANUARY 2006**

**Approval of the Graduate School of Natural and Applied Sciences**

---

**Prof. Dr. Canan Özgen**  
**Director**

**I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.**

---

**Prof. Dr. Semra Kocabıyık**  
**Head of Department**

**This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science**

---

**Prof. Dr. Gülay Özcengiz**  
**Supervisor**

**Examining Committee Members**

**Prof. Dr. Vasıf Hasırcı** (METU,BIO) \_\_\_\_\_

**Prof. Dr. Gülay Özcengiz** (METU,BIO) \_\_\_\_\_

**Prof. Dr. Cihan Öner** (Hacettepe Univ.,BIO) \_\_\_\_\_

**Prof. Dr. Mesude İşcan** (METU,BIO) \_\_\_\_\_

**Prof. Dr. Hüseyin Avni Öktem** (METU,BIO) \_\_\_\_\_

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Last name : Volkan Yıldırım

Signature :

## **ABSTRACT**

### **FIRST REFERENCE MAP FOR *PHANEROCHAETE CHRYSOSPORIUM* PROTEOME**

Yıldırım, Volkan

M.Sc., Department of Biology

Supervisor: Prof. Dr. Gülay Özcengiz

January 2006, 96 pages

In this study, the soluble protein fraction of *P. chrysosporium* grown under standard conditions was analyzed by using 2D-PAGE approach and a 2-D reference map was constructed. 910 spots could be separated and detected on Coomassie-stained 2-D gels by the help of Delta2D image analysis software. 720 spots could be cut from the master gel and were subjected to MALDI-TOF MS analysis followed by MASCOT search. A total of 517 spots out of 720 were assigned to specific accession numbers from the *P. chrysosporium* genome database. Further analysis of the data revealed 314 different gene products (distinct ORFs).

The theoretical  $pI$  and MW values were plotted against the experimental migration distances. Results indicated the existence of 124 protein spots whose horizontal migration differed significantly from the expected migration according to the calculated  $pI$  values and 52 spots with an apparent molecular weight that is significantly different from their theoretical molecular weight. While protein modification could be predicted by these analyses, the main support was the presence of multiple spots of the same

gene product. As much as 118 ORFs yielded multiple spots on the master gel, corresponding to 37.5% of the all distinct ORFs identified in this work.

The relative abundance of each of the 517 identified polypeptides was calculated in terms of spot intensity. The majority of the most abundant proteins were found to be housekeeping ones. When the relative distribution of the proteins into four main functional categories was taken into consideration, "Metabolism" appeared the most important category with a share of 50.6% among identified proteins. However, among the functional classes, "Posttranslational modifications, protein turnover, chaperones" which is listed under the main category "Cellular Processing and Signalling" was represented by the highest number (104) of the identified proteins. Only 6 of the proteins listed in this study were assigned to hypothetical proteins.

Out of the 314 identified gene products shown in *P. chryso sporium*, 29 were predicted to have a signal peptide sequence according to the SignalP algorithm. By making a WoLF PSORT search, subcellular localization of the proteins was predicted. Accordingly, 147 of the proteins were predicted to be located in cytoplasm. The phosphorylated proteins of *P. chryso sporium* were detected by ProQ phosphoprotein staining of the 2-D gel. 380 out of 910 distinct protein spots (40%) were found to be phosphorylated in exponentially growing cells of *P. chryso sporium*. Of these spots, 96 could be matched to the identified proteins.

Key words: Proteome, *Phanerochaete chryso sporium*, Reference proteome map, 2-DE map, MALDI TOF MS, Peptide mass fingerprint.

## ÖZ

### ***PHANEROCHAETE CHRYSOSPORIUM* PROTEOMUNUN İLK REFERANS HARİTASI**

Yıldırım, Volkan

Yüksek Lisans Biyoloji Bölümü

Tez Yöneticisi: Prof. Dr. Gülay Özcengiz

Ocak 2006, 96 sayfa

Bu çalışmada, standart koşullar altında büyütülen *P. chrysosporium* 'un çözünür protein fraksiyonu 2D-PAGE yaklaşımı kullanılarak analiz edildi ve 2-D referans haritası oluşturuldu. Coomassie ile boyanmış 2-D jellerinden Delta2D görüntü analiz yazılımı kullanılarak 910 protein spotu belirlenmiştir. Bu spotların 720 tanesi jelden kesilebilmiş ve herbiri MALDI-TOF MS analizine tabi tutulmuştur. Daha sonra MASCOT taraması yapılarak proteinlerin kimliği belirlenmiştir. 720 protein içerisinde 517 adet protein, *P. chrysosporium* genom veritabanında özgün giriş numarası almış proteinlere karşılık gelmiş ve böylelikle tanımlanmıştır. Bulguların daha ileri analizi, bu proteinlerin toplam 314 farklı gen ürününü (farklı ORF 'leri) temsil ettiğini göstermiştir.

Proteinlerin teorik  $pI$  ve MW değerleri deneysel olarak elde edilen jelde yürüme verilerine karşı incelendiğinde, çok sayıda protein spotu için yatay ve dikey yürüme mesafeleri hesaplanan  $pI$  ve MW değerlerine göre beklenenlerden önemli farklılıklar göstermiştir. Bu analizler protein modifikasyonuna işaret etmiş olmakla birlikte, protein modifikasyonlarına dair ana kanıt aynı gen ürünü için birden çok sayıda protein spotunun varlığı olmuştur. 118 farklı ORF 'nin birden çok sayıda spot verdiği bulunmuştur; dolayısıyla bu tip ORF 'ler tanımlanan tüm farklı ORF 'lerin % 37.5 'ini

oluşturmaktadır. Bunlar içerisinde 38 protein elektrik yükü, 10 protein molekül ağırlığı ve 70 protein hem elektrik yükü, hem de molekül ağırlığı temelinde modifikasyonlar göstermiştir.

Tanımlanan 517 polipeptidin herbirinin göreceli çokluğu spot koyuluğu olarak hesaplanmıştır. En çok bulunan proteinlerin büyük çoğunluğunun "ev idaresinde" görevli proteinler olduğu belirlenmiştir. Proteinlerin 4 ana fonksiyonel kategoriye göreceli dağılımları incelendiğinde, "Metabolizma" 'nın tanımlanan tüm proteinlerin % 50.6 'sını içererek en önemli kategori olduğu görülmüştür. Diğer yandan, fonksiyonel alt sınıflar içerisinde "posttranslasyonel modifikasyonlar, protein çevrimi, şaperonlar" 'ı içeren ve "Hücreyel İşlemler ve Sinyaller" ana kategorisi altında listelenen sınıf en yüksek sayıda tanımlanmış proteinle (104) temsil edilmiştir. Bu çalışmada listelenen proteinler içerisinde sadece 6 'sı hipotetik protein olarak tanımlanmıştır.

SignalP algoritması kullanılarak yapılan biyoinformatik analiz, *P. chryso sporium* hücrelerinde tanımlanan 314 gen ürününden 29 'unun sinyal peptid dizilimine sahip olduğunu öngörmüştür. WoLF PSORT yazılımı kullanılarak proteinlerin hücre içi lokalizasyonları tahmin edilmiştir. Buna göre proteinlerin 147 'si sitoplazmada bulunmaktadır. *P. chryso sporium* 'un fosforile olmuş proteinlerinin teşhisi için 2-D jelle ProQ Fosfoprotein boyası uygulanmıştır. Eksponansiyel olarak üreyen *P. chryso sporium* hücrelerinde gösterilen toplam 910 farklı protein spotu içerisinde 380 'inin fosforile olduğu belirlenmiştir. Bu spotlardan 96 'sı daha önce tanımladığımız proteinlere karşılık gelmiştir.

Anahtar kelimeler: Proteome, *Phanerochaete chryso sporium*, Referans proteome haritası, 2-DE haritası, MALDI-TOF MS, Peptid kütle parmak izi.

***To My Parents***

## ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to my supervisor, Prof. Dr. Gülay Özcengiz for her guidance, supervision, encouragement and patience given during this long process.

I would also like to thank to Asst. Prof. Dr. Servet Özcan for his kindness, guidance, advices and for sharing all his work experiences with me.

I would like to express my great indebtedness to the people in the University of Greifswald Department of Microbiology, especially Susanne Wolff, Andreas Otto, Katrine Gronau, Knut Büttner, Dr. Jörg Bernhardt, Dr. Dirk Albrecht and Prof. Dr. Michael Hecker for their hospitality, kindness, endless help and making available all the equipment for MALDI-TOF MS analysis in this work. A special thank goes to Dr. Dörte Becher for her kindness, unique hospitality and invaluable help both at work and daily life during my stay in Germany.

I would also like to thank Levent Kaya for his endless patience, support, helps whenever I need, great friendship and advices along the road. Soon we will be parted but I am sure we will have a chance to walk together on the same road again.

Very special thanks to my lab mates Dr. Ebru İnce Yılmaz, Çiğdem Yağcıoğlu, Bilgin Taşkın, Ayşe Koca, Aslıhan Kurt, Sezer Okay, Araz Zeyniyev, Erkam Gündoğdu, Orhan Özcan, Emrah Altındış, Burcu Tefon and İsmail Öğülür for their great friendship, invaluable understanding and being with me whenever I needed.

I also would like to thank to Dr. Hasan Oğul of Başkent University for his invaluable help during the certain bioinformatic analysis of our data. My special thank also go to Prof. Dr. Gerhard Wilhelm Weber for his kind and constant interest in this study.

The last but not least, I would like to extend my gratitude to my precious parents, Hatice and Ürfettin YILDIRIM for their endless love, kind understanding, patience and endless support during my studies.

## TABLE OF CONTENTS

PLAGIARISM.....	III
ABSTRACT .....	IV
ÖZ.....	VI
DEDICATION.....	VIII
ACKNOWLEDGEMENTS .....	IX
TABLE OF CONTENTS .....	X
LIST OF FIGURES .....	XIII
1. INTRODUCTION .....	1
1.1. Proteome .....	1
1.2. Why Proteomics? .....	2
1.3. Annotation of Genome .....	3
1.4. Why 2-D PAGE? .....	3
1.5. Steps in Proteome Work.....	4
1.5.1. Isoelectric Focusing and Immobilized pH Gradient.....	4
1.5.2. Visualisation of Protein Spots and Image Analysis .....	6
1.5.3. Mass Spectrometry Analysis.....	7
1.5.4. Proteome and Bioinformatics .....	8
1.6. Recent Reference Proteome Map Studies .....	9
1.6.1. Bacterial Proteome.....	10
1.6.2. Proteome Studies for Eukaryotic Microbes .....	12
1.6.3. Human Proteome.....	14
1.6.4. Plant Proteome .....	15
1.6.5. Posttranslational Modifications (PTM) .....	16
1.7. Stress-Related Proteome Studies .....	20
1.8. Phanerochaete chrysosporium .....	23
1.9. Aim of the Present Study.....	25

2. MATERIALS AND METHODS .....	26
2.1. Microorganism and Its Maintenance.....	26
2.2. Growth Conditions and Biomass Preparation.....	26
2.3. Protein Extraction .....	27
2.4. Protein Estimation.....	27
2.5. Proteome Study .....	28
2.1.1. Isoelectric Focusing.....	28
2.2.1. SDS-PAGE .....	28
2.3.1. Phosphoprotein Detection.....	29
2.4.1. Evaluation of 2-DE Data .....	29
2.5.1. Sample Preparation and Analysis by Mass Spectroscopy .....	29
2.6.1. Protein Identification .....	31
2.7.1. Bioinformatic analyses.....	31
3. RESULTS AND DISCUSSION .....	33
3.1. <i>P. chrysosporium</i> Theoretical 2-DE Map .....	33
3.2. Master Gel.....	34
3.3. Identification of <i>P. chrysosporium</i> Proteins by MALDI-TOF MS.....	35
3.4. Experimental versus Theoretical Data.....	41
3.5. Proteins Appearing in Multiple Spots.....	43
3.6. Functional Assignments .....	45
3.7. Relative Quantitative Distribution of the Proteins into Functional Categories.....	63
3.8. Prediction of Subcellular Localizations and Signal Peptides.....	70
3.9. Phosphorylated Proteins .....	71
CONCLUSION .....	79
REFERENCES.....	81

## LIST OF TABLES

<b>Table 1.</b> Overview of the proteins which appear in multiple spots .....	44
<b>Table 2.</b> Predicted functions of the proteins annotated in the master gel...	46
<b>Table 3.</b> KOG classification and the number of gene models at the JGI website .....	65
<b>Table 4.</b> The 50 most abundant proteins in <i>P. chrysosporium</i> .....	67
<b>Table 5.</b> Distribution of the 517 proteins into four main functional categories .....	70
<b>Table 6.</b> List of phosphorylated spots on the ProQ stained gel .....	76

## LIST OF FIGURES

<b>Figure 1.</b> Predicted 2-D gel separation of the <i>P. chrysosporium</i> proteome (theoretical 2-DE map).....	33
<b>Figure 2.</b> A general view of 2-DE of <i>P. chrysosporium</i> soluble proteins. ....	35
<b>Figure 3.</b> MALDI-TOF MS spectra of the peptides (PMF) obtained from one randomly chosen protein spot.....	36
<b>Figure 4.</b> Representative MASCOT search output.....	37
<b>Figure 5.</b> Representative output of <i>P. chrysosporium</i> database search .....	37
<b>Figure 6 A and B.</b> A reference 2-D map (master gel) of <i>P. chrysosporium</i> .	40
<b>Figure 7.</b> Comparison of the theoretical and experimental molecular weights of the identified proteins in the <i>P. chrysosporium</i> master gel. ....	42
<b>Figure 8.</b> Comparison of the theoretical and experimental pI of the identified proteins in the <i>P. chrysosporium</i> master gel. ....	43
<b>Figure 9.</b> Relative quantitative distribution of the 517 identified proteins to various functional classes of cellular physiology. ....	69
<b>Figure 10.</b> ProQ stained 2-DE of <i>P. chrysosporium</i> soluble proteins.....	73
<b>Figure 11.</b> Spot boundary detection of phosphoproteins . ....	74
<b>Figure 12.</b> The image of the ProQ stained and Coomassie stained spots on the same gel as created by the Delta2D software. ....	75

## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1. Proteome**

Proteome indicates "the entire PROTEin complement expressed by a genOME" (Wilkins *et al.*, 1996; Pennigton *et al.*, 1997). Multiple steps required in proteome analysis include the stringent control of sample preparation, two-dimensional polyacrylamide gel electrophoresis (2-D PAGE or 2-DE), image detection and analysis, spot identification and database search. As to its application, many different arrays of study are now grouped under the rubric of "proteomics". These comprise protein-protein interaction, protein modifications, protein function, protein localization and many others. With another definition, "proteome" is a term used to refer to all different proteins occurring in an organism in space and time. "Different proteins" includes not only the primary polypeptides created by the individual genes, but also all the co- and post-translational modifications of a protein. "Space" considers the fact that to a certain extent, different proteins occur in different cell types and different subcellular compartments. "Time" considers the differences in protein composition and concentrations which may exist for distinct cell types at different stages of embryonic development, at different stages of aging, or during any functional process taking place in the cell. Whereas the terms proteome and genome suggest a certain homology between these two cell components, the definition given above indicates that the proteome is much more complicated than the genome. The terminus genome is "used to refer to all the genes carried by a single gamete", and in

principle, one could say “carried by any cell type of an organism”. In contrast, the proteome differs from one cell type to another qualitatively as well as quantitatively, and moreover, depends on the course of time and provides an invaluable tool for the analysis of the gene expression (Gauss *et al.*, 1999).

## **1.2. Why Proteomics?**

As a term, “proteomics” is the large-scale screening of the proteins of a cell, organism or biological fluids. Many types of information can not be obtained from the study of genes alone. Genomic analyses can provide very limited information with respect to describing how microorganisms adapt to a constantly changing environment. While the presence of certain genes characterized by their sequence identity to a database homology suggests an organism may have the ability to adapt to certain ecological niches and to utilize various substrates, the knowledge of protein expression under these conditions is essential for fully understanding how the organism responds to a given challenge (Cordwell *et al.*, 2001).

In recent years, DNA microarray technology and serial analysis of gene expression have become increasingly popular for the analysis of mRNA expression. However, the analysis of mRNA is not a direct reflection of the protein content in the cell. Gygi *et al.* (1999) has shown that there is a poor correlation between mRNA and protein expression level. The first reason is that mRNA is subject to postranscriptional control in the form of alternative splicing, polyadenylation and mRNA editing. Secondly, in eukaryota, mRNA then can be subjected to regulation at the level of translation. After being formed, proteins are also subjected to posttranscriptional modification, proteolysis and compartmentalization. Hence, the average number of proteins formed per gene is predicted to be one or two in bacteria, three in yeast and three or more for humans (Wilkins *et al.*, 1996).

To reveal the stress related proteomes, such as starvation physiology of different organisms, several proteome projects have been accomplished (Voigt *et al.*, 2006; Yin *et al.*, 2004). For instance, a proteomic approach was used recently to identify genes related with the aging process of human by de Magalhaes and Toussaint (2004). Different tissues of the same organism have the same genome, but different proteomes. Thus, proteomic approach is also useful to define tissue-specific genes. The most significant attempt for investigation of the proteome of different tissues has been initiated by the Human Proteome Project (HUPO). HUPO aims to identify tissue and organ specific proteomes ([www.hupo.org](http://www.hupo.org)).

### **1.3. Annotation of Genome**

Annotation is the allocation of functions to individual genes in genome sequences. Identification of the total number of genes in a given genome is one of the first applications of the proteomics. This functional annotation of a genome is necessary since it is difficult to predict genes accurately from genomic data (Eisenberg *et al.*, 2000). Therefore, genomic information will have to be integrated with data obtained from protein studies in order to confirm the existence of a particular gene.

### **1.4. Why 2-D PAGE?**

To date, proteomics have relied heavily on 2-D PAGE. Though frequently denoted as outdated (Gygi *et al.*, 2000, Hille *et al.*, 2001), alternative sorting procedures that have been introduced so far [e.g. gel free approach which includes consecutive use of different liquid chromatography techniques (2D-RP-HPLC)] could not establish a resolution comparable to that of the 2-DE method. The first dimension of 2-DE is isoelectric focusing. During IEF, proteins are separated in a pH gradient until they reach a stationary position where their net charge is zero. In the second dimension, the proteins

separated by isoelectric focusing are separated orthogonally by electrophoresis according to their molecular weight. Major advantages of this approach are the high information content obtained during the experiments and the high resolution capability (Fey and Larsen, 2001; Giorgianni *et al.*, 2003). The technique that had been introduced almost 30 years ago could be facilitated enormously with respect to both handling and reproducibility by the introduction of immobilized pH gradient in ready cast gels as a first dimension, which opened the path to proteomics even to scientists rather inexperienced in the field (Görg *et al.*, 1988).

## **1.5. Steps in Proteome Work**

### **1.5.1. Isoelectric Focusing and Immobilized pH Gradient**

Isoelectric focusing (IEF) is an electrophoretic method which separates proteins according to their isoelectric points ( $pI$ ) (O' Farrell, 1975). In nature, proteins are amphoteric molecules. They carry either negative, positive or zero net charge, depending on their surrounding pH level. The net charge of a protein is the sum of all the positive and negative charges of R groups (amino acid side chains) and amino- and carboxyl-termini. Isoelectric point ( $pI$ ) is the specific pH where the net charge of a protein is zero. Proteins are positively-charged with pH values below their  $pI$  and negatively-charged with pH values above their  $pI$ . If the net charge of a protein is plotted versus the pH of its environment, the resulting curve intersects the x-axis at its  $pI$ .

The presence of the pH gradient is essential to the IEF method. In the pH gradient, under the influence of an electrical field, a protein will migrate to a location in the gradient where its net charge is zero. Proteins with positive net charges will migrate toward the cathode, becoming progressively less positively charged as it moves through the pH gradient until it reaches its  $pI$ . On the contrary, proteins with negative net charges will migrate toward

anode, becoming less negatively charged as it moves through the pH gradient until it also reaches a zero net charge. When a protein diffuses away from its  $pI$ , it suddenly gains charge and moves back to its  $pI$  position. This is called "focusing" effect of IEF, which concentrates proteins at their  $pI$ s and permits proteins to be separated on the basis of very small charge differences.

IEF is performed under denaturing conditions. Complete solubilization and denaturation is achieved by using a mixture of detergent, urea and ampholines. Employing this mixture ensures that each protein is present in only one configuration, aggregation and intramolecular interaction.

The original method for the first dimension IEF is run in thin polyacrylamide gel rods in glass tubes with the aid of carrier ampholyte-generated pH gradients (O'Farrel, 1975; Klose, 1975). Carrier ampholytes are small, soluble amphoteric molecules with a high buffering capacity near their  $pI$ . Carrier ampholytes are composed of hundreds of individual polymeric molecules with their  $pI$ s spanning a specific pH range. Under an electrical field, carrier ampholytes with the highest  $pI$  (the most negative charge) move toward the anode and the carrier ampholytes with lowest  $pI$  (the most positive charge) migrate through the cathode. Those that are spanning between extremes line up according to their  $pI$ s providing a continuous pH gradient.

Carrier ampholyte-based technique has been used in a number of 2-DE. However, there are several limitations that have prevented its more widespread application. The most common problem is reproducibility of the first dimension (Berkelman and Stenstedt, 1998; Westermeier and Naven, 2002). This is suffered due to the batch to batch variation in manufacturing carrier ampholytes and the instability of pH gradient (i.e., general tendency to drift towards a cathode over time). Another major difficulty is the

mechanical instability of the soft polyacrylamide gel rods that results in stretching or breaking during handling.

To overcome the above mentioned problems, immobilized pH gradients (IPG) have been introduced for the generation of the pH gradient (Bjellqvist *et al.*, 1982). The pH gradient in IPG is prepared by covalently incorporating a gradient of carboxylic and tertiary amino groups as buffering agents at the time of casting. Due to the preformed buffering groups, the gradient can not drift and is not influenced by the sample composition. As commercial products, pre-manufactured gel strips and dedicated IPG instruments are available; a reproducible and more reliable data generation is also possible. Another advantage of the IPG over the traditional carrier ampholyte-mediated IEF is the amount of protein samples which could be increased more than 10 fold.

After the first dimension is completed by either carrier ampholyte-generated system or automated IPG systems, the second dimension is achieved on the base of protein size using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

### **1.5.2. Visualisation of Protein Spots and Image Analysis**

There are a number of visualizing methods for the protein spots resolved by the 2-DE. The highest sensitivity is achieved with radioactivity/fluorography detection. A number of non-radioactive methods (Silver staining, Coomassie Brilliant Blue staining) are applied, with big difference in sensitivity, linearity and dynamic range. Sensitive protein staining, like the ammoniacal silver staining, permits the detection of proteins at or below nanogram quantities. Hoving *et al.* (2000) reported the calculation down to 100 copies of a protein per cell with sensitive silver staining.

Following the staining, evaluation and comparison of the complex 2-DE patterns seems to be very difficult and nearly impossible. Therefore, the gel images have to be converted into digital data with a scanner, camera or more sophisticated gel image documentation systems. This high resolution image (preferably as much as 16 bit or greater) is then evaluated with the aid of the image analysis softwares (e.g. PDQuest, Biorad; Delta2D, Decodon; Phoretix, Nonlinear; Melanie, GeneBio).

### **1.5.3. Mass Spectrometry Analysis**

An array of approaches has been developed for addressing proteome analysis and protein identification. Methods of protein identification have included immunoblotting, peptide sequencing, amino acid composition and more recently the use of mass spectrometry (Celis *et al.*, 2000). One of the earliest methods used for protein identification was microsequencing by Edman chemistry to obtain N-terminal amino acid sequences. Since the invention, little has changed in this technique, but improvements in sequencing technology have increased the sensitivity and ease of it. One of the biggest hamper that has limited the use of the Edman sequencing in the past was N-terminal modification of proteins. In order to bypass this problem, mixed peptide sequencing is employed (Graves and Haystead, 2002).

Mass spectrometry (MS) enables protein structural information (i. e., peptide masses or amino acid sequences) to be obtained. Hence, this information can be employed to identify the protein by searching nucleotide and protein databases. Identification of proteins by mass spectrometry requires three major stages: (i) sample preparation, (ii) sample ionization and (iii) mass analysis. Extraction of a whole protein or its constituent peptide from the gel is commonly very difficult and mostly inefficient. On the other hand, if a protein is "in gel" digested with proteases (e.g. with trypsin) many of the peptides can be extracted from the gel (Andersen and Mann, 2000).

Therefore *in situ* digestion is more efficient at sample recovery than other common methods such as electroblotting. Additionally, the conversion of a protein into its parts provides more information than that can be obtained from a whole protein itself.

First requisite for biological sample to be analyzed by MS is that the molecule must be charged and dry (Graves and Hystead, 2002). Conversion of the samples into desolvated ions is the most commonly used application. Although there are different techniques available, the most preferable methods are Electrospray Ionization (ESI) and Matrix-assisted Laser Desorption/Ionization (MALDI). In both methods, peptides are converted into ions by the addition or loss of one or more protons. Both methods allow the formation of ions without major loss of sample integrity. This is important since mass information about proteins and peptides is obtained in their native state.

Mass analysis follows the conversion of protein or peptides into molecular ions. This is performed by the mass analyzer in a mass spectrometer which resolves the molecular ions on the bases of their mass and charge in a vacuum. Although there are few available techniques, a time-of-flight (TOF) is one of the simplest mass analyzer. It measures the  $m/z$  (mass to charge) ratio of an ion by determining the time required for it to transverse the length of a flight tube. From the MS scan, after subtracting the backgrounds, amino acid sequence information for the peptide is obtained with the aid of software. The information obtained is then used to search DNA and protein databases (SWISS-PROT, FASTA; MOWSE, ProFound, PepFrag, PepSea and so on).

#### **1.5.4. Proteome and Bioinformatics**

Over the last three decades, 2-DE has established itself as the *de facto* approach to separating proteins from cell and tissue samples (Anderson and

Anderson, 2002). Due to the sheer volume of data and its experimental geometric and expression uncertainties, quantitative analysis of these data with image processing and modelling has become an actively pursued research topic. The results of these analyses include accurate protein quantification, isoelectric point and relative molecular mass estimation, and the detection of differential expression between samples run on different gels. Systematic errors such as current leakage and regional expression inhomogeneities are corrected for, followed by each protein spot in the gel being segmented and modelled for quantification (Dowsey *et al.*, 2003).

Computer algorithms identify proteins based on peptide mass fingerprints (PMF) and fragmentation (MS/MS) information to search protein databases. Currently, there are different softwares (e.g. MASCOT, SEQUEST) using different algorithms to identify proteins from PMF and MS/MS information. Several papers have been published using both methodologies (Lewis *et al.*, 2000; Predic *et al.*, 2002; Kanamoto *et al.*, 2002).

Bioinformatic softwares give additional and very useful information by analysing the amino acid sequences of proteins. General approach is to establish a theoretical (*in silico*) proteome map by using the available genome sequence and comparing it with the experimental data (Rosen *et al.*, 2004; Gade *et al.*, 2005). By analyzing the amino acid sequence, bioinformatic softwares can predict several aspects of a protein such as subcellular localization (Dönnes and Höglund, 2004), existence of signal peptide (Nielsen *et al.*, 1997), function (Watson *et al.*, 2005) and pI/MW values (Bjellqvist *et al.*, 1993).

## **1.6. Recent Reference Proteome Map Studies**

For organisms with known genome sequences such as *E. coli* (Blattner *et al.*, 1997) and *S. cerevisiae* (Goffeau *et al.*, 1996), proteome projects can

provide 2-DE reference maps where all proteins are identified, with immediate and considerable details about protein expression. Proteomics is a commonly used method for physiological studies, the latter requiring 2-DE reference maps on which most of the main proteins (up- or down-regulated) are identified (Rosen *et al.*, 2004). Since early 2000s, the articles reporting on 2-DE proteome maps have been accumulating for prokaryotic and eukaryotic organisms to represent several species.

### **1.6.1. Bacterial Proteome**

The availability of entire genome sequences of bacteria since late 90's have facilitated the global analysis of gene expression under various conditions by proteome analysis. *E.coli* is one of the most extensively studied organism in bacteria and its 2-DE map was constructed by Pasquali *et al.*, in 1996 (updated in 1998), the 2-DE map having published in SWISS-2DPAGE format in the expasy website (<http://www.expasy.org/ch2d/publi/ecoli.html>). Recently, a 2-DE based reference map was also constructed for the plant pathogen *Agrobacterium tumefaciens*. Over 250 spots were identified and marked on the reference map (Rosen *et al.* 2004). Heim *et al.* (2003) constructed a 2-DE reference map for *Pseudomonas putida* proteome. Spots on the gels were identified by MALDI-TOF analysis, in conjunction with an in house database developed from the genome sequence of KT2440 strain of *P. putida* and about 200 spots were annotated to genome. The map was established to use for assessing the genomic response of KT2440 to iron limitation stress and to compare this response with that of the closely related facultative human pathogen *Pseudomonas aeruginosa* strain PAO1.

A proteome study was performed for *Ralstonia metallidurans* CH34, an aerobic gram-negative chemolithotrophic bacterium (Noël-Georis *et al.*, 2004). 2-DE analyses using various pH gradients were made and 224 different proteins were characterized from 352 protein spots. Although the proteome map is still not complete, one could appraise the importance of

proteomics for genome analyses through (i) the identification of previously undetected open reading frames, (ii) the identification of proteins not encoded by the already sequenced genome fragments, (iii) the characterization of protein-encoding genes spanning two different contigs, enabling their merging, and (iv) the precise delineation of the N-terminus of several proteins.

On the basis of its recently determined complete genome, a reference 2-DE gel for the soluble protein fraction of *Rhodospirellula baltica* was constructed (Gade *et al.*, 2005). Approximately 1000 protein spots were excised from a colloidal Coomassie-stained gel (pH 4–7), analyzed by MALDI-MS and identified by PMF. The non-redundant data set contained 626 distinct protein spots, corresponding to 558 different genes. Among the identified proteins, 146 contained a predicted signal peptide suggesting their translocation. Some proteins were detected in more than one spot on the gel, indicating post-translational modification.

A proteome reference map for cytosolic proteins of gram-positive bacterium *Lactococcus lactis* was constructed by Anglade *et al.* (2005). By using two pH gradients (pH 4-7, 4.5-5.5), more than 230 spots were identified by peptide mass fingerprints, corresponding to 25% of the predicted acid proteome. The study also indicated that the fermentative metabolism, which characterizes *L. lactis*, is associated with a high expression of glycolytic enzymes. 34 proteins were matched to open reading frames for which there was no assigned function. Another gram-positive bacterium *Streptococcus thermophilus* was subjected to proteome analysis by Arena *et al.* (2006). They have identified 244 spots on 2D gel (pH 4-7) which corresponded to 139 genes of *S. thermophilus* genome.

Proteome reference map construction for disease pathogens was also performed by many other research groups. As being the causative agent of

pneumonia, a 2-DE based proteome map of the bacterium *Mycoplasma pneumoniae* was constructed (Regula *et al.*, 2000). They have identified a total of 350 spots on silver stained gels. These spots were assigned to 224 ORFs among the total 688 ORFs of the *M. pneumoniae* genome. In another study, a 2-DE reference map and a database of *Bacillus anthracis* A16R proteome was constructed (Wang *et al.*, 2005). In total, 534 spots were processed and 406 spots representing 299 proteins were identified. A first proteomic analysis of the respiratory pathogen *Legionella pneumophila* was performed by Lebeau *et al.* (2005) Two-dimensional gel electrophoresis of total cell extracts was carried out. A total of 130 protein spots were identified by MALDI-TOF or by Q-TOF tandemMS, including the proteins correlated with virulence. 2-DE reference map of *Staphylococcus aureus* was established by Cordwell *et al.*, (2002). In this reference map, 377 proteins that corresponded to 266 distinct ORFs were identified. Amongst these identified proteins were 14 potential virulence factors.

### **1.6.2. Proteome Studies for Eukaryotic Microbes**

To date, several eukaryotic proteome studies have been published. As having the first completed genome among eukaryota, *Saccharomyces cerevisiae* was the first eukaryotic organism having a proteome reference map. The map was published in 1996 by Boucherie *et al.* which was updated in 1999 by Perrot *et al.* The web version of the map currently includes 410 spots corresponding to 258 genes (<http://www.ibgc.u-bordeaux2.fr/YPM/>). A recent work established the first proteome map of the fission yeast *Schizosaccharomyces pombe* (Sun *et al.*, 2005). *S. pombe* cytosolic proteins were separated by 2-DE in the pI range 4-7. More than 500 spots were detected by Silver staining and 70 different proteins corresponding to 111 spots were identified by MALDI-TOF and MS/MS analysis. Moreover, in the pI range 6-9, approximately 330 spots were detected and of these, 31 proteins corresponding to 38 spots were identified by mass spectrometry. More than 50% of the identified proteins were involved in amino acid, carbohydrate or

nucleotide metabolism, and energy production. A second large group of identified proteins comprises heat shock and other stress-related proteins and chaperones. Proteome analysis of the causative agent of the "chagas disease", *Trypanosoma cruzi*, was carried out by Parodi-Talice *et al.* (2004) who identified 45 of the 70 excised spots from the 2-DE gels.

For the cellular slime mold *Dictyostelium discoideum*, Yan *et al.* (1997) constructed a reference map. On this map, 150 spots were visualised by Amido black staining and 31 were putatively identified. Cytoplasmic extract from hyphal cells of the opportunistic human pathogen *Candida albicans* was analyzed by 2-DE and a reference map was obtained (Hernandez *et al.*, 2004). Protein identification was carried out. A total of 106 spots excised from 2-D gels were analyzed by MALDI-TOF or MALDI-TOF/TOF MS. This resulted in the identification of 43 proteins involved in metabolism, 13 involved in transcription, protein synthesis and fate, 8 involved in cell rescue, virulence and defense, and 2 proteins of unknown function.

A proteome study aimed at identification of cell envelope-associated proteins of *Trichoderma reesei* was performed by Lim *et al.* (2001). From the 2-D gels, a total of 56 spots were examined by nanoelectrospray tandem mass spectrometry and amino acid sequences were obtained for 32 spots. The identified spots on the 2-D gels included the proteins such as HEX1, vacuolar protease A, enolase, glyceraldehyde-3-phosphate dehydrogenase. Mass results showed an additional 12 spots which could not be annotated in the BLAST search indicating novel proteins.

A proteome map for the soil-borne filamentous fungus *Trichoderma harzianum* that exhibits biological control properties was also constructed (Grinyer *et al.*, 2004). From the 2-DE gels, 96 spots were cut for MS analysis and among these spots, 25 spots could be identified and shown to correspond to 22 different genes.

To carry out an initial analysis of *Entamoeba histolytica* trophozoite proteome, Leitsch *et al.* (2005) prepared 2-D gels and solubilized proteins were separated in the in IPG strips in different pH ranges. Of more than 1500 protein spots visualized, however, only 10 landmark spots were isolated, identified by either tryptic cleavage and subsequent MALDI-TOF MS or by protein sequencing.

Gelhaus *et al.* (2005) have studied the proteome of *Plasmodium falciparum* together with the infected red blood cells by means of 2-D gel and MALDI-TOF MS analyses. After 2-D PAGE separation of various protein samples, a total of 255 protein spots were analyzed by MALDI-TOF MS. Of these, 18% (45) were of nonplasmodial origin and 20% (52) of the protein spots could not be identified. 11% (28) of the analyzed protein spots were found to be annotated as hypothetical proteins. It was of particular interest that these proteins have never been detected before.

### **1.6.3. Human Proteome**

To map the human proteome, "Human Proteome Organisation (HUPO)" ([www.HUPO.org](http://www.HUPO.org)) was initiated after the completion of "Human Genome Project". HUPO includes some proteome research groups such as "Human Plasma Proteome Project" ([www.bioinformatics.med.umich.edu/hupo/ppp](http://www.bioinformatics.med.umich.edu/hupo/ppp)), "Human Liver Proteome Project" ([www.HLPP.org](http://www.HLPP.org)) and "Human Brain Proteome Project" ([www.HBPP.org](http://www.HBPP.org)). The Human Liver Proteome Project is the first initiative of the HUPO and aims to reveal human liver proteome, expression profiles, modification profiles, a protein linkage (protein-protein interaction) map, a proteome localization map as well as defining ORFeome (complete sets of open reading frames), physiome (determination of structure and function at all levels of biological organisation) and pathome (referring the collection of all pathophysiological phenotypes) (He, 2005).

Independently from HUPO, many other research groups have also been working on different tissues, organs and body fluids of human. Finehout *et al.* (2004) have constructed a 2-DE based proteome map of cerebrospinal fluid (CSF) from a single hydrocephalus patient. Their reference map contained 600 identified spots representing 82 different proteins. Of the identified 82 proteins, 25 have not appeared in any previously published 2-DE map of CSF, and 11 have not been previously reported to exist in CSF. As a diagnostic fluid, human urine proteome were also constructed (Oh *et al.*, 2004). On the 2-DE gels of human urine, 157 spots were analyzed and 80% of these spots were annotated to 113 different proteins, while 20% remained not annotated. Proteome reference maps are also beneficial tools for cancer researchers. For constructing a pituitary adenoma proteome map, Zhan *et al.*, (2003) excised 226 spots from 2-DE gels and 135 protein spots that represented 111 proteins were characterized with mass spectrometry. The characterized proteins included pituitary hormones, cellular signals, enzymes, cellular-defense proteins, cell structure proteins, transport proteins, *etc.*

In another work, 2-DE work for protein separation followed by MALDI-TOF analysis and database searches identified 231 protein spots corresponding to 138 individual proteins that were found in gels representing both the follicular and luteal phases of human ovary (Wang *et al.*, 2005). The data were used to construct an online database of human ovary proteome (<http://www.reprod.njmu.edu.cn/2d>).

#### **1.6.4. Plant Proteome**

In recent years, many plants were subjected to proteome analysis. For instance, an extensive work was performed for the construction of 2-DE based proteome database of rice (Komatsu 2005). For different tissues and subcellular fractions of rice, 2-DE followed by spot identification via MALDI-TOF analyses was performed and a total of 5676 spot annotated to 5092 proteins. A 2-DE based reference map of the mature pollen proteome of the dicotyledonous model plant species, *Arabidopsis thaliana*, was constructed

(Noir *et al.*, 2005). 121 different proteins were identified in 145 individual spots. The presence, subcellular localization, and functional classification of the identified proteins were discussed in relation to the pollen transcriptome and the full protein complement encoded by the nuclear *Arabidopsis* genome. The moss *Physcomitrella patens* reference proteome map was established by Sarnighausen *et al.* (2004). Out of 790 spots that could be resolved and displayed on a Coomassie-stained gel, those with moderate to high staining intensity were selected, cut from the gel and analyzed by mass spectrometry. 208 of these spots yielded interpretable mass spectra. From these 208 spots, 306 different proteins were identified. 29% of all analyzed spots comprised of two different proteins, while 17% of the spots comprised of three, exemplifying the limits of 2-DE in cases of such spot overlaps. A proteome reference map for maize (*Zea mays L.*) endosperm by means of 2-DE and protein identification with LC-MS/MS analysis was constructed in the pI range of 4 to 7 (Mechin *et al.*, 2004). Among the processed 632 protein spots, 496 were identified. Metabolic processes, protein destination, protein synthesis, cell rescue, defense, cell death and aging were the most abundant functional categories.

#### **1.6.5. Posttranslational Modifications (PTM)**

One of the first hurdles to be negotiated in the postgenomic era involves the description of the entire protein content of the cell, the proteome. Such efforts are presently complicated by various posttranslational modifications that proteins can experience, including glycosylation, lipid attachment, phosphorylation, methylation, disulfide bond formation, and proteolytic cleavage (Eichler and Adams, 2005). Multiple spots those having the same identification on the 2-DE maps do show the possible PTM of proteins. Rosen *et al.* (2004) found that as much as 10% of the soluble proteins in the 2-DE map of *A. tumefaciens* appeared as multiple spots.

### **1.6.5.1. Phosphoproteomics**

Presently, phosphorylation of proteins is the most studied and best understood PTM. However, the analysis of phosphoproteins and phosphopeptides is still one of the most challenging tasks in contemporary proteome research (Reinders and Sickmann, 2005). Phosphoproteomics have been fueled by the need to monitor simultaneously many different phosphoproteins within the signaling networks that coordinate responses to changes in the cellular environment (Mumby and Brekken 2005).

The variety of functions in which phosphoproteins are involved necessitates a huge diversity of phosphorylations (Reinders and Sickmann, 2005). Several amino acids can be phosphorylated by the four known types of phosphorylation. O-phosphates are the most common class and they are mostly attached to serine-, threonine- and tyrosine-residues but also to unusual amino acids such as hydroxy-proline. N-, S- and acyl-phosphorylation are far less spread and occur mostly on histidine and lysine (N-phosphates), cysteine (S-phosphates) and aspartic and glutamic acid residues (acyl-phosphates). In some cases, the kind of phosphorylation on a specific residue has not been revealed up to now, *e.g.*, phosphorylation of hydroxy-lysine (N- or O-phosphorylation).

It has been estimated that the mammalian genome encodes >2000 different protein kinases. Experiments by 2D-gel electrophoresis indicate that as much as 30-50% of the proteins in a eukaryotic cell may be phosphorylatable (Blom *et al.*, 1998). Since there may be ~10 000-30 000 different protein species in a eukaryotic cell, the average kinase probably has more than one target protein.

To identify the phosphorylated proteins of the organisms, phosphoproteome maps have also been constructed besides the reference proteome maps. The alternative methods for detection of phosphorylated proteins within complex

mixtures were extensively reviewed by Reinders and Sickmann, 2005. It is usually not possible to detect them without prior separation of the proteins such as 2-DE. Phosphoproteins can be visualized either directly in 2-D gels using phosphospecific stains or by Western blotting techniques. The most sensitive method is radioactive labelling of the phosphate-groups using  $^{32}\text{P}$  or  $^{33}\text{P}$  and subsequent radioimmunoblotting. A major advantage of these techniques is that all kinds of phosphorylations are detected and signals can be quantitated absolutely. Introduction of radioactive phosphate-groups can be done *in vivo* or *in vitro*. The latter is carried out by incubating the respective protein or protein mixture with a chosen kinase and [ $\gamma$ - $^{32/33}\text{P}$ ]-ATP. After an appropriate incubation time, the sample can be subjected to downstream analysis steps. Western-blotting using phosphospecific antibodies is also widely used and is able to detect very low amounts—few femtomoles— of phosphoproteins, but specificity and sensitivity of this method is strongly dependent on the respective antibodies. Commercially available phospho-stains are less sensitive than radioactive methods but the handling of these “inactive” reagents is far more convenient. These stains are mostly specific for certain types of phosphorylation such as all O- or just Ser/Thr-phosphorylations depending on the respective staining principle. Furthermore, good compatibility to other staining methods and to MS is reported as no blotting is needed.

As can be seen, detection and sequencing of phosphopeptides is a well-studied problem. Nevertheless, standard techniques of site identification by mass spectrometry or  $^{32}\text{P}$ -labeling coupled with Edman sequencing are inefficient when dealing with highly complex mixtures of proteins or peptides and still require relatively pure starting samples. One solution is to selectively enrich phosphopeptides from unpurified mixtures (Ahn and Resing, 2001; Reinders and Sickmann, 2005). A first-generation technique that is easy and cheap is immobilized metal affinity chromatography (IMAC), in which

phosphopeptides are bound noncovalently to resins that chelate Fe(III) or other trivalent metals, followed by base elution.

As an example to recent phosphoproteome studies, a proteome map and a phosphoproteome map were constructed for rat mesangial cells by Jiang *et al.* (2005). 37 protein spots (representing 28 unique proteins) of the 157 identified protein spots in the silver stained 2-DE map was found to be phosphorylated. In a recent study, Khan *et al.* (2005) has studied a rice phosphoproteome. 44 phosphoproteins were detected on a 2-D gel. Among the phosphoproteins detected, 42 were identified through analysis with MALDI-TOF MS. The largest percentage of the identified phosphoproteins are involved in signaling (30%), while 18% are involved in metabolism. Using a 2-DE-independent approach, Ptacek *et al.* (2005) detected in yeast cells approximately 4,200 phosphorylation events affecting 1,325 proteins by performing 87 protein kinase assays. Their phosphorylation results constituted a first-generation phosphorylation map for yeast.

#### **1.6.5.2. Glycoproteomics**

Glycosylation is perhaps the most common and versatile posttranslational modification of proteins. For instance, it is estimated that over one-half of all mammalian proteins are glycosylated (Apweiler *et al.*, 1999). The field glycomics involves the investigation and documentation of the large collection of glycan structures. Glycan moieties are involved in a wide variety of intracellular, cell-cell and cell-matrix recognition events. An understanding of how glycosylation affects the activities and functions of proteins in health and disease represents a major challenge.

#### **1.6.6. Subcellular Proteome Studies (Subproteome)**

Combination of proteomics with cell biology has produced extensive inventories of the proteins that inhabit several subcellular organelles. Recent

proteomic analysis has identified many new putative subcellular compartments-specific proteins. Subproteome studies were recently focused on cell membranes (Nouwens *et al.*, 2000; Ying *et al.*, 2005). Hanson *et al.* (2001) established a proteome map for mitochondrion of MWC-5 fibroblast cells. The cytosolic, cell surface and extracellular proteomes for *Corynebacterium efficiens* was also constructed very recently by Hansmeier *et al.* (2006).

### **1.7. Stress-Related Proteome Studies**

Currently, stress and stress-related physiological studies are being conducted in many laboratories. As a tool, proteomic studies represent the state-of-the-art for many different stress-inducing applications. The basic goal of a proteomic studies is to identify proteins involved in particular processes (Washburn and Yetas, 2000). When the aim is to find interesting protein(s) whose function might relate to the conditions of interest, investigators often focus on the induced proteins. A dramatic increase in the abundance of a protein indicates the importance of protein in the response to a given environmental change; the heat and cold shock proteins of *E coli* being the best examples of this approach (Tohomas and VanBogelen, 2000). Another study that used the proteomic approach was conducted by Rosen *et al.* (2002) who worked with *Agrobacterium tumefaciens* subjected to heat shock.

Blomberg (1995) reported that when exponentially growing *S. cerevisiae* was challenged with increased salinity (in 0.7 M NaCl-containing medium), a number of proteins displayed changes in their relative rate of synthesis. During the stress, 18 proteins exhibited increases more than eighth-fold and those proteins were named as "highly NaCl-responsive proteins". Given that *S. cerevisiae* genome project was completed in the year 1996 (Goffeau *et al.*), many of such NaCl-responsive protein genes had not been reported elsewhere before this study. The study also revealed that only two proteins

were repressed. The metabolic alterations and adaptations of yeast cells in response to osmotic stress and the relevant proteome studies were later reviewed extensively by the same author (Blomberg, 1997, 2000).

2-DE gel protein database of *Saccharomyces cerevisiae* was last updated by Perrot *et al.*, in 1999. On the same year, a very detailed study on yeast proteome was reported by Futcher *et al.*, (1999) providing quantitative information about 1400 spots. The relative abundance of proteins was measured and compared in glucose and ethanol media. Many heat shock proteins were about twofold more abundant in ethanol medium than in glucose medium which was quite consistent with the increased heat resistance in ethanol medium.

Living organisms use several mechanisms to counter toxic effects of heavy metals. Vido *et al.* (2001) analyzed the proteomic response of *S. cerevisiae* cells to acute cadmium stress. The organism responded to the presence of cadmium with 54 induced and 43 repressed proteins as a total. The induced proteins included 9 sulfur amino acid biosynthetic pathway enzymes, several proteins with antioxidant properties, heat-shock proteins, chaperones and proteases. Their data, altogether, indicated that two cellular thiol redox systems, glutathione and thioredoxin were essential for cellular defense when the cells were exposed to 1 mM cadmium.

Under certain conditions, microbial metabolism produces acetate and formate which are known for their cytotoxic effects. By using 2-DE, Steiner and Sauer (2001) found that at least 50 proteins were specifically induced by adaptation to acetate but not by other stress conditions (heat, oxidative or osmotic stress) in *Acetobacter aceti*. 19 of those proteins were significantly induced under both batch and continuous culture conditions.

In response to heat, oxidative and acid pH stresses, the intracellular pathogen *Brucella melitensis*, regulates 19 out of 676 proteins up or down, some of them being identical to those previously reported for *Brucella*, such as BvrR, DnaK, GroEL, and Cu-Zn superoxide dismutase (SOD) (Teixeira-Gomes *et al.*, 2000). Eight other proteins closely matched to other bacterial proteins (e. g. Fe and/or Mn SOD).

A proteome analysis of *B. subtilis* was conducted under low and high salinity conditions which revealed 16 induced and 2 repressed proteins (Hoffmann *et al.*, 2002). Four of the 16 high salinity-induced proteins corresponded to the enzymes involved in the synthesis of 2,3-dihydroxy benzoate (DHB) and modification and esterification leading to the iron siderophore bacillibactin. The authors strongly believed that *B. subtilis* cells growing in a high salinity environment experience iron limitation. Therefore, as a stress condition, high salinity and osmolarity lead to alteration of iron controlled genes. In an earlier study done by Bernhardt *et al.* (1997) who investigated the changes in protein synthesis profile of wild type and *sigB* mutants in response to heat shock, ethanol stress and glucose or phosphate starvation, it was reported that at least 42 general stress proteins absolutely required the alternate sigma factors. The relative synthesis rate of the general stress proteins made up to 40 % of the total proteins of the stressed cells as "acetate adaptation proteins". In a similar manner, acetate and formate response of *E. coli* was studied by Kirkpatrick *et al.* (2001). Presence of acetate increased steady-state expression of 37 proteins, including periplasmic transporters for amino acid and peptides, metabolic enzymes, the autoinducer synthesis proteins, and so on.

Bebien *et al.* (2001) reported the selenite effects on the growth rate and protein synthesis in *Rhodobacter sphaeroides*. Proteome analysis of selenite response revealed an enhanced synthesis of some chaperones, an elongation factor and the enzymes associated to oxidative stress. Induction of these

antioxidant proteins confirmed that the major effect of selenite involves the formation of reactive oxygen species during growth.

In a very recent study, fungicide exposed *Neurospora crassa* cells were studied in terms of induction of OS-2 protein phosphorylation (Irmiler *et al.* 2006). In the study 2-DE gels was constructed to identify the putative key enzymes involved in the cascade downstream of *os-2*. 2-DE was carried out in order to characterize the changes in the protein and phosphoprotein composition between the mycelia of the wild type and the *os-2* mutant strains of *N. Crassa*. In the 2-DE gels 23 proteins showed different levels of expression. Among the genes that show *os-2*-dependent expression, the vacuolar protease A and the polyketide synthase E69679 appeared to be positively regulated, whereas NADP-glutamate dehydrogenase and keto-acid reductoisomerase are negatively regulated by *os-2*. 17 of these proteins found to be phosphorylated and only one of them, which could not be identified, was expressed in mutant strain.

### **1.8. *Phanerochaete chrysosporium***

Taxonomy of the model wood decay fungus *Phanerochaete chrysosporium* was described by Burdsall and Eslyn (1974) ([http://botit.botany.wisc.edu/toms\\_fungi/may97.html](http://botit.botany.wisc.edu/toms_fungi/may97.html)). The relevant classification is as follows:

Kingdom:	Fungi
Division:	Basidiomycota
Class:	Hymenomycetes
Subclass:	Homobasidiomycetes
Order:	Aphyllorphales
Family:	Corticaceae
Genus:	Phanerochaete

*P. chrysosporium* is one of the resupinate or crust white-rot fungi that decay wood. These fungi never form a mushroom for reproduction, but form effused, very flat, fruiting bodies that appear as no more than a crust on the underside of a log.

The *P. chrysosporium* genome is approximately 30 Mb in size and organized in 10 chromosomes. The *P. chrysosporium* genome sequencing was completed by DOE Joint Genome Institute (Martinez *et al.*, 2004) and constituted the first completed basidiomycete genome project.

*P. chrysosporium* has several features that make it very useful biotechnologically. First of all, unlike some other white rotters, it leaves the cellulose of the wood virtually untouched which is a major asset in paper production. Secondly, it has a high optimum temperature (about 40 °C) making it useful in industrial operations. It has been investigated as a possible biobleaching and biopulping agent to replace the harsh chemicals that are being used in conventional paper bleaching. It produces lignin peroxidases (LiPs), a family of extracellular glycosylated heme proteins, as major components of its lignin-degrading system. Some of the lignin-degrading enzymes of *P. chrysosporium* can also degrade toxic wastes, such as PCB's, PCP's and TNT's, representing very important capabilities from the pointview of bioremediation.

The biomass of *P. chrysosporium* has also shown potential as a metal biosorbent for treating industrial effluents and contaminated resources (Yetiş *et al.*, 1998). Using this organism, biosorption studies were performed for different metals and the kinetics of removal of Pb (II) by live, resting and dead cells were documented earlier, with the cooperation of our research group (Yetiş *et al.*, 2000).

Very recently, a proteome study for secreted proteins (secretome) of *P. chrysosporium* was performed (Wymelenberg *et al.*, 2005). By the use of

bioinformatic tools, it was predicted that the current *P. chrysosporium* protein database likely contains 268 proteins with signal peptide sequences. A total of 182 unique peptide sequences they experimentally obtained from the secretome were matched to 50 specific genes. Identified proteins included functionally interconnected enzyme groups, e.g. glycosyl hydrolases, the multiple endoglucanases and processive exocellobiohydrolases, a hemicellulolytic system included endoxylanases, acetyl xylan esterase, and arabinofuranosidase.

### **1.9. Aim of the Present Study**

The aim of the present study was to establish a reference proteome map for *P. chrysosporium* as based on 2-DE followed by MALDI-TOF MS analysis. Our approach involved the identification of the main spots on the 2-DE gels of *P. chrysosporium* when the organism was grown under normal physiological conditions. Annotation of the identified proteins was made by using the DOE Joint Genome Institute's *P. chrysosporium* genomic database. The data presented by the present work constitutes the first proteome map originated from our country. The map has also formed a prelude for large-scale study of heavy metal response of the organism which is being undertaken in our laboratory.

## CHAPTER 2

### MATERIALS AND METHODS

#### 2.1. Microorganism and Its Maintenance

The white-rot fungus, *Phanerochaete chrysosporium* (ATTC 24725) which was kindly provided by Prof. Filiz B. Dilek (Env. Eng. Dept., METU) was used in this study. The fungus was inoculated to Sabaroud Dextrose agar slants and incubated for 4 days at 35 °C for spore formation. The organism was transferred monthly and stored at 4 °C for further use.

#### 2.2. Growth Conditions and Biomass Preparation

Biomass was prepared as reported by Yetiş *et al.* (2000). *P. chrysosporium* spores were separated from Sabaroud Dextrose agar slant surfaces by scrapping. The biomass was gently suspended to homogeneity by hand using a teflon homogenizer (İldam Cam A.Ş.). A spore suspension was prepared to contain  $2.5 \times 10^6$  spores /mL at an absorbance of 0.5 at 650 nm using a Shimadzu UV-1208 spectrophotometer. The spore suspension was then transferred into 250 mL Erlenmayer flasks each containing 150 mL of the growth medium. The medium used to grow *P. chrysosporium* in this study is that described by Prouty (1990). The medium is composed, in g/L of glucose, 10; KH<sub>2</sub>PO<sub>4</sub>, 2; MgSO<sub>4</sub>, 0.5; CaCl<sub>2</sub>, 0.1; NH<sub>4</sub>Cl, 0.12 and thiamine, 0.001 and adjusted to a pH of 4.5. The cultures were incubated for 40 h at 200 rpm in a rotary shaker at 35 °C. When the growth was terminated at 40<sup>th</sup> h, the cultures were still in exponential growth (Özcan, 2003) and the bead-like mycelia of the fungi had a diameter of 2.5 to 4 mm.

### **2.3. Protein Extraction**

For the extraction of proteins, after crushing the cells with liquid nitrogen, TCA-acetone extraction was performed as described by Damerwal *et al.* (1986). Shortly, after breaking a 500 mg of harvested mycelium in liquid nitrogen, 5 mL of 10 % trichloroacetic acid in acetone containing 0.07 %  $\beta$ -mercaptoethanol was added and vortexed, then incubated at  $-20\text{ }^{\circ}\text{C}$  for 45 min and centrifuged at 15.000 xg for 15 min. The supernatant was decanted, and the pellet was resuspended in 5 mL of acetone containing 0.07 %  $\beta$ -mercaptoethanol, then incubated at  $-20\text{ }^{\circ}\text{C}$  for 1 h (mixed every 15 min intervals by vortexing) and centrifuged as above. Then the supernatant was discarded, the remaining pellet was vacuum-dried and stored as a powder at  $-20\text{ }^{\circ}\text{C}$ . The proteins obtained by using this procedure represented total cellular proteins since the debris contained whole soluble protein complements of the cells.

### **2.4. Protein Estimation**

To determine total protein concentrations, the modified Bradford assay described by Ramagli and Rodrigez (1985) was used. 5 X Bradford reagent (containing 500 mg Coomassie Brilliant Blue G- 250, 250 mL of 96% ethanol and 500 mL of 85 % ortho-phosphoric acid; completed to a 1 L with dH<sub>2</sub>O) was diluted 1:3 with dH<sub>2</sub>O and filtered at least three times using Whatman No. 1 filter paper. For determination of the total protein concentration, 10 mg of a mycelium extract powder was dissolved in 500  $\mu\text{L}$  of rehydration buffer [containing 8 M urea, 2 M thiourea, 0.3 % DTT, 1% (w/v) CHAPS, 0,3 % (w/v) DTT, 0,5% (v/v) ampholyte pH 3-10]. The suspension was mixed and incubated at room temperature for 1 h and then centrifuged at 12 000 xg for 5 min. To 20  $\mu\text{L}$  of aliquots of the supernatant, 80  $\mu\text{L}$  of 0.1 N HCl were added to protonate samples and mixed thoroughly. To this mixture, 3.5 mL of 1:3 diluted 5X Bradford reagent was added, incubated at room

temperature for 10 min and absorbance was measured at 595 nm. Bovine Serum Albumin (BSA) fraction number V was used as a standard for the construction of calibration curves.

## **2.5. Proteome Study**

### **2.1.1. Isoelectric Focusing**

Isoelectric focusing was performed in 18 cm IPG-strips (pH range 3-10 linear, Amersham Biosciences, Uppsala, Sweden). IPG strips were passively rehydrated by applying 300 µl of rehydration buffer including 300 µg protein for 16 h. The isoelectric focusing was performed with the MultiphorII unit (Amersham Biosciences) employing the following voltage profile; linear increase from 0 to 500 V for 2500 Vh, 500 V for 2500 Vh, linear increase from 500 to 3500 V for 10.000 Vh and a final phase of 3500 V for 35 000 Vh.

### **2.2.1. SDS-PAGE**

After IEF, focused strips were equilibrated according to Görg *et al.* (1988). Equilibration included the reduction step; shaking of the IPG strips for 15 min in equilibration buffer (6 M urea, 30% (w/v) glycerol and 2% (w/v) SDS in 0.05 M Tris-HCl buffer, pH 8.8.) containing 100mg/10ml DTT and an alkylation step; shaking of the IPG strips for 15 minutes in equilibration buffer containing 250mg/10ml Iodoacetamide.

Equilibrated IPG strips was placed on top of the 1 mm thick, continuous polyacrylamide gels of 12.5% T and 2.6% C according to Leammli (1975). SDS-PAGE was performed by using the Millipore electrophoresis system (Millipore, MA. USA) by applying approximately 2 W per gel.

After electrophoresis, protein spots were stained with colloidal Coomassie blue (CCB) by using the procedure of Neuhof *et al.* (1988) (Appendix A).

CCB-stained gels were stored in a little amount of water in dark cold room for image analysis.

### **2.3.1. Phosphoprotein Detection**

Fluorescent staining of 2-DE gels using Pro-Q Diamond phosphoprotein gel stain (Molecular Probes, Eugene, OR) was performed according to the manufacturer guidelines. Briefly, the gels were fixed in 45% methanol, 5% acetic acid overnight, washing with three changes of deionized water for 10–20 min per wash, followed by incubation in Pro-Q Diamond phosphoprotein gel stain for 180 min, and destaining with successive washes of 4% ACN in 50mM sodium acetate, pH 4.0. The resulting gels were scanned with a fluorescent laser scanner (Biorad). After scanning, the gels were Coomassie-stained for visualization of the spots for comparative analysis with Delta2D.

### **2.4.1. Evaluation of 2-DE Data**

Scanning of preserved gels was done by using a X-Finity scanner (Quato Graphics Braunschweig, Germany). Spot pattern analyses and labeling of the spots to be cut were accomplished by using the 2D image analysis software Delta2D version 3.3 (Decodon, Germany).

### **2.5.1. Sample Preparation and Analysis by Mass Spectroscopy**

The protein spots were excised from the stained 2D gels using a spot cutter (Proteome Works<sup>TM</sup>, Biorad, Hercules, CA, USA). Cut spots were transferred into 96 well microtiter plates. The tryptic digest with subsequent spotting on a matrix assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF MS target) was carried out automatically with the Ettan Spot Handling Workstation (Amersham Biosciences, Uppsala, Sweden) using the following protocol: The gel pieces were washed twice with 100µl of a

solution of 50% CH<sub>3</sub>CN and 50 % 50 mM NH<sub>4</sub>HCO<sub>3</sub> for 30 min and once with 100µl 75 % CH<sub>3</sub>CN for 10 min. After drying at 37 °C for 17 min 10µl trypsin solution containing 20 ng/µl trypsin (Promega, Madison, WI, USA) was added and incubated at 37 °C for 120 min. For extraction, the gel pieces were covered with 60 µl 0.1 % trifluoroacetic acid (TFA) in 50 % CH<sub>3</sub>CN and incubated for 30min at 40 °C. The peptide containing supernatant was transferred into a new microtiter plate and the extraction was repeated with 40 µl of the same solution. The supernatants were dried at 40 °C for 220 min completely. The dry residue was dissolved in 3 µl of 0.5 % TFA in 50 % CH<sub>3</sub>CN and 0.4 µl of this solution were directly spotted on the MALDI target. Then 0.4 µl of a saturated α-cyano-4-hydroxy cinnamic acid solution in 70 % CH<sub>3</sub>CN were added and mixed with the sample by aspirating the mixture five times. The samples were allowed to dry on the target for 10 to 15 min before measurement in MALDI-TOF.

The MALDI-TOF measurement was carried out on the Proteome-Analyzer 4700 (Applied Biosystems, Foster City, CA, USA). This instrument is designed for high throughput measurement, being automatically able to measure the samples, calibrating the spectra and analyzing the data using the 4700 Explorer™ Software. The spectra were recorded in a mass range from 900 to 3700 Da. If the autolytical fragment of trypsin with the mono-isotopic (M+H)<sup>+</sup>m/z at 2211.104 reached a signal to noise ratio (S/N) of at least 10, an internal calibration was automatically performed as one-point-calibration using this peak. After calibration the peak lists were created by using the "peak to mascot" script of the 4700 Explorer™ Software. The three highest peaks were chosen automatically for recording an MS/MS spectrum. The peak lists of the MS/MS experiments were included to the mascot search to improve the reliability of the search result.

### **2.6.1. Protein Identification**

*P. chrysosporium* whole genomic DNA sequence data which are available in the organism's genome project (Joint Genome Institutes) web address (<http://genome.jgi-psf.org/Phchr1/Phchr1.home.html> database version 2.1) have already been translated to amino acid sequences in FASTA format and published at the same web address. In order to construct a database for PMF search for *P. chrysosporium* proteins, the whole database was downloaded and processed for theoretical trypsin digestion of amino acid sequences by the use of MASCOT program to construct a theoretical peptide mass fingerprint (PMF) database. The peak lists of each protein spot (peptide mass fingerprint and MS/MS data) obtained from MALDI TOF MS measurement were analyzed with the aid of "Peptide Mass Fingerprint" and "MS/MS Ion Search" engines of MASCOT software (Matrix Science Inc., Boston, MA, USA) against the above-mentioned *P. chrysosporium* PMF database. Of the results given by the MASCOT software, those having a probability score value higher than 53 (which was calculated statistically by the software as base score for a meaningful identification according to genome size of *P. chrysosporium*) were considered for protein identification. To find out putative functions, protein accession numbers of the identified spots which were included in the MASCOT output were searched in the website of the Joint Genome Institutes for *P. chrysosporium*. The JGI database uses InterPro approach which includes an integrated documentation resource for protein families, domains and sites, for putative function assignment for each protein in their databases. For the proteins which were certainly assigned to the particular genes, i.e. the identified ones, the functional classification was made by consulting to the functional categories list contained the JGI website.

### **2.7.1. Bioinformatic analyses**

For each identified protein, the theoretical  $pI$  and MW values were calculated by the use of ExPASy's MW/ $pI$  calculation webpage

([http://www.expasy.org/tools/pi\\_tool.html](http://www.expasy.org/tools/pi_tool.html)). The amino acid sequence of each identified protein was submitted to this webpage and the resulting MW/ $pI$  values were assigned.

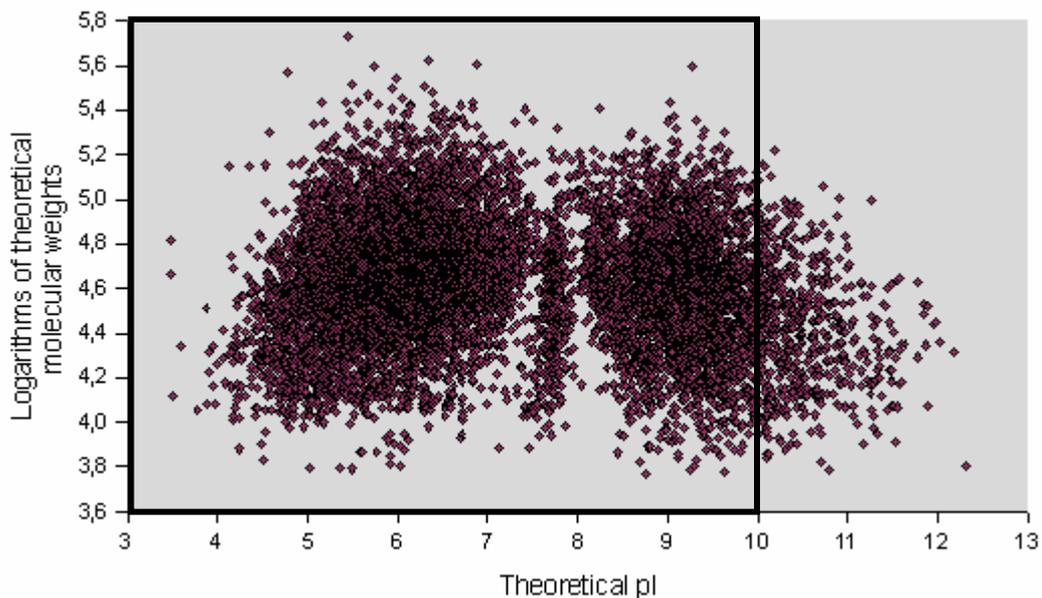
Protein subcellular localization predictions were obtained from PSORT web server (<http://wolfpsort.seq.cbrc.jp/>) by submitting amino acid sequences of the identified proteins to this server. The existence of signal peptide sequences was checked in the public web server of signal peptide prediction program SignalP version 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>).

## CHAPTER 3

### RESULTS AND DISCUSSION

#### 3.1. *P. chrysosporium* Theoretical 2-DE Map

In the basic 2-D separation method for resolving proteome, proteins with extended hydrophobic regions, such as membrane proteins are neglected under the standard 2-D gel conditions. In addition, one has to consider the tradeoff between resolving power and the width of the gel pH gradient. To determine the optimal range for standard 2-D gel-based proteome analysis of *P. chrysosporium*, the theoretical gel was calculated from the genomic data in the JGI database (<http://genome.jgi-psf.org/Phchr1/Phchr1.home.html>). The data for the pseudo 2-D gel consisted of theoretical pI and MW values which were obtained from the ExPASy's pI/MW calculation tool by submitting 10048 protein entries of *P. chrysosporium* database by the help of a Java script.



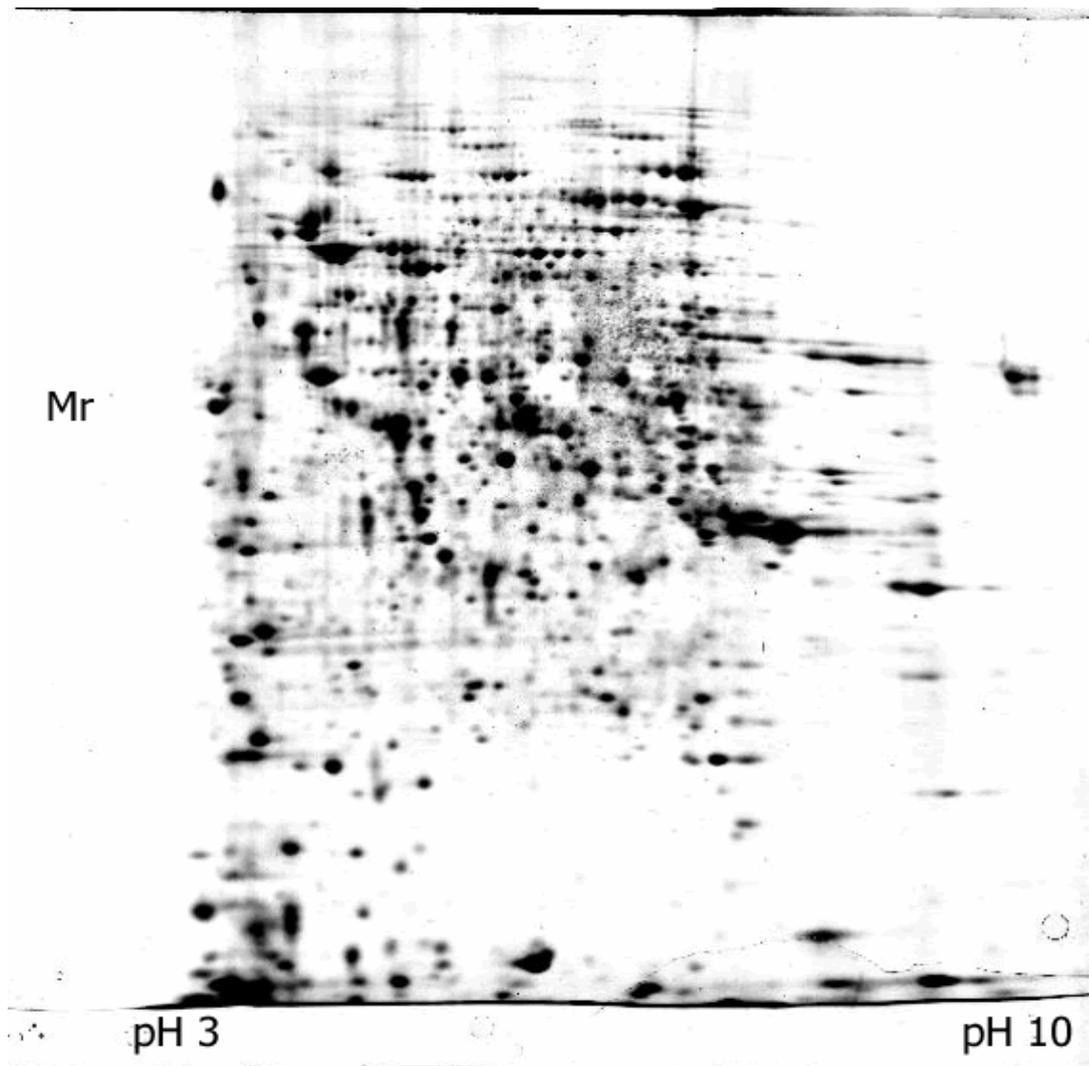
**Figure 1.** Predicted 2-D gel separation of the *P. chrysosporium* proteome (theoretical 2-DE map). The rectangle indicates the analytical window of the experimental master gel.

The resulting theoretical map revealed the optimal *pI* range for a standard 2-D gel-based proteome analysis of *P. chrysosporium* (Figure 1). The *pI* values ranged from 3,48 to 12,38 and the molecular weights were between 5,845 kDa and 536,200 kDa. The majority of the proteins of *P. chrysosporium* were found to have a theoretical *pI* between 4 and 12. However, since the commercially available IPG strips include only those having a pH range of 4 to 7, 3 to 6, or 3 to 10, respectively, the latter one was chosen to be the standard analytical window in our 2-D gel electrophoresis experiments.

### **3.2. Master Gel**

Physiological proteomics requires 2-D reference maps on which main protein components of a physiological response are identified. In this study, our aim was the construction of a reference proteome map of the cells growing exponentially under normal physiological conditions, hence to establish a base for further physiological proteome studies on *P. chrysosporium*. According to the typical growth curve of the organism, the culture was in mid-exponential phase of growth when the cells were collected for protein extraction at 40<sup>th</sup> h of incubation. The soluble protein fraction of *P. chrysosporium* was then visualized by using 2-DE approach by using IPG strips in a pH range from 3 to 10 for the first dimension.

A total of 910 protein spots could be separated and detected on Coomassie-stained 2-D gels of *P. chrysosporium* by the help of Delta2D (Decodon, Germany) image analysing software. However, due to the technical limitations of the spot cutter, only 720 spots could be cut from each master gel (1440 spots as a total from the duplicate master gels). Figure 2 shows the colloidal Coomassie-stained master gel for *P. chrysosporium* cells grown under standard conditions.

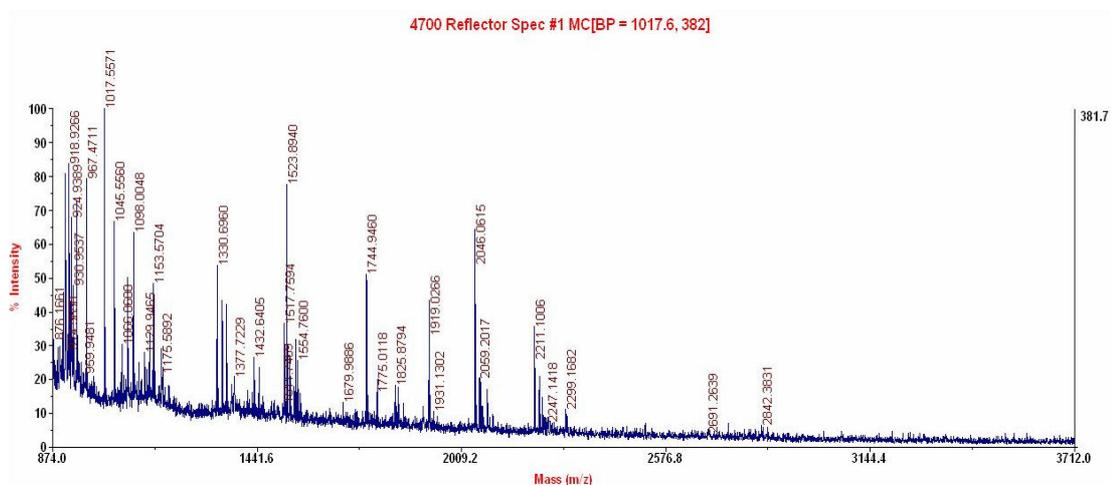


**Figure 2.** A general view of 2-DE of *P. chrysosporium* soluble proteins.

### **3.3. Identification of *P. chrysosporium* Proteins by MALDI-TOF MS**

For each cut spot, in gel digestion with trypsin was performed and the spots were processed for MALDI TOF MS analysis. The protein database of *P. chrysosporium* was processed with theoretical trypsin digestion of amino acid sequences by the use of MASCOT program to obtain a theoretical peptide mass fingerprint (PMF) database. After MALDI-TOF MS analysis, the peak list of each protein spot, i.e. PMF spectra was analyzed with the aid of MASCOT Search engine for *P. chrysosporium* database. The distinct proteins identified by their accession numbers were then searched in JGI database

for their specific functions. To exemplify data processing, the PMF obtained from the MALDI-TOF MS spectra of one randomly chosen protein (Figure 3), its MASCOT search result showing the respective protein accession number along with its probability score (identification; Figure 4) and the output of *P. chrysosporium* database search for assignment of function (annotation; Figure 5) are illustrated. As can be seen, the protein was identified with the accession number 10340 and then shown to be the heat shock protein Hsp70.



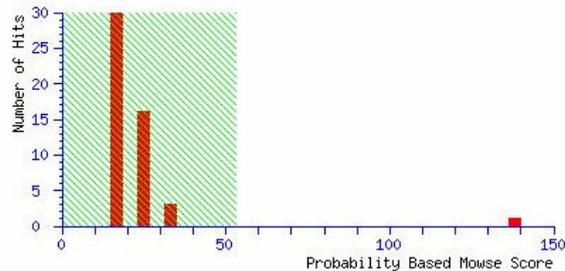
**Figure 3.** MALDI-TOF MS spectra of the peptides (PMF) obtained from one randomly chosen protein spot.

## *MATRIX* SCIENCE Mascot Search Results

User : Mascot Daemon  
 Email :  
 Search title : Project: Hecker\050727, Spot Set: Hecker\050727\Volkan\_III-Daniela+Holger\_TOF4  
 MS data file : Z:\volkan\Volkan\_III\ppw\_A41-A6\_112249954840.txt  
 Database : Phanerochaete 050727 (10048 sequences; 4564713 residues)  
 Timestamp : 29 Jul 2005 at 21:07:57 GMT  
 Warning : A Peptide summary report will usually give a much clearer picture of MS/MS sea:  
 Top Score : 138 for gi|10340|fgenes1, pm.C scaffold 6000005

### Probability Based Mowse Score

Ions score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event.  
 Protein scores greater than 53 are significant ( $p < 0.05$ ).



**Figure 4.** Representative MASCOT search output revealing that the protein given in Figure 3 is encoded by the ORF "pm.C scaffold 6000005" (protein accession number 10340).

**JGI** *Phanerochaete chrysosporium* v2.0  
 Search | BLAST | Browse | GO | KEGG | KOG | AdvancedSearch | Download | Info | Home | **HELP!**

Name: fgenes1\_pm.C\_scaffold\_6000005  
 Location: Phchr1/scaffold\_6:138921-141446  
 Description:  
 Best Hit: jgi|whiterot1|65251|pc.124.6.1 (model%: 100, hit%: 100, score: 3574, %id: 91) [Phanerochaete chrysosporium]  
 total hits(shown) 157 (10)

View nucleotide and 3-frame translation To Genome Browser  
 NCBI blastp Predicted number of transmembrane domains: 0

fgenes1\_pm.C\_scaffold\_6000005 To Genome Browser

1 42 65 199 1908 2526 78  
 52 59 59 54

1 77 154 230 307 383 459 536 612 689 764

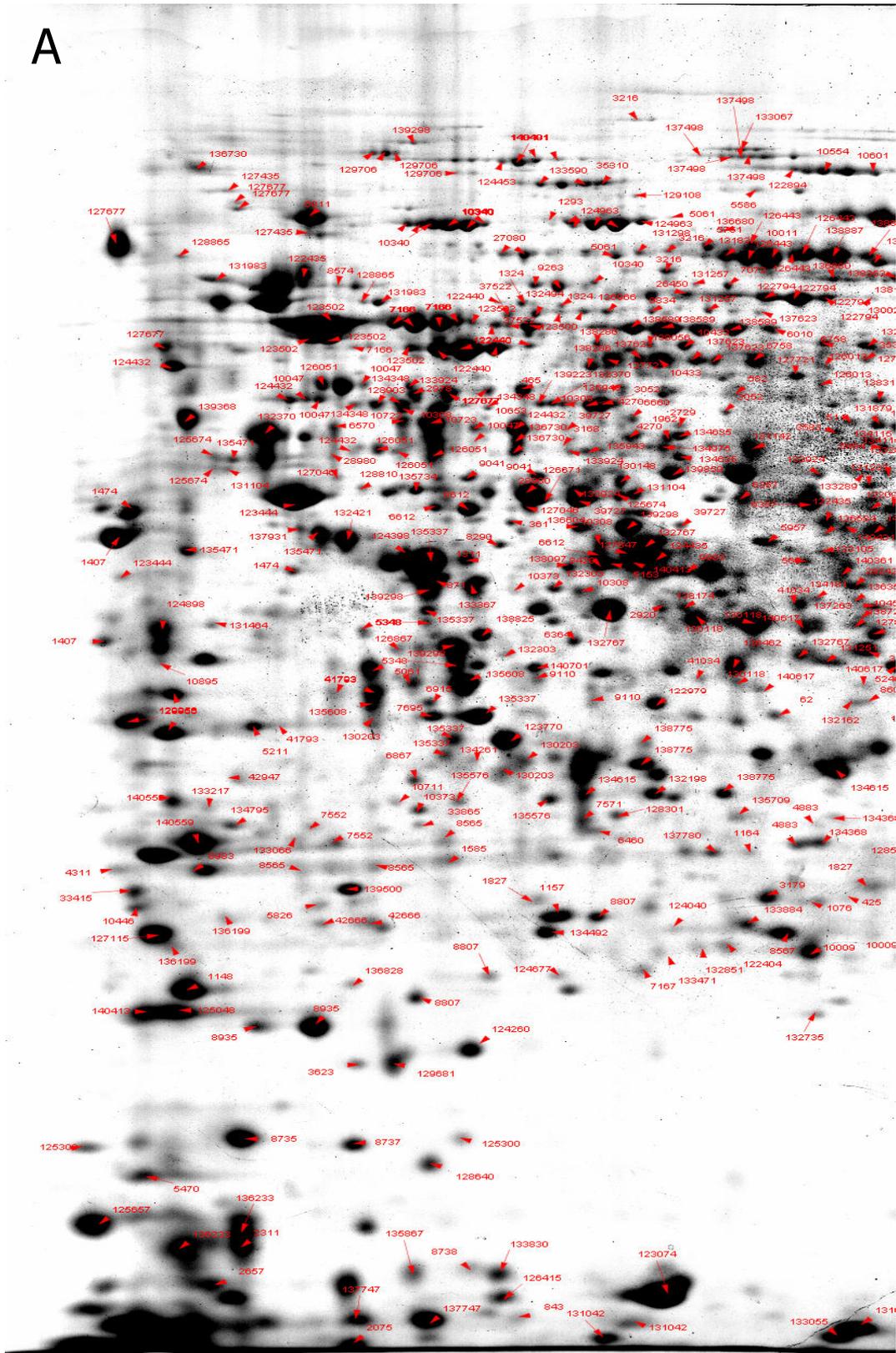
IPR001023: Heat shock protein Hsp70 [BlastProDom]  
 IPR001023: Heat shock protein Hsp70 [HHMPfam]  
 IPR001023: Heat shock protein Hsp70 [FPrintScan]  
 IPR001023: Heat shock protein Hsp70 [ScanRegExp]

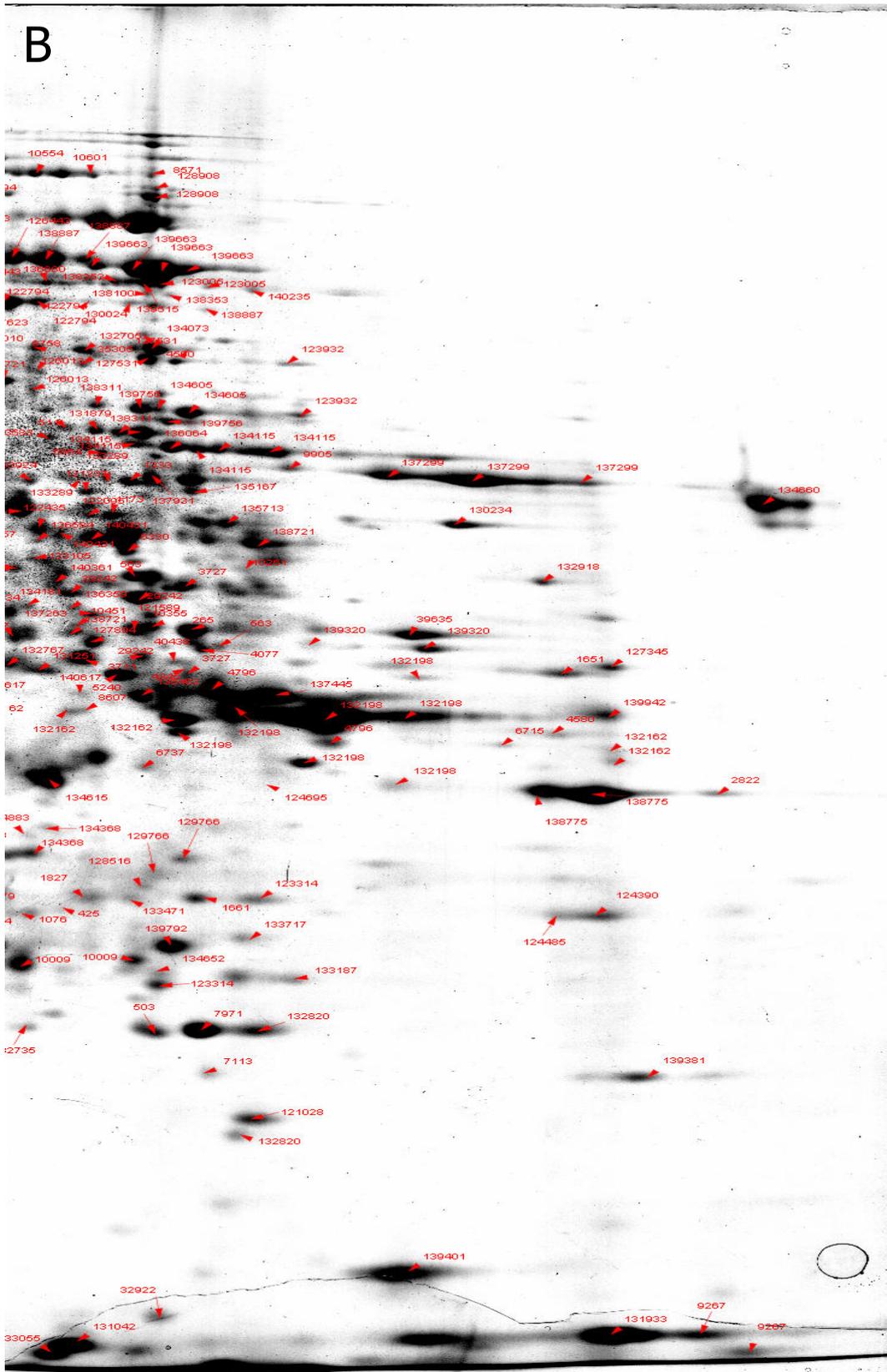
**Figure 5.** Representative output of *P. chrysosporium* database search referring the protein with the accession number 10340 (shown in Figure 4) to the heat shock protein Hsp70.

In analysis of 2-D gels, there are always samples that can not be identified exclusively with the information available in the PMF and hence need to be analyzed with a method that is able to provide primary structure information on a peptide, such as tandem MS systems. In this work, such proteins of *P. chrysosporium* were analyzed by tandem mass spectrometry using MALDI-TOF/TOF. According to the MASCOT results, 517 spots out of 720 gave statistically significant scores (only the identification hits above a significance level of 53 were considered) and assigned to respective protein accession numbers from the *P. chrysosporium* genome database. Figure 6. A and B show the proteins in the master gel sections identified and annotated according to their accession numbers.

Further analysis of the data indicated that the number of distinct ORFs was 314, after excluding the ones occurring as two or more spots. In other words, 314 different gene products (distinct ORFs) could be characterized from 517 protein spots that gave statistically significant scores. Of these, as much as 118 yielded multiple spots on the master gel, corresponding to 37.5% of the all distinct ORFs identified in this work.

A



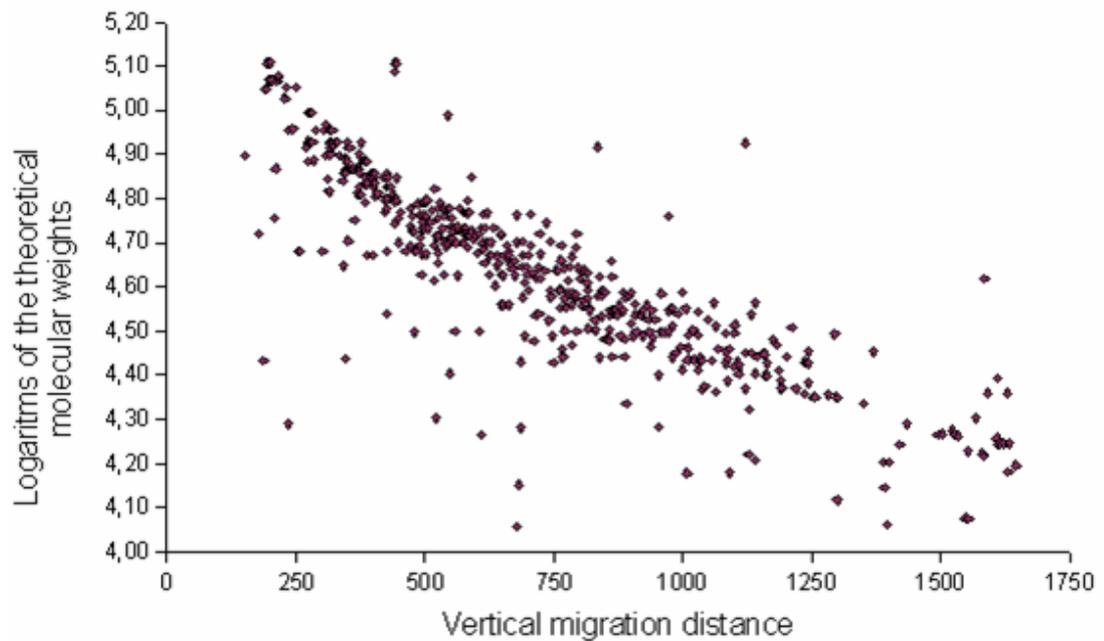


**Figure 6 A and B.** A reference 2-D map (master gel) of *P. chrysosporium*. For better viewing, the Figure is represented as two enlarged (A and B) overlapped captions.

### 3.4. Experimental versus Theoretical Data

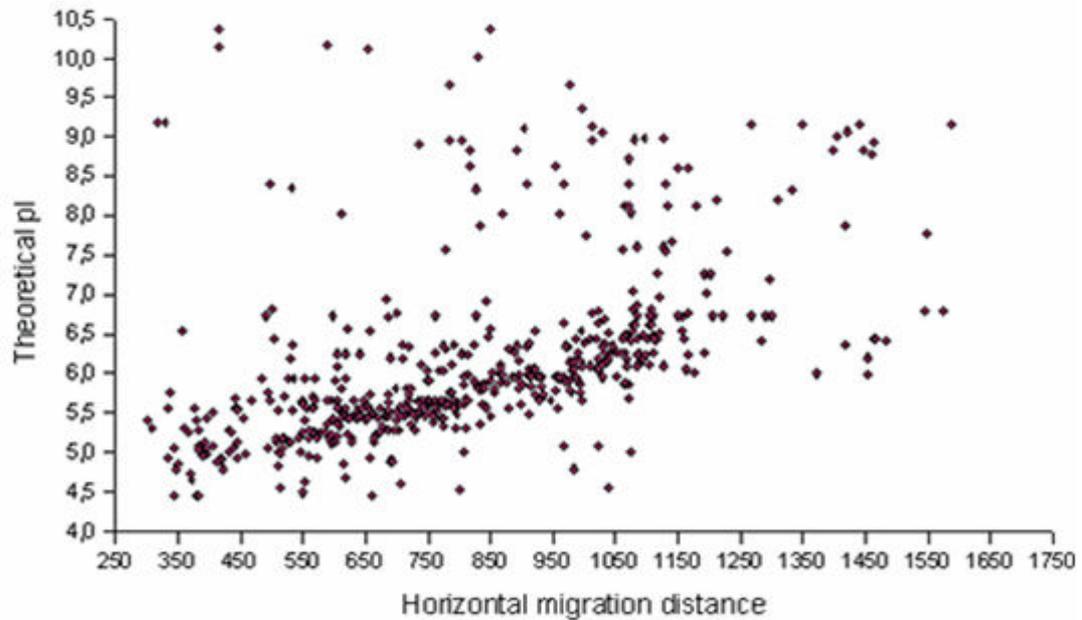
Experimental versus theoretical data analysis involved a comparison of the theoretical molecular weight (MW) of the 517 proteins and their theoretical  $pI$  with their vertical and horizontal migration distances, respectively. This analysis pointed to the existence of several proteins that strongly diverted from the graph trend-line, as explained below.

The natural logarithms of the theoretical molecular weight values of the proteins were plotted against the vertical migration distance of the spots (Figure 7). The graph revealed the existence of ca. 52 spots with an apparent molecular weight that is significantly different from their theoretical molecular weight, suggesting a major change in the protein mass due to post-translational modification. Most of such changes probably resulted from cleavage of part of the proteins. However, in some cases, a protein migrated considerably slower than expected from its molecular weight, a migration distance that corresponded to the size of its dimer. Such an observation was also made by Rosen et al. (2004) with *A. tumefaciens* proteome. The migration of a protein as an oligomer was also the case in this study, but very rare as expected, due to the strong denaturing conditions of the buffers used for 2-D electrophoresis.



**Figure 7.** Comparison of the theoretical and experimental molecular weights of the identified proteins in the *P. chrysosporium* master gel. The theoretical molecular weights were calculated using the EXPASY compute pI/MW tool ([http://www.expasy.org/cgi-bin/pi\\_tool](http://www.expasy.org/cgi-bin/pi_tool)). The migration distances of the spots was measured by Delta2-D as Y-axis coordinates (in pixels) of the spot centers on the gel image.

The theoretical  $pI$  values were plotted against the horizontal migration distance of the spots as calculated by the Delta2D as X-axis coordinate of the spots centers on the gel (Figure 8). This analysis indicated the existence of ca. 124 protein spots whose horizontal migration differed significantly from the expected migration according to the calculated  $pI$  values. The deviation from the theoretical  $pI$  values for such a high number of protein spots suggested extensive post-translational charge modifications in exponentially-growing *P. chrysosporium* cells.



**Figure 8.** Comparison of the theoretical and experimental pI of the identified proteins in the *P. chrysosporium* master gel. The theoretical pI was calculated by using the EXPASY compute pI/MW tool ([http://www.expasy.org/cgi-bin/pi\\_tool](http://www.expasy.org/cgi-bin/pi_tool)). The migration distances of the spots was measured by Delta2-D as X-axis coordinates (in pixels) of the spot centers on the gel image.

### 3.5. Proteins Appearing in Multiple Spots

While protein modification could be predicted by experimental versus theoretical data analysis, the main support of this was the presence of multiple spots of the same gene product. Such spots corresponding to 118 distinct ORFs as a total are summarized in Table 1. The migration of the products of the same gene to several distinct spots suggested different properties of the proteins, due to either mistranslation or post-translational modification events, such as cleavage or chemical modification.

For 38 gene products, the spots were found to differ in their charge, for 10 gene products the spots were found to differ in their apparent molecular weight and for 70 gene products the spots were found to differ both in their charge and their apparent molecular weight. These modifications remain to be characterized. Although the phosphoproteins of *P. chrysosporium* were detected on the 2-D gels by applying the ProQ phosphoprotein stain (Section

3.9), at present, we know very little about the proteins modified biologically since artificial chemical modification is not ruled out.

**Table 1.** Overview of the proteins which appear in multiple spots

Accession number	Modification*	Number of spots	Accession number	Modification*	Number of spots
563	C-M	3	127677	C_M	6
1324	C	2	127721	C_M	2
1407	M	2	128865	C_M	2
1474	C-M	2	128908	M	2
1827	C	2	129706	C	3
3052	C	2	129766	C-M	2
3216	C C-M	3	129956	M C-M	3
3727	C	2	130118	M C-M	3
4270	C-M	2	130203	C-M	2
4796	C-M	2	131042	C C-M	3
4883	C-M	2	131104	C-M	2
5061	C-M	3	131257	C-M	2
5348	C M C-M	4	131983	C-M	2
6357	C	2	132162	C C_M	4
6612	C C-M	3	132198	C M C-M	8
6758	C-M	2	132303	C-M	2
7166	C C-M	5	132370	C-M	2
7552	C-M	2	132607	C-M	3
8565	C C-M	3	132820	M	2
8807	C-M	3	133289	C	2
8935	C	2	133471	C-M	2
9041	C	2	133590	C	2
9110	C	2	133924	C-M	4
9263	C	2	134115	C M	5
9267	C-M	2	134348	C	3
10009	C	2	134368	C_M	2
10047	C C-M	4	134605	C	2
10308	C-M M	4	134615	C-M	2
10340	C	7	134635	M	2
10373	C-M	2	135337	M C-M	5
10433	C-M	2	135471	C C-M	3
10723	M	2	135576	C	2
27080	C	2	135608	C	2
28980	C-M	2	135943	C-M	2
29242	C M	3	136199	C-M	2
37522	C	2	136233	C-M	2
39727	C-M	3	136462	C-M	2
41034	C-M	2	136680	C-M	2
41793	C C-M	3	136730	C-M	2
42666	C	2	137299	C	3

**Table 1.** continued

Accession number	Modification*	Number of spots	Accession number	Modification*	Number of spots
122435	C_M	2	137498	C	3
122440	C M	6	137623	C C-M	4
122794	C	4	137747	C	2
123005	C	2	137931	C	2
123314	C-M	2	138286	M	2
123444	C-M	2	138298	C-M	4
123500	C C-M	3	138311	M	2
123502	C C-M	5	138589	C	3
123932	M	2	138721	C-M	2
124432	C-M	4	138775	C C-M	5
124963	C	5	138887	C M	3
125300	C	2	139320	C	2
125674	M C-M	3	139663	C	4
126013	M	2	139756	C-M	2
126051	C C-M	4	140413	C-M	2
126443	C	4	140431	C	2
127046	C_M	2	140491	C	3
127435	C_M	2	140559	C-M	2
127531	M	2	140617	C-M	3

\* C, charge difference; M, mass difference; C-M, both charge and mass difference

### 3.6. Functional Assignments

As explained in Section 3.3, the protein accession number of each of the identified proteins was used to make search in the publicly available *P. chrysosporium* database in order to assign a putative function. The 314 proteins identified and marked on the reference map (Figure 6. A, B) are listed in supplementary Table 2 to show their functions. The pattern of identified proteins was typical of exponentially growing cells.

**Table 2.** Predicted functions of the proteins annotated in the master gel

Accession number	pI	Mr (Da)	Hit score	Gene scaffold	Putative function	Subcellular localization	Signal peptide
62	5,65	36155,93	<b>134</b>	pg.C scaffold 1000062	26S proteasome regulatory subunit rpn-8.	C	
265	6,43	45763,27	<b>199</b>	pg.C scaffold 1000265	Aspartate/other aminotransferase	C	
361	5,29	46691,31	<b>127</b>	pg.C scaffold 1000361	AAA ATPase	N	
425	7,76	28826,09	<b>139</b>	pg.C scaffold 1000425	Short-chain dehydrogenase/reductase SDR	M	+
465	4,61	31452	<b>58</b>	pg.C scaffold 1000465	DNA-directed RNA polymerase	N	
503	8,97	26713,79	<b>78</b>	pg.C scaffold 1000503	Glutathione S-transferase, C-terminal	C	
511	5,66	41214,98	<b>108</b>	pg.C scaffold 1000511	Enoyl-CoA hydratase/isomerase	C	
563	8,41	44162,29	<b>122</b>	pg.C scaffold 1000563	Pyruvate dehydrogenase E1 component alpha subunit	M	
582	6,39	57657,31	<b>87</b>	pg.C scaffold 1000582	Anthranilate synthase.	C	
843	6,76	18066,15	<b>203</b>	pg.C scaffold 1000843	TonB-dependent receptor protein	N	
871	5,13	30015	<b>105</b>	pg.C scaffold 1000871	E3 binding	C	
1076	6,65	32646,64	<b>178</b>	pg.C scaffold 2000054	Fumarylacetoacetate (FAA) hydrolase	P	+
1148	5,43	32140,67	<b>104</b>	pg.C scaffold 2000126	26S proteasome non-ATPase regulatory subunit Nin1/mts3	N	
1157	8,91	84118,98	<b>358</b>	pg.C scaffold 2000135	RHO protein GDP dissociation inhibitor	M	
1164	5,74	23333,05	<b>109</b>	pg.C scaffold 2000142	Uracil phosphoribosyl transferase	C	
1173	6,35	42298,42	<b>63</b>	pg.C scaffold 2000151	Oxidoreductase, N-terminal	M	+

**Table 2.** continued

1293	5,48	84982,35	<b>142</b>	pg.C scaffold 2000271	Peptidase S9A, prolyl oligopeptidase	M	
1311	10,12	19067,53	<b>96</b>	pg.C scaffold 2000289	RNA-binding region RNP-1 (RNA recognition motif)	N	
1324	5,58	64247,41	<b>65</b>	pg.C scaffold 2000302	TPR repeat, heat shock chaperonin-binding	C	
1333	6,44	48848,43	<b>139</b>	pg.C scaffold 2000311	NAD-dependent epimerase/dehydratase	N	
1407	9,18	36217,71	<b>330</b>	pg.C scaffold 2000385	RNA-binding region RNP-1 (RNA recognition motif)	C	
1474	5,05	46954,24	<b>224</b>	pg.C scaffold 2000452	Proteasome component region PCI	C_N	
1585	5,45	33529,84	<b>164</b>	pg.C scaffold 2000563	Proteasome endopeptidase complex	M	+
1651	9,07	40576,65	<b>104</b>	pg.C scaffold 2000629	Isocitrate dehydrogenase NAD-dependent, mitochondria	M	+
1661	7,28	28408,91	<b>85</b>	pg.C scaffold 2000639	Puromycin N-acetyltransferase	M	
1827	6,34	31263,73	<b>203</b>	pg.C scaffold 2000805	20S proteasome, A and B subunits	M	
1962	5,81	50041,27	<b>146</b>	pg.C scaffold 2000940	Replication factor C conserved region	M	
2075	5,62	15663,76	<b>72</b>	pg.C scaffold 3000069	Ubiquitin-conjugating enzymes	C	
2311	4,94	18815,04	<b>109</b>	pg.C scaffold 3000305	Tropomyosins	N	
2657	10,36	20104,33	<b>91</b>	pg.C scaffold 3000651	Mitochondrial ribosomal protein L5	M	
2729	6,23	57000,51	<b>177</b>	pg.C scaffold 3000723	Probable oxidoreductase	C	
2822	7,78	38407,07	<b>306</b>	pg.C scaffold 4000081	Zinc-containing alcohol dehydrogenase superfamily	C	
2920	5,36	37742,94	<b>53</b>	pg.C scaffold 4000179	Putative dyhydroflavanol-4-reductase	C	+

**Table 2.** Continued

2978	5,20	61489,56	<b>95</b>	pg.C scaffold 4000237	Chaperonin ATPase.	C	
3052	6,35	53942,8	<b>96</b>	pg.C scaffold 4000311	Nucleotidyl transferase, Mannose-6-phosphate isomerase, type II	Cs	
3168	6,03	49858,23	<b>227</b>	pg.C scaffold 4000427	Transforming growth factor beta	C	
3179	5,83	26242,79	<b>56</b>	pg.C scaffold 4000438	Hypothetical protein	C	
3216	5,85	78822,96	<b>141</b>	pg.C scaffold 4000475	tRNA synthetases, class II (G, H, P and S), Prolyl-tRNA synthetase, class Iia	C	<b>+</b>
3583	6,17	61776	<b>58</b>	pg.C scaffold 5000145	Phosphoglucose isomerase (PGI)	C	
3623	4,62	13074,28	<b>144</b>	pg.C scaffold 5000185	hypothetical protein, elongation factor EF-Tu, tufA	N	
3711	6,80	38183,51	<b>64</b>	pg.C scaffold 5000273	26S proteasome non-ATPase regulatory subunit	C	
3727	6,61	42914,69	<b>61</b>	pg.C scaffold 5000289	GDP-mannose 4,6-dehydratase	C	
4077	6,51	46655,61	<b>122</b>	pg.C scaffold 5000639	Adenylosuccinate synthetase	C	
4270	5,62	57768,75	<b>162</b>	pg.C scaffold 7000083	Metalloenzyme, phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	C	
4311	4,93	23029,09	<b>111</b>	pg.C scaffold 7000124	Amidase from nicotinamidase family	C	
4345	6,81	36013,2	<b>112</b>	pg.C scaffold 7000158	Arp2/3 complex, 34kDa subunit p34-Arc	N	
4500	6,48	62830,51	<b>105</b>	pg.C scaffold 7000313	G-protein beta WD-40 repeat, corin like proteinon	E	
4580	6,37	37621,24	<b>83</b>	pg.C scaffold 7000393	Adipophilin (adipose differentiation-related protein)	C	
4796	7,54	37859,65	<b>189</b>	pg.C scaffold 7000609	Zinc-containing alcohol dehydrogenase superfamily	C	
4883	5,79	27190,1	<b>336</b>	pg.C scaffold 8000058	Short-chain dehydrogenase/reductase SDR	M	

**Table 2.** Continued

5061	5,57	82540,98	<b>136</b>	pg.C scaffold 8000236	Dipeptidyl-peptidase V precursor	E	+
5153	5,75	37632,65	<b>146</b>	pg.C scaffold 8000328	Putative dyhydroflavanol-4-reductase	C	+
5211	5,43	21646,73	<b>111</b>	pg.C scaffold 8000386	Suppressor protein SRP40	M	+
5240	5,96	37391,83	<b>182</b>	pg.C scaffold 8000415	WD-40 repeat	N	
5348	5,55	35596,9	<b>58</b>	pg.C scaffold 8000523	Hypothetical protein	M	
5470	6,54	19452,18	<b>149</b>	pg.C scaffold 9000074	ATP synthase subunit h	M	
5586	6,32	92942,82	<b>116</b>	pg.C scaffold 9000190	Heat shock protein Hsp70,	E.R.	+
5721	5,93	121804,59	<b>111</b>	pg.C scaffold 9000355	Glycosyl S-transferase 35	C	
5811	4,82	76468,06	<b>53</b>	pg.C scaffold 9000415	Required for invasion and pseudohyphae formation in response to nitrogen starvation	N	
5826	4,55	25298,32	<b>193</b>	pg.C scaffold 9000430	Elongation factor 1, beta/beta'/delta chain, Inorganic pyrophosphatase	C	
5864	6,25	52338,34	<b>129</b>	pg.C scaffold 9000468	Glutamine synthetase, catalytic region	M	
5957	5,97	46735,94	<b>83</b>	pg.C scaffold 1000040	O-acetylhomoserine/O-acetylserine sulfhydrylase	C	
6010	5,74	63783,2	<b>81</b>	pg.C scaffold 1000093	Glucose-methanol-choline oxidoreductase	C	
6330	6,27	44842,66	<b>213</b>	pg.C scaffold 1100015	NADH:flavin oxidoreductase/NADH oxidase	C	
6357	6,35	54079,82	<b>177</b>	pg.C scaffold 1100042	6-phosphogluconate dehydrogenase, C-terminal	C	
6364	5,50	29366,45	<b>53</b>	pg.C scaffold 1100049	6-phosphogluconolactonase	C	
6460	6,32	31774,02	<b>100</b>	pg.C scaffold 11000145	Esterase/lipase/thioesterase	M	

**Table 2.** Continued

6570	8,36	51829,57	<b>177</b>	pg.C scaffold 11000255	Translation elongation factor Tu; Protein synthesis factor, GTP-binding	M	
6612	5,44	50183,13	<b>123</b>	pg.C scaffold 11000297	Rab GDI protein, probable secretory pathway GDP dissociation inhibitor 1	C	
6660	5,96	59388,58	<b>172</b>	pg.C scaffold 11000345	Phosphoglucomutase/phosphomannomutase C terminal	C	
6715	6,00	32905,18	<b>66</b>	pg.C scaffold 11000400	Aldose 1-epimerase	P	
6737	8,72	29194,68	<b>103</b>	pg.C scaffold 11000422	Ribosomal protein S3, C-terminal, Eukaryotic/archaeal ribosomal protein S3	M	
6758	6,05	62682,5	<b>230</b>	pg.C scaffold 11000443	Aspartyl-tRNA synthetase, class IIb	N	
6867	5,36	25157,44	<b>144</b>	pg.C scaffold 12000051	Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen, similar to peroxiredoxin Q	M	
6915	4,68	34935,6	<b>60</b>	pg.C scaffold 12000099	F-actin capping protein, alpha subunit	N	
7072	5,92	73773,92	<b>156</b>	pg.C scaffold 12000256	Glycyl-tRNA synthetase, alpha2 dimer, tRNA synthetases, class-II (G, H, P and S)	C	
7113	8,99	31142,96	<b>86</b>	pg.C scaffold 12000297	2Fe-2S ferredoxin, iron-sulfur binding site, Ferredoxin	M	+
7166	5,65	72054,86	<b>443</b>	pg.C scaffold 12000350	Heat shock protein Hsp70	M	
7167	6,36 /	25748,5	<b>85</b>	pg.C scaffold 12000351	Glutathione S-transferase	C_N	
7552	5,53	30918,73	<b>122</b>	pg.C scaffold 13000340	Glycine-rich cell wall structural protein 1.8 precursor	N	
7571	6,27	35555,99	<b>126</b>	pg.C scaffold 13000359	Aldo/keto reductase	C	
7695	4,86	27620,43	<b>63</b>	pg.C scaffold 14000081	Esterase/lipase/thioesterase	C	
7971	6,97	24145,62	<b>253</b>	pg.C scaffold 15000040	Glutathione transferase	M	

**Table 2.** Continued

8290	5,53	45627,45	<b>87</b>	pg.C scaffold 16000072	Heat shock protein Hsp20,	C	
8420	7,58	50322,63	<b>53</b>	pg.C scaffold 16000202	Cysteine desulfurase (Selenocysteine lyase)	C	
8527	5,43	34868,58	<b>106</b>	pg.C scaffold 17000092	Phospholipase D/Transphosphatidylase, FAD-dependent pyridine nucleotide-disulphide oxidoreductase	M	
8565	5,67	36631,49	<b>68</b>	pg.C scaffold 17000130	Short-chain dehydrogenase/reductase SDR	M	
8571	6,54	118110,26	<b>161</b>	pg.C scaffold 17000136	Glu/Leu/Phe/Val dehydrogenase, C terminal	C	
8574	5,94	50577,21	<b>180</b>	pg.C scaffold 17000139	Insulinase-like	M	+
8607	6,05	30077,44	<b>215</b>	pg.C scaffold 17000172	RNA-binding region RNP-1 (RNA recognition motif)	C_N	
8735	5,53	13968,43	<b>216</b>	pg.C scaffold 18000106	TonB box, N-terminal	M	
8737	5,64	11511,77	<b>206</b>	pg.C scaffold 18000108	TonB-dependent receptor protein	M	
8738	6,53	11919,25	<b>233</b>	pg.C scaffold 18000109	TonB-dependent receptor protein	M	
8807	5,34	23413,77	<b>363</b>	pg.C scaffold 19000036	Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	C	
8935	4,99	22425,29	<b>272</b>	pg.C scaffold 20000025	GCN5-related N-acetyltransferase	M	
8983	4,88	32414,2	<b>101</b>	pg.C scaffold 20000073	Adhesion regulating molecule	M	
9041	5,74	52006,42	<b>74</b>	pg.C scaffold 21000008	Aldose 1-epimerase	E	+
9110	5,67	31772,29	<b>178</b>	pg.C scaffold 21000077	Glucosamine-6-phosphate isomerase	M	
9263	5,56	82299,45	<b>127</b>	pg.C scaffold 23000011	Spermine synthase, saccharopine dehydrogenase	C	
9267	6,80	22859,85	<b>210</b>	pg.C scaffold 23000015	Manganese and iron superoxide dismutase	P	

**Table 2.** Continued

9834	5,90	76667,32	<b>164</b>	1000004 pm.C scaffold	Tryptophan synthase, alpha chain, pyridoxal- 5'-phosphate-dependent enzyme, beta subunit	C	
9905	6,26	59882,82	<b>85</b>	pm.C scaffold 1000075	Chaperonin ATPase.	C	
9983	6,56	39119,96	<b>92</b>	pm.C scaffold 2000026	Isocitrate/isopropylmalat e dehydrogenase	M	
10009	5,89	25048,65	<b>269</b>	pm.C scaffold 2000052	Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	C	
10011	5,69	82917,51	<b>83</b>	pm.C scaffold 2000054	D-xylulose 5- phosphate/D-fructose 6- phosphate phosphoketolase	Cs	
10047	5,18	49212,33	<b>147</b>	pm.C scaffold 2000090	Tubulin/FtsZ, C-terminal	Cs	
10261	6,54	47054,09	<b>54</b>	pm.C scaffold 4000073	Isocitrate dehydrogenase NADP- dependent, eukaryotic	Cs	
10308	5,66	47250,19	<b>162</b>	pm.C scaffold 5000040	S-adenosyl-L- homocysteine hydrolase	C	
10340	5,43	84821,13	<b>257</b>	pm.C scaffold 6000005	Heat shock protein Hsp72	C	
10355	7,05	43339,47	<b>155</b>	pm.C scaffold 6000020	GHMP kinase, Diphosphomevalonate decarboxylase	M	
10373	5,80	34647,02	<b>76</b>	pm.C scaffold 6000038	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1, WD-40 repeat	C_N	
10433	5,82	70314,27	<b>106</b>	pm.C scaffold 7000011	Transketolase	C	
10446	4,86	26413,15	<b>63</b>	pm.C scaffold 7000023	Thioredoxin-related	Cs	
10451	6,44	43445,71	<b>63</b>	pm.C scaffold 7000028	Conserved hypothetical protein 92, GTP1/OBG	C	
10554	6,14	116217,35	<b>141</b>	pm.C scaffold 8000032	ATP/GTP-binding site motif A (P-loop), ABC transporter, HPr serine phosphorylation site	C	+
10601	6,12	119578,32	<b>69</b>	pm.C scaffold 9000025	Leucyl-tRNA synthetase archae/euk cytosolic, class Ia, Aminoacyl- tRNA synthetase,	C	

**Table 2.** Continued

10653	4,45	47653,62	<b>80</b>	pm.C scaffold 10000028	Nucleosome assembly protein (NAP)	N	+
10711	5,36	32414,23	<b>124</b>	pm.C scaffold 11000052	Helix-turn-helix, Fis-type	N	+
10723	5,23	66532,77	<b>76</b>	pm.C scaffold 11000064	Serine/threonine specific protein phosphatase	Cs	
10895	4,66	49007,72	<b>158</b>	pm.C scaffold 16000011	ADP-ribosylation factor	C_N	
26450	5,75	56331,41	<b>109</b>	gwh2.10.6 9.1	Serine/threonine specific protein phosphatase.	C	
27080	5,45	65379,47	<b>139</b>	gwh2.1.30 7.1	HMG-CoA lyase-like, Yeast 2-isopropylmalate synthase, Alpha-isopropylmalate/homocitrate synthase	C	
28980	6,18	31600,53	<b>74</b>	gwh2.11.2 04.1	RNA recognition, region 1	Cs	
29242	6,10	41533,97	<b>359</b>	gwh2.9.25 3.1	Ubiquinol-cytochrome-c reductase complex core protein 2, Insulinase-like	C	
32922	9,33	19942,73	<b>369</b>	gwh2.6.52 3.1	H <sup>+</sup> -transporting two-sector ATPase, delta (OSCP) subunit	M	
33415	4,79	15108,05	<b>70</b>	gwh2.1.10 50.1	Nascent polypeptide-associated	M	+
33865	8,02	15062,11	<b>74</b>	gwh2.7.66 2.1	Actin-binding, cofilin/tropomyosin type	N	
35305	6,04	122232,36	<b>67</b>	gww2.9.26 .1	Glycoside hydrolase, family 38	C	
35310	5,58	106524,09	<b>82</b>	gww2.4.22 .1	Glycoside hydrolase, family 31	C	
37522	4,89	46977,65	<b>103</b>	gww2.17.3 4.1	Glycoside hydrolase, family 21	C	
39635	7,20	43909,96	<b>254</b>	gww2.11.2 31.1	Aspartate/other aminotransferase	P	
39727	5,45	44997,1	<b>188</b>	gww2.9.26 6.1	Elongation factor 1, gamma chain	C	
40438	6,24	41475,39	<b>102</b>	gww2.1.70 6.1	G-protein, beta subunit, WD-40 repeat	E	
41034	8,03	33383,89	<b>121</b>	gww2.1.75 5.1	Porin, eukaryotic type	E	
41793	5,94	30315,12	<b>62</b>	gww2.7.51 3.1	Uricase	C	
42666	5,72	16573,65	<b>232</b>	gww2.2.79 0.1	RanBP1, EVH1,	N	
42947	5,68	19187,8	<b>61</b>	gww2.1.98 5.1	GrpE protein	C	
121028	6,05	21615,67	<b>71</b>	gwh2.5.44 6.1	Flavodoxin/nitric oxide synthase	M	

**Table 2.** Continued

121237	6,13	52186,04	<b>163</b>	gwh2.5.19 2.1	Aminotransferase, class I and II	C	
121589	6,50	39370,44	<b>57</b>	gwh2.22.8 2.1	G-protein, beta subunit, WD-40 repeat	M	
122095	6,09	46792,15	<b>86</b>	gwh2.9.26 3.1	Flavoprotein pyridine nucleotide cytochrome reductase, Oxidoreductase FAD/NAD(P)-binding	C	
122404	6,16	28047,14	<b>89</b>	gwh2.1.88 5.1	Lipolytic enzyme, G-D-S- L	C_N	
122435	6,44	44458,31	<b>211</b>	gwh2.1.61 2.1	Phosphoglycerate kinase	C	
122440	5,46	67177,01	<b>450</b>	gwh2.1.30 1.1	Heat shock protein Hsp70	C	
122794	5,95	73751,47	<b>245</b>	gwh2.1.22 2.1	Bacterial transketolase	C	
122894	6,14	116299,5	<b>81</b>	gwh2.1.56. 1	ABC transporter	C	
122979	8,34	45521,02	<b>59</b>	gwh2.1.60 7.1	Acyl-CoA dehydrogenase, central region	M	
123005	7,61	72220	<b>198</b>	gwh2.1.21 8.1	Polyadenylate binding protein, human types 1, 2, 3, 5	Cs	
123074	7,87	16672,21	<b>101</b>	gwh2.1.10 71.1	Nucleoside diphosphate kinase	M	
123314	6,23	24372,91	<b>95</b>	gwh2.1.99 0.1	Ras GTPase, GTP- binding nuclear protein Ran	C	
123444	5,77	58106,3	<b>53</b>	gwh2.1.38 0.1	H <sup>+</sup> -transporting two- sector ATPase, alpha/beta subunit, central region	M	+
123500	5,58	68119,83	<b>341</b>	gwh2.1.31 4.1	H <sup>+</sup> -transporting two- sector ATPase, alpha/beta subunit, central region	C	
123502	5,13	70597,16	<b>63</b>	gwh2.1.24 7.1	Peptidase, eukaryotic cysteine peptidase active site	C	
123770	6,20	36357,46	<b>122</b>	gwh2.6.32 6.1	Amino acid-binding ACT	M	
123932	7,26	62606,1	<b>244</b>	gwh2.6.13 1.1	Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal	C	+
124040	6,47	28228,47	<b>70</b>	gwh2.6.47 5.1	HAD-superfamily hydrolase, subfamily IB (PSPase-like)	N	
124260	4,94	22691,02	<b>162</b>	gwh2.12.2 67.1	20S proteasome, A and B subunits	Cs	

**Table 2.** Continued

124390	5,99	26895,84	<b>252</b>	gwh2.12.2 71.1	Triose phosphat isomerase, H+- transporting two-sector ATPase, delta (OSCP) subunit	M	
124398	6,73	57905,33	<b>181</b>	gwh2.12.1 14.1	Catalase	C_N	
124432	6,14	116299,5	<b>124</b>	gwh2.12.1 16.1	ABC transporter	C	
124435	10,0 2	14148,49	<b>81</b>	gwh2.12.3 35.1	Ribosomal protein L14b/L23e	M	
124453	5,74	128170,3	<b>64</b>	gwh2.12.1 3.1	Carbamoyl-phosphate synthase L chain, ATP- binding	M	
124485	7,87	26933,62	<b>198</b>	gwh2.12.2 63.1	Short-chain dehydrogenase/reductas e SDR, Glucose/ribitol dehydrogenase	M	
124677	6,11	23530,69	<b>239</b>	gwh2.11.3 88.1	Rieske iron-sulfur protein, Ubiquinol cytochrome reductase transmembrane region	M	
124898	4,72	31633,26	<b>84</b>	gwh2.11.3 25.1	TPR repeat	C_N	
124963	5,66	98603,56	<b>384</b>	gwh2.11.4 1.1	AAA ATPase	C	
125048	5,03	22720,47	<b>73</b>	gwh2.11.4 07.1	HSP20-like chaperone	C	
125300	5,40	15957,4	<b>193</b>	gwh2.27.3 0.1	Ribosomal protein S12E	C	
125657	5,31	18401,05	<b>97</b>	gwh2.10.2 85.1	Putative peroxiredoxin (Thioredoxin reductase) (Allergen Mal f 3)	C	
125674	4,79	50130,35	<b>205</b>	gwh2.10.1 56.1	Tubulin/FtsZ, C-terminal	N	
126013	6,04	59874,89	<b>73</b>	gwh2.13.1 25.1	Chaperonin Cpn 60/TCP- 1	C	
126051	5,19	54463,7	<b>107</b>	gwh2.13.1 15.1	UTP--glucose-1- phosphate uridylyltransferase	M	
126415	6,71	16398,81	<b>89</b>	gwh2.7.65 7.1	Ubiquitin-conjugating enzymes	C	
126443	5,96	90074,56	<b>230</b>	gwh2.7.87. 1	Aconitate hydratase, N- terminal, 3- isopropylmalate dehydratase large subunit	C	
126584	6,19	46719,21	<b>53</b>	gwh2.7.36 1.1	NADH:flavin oxidoreductase/NADH oxidase	C	
126671	5,81	48420,68	<b>123</b>	gwh2.7.32 7.1	Actin/actin-like	Cs	
126867	5,38	37632,06	<b>358</b>	gwh2.7.40 8.1	Homoserine dehydrogenase	E	<b>+</b>

**Table 2.** Continued

127046	6,37	58444,26	<b>53</b>	gwh2.14.1 22.1	Helicase, C-terminal	N	
127115	5,25	25321,65	<b>132</b>	gwh2.14.2 02.1	SAM (and some other nucleotide) binding motif, O-methyltransferase, family 3	C	
127345	8,79	37126,04	<b>88</b>	gwh2.3.40 7.1	2-nitropropane dioxygenase, NPD	P	
127435	5,00	90194,27	<b>133</b>	gwh2.3.90. 1	AAA ATPase	N	
127531	6,42	64684,68	<b>123</b>	gwh2.3.17 1.1	Methylglyoxal synthase-like, AICARFT/IMPCHase bienzyme	C	
127647	4,53	11400,66	<b>135</b>	gwh2.3.63 5.1	Ribosomal protein P2	C	+
127677	5,55	47806,45	<b>172</b>	gwh2.3.33 8.1	Argininosuccinate synthase	Cs	
127721	6,11	55493,55	<b>383</b>	gwh2.3.14 3.1	G-protein beta WD-40 repeat, corin like proteinon	M	
127894	6,51	37371,12	<b>74</b>	gwh2.4.36 7.1	Zinc-containing alcohol dehydrogenase superfamily	C	
128301	6,37	29150,27	<b>96</b>	gwh2.4.44 5.1	H <sup>+</sup> -transporting two-sector ATPase, gamma subunit	M	
128516	5,68	28751,85	<b>74</b>	gwh2.4.48 2.1	Glucose/ribitol dehydrogenase	C	
128640	6,56	17456,64	<b>186</b>	gwh2.8.41 1.1	Actin-binding, cofilin/tropomyosin type	N	
128810	5,28	52165,88	<b>76</b>	gwh2.8.21 3.1	Proteasome component region PCI	N	
128865	4,95	79414,18	<b>150</b>	gwh2.8.68. 1	Heat shock protein Hsp90; ATP-binding region, ATPase-like	C_N	
128903	5,12	47363,26	<b>110</b>	gwh2.8.22 3.1	AAA ATPase; 26S proteasome subunit P45; AAA ATPase, central region	N	
128908	6,62	112587,8	<b>164</b>	gwh2.8.39. 1	Oxoglutarate Dehydrogenase, E1 component	M	
129108	5,68	90624,63	<b>68</b>	gwh2.2.10 7.1	Phosphoribosyl-AMP cyclohydrolase, Histidinol dehydrogenase	C	
129681	10,1 7	22427,04	<b>171</b>	gwh2.2.86 4.1	Ribosomal protein S7E	N	
129706	5,26	111675,77	<b>227</b>	gwh2.2.50. 1	UBA/THIF-type NAD/FAD binding fold, Ubiquitin-activating enzyme	C	

**Table 2.** Continued

129766	6,12	27108,72	<b>152</b>	gwh2.2.74 9.1	20S proteasome, A and B subunits	C	
129956	4,45	38599,45	<b>234</b>	gwh2.2.11 14.1	Peptidase aspartic, active site, Peptidase A1, pepsin	C	
130024	7,58	64810,01	<b>130</b>	gwh2.2.15 2.1	Yeast eukaryotic release factor, Protein synthesis factor, GTP-binding, Elongation factor Tu, C-terminal	M	
130118	6,06	39164,02	<b>169</b>	gwh2.2.58 8.1	Glutamine synthetase, catalytic region	C	
130148	5,49	70549,77	<b>53</b>	gwh2.2.19 9.1	Phenylalanyl-tRNA synthetase, beta subunit archae/euk cytosolic	N	
130203	5,29	33505,42	<b>202</b>	gwh2.16.1 26.1	Ribosomal protein 60S	C	
130234	8,33	40001,12	<b>68</b>	gwh2.16.9 8.1	Peptidase M24	C	
131042	8,96	17601,9	<b>216</b>	gww2.5.48 7.1	Peptidyl-prolyl cis-trans isomerase, cyclophilin type	C	
131104	5,00	52421,53	<b>170</b>	gww2.5.20 8.1	Peptidase M20	C	
131142	5,84	51636,38	<b>89</b>	gww2.5.22 9.1	UDP-glucose/GDP-mannose dehydrogenase	C	
131251	5,91	42994,6	<b>157</b>	gww2.5.31 2.1	Proteasome component region PCI	C	
131257	5,88	76811,17	<b>234</b>	gww2.5.65 .1	4Fe-4S ferredoxin, iron-sulfur binding, Respiratory-chain NADH dehydrogenase 75 kDa subunit	C	
131298	6,24	69956,03	<b>298</b>	gww2.5.12 4.1	Actin-binding, actinin-type, Calcium-binding EF-hand	C	
131464	4,90	28704,49	<b>55</b>	gww2.5.39 9.1	Proliferating cell nuclear antigen, PCNA	N	
131837	5,81	83517,86	<b>192</b>	gww2.19.3 6.1	Hydantoinase/oxoprolinase	C	
131879	6,69	51737,43	<b>181</b>	gww2.9.19 5.1	Dihydrolipoamide dehydrogenase, FAD-dependent pyridine nucleotide-disulphide oxidoreductase	C	
131933	8,93	24692,09	<b>149</b>	gww2.9.40 8.1	Manganese and iron superoxide dismutase	M	+
131983	4,93	72376,76	<b>262</b>	gww2.9.92 .1	Heat shock protein Hsp70	E	+
132162	6,45	34366,89	<b>278</b>	gww2.9.34 7.1	Malate dehydrogenase, active site	P	

**Table 2.** Continued

132198	6,73	35153,06	<b>154</b>	gww2.9.32 8.1	Glyceraldehyde-3-phosphate dehydrogenase, type I	C	
132303	5,63	43951,51	<b>153</b>	gww2.9.28 8.1	Proteasome endopeptidase complex.	C_N	
132370	5,67	42337,19	<b>104</b>	gww2.1.68 4.1	S-adenosylmethionine synthetase	C	
132421	5,19	47759,59	<b>312</b>	gww2.1.65 4.1	Helicase, C-terminal	N	
132494	5,55	76598,66	<b>247</b>	gww2.1.20 8.1	Peptidase M49	C	
132705	6,65	70226,66	<b>199</b>	gww2.1.24 6.1	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	M	
132735	9,66	26794,51	<b>156</b>	gww2.1.97 6.1	Fumarylacetoacetate (FAA) hydrolase	M	+
132767	8,64	43791,52	<b>341</b>	gww2.1.66 6.1	Acetohydroxy acid isomeroreductase	M	+
132820	8,60	28413,87	<b>106</b>	gww2.1.93 8.1	Adenylate kinase, lid region	C	
132851	6,12	27922,35	<b>82</b>	gww2.1.87 4.1	Beta Lactamase Like	C	
132918	9,00	49452,29	<b>128</b>	gww2.1.56 6.1	Sulfide:quinone oxidoreductase, mitochondrial precursor (Heavy metal tolerance protein 2) (Cadmium resistance protein 1)	M	
133055	9,36	15178,28	<b>227</b>	gww2.1.11 22.1	Cytochrome bd ubiquinol oxidase, 14 kDa subunit	M	
133066	8,41	27142,15	<b>101</b>	gww2.1.87 7.1	NADH dehydrogenase (ubiquinone), 24 kDa subunit	M	
133067	5,94	126844,9	<b>53</b>	gww2.1.45 .1	Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2	C	
133105	6,15	45518,78	<b>59</b>	gww2.1.62 6.1	Aminotransferase class V, Phosphoserine aminotransferase	M	
133187	7,02	29441,5	<b>113</b>	gww2.1.85 0.1	20S proteasome, A and B subunits	C	
133217	10,1 5	28193,92	<b>78</b>	gww2.1.92 3.1	Isopentenyl-diphosphate delta-isomerase	M	
133289	6,19	54104,82	<b>82</b>	gww2.1.45 5.1	Aldehyde dehydrogenase	C	
133367	5,49	34347,5	<b>132</b>	gww2.1.76 2.1	Uncharacterized conserved protein	M	
133471	5,86	26689,47	<b>101</b>	gww2.1.97 3.1	Adenylate kinase	Cs	
133590	5,54	106235,78	<b>316</b>	gww2.6.22 .1	Alanyl-tRNA synthetase	C	
133717	6,71	34460,25	<b>97</b>	gww2.6.47 3.1	20S proteasome, A and B subunits	M	

**Table 2.** Continued

133830	6,95	11864,54	<b>96</b>	gww2.6.58 0.1	Heat shock protein Hsp20, ibpA, n molecular chaperone (small heat shock protein)	C_N	
133884	5,49	20967,97	<b>135</b>	gww2.6.54 0.1	het-c protein	C	
133924	5,91	53293,05	<b>151</b>	gww2.6.24 3.1	Aldehyde dehydrogenase	C	
134073	6,45	59700,76	<b>94</b>	gww2.6.18 7.1	Chaperonin Cpn60	C	
134115	8,13	59956,97	<b>307</b>	gww2.6.22 9.1	UTP--glucose-1- phosphate uridylyltransferase,	C_N	
134181	6,34	43206,22	<b>187</b>	gww2.12.1 61.1	Lipocalin-related protein and Bos/Can/Equ allergen	M	
134261	5,54	36192,39	<b>153</b>	gww2.12.2 07.1	Transaldolase AB	C	
134348	5,26	61777,58	<b>154</b>	gww2.12.1 05.1	Heat shock protein 60, mitochondrial precursor (60 kDa chaperonin); Chaperonin Cpn61	M	
134368	5,75	25816,22	<b>115</b>	gww2.12.2 80.1	H <sup>+</sup> -transporting two- sector ATPase, E subunit	N	
134492	5,29	16120,09	<b>62</b>	gww2.12.3 31.1	Heat shock protein Hsp20, Heat shock protein 16 (16 kDa heat shock protein)	C_N	
134605	6,73	57686,35	<b>152</b>	gww2.11.1 58.1	Pyruvate kinase	C	
134615	5,85	30496,68	<b>85</b>	gww2.11.3 51.1	Dienelactone hydrolase	C	
134635	5,91	53963,71	<b>205</b>	gww2.11.1 76.1	G-protein beta WD-40 repeat,	C	
134652	6,86	30196,64	<b>158</b>	gww2.11.3 52.1	Band 7 protein, Prohibitin,	E	<b>+</b>
134660	9,16	50083,82	<b>112</b>	gww2.11.2 20.1	Translation elongation factor EF-1, alpha subunit	C	
134795	5,25	28902,86	<b>70</b>	gww2.11.3 62.1	Esterase/lipase/thioester ase	M	
134975	5,79	51226,91	<b>114</b>	gww2.11.2 51.1	Cell division/GTP binding protein	N	
135167	6,41	51178,2	<b>53</b>	gww2.20.4 6.1	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	C	
135337	5,48	30797,85	<b>161</b>	gww2.21.4 4.1	Ribosomal protein S2, eukaryotic and archaeal form	M	
135471	5,05	42231,62	<b>74</b>	gww2.13.2 24.1	Cell division/GTP binding protein	M	<b>+</b>

**Table 2.** Continued

135576	5,43	35719,07	<b>251</b>	gww2.13.3 00.1	Bacterial/Archaeal inorganic pyrophosphatase	N	
135608	5,58	41964,65	<b>191</b>	gww2.13.2 48.1	Peptidase aspartic, active site, Peptidase A1, pepsin	E	+
135709	9,11	38673,82	<b>77</b>	gww2.13.2 87.1	Flavoprotein pyridine nucleotide cytochrome reductase, NADH:cytochrome b5 reductase (CBR), Oxidoreductase FAD/NAD(P)-binding	C	+
135713	7,68	54122,04	<b>78</b>	gww2.13.1 55.1	Fumarate hydratase, class II	M	
135734	5,56	47387,72	<b>230</b>	gww2.13.1 80.1	Enolase	C	
135867	6,08	16899,41	<b>97</b>	gww2.10.3 35.1	Ubiquitin-conjugating enzymes	P	
135943	5,63	53041,34	<b>72</b>	gww2.10.1 30.1	Peptidase M17, cytosol aminopeptidase	N	
136064	8,99	97266,1	<b>145</b>	gww2.10.3 8.1	Protein of unknown function DUF619	M	
136199	5,28	28151,92	<b>80</b>	gww2.7.55 8.1	HAD-superfamily hydrolase, subfamily IB (PSPase-like), Haloacid dehalogenase-like hydrolase	C	
136233	5,14	18467,68	<b>248</b>	gww2.7.65 2.1	Nascent polypeptide- associated complex NAC, similar to RIKEN cDNA 5730434I03 gene	C	
136359	9,13	55670,22	<b>88</b>	gww2.7.25 9.1	Citrate synthase	M	
136462	5,85	39089,4	<b>114</b>	gww2.7.42 6.1	Ketose-bisphosphate aldolase, class-II, Fructose-bisphosphate aldolase, class II, yeast/E. coli subtype	C	
136604	5,38	42429,32	<b>54</b>	gww2.7.36 0.1	Disulphide isomerase, Thioredoxin-related	C	+
136680	6,26	77041,71	<b>270</b>	gww2.7.99 .1	Glucosamine-fructose-6- phosphate aminotransferase, Sugar isomerase (SIS), Bacterial extracellular solute-binding protein, family 3,	Cs	
136730	5,51	56883,98	<b>71</b>	gww2.7.20 7.1	Pyridoxal-5'-phosphate- dependent enzyme, beta subunit	C	
136828	4,49	27683,95	<b>57</b>	gww2.7.54 7.1	Mitochondrial glycoprotein	M	+

**Table 2.** Continued

136966	5,77	70336,92	<b>65</b>	gww2.14.3 7.1	Peptidase M1, membrane alanine aminopeptidase, Peptidase M, neutral zinc metallopeptidases, zinc-binding site	C	
137263	9,06	50311,31	<b>144</b>	gww2.3.36 3.1	26S proteasome subunit P45	C	
137299	9,15	58748,43	<b>173</b>	gww2.3.21 1.1	ATP Synthase F1, H <sup>+</sup> - transporting two-sector ATPase, alpha/beta subunit, central region	M	
137445	6,02	34132,87	<b>87</b>	gww2.3.45 9.1	SAICAR synthetase	M	
137498	5,98	127783,96	<b>189</b>	gww2.3.52 .1	HMG-CoA lyase-like, Carbamoyl-phosphate synthase L chain, ATP- binding, Biotin-binding site	M	
137623	5,96	61675,62	<b>444</b>	gww2.3.19 6.1	Phosphoglucomutase/ph osphomannomutase C terminal	C	
137747	5,94	17491,4	<b>269</b>	gww2.3.60 0.1	Heat shock protein Hsp20	C_N	
137780	5,56	23737,26	<b>98</b>	gww2.3.51 8.1	Mitochondrial ATP synthase B chain	M	
137931	6,81	52201,6	<b>115</b>	gww2.4.23 6.1	Glycine hydroxymethyltransferas e	C	
138097	5,61	40734	<b>110</b>	gww2.4.34 2.1	Peptidyl-prolyl cis-trans isomerase, cyclophilin type	C	
138100	8,04	73406,69	<b>104</b>	gww2.4.90 .1	Homoaconitase	M	<b>+</b>
138174	6,01	39043,44	<b>90</b>	gww2.4.36 0.1	Peptidase aspartic, active site, Oxidoreductase, N- terminal	E	
138286	6,03	58731,71	<b>169</b>	gww2.4.14 7.1	Chaperonin ATPase.	M	
138311	6,26	51398,65	<b>66</b>	gww2.4.25 5.1	tRNA synthetases, class- II (G, H, P and S), Seryl-tRNA synthetase, class Iia	C	
138353	6,25	69245	<b>277</b>	gww2.4.11 9.1	AMP-dependent synthetase and ligase	C	
138589	5,62	64774,3	<b>181</b>	gww2.8.12 1.1	Pyruvate decarboxylase.	C	
138721	6,77	49437,21	<b>291</b>	gww2.8.21 7.1	Glu/Leu/Phe/Val dehydrogenase	C	
138775	8,84	31467,42	<b>265</b>	gww2.8.34 6.1	Porin, eukaryotic type	C	
138825	5,23	41139,9	<b>120</b>	gww2.8.23 1.1	Iron-containing alcohol dehydrogenase	C	

**Table 2.** Continued

138887	6,10	84478,11	<b>481</b>	gww2.8.56 .1	Aconitate hydratase, mitochondrial	C	
139223	5,35	53878,28	<b>98</b>	gww2.2.40 5.1	Hexokinase	C	
139298	6,25	52515,75	<b>57</b>	gww2.2.54 9.1	Actin/actin-like	Cs	
139320	8,21	38580,22	<b>133</b>	gww2.2.54 1.1	Isocitrate dehydrogenase NAD- dependent, mitochondrial	P	
139368	4,97	49284,55	<b>150</b>	gww2.2.45 1.1	Tubulin/FtsZ, C-terminal	Cs	
139381	6,42	22368,76	<b>95</b>	gww2.2.83 5.1	hypothetical protein, conserved	M	
139401	6,41	18261,86	<b>310</b>	gww2.2.91 0.1	ATP synthase D chain, mitochondrial	C	
139500	5,01	27315,81	<b>246</b>	gww2.2.76 5.1	14-3-3 protein	N	
139515	6,08	75186,63	<b>71</b>	gww2.2.15 6.1	Peptidase S16, Ion protease, AAA ATPase, central region, Cell division protein 48, CDC48, domain 2	Cs	
139663	6,26	84830	<b>291</b>	gww2.2.12 4.1	Methionine synthase, vitamin-B12 independent, 5- methyltetrahydropteroylt riglutamate-- homocysteine S- methyltransferase	C	
139756	6,50	55871,76	<b>265</b>	gww2.2.33 5.1	Peptidase, eukaryotic cysteine peptidase active site	C	
139792	6,20	27586,55	<b>53</b>	gww2.2.11 22.1	20S proteasome, A and B subunits	C	
139859	6,92	62479,23	<b>63</b>	gww2.2.23 3.1	1-(5-Phosphoribosyl)-5- amino-4-imidazole- carboxylate (AIR) carboxylase	M	
139942	6,20	27586,55	<b>168</b>	gww2.2.65 7.1	26S proteasome non- ATPase regulatory subunit 14	Cs	
140235	6,43	77730,25	<b>103</b>	gww2.16.2 3.1	Methionyl-tRNA synthetase, class Ia	C	
140361	6,53	42600,27	<b>181</b>	gww2.25.2 2.1	Chorismate synthase	C	
140413	5,31	26942,8	<b>193</b>	gww2.18.1 11.1	20S proteasome, A and B subunits	C_N	
140431	6,39	36256,3	<b>103</b>	gww2.18.8 6.1	Putative dioxygenase C576.01c	C	
140491	5,51	116916,73	<b>101</b>	gww2.18.1 .1	Animal haem peroxidase	E	
140559	5,09	27550,2	<b>166</b>	gww2.17.1 09.1	EB1 protein	C	

**Table 2.** Continued

140617	5,97	38084,35	<b>134</b>	gww2.17.6 4.1	Coproporphyrinogen III oxidase	C	
140701	5,82	31991,98	<b>310</b>	gww2.15.1 45.1	hypothetical protein PA2095	N	

\* C: Cytoplasm, Cs: Cytoskeleton, C\_N: Cytoplasm or Nucleus. E: Extracellular, M: Mitochondria, N: Nucleus, P: Peroxisome

Only 6 of the proteins listed in Table 2 were assigned to hypothetical proteins. EXPASY Blast search for these proteins indicated that none of them are unique to *P. chrysosporium*. 5 of them had conserved domain matches to hypothetical proteins from other organisms, and one was very recently identified as an acetyltransferase from *A. fumigatus*. On the other hand, our study indicated that genes coding for hypothetical proteins are actually expressed under normal physiological conditions and consequently have to be considered relevant for the physiology of the *P. chrysosporium*. The first experimental evidence on the expression of such proteins under normal physiological conditions was provided by Gade *et al.* (2005) for a prokaryotic organism, the marine bacterium *Rhodopirellula baltica*. The investigators reported that about 18% of the identified proteins represented predicted hypothetical proteins. Such proteins constituted only 2% of the identified proteins in exponentially-growing *P. chrysosporium* cells. Yet, *P. chrysosporium* genome project revealed 331 gene models that belong to the “function unknown” class (Section 3.7), therefore proteome analyses at different stages of growth and development of the organism are expected to detect much more hypothetical proteins.

### **3.7. Relative Quantitative Distribution of the Proteins into Functional Categories**

The distribution of the identified proteins among the functional categories (KOG Classification as based on EuKaryotic Orthologous Groups) listed in the JGI website was searched by entering keywords. The four main KOG

headings in this classification are "Cellular Processes and Signaling", "Information Storage and Processing", "Metabolism" and "Poorly Characterized" which are being represented by 2132, 1456, 2044 and 1588 total gene counts in *P. chrysosporium* genome, respectively. The major protein classes and the number of respective gene models under these headings are tabulated in Table 3.

**Table 3.** KOG classification and the number of gene models at the JGI website

**CELLULAR PROCESSES AND SIGNALING**

Cell wall/membrane/envelope biogenesis	73
Cell motility	4
Posttranslational modification, protein turnover, chaperones	600
Signal transduction mechanisms	604
Intracellular trafficking, secretion, and vesicular transport	305
Defense mechanisms	133
Extracellular structures	87
Nuclear structure	100
Cytoskeleton	226
Total Gene Count	2132

**INFORMATION STORAGE AND PROCESSING**

RNA processing and modification	373
Chromatin structure and dynamics	162
Translation, ribosomal structure and biogenesis	334
Transcription	381
Replication, recombination and repair	206
Total Gene Count	1456

**Table 3.** Continued**METABOLISM**

Energy production and conversion	317
Cell cycle control, cell division, chromosome partitioning	191
Amino acid transport and metabolism	250
Nucleotide transport and metabolism	146
Carbohydrate transport and metabolism	301
Coenzyme transport and metabolism	94
Lipid transport and metabolism	282
Inorganic ion transport and metabolism	148
Secondary metabolites biosynthesis, transport and catabolism	315
Total Gene Count	2044

**POORLY CHARACTERIZED**

General function prediction only	1257
Function unknown	331
Total Gene Count	1588

The available data set was next analyzed for a correlation between relative quantitative levels of the proteins and their physiological function in exponentially growing cells. Accordingly, the relative abundance of each of the 517 identified polypeptides was calculated in terms of spot intensity (in relative volume with respect to all detected spots on the gel) by using Delta2D software. A list of the 50 most abundant proteins in *P. chrysosporium* and their relative abundance is presented in Table 4. For proteins which occurred in two or more spots, the sum of the relative intensity of such spots was taken and thus as expected, 42 out of 50 most

abundant proteins were those that gave multiple spots. The majority of the 50 most abundant proteins were found to be involved in housekeeping functions.

**Table 4.** The 50 most abundant proteins in *P. chrysosporium*

<b>Protein ID</b>	<b>Relative abundance</b>	<b>Multiple spots</b>	<b>Putative function</b>
132198	2,958	+	Glyceraldehyde-3-phosphate dehydrogenase, type I
135337	1,620	+	Ribosomal protein S2, eukaryotic and archaeal form
138775	1,304	+	Porin, eukaryotic type
123074	1,251		Nucleoside diphosphate kinase
136233	1,250	+	Nascent polypeptide-associated complex NAC, similar to RIKEN cDNA 5730434I03 gene
123444	1,069	+	H <sup>+</sup> -transporting two-sector ATPase, alpha/beta subunit, central region
122440	1,036	+	Heat shock protein Hsp70
132767	0,908	+	Acetohydroxy acid isomeroreductase
129956	0,892	+	Peptidase aspartic, active site, Peptidase A1, pepsin
131298	0,875	+	Actin-binding, actinin-type, Calcium-binding EF-hand
139298	0,875	+	Actin/actin-like
137299	0,863	+	ATP synthase F1, H <sup>+</sup> -transporting two-sector ATPase, alpha/beta subunit, central region
131933	0,857		Manganese and iron superoxide dismutase
131042	0,836	+	Peptidyl-prolyl cis-trans isomerase, cyclophilin type
140559	0,812	+	EB1 protein
2075	0,802		Ubiquitin-conjugating enzymes
130118	0,793	+	Glutamine synthetase, catalytic region
127677	0,759	+	Argininosuccinate synthase
7166	0,704	+	Heat shock protein Hsp70
137747	0,644	+	Heat shock protein Hsp20
138589	0,638	+	Pyruvate decarboxylase.
124432	0,597	+	H <sup>+</sup> -transporting two-sector ATPase, alpha/beta subunit, central region
139401	0,587	+	ATP synthase D chain, mitochondrial
140413	0,574	+	20S proteasome, A and B subunits
134115	0,544	+	UTP--glucose-1-phosphate uridylyltransferase,
132370	0,531	+	S-adenosylmethionine synthetase
134615	0,520	+	Dienelactone hydrolase
137445	0,514		SAICAR synthetase
139663	0,513	+	Methionine synthase, vitamin-B12 independent, 5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase
135608	0,509	+	Peptidase aspartic, active site, Peptidase A1, pepsin
5811	0,507		Hypothetical protein

**Table 4.** Continued

1148	0,497		26S proteasome non-ATPase regulatory subunit Nin1/mts3
1407	0,495	+	RNA-binding region RNP-1 (RNA recognition motif)
8935	0,494	+	GCN5-related N-acetyltransferase
10340	0,449	+	Heat shock protein Hsp72
4796	0,445	+	Zinc-containing alcohol dehydrogenase superfamily
135471	0,445	+	Cell division/GTP binding protein
29242	0,443	+	H <sup>+</sup> -transporting two-sector ATPase, delta (OSCP) subunit
125657	0,441	+	Thioredoxin reductase, Allergen Mal f 3
127115	0,434		SAM (and some other nucleotide) binding motif, O-methyltransferase, family 3
138887	0,432	+	Aconitate hydratase, mitochondrial
6357	0,424	+	6-phosphogluconate dehydrogenase, C-terminal
8735	0,424		TonB box, N-terminal
122794	0,419	+	Bacterial transketolase
126443	0,418	+	Aconitate hydratase, N-terminal, 3-isopropylmalate dehydratase large subunit
136462	0,413	+	Ketose-bisphosphate aldolase, class-II, Fructose-bisphosphate aldolase, class II, yeast/E. coli subtype
124963	0,405	+	AAA ATPase
132162	0,393	+	Malate dehydrogenase, active site
123502	0,385	+	Peptidase, eukaryotic cysteine peptidase active site

Figure 9 shows relative distribution of the 517 identified proteins to various functional classes of cellular physiology. The functional class "Posttranslational modifications, protein turnover, chaperones" was represented with the highest number (104) of the identified proteins. This finding was not surprising recalling extensive *pI* and MW shifts seen on the 2-D gel. "Energy production and conversion", "Carbohydrate transport and metabolism", "Amino acid transport and metabolism", and "Translation, ribosomal structure and biogenesis" were the other important classes with 64, 50, 49 and 34 polypeptides, respectively. It follows that more than half of the identified proteins perform the functions of these five classes.



Posttranslational modifications of proteins play critical roles in most cellular processes through their unique ability to cause rapid alterations in the functions of preexisting proteins, multiprotein complexes, and subcellular structures and are key regulatory events in many cellular processes, including recognition, signaling, targeting and metabolism (Jensen, 2000). The present study revealed that the most abundant proteins found in exponentially growing cells of *P. chrysosporium* are involved in posttranslational modifications, protein turnover and folding. On the other hand, such functions constitute just a class of the main category “Cellular Processes and Signaling” and when the distribution of the proteins into four main functional categories is taken into consideration (Table 5), “Metabolism” appears the most important category with a share of 50.6% among identified proteins.

**Table 5.** Distribution of the 517 proteins into four main functional categories

<b>Functional Categories</b>	<b>Number of identified spots</b>	<b>Share (%) among identified proteins</b>
METABOLISM	245	50,6
CELLULAR PROCESSES AND SIGNALING	164	30,8
INFORMATION STORAGE AND PROCESSING	67	12,8
POORLY CHARACTERIZED	41	5,8
TOTAL	517	100

### **3.8. Prediction of Subcellular Localizations and Signal Peptides**

Prediction of subcellular localization of the proteins was accomplished by submitting the amino acid sequences of the 314 identified proteins to the WoLF PSORT Protein Subcellular Localization Prediction Program (<http://wolfpsort.seq.cbrc.jp>). According the results obtained from the WoLF PSORT search, 147 proteins were predicted to be located in cytoplasm, 17 proteins were found to be located on cytoskeleton, 15 were found to be in either nucleus or cytoplasm, 1 was found to be in endoplasmic reticulum, 11

were found to be extracellular, 80 were found to be mitochondrial, 35 were found to be nuclear and 8 were found to be located in peroxisome (Table 2).

According to the signal hypothesis, the majority of secreted proteins have a signal peptide (Blobel 2000). Out of the 314 gene products shown in *P. chrysosporium*, 29 were predicted to have a signal peptide sequence according to the SignalP algorithm (<http://www.cbs.dtu.dk/services/SignalP>) (Table 2). We therefore postulate that these 29 signal peptide-containing proteins are potentially being secreted.

### **3.9. Phosphorylated Proteins**

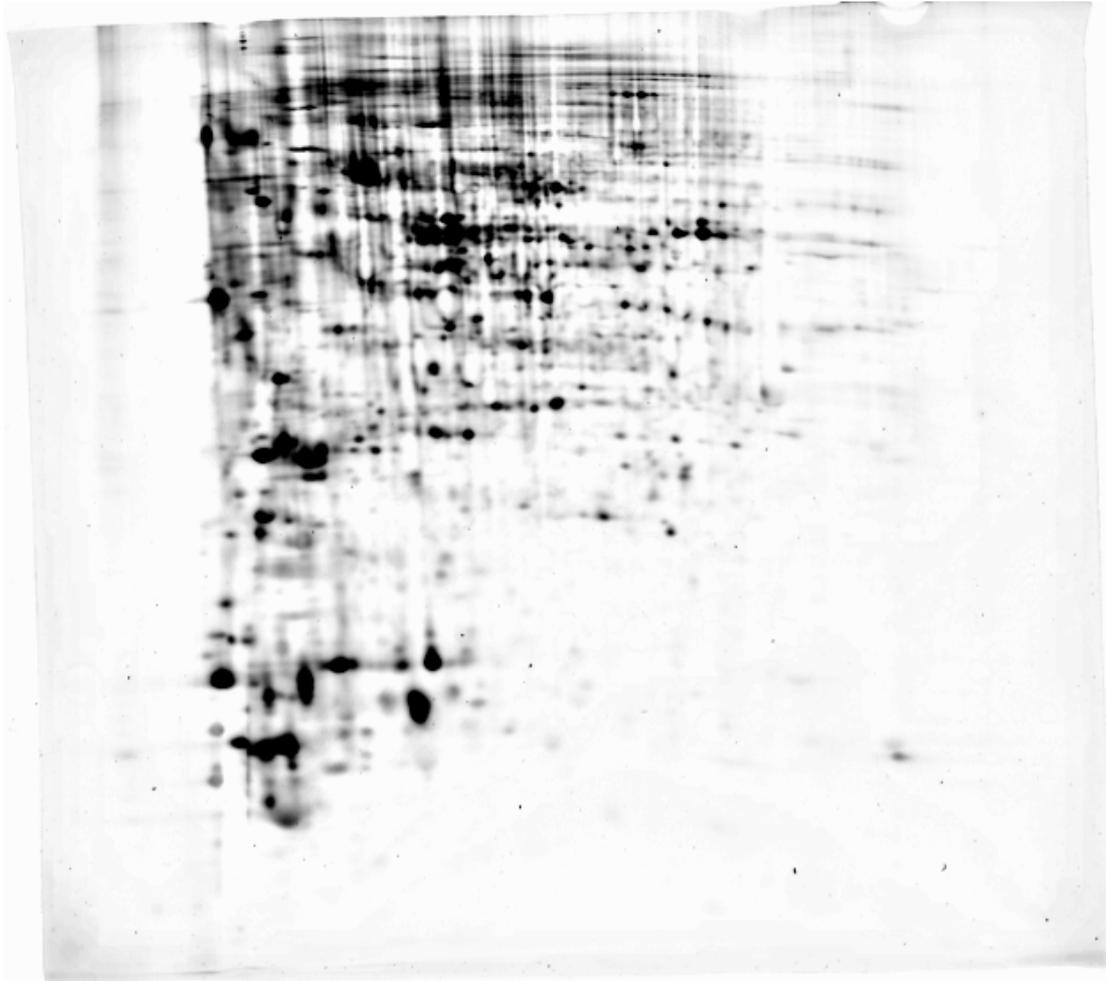
The phosphorylated proteins of *P. chrysosporium* were next detected on the 2-D gels by ProQ Diamond phosphoprotein staining (Figure 10). Spot boundary detection of the phosphoproteins by the use of Delta2D software revealed that 380 out of 910 distinct protein spots (40%) were phosphorylated in exponentially growing cells of *P. chrysosporium* (Figure 11). The image of the ProQ stained and Coomassie stained spots (on master gel) on the same gel was created by using Delta2D software and two images were warped in order to match the spots present on both gels. As can be seen in Figure 12, such spots appeared yellow whereas the ones visible only upon phosphostaining or Coomassie staining appeared red and green, respectively. The red ones missing from the master gel remain to be identified by MALDI TOF MS. Of the yellow spots that could be matched to the previously identified polypeptides on the master gel were then marked and a list of phosphorylated proteins of *P. chrysosporium* was prepared (Table 6).

Over 50% of all proteins are supposed to be phosphorylated and there are more than 100 000 estimated phosphorylation sites in the human proteome (Kalume *et al.*, 2003). Indeed, about 30% of eukaryotic proteins are

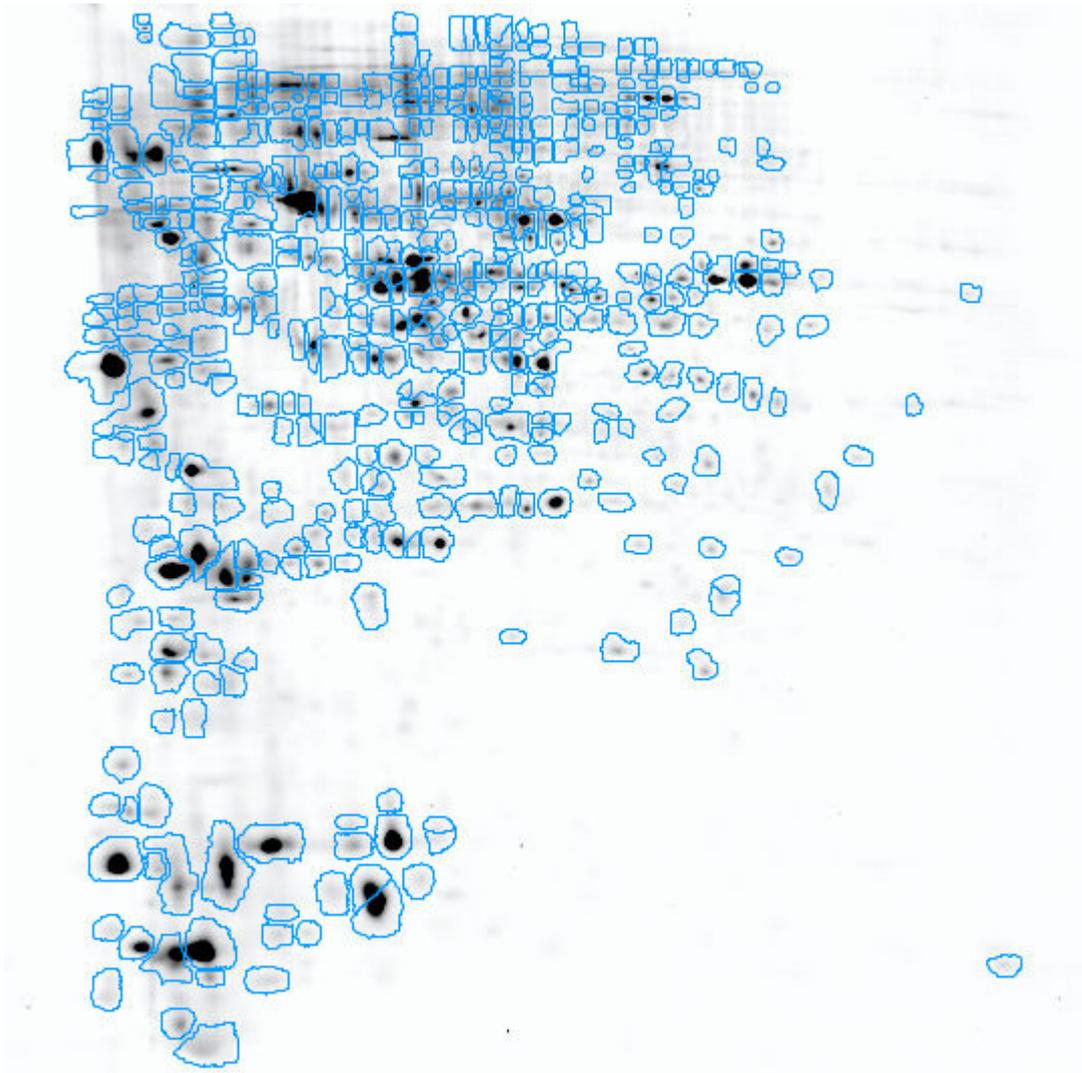
supposed to undergo phosphorylation or dephosphorylation processes (Blom *et al.*, 1998). As to prokaryotic protein phosphorylation, a recent study on *Corynebacterium glutamicum* phosphoproteome indicated that 22% of whole proteome is phosphorylated (Bendt *et al.*, 2003).

Interest in phosphoproteome is apparently growing nowadays, but there has been only a few literature reports published in this field. Ficarro *et al.* (2002) made phosphoproteome analysis of yeast cells. The investigators assumed that if 30% of expressed proteins contain at least one covalently bound phosphate, the total number of phosphoproteins in the sample could easily exceed 1,000. To evaluate the possibility of identifying all these phosphoproteins, trypsin digested proteins were immobilized metal affinity chromatography (IMAC) and analysed by nanoflow HPLC/ESI MS. They could make 216 assignments as a total, of which 60 were singly phosphorylated, 145 were double phosphorylated, and 11 were triply phosphorylated. In a more recent global analysis of protein phosphorylation in yeast which did not involve phosphoproteome, (Ptacek *et al.*, 2005) approximately 4,200 phosphorylation events affecting 1,325 proteins were identified from the 87 yeast protein kinase assays. Each kinase recognized between 1 and 256 substrates with an average of 47 substrates per kinase. Furthermore, an integration of the phosphorylation results with former protein–protein interaction and transcription factor binding data revealed novel regulatory modules. The phosphorylation results of this group have been assembled into a first-generation phosphorylation map for yeast. In a phosphoproteome map constructed for rat mesangial cells by employing ProQ staining protocol (Jiang *et al.*, 2005), 37 protein spots (representing 28 unique proteins) of the 157 identified protein spots (23.5%) in the silver stained 2-DE map were found to be phosphorylated. In a study with rice, about 900 protein spots were detected on 2-D gel by silver staining and only 44 of these were phosphoproteins as defined by their being detected on X-ray film. This corresponded to 4.9% of the total number of proteins (Khan *et al.*, 2005). In

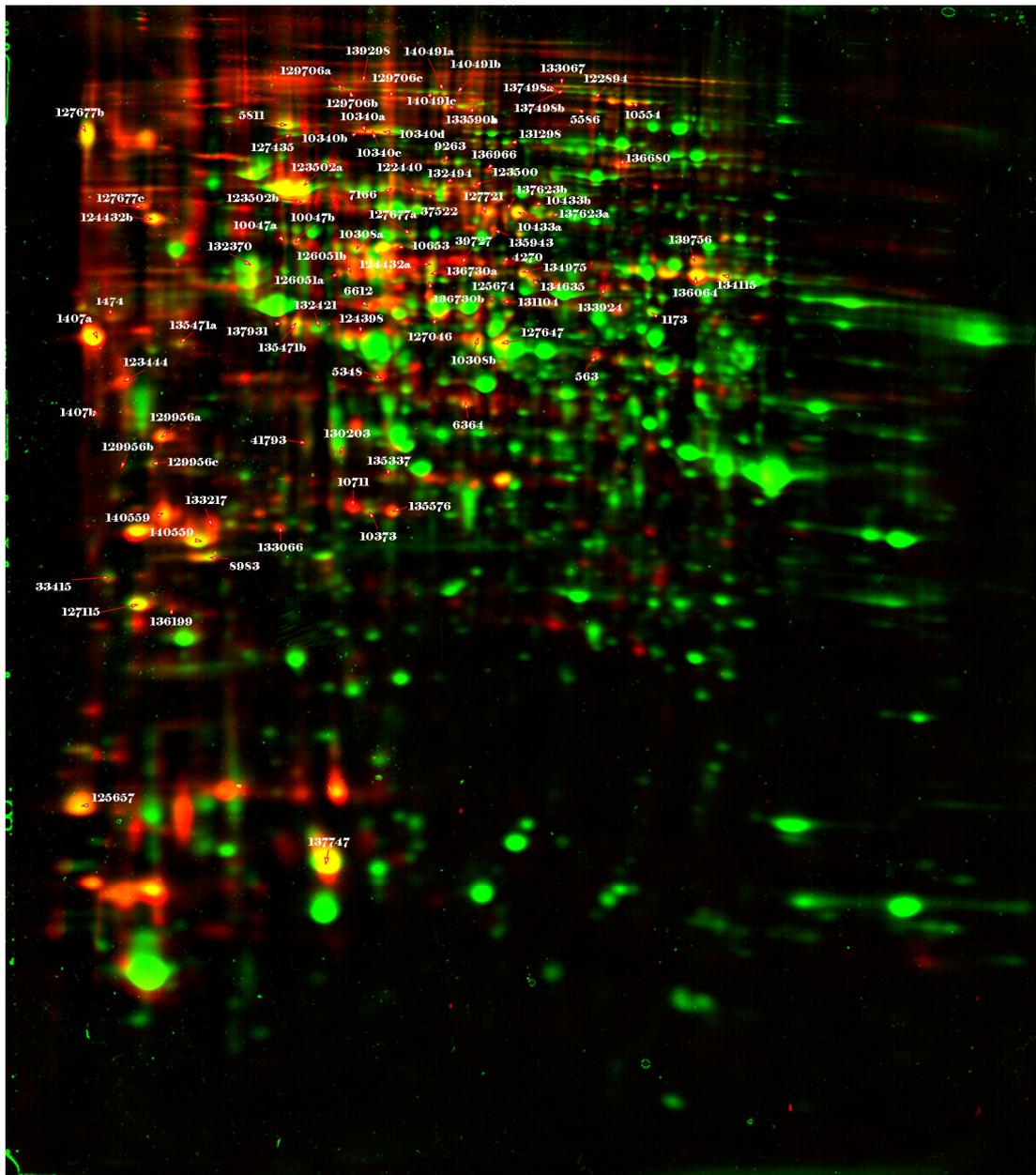
phosphoproteomic analysis of rat liver, Moser and White (2006) made phosphoproteomic analysis of rat liver by high capacity IMAC and LC MS/MS, resulting in the identification of 299 phosphopeptides from over 200 proteins in rat liver. In total, they were able to identify 339 phosphorylation sites and unambiguously assign 90% of these phosphorylation events to a single residue.



**Figure 10.** ProQ stained 2-DE of *P. chrysosporium* soluble proteins.



**Figure 11.** Spot boundary detection of phosphoproteins by the use of Delta2D software.



**Figure 12.** The image of the ProQ stained (Red spots) and Coomassie stained (Green) spots on the same gel as created by the Delta2D software. The spot identifications were shown as protein accession numbers.

In the present study, 40% of the distinct protein spots were found to be phosphorylated in *P. chrysosporium*. As an eukaryotic microorganism, the ratio of phosphoproteins seems to be higher than expected in view of the recent related reports based on higher eukaryotes. Nevertheless, it is to be noted that our results are not comparable to previously published ones, since our master gel was stained with Coomassie blue in order to make protein spots

amenable to further MS analysis. In most other studies including that of Jiang *et al.* (2005) who also employed ProQ staining for phosphoproteins, the master gel was stained with silver. As it is well known, Coomassie staining is about 1/50<sup>th</sup> less sensitive than silver staining and hence we might have missed several unphosphorylated spots in our 2-D master gels. Before comparing our results to those from other organisms, we have to repeat the same experiments by staining master gel with silver, instead of Coomassie. Indeed, *P. chrysosporium* phosphoproteome work is yet unfinished in our laboratory and our future studies including determination of phosphorylation sites and stress-induced phosphoproteins will contribute significantly to deeper insights into cellular processes and cell regulation in this organism.

**Table 6.** List of phosphorylated spots on the ProQ stained gel

<b>Protein ID</b>	<b>Putative Function</b>
563	Pyruvate dehydrogenase E1 component alpha subunit
1173	Oxidoreductase, N-terminal
1407a	RNA-binding region RNP-1 (RNA recognition motif)
1407b	RNA-binding region RNP-1 (RNA recognition motif)
1474	Proteasome component region PCI
4270	Metalloenzyme, phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent
5348	Hypothetical protein
5586	Heat shock protein Hsp70,
5811	Required for invasion and pseudohyphae formation in response to nitrogen starvation
6364	6-phosphogluconolactonase
6612	Rab GDI protein, probable secretory pathway GDP dissociation inhibitor 1
7166	Heat shock protein Hsp70
8983	Adhesion regulating molecule
9263	Spermine synthase, saccharopine dehydrogenase
10047a	Tubulin/FtsZ, C-terminal
10047b	Tubulin/FtsZ, C-terminal
10308a	S-adenosyl-L-homocysteine hydrolase
10308b	S-adenosyl-L-homocysteine hydrolase
10340a	Heat shock protein Hsp72
10340b	Heat shock protein Hsp72
10340c	Heat shock protein Hsp72
10340d	Heat shock protein Hsp72
10373	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1, WD-40 repeat

**Table 6.** Continued

10433a	Transketolase
10433b	Transketolase
10554	ATP/GTP-binding site motif A (P-loop), ABC transporter, HPr serine phosphorylation site
10653	Nucleosome assembly protein (NAP)
10711	Helix-turn-helix, Fis-type
33415	Nascent polypeptide-associated
37522	Glycoside hydrolase, family 21
39727	Elongation factor 1, gamma chain
41793	Uricase
122440	Heat shock protein Hsp70
122894	ABC transporter
123444	H <sup>+</sup> -transporting two-sector ATPase, alpha/beta subunit, central region
123500	H <sup>+</sup> -transporting two-sector ATPase, alpha/beta subunit, central region
123502a	Peptidase, eukaryotic cysteine peptidase active site
123502b	Peptidase, eukaryotic cysteine peptidase active site
124398	Catalase
124432a	ABC transporter
124432b	ABC transporter
125657	Putative peroxiredoxin (Thioredoxin reductase) (Allergen Mal f 3)
125674	Tubulin/FtsZ, C-terminal
126051a	UTP--glucose-1-phosphate uridylyltransferase
126051b	UTP--glucose-1-phosphate uridylyltransferase
127046	Helicase, C-terminal
127115	SAM (and some other nucleotide) binding motif, O-methyltransferase, family 3
127435	AAA ATPase
127647	Ribosomal protein P2
127677a	Argininosuccinate synthase
127677b	Argininosuccinate synthase
127677c	Argininosuccinate synthase
127721	G-protein beta WD-40 repeat, corin like proteinon
129706a	UBA/THIF-type NAD/FAD binding fold, Ubiquitin-activating enzyme
129706b	UBA/THIF-type NAD/FAD binding fold, Ubiquitin-activating enzyme
129706c	UBA/THIF-type NAD/FAD binding fold, Ubiquitin-activating enzyme
129956a	Peptidase aspartic, active site, Peptidase A1, pepsin
129956b	Peptidase aspartic, active site, Peptidase A1, pepsin
129956c	Peptidase aspartic, active site, Peptidase A1, pepsin
130203	Ribosomal protein 60S
131104	Peptidase M20
131298	Actin-binding, actinin-type, Calcium-binding EF-hand
132370	S-adenosylmethionine synthetase
132421	Helicase, C-terminal
132494	Peptidase M49
133066	NADH dehydrogenase (ubiquinone), 24 kDa subunit

**Table 6.** Continued

133067	Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2
133217	Isopentenyl-diphosphate delta-isomerase
133590a	Alanyl-tRNA synthetase
133590b	Alanyl-tRNA synthetase
133924	Aldehyde dehydrogenase
134115	UTP--glucose-1-phosphate uridylyltransferase,
134635	G-protein beta WD-40 repeat,
134975	Cell division/GTP binding protein
135337	Ribosomal protein S2, eukaryotic and archaeal form
135471a	Cell division/GTP binding protein
135471b	Cell division/GTP binding protein
135576	Bacterial/Archaeal inorganic pyrophosphatase
135943	Peptidase M17, cytosol aminopeptidase
136064	Protein of unknown function DUF619
136199	HAD-superfamily hydrolase, subfamily IB (PSPase-like), Haloacid dehalogenase-like hydrolase
136680	Glucosamine-fructose-6-phosphate aminotransferase, Sugar isomerase (SIS), Bacterial extracellular solute-binding protein, family 3,
136730a	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit
136730b	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit
136966	Peptidase M1, membrane alanine aminopeptidase, Peptidase M, neutral zinc metallopeptidases, zinc-binding site
137498a	HMG-CoA lyase-like, Carbamoyl-phosphate synthase L chain, ATP-binding, Biotin-binding site
137498b	HMG-CoA lyase-like, Carbamoyl-phosphate synthase L chain, ATP-binding, Biotin-binding site
137623a	Phosphoglucomutase/phosphomannomutase C terminal
137623b	Phosphoglucomutase/phosphomannomutase C terminal
137747	Heat shock protein Hsp20
137931	Glycine hydroxymethyltransferase
139298	Actin/actin-like
139756	Peptidase, eukaryotic cysteine peptidase active site
140491b	Animal haem peroxidase
140491c	Animal haem peroxidase
140559	EB1 protein

## CHAPTER 4

### CONCLUSION

- In this study, a 2-D reference map was constructed for the vegetative cells of the white-rot fungus *P. chrysosporium* grown under normal physiological conditions. From this map, 720 spots in duplicates (a total of 1440 spots) were processed and analyzed for their PMF. Data analyses by bioinformatic tools identified 517 polypeptides which represented 314 distinct ORFs.
- 118 out of 314 ORFs yielded multiple spots. Of these, 38 were found to differ in their charge, 10 differed in their apparent molecular weight and 70 were found to differ both in their charge and their apparent molecular weight. These proteins should have been formed as a result of post-translational modification events, such as enzymatical cleavage or modification, though mistranslation and artificial chemical modification are not ruled out.
- Relative abundancies of the polypeptides on the map revealed that the most abundant ones are housekeeping proteins.
- Relative distribution of the 517 identified proteins to functional categories showed that "Metabolism" was the main functional category in exponentially-growing *P. chrysosporium* cells with a share of 50.6% among identified proteins.
- More than half of the identified proteins were found to perform the functions of the following five classes: "Posttranslational modifications, protein turnover, chaperones", "Energy production and conversion", "Carbohydrate transport and metabolism", "Amino acid transport and

metabolism”, and “Translation, ribosomal structure and biogenesis”. Of these, “Posttranslational modifications, protein turnover, chaperones” was represented with the highest number (104) of the polypeptides.

- Out of the 314 gene products, 147 were predicted to be located in cytoplasm, 17 proteins were found to be located on cytoskeleton, 15 were found to be in either nucleus or cytoplasm, 1 was found to be in endoplasmic reticulum, 11 were found to be extracellular, 80 were found to be mitochondrial, 35 were found to be nuclear and 8 were found to be located in peroxisome. In the present reference map, 29 proteins were predicted to have a signal peptide sequence.
- 380 out of 910 distinct protein spots (40%) were phosphorylated in exponentially growing cells of *P. chrysosporium*. 96 of these spots were matched to the identified proteins on the reference map.
- As the first presentation of a proteome map of *P. chrysosporium*. with 517 identified gene products, the present study established a solid proteomic framework for further analysis of differential gene expression in *P. chrysosporium*.

## REFERENCES

- Ahn. N.G. and Resing. K.A. (2001).** Toward the phosphoproteome. *Nat Biotechnol.* 19(4):317-8.
- Andersen, J. S. and Mann, M. (2000).** Functional genomics by mass spectrometry. *FEBS Lett.* 480: 25-31.
- Anderson. N.L. and Anderson N.G. (2002).** The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 1(11):845-67.
- Anglade. P., Demey. E., Labas. V., Le Caer. J.P. and Chich. J.F. (2000).** Towards a proteomic map of *Lactococcus lactis* NCDO 763. *Electrophoresis.* 2000 Jul;21:2546-9.
- Apweiler. R., Hermjakob. H. and Sharon. N., (1999).** On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta.* 6;1473:4-8.
- Arena. S., D'Ambrosio. C., Renzone. G., Rullo. R., Ledda. L., Vitale. F., Maglione. G., Varcamonti. M., Ferrara. L. and Scaloni. A. (2006).** A study of *Streptococcus thermophilus* proteome by integrated analytical procedures and differential expression investigations. *Proteomics.* 6(1):181-92.
- Bebien, M., Chauvin, J-P., Adriano, J-M., Grosse, S. and Vermeoglio, A. (2001).** Effect of selenite on growth and protein synthesis in the phototrophic bacterium *Rhodobacter sphaeroides*. *Appl. Environ. Microb.* 67: 4440-4447.

**Bendt, A.K., Burkovski, A., Schaffer, S., Bott, M., Farwick, M., and Hermann, T. (2003)** Towards a phosphoproteome map of *Corynebacterium glutamicum*. *Proteomics*.3:1637-46.

**Berkelman, T. and Stensted, T. (1998).** 2-D Electrophoresis; using immobilized pH gradient, Principles and Methods. 80-6429-60 Amersham Bioscience. pp. 27-76.

**Bernhardt, J., Volker, U., Volker, A., Antelmann, H., Schmih, R., Marc, H. and Hecker, M. (1997).** Specific and general stress proteins in *Bacillus subtilis*-a two dimensional protein electrophoresis study. *Microbiology*. 143: 999-1017.

**Bjellqvist, B., Ek, K., Righetti, P. G., Giannazza, E., Görg, A., Westermeier, R. and Postel, W. (1982).** Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Bioph. Meth.* 6: 317-339.

**Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S. and Hochstrasser, D. (1993).** The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*. 14:1023-31.

**Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997).** The complete genome sequence of *Escherichia coli* K-12. *Science*. 5;277:1453-74.

**Blom, N., Kreegipuu, A., and Brunak, S. (1998).** PhosphoBase: a database of phosphorylation sites. *Nucleic Acids Res.*: 1;26:382-6.

**Blomberg, A. (1995).** Global changes in protein synthesis during adaptation of the yeast *Saccharomyces cerevisiae* to 0.7 M NaCl. *J. Bacteriol.* 177: 3563-3572.

**Blomberg, A. (1997).** Osmoresponsive proteins and functional assessment strategies in *Saccharomyces cerevisiae*. *Electrophoresis*. 18: 1429-1440.

**Blomberg, A. (2000).** Metabolic surprises in *Saccharomyces cerevisiae* during adaptation to saline conditions: questions, some answers and a model. *FEMS Microbiol. Lett.* 182: 1-8.

**Blum, H., Beier, H. and Gross, H. J. (1987).** Improved silver staining of plant proteins, RNA, and DNA in polyacrylamide gels. *Electrophoresis*. 8: 93-99.

**Boucherie. H., Sagliocco. F., Joubert. R., Maillet. I., Labarre. J. and Perrot. M. (1996).** Two-dimensional gel protein database of *Saccharomyces cerevisiae* *Electrophoresis*. 17: 1683-1699.

**Bradford, M. M. (1976).** A rapid and sensitive method for the quantitation of micrograms quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-244.

**Bumpus, J. A., Tien M., Wright D. and Aust S. D. (1985).** Oxidation of persistent environmental pollutants by a white rod fungus. *Science*. 228: 1434-1436.

**Burdsall. H. H., Jr, and Esllyn. W. (1974).** A new *Phanerochaete* with a *Chrysosporium* imperfect state. *Mycotaxon* 1:124

**Celis. J.E., Kruhoffer. M., Gromova. I., Frederiksen. C., Ostergaard. M., Thykjaer. T., Gromov. P., Yu. J., Palsdottir. H., Magnusson. N. and Orntoft. T.F. (2000).** Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* 25;480:2-16.

**Cordwell. S.J., Larsen. M.R., Cole. R.T. and Walsh. B.J. (2002).** Comparative proteomics of *Staphylococcus aureus* and the response of methicillin-resistant and methicillin-sensitive strains to Triton X-100. *Microbiology*. 148:2765-81.

**Cordwell. S.J., Nouwens. A.S., and Walsh. B.J. (2001).** Comparative proteomics of bacterial pathogens. *Proteomics*. 1(4):461-72.

**Damerval. C., de Vienne. D., Zivy. M. and Thiellement. H. (1986)** Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis* 7: 52 - 54

**de Magalhaes. J.P. and Toussaint. O., (2004).** GenAge: a genomic and proteomic network map of human ageing. *FEBS Lett.* 30;571:243-7.

**Donnes. P. and Hoglund. A. (2004).** Predicting protein subcellular localization: past, present, and future. *Genomics Proteomics Bioinformatics*. 2(4):209-15.

**Dowsey. A.W., Dunn. M.J. and Yang G.Z. (2003).** The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics*. 3:1567-96.

**Eichler. J. and Adams. M.W. (2005).** Posttranslational protein modification in Archaea. *Microbiol Mol Biol Rev.* 69:393-425.

**Eisenberg, D., Marcotte, E. M., Xenarios, I. and Yetas, T. O. (2000).** Proteins function in the post-genomic era. *Nature*. 405: 823-826.

**Fey. S.J. and Larsen. P.M. (2001)** 2D or not 2D. Two-dimensional gel electrophoresis. *Curr Opin Chem Biol.* 2001 5:26-33.

**Ficarro. S.B., McClelland. M.L., Stukenberg. P.T., Burke. D.J., Ross. M.M., Shabanowitz. J., Hunt. D.F. and White. F.M. (2002)** Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol.* 20:301-5.

**Finehout. E.J., Franck. Z. and Lee. K.H. (2004).** Towards two-dimensional electrophoresis mapping of the cerebrospinal fluid proteome from a single individual. *Electrophoresis*. 25:2564-75.

**Futcher, B., Latter, G. I., Monardo, P., McLaughlin C. S. and Garrels, J. I. (1999).** A sampling of the yeast proteome. *Mol. Cell. Biol.* 19: 7357-7368.

**Gade, D., Theiss, D., Lange, D., Mirgorodskaya, E., Lombardot, T., Glöckner, F.O., Kube, M., Reinhardt., R., Amann, R., Lehrach, H., Rabus, R. and Gobom, G. (2005).** Towards the proteome of the marine bacterium *Rhodospirillum rubrum*: Mapping the soluble proteins. *Proteomics*, 5, 3654–3671

**Gauss, C., Kalkum, M., Löwe, M., Lehrach, H. and Klose, J. (1999).** Analysis of the mouse proteome. (I) Brain proteins: separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation. *Electrophoresis*. 20: 575-600.

**Gelhaus. C., Fritsch. J., Krause. E. and Leippe. M. (2005).** Fractionation and identification of proteins by 2-DE and MS: towards a proteomic analysis of *Plasmodium falciparum*. *Proteomics*. 5:4213-22.

**Giorgianni. F., Desiderio. D.M. and Beranova-Giorgianni. S. (2003).** Proteome analysis using isoelectric focusing in immobilized pH gradient gels followed by mass spectrometry. *Electrophoresis*. 24:253-9.

**Goffeau. A., Barrell. B.G., Bussey. H., Davis. R.W., Dujon. B., Feldmann. H., Galibert. F., Hoheisel. J.D., Jacq. C., Johnston. M., Louis. E.J., Mewes. H.W., Murakami. Y., Philippsen. P., Tettelin. H. and Oliver. S.G. (1996).** Life with 6000 genes. *Science*. 274:546, 563-7.

**Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. and Weiss, W. (2000).** The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*. 21: 1037-1053.

**Gorg. A., Postel. W., Gunther. S., Weser. J., Strahler. J.R., Hanash. S.M., Somerlot. L. and Kuick. R. (1988).** Approach to stationary two-dimensional pattern: influence of focusing time and immobiline/carrier ampholytes concentrations. *Electrophoresis*. 9:37-46.

**Graves, P. R. and Haystead, T. A. (2002).** Molecular biologist's guide to proteomics. *Microbiol. Mol. Biol. Rev.* 66: 39-63.

**Grinyer. J., McKay. M., Nevalainen H. and Herbert. B.R. (2004).** Fungal proteomics: initial mapping of biological control strain *Trichoderma harzianum*. *Curr Genet*. 45:163-9.

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

**Gygi. S.P., Rist. B. and Aebersold. R. (2000).** Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol.* 11:396-401.

**Hansmeier. N., Chao. T.C., Puhler. A., Tauch. A., Kalinowski. J. (2006).** The cytosolic, cell surface and extracellular proteomes of the biotechnologically important soil bacterium *Corynebacterium efficiens* YS-314 in comparison to those of *Corynebacterium glutamicum* ATCC 13032. *Proteomics*. 6:233-50.

**Hanson. B.J., Schulenberg. B., Patton. W.F., Capaldi. R.A. (2001).** A novel subfractionation approach for mitochondrial proteins: a three-dimensional mitochondrial proteome map. *Electrophoresis*.22:950-9.

**He. F. (2005).** Human Liver Proteome: Project Plan, Progress, and Perspectives *Molecular & Cellular Proteomics*. 4:1841-1848.

**Heim. S., Ferrer. M., Heuer. H., Regenhardt. D., Nimtz. M. and Timmis K.N. (2003).** Proteome reference map of *Pseudomonas putida* strain KT2440 for genome expression profiling: distinct responses of KT2440 and *Pseudomonas aeruginosa* strain PAO1 to iron deprivation and a new form of superoxide dismutase. *Environ Microbiol*. 5:1257-69.

**Hernandez. R., Nombela. C., Diez-Orejas. R. and Gil. C. (2004).** Two-dimensional reference map of *Candida albicans* hyphal forms. *Proteomics*. 4:374-82.

**Hille. J.M., Freed. A.L. and Watzig. H. (2001).** Possibilities to improve automation, speed and precision of proteome analysis: a comparison of two-dimensional electrophoresis and alternatives. *Electrophoresis*. 22:4035-52.

**Hochstrasser. D. F., Harrington. M. G., Hochstrasser. A. C., Miller. M. J. and Merril. C. R. (1988).** Methods for increasing the resolution of two-dimensional protein electrophoresis. *Anal. Biochem*. 173: 424-435.

**Hoffmann, T., Schutz, A., Brosius, M., Volker, A., Volker, U. and Bremer, E. (2002).** High-salinity-induced iron limitation in *Bacillus subtilis*. *J. Bacteriol*. 184: 718-727.

**Hoving, S., Voshol, H., and van Oostrum, J. (2000).** Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. *Electrophoresis*. 21: 2612-2621.

**[http://botit.botany.wisc.edu/toms\\_fungi/may97.html](http://botit.botany.wisc.edu/toms_fungi/may97.html)** last accessed on 18.01.2006

**<http://genome.jgi-psf.org/Phchr1/Phchr1.home.html>** last accessed on 18.01.2006

<http://wolfsort.seq.cbrc.jp> date: 18.01.2006 last accessed on 18.01.2006

<http://www.cbs.dtu.dk/services/SignalP> last accessed on 18.01.2006

<http://www.expasy.org/ch2d/publi/ecoli.html> last accessed on 18.01.2006

[http://www.expasy.org/tools/pi\\_tool.html](http://www.expasy.org/tools/pi_tool.html) last accessed on 18.01.2006

<http://www.hupo.org> last accessed on 18.01.2006

<http://www.ibgc.u-bordeaux2.fr/YPM/> last accessed on 18.01.2006

<http://www.bioinformatics.med.umich.edu/hupo/ppp> last accessed on 18.01.2006

<http://www.HUPO.org> last accessed on 18.01.2006

**Jensen, O.N. (2000).** Modification-specific proteomics: systematic strategies for analysing post-translationally modified proteins. *Proteomics: A Trends Guide*. 36-42

**Jiang. X.S., Tang. L.Y., Cao. X.J., Zhou. H., Xia. Q.C., Wu. J.R. and Zeng. R., (2005).** Two-dimensional gel electrophoresis maps of the proteome and phosphoproteome of primitively cultured rat mesangial cells. *Electrophoresis*, 26:4540-62.

**Kalume. D.E., Molina. H. and Pandey. A. (2003).** Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol*. 7:64-9.

**Kanamoto T, Hellman U, Heldin CH, Souchelnytskyi S. (2002).** Functional proteomics of transforming growth factor-beta1-stimulated Mv1Lu epithelial cells: Rad51 as a target of TGFbeta1-dependent regulation of DNA repair. *EMBO J*. 1;21:1219-30.

**Khan. M., Takasaki. H. and Komatsu S. (2005).** Comprehensive phosphoproteome analysis in rice and identification of phosphoproteins responsive to different hormones/stresses. *J Proteome Res.* 4:1592-9.

**Kirkpatrick, C., Maurer, L. M., Oyelakin, N. E., Yoncheva, Y. N., Maurer, R. and Slonczewski, J. L. (2001).** Acetate and formate stress: opposite responses in the proteome of *Escherichia coli*. *J. Bacteriol.* 183: 6466-6477.

**Klose, J. (1975).** Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutation in mammals. *Hum. Genet.* 26: 231-243.

**Komatsu. S. (2005).** Rice proteome database: a step toward functional analysis of the rice genome. *Plant Mol Biol.* 59:179-90.

**Laemmli, U. K. (1970).** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227: 680-685.

**Lebeau. I., Lammertyn. E., De Buck. E., Maes. L., Geukens. N., Van Mellaert. L., Arckens. L., Anne. J. and Clerens. S. (2005).** First proteomic analysis of *Legionella pneumophila* based on its developing genome sequence. *Res Microbiol.* 156:119-29.

**Leitsch. D., Radauer. C., Paschinger. K., Wilson. I.B., Breiteneder. H., Scheiner. O. and Duchene. M. (2005).** *Entamoeba histolytica*: analysis of the trophozoite proteome by two-dimensional polyacrylamide gel electrophoresis. *Exp Parasitol.* 110:191-5.

**Lewis, T. S., Hunt, J. B., Aveline, L. D., Jonscher, K. R., Louie, D. F., Yeh, J. M., Nahreini, T. S., Resing, K. A. and Ahn, N. G. (2000).** Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol. Cell.* 6:1343-1354.

**Lim. D., Hains. P., Walsh. B., Bergquist. P. and Nevalainen. H. (2001).** Proteins associated with the cell envelope of *Trichoderma reesei*: a proteomic approach. *Proteomics*.1:899-909.

**Martinez. D., Larrondo. L.F., Putnam. N., Gelpke. M.D., Huang. K., Chapman. J., Helfenbein. K.G., Ramaiya. P., Detter. J.C., Larimer. F., Coutinho. P.M., Henrissat. B., Berka. R., Cullen. D. and Rokhsar. D. (2004).** Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol.* 22:695-700.

**Mechin. V., Balliau. T., Chateau-Joubert. S., Davanture. M., Langella. O., Negroni. L., Prioul. J.L., Thevenot. C., Zivy. M. and Damerval. C. (2004).** A two-dimensional proteome map of maize endosperm. *Phytochemistry.* 65:1609-18.

**Moser. K., and White FM. (2006).** Phosphoproteomic Analysis of Rat Liver by High Capacity IMAC and LC-MS/MS. *J Proteome Res.* 5:98-104.

**Mumby. M. and Brekken. D. (2005).** Phosphoproteomics: new insights into cellular signaling. *Genome Biol.* 6:230.

**Neuhof, V.N., Arold, N., Taube, D. & Erhardt, W. ( 1988).** Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9, 255– 262.

**Nielsen. H., Engelbrecht. J., Brunak. S. and von Heijne. G. (1997).** Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10:1-6.

**Noel-Georis. I., Vallaey. T., Chauvaux. R., Monchy. S., Falmagne. P., Mergeay. M. and Wattiez. R. (2004).** Global analysis of the *Ralstonia metallidurans* proteome: prelude for the large-scale study of heavy metal response. *Proteomics.* 4:151-79.

**Noir. S., Brautigam. A., Colby. T., Schmidt. J. and Panstruga. R. (2005).** A reference map of the *Arabidopsis thaliana* mature pollen proteome. *Biochem Biophys Res Commun.* 2;337:1257-66.

**Nouwens. A.S., Cordwell. S.J., Larsen. M.R., Molloy. M.P., Gillings. M., Willcox. M.D. and Walsh. B.J. (2000).** Complementing genomics with proteomics: the membrane subproteome of *Pseudomonas aeruginosa* PAO1. *Electrophoresis.* 21:3797-809.

**Nouwens. A.S., Willcox. M.D., Walsh. B.J. and Cordwell. S.J. (2002).** Proteomic comparison of membrane and extracellular proteins from invasive (PAO1) and cytotoxic (6206) strains of *Pseudomonas aeruginosa*. *Proteomics.* 2:1325-46.

**O'Farrell, P. H. (1975).** High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250: 4007-4021.

**Oh. J., Pyo. J.H., Jo. E.H., Hwang. S.I., Kang. S.C., Jung. J.H., Park. E.K., Kim. S.Y., Choi. J.Y. and Lim. J. (2004).** Establishment of a near-standard two-dimensional human urine proteomic map. *Proteomics.* 4:3485-97.

**Özcan, S. (2003).** Proteome analysis of Pb (II), Ni (II) and Cr (III) response in *Phanerochaete chrysosporium*. PhD thesis, METU.

**Parodi-Talice. A., Duran. R., Arrambide. N., Prieto. V., Pineyro. M.D., Pritsch. O., Cayota. A., Cervenansky. C. and Robello. C. (2004).** Proteome analysis of the causative agent of Chagas disease: *Trypanosoma cruzi*. *Int J Parasitol.* 34:881-6.

**Pasquali. C., Frutiger. S., Wilkins. M.R., Hughes. G.J., Appel. R.D., Bairoch. A., Schaller. D., Sanchez. J.C. and Hochstrasser. D.F. (1996).** Two-dimensional gel electrophoresis of *Escherichia coli* homogenates: the *Escherichia coli* SWISS-2DPAGE database. *Electrophoresis.* 17:547-55.

**Pennigton, S. R., Wilkins, M. R., Hochstrasser, D. E. and Dunn, M. J. (1997).** Proteome analysis: from protein characterization to biological function. *Trends. Cell. Biol.* 7: 168-173.

**Perrot. M., Sagliocco. F., Mini. T., Monribot. C., Schneider. U., Shevchenko. A., Mann. M., Jenö. P. and Boucherie. H. (1999).** Two-dimensional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis.* 20:2280-98.

**Predic. J., Soskic. V., Bradley. D. and Godovac-Zimmermann. J. (2002).** Monitoring of gene expression by functional proteomics: response of human lung fibroblast cells to stimulation by endothelin-1. *Biochemistry.* 41:1070-8.

**Prouty, A. L. (1990).** Bench-scale development and evaluation of a fungal bioreactor for color removal from bleach effluence. *Appl. Microbiol. Biotechnol.* 32: 490-493.

**Ptacek. J., Devgan. G., Michaud. G., Zhu. H., Zhu. X., Fasolo. J., Guo. H., Jona. G., Breitkreutz. A., Sopko. R., McCartney. R.R., Schmidt. M.C., Rachidi. N., Lee. S.J., Mah. A.S., Meng L., Stark. M.J., Stern. D.F., De Virgilio. C., Tyers. M., Andrews. B., Gerstein. M., Schweitzer. B., Predki. P.F., and Snyder. M. (2005).** Global analysis of protein phosphorylation in yeast. *Nature.* 1;438:679-84.

**Ramagli, L. S. and Rodriguez, L. V. (1985).** Quantification of microgram amount of protein in 2-dimensional PAGE sample buffer. *Electrophoresis.* 6: 559-563.

**Regula. J.T., Ueberle. B., Boguth. G., Gorg. A., Scholzer. M., Herrmann. R. and Frank. R. (2000).** Towards a two-dimensional proteome map of *Mycoplasma pneumoniae*. *Electrophoresis.* 21:3765-80.

**Reinders. J. and Sickmann. A. (2005).** State-of-the-art in phosphoproteomics. *Proteomics.* 5:4052-61.

**Rosen, R., Buttner, K., Becher, D., Nakahigashi, K., Yura, T., Hecker, M. and Ron, E. Z. (2002).** Heat shock proteome of *Agrobacterium tumefaciens*. *J. Bacteriol.* 184: 1772-1778.

**Rosen, R., Sacher, A., Shechter, N., Becher, D., Büttner, K., Biran, D., Hecker M., and Ron, E.Z, (2004)** Two-dimensional reference map of *Agrobacterium tumefaciens* proteins. *Proteomics*, 4: 1061–1073.

**Sarnighausen. E., Wurtz. V., Heintz. D., Van Dorselaer. A. and Reski. R. (2004).** Mapping of the *Physcomitrella patens* proteome. *Phytochemistry*. 65:1589-607.

**Steiner, P. and Sauer, U. (2001).** Proteins induced during adaptation of *Acetobacter acetii* to high acetate concentrations. *Appl. Environ. Microbiol.* 67: 5474-5481.

**Sun. N., Jang. J., Lee. S., Kim. S., Lee. S., Hoe. K.L., Chung. K.S., Kim. D.U., Yoo. H.S., Won. M. and Song. K.B. (2005).** The first two-dimensional reference map of the fission yeast, *Schizosaccharomyces pombe* proteins. *Proteomics*. 5:1574-9.

**Teixeira-Gomes, A. P., Cloeckert, A. and Zygmunt, M. S. (2000).** Characterization of heat, oxidative and acid stress responses in *Brucella mellitensis*. *Infect. Immun.* 68: 2954-2961.

**Tohomas, J. D. and VanBogelen, R. A. (2000).** Experimentalism in proteomics. *Proteomics: Trends Guide.* 7-11.

**Tonella. L., Walsh. B.J., Sanchez. J.C., Ou. K., Wilkins. M.R., Tyler. M., Frutiger. S., Gooley. A.A., Pescaru. I., Appel. R.D., Yan. J.X., Bairoch. A., Hoogland. C., Morch. F.S., Hughes. G.J., Williams. K.L. and Hochstrasser. D.F. (1998).** '98 *Escherichia coli* SWISS-2DPAGE database update. *Electrophoresis*. 19:1960-71.

**Vido, K., Spector, D., Lagniel, G., Lopez, S., Tolodano, M. B. and Labarre, J. (2001).** A proteome analysis of the cadmium response in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 276(11): 8469-8474.

**Voigt, B., Schweder, T., Sibbald, M.J., Albrecht, D., Ehrenreich, A., Bernhardt, J., Feesche, J., Maurer, K.H., Gottschalk, G., van Dijk, J.M., and Hecker, M. (2006).** The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions. *Proteomics*. 6:268-81.

**Wang, J., Ying, T., Wang, H., Shi, Z., Li, M., He, K., Feng, E., Wang, J., Yuan, J., Li, T., Wei, K., Su, G., Zhu, H., Zhang, X., Huang, P. and Huang, L. (2005).** 2-D reference map of *Bacillus anthracis* vaccine strain A16R proteins. *Proteomics*. 5:4488-95.

**Wang, L., Zhu, Y.F., Guo, X.J., Huo, R., Ma, X., Lin, M., Zhou, Z.M. and Sha, J.H. (2005).** A two-dimensional electrophoresis reference map of human ovary. *J Mol Med.* 83:812-821.

**Washburn, M. P. and Yates, J. R. (2000).** Analysis of the microbial proteome. *Curr. Opin. Microbiol.* 3:292-297.

**Watson, S.P., Bahou, W.F., Fitzgerald, D., Ouwehand, W., Rao, A.K. and Leavitt, A.D. (2005).** Mapping the platelet proteome: a report of the ISTH Platelet Physiology Subcommittee. *J Thromb Haemost.* 3:2098-101.

**Westermeier, R. and Naven, T. (2002).** Proteomics in practice: a laboratory manual of proteome analysis. 1<sup>st</sup> ed. Wiley-VCH Verlag GmbH. pp. 1-150.

**Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, X., Gooley, A. A., Hughes, G., Humphrey-Smith, I., Williams, K. L. and Hochstrasser, D. E. (1996).** From proteins to proteomes: Large scale protein identification by two dimensional electrophoresis and amino acid analysis. *Bio. Technol.* 14: 61-65.

**Wilkins, M. R., Sanchez, J. C., Williams, K. L. and Hochstrasser, D. F. (1996).** Current challenges and future applications for protein maps and pos-translational vector maps in proteome projects. *Electrophoresis*. 17: 830-838.

**Wymelenberg. A.V., Sabat. G., Martinez. D., Rajangam. A.S., Teeri. T.T., Gaskell. J., Kersten. P.J. and Cullen D. (2005).** The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. *J Biotechnol*. 21;118:17-34.

**Yan. J.X., Tonella. L., Sanchez. J.C., Wilkins. M.R., Packer. N.H., Gooley. A.A., Hochstrasser. D.F. and Williams. K.L. (1997).** The *Dictyostelium discoideum* proteome--the SWISS-2DPAGE database of the multicellular aggregate (slug). *Electrophoresis*. 18:491-7.

**Yetiş, Ü., Dölek, A., Dilek, F. B. and Özcengiz, G. (2000).** The removal of Pb (II) by *Phanerochaete chrysosporium*. *Wat. Res*. 34: 4090-4100.

**Yetiş, Ü., Özcengiz, G., Dilek, F. B., Ergen, N. Erbay, A. and Dölek, A. (1998).** Heavy metal biosorption by *Polyporous versicolor* and *Phanerorochaete chrysosporium*. IAWQ 19<sup>th</sup> Biennial Conference, Water Quality Internanatinal 98, June 21-26, Vancouver, pp. 324-331.

**Yetiş, Ü., Özcengiz, G., Dilek, F. B., Ergen, N. Erbay, A. and Dölek, A. (1998).** Heavy metal biosorption by white-rot fungi. *Water. Sci. Technol*. 38: 323-330.

**Yin. Z., Stead. D., Selway. L., Walker. J., Riba-Garcia. I., McLnerney. T., Gaskell. S., Oliver. S.G., Cash. P. and Brown. A.J. (2004).** Proteomic response to amino acid starvation in *Candida albicans* and *Saccharomyces cerevisiae*. *Proteomics*. (8):2425-36.

**Ying. T., Wang. H., Li. M., Wang. J., Wang. J., Shi. Z., Feng. E., Liu. X., Su. G., Wei. K., Zhang. X., Huang. P. and Huang. L (2005).** Immunoproteomics of outer membrane proteins and extracellular proteins of *Shigella flexneri* 2a 2457T. *Proteomics*. 5:4777-93.

**Zhan. X. and Desiderio. D.M. (2003).** A reference map of a human pituitary adenoma proteome. *Proteomics*. 3:699-713.

## APPENDIX A

### Colloidal Coomassie Blue (CCB) Staining and Stock Solutions

#### **1-Coomassie Brilliant Blue (CBB) Stock**

Coomassie Brilliant Blue G-250	5 g
dH <sub>2</sub> O	100 mL

#### **2-Fixation**

40 % Ethanol	125 mL
10 % Acetic Acid	25 mL
50 % dH <sub>2</sub> O	100 mL

The gel is shaken in this solution for 1 to 2 hours

#### **3-Washing**

The gel is washed with dH<sub>2</sub>O two times for 10 min.

#### **4-CCB dye solution**

Ammonium sulfate	100 g
85% phosphoric acid	12 mL
CBB stock solution	20 mL
Distilled water add to	1000mL

### **5-CCB staining**

CCB dye solution	200 mL
Methanol	50 mL

The gel is put in 200 mL CCB dye solution then 50 mL methanol is added and the gel is kept in this solution for 24 to 48 hours.

### **6-Washing**

The gel is washed with dH<sub>2</sub>O until protein spots are clearly visible.

## APPENDIX B

### Chemicals and Their Suppliers

<b><u>Chemicals</u></b>	<b><u>Supplier</u></b>
$\beta$ -mercaptoethanol	Merck
Acetic acid	Merck
Acetone	Merck
Acrylamide	Sigma
Ammonium sulfate	Merck
Ampholines pH (3-10)	Fluka
Bis-acryamide	Sigma
Bovine Serum Albumin (BSA)	Sigma
CaCl <sub>2</sub> H <sub>2</sub> O	Merck
CH <sub>3</sub> CN	Applichem
CHAPS	Merck
Comassie Brilliant Blue G 250	Sigma
DTT	Fluka
Ethanol	Merck
Glucose	Merck
Glycerol	Merck
Glycine	Merck

H <sub>3</sub> PO <sub>4</sub>	Merck
HCl	Merck
IPG strips	Amersham
KH <sub>2</sub> PO <sub>4</sub>	Merck
Methanol	Merck
MgSO <sub>4</sub> 7H <sub>2</sub> O	Merck
Molecular Weight Standard (14,400-116,000)	Fermentas
NaOH	Merck
NH <sub>4</sub> Cl	Merck
NH <sub>4</sub> HCO <sub>3</sub>	Applichem
SDS	Sigma
TEMED	Sigma
TFA	Applichem
Thiamine	Sigma
Thiourea	Fluka
Trichloroacetic acid (TCA)	Merck
Tris- HCl	Sigma
Urea	Fluka