Proteome Profiling of *Saccharomyces cerevisiae* stress response to Cumene Hydroperoxide (CHP)

Leepika Tuli

A Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
In partial fulfillment of the requirement for the degree of

Doctor of Philosophy
in
Genetics, Bioinformatics and Computational Biology

Dr.Vladimir Shulaev, Chair
Dr. David Bevan
Dr. Ina Hoeschele
Dr. Pedro Mendes
Dr. Reinhard C. Laubenbacher

Blacksburg, Virginia
July 25, 2008

Keywords: *Saccharomyces cerevisae*, oxidative stress, proteomics,
bioinformatics, systems biology, mass spectrometry

Proteome Profiling of *Saccharomyces cerevisiae* stress response to Cumene Hydroperoxide (CHP)

Leepika Tuli

ABSTRACT

Oxidative stress, described as the state of disturbed intracellular redox balance, has been associated with several human conditions including ageing, apoptosis, cancer, autoimmune and neuro-degenerative diseases. Stress studies have shown that reactive oxygen species (ROS) and reactive nitrogen species (RNS) along with its intermediates can attack essential cell targets such as: DNA, proteins, lipids and carbohydrates, leaving behind dysfunctional biologic molecules. In effect, a cell's primary response is to involve several defense mechanisms that are under a complex and intricate regulatory control to repair any damages that may have occurred. Although several stress studies have been conducted in the past that have approached this biologically complex process step by step, application of a *Systems Biology* towards a comprehensive understanding is still emerging.

The current objective of this project is to identify proteins that change in response to cumene hydroperxoide (CHP) treatment and in parallel make an attempt to uncover events and processes that are a part of CHP-induced oxidative stress response. From a systems biology viewpoint, the Yeast Oxidative Stress project will monitor response at three different levels: transcriptomics, proteomics and metabolomics, with dynamic changes being measured from 3 to 120 min after CHP addition. Data collected from the different levels will be integrated to accomplish a holistic viewpoint of stress response in the given system and to develop mathematical tools for modeling biochemical networks.

Saccharomyces cerevisiae was chosen as a model, based on its availability of a completely mapped genome sequence with a collection of null mutants that was relevant to our fundamental research of stress response mechanism. Yeast, a simple unicellular eukaryote has been extensively used for applied studies and has proven to be indispensable for stress research. Information derived from this project can reveal

response mechanisms used by higher eukaryotes, especially if via analogous signaling cascades that are comparable between organisms.

Current research investigates an optimal workflow for generating 2D gel based protein expression data and identifying proteins that are induced by cumene hydroperoxide treatment. A non-targeted protein profiling followed by a 2-way ANOVA analysis provided a list of proteins that differ significantly between treatments. Protein identification provided relevant information on which proteins are affected by CHP induced stress response, including posttranslational modifications of peroxiredoxins. Redox active protein, Ahp1, was regulated post-translationally with sulfonic acid modification observed for its active Cys(62) residue.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to Dr. Vladimir Shulaev as my PhD advisor, for his invaluable support and constant guidance during my four years of doctoral program. Working under his supervision has been an amazing learning experience. His hard work, determination, commitment to the advancement of science, and kindness has been very inspiring to me. I thank him for his patience and insightful suggestions for my research.

I am grateful to my graduate committee members, Dr. Bevan, Dr. Hoeschele, Dr. Mendes and Dr. Laubenbacher for their constant support, encouragement and detailed guidance that has been an invaluable input for this project.

I would also like to extend special thanks to fellow members of our lab- Dr. Joel Shuman, Dr. Ana Martins, Deigo, Dr. Teruko Oosumi, Hope, Sarah, Dr. Sunil Bajad, Debbie and members from the Biochemical Networks Modeling Group and Yeast Systems Biology Group, including Dr. Wei Sha, Dr. Bharat Mehrotra, Alberto de la Fuente, Dr. Diogo Camacho, Autumn Clapp, Elena Dimitrova, Dr. Dustin Potter, Dr. Brandy Stigler and the Principal Investigator and co-PI's of the yeast project Dr. Reinhard Laubenbacher and Dr. Pedro Mendes, Dr. Vladimir Shulaev respectively.

I would like to acknowledge Dr. Amrita Cheema, Dr. Helm, Dr. Keith Ray, Dr. Iuliana Lazar, Dr Jenny Armenta, Kris Lee, Clive Evans and Don for their valuable comments and for sharing their knowledge on my research.

My special thanks go to my family and friends that have been a constant source of support and fun and for being there when I needed them the most; especially my nephews for making a difference in my life by their presence.

Finally I would like to acknowledge the National Institute of General Medical Sciences (grant R01 GM068947-01), National Science Foundation (grant DBI-0109732) and Virginia Bioinformatics Institute for their financial support to conduct this research

TABLE OF CONTENTS

ABSTRACT .... **II**

ACKNOWLEDGEMENTS .... **IV**

LIST OF TABLES .... **VII**

LIST OF FIGURES .... **VIII**

LIST OF ABBREVIATIONS .... **X**

SUMMARY .... **XI**

CHAPTER 1 .... **1**

A. INTRODUCTION **2**

A1. FREE RADICAL MECHANISM AND REDOX REGULATION 3
A2. MONITORING AND MEASURING STRESS RESPONSE 6
A3. RELEVANCE OF OXIDATIVE STRESS AND CHP 8
A2.1 Human Disease: *9*
A2.2 Crop Protection *10*
A2.3 Cumene Hydroperoxide (CHP) *11*
A4. MODEL ORGANISMS FOR STRESS STUDIES 12
A4.1. Saccharomyces cerevisae: *12*
A4.2 Other model organisms *23*
A5. TOWARDS INTEGRATED UNDERSTANDING OF STRESS RESPONSE 27

CHAPTER 2 .... **30**

B. PROTEOMICS AND SYSTEM BIOLOGY **31**

B1. PROTEOMICS 31
B1.1 Types of Proteomics *32*
B1.2 Current Proteome Profiling Technologies *34*
B1.3 2D or not 2D *44*
B1.4 Novel Approaches in Proteomics *44*
B1.5 Mass Spectrometry *47*
B1.6 Database Utilization *52*
B.2 SYSTEMS BIOLOGY 55

CHAPTER 3 .... **60**

C. EXPERIMENTAL DESIGN, METHOD DEVELOPMENT AND DATA ANALYSIS **61**

C1. EXPERIMENTAL DESIGN 61
C1.1 Reference Gel/Master Gel *62*
C1.2 Optimal Sample Extraction, Separation and Visualization Technique: *66*
C1.3 Two-dimensional gel electrophoresis: *70*
C2. PROTEOME INFORMATICS 71
C2.1 Analysis of gel images in PDQuest: *72*
C3. STATISTICAL ANALYSIS OF PROTEIN EXPRESSION DATA 78
C3.1 Normalization: *79*
C3.2 Missing Value Imputation: *84*
C3.3 Data Transformation: *85*
C3.4 2-way ANOVA model: *86*
C4 PROTEIN IDENTIFICATION WITH MASS SPECTROMETRY 88

CHAPTER 4 .... **91**

D. RESULTS AND DISCUSSION **92**

D1 PEROXIREDOXINS 97
D2 OTHER SIGNIFICANT PROTEINS: 105
D3 CONCLUSIONS 111

CHAPTER 5 .... **114**

E. FUTURE STUDIES **115**
F. APPENDICES **117**
G. REFERENCES **119**

LIST OF TABLES

Table 1: Comparison of commonly used gel stains for visualization of separated proteins........... 69

Table 2: Comparison of 2-way ANOVA results from log-transformed intensities normalized using four different methods. Protein annotations that are significantly differentially expressed (pFDR< 0.05) for quantile normalised method are shaded in yellow. TQVS*: total quantity of valid spot, LOESS[56]: Local regression model, (): :number of spot significant.. 87

Table 3: List of 2D resolved proteins identified by in-gel trypsination and sequencing of generated peptides. Data on the encoded proteins come from the SGD (Saccharomyces Genome Database). Proteins are presented in alphabetical order of their annotation......... 95

LIST OF FIGURES

Figure 1: Mechanism of reactive oxygen species production..3
Figure 2: Involvement of Grx3p and yap1p in sensing hydrogen peroxide induced damage to activate transcription of antioxidant defenses. Reprinted by permission from Elsevier Publishers Ltd: (Trends in Cell Biology) [27], copyright (2005)..19
Figure 3: PKA regulation of genes involved in stress, nutrient and glycogen accumulation. Reprinted by permission from Mcmillan Publishers Ltd: (EMBO Journal) [41], copyright (1998)...24
Figure 4: Schematic depiction of ROS sensing and the role of MAPK signaling cascade Reprinted, with permission, from the Annual Review of Plant Biology, Volume 55 © 2004 by Annual Reviews www.annualreviews.org [2] 24
Figure 5: Proteomics workflow of 2D gel based approach experiment 45
Figure 6: A simple complexity pyramid composed of DNA, RNA, proteins and metabolites. From Oltvai and Barabasi, SCIENCE 298: 763 (10/25/2002). Illustration k. Sutliff Reprinted with permission from AAAS [68] 58
Figure 7: Experimental design of gel triplicates for each of the eight time points and two groups (control vs treatment). Ethanol (EtOH) was added to the control fermentors, while CHP was added to the treated samples Samples were collected immediately before addition of CHP or EtOH and at: 3, 6, 12, 20, 40, 70 and 120 minutes following addition of CHP or EtOH 65
Figure 8: Gels (7 cm IPG strip) loaded with 15 µg of total soluble protein extracted from wild type yeast using: A) Yeast extraction kit, B) RIPA buffer, C) Thiourea/Urea. D) E. coli proteins were extracted as a control for comparison using E.coli extraction kit 66
Figure 9: Gels loaded with 150 µg total soluble proteins from wild type yeast using A) Yeast extraction kit B) RIPA buffer. Samples were run on 24 cm IPG strips: 1) pH 3-6 and 2) pH 5-8. 67
Figure 10: Comparison of four different stains for protein detection using 24 cm IPG strips A) Coomassie Blue, pH 3-10; B) Silver stain, pH 5-8; C) Sypro Ruby pH 3-10; D) Flamingo Pink pH 3-10 68
Figure 11 : Screenshot of a reference gel loaded with 300 µg of total soluble protein on a 24 cm IPG strip, pH 3-10 71
Figure 12: Gel warping step: Alignment of two images with their distortion mesh. A) Sample image B) Warped version of sample gel to align with, C) Reference image..........................75
Figure 13: Screenshot of a wt-T0-chp matchset using PDQuest software with Ref: reference gel; R1,R2 andR3: three biological replicates. A) Manual matching (pink square) with Landmark spots (green square) B) Automated spot matching error (blue oval in R1) ; observed as missing value in data set ,which can be corrected by manual validation. 77
Figure 14: Normalization of expression data. Boxplots of spot intensities from each experimental group are plotted for A) log transformation of total quantity of valid spot method B) log transformation of total quantity of valid spot (406 spots) , C) log transformed LOESS and D) Qunatile Normalization method 83
Figure 15: (A) Q-Q plot of normalized, untransformed spot intensities for three biological replicates of wt_t0_chp treatment matchset. (B) QQ plot of normalized spot intensities after log transformation for the same three replicates. The straight line represents a normal distribution and the log transformation of normalized intensities lays all points close to this line. 85
Figure 16: Hypothetical model of cysteine modifications of peroxiredoxins via Srx1..................98
Figure 17: 2D-PAGE analysis of oxidized and reduced forms of Ahp1p and Tsa1p in yeast cells after exposure to CHP. CHP was added to the cultures immediately after sample collection at t=0 and a time-course was followed up to 20 min, as indicated. Results shown are one of the three biological replicates..99
Figure 18: Dynamics of changes in the levels of Ahp1 and Tsa1 at the transcriptional (A) and proteome (B,C) levels. Bold symbols correspond to control (non-treated) cultures, while open symbols are used for CHP-treated cultures. Results shown are mean ± standard deviation of three biological replicates. Column (B) shows the time-course of changes for the oxidized forms of Ahp1p and Tsa1p, while (C) shows the same time-course for the oxidized forms (one Ahp1p^{ox}, and two Tsa1p^{ox}). 100

Figure 19: The MALDI-TOF mass spectra corresponding to A) AHP1 identification and B) two forms of Ahp1p are shown. A1) Database query result with significant identification made to Ahp1p; A2) Representative protein view with list of 8 peptides used towards Ahp1p identification; and A3) Representative product ion spectrum of *m/z* 3337±1 and its *b* and *y* ions. B1) Peak near 3338±1 corresponds to the peptide containing the active-site cysteine (modified with sulfonic acid) and peak near 3347 corressponds to alkylated cysteine (modified with iodoacetamide 104
Figure 20: Dynamics of changes in the levels of TIF51A, TDH1 and EAF3 at the transcriptional (A) and proteome (B) level. Pink symbols correspond to control (non-treated) cultures, while blue diamond is used for CHP-treated cultures. Results shown are mean± standard deviation of three biological replicates 107
Figure 21: Dynamics of changes in the levels of IMD2/3/4 at the transcriptional (A) and proteome (B) level. Pink symbols correspond to control (non-treated) cultures, while blue diamond is used for CHP-treated cultures. Results shown are mean± standard deviation of three biological replicates. 109
Figure 22: Spot profile of TIF51A (A) ; TDH1 (B); EAF3 (C) and IMD2/3/4 (D) in yeast cells in both control vs CHP treatment groups. CHP was added to the cultures imeediately aftre sample collection at t=0 and a time-course was followed up to 20 min, as indicated. Results shown are one of the three biological replicates. 110
Figure 23: Time series of significant protein spots (pFDR <0.05). CHP sample is labeled in blue and the control samples are labeled in blue. Proteins identified for each spot include : A) RPN 13, B) TDH2, C) PRE8, D)TPI1, E)EGD2, F)HPT1, G)CAP1, H) TDH1/2/3 and I)PEP4 117
Figure 24: Time series of significant protein spots (pFDR <0.05). CHP sample is labeled in blue and the control samples are labeled in blue (continued). Proteins identified for each spot include: J) TKL1, K) YBR026C (ETR1), L)AAT2, M)NPT1, N)SAM1, O)ERG13/ PGK1, P)KGD2 and Q) CYS4 AND AAT2. 118

LIST OF ABBREVIATIONS

2DE	Two-dimensional electrophoresis
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
AHP1	Alkyl Hydroperoxide
ANOVA	Analysis of Variance
ATCC	American Type Culture Collection
CHCA	Cyano-4-hydroxy cinnamic acid
CHP	Cumene Hydroperoxide
COH	Cumyl alcohol
C_p	Peroxidatic cysteine
C_r	Resolving cysteine
CTR	Control
CYS	Cysteine
Da	Dalton
DTT	Dithiothreitol
ESI	Electrospray Ionisation
gpx1Δ	GPX1 gene mutant
GSH	Glutathione(reduced)
GSSG	Glutathione(oxidised)
H_2O_2	Hydrogen Peroxide
IEF	Isoelectric focussing
IPG	Immobilized pH gradient
LC	Liquid chromatography
MALDI	Matrix Assisted Laser Desorption Ionization
MAPK	Mitogen-activated protein kinase
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MW	Molecular Weight
Nano-LC	Nano liquid chromatography
OS	Oxidative Stress
PCA	Principal Component analysis
PCD	Programmed Cell Death
PD	Parkinson's disease
pI	Isolectric point
PMF	Peptide mass fingerprint
PMSF	Phenyl methyl sulfonyl flouride
Ppm	Parts per million
PRX	Peroxiredoxin
RNS	Reactive Nitrogen Species
ROI	Reactive oxygen intermediates
ROS	Reactive Oxygen Species
RP-HPLC	Reverse-phase high performance liquid chromatography
SDS	Sodium dodecyl sulfate
SOD	Superoxide Dismutase
t-BOOH	Tert-butyl alcohol hydroperoxide
TF	Transcription Factor
TOF	Time-of flight
TPX	Thioperoxiredoxin
TSA1	Thioredoxin Peroxidase
WT	WildType
yap1Δ	YAP1 gene mutant

SUMMARY

The objective of this project is to elucidate underlying mechanisms of oxidative stress response in yeast induced by cumene hydroperoxide (CHP) oxidant. The specific aims include:

(1) Develop an optimal protein extraction, separation and visualization method for generating two-dimensional (2D) gel electrophoresis-based protein expression data set that can be used for modeling purposes.

(2) Generate an annotated “reference gel” map that represents a spot pattern for all spots, possibly detected on a given sample gel in this experiment.

(3) Establish differential protein expression using a suitable statistical model.

(4) Identify differentially expressed proteins and classify biochemical pathways significantly affected by CHP inducement.

To address the above mentioned objectives, this dissertation is divided into five chapters. Chapter 1 describes the oxidative stress response and discusses its relevance in *Saccharomyces cerevisae* and other model organisms. Chapter 2 describes the protein profiling technologies and the novelty of *systems biology* approach and its application to investigative studies from a holistic viewpoint. Chapter 3 discusses the experimental design and our approach to generating 2D gel electrophoresis based protein expression data and current workflow for systematic statistical analysis of protein expression data. Chapter 4 outlines results for significantly induced proteins and its biological implications during oxidative stress response. Suggestions for future work are presented in Chapter 5.

CHAPTER 1

A. INTRODUCTION

As molecular oxygen (O_2) is needed for respiration and nutrient utilization purposes, all aerobic organisms get exposed to reactive oxygen intermediates eventually, at some point of time. An incomplete reduction of oxygen during respiration can lead to *in vivo* production of reactive oxidizing species (ROS). ROS can be extremely toxic and, therefore, cells developed defense systems that are critical for their survival. Plants are known to have an additional source of oxygen species production via photosynthesis which puts additional demands in terms of oxidative damage. Sometimes, reactive oxygen intermediates (ROI) generated by radiation, redox active drugs, exposure to light, and/or β oxidation of fatty acids are more toxic by nature than their parent species [1]. Earlier stress studies have focused more upon the damaging effects of ROS, than ROS as beneficial signaling molecules for intracellular communications [2]. With massive amounts of exploratory stress studies carried out in different organisms, it is evident that ROS can have different roles to play, depending upon the cell type, developmental stage and/or kind of stress factor involved [2]. Under normal conditions, intra-cellular antioxidants with scavenging abilities can restrict ROS (produced as a byproduct of aerobic metabolism) accumulation, however, when produced in response to host-pathogen interactions, ROS can serve as a signaling molecule with the host investing minimal scavenging properties [3]. In plants, certain environmental factors can activate oxidases and peroxidases to generate ROS that regulate programmed cell death (PCD) and stomatal behavior [3]. Reactive oxygen species can play a critical role against pathogens as a plant defense mechanism. For example, ROS produced by specialized NADPH oxidase, present on macrophage membranes; referred to as oxidative burst in certain cell types; can be beneficial against microbial attack. With different streaks of beneficial and harmful activities clubbed together, defining oxidative stress (OS) has become difficult as it is subjective to the current state of a cell. For example, some healthy cell types could indulge in oxidative challenge but not qualify as being oxidatively stressed [3]. Likewise, loss of antioxidants with no resulting damage to cell would represent a balanced redox system. Under such diverse conditions, oxidative stress can

be defined as a state of disturbed redox balance that is in favor of reactive oxidizing species, capable of causing extensive damage [3]. In other words, oxidative stress occurs when internal and/or external stimuli can influence the equilibrium between reactive oxidizing species and host scavenging molecules such that it leads to extensive biomolecular damage [4]

A1. Free Radical Mechanism and Redox Regulation

Different kinds of endogenously produced oxidizing species observed across organisms include: oxygen (ROS) or nitrogen (RNS) derived species and/or free transition metal ions released from ROS attack of metalloproteins [5]. Dioxygen, when physically or chemically activated can generate reactive oxygen species such as: superoxide anion, singlet oxygen, hydrogen peroxide and highly reactive hydroxyl radical ($OH^{\cdot}$), as seen in Figure 1. As is commonly believed, not all reactive oxidizing species are free radical ions, with hydrogen peroxide being one exception. The reactivity of a given radical depends upon its life span and its diffusability across membrane and to other

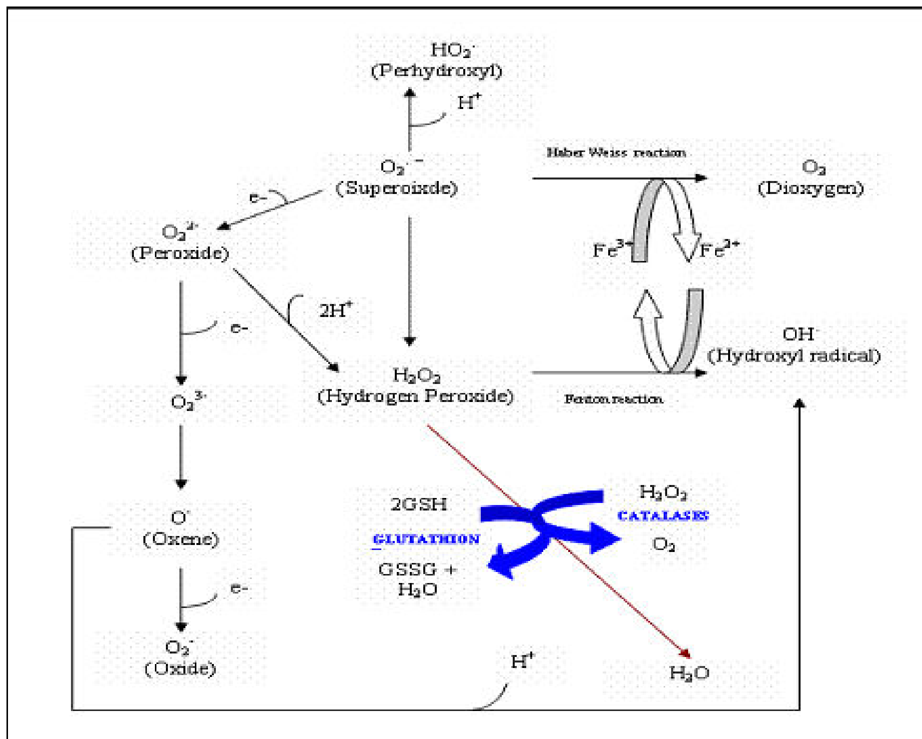


Figure 1: Mechanism of reactive oxygen species production.

cell locations. Oxidizing species such as H_2O_2, are less reactive as they are highly diffusible and can penetrate cells, imparting them more access to cellular compartments. On the other hand, OH^- radical is highly reactive due to its short half life and its presence of an unpaired electron that reacts with other molecules to generate new radical species further. Figure 1 describes some of the commonly used mechanisms used in generating reactive oxygen species (Fig.1).

The reactive species target essential biomolecules like DNA, RNA, proteins and lipids, yielding oxidation products and/or modified intermediates that are indicative of a stressed state. For DNA molecules, as many as sixty modifications, initiated by hydroxyl radical have been documented. These modifications range from strand lesions and base modifications to DNA cross-linking etc. [6]. Depending upon the interaction of hydroxyl radical with DNA, the radical can either add hydrogen atoms to DNA base or remove hydrogen from the backbone [7]. Studies on *E.coli* cells treated with H_2O_2 reveal three kinds of oxidants produced by Fe-mediated Fenton reaction that was responsible for DNA modification and as a result cell death [7] For some DNA modifications, iron-colloid formations have been observed where Iron has the ability to open supercoiled DNA and integrate itself in between [7]

Besides oxygen species, metal ions too have the ability to generate oxidative modifications of amino acids residues in protein molecules [7]. The most common sites for modifications include histidine, cysteine, arginine, proline and methionine residues. The modifications are typically site specific with only one residue frequently being modified. Residues with metal binding sites are viewed as specific targets where metal ions once bound make the protein site unavailable to free radical scavengers [7]. Chemical modifications of amino acids lead to prosthetic group modifications and change in protein charge states which combined with inter and intra-molecular cross links can account for protein aggregation observed. Cysteine is one residue that is known to occur in thiol, thiyl, disulfide, sulfenic, sulfinic and sulfonic forms, each being chemically distinct from others, making it more suitable for regulatory purposes, where small changes in chemical entities can regulate various functions [5]. Some of these transformations are irreversible whereas the ones that are reversible are better suited for sensor purposes and viewed

as regulatory switches [5] Besides oxidative modifications, mechanisms such as hydroxylation, electron transfer and exchange reactions can also lead to redox transformations of proteins, which seem to vary in response to changing intracellular conditions [5]. Damaged proteins generated as a result, are then targets for degradation by proteases or end up as regulatory molecules for a while, before being converted back to their native forms [6].

For lipids, hydroxyl radical is shown to initiate peroxidation reaction by removing hydrogen atoms from poly-unsaturated fatty acids embedded in biological membranes [7]. The cell membrane is disrupted and lipid peroxidation products including aldehydes such as malonaldehyde (MDA), 4-hydroxyalkenals, and 4-hydroxynoneal are generated [8]. Aldehydes are reactive, as they are capable of interacting with biological molecules to form DNA adducts and protein aggregates [7]. Aldehydes can also act as secondary messengers due to their long half life and can diffuse to other locations, besides their original site of production [8]..

Since cells are constantly sensing and adapting to the redox perturbations in their environment, induction of genes or other components, to help maintain a reduced cellular redox environment is typically how cells would respond [8]. Maintaining critical intracellular environment, by means of redox regulation, can be highly complex, as any given biological system is dynamic by nature. Redox regulation can occur in two ways using direct or indirect mechanisms in direct method, transcription factors (TF) interact directly with the oxidizing species or its ROS intermediates. ROS can directly interact with the cysteine thiol groups present on the factor to transform its properties and initiate a downstream process. Whereas, in an indirect mechanism, the MAPK cascade intervenes to alter transcription factor activity, before the factor can induce any gene regulation [3]. It is observed that MAP kinase phophorylates serine and threonine residues of transcription factor to modulate its regulation [3]. In yeast, Yap1 is the most extensively studied transcription factor also recognized to be crucial for oxidative stress signaling [3]. Oxidative stress studies induced by hydrogen peroxide have shown that Yap1 factor forms a disulfide-linked complex with the sulfenic acid intermediate of glutathione peroxidase (Gpx3), indicating to its role activity of an oxidant sensor in yeast [9]. The transient complex formed between yap1-

Gpx3, is known to rearrange itself to generate oxidized Yap1 which gets confined inside the nucleus, thereby inducing transcription of defense genes [9]. A similar mechanism exists in plants where nonexpresser of PR genes 1 (NPR1) is the redox-controlled transcriptional cofactor that is activated upon salicylic acid (SA) accumulation as a part of the systemic acquired response (SAR) [9]. NPR1 typically exists as an intermolecular oligomer in the cytoplasm but rearranges itself to a monomeric form during systemic acquired resistance (SAR) [9]. The monomeric forms are capable of traveling across the nuclear membrane and interacting with reduced TGA transcription factor to regulate gene expression [9]. For higher eukaryotes, it is the AP-1 (YAP1 homolog) factor and NFkB factors that are drawn in by a variety of oxidants and biological inducers for gene expression regulation [9].

Besides transcription factors, redox regulation can also be controlled via heat shock proteins, protein disulfide isomerases, endoplasmic reticulum (ER) glycoproteins, immunoglobin heavy chain binding proteins etc. [9]. These proteins act as molecular chaperones and are involved with protein folding and/or refolding practices or degradation of mis-folded proteins [9]. Comparative stress studies across various organisms prove that yeast, plants and higher eukaryotes share a somewhat similar sketch of cytoplasmic and ER stress response that can be categorized as being preventative, interceptive or repairing by nature, but differs only in the kind of components used or initiated. A description of similarity and differences observed for stress response mechanisms and components in model organisms (Yeasts, *E.coli and Arabidopsis thaliana*) are discussed in section A4 to formulate a better idea of what is already known [9].

A2. Monitoring and Measuring Stress Response

Monitoring activity of enzymes along with levels of anitioxidants and changes in gene, protein and metabolite expression levels, in a stress response provides clues to the adaptation and survival mechanism of a stressed organism.

In unicellular organisms, decrease in growth rate induced by adverse environmental conditions is one indicator of a cell being in a stressed state

[10]. Determined as CFU (colony forming unit), the cell viability of a population can estimate the concentration of a stressor that is more likely to initiate a stress response. Another assay that can monitor growth inhibition is measuring DNA content of cell populations by flow cytometry [6]. Since ROS are highly reactive and cannot be measured directly, indirect methods for quantifying antioxidant enzyme activity exist that can assess the toxicity mechanism of a stress factor [10]. For non-enzymatic systems, analytical methods developed include measurement of glutathione, metallothionein, glutaredoxin and/ or thioredoxin levels [10]. Monitoring global gene expression in response to stress factors is achieved by mRNA analysis, real-time PCR, serial analysis of gene expression and microarray chips. Similarly, monitoring of proteome changes is needed to receive accurate information on which proteins are involved in the stress mechanism. Techniques used towards such objectives include: 2D gel protein separation, isotope coded affinity tagging ICAT), surface enhanced laser desorption ionization-time of flight (SELDI-TOF), etc. [10]. Detection of protein carbonyl and reduced cysteine SH status can be monitored by dinitrophenyl hydrazine (DNPH) antibody reactivity and electrophoretic mobility respectively [6]. Lipid peroxidation products can also be measured but by means of chromatographic methods [10].

Since oxidizing species are highly reactive with a short life span, it is difficult to get a footprint of all intermediates/radicals present at a given time. However, the products of reactive species such as modified DNA, lipid peroxidation products and modified proteins are more stable in nature which are utilized for assessing a chemical pattern of stress existence and used as biomarkers for stress occurrence [11]. With the crucial role that oxidizing species play, it was essential to develop mechanisms by which reactive oxygen/ nitrogen species in cells/tissues or body fluids could be measured. In order to achieve this, it was important to devise techniques that could detect low intracellular concentrations of intermediate species in pace with the continuous scavenging activities of antioxidants [10] For some oxidizing species it was easy, for example reactive nitrogen species (RNS) can generate stable metabolites, such as nitrate and nitrite’s which are easily

measured in comparison to the parent species, but for the remaining reactive species, new methods are being devised and developed.

Within a human pathological scheme, a concept of biomarker was established that could provide the following information: i) extent of damage to any biomolecule 2) functional markers for blood flow, platelet aggregation or cognitive function and 3) endpoints for specific disease [11]. Glutathione (GSH) is one of the most commonly used biomarkers suggested for monitoring chronic disease progress. Measuring reduced and oxidized GSH levels in blood has been a useful indicator for stress in humans [11]. HPLC, GC-MS and HPLC-ESI-MS are some methods that have been optimized to identify and quantify glutathione forms in human fluids. However, even though DNA base adducts have been reported in neurodegenerative diseases such as: Alzheimer's disease (AD), amyloptropic sclerosis (ALS) etc., no valid nucleic acid based biomarker has been established as yet [11].

Proteins being the major components of a system were also examined for biomarker attributes but it was observed that not all proteins were equally sensitive and susceptible to oxidative modifications [11]. But a few that have surpassed the challenges of valid biomarker requirements include carbonylated proteins currently used for assessment of oxidative stress conditions and damage in several age-related diseases and neuro-degenerative diseases [11]. The need for reliable biomarkers at the protein level has seen a dramatic improvement in the area of redox proteomics which appears to be an exciting way of examining pathological processes at the biomolecular level. Widespread screening of modified proteins promises an insight into the complex phenomenon of stress, signaling and oxidative damage [11].

A3. Relevance of Oxidative Stress and CHP

With our knowledge of oxidative stress rapidly increasing in pace with the growing evidence generated from exploratory studies, the role of oxidative stress and its potential in human health optimization and plant protection is becoming powerful. This section emphasizes the relevance of investigative

oxidative stress studies for human healthcare and its implication in plants, as in the latter case, stress can have a powerful impact on crop protection.

A2.1 Human Disease:

Over the years, oxidative stress has received huge degree of scientific interest not because of its ubiquity, but for its profound and deleterious effects on human health. Parkinson's disease (PD), cancer, amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) are a few of the several human conditions that are known to progress with stress escalation [12]. Alzheimer's disease , known to affect 2% of the current US population, is estimated to affect 22 million people worldwide in the near future [13]. Several studies investigating this syndrome report a strong correlation of protein oxidation and markers of AD histopathology; with apparent evidence of protein oxidation condition for AD progression [13]. Besides AD, oxidative stress is known to be involved in cardiovascular diseases, based on evidence of antioxidant deficit observed in diabetic cardiomyopathy patients [14]. The study illustrates improved cardiac function in diabetic patients upon antioxidant administration [14]. A progressive loss of specific neuronal cell populations with deposition of protein aggregates has been characteristically associated with many human neurodegenerative diseases. It is speculated that oxidative stress causes death of neuronal cells which contributes to disease progression and pathogenesis [12]. HIV infected patients also demonstrate AIDS progression with increased levels of oxidative stress [15].

Other prevalent pathological conditions affected by oxidative stress include diabetes, allergy and age-related diseases [12]. Activated leukocytes observed in inflammatory response are known to generate hydroxyl radicals, which are an important resource of oxygen free radicals [7] At some level, age-associated decline in health conditions has been associated with accumulation of irreparable oxidative damage, a possible cause of decreased antioxidative defenses observed. It is a well known fact that biological aging progresses with the increase in oxidized biomolecule accumulation in most tissues [16] [11]. Age related conditions based on mice studies, reveal a correlation between oxidative damage in cerebral cortex and loss of cognitive

and motoric abilities [16]. Studies with *C. elegans* model depict that besides environment, the genetic composition of an organism can also influence its response to stress response conditions, which has initiated an interest in human population screening for genetic susceptibility to disease [16]. Respiratory distress syndrome, muscular dystrophy, and cataractogenesis are some additional conditions that demonstrate accumulation of protein carbonyls (oxidized proteins) over time [17]. Previous studies suggest that oxygen free radicals can contribute to cancer development via accumulation of radical-related DNA and protein lesions that can initiate and/or promote development of cancer [7]. It is believed that carcinogenesis requires permanent modifications of DNA which can lead to DNA aggregates, which are beyond repair and can sometimes accrue with age [7].

And while researchers worldwide are generating evidence for the involvement of oxidative stress in disease progression, some human diseases characterize oxidative stress as a consequence and not as an initiating event of primary disease development. Parkinson's disease (PD) is one example where oxidative stress is considered to be a central component but not a source of disease development. Likewise, pre-eclampsia is another example where activated neutrophils and endothelial cells release cytokines that add to the existing oxidative stress condition but the stress is not a cause of disease development [7].

A2.2 Crop Protection

Besides healthcare, another area where oxidative stress can have a major impact is in biotic and abiotic damages to plants and crops [7] Apparently reactive oxygen species can play multiple roles as in programmed cell death (PCD), in substrate lignification, as antimicrobial agents and signaling molecules [7]. Pathogen invasion of plant cells leads to rapid production of reactive radicals around the invasion site which induces localized cell death. NADPH oxidases, peroxidases (cell membrane) and amine oxidases (apoplast) are a few of the many reactive oxygen species, that allow for the onset of programmed cell death (PCD) in response to pathogen attack [2] [7]. During biotic stress, plant cells produce more ROS and lower the reactive

species scavenging activities. PCD, known to limit the spread of disease, typically starts off with a primary burst around the site of invasion followed by a secondary "microburst" required for systemic resistance [2]. The differences observed between plant and animal biotic defense is that plants have peroxisomes and chloroplasts as the main source of ROS production whereas in animals mitochondrial ROS production is implicated in eliciting PCD [2]. PCD can occur even during abiotic stress and even though the exact mechanism of ROS production and scavenging is not well understood for plants, its role in defense and signaling mechanism is well established [2]

One aspect that is still not well elucidated is the proposed mechanism of distribution of ROS to distal parts of plant cells [2]. Considerable evidence exists for the involvement of salicylic acid (SA) in the induction of systemic acquired resistance (SAR) or for *Agrobacterium* tumor formation. [7]. Besides signaling, ROS also appear to participate in plant-pest interaction, where external changes in host plant cause oxidative damage to the midgut of insects feeding on plants [7]. It is observed that recovery from drought appears to be accompanied by oxidative stress, requiring some kind of defense system induction. Studies on sensitive species such as tomato and spinach show that during cold and dark treatment, leaves become depleted of Cu^{+} ion and Cu/Zn SOD activity, which exacerbates oxidative stress injury upon illumination [7].

A2.3 Cumene Hydroperoxide (CHP)

Cumene hydroperoxide (CHP), is an organic peroxide that is hydrophobic by nature and can react with components of the cell membrane [6, 18]. It is a free radical generating compound that has been investigated for its oxidative stress effects on protein synthesis, cell lysis, organ necrosis and can initiate lipid peroxidation products [6, 18]. It is believed that CHP generates free radicals such as superoxide radical in the presence of transition metals [19]. CHP is similar to natural organic compounds, as it can be easily taken up by cells and cannot be metabolized by catalase. Comparative studies of four different oxidants: menadione, 1-chloro-2,4-dinitrobenzene, hydrogen peroxide and CHP in *S. pombe,* reveal different

kinetics for CHP regulated genes that can be induced to 2 hrs [19]. Cytotoxic aldehydes derived from CHP are capable of inhibiting protein synthesis in human skin fibroblasts and have been investigated for molecular mechanisms of chemical carcinogenesis in human cell cultures [18]. In an attempt to study the hydroperoxide metabolizing system, it was found that CHP treatment of hemoglobin free rat liver perfusion not only led to altered glutathione peroxidase activity but also modified cytochrome b5 present in hepatocytes [20]. Studies with rat liver show that CHP treatment oxidatively modified elongation factor 2 (EF2) thereby affecting the elongation step of protein synthesis [6]. Carbonyl content, ADP ribosylation of EF2 and its decreased amount in treated cell cultures (7 day CHP) indicate that CHP treatment led to polyribosomal aggregation accompanied by a decline in the rate of polypeptide synthesis [6]. Inhibition of polypeptide chain initiation led to monoribosomal aggregates which would lead to polyribosomal breakdown.[18]. No change in total amount of EF2 was observed in treated cells, which reinforced that the decline in protein synthesis rate was due to post translational modifications of EF2, instead of a decrease in its steady state level [6].

CHP has been used often for stress studies as an intracellular resource for reactive oxygen intermediate [18]. This project attempts to capture the effect of CHP treatment on yeast cells in a dynamic context. In view of other stress studies induced by H_2O_2, CHP has an advantage as it converts to cumyl alcohol (COH) when degraded which can be easily measured as opposed to H_2O which is the degradation product of H_2O_2.

A4. Model organisms for stress studies

A4.1. *Saccharomyces cerevisae*:

Saccharomyces cerevisae has been widely used as a eukaryotic cell model for stress studies. The choice stems from the wide use of molecular biology, initially developed on yeast system, its completely mapped genome sequence, easy laboratory handling and the availability of null mutants [21]. Additionally, the plausibility of comparing yeast ROS sensing and signaling to

higher eukaryotes makes it an indispensable model for studying redox control and metabolism.

This section provides a background on various oxidative stress studies carried out in yeast and their findings with partial overlap with many other diverse stress responses observed in the same organism. A vast range of oxidants that have been investigated for their stress response mechanisms include organic peroxide derivatives such as: tert-butyl hydroperoxide (t-BOOH), linolenic and linoleic acid and others like H_2O_2, menadione, diamide, diethylmaleate, lithium and etc. [6]. Yeast cells treated with oxidants, have shown varying tolerance to ROS activity depending upon the growth stage of their cells. Stationary phase cells (where all nutrients are exhausted) portray a higher resistance to temperature and oxidative stress in comparison to exponentially growing cells [6]. Likewise, a higher basal level of protein carbonyl content was observed for respiring cells, associated with higher production of endogenous ROS [6] . Apparently, respiration can have beneficial effects on oxidative stress tolerance as some catabolites (glucose) exert a negative regulatory effect on yeast antioxidants. ROS are also known to be involved in cell cycle and apoptosis mechanisms with menadione and H_2O_2 both causing cell arrest at G1 and G2 phases, respectively [6]. Studies with H_2O_2 treatment and UV irradiation show cell death aggravation with apoptosis properties [6].

Since all organisms have evolved some means of cellular defense against reactive oxidizing species, these defenses can be categorized as preventative and/or protective measures. Yeast cells possess both enzymatic and non-enzymatic defense systems that are used towards maintaining a balanced cellular redox state, as discussed below

Antioxidant enzymatic defense

This system includes several enzymes that have protective properties and are capable of removing oxygen species and their products, in addition to repair mechanisms for any damage that may have occurred. A systematic study of H_2O_2 treated yeast cells, reveal induction of several antioxidant enzymes such as: catalase, cytochrome peroxidase, superoxide dismutase,

thioperoxidases, thioredoxin reductases and glutathione reductases in response to stress conditions, some of which are described below [1] [6].

Catalase is an iron containing enzyme occurring in two forms, *viz*. cytosolic catalase A and peroxisomal catalase T, encoded by *CTA1* and *CTT1,* respectively [6]. These enzymes break H_2O_2 into O_2 and H_2O making *cta1* and *ctt1* mutant strains highly sensitive to H_2O_2. It is shown that while catalase genes are moderately induced to H_2O_2 treatment, both forms are needed for resistance against H_2O_2 [6]. Superoxide dismutase, is a metal containing enzyme that is present in nearly all aerobic organisms and in yeast occurs in two forms as cytoplasmatic Cu/Zn-superoxide dismutase (encoded by *SOD1*) and a mitochondrial Mn-superoxide dismutase (encoded by *SOD2*) [21]. The Sod1p is known to protect against O_2 toxicity whereas Sod2p acts as a primary scavenger protecting against O^-_2 generated radicals during respiration. Metal toxicity studies demonstrate an association between superoxide dismutase proteins and metal metabolism [21]. Another enzyme that is also involved in response to stress is the thioredoxin peroxidase (encoded by *TSA*), from the thiol redox pathway. *TSA* was found to be more specific to H_2O_2-induced stress in exponentially growing yeast cells [21]. It reduces H_2O_2 and other hydroperoxide derivatives, in conjunction with NADPH and thioredoxin reductase (encoded by *TRR1*) [21]. For example, *Trr1* mutants were found to accumulate the oxidized form of thioredoxin which depicted a slow growth phenotype [6]. Likewise, glutathione reductase, encoded by *GLR1*, is an enzyme that helps maintain a high GSH/GSSG ratio critical for the role of glutathione as an antioxidant as it converts oxidized form of glutathione to its reduced form. Strains mutant for this enzyme led to increased levels of oxidized glutathione (GSSG) making it hypersensitive to oxidants.

While these antioxidants are induced in oxidative stress response, considerable overlap has been seen between oxidative stress and other stress responses such as heat-shock, starvation, osmotic shock and resistance to heavy metals, which indicate the existence of common nodal points as regulatory targets.

Antioxidant non–enzymatic molecules.

The non-enzymatic defense system consists of small molecules that act as radical scavengers by interacting with ROS to be able to remove them from the system. Some examples of these molecules include: glutathione, trehalose, metallothioneins, thioredoxin, glutaredoxin that are involved in repairing mechanisms [6].

Glutathione (GSH), considered to be the first line of defense in tissue injury against electrophiles, maintains low redox potential inside cells by sustaining a high GSH/GSSG ratio. Studies in yeast cells show mutants with low GSH content are hypersensitive to H_2O_2, menadione and other stress-inducing agents such as cadmium and methyl glyoxal [6]. GSSG formed directly by glutathione peroxidase catalysed reduction of peroxide is known to be actively secreted in human liver cells in response to drug treatment which can also participate in non-enzymatic DNA repair [22]. Trehalose is another biomolecule with antioxidant properties that is also vital for resistance to environmental stresses. Similarly metallothioneins, a class of cysteine-rich small proteins with antioxidant properties, are known to bind to a number of free radicals in response to metal ion induced stress. In yeast, both *CUP1* and *CRS5* (encoding metallothioneins) have been found to be involved in protecting cells against various oxidants. Deletion of the TRX2 gene, encoding a cytoplasmic thioredoxin was seen to make yeast cells hypersensitive to H_2O_2 [21]. Similarly, studies focusing on glutaredoxins recognize its importance for resistance to oxidative stress induced by H_2O_2 [21]. Out of the five glutatredoxins, *GRX1* and *GRX2* were the only two induced upon exposure to H_2O_2 and menadione, however, no induction was observed for *GRX3, GRX4* and *GRX5*, genes that have only one cysteine residue at their active site [23]. It was observed that both Grx1 and Grx2 had similar patterns of glutathione S – transferase (GST) and peroxidase activity with H_2O_2, t-BOOH and CHP [24]. Grx5, established as the yeast glutaredoxin whose absence is responsible for the most dramatic effects on oxidative damage show a negative effect on the activity of mitochondrial protein with Fe/S clusters but not heme-containing proteins as mutants [25].

Besides inactivation of Fe/S containing clusters such as succinate dehydrogenase and aconitase, *grx5* mutants also share Fe/S deficiency phenotypes such as iron accumulation and the inability to grow under respiratory conditions [25]. In general, yeast has two parallel and inter-dependent reducing systems: glutathione and thioredoxin pathways that are linked to antioxidant defense, however differences occur only in components that make it an interesting regulatory point [5].

Regulation of gene expression.

Besides focusing on components of a defense system, studies conducted in the past show that oxidative stress response in yeast is regulated at the transcriptional level as well. Gene expression studies on yeast response to H_2O_2 and menadione depict induction of target genes that are under transcription factor regulon of stress response elements [26]. The target genes include elements from glutathione and thioredoxin detoxification system as well as other antioxidant components such as catalase, superoxide dismutase etc as described earlier.[26] Out of the many transcription factors identified in yeast, Yap1, Yap2 and Gcn4 are few of the bZip transcription factors, extensively investigated in stress responses [21]. The involvement of Yap1 in regulating gene expression in response to oxidants was first suggested when yap1 mutants of yeast were found to be hypersensitive to oxidants [21]. Presence of Yap1p-binding sites in the promoter region of *SOD1 ZWF1, TRX2, GLR1* and *GSH1* genes, all encoding proteins with antioxidant activities, suggested that their expression may be Yap1-dependent and/or oxidant inducible. Besides oxidative stress, the Yap family of factors have been implicated in other stresses such as: osmotic, arsenic, drug and heat induced [21]. Yap1 activity, primarily found to be regulated at the level of intracellular localization is confined to cytoplasm under non-stress conditions. Upon exposure to diamide, diethylmaleate or peroxides, the two cysteine residues in Yap1p change their redox state forming an intermolecular disulfide bond (Fig. 1). This masks the nuclear-export signal (NES) present which is responsible for its export across the nucleus [25]. In a similar fashion,

the peroxiredoxin Gpx3p, found to be highly sensitive to H_2O_2, transfers its redox modification to Yap1p thus regulating its activity [25]. Although a lot is known for Yap1, further work is needed to elucidate the physiological function of Yap2 and other Yap factors, as their overexpression was observed to build resistance to 1, 10-phenanthrolin and cadmium treatments [21].

Msn2 and Msn4 are another set of transcription factors (zinc finger) that mediate stress responses by inducing gene expression through the stress responsive element (STRE) [27]. Genes induced by Msn2/4 include *CTT1, DDR2* and *HSP12*. These genes respond to a variety of stresses including heat-shock, osmotic, hydrogen peroxide and DNA damage [21]. Msn2/4 regulon is known to be involved with ubiquitin and proteasome degradation pathways and negatively regulated by ras-cAMP-protein kinase (PKA) pathway, also suggested for yap1 regulated gene transcription [28]. Msn2p acts very similar to Yap1 as its regulation is influenced by its cellular localization as well [21].

Besides Yap1 and Msn2/4, Skn7 is another transcription factor drawn in by yeast, in response to oxidative stress [29]. Skn7, implicated in regulation of cell wall biosynthesis, cell cycle and osmotic stress response, displays similar sensitivity to H_2O_2 as observed for yap1 mutants, upon mutation. [29] [30]. 2D gel electrophoresis applied towards investigating the role of Yap1 and Skn7 factors in defense genes induction, show cooperation between Yap1 and Skn7 factors to activate *TRX2* (thioredoxin) and *TRR1* (thioredoxin reductase) in response to H_2O_2 [29]. While defense functions suggest both Yap1 and Skn7 as essential for resistance to peroxides (H_2O_2, t-BOOH), only Yap1 was critical for resistance to cadmium stress [29]. It is observed that the partition between the Yap1 and Skn7 regulon correlates with two distinct classes of defense activity: i) Skn7 independent subset of genes includes ones with ROI scavenging activities whereas, ii) Skn7 dependent genes involved in glutathione and pentose phosphate pathways [29] Genes required for cadmium resistance were seen confined to Skn7-indpendent subset of genes under Yap1 regulon [29].

Similarly, studies with grx5 mutants that in a way simulate continuous oxidizing conditions depict down-regulation of Hap4 regulon genes involved in respiratory metabolism. The inhibition of respiratory metabolism during

moderate but continuous oxidative conditions was viewed as a means protective measure used by the cell [31]. grx5Δ led to 1) iron accumulation that is not available for cell utilization but promotes oxidative damage and 2) inactivation of enzymes that require Fe-S cluster activity [25]. Genes down-regulated in absence of functional Grx5 were found to be Hap4 regulated, which along with Hap2/3/5, is required for respiratory gene expression. A plausible mechanism suggested was depletion of metabolically available iron unregulated HMX1 (heme oxygenase) gene leading to heme depletion and thereby Hap4 inactivation [31]. However when no difference was observed between heme levels of the mutant or wt strain, the role of additional factors could no longer be denied [31]. Grx5 deficient cells, representing an incomplete respiratory chain, suggest that down-regulation of respiratory chain components via Hap4 activity is one mechanism by which ROS production due to respiration could be prevented [31]. Even though the mechanism of iron entry into mitochondria is still not clearly elucidated, a direct interaction between Grx5 and Fe-S assembly is definitely supported [31].

Thus, a comparison between different oxidants revealed that yeast cells respond differently depending upon the kind of stressor used. For example, cells responded differently to menadion and hydrogen peroxide induced stress in yeast cells, as *rad9* mutants were more sensitive to hydrogen peroxide but resistant to menadione [32] Microarray analysis of lithium-treated cells revealed upregulation of genes involved with energy reserve and monosaccharide metabolism and downregulation of genes involved with transcription, translation and nucleotide metabolism [33]. Reduced protein synthesis and increased detoxification were some of the various metabolic changes associated with linoleic acid hydroperoxide (LoaOOH) treatment of yeast cells [34] Loss in free glutathione was observed for LoaOOH-conjugate treated cells due to induced treatment permeabilization [34].

It is however, important to map out any regulation occurring at the transcriptional level to its end products, to be able to sketch a detailed and complete picture of stress response mechanism, as an example in this case.

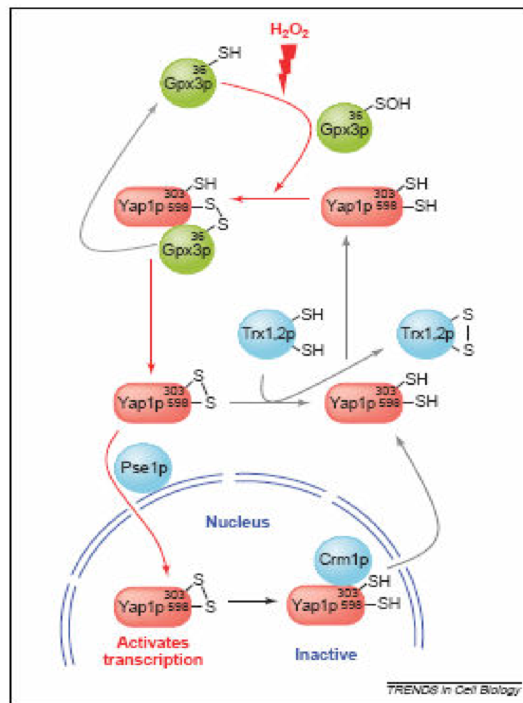


Figure 2: Involvement of Grx3p and yap1p in sensing hydrogen peroxide induced damage to activate transcription of antioxidant defenses. Reprinted by permission from Elsevier Publishers Ltd: (Trends in Cell Biology) [27], copyright (2005)

Regulation of protein expression

The inability to correlate differential mRNA expression to subsequent protein expression change has not only intrigued researchers but redirected their interest to the protein identification process for a given stress response [35]. Several studies conducted to this end have identified proteins as potential molecular effectors that were not regulated at the transcriptional level. For cadmium-induced stress, a proteomics approach revealed induction of sulfur amino acid biosynthetic pathway enzymes and heat shock proteins, proteases and enzymes with anti-oxidant properties enzymes [36]. Metabolic enzymes and proteins comprising the translational apparatus were found to be repressed by cadmium administration [36]. Likewise, studies investigating effects of different metal salts (including cobalt, copper, silver salts) demonstrated inducement of heat shock proteins, proteins from carbohydrate metabolism pathway and proteases [37]. Depending upon the kind of response generated and the number of proteins up/down regulated to fifteen different metal treatments, the metals were divided into three groups: i) Group

A: toxic metals such as Cu^{2+}, Cd^{2+}, Co^{2+}, Mn^{2+} that were found to upregulate more proteins ii) Group B includes essential metals such as Mg^{2+}, K^+, Na^+ that had an almost equal number of up/down regulated proteins and iii) Group C (Rb^+ and Fe^{3+}) that had more downregulated proteins [37]. With SOD1 expression changing for most metal treatments, it was clearly established that superoxide dismutase had a vital role to play in heavy metal detoxification [37]. A proteomic approach used quite successfully for studying effects of H_2O_2 on yeast cells showed inducement of heat shock proteins, proteases, enzymes from carbohydrate and amino acid metabolism, enzymes with antioxidant/reactive intermediate scavenging properties, and proteins from the translational apparatus and repression of metabolic enzymes involved in either glycolysis, Krebs cycle, purine and polyamine biosynthesis [1]. Although these results were in agreement with proteins typically involved for a stress mechanism, some elements were found to be oxidant specific. Heat shock proteins that are induced under various stress conditions have an additional advantage of acting as molecular chaperones that assist damaged proteins by refolding them or by subjecting them to proteolytic degradation [1]. After exposure to sublethal levels of oxidants, cells develop an adaptive response, which in the case of H_2O_2 studies appeared to be available until 45 min after treatment but was lost after 4 hrs of treatment [27].

In addition, studies on yeast cells with some peroxides revealed cross adaptation between oxidants. For example, adaptation to H_2O_2 protects cells against menadione and paraquat (superoxide-generating compounds) stressors [27]. Lithium toxicity of cells growing in glucose revealed inhibition of protein biosynthesis machinery including components such as ribosome proteins, proteins from nucleotide metabolism and methionine metabolism [33]. On the other hand, proteins involved with carbohydrate metabolism were found be at increased expression levels [33]. A proposed overview of the regulatory network controlling diauxic shift in yeast for *snf4*Δ mutant, reveal a fewer number of protein spots detected, which in a way suggests that protein synthesis transiently decreases after glucose exhaustion [38]. Deletion of *snf*4 causes a defect in expression of genes regulated by Cat8p which includes a variety of processes including protein degradation and mitochondrial function [38]. Glycolytic enzymes were found to be enhanced in the snf4 mutant strain

including hexokinase, enolase, alcohol dehyrogenase, transketoalse etc. which suggests that Snf4 regulates gene expression in response to glucose depletion [38]. cAMP, acting as a regulatory signal upstream of MSn2/4 and Snf1/4 pathway, was suggested to drop with glucose exhaustion during diauxic shift [38]. Previous studied with limited nitrogen showed increased expression of H^+-ATPase complex whereas cytochrome c oxidase expression increased under carbon limitation [39]. It was observed that GDH1 transcription was repressed by glucose and not at all induced by nitrogen limitation. Similar changes were observed for GLN1 transcript and protein levels and it is suggested that due to fast protein synthesis or slower protein degradation process, amino acids are in a position to control translation and degradation of protein accordingly [39] .

Redox Sensing and Signalling:

With the elements of antioxidant defense system identified and information on genes and proteins involved for any stressor, an understanding of the signaling transduction pathways involved is crucial as they are the regulators and coordinators of stress regulons for a given organism. An interest in the inherent mechanisms of redox control, sensing and maintenance, has led many researchers to consider similarities observed between organisms in terms of their redox pathways influenced.

In yeast, the Orp1- H_2O_2 redox relay system is similar to its OxyR-H_2O_2 counterpart of *E. coli*. Orp1 (oxidant receptor peroxidase), identified as the homolog of the glutathione peroxidase (GPX) family of scavengers, was recognized as a sensor and signal transducer that could interact with the peroxide via its cellular cysteines [40]. Orp1, known to carry three cysteine residues, had Cys^{82} engaged in peroxidase activity and was recycled by thioredoxin instead of GSH [40]. Cys^{36} of Orp1 senses the H_2O_2 signal and oxidizes to Cys^{36}-SOH which transduces its signal to Yap1 (via Ybp1) by forming a intermolecular disulfide bond between its own Cys^{36} and Cys^{59} of Yap1. Ybp1, known to act as a chaperone that brings both Orp1 and the transcription factor together, is also implicated in a scaffold for a non-redox

complex [40]. The Orp1-Yap1 switch is specific in being a on-off mechanism that can accommodate one redox exchange reaction between oxidized Orp1 and reduced Yap1. Another difference in comparison to its *E. coli* counterpart is that Orp1 responds only to peroxides, whereas OxyR can be elicited by various oxidants such as nitrosothiols, peroxides, thiol oxidant diamide and disulfide stress [40]. The distinction lies in the effect of various oxidants on Yap1 but independent of Orp1 and modification of the latter [40]. Besides the Orp1-Yap1 sensor, other redox switches include: reduced glutathione/oxidized glutathione; reduced thioredoxin/oxidized thioredoxin; NADPH/NADP+; protein-SH/protein-S-SR and ascorbate /dehydroascorbate that can buffer the redox environment of a cell [41].

As compared to *S. pombe*, regulation of the yeast signal transduction pathway in response to oxidants is still not well understood. Hog1 was the first MAPK found to be involved in high osmolarity stress [42]. Its role in oxidative stress was suggested when mutants of hog1 were found sensitive to diamide and H_2O_2 treatments. The mitogen-activated protein kinase (MAPKs) and c-AMP pathways are clearly important in signaling mechanisms that have been extensively studied for their role in stress response and regulation of gene expression [41]. MAPK, linked to its upstream receptors/ROS sensors is a linear cascade of three consecutively acting protein kinases that transduces extracellular signals by modulating activity of intermediary proteins (Figure 3) [43]. The three kinases functionally interlinked in a linear cascade have been identified in other eukaryotic organisms including yeast and mammals. In eukaryotic cells, MAP kinase can be regulated transcriptionally, translationally and post-translationally [44]. Typically the MAPKs are activated by serine/threonine and /or tyrosine phosphorylation whereas inactivation requires activity of specific phosphatases [44]. Similarly, genes regulated by oxidants are responsive to cAMP levels which are known to involve transcription factors, STRE and PDE elements [21]. The role of MSn2/4 and Yap1 factors and their interaction with the ras-cAMP pathway is well established [45]. CTT1, SSA3, HSP26 and HSP12 are a few of the genes induced by decreased levels of intracellular cAMP [21]. Heat shock, carbon and nitrogen starvation, cell entry into stationary phase and tolerance to freeze thawing of yeast cells are some of the stresses regulated by cAMP

levels. However, the exact mechanism of transcription factor regulation by cAMP and protein kinase is not well understood [21].

Sln1 is one of the components of a multistep posphorelay system, identified upstream of all yeast stress-activated protein kinases (SAPK) cascades, that can sense the environment and pass on signal downstream. It has been characterized well in eukaryote systems [42]. The Sln sensor kinase is transmembrane protein that has a cytoplasmic histidine domain which is active (under low osmolarity conditions) enough to autophosphorylate its histidine residue [42]. The phosphate is transferred to Ssk1 regulator which is inactive in this form and can directly bind with Ssk2. Another response regulator that also acts as a transcription factor is SKN7, which has a phospho-accepting aspartate at its receiver domain along with a DNA binding domain. Skn7 mutants were seen to be hypersensitive for H_2O_2 treatment [42].

A4.2 Other model organisms

E.coli:

In *E.coli*, defense against peroxides is mediated by the OxyR transcriptional regulator whereas for superoxide, it is the SoxRS regulon. The OxyR gene triggers expression of reductive activities such as enzymes and other antioxidants that can degrade the oxidant and reduce disulfide bridges (with glutaredoxin). Some examples of genes regulated by OxyR include; NADPH – dependent alkyl hydroperoxide, glutaredoxin, catalase (HP1 and HP2) and protective DNA binding protein (Dbs) [8]. DNA binding protein is known to protect DNA damage from oxidative stress and supports the DNA-repair system. The SoxR regulon induces genes such as: Mn-SOD, aconitase, fumarase , Gluc-6-P-DH and endonuclease IV [8].

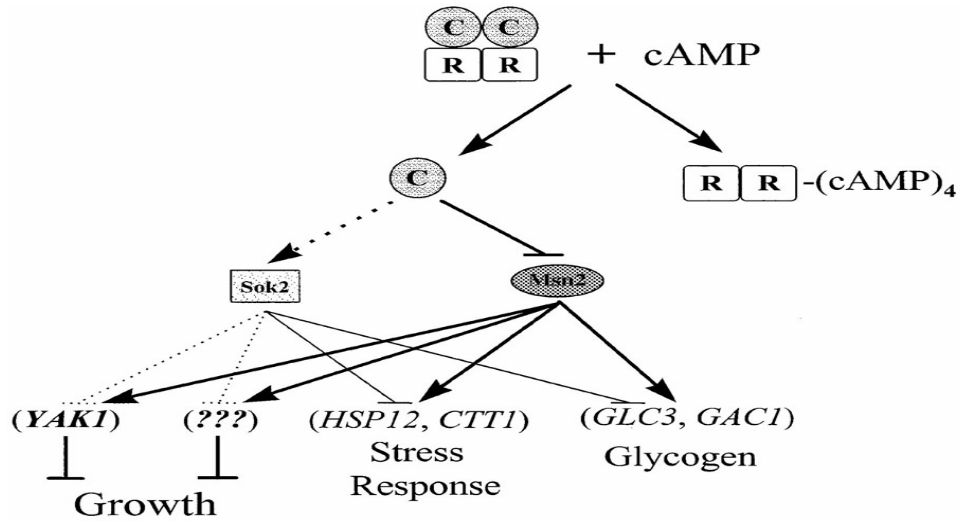


Figure 3: PKA regulation of genes involved in stress, nutrient and glycogen accumulation. Reprinted by permission from Mcmillan Publishers Ltd: (EMBO Journal) [41], copyright (1998)

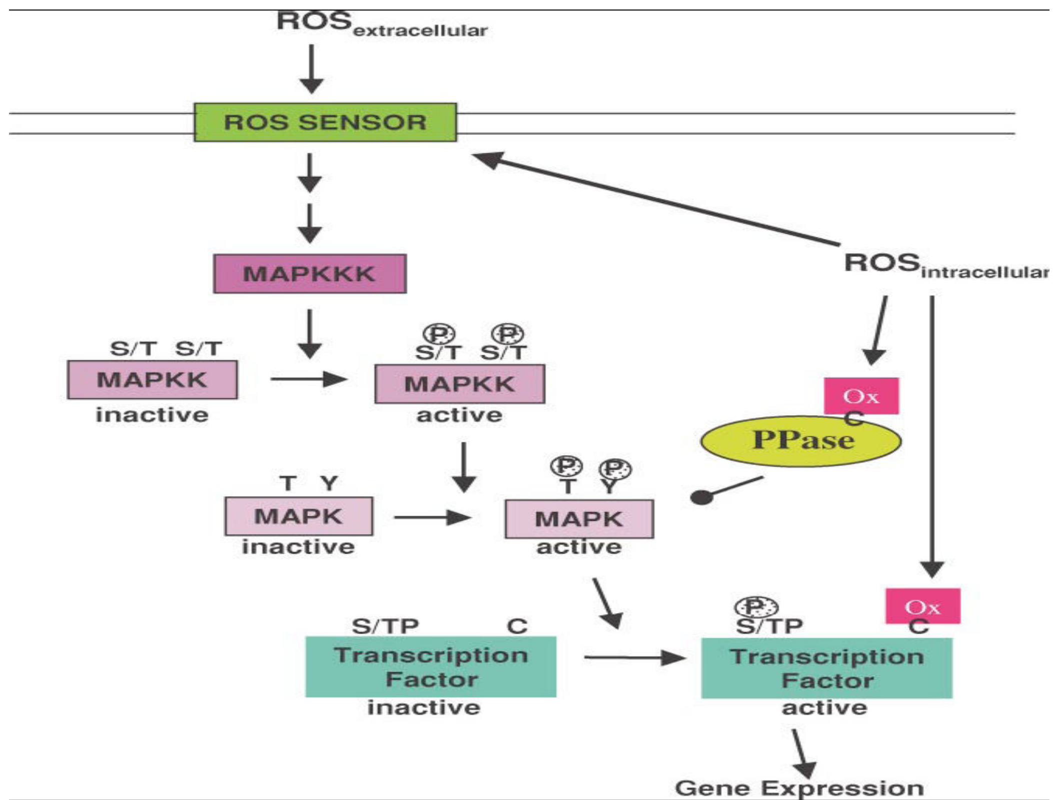


Figure 4: Schematic depiction of ROS sensing and the role of MAPK signaling cascade Reprinted, with permission, from the Annual Review of Plant Biology, Volume 55 © 2004 by Annual Reviews www.annualreviews.org [2]

Other regulatory systems known to exist for oxyR mutants show induction of many proteins that make cells more susceptible to H_2O_2. SoxR is proposed to be activated by a reversible one-electron oxidation of its Fe-S clusters whereas OxyR is reversibly activated by the formation of an intra-molecular disulfide bond that results from altered redox state. It was observed that for H_2O_2 induced stress, the response depends upon the magnitude of stimulus along with the current growth phase of the cells [8]. Cells in stationary phase showed negligible stress in comparison to logarithmic phase cells. Another widespread phenomenon is the adaptive response of prokaryote cells (also observed in yeast, mammals and plants) to lower doses of oxidants. *E.coli* is known to adapt to lower doses of H_2O_2 and after some duration becomes resistant to higher concentrations of H_2O_2. The main benefit is to protect cells from a subsequent lethal dose [8]. At the cellular level oxidative modification of amino acids leads to altered protein structure, also represented as "disulfide stress" in *E.coli* cells. The modifications are either reversible or irreversible in nature depending upon the mechanism of reaction exerted and typically lead to carbonyl formation, change in acidity, protein activity, thermal stability, fragmentation, formation of protein-protein cross links and increased susceptibility to proteolysis [8]. Ionizing radiation and metal catalyzed reaction have an irreversible effect with the superoxide ion, known to inactivate enzymes with an Fe-S cluster including aconitase [8]. Inhibition occurs by the release of free iron from the enzyme which adds up to the oxygen stress already present. Proteins modified as a result of oxidative damage displayed increased hydrophobicity which were later taggged for proteolytic degradation [8]. Although several mechanisms for protein modification have been suggested, most of the studies took place under *in vitro* conditions [8]. Proteins involved in cell processes such as glucose metabolism, protein synthesis, chaperone activity and ATPase activity were identified as main targets of menadione and H_2O_2 induced stress modifications.

Arabidopsis thaliana:

In comparison to yeast and prokaryotes, plants are more disposed to oxidative damage because of their constant exposure to environmental

stresses. These stressors typically categorized as biotic or abiotic can produce an imbalance in ROS production and its scavenging activities that are elegantly maintained under steady state physiological conditions. In fact, at times it is difficult to differentiate between stress and physiological response, as plants are continuously trying to adapt to environmental conditions. It is evident that plants under biotic invasion produce ROS as their defense against pathogens[2]. Under abiotic stress, like extreme temperature, drought, high light and wound/mechanical damage, plants produce reactive oxygen species that act as signaling molecules to induce gene expression characteristic of the response. Hydrogen peroxide generated in response to wound development is known to induce specific defense genes that allow plants to survive any mechanical damage [2].

ROS production as a result of photosynthetic or mitochondrial electron transport chain has been extensively studied in *Arabidopsis* at its proteome level [46]. Substantial amounts of superoxide and H_2O_2 as intermediates are produced as a result of photosynthesis and photorespiration [46]. In chloroplasts, superoxide radical (produced by photosynthetic electron transport chain) is initially reduced to hydrogen peroxide by a Cu/Zn superoxide dismutase (SOD) and finally to water by other reductive enzymes [46]. H_2O_2 is a strong inhibitor of photosynthetic CO_2 assimilation, but since catalases are absent from the chloroplasts, rapid destruction of H_2O_2 takes place by the action of ascorbate peroxidases (APX) [46]. APXs are present in both the stromal and thylakoid part of the chloroplast. Besides APX, another enzyme associated with stress is the glutathione peroxidase (GPX) which is also seen in the chloroplasts [46]. Electrons donated by NADPH are supplied via glutathione, ascorbate or thioredoxin redox regulators [46]. ROS detoxification is achieved by some non-enzymatic mechanisms that involve tocopherol, flavonoids, alkaloids and carotenoids in plants [2]. Increased xanthophyll, in response to over-expression of β-carotene hydroxylase under elevated light influence was evident in one such study of adaptive tolerance to oxidative stress [2].

Since ROS can also act as signaling molecules, two possibilities for their roles are i) Act as a second messenger and modulate target molecule activity that could be involved downstream in signaling or transcription activity

2) Propagates its cytotoxicity to polyunsaturated fatty acids, the oxygenation products of which regulate specific gene expression. Components of the MAPK pathway have been identified to play a vital role in various plant cell signal transduction pathways, such as cell division and differentiation, cell wall biosynthesis, hormone signaling (auxin, absicissic acid (ABA)) and ethylene – responsive processes [44]. MAPK cascade initiation is also implicated in signaling defense responses of plants that are under abiotic and biotic stresses. Looking at the different kinds of signals transmitted by this single pathway, a complex side of the MAPK cascade and its ability to cross-talk with other different plant networks is revealed [44].

Different plant species have evolved diverse strategies to acclimate to stress response. For instance, H_2O_2 can induce various processes such as the hypersensitive response (discussed earlier), systemic acquired resistance, and low temperature acclimation. However, glutathione generated an opposite regulatory effect on expression of genes encoding enzymatic ROI scavenging systems [7]. The interaction between signals mediated by Ca^{2+}, reactive oxygen intermediates (ROI), NO, ethylene, phytohormones and modified proteins during stress conditions still need to be elucidated [7]. Similarly ROS form the oxidative burst have been identified as a key component of plant defense against pathogen, however, little is known about their site of formation or compartmentalization, mechanism of interacting with biomolecules and their immediate reaction products. Therefore, it is important to elucidate these processes as it will not only expand our knowledge on basic plant-pathogen interactions but also allow for development of plant protection management [7]

A5. Towards Integrated Understanding of Stress Response

While organisms have developed adaptive responses to reactive oxygen species and oxidative stress induced by them, reactive species have been proposed to function as messenger molecules to signal cellular processes such as necrosis, apoptosis and cell death [3]. And while the transcriptional network of stress mechanism in eukaryotes is still unraveling, information derived from prokaryotic systems have suggested innovative

approaches to understanding evolving antioxidant defenses. With the discovery of ROS, particularly H_2O_2, as signaling molecules that can interact with transcription factors to regulate gene expression, several expression profiling studies have followed that broaden our knowledge on oxidative–stress response. Through the use of microarrays, it was evident that many more genes could respond to ROS than known earlier and that in addition to a global set of genes, some genes were stressor specific [3]. H_2O_2 and $O_2^{\cdot -}$ studies was one instance where identical expression profiles were observed, despite the fact that different oxidants were involved [3]. From yeast cells exposed to H_2O_2 came the revelation that most transcriptomes return to pre-stress levels within 2 hrs of its exposure to the peroxide [3]. In spite of identifying ROS responsive genes for each environmental signal investigated in various model organisms, a lack of complete understanding was felt as information for most proteins and metabolites affected was not available. Since proteins are ultimate effectors of a system along with metabolites, a question that still remains unanswered is: *How do reactive oxygen species drive and regulate components relative to each other in a physiological phenomenon such as stress response*? The network of interactions between oxidizing species and bimolecular components is highly complex and one possible solution towards this enormous challenge would be to derive data at all levels of biological information available and integrate it using mathematical tools to obtain a holistic view of a given system which brings us to compare benefits of two strategies currently used for deciphering complex systems: the reductionist and holistic approach.

The reductionism view approaches the problem by breaking a system into its components and understanding each component and their interactions *in vitro*. From a philosophical point of view, this approach views a complex system as a sum of its parts. This is also seen as a “bottoms up” approach where the researcher reconstructs the system based on the knowledge derived from function of its components, a chief attribute of molecular biology [47].

On the other hand, a holistic approach views the system as a whole and studies its interactions and complexities while it is intact-also referred to as *Systems Biology* [48]. The holistic approach suggests that systemic

interactions arising for each stage of integration can introduce new characteristics that can be easily missed with a reductionist approach [47]. The post-genomic era comes with a strategy that employs quantitative measurements of essential biomolecules at the transcriptome, proteome and metabolome levels and utilizes mathematical methods and modeling techniques to infer functions for unknown components and to determine interrelationships between elements [48]. The key to understanding a biological system involves perturbing the system, such that the relationship of the individual components to one another can be determined [47]. Since inferences from the human genome project and other previous studies illustrate that all living organisms grew evolutionarily from a single common ancestor, thus indicating same genetic code and molecular machinery; it would be incorrect to generalize major mechanisms of stress response across organisms, which was observed for some of the organisms *Yeast*, *E.coli* and *Arabidopsis*, discussed earlier [47].

A model, described as a "complex biological processes defined by mathematical equations", requires information on its components such as identity, function, quantity and other descriptive features [49]. According to Hood, a genome can provide two kinds of digital information: 1) information present in the genes that encode specific proteins and 2) regulatory networks that drive the behavior of genes [47]. Regulatory networks are critical for understanding development, physiology and evolution as its specification differs between organism and it can evolve rapidly with time [47]. The hierarchical nature of different data types, provides new insights into the operation of a system, which requires collecting different data types [47]. The complexity of a system could vary from something as simple as a set of protein molecules that come together to serve on a defined task or a group of cells that performs a specified function.

The "Yeast Oxidative Stress Project" will use a similar approach, where massive amounts of "omics" data collected from a system, at all three levels will be utilized towards building models that can further expand our viewpoint on oxidative stress regulatory networks. Consequently, findings from this study can help answer some of the questions concerning disease progression in higher eukaryotes, and potential drug design.

CHAPTER 2

B. PROTEOMICS AND SYSTEM BIOLOGY

With the need to obtain a fundamental, comprehensive and systematic understanding of a complex biological system, the inductive and hypothesis driven studies are currently viewed as complementary to each other, instead of being handled as unrelated approaches [48]. While hypothesis-driven studies have been emphasized for quite sometime, inductive programs also referred to as ”Data driven“ approach, are rapidly emerging in various investigative studies [49]. Genomics era, focused more on the sequence of the gene and led to a map of the entire genome. The post genomic era, however, focuses upon a holistic view of a complex biological system, by integrating data collected at all levels: transciptome , metabolome and proteome, to derive complete understanding on a given process [48]. With advances in techniques such as serial analysis of gene expression (SAGE) and DNA microarray technology, the high throughput production of mRNA profiles has increasingly become popular. However, since mRNA analysis may not directly reflect the protein content of a cell, an interest in the ultimate sequence of events down to the proteomics level, is widely rising . This section discusses the current protein profiling techniques and the impact of systems biology as a plausible solution for understanding an organism, as a whole. Since proteomic studies have become an important component in describing general properties of a system, different steps involved to facilitate biological modeling of proteomics data is also discussed in this chapter.

B1. Proteomics

Proteomic studies rely on technological advances to derive information on protein changes that may have occurred at the level of a cell, tissue or a complete organism. These changes can vary from post-translational modifications to quantitative changes in expression levels or formation of protein complexes . Extensively used for monitoring both physiological phenomena and pathological conditions, proteomics mainly focuses upon identifying, characterizing and quantifying proteins and their relative expression levels in a given sample. The genome sequencing project has

added to the massive volumes of protein data which assists with the acquisition of protein sequence information at the proteome analysis level.

Post-translational modifications induced in response to oxidative stress have been shown to affect several defense proteins within a cell [50]. Systemic increase in low pI (acidic) peroxiredoxins observed for primary culture of human Leydig cells, challenged with tumor necrosis factor α is one such example that would not have been evident from gene function alone [50] .It is evident that transcript levels cannot provide comprehensive information on final gene products, as regulation can occur at transcriptional, translational or post-translational levels. Lastly, proteins have a dynamic nature of expression in response to their environment, which makes it more suited for describing dynamic complexity of a biological process. The importance of proteomics is thus evident, as it focuses upon the main effectors of a cell.

Several areas of investigation including cancer research have greatly benefited from identification and application of the proteomic approach towards identifying various tumor markers as they are critical in the early stage of cancer diagnosis. Likewise, bladder carcinoma, breast cancer, renal cell carcinoma and lung tumors are some of the various oncology-related sub-areas that have been extensively investigated at the proteome level [51]. Similarly, proteomic approaches have been applied towards investigation of Alzheimer's disease, schizophrenia (neuropathology), cardiac hypertrophy and cardiomyopathy (cardiovascular diseases) progression [51]. Most of these studies have employed gel-based techniques, and demonstrated this technique as an exciting tool to examine pathological processes.

B1.1 Types of Proteomics

Proteomics studies have been divided into three categories namely; expression, structural and functional proteomics based upon their underlying approach. Expression proteomics deals with quantitative comparison of protein expression data for samples that differ by an experimental condition [52]. This approach utilizes the entire proteome to reveal information on proteins associated with a diseased state or any other biological phenomenon being investigated. Proteins are profiled on expression level changes or any

modifications that may have occurred between two groups of samples [52]. Structural proteomics aims at mapping out structure of a protein complex or any specific protein of interest isolated from a specific cell location. Identification of proteins within a complex, characterization of protein-protein interactions and location of proteins are some of the main goals of structural proteomics [52]. Functional proteomics studies characterize a selective group of proteins and assign a function derived from protein signaling and/ or protein drug interactions [52]. Current project is a protein expression study that aims at: 1) profiling proteins that are differentially expressed between treatments and time and 2) identify proteins and their possible post-translational modifications that may have occurred due to treatment effect.

For determining quantitative parameters such as expression-level, or qualitative features such as localization and protein modifications, one may use a forward or reverse proteomics approach. The forward approach relies heavily on sample preparation technologies and mass spectrometry analysis for protein identification and function [53]. It starts at the sample level, with the separation method being the most critical step of its workflow. Extracted proteins are separated before being subjected to mass spectrometry analysis and bioinformatics-mediated protein annotation [53]. The reverse approach on the other hand is directed more towards protein function and protein localization that is initially dependent upon a biological system and starts at the genome sequence level. The transcriptome and proteome are predicted *in silico*, based on which cDNA clones and protein expression systems are generated with specific functional analysis being carried out. The expression systems allow antibody production, structural and functional studies, and include bacterial to mammalian models. The most commonly used expression system includes bacteria, whereas mammalian systems if used can offer the advantage of correct folding and post-translational modifications [53]. Functional assays that reveal gene product function, include techniques such as protein-protein interaction, deriving enzymatic activities, gene silencing or integration of all these methods to unravel protein function [53].

Even though forward and reverse proteomics approaches are effective individually, at some level, their integration can lead to an advanced understanding of protein characterization. Nevertheless, the integration is not

simple, as parameters from these two experimental pipelines are still undefined and a common platform for converging different levels of data is undecided [53]

B1.2 Current Proteome Profiling Technologies

Since protein technologies are an integral part of any proteomics study undertaken, techniques are constantly being revised and improved to overcome the existing limitations. In recent years, several protein profiling technologies have been developed that allow identification, characterization and quantitative comparison of hundreds of proteins present in a cell, tissue or body fluids [52]. These advances have primarily resulted from an initial development and continuous use of instrumentation and software and data analysis tools such as chromatographic and electrophoretic separation, mass spectrometry and bioinformatics applications. Typically a proteomics experiment would include the following steps: i) isolation and separation of proteins in a given sample ii) acquisition of protein fragmentation data such as mass spectra and iii) database searches against protein sequence data to reveal protein identity, each being improved and developed further to accommodate large-scale demands [52]. A common approach used towards understanding the role of individual proteins in a cellular process is to correlate protein expression levels to biological changes occurring externally or internally within a system. External modifiers, such as environmental factors, or internal stimulus, such as mutation, cell differentiation or diseased state, can induce complex cellular processes that are reflected by changes in the proteome.

Discussed below are some of the most advanced and effective protein profiling technologies currently used in the area of expression proteomics. And while protein profiling can allow monitoring of protein levels in response to its modifiers, some technological limitations still exist that contribute to its slow rate of progress [52]. Two factors that contribute to the enormity of studies investigating proteins are the broad dynamic range of protein concentrations found in the sample and the need to capture the ethereal dynamics of any proteome (in comparison to its genome) [52]. Even though a

lot of progress has been made on improving and refining standard techniques, we are still far from a point where a single technique could fully characterize a cell proteome. Currently available protein profiling pipeline can be divided into three steps: I) generation and extraction/enrichment of proteins (sample preparation before separation), ii) generation of peptides from intact proteins (separation and resolution) followed by iii) mass spectrometry analysis and identification (section B1.5).

Due to great diversity observed for different protein sample types, an appropriate sample preparation is essential for any technique (applied downstream) to produce reliable data. The optimal procedure depends a lot upon the type of sample and the physicochemical properties of the proteins being extracted and/or solubilized. In the absence of a universal sample preparation procedure, diverse proteins from different cell locations can be isolated by a combination of mechanisms such as lysis, centrifugation, fractionation, filtration and precipitation [54]. Although each sample preparation step can enhance the quality of the final result, additional steps can result in selective loss of protein entities as well. Cell lysis is carried out in appropriate solubilization buffer to extract proteins of interest with mechanical and chemical disruption methods. Gentle lysis methods include osmotic, freeze thaw and detergent based lysis whereas vigorous methods include sonication, grinding, mechanical homogenization and glass bead disruption. Sample fractionation reduces sample complexity and enriches a specific subset of proteins before separation and analysis [54]. This approach can be used towards isolating specific protein complexes and/or other proteins from particular subcellular compartments such as nuclei and mitochondria [52]. Subcellular fractionation can be achieved by cell/tissue homogenization and or density gradient centrifugation or by reducing abundant proteins that may mask the existence of low abundant proteins, critical for signaling purposes. Protein precipitation is optional as it is employed to remove contaminating species such as nucleic acids, lipids, salts and concentration of protein sample. As a substitute, clean up kits are used to remove any insoluble components before protein separation.

Typically protein separation is achieved by gel or a non-gel based method, with 2-dimensional electrophoresis (2DE) being most popular gel

based approach. It is important to realize that even though each technique is efficient in its own way, all techniques have their own set of drawbacks. For 2D gels the limitation exists at the solubilization scheme where an optimal composition of buffer/solution drives the extraction of all proteins reflecting an entire proteome. One of its major strengths however, is the ability to resolve proteins that have undergone post-translational modifications [52]. In addition, 2-dimensional electrophoresis can present proteins that arise from mRNA splicing or proteolytic processing [52]. The non-gel based separation techniques include chromatography tools such as strong cation exchange (SCX), reverse phase (RP)-HPLC and ESI-MS that have the ability to separate intact proteins and can thus provide better coverage [53]. For quantitative assessment, proteins can be quantitated with stable isotope labeling of amino acids (SILAC), prior to trypsin digestion or wth isotope coded affinity tags (ICAT), which is done after protein isolation using heavy deuterium for cysteine alkylation [55]. Isobaric tag for absolute quantitation (iTRAQ), is another technique for peptide quantification that can later provide a protein amount estimate. It is observed that all these techniques employ peptide fractionation by SCX chromatography or any other resolution by means of reverse-phase (RP) and/or liquid chromatography (LC) before being analyzed by mass spectrometry [53]

Described below are some of the most commonly used techniques employed in protein profiling studies:

i) SELDI/MALD-TOF: Surface enhanced laser desorption ionization/ matrix assisted laser desorption ionization method is a high throughput technique for differentiating control and experimental samples with a relatively high number of peptide identifications [56]. SELDI uses selective adsorption of peptides on chromatographic surfaces based on their biophysical properties. Depending upon the surface chemistry, proteins with specific physico-chemical properties can separate on the matrix prior the MS spectra acquisition [56]. MALDI utilizes an organic compound that acts as an energy absorber and facilitates desorption and ionization of

protein digests (peptides) before being detected by MS or tandem MS [56]. The peptides are ionized by simple protonation assisted by the photoexcited matrix and these analyte ions are then accelerated by an electrostatic field and separated in the flight tube based on their mass to charge ratio [56].

ii) 2D Gel Electrophoresis: Two-dimensional gel electrophoresis (2DGE), one of the most widely used techniques in proteomics, involves an in-gel digestion step of protein spots (after spot detection) which generate peptides at specific cut off points that are subsequently put through mass-spectrometry analysis. The proteins are first separated by their isolelectric point (pI), followed by separation based on molecular weight (MW). This two-dimensional separation resolves proteins as spots, each spot being a protein isoform specified by its pI and MW, also referred to as X and Y coordinates on a gel [52]. Within an experiment the normal and treated cellular state can be compared to find proteins relevant to a certain state. Two-dimensional electrophoresis provides a better resolution of complex protein mixtures, however, in comparison to other proteomic platforms available, this technique faces difficulty in accommodating hydrophobic proteins or ones with extreme acidic or basic isolelectric points [52]. For example, membrane proteins are difficult to solubilize due to their physiochemical properties which makes representation of an entire proteome, difficult to achieve [55]. Likewise, the most abundance proteins are easily detected in comparison to lower abundant proteins, which introduces a distinct bias in terms of global proteome studies [52]. An alternative approach would be sequential extraction of proteins that are fractionated based on common biophysical properties which improves the loading amount for a category of protein subset. Generating protein fractions that are enriched with certain subset of proteins, can broaden the range and kind of proteins covered and reduce any complexity issue in total protein mix. Another limitation that exists for 2D technique is at the protein detection and quantitation level [56]. Linear broad range of protein detection on

2D gels has greatly improved with the use of fluorescent stains such as Sypro Ruby which perform better than standard stains like Coomassie Brilliant Blue and Silver Stain [56]. In view of the current limitations listed, many gel-free techniques have become available that are less biased and more global by nature.

iii) DIGE: Differential in gel electrophoresis is an improved technique of gel electrophoresis that uses three mass- and charge matched fluorescent dyes that are spectrally resolved [57]. Due to the availability of these dyes, three samples can be run in parallel on the same gel that overcomes any qualitative and quantitative reproducibility issues commonly observed with conventional 2D gel electrophoresis [56]. And while both control and experimental samples are run on the same gel, images are scanned separately which are perfectly aligned without any concerns for gel distortion issues and/or gel warping requirements. Although 2D gel provides a comprehensive profile of a proteome at a given time, the image analysis part of the workflow is a major bottleneck [56]. For single stain experiments, fluorescence dye can detect protein amounts as low as 1 ng, which is good for low copy number proteins. With the recent introduction of saturated Cy (cyanine) dye, the sensitivity of DIGE has become comparable to silver and Sypro Ruby stained gels [56]. The Cy dye labeling has a dynamic range of 10^4 – 10^5 with detection of low abundant proteins [57]. One substitute for 2D gels is the one dimensional separation (1D gels) which can be useful but would need tandem MS in conjunction with one dimensional reverse phase (RP) LC for identification purposes [57].

iv) PF2D: Is an alternative to the classical proteome approach where protein fractions are collected based on isoelectric point in the first dimension and by hydrophobicity in the second dimension [58]. In simpler terms it is a 2D LC based protein fractionating technique where fractions can be later analyzed with mass spectrometry for protein composition [58]. Samples separated on a column are put through second dimension reverse phase HPLC [58]. PF2D, however, has some limitations, including the large volume of

sample needed and short lifetime of a separation column [58]. Even though PF2D is not as popular as other protein separation schemes, one example of its accepted application was for the characterization of immunogens of nonpathogenic bacteria *Baciliius subtilis* using PF2D in conjunction with polyclonal antibody and tandem mass spectrometry [58]. The approach also referred to as "i-PF2D-MS/MS" for the first time integrates analytical 2D LC (PF2D) to immuno-blotting and mass spectrometry [58].

v) MudPIT: Multidimensional protein identification technology is one additional protein profiling technique that can directly identify protein and/or peptides from a complex protein digest. Tandemly coupled liquid chromatography columns are used, with strong cation exchange and reverse phase HPLC separation, followed by tandem MS analysis [57]. Depending upon the sample complexity and amount of protein for use, peptides eluted from cation exchange column, are separated on a reversed phase HPLC columns repeatedly before being analyzed by mass spectrometry. This technique has been applied on the yeast proteome, where ~1400 proteins with a demonstrated dynamic range of detection of ~10,000 were observed [57].

vi) ICAT: Isotope coded affinity tags is a commonly used quantitative technique that allows protein quantification by using light or heavy isotopes that bind to alkylation sites of cysteine residues and are identifiable by micro-capillary LC/ESI/MS/MS [56]. Chemical incorporation of isotope tags is done after protein extraction. The control and experimental samples are derivatized with light and heavy ICAT reagent followed by trypsin digestion [56]. Labeled peptides are fractionated using strong-cation exchange liquid chromatography followed by RP-HPLC and tandem mass spectrometry analysis to identify ICAT peptide pairs [57]. ICAT is analogous to microarray use of two different dyes or DIGE protein expression analysis, as changes in expression are determined by differences in intensity observed. Labeling approaches provide relative expression ratios of samples being investigated under two

experimental conditions [56]. Labeling is dependent upon the presence of cysteine residues as its sulfydryl groups are chemically labeled in proteins. The limitation arises at the mass spectra and database search levels, as proteins that lack cysteine cannot be included in the analysis [56]. ICAT technology has been reportedly applied to several proteomic studies including total proteome characterization of yeast and P*seudomonas aeruginosa* [56].

vii) SILAC: Stable isotope labeling by amino acids in cell culture Is a metabolic labeling approach where isotopes are added to the culture media during cell growth [56]. This approach was first described in yeast and since then has been applied for studying protein-protein interactions and differentially expressed proteins in EGF (epidermal growth factor) pathway and prostrate cancer studies [56]. This approach makes tandem MS interpretation much easier as labeled and unlabelled peptides mass differences are easily predictable [56].

viii) Emerging technologies: In order to simplify protein quantification downstream, stable isotope techniques such as iTRAQ have recently been developed, that include labeled amine modifying chemistry with MS/MS based quantification mode. With the advances in microfabrication technique and current bottlenecks perceived in proteome profiling schemes, several chip based systems of large scale protein analysis have been developed including protein chip technology, microfluidic arrays, antibody arrays and tissue arrays [55]. However, whatever limitations exist with these technologies is only due to lack of standard methods that can generate large amounts of proteins. Since proteins are unstable and chemically heterogenous, they are subject to denaturation, which makes maintenance of protein conformity on synthetic surfaces more critical, so as to allow probe activity [56]. Antibody microarrays include immobilized antibodies that are incubated with a cell extract. Covalent attachment of antibodies to protein antigens present in the sample can be detected by various by fluorescent dyes and has been used as a means to study receptor tyrosine

kinases in human tumor cell lines [57]. With the dramatic advances observed for microfabrication and microfluidic application, several protein profiling strategies such as microfluidics-based isoelectric focusing system, microdialysis of small volumes of proteins have emerged that have the potential for delivering improvements needed for this area [55]. Even though a fully integrated proteomic system has not been developed, attempts at achieving this is clearly progressing [55].It is evident that a single technology cannot address all different aspects constituting a proteomics research; however, the improvements made so far, have definitely broadened the utility of proteomics approach for “biological understanding”.

It seems that in order to increase the number of proteins identified and being quantitated, one may need to use multiple complementary protein profiling approaches that would not only result in massive amounts of data but would require bioinformatics schemes to assimilate biological information

2-D Gel Image Analysis Software: Following separation of proteins on a two dimensional platform, the next step in 2DE proteomics workflow is the spot detection and quantization procedure, mainly achieved by image analysis software. Several software packages are available that can analyze stained images of resolved protein spots on 2D gels and allow matching between different images including PDQuest (BioRad), Progenesis (Nonlinear Dynamics), Melanie (Swiss Institute of Bioinformatics) andZ3 (Compugen) [59]. The packages offer various levels of automation with some degree of manual matching and parameter setting. However, there are still some problems associated with the spot recognition, matching and normalization which need to evaluated before choosing a single image analysis software [59]. Typically, automatic alignment of 2D gel images starts off with image warping, which can remove any variations that may have occurred due to running conditions [59]. Sometimes, warping alone is not sufficient which reveals mismatching of spots that can later be removed with manual matching. Some image packages use x, y coordinates for matching purposes, while others can utilize size, shape and intensity of a spot. In an advanced

version, spots that are low intensity (below detection threshold) or close to the level of noise are filtered out of the analysis. A computer assisted analysis of 2D gels includes three steps: i) protein spot detection, ii) spot quantitation and iii) matching spot patterns between gels [59]. Parameter settings with details of image analysis software used for the current project are discussed in Chapter 3. Due to the availability of two image analysis software, a comparison conducted between PDQuest and Progenesis (SameSpots) package reveal:

i) PDQuest: In this software, spot detection is carried out by selecting the most intense and faint spot first, to define a dynamic range for spot detection sensitivity. Candidates (well resolved spots) that are common between sample and reference image are marked as landmarks [59]. Preprocessing is done with a background subtraction (floating ball method), followed by adjusting the detection parameters. Unfortunately, landmarking is not propagated on all gels as compared to Progenesis, as a result of which each image has to be covered individually. Spot boundary tools allow merging of two spots with redefining facilities for spot boundary. PDQuest claims to handle multiple gels with a 3D montage available for spot contour. However, with experience it is observed that more than 50 gels can dramatically reduce the speed of image processing[59]. For details on spot detection parameters and manual matching methods, refer to section C2 for further details. PDQuest offers at least 5 different normalization methods, with the availability of few statistical tools such as Mann-Whitney test, 1-way ANOVA etc.

ii) Progenesis (SameSpot) : Progenesis from NonLinear Dynamics, is one of the most comprehensive software package with an advanced version that allows automatic detection of spots for a large group of gels. The software can at a given time handle more than 50 gels with ease and can generate comparisons between different experiments, provided each experiment (set of gels compared) shares the same reference gel. The software cannot generate a cumulative virtual gel of all gels compared, which is typically how PDQuest operates when comparing two different experiments [57] Automatic detection of protein spots is achieved by determining background levels. However,

filtering is performed automatically without any requirement of parameter entry. Some of the advantages that exist in comparison to PDQuest include: mass deletion of spots based on chosen area of interest, a three dimensional montage useful for viewing total area of spot over multiple gels and the ability to split a spot, the latter being absent from PDQuest software. The normalization method available for Progenesis is very similar to median normalization, though a detailed mechanism of normalization was not available in the manual. SameSpot software offers PCA clustering and 1-way ANOVA as statistical tools.

Thus, after completion of spot detection and accurate spot matching across gels, proteins (as gel spots) of interest are excised either manually or by means of robotic assistance and digested into peptides for mass spectrometry analysis. Since mass spectrometry is a sensitive technique, all protein samples (digested or undigested depending upon the nature of protein content) are cleaned before effective sample ionization can be achieved. Substances such as dyes and buffers can interfere with the identification process and can remarkably reduce instrument sensitivity for low abundance proteins present in a complex mixture [59]. Contaminants such as detergents, lipids, nucleic acids present in low quantities need to be removed, without eliminating biologically significant proteins. Moreover, micropipette tips like ZiptTps can be used for purifying and concentrating protein samples to a small volume [54].

B1.3 2D or not 2D

Modulation in protein expression levels represents the fine tuning of a cell in relation to its environment. The genome of a cell is essentially static, whereas the proteome is a dynamic entity which makes its analysis more pressing [59]. Even though significant advances have been made in the area of 2D technologies and has been used for generating protein expression data for this project; it is important to consider the effect of 2DE limitations in context of results obtained. Listed below are a few existing drawbacks: i) Individual 2D gel patterns differ between labs and therefore reliable deductions are possible only after a protein spot has been identified in both labs 2) Multiple gels with narrow pH range can increase the resolution of protein separation but a the same time increase the amount of work (in terms of image analysis. 3) Detection of low abundant proteins including receptor molecules, signal transduction or regulatory proteins on 2D gels is difficult to achieve with a single stain experiment. And while, membrane proteins, critical for signal transduction, cell adhesion, metabolite and ion transport, are not well resolved on 2D gels, this technique offers an advantage of resolving hundreds of proteins at a given time. A global level of profiling without intensive data analysis at the MS level is definitely achievable with 2D gel, which supports our choice of profiling technique.

B1.4 Novel Approaches in Proteomics

Protein Microarrays: As mentioned earlier, current bottlenecks of 2DE platform has led to the development of new framework of protein characterization. methods Protein microarrays is one such emerging technique, that can be used for functional or expression profiling [55]. Each microarray is individually hybridized with a complex labeled sample and a protein expression profile , defined by the library of capture reagents immobilized on the surface of an array is generated [60]. The microarrays consist of a library of peptides, proteins or any target that is spotted on a solid support. Protein samples used for analysis are labeled with a fluorescent tag binding to the individual targets for quantification and measurement [60]. In addition to flourscence, chromogen , chemiluminescent and radioisotopic labeling has been used as a detection method for protein microarrays [60] Even though protein array

are flexible and have a great potential to complement other prevalent proteomic technologies, its utilization and development have been limited due to technical challenges [60]. Some areas that need further improvement include: a) Developing a wider variety of affinity reagents (besides monoclonal antibodies, recombinant proteins) , b) Improved array surface chemistry to facilitate immobilization and capture of affinity reagents and c) need for self-assembling protein array platform [60] .

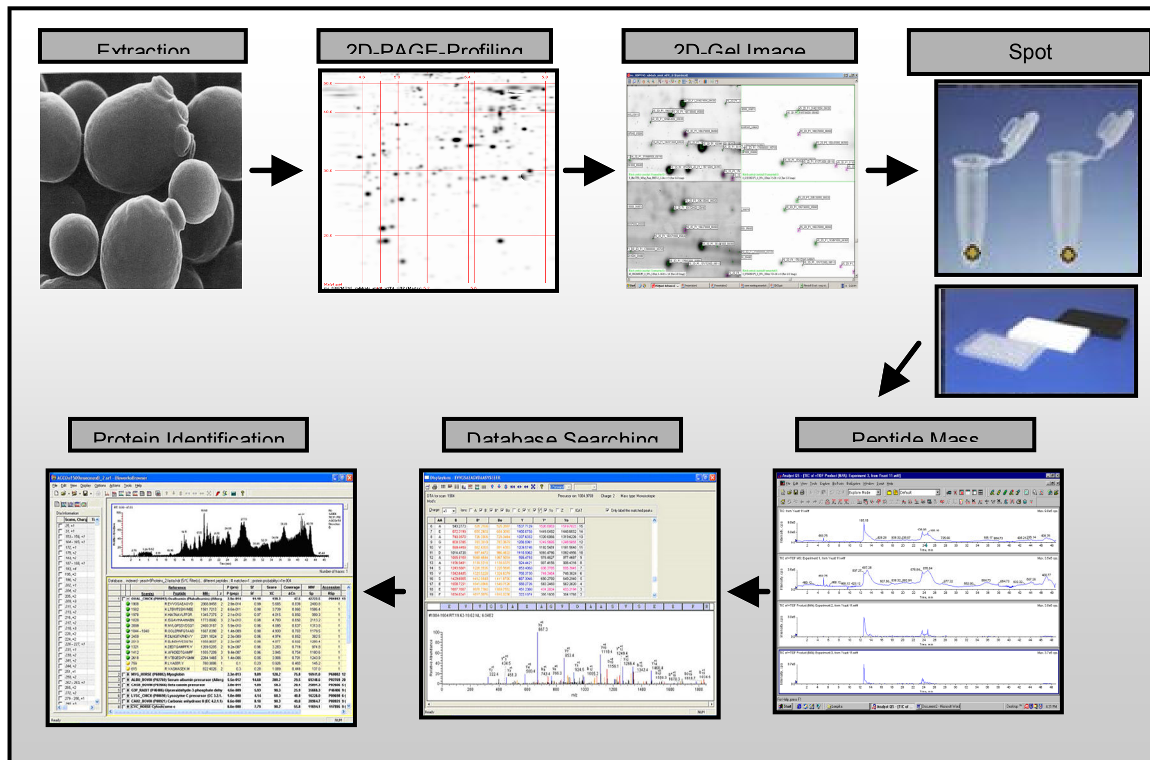


Figure 5: Proteomics workflow of 2D gel based approach experiment

Microfluidics: Microfluidc devices have advanced rapidly with increased application in proteomics with the aim of analyzing small volumes of proteins [60]. In particular, the advantage is miniaturization of this analytical tool for proteome sample treatments. Recent advances have been made towards implementing microfluiidcs for protein sample treatment, cell manipulation, sample cleanup and fractionation as well as on chip proteolytic digestion [59]. Some groups are developing interfaces for microfabricated microfluidic devices, to nanoelectrospray mass spectrometry, for

high throughput delivery. [55] While ESI-MS emitters are effective for an infusion analysis, the tips can contribute to peak broadening not sufficient for microfluidics based separation. Similarly for MALDI purposes, crystallized peptides have been presented along the edge of a disc for MS analysis. In such cases, the device has an increased surface-to-volume ratio which allows protein digestion by means of proteolytic enzymes immobilized on the chip surface [60]. Microfluidics has been successfully applied for protein/peptide separation using chromatographic and/or electrokinetic-based principles [60]. Separation of complex samples such as yeast cell lysate has been demonstrated using multi-dimensional system[60]. On one side micrifluidics stages an advantage as it is capable of dealing with minute sample amounts but it would be unrealistic to expect that certain problems encountered for un-miniaturized setups will be resolved with microfluidics [60]. One such example is the co-migration or co-elution of proteins occurring for gels or LC separation. Since this pretty much depends upon the intrinsic nature of the proteins and not the analytical tool itself, such problems would still exist in miniaturized technology.

Single cell Proteomics; is another approach applicable for proteomics research, which aims at defining protein function depending upon its spatial and temporal location [60] Single cell proteomics is valuable for both basic and clinical studies, for eg in oncology, neuroscience and cell development studies, as it provides a complete description of a cell proteome in spite of the fact that a cellular proteome. Is in constant flux [60]. Cell proteomics is relevant as distribution of protein expression is very heterogenous which also makes it more challenging in comparison to other approaches. Mass spectrometry can definitely help in protein analysis but besides mass spectrometry, some separation methods that have been used include capillary electrophoresis and capillary sieving electrophoresis [60]. These methods have been applied for cell injection, lysis and protein content analysis. The multiple CE system has been successful in analyzing a number of samples and cell types including separation of proteins form MC3T3 osteo-precursor cells. What is important at the mass spectrometry level is the use of high accuracy and resolution instruments that can perform intact protein analysis [60]. Following

section is a description of the various types of MS platforms currently available and used in this area, with a focus upon its utilization and mode of operation [60].

Bioinformatics Tools for Proteomics: In view of the massive volume of MS data, currently generated in a typical proteomic analysis run, bioinformatic tools are needed to expedite identification, quantitation and mapping protein modification part of the scheme [60]. Typically bioinformatic tools include algorithms and softwares that allow comparisons of MS data between them and *in silico* data stored in a database, identification of protein and its modifications from MS data and evaluation and organization of results on protein identifications. If such tools were not available these operations would be conducted manually. However, with high throughput production of mass spectrometry data, such manual analysis is not feasible Several programs that provide protein identification using peptide mass fingerprinting (PMF) include Mascot, Profound, ProteinLynx, MS-Fit etc [60]. One thing common to all these programs is the input list of peptide peaks; however each program differs in their scoring mechanism. While some rank a protein hit based on the number of potential matches between experimental spectrum and database spectra, other use a probability based scoring that distinguishes correct matches from random occurrences [60]. Some software tools exist for interpretation of MS/MS data as well. SEQUEST (BioWorks) was one of the earlier one developed in 1994, that uses a cross-correlation scoring for comparing actual spectra to the database MS/MS spectra. However, all identifications are not reliable as poor quality spectra and unaccounted modifications can confer incorrect identifications [60]

B1.5 Mass Spectrometry

Mass Spectrometry (MS) is a widely used technique that has been combined with 2D-GE for direct protein identification and/or protein/peptide sequence analysis. A mass spectrometer consists of three main components: 1) Ion source: that analyzes the nonvolatile molecules such as proteins, peptides, nucleic acids etc. 2) Ion analyzer: that separates ionized bio-analytes based on mass-to-charge ratio and 3) Detector: that counts the number of ions characteristic of a specific *m/z* value [61]. Two important ionization methods that have totally revolutionized the use of mass

spectrometry in protein sequencing and identification include: matrix-assisted laser desorption /ionization (MALDI) and electrospray ionization (ESI) [60]. Both these soft-ionization techniques allow formation of intact gas-phase ions needed before molecular masses of biomolecules can be measured by the mass analyzer [62]. Several ion analyzers currently available in different mass spectrometry instruments include: quadropole ion analyzer (Q), time-of-flight (TOF), quadropole ion-trap analyzer (QIT), magnetic sector-type ion analyzer and Fourier transform ion cyclotron resonance analyzer (FTICR). Theoretically a mass spectrometer can have any combination of ion source and analyzer, however, the most common combination include ESI-tandem-Q, ESI-Q-TOF or QIT, MALDI-TOF and MALDI-QIT-TOF [63]. The experimental basis for protein identification depends upon the ability of a mass spectrometer (MS) to measure m/z value of peptides with high sensitivity and accuracy. With the exponential growth of genome sequence databases, considerable improvement has been observed for the two mass spectrometric methods and protein discovery in translational research [62].

MALDI: As the name suggests MALDI relies on laser desorption ionization (LDI) technique for ionizing samples of interest. The only improvement from its previous version LDI, is that instead of a soluble analyte being air-dried on a metal surface, the sample is now assisted by a matrix compound that absorbs and transfers energy from the laser [62]. A variety of matrices including aromatic acids can be used towards this objective. The aromatic group absorbs energy at the laser light wavelength which results in proton transfer to the analyte [62]. The MALDI technique is fast and good for ionizing peptides and proteins, but the quality of the MALDI-MS spectra pretty much depends on the matrix preparation. The sample co- crystallizes with an excess matrix solution which has led to several methods on sample-matrix preparation. All methods aim towards a homogenous layer of small analyte crystals as non-uniform sample/matrix crystals give low resolution with low correlation between analyte concentration and its intensity [62]. In general, MALDI ion source interfaced with TOF mass analyzer, measures tryptic digest of a target protein, obtaining 30-40% protein sequence coverage. In contrast to ESI, MALDI produces mainly singly charged peptide ions which make mass spectra interpretation very

straightforward. Upon obtaining peptide mass fingerprint (PMF), the peptides of interest can be further fragmented which will derive the protein sequence and its coverage. The mass numbers derived for the fragment, after tandem MS (MS/MS), can be searched against the database for spectra matching which will impart identification with a probability score. Since, Arg- containing peptides have higher affinity for protons than its Lysine containing counterparts, one can see these former peptides more frequently in MALDI-MS spectra [62]. The main advantage with MALDI-TOF MS is its high sensitivity and its extensive use with 2D electrophoresis. However, one of the major issues with protein identification and this technique is the difficulty in obtaining high significant search results. This is attributed to several factors such as : i) Some peptides are not able to co-crystallize with the matrix ii) All expected tryptic peptides do not show up on spectra due to their respective ionization affinity iii) Homogenous sample-matrix crystals need to be searched as a good sample spot within a target [62].

ESI: another ionization method that has found its relevance in proteomic studies, forms ions at atmospheric pressure followed by droplet evaporation. A solution of peptides or proteins is passed through a fine needle at high potential to generate analyte ions [60]. The electrical potential produces charged droplets which shrink by evaporation resulting in charge density. The ESI ion source has a tendency to produce multiply charged peptide ions depending upon the number of groups on a polypeptide chain that are available for ionization. Tandem MS of polypeptides are most often done in positive ionization, though negative ionization can be applied towards identifying sulfated or phophorylated peptides [63]. A common setup for ESI includes reverse phase-liquid chromatography (RP-LC) coupled to ESI-MS/MS. The flow rates of the solution can be adjusted depending upon the nano or micro bore RP columns. Typically the flow rates used for LC systems are 100-300nL/min and 1-100μL/min for the nano or micro LC, respectively [62]. Gel isolated proteins, available in picomole amounts can be easily detected with a nano-LC system that has the following advantages: low flow rate, high ionization efficiency and high sensitivity detection.[62].

After peptides/proteins are converted to molecular ions, the mass analyzer of a mass spectrometer separates the ions in a vacuum based on mass and charge. Described below are the different kinds of mass analyzers:

TOF Mass analyzer: In TOF-MS, ions accelerated by an electrical potential, pass through a field free region traveling with a speed characteristic of its m/z value. Time of flight is calculated as sum of t_a + time of flight in acceleration region , t_d +flight time in drift free region and t_d = detection time [60]. An ion reflector can compensate for kinetic energy differences as its electric field returns ions at an angle opposite such that more energetic ions penetrate more deeply and receive a longer flight path [62].

Ion Trap Mass Analyzer: This mass analyzer uses a quadropole field to trap ions of low m/z or high m/z range. The trapped ions (in 3D electric field) are ejected according to their m/z and MS spectrum is obtained on detected ions that are ejected selectively at different dipolar field frequency [60]. The precursor ion of interest is stored which is fragmented to derive sequence information. The energy of the precursor ion is increased that leads to frequent collision with a inert gas for example, helium that causes the fragmentation [60]. The fragment ion ejected at different dipolar frequencies are detected by producing MS/MS spectrum [62].

Quadropole mass analyzer: Is one of the most common mass analyzers and typically share three quadropoles, where Q1 and Q2 serves as ion guide elements and the TOF reflectron separates the ions based on m/z values [60]. A quadropole comprises of 4 metal rods that can transmit all ions or filter based on certain mass-to-charge ratio [60]. Precursor ions transmitted form Q1 are accelerated to Q2 where they undergo collision induced dissociation (CID) and the fragment ions are thereof focused in Q3. Different mass analyzers can produce different MS/MS spectra. Fragmnets generated from Q TOF analyzers are most stable as they fragment carry some kinetic energy even after collision, resulting in y ions mostly. For ion trap the fragmentation of the precursor ion occurs because of its frequency, which means all *b*, *y* and *a* ions can be observed. The Q-TOF ion is easy to interpret but of low significance as it is missing other ions, in comparison to ion trap [62]

With the combination of different mass analyzers available and ionization source, a new set of hybrid analyzers have rapidly emerged, some of which are discussed here. One example is the quadropole –TOF mass spectrometer, which contains Q1, quadropole collision cell (q) and a time-of-flight analyzer [60]. Since this instrument is typically coupled to electrospray, it is not a popular choice for large scale proteomics. MALDI-Q-TOF was developed to allow peptide mass fingerprinting along with amino acid sequencing and is a combination of MALDI ion source [60]. A Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer is a high resolution and mass accurate instrument that can analyze complex mixtures. It is an ion-trapping instrument and when coupled to ESI has been applied towards studying protein interactions and conformations [60]. It is evident that even though vast improvements have been made in separation and 2D chromatography, analysis of thousand components within a complex peptide mixture still remains a challenging task. As a result a need for MS component to interact with other proteomic components such as separation, bioinformatics and enrichment protocols can definitely provide more meaningful data.

B1.6 Database Utilization

Protein identification by mass spectrometry depends upon the peptide mass fingerprint (PMF) specific to a particular protein and its pattern of masses generated by an instrument. The fingerprint is then searched against a proteome database; best matches of experimentally obtained peptide map to theoretical PMF of individual proteins within a database, leads to the identity of the unknown protein. Several factors that can affect peptide mapping results, can be grouped together as either i) Fingerprint–constructing or ii) Fingerprint-searching factors. Factors that influence fingerprint construction include i) Noise level of a peptide set ii) The number of peptides in a given fingerprint and iii) Mass accuracy based on instrument calibration. When searching the database, i) Characteristics of the organism being studied and ii) Specific post-translational modifications are two factors that need attention .

The query PMF is compared to every sequence that exists in the specified database. A match is evaluated based on various algorithms which returns a probability based score. If the fragment is a result of tryptic digest, every fragment between K and R in a protein's theoretical sequence is quantified by the weight of amino acids in that fragment (in Daltons) and if the mass of peak submitted as a query matches to this calculated mass, a hit is declared. Ultimately, each protein match is assigned a probability-based score which helps in concluding whether match was due to a random chance or true protein identification [60]. The accuracy for the mass of an unknown sample can vary anywhere from between 1 to 1000 ppm , with 100-to 400 ppm being the most typical for many laboratories. Depending upon how well is the instrument calibrated; a more stringent search will lead a searching algorithm to fail to match some observed peaks to database fragment masses resulting in no identification. The percentage length coverage of a peptide is an index of how well is a protein represented in the query PMF. Fragments outside the allowable range of 600Da and above 3000Da and below an intensity threshold are not included in the query PMF in database search. PMF can sometimes have limitations such as some peptides tend to ionize over the expense of others or signal for modified peptides can be observed that are not predicted by *in silico* digestion

which cannot be matched unless accounted for the modification [53] Features such as database size, distribution frequency of a peptide mass for a given protein, and distribution of mass accuracy are some parameters that can influence the specificity of a search Therefore the choice of a parameter helps user to get high specificity without missing a true protein positive [60]. The intensities of spectrum do not correlate with the amount of peptide in the samples due to suppression effect and ionization bias, therefore it is relevant to evaluate if the intense peaks have been used towards protein identification or not. For large proteins, large amount of peptides should be matched whereas for a small protein low number of peptide matches can result in reasonable coverage [60].

Tandem MS (MS/MS), on the other hand reveals additional information on peptide sequences. Peptide samples can be separated by one or multi dimensional liquid chromatography (LC) and subjected to tandem mass spectrometry for peptide sequencing. Database search parameters: 1) Types of ions selected for generating theoretic data can depend upon the kind of instrument used for fragmentation. Mass spectrometers such as ion trap, quadropole and Q-TOF result in *b* and *y* ions whereas instrument with high energy CID can generate a, c, x and z ions as well. 2) For the calculation of peptide mass, monoisotopic or average method can be used as mass spectrometers do no measure mass of peptides but instead mass to charge values [60]. For a given protein, the monoisotopic mass is the mass of the isotopic peak whose elemental composition is composed of the most abundant isotopes of those elements Average mass is the weighted average of all the isotopic masses abundant of that element. High resolution mass spectrometers can use monoistopic determination for mass whereas with ion traps it is better to use average mass (low resolution). 3) Peptide ion charge state can be determined in high resolution instruments by the isotopic distribution patterns observed in MS spectrum. With low resolution instruments it is not possible to tell the exact charge, though single and multiple charged ions can be easily distinguished 4) Parent ion tolerance allows certain measured peptides selected from sequence database to be scored against the experimental spectra along with the choice of enzyme used for peptide digestion. The number of candidate peptides needed for analysis is reduced with the specification of an enzyme which reduces search time significantly. Modifications

such as reduction and alkylation for gel based proteins are incorporated prior to analysis; Other modifications categorized as static or variable are incorporated as search parameters in database search.; static modifications are where all occurrences of a residue are modified whereas variable modifications are when some residues may or may not be modified [60].

For protein mixtures as complex as> 10, 000 proteins, fragment-ion matching technology is used instead of PMF. Peptides from protein digest are dissociated into fragments using mass spectrometers, the mass spectral of which is measured and searched against a database to determine the resulting precursor peptide mass. This approach is particularly useful when a peptide sequence is unique to possibly identify the protein origin based only on MS/ MS fragmentation. Researchers have found that tandem mass spectrometry has a higher success rate in protein identification than MS-based identification [57]. Another method for database searching involves the “sequence tag” approach which uses short amino acid tags generated after tandem MS interpretation against peptides in protein databases for the same enzymatic cleavage. For a protein with no previous sequence information, *de novo* interpretation is considered useful at the tandem MS level [64].

In order to search databases, many MS search engines have been developed that are tools that allow peptide identification by searching experimental mass spectra against MS data of *in silico* digested protein databases. SEQUEST (BioWorks), MASCOT and ProteinProspector are some of the algorithms used for peptide identification. Peptide matching a protein entry are clustered together and reported as a protein hit [53]. The database score is computed according to some scoring function that measures the degree of similarity between experimental spectra and the peptide pattern observed for theoretical fragmentation. SEQUEST one of the most commonly used programs calculates cross correlation score for all peptides queried. In addition to X-corr a derivative score which computes the relative difference between the best and second best X-corr is computed which is useful for discriminating between correct and incorrect identifications. MASCOT, a probability based score estimates the probability of matches occurring by chance for the number of peaks in an experimental spectra and the distribution of a predicted ions. With the SEQUSET algorithm, manual review of data is needed to avoid choosing

false positives. Since MASCOT uses probability based scoring that assigns score to all identifications, it depends entirely upon the researcher to decide which protein identification to be taken as significant. [57].

To determine the function and regulation of a protein, it is necessary to identify types and location of modifications, which occurs either co- or post translationally. Hundreds of protein modifications are known; among which phosphorylation is the most widespread and functionally important due to its involvement in signaling cascades (e.g. MAPKK). Other modifications include: glycosylated proteins that are ubiquitously present on cell surfaces and facilitate cell-cell interaction; ubiquitinated proteins that have ubiquitin moieties as tags for degrading specific proteins; modifications involved with protein regulation include disulfide bonds, acetylation, methylation, oxidation of methionines etc. Sometimes it is difficult to analyze modified peptides in presence of non-modified background, which emphasizes the need to identify the type of modification before enriching proteins with that particular modification, as an additional step in the experimental scheme.

B.2 Systems Biology

With the number of genome sequencing and exploratory studies being conducted, Systems biology is one approach that aims at obtaining a comprehensive and quantitative understanding of a living systems by integrating data collected at DNA, RNA, protein and metabolite levels and using mathematical models to describe interactions [65]. Building and analyzing models of biological system requires many modeling tools and extensive computational analysis before reliable biological understanding can be attained. It is interesting to see how modeling navigates between the need to reveal details of a system and the requirement for abstraction of data, such that a biological insight can be accomplished. Modeling tools cover a broad range of mathematical methods like differential equations, statistical correlation tools etc. however, one thing common for all these tools is the requirement of identity and quantity information on components and the dynamics of their inter relationship. Modeling processes are categorized as either bottom-up, top-

down and middle-out [65]. Bottoms up modeling start of at the levels of DNA and proteins and moves upwards characterizing higher level process whereas top-down works in an opposite way starting of at high-level function. In view of the above two approaches, middle out method may seem as a convenient solution in principle due to problems associated with the other two methods, however, a consensus on what defines a midlevel is still to be reached [65]. In the end, it is important to realize that the components of a system should not studied in isolation but instead within a system, as a whole as some components such as metabolites and proteins are not static as they are continuously synthesized and degraded, adding to the complexity of an arrangement [65]. This an integral part of Systems biology approach is the building and analysis of models that have the ability to capture biological complexity of a process at its different levels [65].

Model organisms such as yeast, mouse and fruit fly are beneficial for application of system biology approaches, as general scientific principles were developed on these organisms and studies performed are rapid and reproducible for these organisms[66]. Yeast in particular has been an excellent model for system biology studies due to the massive amounts of biological information available and it being a simple eukaryote organism. [66] With the emergence of Systems biology in the field of biological sciences, its application has varied from characterization of glucose metabolism (in yeast) to description of functioning of the human heart [66]. For this approach, qualitative date is useful in describing functions and defining pathways whereas quantitative data is needed to describe any given process. However, it has been difficult to capture absolute levels of proteins, mRNA and certain metabolites inside cells as only relative values are generated for some investigative studies [66].

Systems biology has been successfully applied to various biological systems and have been critical for providing insights into the physiological aspects of many human diseases including cancer, diabetes , asthma, heart conditions etc. [49]. With the emergence of systems biology, it is becoming increasingly apparent that systems biology and proteomics, one of its major disciplines will have a major impact on future health care by creating a predictive, preventative and personalized vision of medicine. Since genomic information does not provide any knowledge on the post

translational modification as well as on other dynamic responses, proteomic approach has become more appealing to disease state studies, over time. With the improvement in protein analysis techniques as development of several bioinformatics tools such as database and software's, use of protein expression studies for investigating diseases conditions is not only feasible but reliable too [51]. Proteomics has rapidly covered areas such as cancer, cardiovascular disease, neuropathology etc in context with human conditions [51]. 2D database of bladder carcinoma proteins have been established along with information for proteins from breasts ductal carcinomas, human kidney proteins etc. Proteomics has been used for studying tumor development and progression studies specifically for factors affecting apoptosis failure. Total brain tissue has been used to generate a 2D gel database of proteins expressed form the parietal lobe cortex which can be used to compare to brain extracts from abnormal individuals [51]. Neurodegenerative disease such as Alzheimer's and schizophrenia has been compared along with the process of demyleination in mouse model based on 2D gels. Studies of dilated cardiomyopathy with the application of proteomics has lead to establishment of human myocardial database with 150 identified proteins, with some proteins present only for patients with dilated cardiomyopathy [51].

Since proteomics is a large-scale technology that can provide such relevant information on the effector molecules in a living organism, its applicability for holistic studies cannot be denied. However, it can be used to model biological processes only when hundreds of proteins are simultaneously analyzed [49]. The reason because proteomes are complex systems and in order to work with complete datasets every single molecules should be covered, which is impossible to achieve experimentally [65] One way to resolve this would be to work with protein complexes and signaling pathways as functional modules and integrate such datasets together (Fig. 6) [65] Different steps involved in biological modeling of proteomics data can be described as: Acquisition of data on protein species that includes information on its identity, quantification, functional status, localization and dynamics of changes [65]. Following acquisition of protein information, datasets are converted into formats compatible with modeling tools. The data is usually adapted to the requirements of systems biology approach by converting it into XML-supported formats [49]. Studies

like “correlation analysis of proteomics data describing interferon signaling in liver cells” and “distribution of reaction flux in mitochondrial metabolic network” present some sort of a merger between proteomics and modeling, for integrated understanding of a biological phenomenon [67]. It would be interesting to see how protein expression data generated for the yeast oxidative stress project can undergo a similar mechanism of model building with the help of mathematical tools.

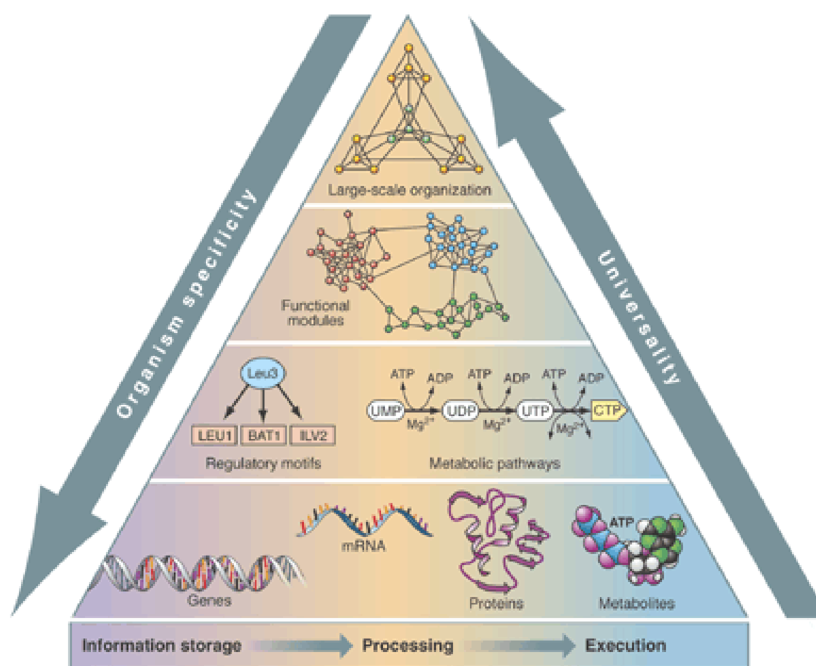


Figure 6: A simple complexity pyramid composed of DNA, RNA, proteins and metabolites. From Oltvai and Barabasi, SCIENCE 298: 763 (10/25/2002). Illustration k. Sutliff Reprinted with permission from AAAS [68]

Systems biology and its provision of a holistic view of system have already begun to heave impact on the drug discovery procedure, therapeutic methods and improvement of many industrial biotechnological processes. Identification of novel drug targets by means of model simulation is being considered as a potential application in drug discovery, as improved drug design can lead to better pharmaceutical properties to avoid side effects [66]. In addition to this, the ability to validate new chemical entities early in the process of drug discovery has led to saving capital otherwise spent on subsequent steps of drug discovery [66].

Application of system approach can have a profound effect on successful identification of pharmaceuticals as any biological problem or disease reflects the operation of a perturbed network. A comparison of diseased network to a control network can lead to identification of nodal points that can be reconfigured to change a perturbed system back to its normal state. The nodal point can be proteins as potential drug targets or other relevant targets that contribute to disease pathogenesis and /or progression [69]. It is clearly evident that with the existing approaches to drug discovery, the pharmaceutical industry has a lot to lose in terms of time and capital. Systems biology offers a novel approach to the two of the several issues currently challenging the drug discovery pipeline: by profiling a relevant target and its assessment for pharmaceutical properties as a drug, before administering for clinical trials [65] Besides biomedical research, systems biology approach has been applied towards development of metabolic engineering of cell factories for food, chemical and pharmaceutical industries [66].The industrial production of compounds is estimated to produce 10-12% chemicals by the year 2010. Use of renewable feedstock as a bulk material for chemical, fuel or bio plastic production has had an influential impact on plant biotechnology industry [66]. Systems biology approach is thought to enable informed risk assessment of new biological entities in the environment to improve waste management and treatment process.

CHAPTER 3

C. EXPERIMENTAL DESIGN, METHOD DEVELOPMENT AND DATA ANALYSIS

Two-dimensional gel electrophoresis (2-DE) in conjunction with mass spectrometry has been a core technology in proteomics. Since a cell proteome is in constant flux with respect to its environment, the area of *expression proteomics* involved with observing such changes becomes more intensive and challenging, as each technique tries its best to capture the dynamics of protein expression in between time points. This chapter discusses 2D-PAGE methodology in context of this project and evaluates an optimal method for detecting changes in protein expression. Since data generated from 2-DE is similar in characteristics to gene expression microarrray data, evaluation of different normalization methods followed by statistical analysis is also discussed. The comprehensive analysis of a large number of spots and its comparison between samples is simplified by the use of a reference gel. The reference gel acts a common platform which facilitates comparison between sample gels to reveal proteins that are involved in oxidative stress response (CHP treatment). Parameters used at each step of this proteomics workflow and the reasoning for their use is discussed in this chapter.

C1. Experimental Design

An experimental design proves critical for evaluating a given proteomic technology and its suitability for accomplishing the objectives of a given project. The yeast oxidative stress project aims at developing mathematical tools to infer biochemical networks and make predictions for the inferred network model. In order to achieve this relevant information on i) Protein quantity and identity ii) Localization and functional status and, iii) Dynamics of protein changes in response to treatment is needed. For modeling purposes, the time scale used for spanning cellular processes becomes a critical component of an experimental design. The time interval between data points defines the time points for sample collection [49]. For a given experiment the choice of time points is decided based on previous knowledge of time course needed to observe any significant changes. Sample collection at specific time points is then incorporated into the experimental design assuming that the essential

elements of the proteome dynamics such as various modifications will be effectively captured [49]. For this project, the logarithmic time scale chosen for sample collection was based upon the kind of mathematical tools proposed for analyzing time series expression data. Based on previous *S. pombe* studies on CHP kinetics, a time course of two hours was found sufficient to capture any changes that may occur in response to stress conditions. Figure 7 summarizes the experimental design for the 2D PAGE experiments with ethanol (EtOH) added to the control fermentors and CHP (in EtOH solvent) added to treatment fermentors. Samples were collected immediately (0 min.) before addition of CHP or EtOH and at 3, 6, 12, 20, 40, 70 and 120 minutes following addition of these solvents. Proteins extracted (section C1.2) from the samples collected at each time point and fermentor was run on a 24cm 2D gel, explained further in following sections. Each time point in a group (CHP vs CTR) was represented by three biological replicates. As mentioned earlier in microaary studies from the same samples (Sha, 2008), some artifacts were found to affect control samples after 20 minutes time point. The changes manifested as altered gene expression levels for many number genes, which led to the notion that limited media supplement was causing some kind of gene regulation Since not many genes were altered in control until 20 minutes, for data analysis purposes only the first five time points (0 to 20 mins) were utilized, as discussed later.

C1.1 Reference Gel/Master Gel

For a proteomics project of such scale where numerous comparisons need to be made in order to elucidate time and treatment effect, an idea of reference design was incorporated in the experimental design, falling along the lines of “sample pooling” approach, initially used for differentiating cell lines [70]. This approach, known for its cost effective means (as samples from same pathological background could be mixed together on one gel) and reduced non-specific expression background, was applied to the yeast project. A reference gel was generated by pooling together protein samples from all time points of wt chp treatment and time pint 0 of control group in equal concentrations and run as 300 ug load on a separate gel. This gel is assumed to contain all spots appearing at every time point and group of wild type samples in addition to spots restricted to individual samples. Most 2D

packages rely upon a virtual reference gel for spot matching purposes, which is actually a cumulative synthetic image generated from a set of gels being compared [71]. Having a reference gel is useful as it helps at the "image analysis" part of the workflow as any number of sample images can be aligned to the reference gel. Typically an algorithm would generate a spot pattern from the reference gel which is then matched up to sample gels. Figure 11 shows the reference map used for generating match sets for our given set of experiments (WT). Since reference gel was annotated, spots on sample gels would derive their annotations from corresponding spots on the reference map, after spot mactshing was completed.

Cell growth and stress conditions: The *Saccharomyces cerevisiae* strain used in this work was the BY4743 (; *MATa/MATα his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 lys2Δ0/+ met15Δ0/+ ura3Δ0/ura3Δ0*), derived from the S288C strain and acquired from the American Type Culture Collection (ATCC). Initial cultures were batch grown overnight at 30°C, 150 rpm, in minimal medium with sucrose (MMS) 2% (w/v) supplemented with uracil 20 mg/l, L-leucine 60 mg/l and L-histidine 20 mg/l. The cultures were then used to inoculate 1 liter New Brunswick BioFlo fermentors (biological replicates) and grown to mid-exponential phase (OD_{600}~1.5) in MMS 4% (w/v) supplemented with uracil 40 mg/l, L-leucine 120 mg/l and L-histidine 40 mg/l. Cultures were grown at 30° C, pH 6.0 and dO2 > 80%. Oxidative stress conditions were achieved by adding a solution of CHP of known concentration (determined by HPLC), to the fermentors, to a final concentration of 190 μM. Triplicates of each strain as well as controls (no added oxidant) were also performed.

Yeast cells used for experimental data were grown in fermentors instead of flask to be able to monitor and maintain cell growth conditions and accommodate massive volume of culture. The samples were collected at different time intervals: 0, 3, 6, 12, 20, 40, 70 and 120 min. and divided into three parts: for transcriptomic, metabolomic and proteomics data levels and experiments. Samples (~60 ml) were collected into ~ 300ml buffered methanol flasks (-40°C) at different time intervals, centrifuged (at -20°C) to collect cells that were washed in water and then freeze dried for 48 hours. The freeze dried samples were stored at -80° C until further use.

The overall process for proteomics comprises of three main steps: sample preparation, protein separation and characterization based on protein function. For high throughput analysis of protein expression and function data, it is important to implement strategies that are efficient for each step of a proteomics workflow. Here, for example we examine an optimal protein extraction, separation and visualization technique followed by subsequent steps of data analysis and protein identification.
Sample preparation isolates proteins from a given organism but becomes critical when aiming towards achieving accurate and reproducible results. Likewise, efficient separation can improve reproducibility in 2D PAGE technique with increased resolution. Since each area of functional, expression and structural proteomics comes with its set of protocols for sample preparation and characterization, it is important to evaluate relevant protocols before producing massive amounts of data.

For a sample preparation step that is aimed towards non targeted profiling, an extraction method that is representative of total proteome is very important. The ability to keep extraneous materials to a minimum along with the extrication of low abundant proteins can dramatically improve the performance of a 2D PAGE technique and influence findings downstream. Since proteins occur at different cell locations and have varying solubility, an extraction buffer that is excellent for solubilizing most proteins is definitely needed. 2-D gel electrophoresis separates proteins in two dimensions based on given features : i) Protein when placed in an appropriate medium will migrate in a gradient of increasing pH, till they reach their isoelectric point (pI) where the net charge of side chains is zero (horizontally). ii) Molecular weight using sodium dodecyl sulfate –polyacrylamide (vertically). For efficient comparison and identification purposes, it is important that the proteins are well resolved. A larger dimension (24cm) gel can resolve proteins (as spots) better in comparison to a smaller gel (7cm), however, certain limitations exist in terms of minimum amount of protein needed for a given dimension of a gel as well inefficient resolution at extreme acidic and basic pH values .

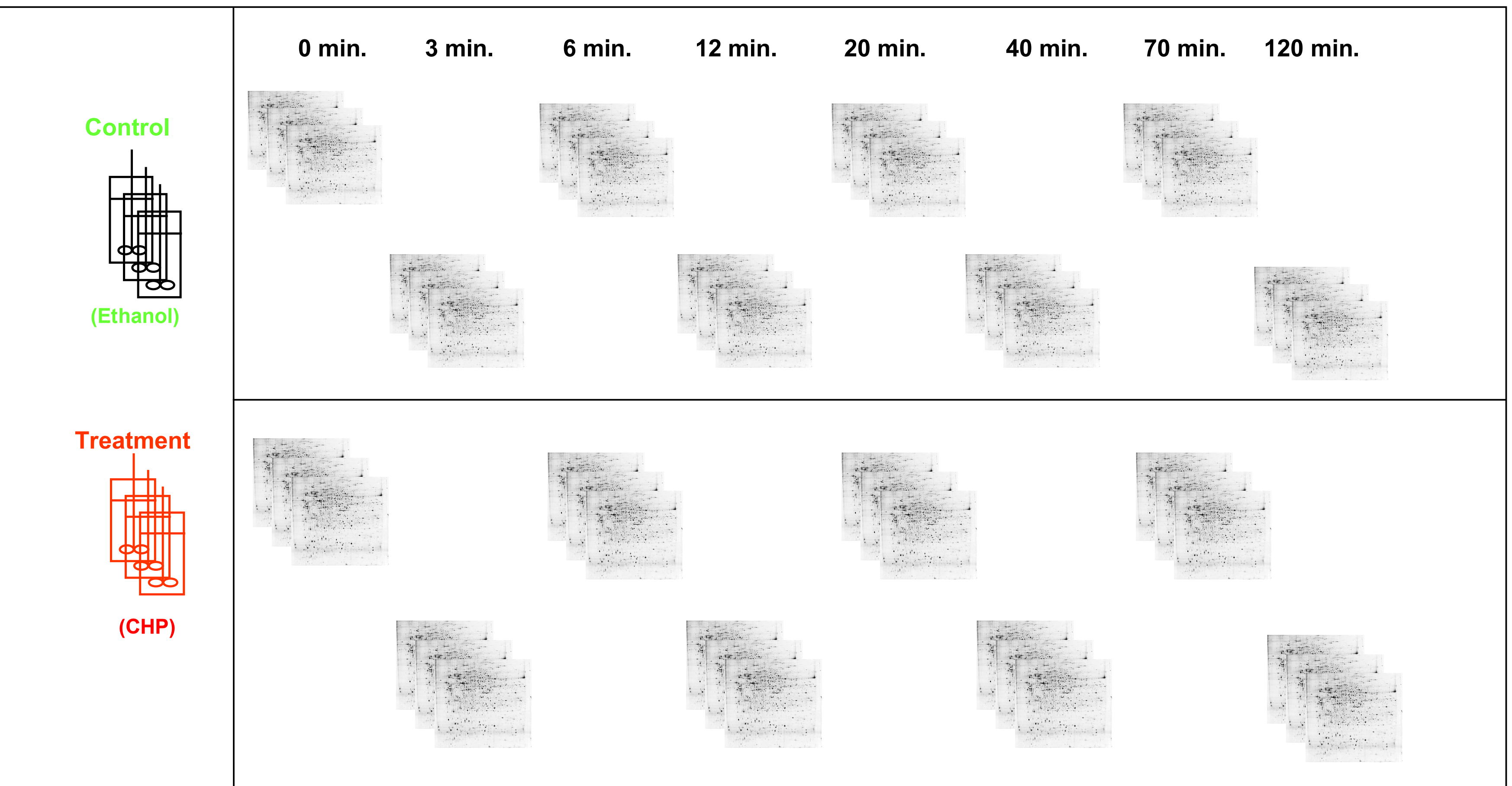


Figure 7: Experimental design of gel triplicates for each of the eight time points and two groups (control vs treatment). Ethanol (EtOH) was added to the control fermentors, while CHP was added to the treated samples Samples were collected immediately before addition of CHP or EtOH and at: 3, 6, 12, 20, 40, 70 and 120 minutes following addition of CHP or EtOH .

C1.2 Optimal Sample Extraction, Separation and Visualization Technique:

In absence of a universal sample preparation method, that can isolate all proteins of a cell, it is critical to choose an extraction method that would allow a global representation of the cell proteome and at the same time maintain the conformation and integrity of all proteins extracted. Typically a protein sample preparation method would involve cell disruption (physical or non-physical) step with a solubilization scheme that is representative of a protein sample. For cell lysis purposes we compared two methods and evaluated their performance based on total protein yield (for a 7cm gel). The Covaris protocol, an acoustic based cell disruption technique that utilizes high frequency ultrasonic waves failed to provide the required concentration of protein needed to run a 7 cm gel; whereas the beads method (0.5mm) was effective, as it did not cause any sample "frothing" upon cell lysis, facilitating protein solubilization (data not shown),

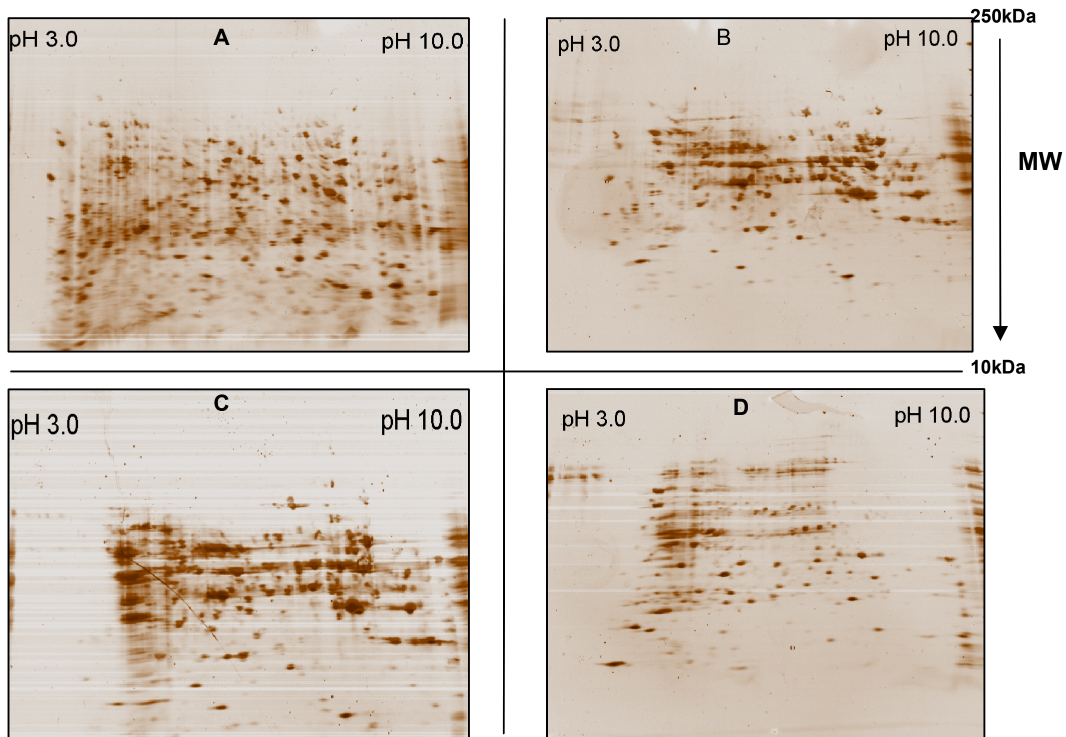


Figure 8: Gels (7 cm IPG strip) loaded with 15 µg of total soluble protein extracted from wild type yeast using: A) Yeast extraction kit, B) RIPA buffer, C) Thiourea/Urea. D) E. coli proteins were extracted as a control for comparison using E.coli extraction kit

To choose an optimal extraction buffer for our project we compared the three most commonly used protein extraction protocols for yeast cells; namely Yeast extraction kit (Calbiochem), RIPA buffer and Thioure/Urea buffer (Figure 8). Proteins extracted with these three buffers were cleaned for any contaminants such as salts, detergents etc. before running them on a 7 cm gel. Results show that both Yeast extraction kit and RIPA buffer generated well resolved spots with minimal streaking and image background. However, when the same extracts were separated on a 24 cm IPG strip with pH overlap (3-6 and 5-8), RIPA buffer fared better. As seen in Figure 9A1 in comparison to 9B1, Yeast extraction kit was not at all efficient in solubilizing low pH proteins.

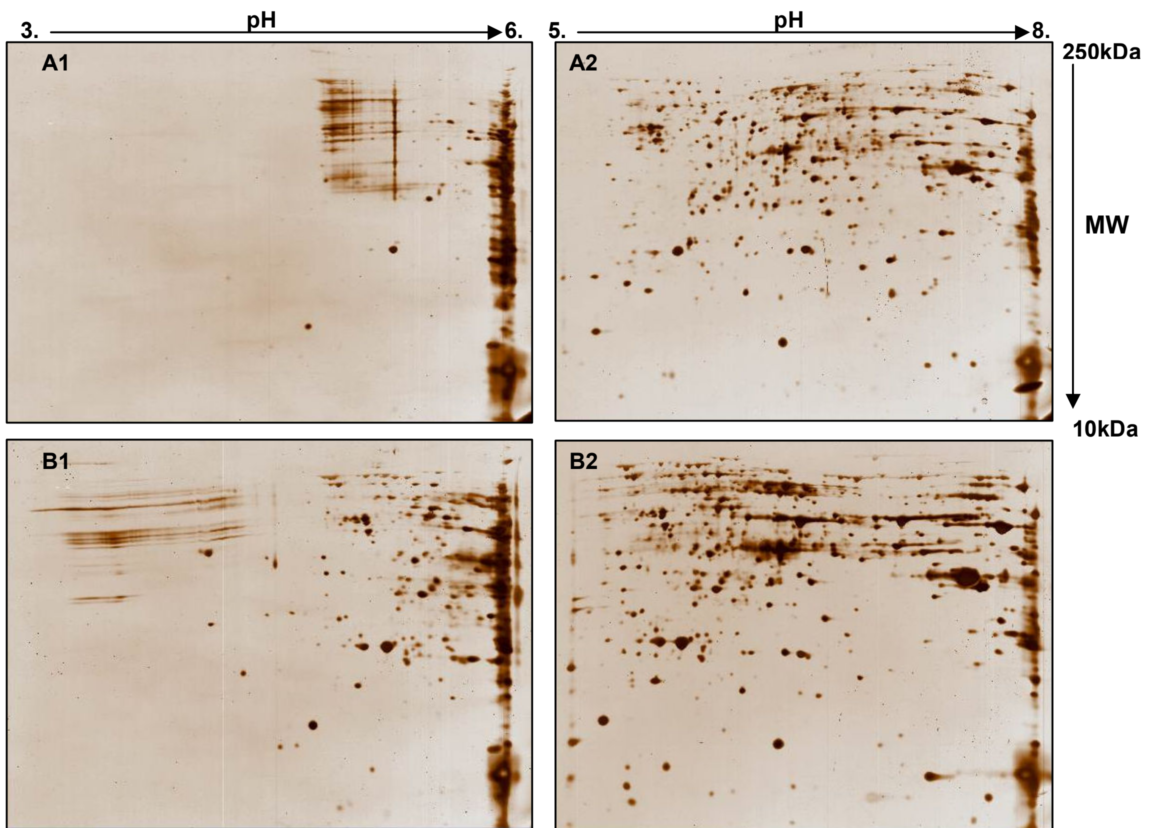


Figure 9: Gels loaded with 150 μg total soluble proteins from wild type yeast using A) Yeast extraction kit B) RIPA buffer. Samples were run on 24 cm IPG strips: 1) pH 3-6 and 2) pH 5-8.

Proteins separated by electrophoresis can be visualized by the use of stains, followed by high resolution 2-D imaging to summarize target proteins for excision and identification. Staining of protein spots is a critical step in gel-based proteomics, as protein identification and quantification depends upon the

ability of certain dyes to bind specifically to proteins. Since some proteins spots may show up with one stain but are not visible with another stain it is essential to choose a stain that can detect most, if not all, proteins present on a gel [72]. A comparison between four different gel stains: Coomassie Brilliant Blue, Silver stain, Flamingo Pink and Sypro Ruby were made. Coomassie Brilliant Blue with a protein detection range of 50-1000 ng suffered from poor sensitivity. Silver stain, with a range of 10-100 ng of protein detection was used for the preliminary method development but was found incompatible with mass spectrometry and hence discarded from further use [73] Fluorescent stains demonstrates an advantage over Coomassie and Silver stains as they did not detect any non-specific interactions, since spectral properties of fluorescent stains change only upon protein binding and not to other contaminants [73].

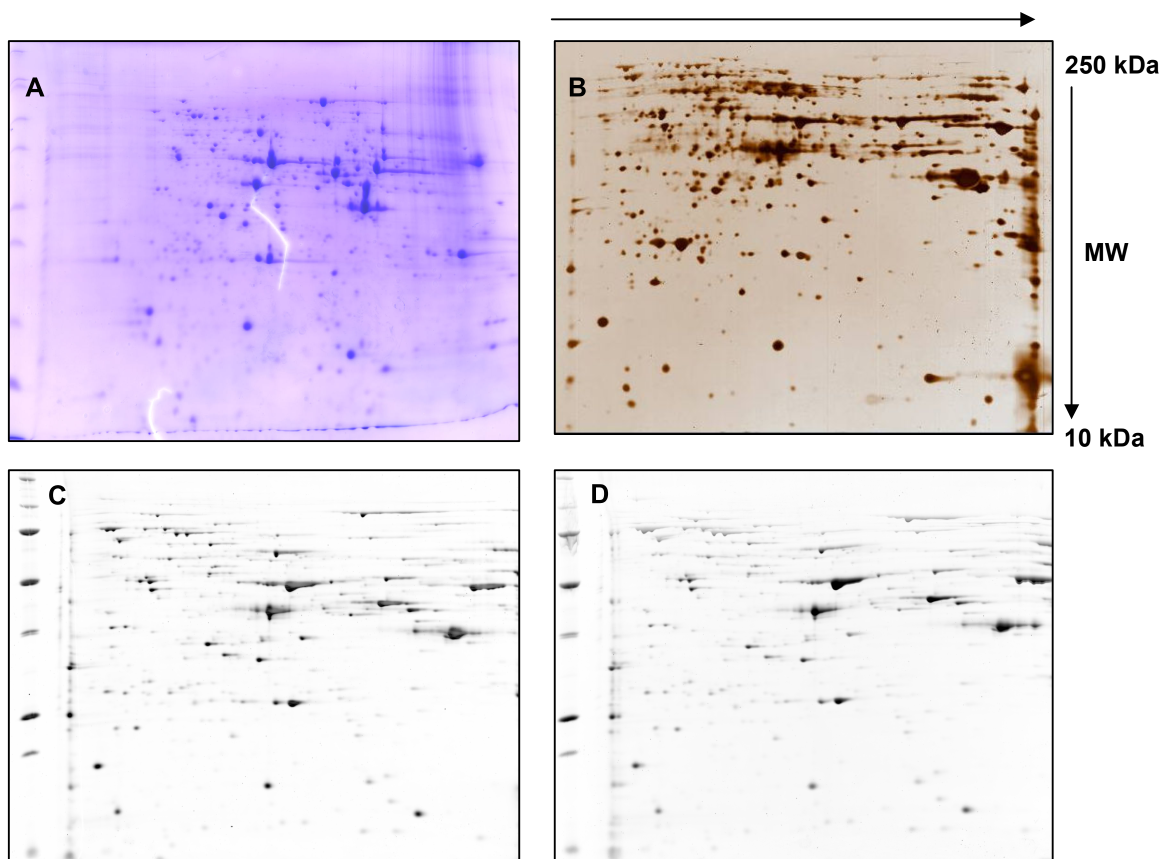


Figure 10: Comparison of four different stains for protein detection using 24 cm IPG strips A) Coomassie Blue, pH 3-10; B) Silver stain, pH 5-8; C) Sypro Ruby pH 3-10; D) Flamingo Pink pH 3-10 .

Upon comparison between Coomassie R-250, Silver, Sypro Ruby and Flamingo Pink stains (Table 2 and Figure 10), it was apparent that fluorescent stains worked best for yeast samples due to their reduced background, less time consuming staining protocols and Mass Spectrometry compatibility. Flamingo pink (BioRad) was chosen for our single stain experiments due to its excellent linearity, broad range of protein detection (2ng-2μg), and other advantages discussed below in Table 2 [73].

Table 1: Comparison of commonly used gel stains for visualization of separated proteins

Stain	Time (hr)	Detection Limit (ng)	Comments
Coomassie Blue			Widely used; Low sensitivity; Non-covalent binding [71]
G-250	15	1-16	
R- 250	12-48	30-100	
Silver Stain	0.25-3	0.5	Sensitive; MS incompatible; oxidative attack of silver ions on proteins [73]
Acidic silver nitrate			
Alkaline silver diamine			
Fluorescent dyes:			Compatibility with MS; High staining reproducibility; Staining in dark [73]
Sypro Ruby	5.5-12	0.5-5	Non covalent interaction
Flamingo pink	<5	0.25	High sensitivity; Fluoresces only when bound to proteins; broad linear range (3 orders of magnitude); Low background; Cost effective
Deep purple	3.5	0.1	Sensitive florescent-based stain, compatible with MALDI-TOF; staining saturation is possible
DIGE	0.75	0.025	Use of three labeled Cy dyes ; Expensive

Protein Extraction: For protein experiments, freeze dried yeast cells (11mg) were extracted in RIPA buffer containing: 50mM Tris-HCl (pH 7.2) , 50mM NaCl, 1% (v/v) NP-40, 0.5% Na Deoxycholate, 0.1% SDS with the addition of 1 μg/ml aprotinin, 1 μg/ml leupeptin, 1 μg/ml pepstatin and 50 μg/ml PMSF (phenyl methylsulfonyl flouride) to the buffer, just before use. The cell pellets were washed (twice) with a cold solution of "Wash Buffer" (Complete Yeast Proteome Extraction Kit/ Calbiochem) and vortexed (4 times) with 0.5 mm beads (200μl) for 1 minute duration, while on ice. Following protein extraction, benzonase (2 μl) was added to remove any traces of DNA and RNA and samples were incubated at -10°C for 30 minutes. Protein was recovered by

centrifugation at 20,000 x *g* for 30 min at 10°C. The protein concentration was assayed using a Non-Interfering protein assay kit from CalbIochem (Catalog # 488250).

C1.3 Two-dimensional gel electrophoresis:

Samples were cleaned from any substances, such as detergents or salts that can interfere during focusing using ReadyPrep 2D cleanup kit (BioRad) and resolubilised in rehydration buffer containing 8M urea, 2% CHPAS and 50mM dithiothreitol (DTT). Protein extraction and gel runs were scheduled randomly to remove any batch effects that may have occurred during extraction kits or gel runs.

Immobilized pH gradient (IPG) strips (ReadyStrip^{TM} 24cm, pH 3-10 non-linear, BioRad), were passively rehydrated overnight with protein solution (150 μg for samples and 300 μg for reference gel in 410 μl of rehydration buffer) in PROTEAN IEF focusing tray. Isoelectric focusing (IEF) of proteins was performed using the following step gradient: from 0 volt to 8000 volt and ramps up until 60,000 volt-hours has been reached. Following IEF, strips were equilibrated in three steps with a buffer containing 6M urea, 4% SDS, 0.375 M Tris-HCl (pH 8.8) and 20% glycerol made up in purified water with 130 mM DTT for reduction or 130mM IAA for alkylation. Equilibrated IPG strips were loaded on a 24 cm pre-cast / Tris-HCl /12 % acrylamide gel, sealed with 0.75% agarose, and electrophoresed overnight (~20 hrs) at 125 V per gel. All gels were stained with BioRad Flamingo Fluorescent Gel Stain and images acquired on a BioRad Molecular Imager FX-Pro at 532nm. A pixel size of 100 μm (for sample gels) and 200 μm (for master gel), was applied at two intensity settings: 39% PMT and 50% PMT and saved as .general scan (.gsc) files.

The resolution of an image drives the extent of detail one needs to observe. Small pixels (high resolution) were chosen as more detail was needed on sample gels. The photomultiplier (PMT) tube voltages reflect the sample intensity settings which is a rough estimate of how much sample is visible on a gel. Many FX-Pro (scanner) applications require a sample intensity selection of Low (55), Medium (45) or High (35). Low intensity setting on an image results

in too many saturated pixels whereas a high intensity setting makes spots appear fainter.

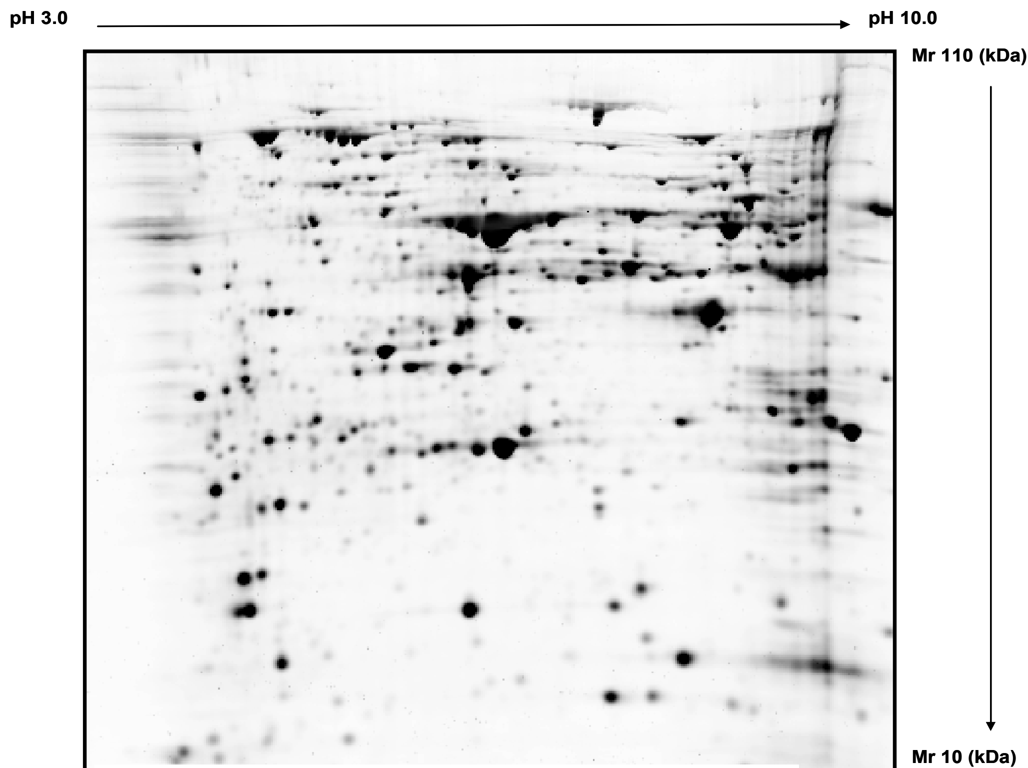


Figure 11 : Screenshot of a reference gel loaded with 300 µg of total soluble protein on a 24 cm IPG strip, pH 3-10 .

As mentioned earlier, for sample gels, a PMT setting between low and medium was optimal. The molecular weight (MW) and pI grid for a gel was calculated based on information on MW and pI of 2-D SDS-PAGE standard proteins, run separately for each batch of gels. Visualization and detection of protein spots separated on 2-DE gels were performed using PDQuest software version 8.0 (BioRad). ProteomeWorks Spot Cutter System (BioRad) was used to cut all protein spots for a given gel. The gel plugs were transferred to polypropylene 96-well plates with 0.1% acetic acid, sealed and stored at -20° C until further use.

C2. Proteome Informatics

2-DE in general is faced with many technical limitations and challenges some if which are at the Image Analysis part of the workflow, requiring several hours of

manual spot validation and gel alignment to facilitate sample comparison. Currently, several 2-D image analysis software packages are available that have been continuously developed and improved with time, in terms of reliable matching algorithms and less manual intervention. The commercially available software packages include: Delta 2D (Decodon), PDQuest (BioRad), Phoretix2D and Progenesis (Nonlinear Dynamics), Image Master 2D and DeCyder (Amersham), Melanie3 (GeneBio) etc. New packages that have recently emerged, offer robust statistical tools with different normalization methods to choose from and the ability to integrate data from various source and/or formats. In addition to these, some open source systems exist that can be used for storage, organization and analysis of protein experimental data purposes. PROTEIOS platform is one such system that is based on proteomics data publishing principles. Global proteome machine (GPM) is another system that is used for analyzing and validating proteomics information derived at tandem mass spectrometry level. Discussed below are the various steps of image analysis using PDQuest v 8.0 software and its implications to the quality and nature of data generated.

C2.1 Analysis of gel images in PDQuest:

The quality of two-dimensional separation and the algorithm used for spot detection drives the success of a computer assisted 2-D image analysis. The gel images are subjected to cleanup to reduce background smear and remove horizontal and /or vertical streaks, before spots can be detected and quantified. Spot detection is concerned with the individual resolution of each spot into its x, y coordinates, intensity and geometric properties. After spot detection, information characteristic for each spot is extracted to both quantify and assist in spot matching process. With a reference gel included, each spot can be matched to its counterpart across sample gels [74]. To be consistent with spot detection parameters, each gel (including master gel) was cropped to a dimension of 267.7mm x 222.7mm. The gel images were selected to be warped before automated matching was applied.

1) Background Subtraction: For background subtraction, PDQuest offers only one option which is the "floating ball" method which calculates the default radius size based on the largest spot cluster that is selected. The software first selects a "Faint spot" size that sets the sensitivity and minimum peak value parameters, followed by the "Large Spot Cluster" which sets the radius of the background subtraction rolling ball and streak removal rolling disk [59].These controls take care of vertical and horizontal streaks and allow background subtraction with noise smoothing. For PDQuest five "smoothing" filter types exist: Median, Weighted Mean, Power Mean, Contra Mean and Adaptive along with three different kernel sizes: 3x3, 5x5 and 7x7. These smoothing algorithms have to be selected before spot detection occurs. The use of smoothing algorithms in PDQuest is known to decrease the variability in comparison to excluding the smoothing algorithm completely. Likewise, the choice of filter type does not appear to influence the variability in spot quantification [75] Parameters, selected during a preliminary test of optimal spot detection and background subtraction, led to the selection of following parameters for the current project. A background subtraction with 99 pixels floating ball radius and a smoothing median filter with 5x5 kernel size, with speckles and streak filter was applied. The sensitivity parameter was kept as 47.58 in between the gels.

2) Spot Detection: The spot detection parameters starts off with defining faint and largest spot clusters for each gel image, followed by application of Gaussian modeling to find spot centers. Some of the spot detection techniques commonly used include: i) Derivative based ii) Watershed transform (WTS) iii) Linear programming (recently adapted for proteomics purposes) etc. [76]. The derivative method is based on two assumptions: a) Spots comprising the foreground are of different intensity than the gel, which is the background. b) Intensity is relatively uniform in areas far from borders of spots and gel [76]. However, the most popular technique for spot detection is the WTS method mainly because of its robustness for noise [77]. This method treats a gel as a topographic relief where spots are designated as depressions. Different catchment basins are assigned a unique label with a special label for pixels of the watershed [77]. In short, spot detection using WTS algorithm undergoes the following steps: 1) Detects centers for as many spots as possible 2) Segment

the gel such that each region contains one such spot 3) Model each region by a parametric spot model to extract a characteristic vector for each spot and detect and separate co-migrated spots [77]. After spot detection, the spot features generated help with registration of the 2-DE either automatically or via manual intervention (discussed below). Spot matching between reference and sample gels requires automatic pairing of spots that requires land marking [77]. Landmarks are reference spots used to align and position member gels for matching purposes and to compensate for any distortion observed. These spot are consistently present on reference and all sample gels. A total of 21 "landmarks", based on its resolution, intensity, location in all quadrants of a gel and its presence in between images were chosen, with an automated matching done after addition of each landmark.

3) Warping: Image warping is described as a function that deforms images such that it removes any image distortions that may have occurred due to running conditions, image viewing perspective, or to facilitate alignment of two or more images[59]. A major factor for geometric distortions occurring in protein pattern is due to current leakage which is a result of global change in electric field [59] In PDQuest one needs to select the warping feature before achieving any spot detection. Figure 12 shows a typical image before and after warping to the chosen reference gel.

4) Spot Editing: For spots that are saturated or irregularly defined, Gaussian modeling is not applicable, and as a result of determining spot quantity becomes difficult. Such spots along with overlapping spots can be defined manually using "Spot Boundary tools". PDQuest allows copying of spot boundary between images which is helpful as spots with modified boundaries can retain the same boundary all through its counterparts. However, this process can be time consuming as it needs to be done for each gel. Progenesis on the other hand has an advantage, as any modifications including

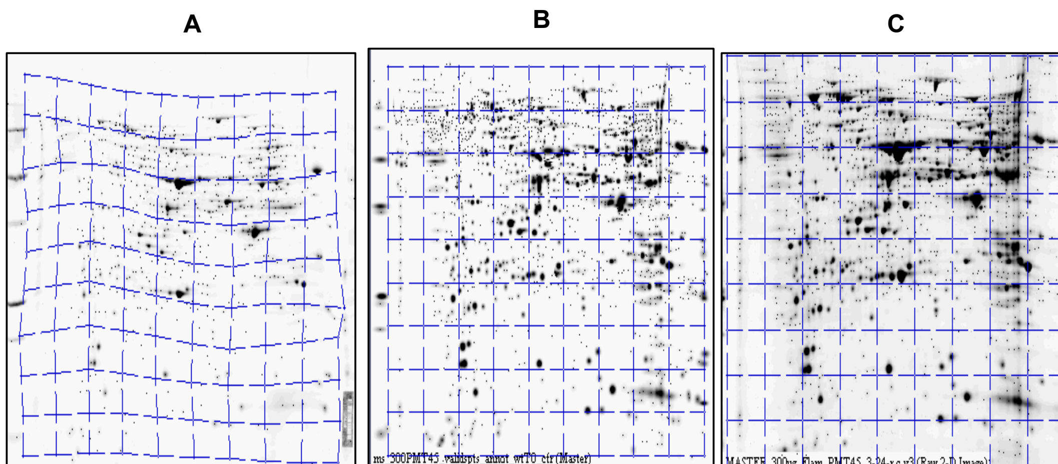


Figure 12: Gel warping step: Alignment of two images with their distortion mesh. A) Sample image B) Warped version of sample gel to align with, C) Reference image

spot editing, adding manual vectors and spot matching done on reference gel is propagated to the remaining images setup for an experiment. The correlation coefficient is a measure of the association between spot intensities on each gel for a given matchset and can range from 0 to 1, with a zero indicating no agreement and a one indicating perfect agreement. A matchset defines a group of sample gels and reference gels being compared and a synthetic image that is cumulative of all spots detected on its members (gels). Rigorous spot editing and manual matching was seen to improve the correlation coefficient between gels (data not shown).

5) Manual Spot Validation:

Manual spot validation was required in view of inefficient automated matching performed by the PDQuest software and to account for any technically generated missing data (Fig 13). The given below protocol was developed to allow manual spot validation in the match sets of the Yeast Oxidative Stress Project. Matchsets for the two treatments (control and CHP) and the first five time points (T0 - T4) were used for data analysis. Each gel had close to ~1200 spots; therefore, ~406 spots based on the ascending quality scores were initially chosen for spot validation as discussed below:

1. A spot quality score list was generated for each matchset from wt T0_CHP through wt T4-CHP and integrated together. The list was then sorted in a descending manner showing spots with high quality score on top for each matchset. The quality score is a number that ranges within 0-100 based on five attributes of a given spot as follows: i) Gaussian fit: Determines how well a spot fits a Gaussian model ii) X streaking: Gel streaking effect on a spot, in a horizontal direction iii) Y streaking: Gel streaking effect in on a spot in the vertical direction iv) Overlap between spots and v) Linear range of a scanner: the range of the scanner where the peak intensity of a spot is linear. In the last case, it is important to remember that saturated spots are outside the range and thus are poorly scored.
2. Spots were then evaluated, one by one based on their resolution, presence/absence in a replicate/ boundary etc. Spots identified by their SSP and *mrpI* annotation were then manually validated across all sample gels one by one.
3. Spots present on the master gel with annotation were manually matched to their corresponding spots in the replicate gels of a given match set. In case of absence in a replicate, a 3D montage could inspect the spot area. If a gaussian curve was observed, a spot was added, followed by manual match; otherwise, it was listed as “NA” in the excel sheet. Sometimes a spot can be seen in 2 out of 3 replicates in which case a placeholder was positioned for the missing replicate.
4. Upon correct matching of a spot to a reference gel spot the SSP changes. All changes were documented as PDQuest works around with SSP only and not mrpi annotations. Any spots that were cancelled at this step were also documented in the excel sheet.
5. If overlapping spots could not be redefined or split, they were consistently combined together into one spot. Spots that were difficult to comprehend were color coded. Some spots were displaced and needed Mass Spectrometry (MS) validation and were commented for future reference.
6. Boundaries for spots that were redefined were saved and applied to its counterpart across all gels. Since PDQuest could memorize only one

boundary at a time, redefining spots across matchsets could be achieved one at a time till all spots of interest were covered, which time was consuming.

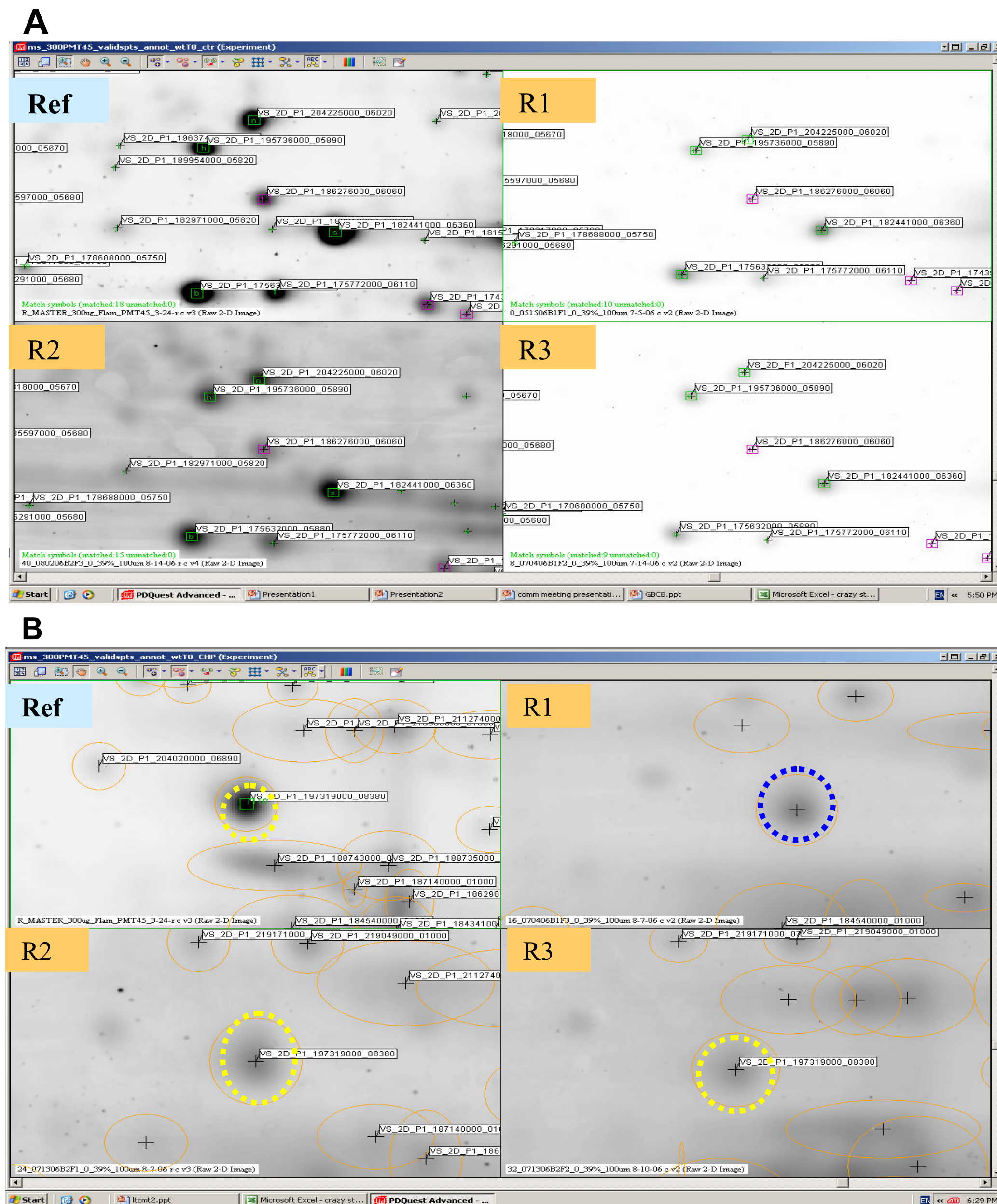


Figure 13: Screenshot of a wt-T0-chp matchset using PDQuest software with Ref: reference gel; R1,R2 andR3: three biological replicates. A) Manual matching (pink square) with Landmark spots (green square) B) Automated spot matching error (blue oval in R1) ; observed as missing value in data set ,which can be corrected by manual validation.

C3. Statistical Analysis of Protein Expression Data

Two dimensional polyacrylamide gel electrophoresis (2D- PAGE) technique developed in early 1970's, has improved over the past few years with a potential for high-throughput data production. Very few studies exist that have investigated appropriate statistical approaches to analyze 2D gel based protein expression data [78]. And any comparisons that have bee made, for statistical pre-processing protocols, portray that systematic data analysis of 2D gel based data needs to deal with the following features :i) availability of limited number of gels; ii) need to focus upon a large number of spots and iii) statistical protocol used [75]. Data typically obtained from gel image processing is a large matrix of variables and parameters) which represent protein spot intensities and different experimental conditions respectively. Depending upon the kind of software used, parameters such as spot pattern, spot quality, spot volume, area and peak height can be determined which are software driven [78]. The intensity of the spot is directly assumed to be related to the amount of protein on a gel and any changes in protein intensity, is therefore estimated by tracking protein spot intensities in the gel images. In spite of its impressive ability to separate thousands of proteins as spots, 2DE faces some shortcomings when it comes to differentiating between true spots from spurious noise. Before any comparisons can be made, the systemic variations across gels need to be adjusted [78]. This and many other limitations have led to the development of several image processing tools that cater to this "bottleneck" in their own specific ways. In order to conduct a systematic analysis on 2D gel data, pre-processing is needed which includes: 1) Normalizing spot intensities to remove technical or developmental effects 2) Imputing missing values or spot intensities and 3) Transforming raw data to normally distributed variables.

Discussed below are the various steps of preprocessing currently used and its relevance to our yeast project and protein expression data.

C3.1 Normalization:

A large part of spot variation observed in 2D gel experiments is related to either gel effect or uncontrolled factors such as differences in sample preparation/running conditions, differences in sample loading and/or staining or differences due to image acquisition. These sources of systematic variation can not only obscure biological changes but also complicate identification of true expression changes. In order to accurately compare protein expression changes across gels, it is essential to normalize data that compensates for these variations by bringing all gels to a common level before any comparison can be accomplished. Image analysis packages like PDQuest and Progenesis (SameSpot) have in-built normalization methods that have been extensively used by researchers worldwide. However, when quantitative performance of these commercially available software's were evaluated in previous studies , on a same set of gels, results showed that the total variance observed for protein spot quantification between replicates differed depending on the kind of the package used [79]. Therefore it is critical to compare several methods before choosing a method that works best for one's data. Summarized below are several normalization methods commonly used for proteomics data; some of which are offered as a part of PDQuest software [71]. Progenesis (SameSpot) offers only one normalization algorithm which does not allow a choice of applying other normalization methods for its dataset of normalized spot intensities generated.

A) Single Spot Normalization:

This method is based on linear scaling of spots to a known value of a single spot which could be a marker protein in the sample, such as an internal standard that is present in all the members of an experiment or an arbitrary value used for comparison purpose [71]. This spot is also referred to as base spot and is used for all volume calculations. The formula takes the volume of a spot n (V_n) and divides it with the volume of the selected base spot (V_{bs}) and multiplies it by the user-defined scaling value (S).

$$NV_n = \{V_n/V_{ss}\} \times S$$

This method is useful if the samples are spiked with a known protein marker which was not how gels for yeast experiment were designed.Therefore we did not use this method of normalization.

B) Total Quantity of Valid Spots (PDQuest) :

This method represents each individual spot for a given gel relative to all valid spots present on that gel. Typically the calculation is as such: (Raw IOD value of spot n/ Sum of all valid protein spot IOD's on that gel) X S (scaling factor). Due to small values, the ratio can be multiplied by a scaling factor defined by the user. IOD corresponds to a volume V= area X mean optical density (OD). The underlying assumption for this method as follows: i) Few protein spots change within the experiment, and the changes average out across the whole gel ii) Spots that do change do not contribute much towards the total volume of spots. This method is best applicable when little information is known about the possible source of sample variation [80] [81]. A modification to this method would be to use "same number of valid spots present across all members" which is applicable to an experiment where different numbers of spots can be detected for sample gels and can be counted towards the number of valid spots. This was relevant to our case where we had manually validated ~ 406 spots, matched all across sample gels and used these in data analysis. We chose this method for data normalization with parts per million (ppm) scaling factor. Box plots generated using this normalization method and its modified version are shown in Figure 14.

C) Total Density on Gel Image (PDQuest):

This method relies on the assumption that the total density of an image (i.e., background density plus spot density) remains relatively consistent from one sample to another. The OD value of all pixels for any spot in a member gel is divided by the OD value of all the pixels within the given image. A qualitative evaluation of our sample gels showed inconsistency in background intensity of replicates, which led us to believe that this method is not applicable in our case. The background intensities of our sample gels varied a lot which can be

attributed to differences in staining procedure in spite of using the same protocol for all gels. Figure 13 A depicts gel intensity variation observed between biological replicates (with no changes in contrast applied).

D) Specified value (PDQuest):

This method corrects each spot within an individual gel by a user-definable normalization factor for that gel. It is utilized when one needs to correct for known variations across gels for example- the amount of cells that went into an experiment, the amount of extract loaded, etc. Value Normalization is a modification of the above method, where the volume of each spot is multiplied by a value and then by the area of a single pixel in mm^2 units (provided scaling information is available). The area component of the equation corrects for any resolution differences that may have occurred for sample gels. This method was, therefore, not used for our dataset, as it was difficult to decide what single value to use that could account for any variations that may have occurred as part of the process. Since all protein extractions were done using 11 mg of freeze dried yeast, in the same volume of extraction buffer, with 150 ug protein load/ IPG strip, theoretically this method was not applicable in our case.

E) Mean of Log Ratios/ Match Ratio Method (PDQuest):

This method calculates what is known as a complex ratio (R) for every spot pair common between the two images; one of which is a base image (also known as a reference gel). R is the mean of all log ratios (log spot quantity of gel / log spot quantity of master gel) and is calculated as follows:

i) The natural logs of each volume ratio (matched spot in sample gel/ matched spot in master gel) is calculated and sorted into ascending order.
ii) These values are then divided into (X-(bin size-1)) overlapping bins, each having a bin size of X/4(X being the number of matched spots)
iii) Upon identifying the bin with the lowest range, that is the bin with the smallest difference between the first and the last item, the average value is calculated. the antilog of which is represented by R.

F) Local regression model LOESS (PDQuest):

A sophisticated normalization method, also commonly used for microarray data, is an intensity dependant (mean) normalization that is less susceptible to outliers than a simple linear regression and corrects for any differences in labeling efficiency occurring for different concentration levels in the gel. This model calculates a curve in the scatter plot which minimizes the distance to all points in the plot. The curve is then used to calculate the normalization factor for each spot. It is perceived that a pairwise comparison occurs between a sample gel and a reference image, Box plots generated using this normalization method are shown in Figure 14..

G) Total volume ratio method:

In this approach each spot volume is multiplied by the ratio volume (total volume of all spots in base image /total volume of all spots in study image) and then by the area of a single pixel in mm^2 (A_{px}) to take into account any differences in image resolution.

$$NV_n = V_n \text{ X } \{V_n/V_{ss}\} \text{ x } A_{px}$$

H) Ratiometric Normalization Method:

This method is relevant to differential in gel expression (DIGE) experiments that use an internal standard. Ratiometric spot volumes are calculated by expressing the volume of the spot relative to the spot volume of the internal standard. This method assumes that not many changes occur in the spots across the images. Since gels used for this project were single stain experiments, this was method was not applicable in our case.

I) Quantile Normalization:

This method is commonly used for microarray data, but its relevance is being recognized for proteomics data [82]. Quantile normalization is a powerful procedure which lines up the intensity values not only at the median, but also

enforces identical distribution across gels (Fig 14) [80]. The goal of quantile normalization is to make the distribution of intensities equal across all gels. Studies that have investigated protein content of knock out mouse strains show that quantile normalization is a powerful method and can remove any variations measured in fluorescent spot intensities derived from 2D gels [82].

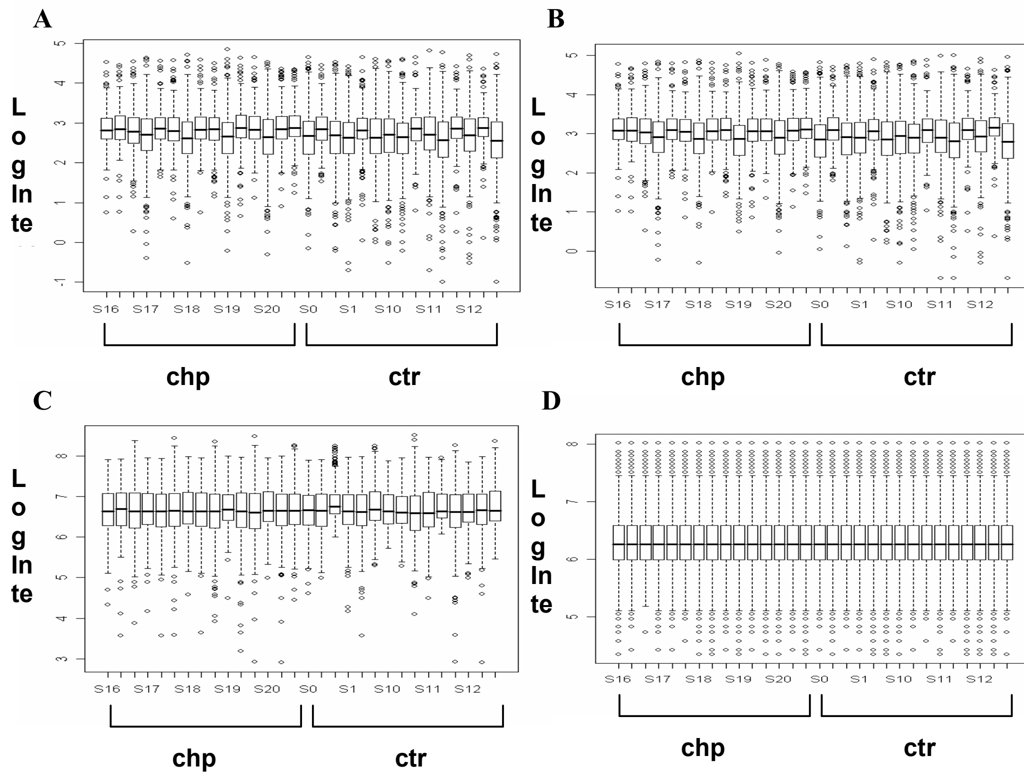


Figure 14: Normalization of expression data. Boxplots of spot intensities from each experimental group are plotted for A) log transformation of total quantity of valid spot method B) log transformation of total quantity of valid spot (406 spots) , C) log transformed LOESS and D) Qunatile Normalization method

All four normalization methods were compared after log transformation was applied. On the basis of boxplots, spot intensities for all gels lined up well at the median for both LOESS (S10 is an exception) and Quantile method. On the basis of the 2-way ANOVA model applied to the different normalization methods described above, we found that the quantile method provided the highest number (26) of spots showing a significant change in expression level at pFDR < 0.05 level (see Table 3). Significant proteins found with the quantile method overlapped with the findings of other normalization methods compared, which reinforces the fact that strong patterns can be picked up by any

normalization method applied. Quantile normalization has been previously applied to fluorescent stained 2DE experiments, and has been recognized as an optimal choice of normalization, in comparison to median and total spot volume methods. Studies using quantile method for protein expression data, were able to detect Superoxide dismutase (MnSOD) protein as a means of validation, expected to be differentially expressed between groups under study [82]. With our dataset we were not sure what to expect but as discussed later (results section) significant proteins identified with quantile method seem biologically relevant to stress conditions. Quantile method assumes an identical spot intensity distribution across gels, which is complicated as determining intensity distribution for a given dataset can be difficult. But given the fact that it is a sensitive method and with an overlap of significant proteins observed with other normalization methods, the quantile method was our choice of normalization method

C3.2 Missing Value Imputation:

Missing spot intensities can be caused by two sources; namely, experimental variation observed due to variability in staining, or low spot intensity detected by the image analysis software etc. and biological variation, which is associated with either a true missing spot from one group or appears missing due to migration to a different location [78]. It is important to approach the missing value issue by first defining the source of missing intensities and also evaluating the proportion of missing intensities for a given dataset. For our study we treated all missing values as undetectable spots and, therefore, replaced missing values, with the lowest "detectable intensity" observed for a given gel. In PDQuest, missing spot intensities can be replaced with a specific value or with a zero but was not applied to our project. Another option would be to remove observations with more than 20% missing values but that can prove expensive as a lot of data is lost [78]. A third option, recently suggested in a systematic analysis study is to use a random process to substitute missing values with a plausible "set of detectable" intensities [78]. Imputing values for missing spot intensities can thus be achieved in two ways: i) Single value or ii) Multiple value imputation. Therefore, one could either use the lowest intensity

detected for a gel or randomly impute dataset with multiple low spot intensities, organized as a list.

C3.3 Data Transformation:

The normalization technique is known to keep the non-Gaussian distribution of the raw data, without altering spot intensities. Data specifically generated by PDQuest and Progenesis image analysis software have been specifically used for comparative studies to evaluate effects of log transformation of normalized and non-normalized spot intensities data . PDquest allows a user to remove systematic biases from expression data; however, the distribution of the outcome variable cannot be tested within the software, which creates a need to transform data outside PDQuest. Results show that log transformation reduces the skew in spot intensity distribution and converts the data to a normal distribution. Effect of log transformation on dataset from wt_T1_CHP treated matchset (randomly chosen for this purpose) shows the change in distribution after transformation. Gel# 16, 24 and 32 are the three biological replicates in this given matchset (figure 15).

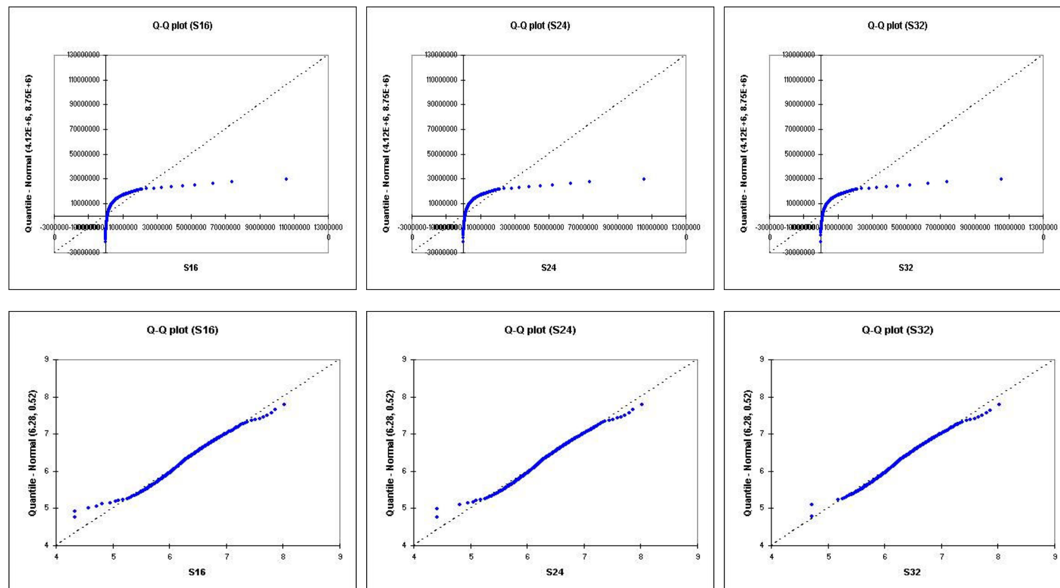


Figure 15: (A) Q-Q plot of normalized, untransformed spot intensities for three biological replicates of wt_t0_chp treatment matchset. (B) QQ plot of normalized spot intensities after log transformation for the same three replicates. The straight line represents a normal distribution and the log transformation of normalized intensities lays all points close to this line.

A normal probability plot also known as qunatile-quantile (Q-Q) plot is a graphical representation to determine if a data set comes from a population with a normal distribution. The data is plotted on the X axis whereas the normally distributed quantile is plotted on the Yaxis. If the data is normally distributed, then all data points fall along the straight line, as shown in Figure 15b. It was found that both PDQuest and Progenesis (SameSpots) do not provide a tool to test the distribution of a data, which emphasizes the fact that data generated from this software should not be assumed for normality of spot intensity distribution. It is important to realize that this assumption needs to be fulfilled before applying ANOVA tool to differentiate between groups that are investigated.

C3.4 2-way ANOVA model:

To assess the significance of differences between spot intensities across two different experimental conditions (time and treatment), a 2-way ANOVA spot-by-spot model was used in SAS version 9 (SAS Institute Inc., Cary, NC, USA):

$$y_{ijk} = \mu + T_i + V_j + (TV)_{ij} + \varepsilon_{ijk} \ ,$$

where y_{ijk} is the intensity measured on the gel for time i (in this case, i=0,2,..,20), treatment j (in this case, treatment is control or CHP) and replicate k; μ is the overall mean intensity of this spot; T_i is the effect of the ith time; V_j is the effect of the jth treatment; $(TV)_{ij}$ is the interaction effect between time i and treatment j; ε_{ijk} is the residual for time i, treatment j and replicate k. The positive False Discovery Rate (pFDR, cutoff 0.05) with multiple-testing adjustment was applied (time, treatment and interaction effect) to correct *p* values [83].

Table 2: Comparison of 2-way ANOVA results from log-transformed intensities normalized using four different methods. Protein annotations that are significantly differentially expressed (pFDR< 0.05) for quantile normalised method are shaded in yellow. TQVS*: total quantity of valid spot, LOESS**: Local regression model, (): number of spot significant

Master Spot No. (wtt3chp)	Master Spot No.	Annotation	TQVS(17)*	TQVS(406 spots(14)	Quantile(26)	LOESS(18) **
1016	1	VS_2D_P1_186785000_04800	-	-	0.032	-
1105	2	VS_2D_P1_198852000_04560	-	-	0.027	-
5116	3	VS_2D_P1_207000000_05900	0.07	-	-	-
1118	4	VS_2D_P1_209031000_04830	0.007	0.036	0.043	0.035
8116	5	VS_2D_P1_209730000_01000	-	-	0.04	0.022
6102	6	VS_2D_P1_240055000_06080	-	-	0.03	-
2103	7	VS_2D_P1_248339000_04930	0.04	0.037	0.01	0.04
2112	8	VS_2D_P1_249878000_05140	0.005	0.01	0.0007	0.0007
2110	9	VS_2D_P1_250716000_05040	0.028	0.042	0.03	0.011
5204	10	VS_2D_P1_280477000_05780	-	-	0.0365	0.033
4221	11	VS_2D_P1_283141000_05720	<0.01	-	-	-
2203	12	VS_2D_P1_293546000_04970	0.014	0.039	0.0006	0.006
3211	13	VS_2D_P1_305391000_05360	0.046	-	0.0268	0.041
3209	14	VS_2D_P1_308730000_05360	0.048	0.04	0.028	-
2312	15	VS_2D_P1_317998000_05170	-	-	0.047	-
3603	16	VS_2D_P1_327007000_08340	0.04	-	0.017	0.0226
4324	17	VS_2D_P1_327640000_05700	-	-	0.04	0.0323
4309	18	VS_2D_P1_332611000_05700	0.048	0.04	0.047	0.04
5301	19	VS_2D_P1_336275000_05750	-	-	0.049	-
5406	20	VS_2D_P1_355108000_05790	-	-	-	0.045
7804	21	VS_2D_P1_358440000_05670	0.008	0.011	0.03	0.009
6411	22	VS_2D_P1_358891000_06690	-	-	0.03	-
2402	23	VS_2D_P1_369586000_04930	-	-	0.032	-
1507	24	VS_2D_P1_410349000_04550	-	-	0.043	0.033
6526	25	VS_2D_P1_413809000_06920	0.0002	0.00057	0.000009	0.00005
4512	26	VS_2D_P1_414590000_05700	-	0.036	-	-
8512	27	VS_2D_P1_430638000_01000	-	-	-	0.049
7512	28	VS_2D_P1_440102000_01000	-	-	0.018	-
6602	29	VS_2D_P1_467567000_06150	0.02	0.042	0.045	-
2612	30	VS_2D_P1_469722000_05100	0.018	-	-	-
6605	31	VS_2D_P1_473144000_06420	0.018	0.034	0.025	0.03
3611	32	VS_2D_P1_492318000_05390	0.022	0.034	-	-
6703	33	VS_2D_P1_544264000_06160	-	0.037	0.047	0.047

C4 Protein Identification with Mass Spectrometry

Spots with significant differential expression between treatments were selected for protein identification using MALDI-TOF/TOF (Applied BioSystems 4800) and /or LC/MS-MS.

***In-gel trypsin digestion**:* The selected protein spots were excised from gels using the ProteomeWorks Spot Cutter System (BioRad). The excised gel plugs were transferred to polypropylene 96-well plates (with 0.1% acetic acid), sealed and stored at -20° C until further use. The protein spots were washed with 200 µl of 1:1 (v/v) solution of 25 mM ammonium bicarbonate and 50% acetonitrile (ACN) and vortexed overnight. The wash step was repeated for 2 hr mixing until gel particles were completely clear. The gel pieces were then dehydrated by adding 100 µl of ACN for 15 minutes at room temperature. After ACN removal, the gel pieces were air dried (5-10 minutes) and rehydrated with 10 µl of freshly prepared trypsin solution (10 ng/µl in 25mM ammonium bicarbonate, Promega Trypsin Gold) on ice. After rehydration for 15 min, the gel particles were immersed in 25 µl of 25mM NH_4HCO_3 digested overnight at 37°C. The peptides were recovered and extracted twice with 30 µl and 50 µl of 0.5% formic acid in 40% ACN respectively. The extracts were combined and dried in a Centrivap evaporator (without heat).

Depending upon the kind of mass spectrometry used for sample identification, the digests were either mixed with freshly prepared matrix solution containing 2mg/ml of alpha cyano-4-hydroxyxcinnamic acid (CHCA) in 0.5 % formic, 40% ACN solution (MALDI/TOF/TOF) or reconstituted in 20 µL of 0.1% triflouroacetic acid (TFA) (LC/MS/MS separation). For electrospray MS, the peptides were ZipTip cleaned before analysis. The $ZipTip_{\mu\text{-}C18}$ was prewet by aspirating the wetting solution (50 % methanol) three times before dispensing to waste. The ZipTips were then equilibrated in 0.% TFA (made in MilliQ water) three times with slow aspiration. Peptides were bound to ZipTips by aspirating and dispensing the re-suspended digest, a total of ten times. The tips were washed in 0.1% TFA, ten times, with the purified peptides eluted from the ZipTips into 20 µL of 50% methanol and 0.1% TFA.

***Protein Identification**:* Several reasons exist for the limited identification rates observed for advanced proteomic studies, including the inability of one MS method to identify all proteins from a given gel due sensitivity, or inherent limitations of a principle used in one method. Two individual mass spectrometric techniques: MALDI-MS and (nano-LC) - MS/MS were combined, which led to high identification rate of gel separated proteins.

(i) MALDI/TOF/TOF: The efficiency of desorption and the ionization of a given sample in MALDI is very sensitive to the sample preparation technique employed, as contaminants such as detergents and salts can disrupt matrix-sample crystallization. The matrix-sample deposition as well as the matrix-solvent choice can contribute towards the performance of MALDI analysis [84]. For MALDI-MS analysis, 2 μl of sample digest were mixed with freshly prepared matrix solution (2mg/ml of alpha cyano-4-hydroxyxcinnamic acid (CHCA) in 0.5 % formic acid in 40% ACN solution). Depending upon signal strength, the sample: matrix volume was adjusted (1-2 μl), spotted onto a metal target and allowed to air dry. For MS analysis the target plate was loaded into an Applied Biosytems 4000 MALDI mass spectrometer and spectra was acquired in positive ion reflectron mode with an accelerating voltage of 18 kV. Masses were calibrated with external standards which can include 1 μl of dilute glu-fibrinopeptide (5 pg/ul) to the MALDI target on spots. Protein spots were identified by searching experimental MS/MS spectra against NCBI-non redundant *Saccharomyces cereviase* database. For MALDI/TOF/TOF generated data, peptides were queried using MASCOT searching engine ((http://www.matrixscience.com) against NCBI-nr (non redundant) database, with an allowed peptide mass tolerance of 90 ppm, two missed cleavages and 0.4Da parent ion mass tolerance. Variable modifications of methionine oxidation (± 16 Da) and carbamidomethylation of cysteines (± 57 Da) was applied. Identifications with a MASCOT score of >50 (accepted as significant threshold) or a confidence interval of 95% with at least 2 peptide matches are reported in the result section (table 4).

(ii) Nano-LC/MS/MS: The analysis was performed on an TEMPO MDLC nanoflow LC system (Exsigent) interfaced with a QSTAR ELITE mass

spectrometer (Applied Biosystems/MDS Sciex). The samples were reconstituted in Solvent A (98% Acetonitrile 2% water and 0.1% Formic Acid). The nanoflow HPLC system (TEMPO MDLC system, Applied Biosystems, USA) was used to deliver a flow rate of 300nl/min. Chromatographic separation was accomplished by loading peptide samples onto a 15cm column (75um inner diameter, packing material, Vydac MS 300 C18, 5um, Alltech, USA).Sequential elution of peptides was accomplished using a linear gradient from 5% solvent A to 60% solvent B (98% acetonitrile, 2% water and 0.1% formic acid) in 60 minutes. The resolving column was connected using a fused silica transfer line (20um inner diameter, 360um outer diameter, 10um tip inner diameter, New Objective, Cambridge, MA, USA). The mass spectrometer was operated in positive ion mode with a resolution of 10,000 -12,000 at full width half maximum (FWHM) for the Q STAR Elite using a source temperature of 200°C. Information dependant analysis was employed (three most abundant ions in each cycle), 1 second MS (m/z 350- 1500) and max 2 seconds MSMS (m/z 50-2000), 30 seconds dynamic exclusion. For MS/MS analysis, survey scans were acquired from *m*/*z* 300 to 1500 with up to five precursors selected for MS/MS from *m*/*z 50* to 2000 using dynamic exclusion, and the rolling collision energy was used to promote fragmentation.

Protein identification was performed using Protein- PilotTM Software 2.0 (Applied Biosystems) which uses the Paragon algorithm as a search engine. Each MS/MS spectrum was searched for species of *Saccharomyces cerevisiae* against the NCBI_200803_23 database. The unused protein score is ProteinPilot's measurement of protein identification confidence taking into account all peptide evidence for a protein, excluding any evidence that is better explained by a higher ranking protein. A value of 2 or greater indicates 100% confidence; a value of 1.5 or more is 95% confidence and >1 is equivalent to 90% confidence for protein identification. Sequence coverage was calculated by dividing the number of amino acids observed by the protein amino acid length. Identifications with ProteinPilot with an unused score of ≥ 2 (accepted as significant) are reported in the result section (table 4).

CHAPTER 4

D. RESULTS AND DISCUSSION

Recent years have seen a dramatic transformation of redox proteomics and its involvement in examining pathological conditions at the molecular level. Cells can constantly sense and adapt to disturbed redox status by inducing and/or repressing genes and by modifying other cellular components that are critical for maintaining a reduced cellular environment. This process is generally called stress response. Insights into the molecular mechanisms of this process have been investigated at the genetic, proteomics and metabolomics levels in yeast *Saccharomyces cerevisiae* (see for example [37], [33]).

Since proteins are the effector molecules of a cell, data obtained at the protein level can add new dimensions to the information derived from gene expression [85]. In particular, an extensive wide screening of modified proteins is exciting and challenging since most of it still remains to be explored [17]. A number of diseased states show evidence of protein modification (i.e. post-translational modifications). Identifying oxidatively modified proteins and their function can thus provide insights into the underlying mechanisms of cellular dysfunction [86]. With recent advances in methodological techniques, many protein modifications have been categorized with a large number of sulphur-containing residues being viewed as redox sensing/signaling switches [87]. Reactive oxygen species such as superoxide anion, hydrogen peroxide (H_2O_2) and hydroxyl radical generators have been extensively studied in various model organisms for their interaction with proteins and their effect on molecular modifications [86]..

In this study, we used comparative proteome and transcriptome analysis to identify proteins and genes involved in *S. cerevisiae* response to oxidative stress induced by the aromatic peroxide CHP. The genome-wide transcriptomics data is reported and discussed in another thesis (Sha, 2006). Here we focus upon the proteome analysis of oxidative stress induced by CHP.
Protein modifications include oxidation of amino acid residues, polypeptide fragmentation or formation of protein-protein aggregates [88]. Although all amino residues are subject to oxidative modifications, some are more susceptible due to special groups present, such as sulphydryl (-SH) containing

amino acids like cysteine, methionine, tryptophan, histidine, tyrosine and phenylalanine. These –SH groups when oxidized can convert to forms such as sulfenic (SOH), sulfinic (SOOH) and sulfonic (SOOOH) acid formation [88]. Other amino acid modifications include radiation and metal catalyzed induced formation of carbonyl groups (C=O), most of which lead to altered protein structure and increase in modified protein levels. Earlier studies have shown that while mildly oxidized proteins are easily degraded by the proteasome, prolonged oxidation of proteins can convert them into forms that are beyond repair and inhibit proteasome ability to degrade other proteins [12]. In spite of several mechanisms known for protein oxidation and modifications, most of the studies in the past have been conducted in *in vitro* conditions which have stimulated a recent swicth to *in vivo* models, due to to its indisputable relevance. An interest in oxidative inactivation of proteins has not only broadened perspective of its role in progression of several diseases but have lead to development of several techniques including Western blot in the past [89]. Several pathological conditions such as diabetes, cirrhosis, uremia, cardivascular conditions and inflammation have shown increased accumulation of oxidized proteins and its damaging effects over time [12]

Since all proteins are not equally sensitive to oxidative modification, the ones that are damaged are susceptible, due to their protein structure, certain residues exposed on their molecular surface and /or presence of certain groups such as –SH or metal bound atoms, as mentioned earlier. Therefore it is critical to identify proteins that are most affected by stress conditions to not only elucidate the exact mechanism of oxidative damages but in addition estimate any functional consequences that may be occurring downstream.

All protein spots that were significantly changing in their expression levels (FDRp< 0.05) (Table 3) were excised and trypsin digested before being analyzed with mass spectrometry. Protein digests were initially put through a MALDI-TOF/TOF(AB) instrument for identification purposes; representative MALDI-TOF/TOF data of an experimentally identified protein spot and database search result are shown in Fig. 19. A GPS protein confidence score of > 40 and a MASCOT score of > 60 (except for TIF51A) was reported for protein identifications from MADLI setup. Spots that were not identified with MALDI were then put through a LC/MS-MS setup (Q-TOF) (AB) to achieve

maximum identifications for the list of spots of interest. An unused score of =>2 (1005 confidence) in protein identifications, was reported for the nano-LC/MS-MS setup. Many protein spots yielded one confident identification but there were some spots that identified more than one protein for a given spot: 6416,4512,8512,6605 and 6703, indicating co migration of proteins (as different isoforms of the same protein, or multiple unrelated proteins), in the first and second dimension. A comprehensive list of all identified proteins with a short description, their accession numbers, scoring mechanism, experimental pI's and molecular weights (Mr) and with their spot annotations are provided in Table 4.

Table 3: List of 2D resolved proteins identified by in-gel trypsination and sequencing of generated peptides. Data on the encoded proteins come from the SGD (Saccharomyces Genome Database). Proteins are presented in alphabetical order of their annotation.

Master Spot No. (wtt3chp)	Annotation	Protein	Description	Method of identification	Accession No.	Theoretical Mr[a](kDa/pI)	Scoring
1016[b]	VS_2D_P1_186785000_04800	TIF 51 A	Translation initiation factor eIF-5A	MALDI/TOF-TOF	gi|603645	17.1/4.64	20*
1105[b]	VS_2D_P1_198852000_04560	RPN 13	Subunit of the 19S regulatory particle of the 26S proteasome lid	LC/MS-MS	gi|6323453	17.9/4.28	**2****
5116[c]	VS_2D_P1_207000000_05900	TDH2	Glyceraldehyde-3-phosphate DH, isozyme 2	LC/MS-MS	gi|6322468	35.8/6.96	**2.27****
1118[b c]	VS_2D_P1_209031000_04830	AHP1	Alkyl hydroperoxide	MALDI/TOF-TOF	gi|6323138	19.1/4.87	99.9*
8116[b]	VS_2D_P1_209730000_01000	-	-	-	-	-	-
6102[b]	VS_2D_P1_240055000_06080	PRE8	20S proteasome beta-type subunit	LC/MS-MS	gi|6323547	27.1/5/5	**2****
2103[b c]	VS_2D_P1_248339000_04930	TSA1	Thioredoxin peroxidase	MALDI/TOF-TOF	gi|6323613	21.5/4.87	99*
2112[b c]	VS_2D_P1_249878000_05140	TSA1	Thioredoxin peroxidase	MALDI/TOF-TOF	gi|6323613	21.5/4.87	99*
2110[b c]	VS_2D_P1_250716000_05040	TSA1	Thioredoxin peroxidase	MALDI/TOF-TOF	gi|6323613	21.5/4.87	99*
5204[b]	VS_2D_P1_280477000_05780	-	-	-	-	-	-
4221[c]	VS_2D_P1_283141000_05720	TPI1	Chain A; Structure of Triose phosphate isomerase	MALDI/TOF-TOF	gi|28374000	26.7/5.86	**656**[#]
2203[b c]	VS_2D_P1_293546000_04970	EGD2	Alpha subunit of the heteromeric nascent polypeptide-associated complex (NAC) involved in protein sorting and translocation	MALDI/TOF-TOF	gi|6321987	18.7/4.67	39[#]
3211[b c]	VS_2D_P1_305391000_05360	-	-	-	-	-	-
3209[b c]	VS_2D_P1_308730000_05360	HPT1	Dimeric hypoxanthine-guanine phosphoribosyltransferase	MALDI/TOF-TOF	gi|6320607	25.1/5.48	99*
2312[b]	VS_2D_P1_317998000_05170	CAP1	Alpha subunit of the capping protein (CP) heterodimer	LC/MS-MS	gi|6322845	30.6/5.06	**2****
3603[b c]	VS_2D_P1_327007000_08340	-	-	-	-	-	-
4324[b]	VS_2D_P1_327640000_05700	-	-	-	-	-	-
4309[b c]	VS_2D_P1_332611000_05700	EAF3	Esa1p-associated factor, nonessential component of the NuA4 acetyltransferase complex	LC/MS-MS	gi|6325280	45.2/8.43	**1.4****

Master Spot No. (wtt3chp)	Annotation	Protein	Description	Method of identification	Accession No.	Theoretical Mr[a](kDa/pI)	Scoring
5301[b]	VS_2D_P1_336275000_05750	TDH1	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1	LC/MS-MS	gi|6322409	35.7/8.59	**2****
5406	VS_2D_P1_355108000_05790	-	-	-	-	-	-
7804[b c]	VS_2D_P1_358440000_05670	-	-	-	-	-	-
6411[b]	VS_2D_P1_358891000_06690	TDH 1/2/3	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1/2/3[b]	MALDI/TOF-TOF	gi|6322409/gi|6322468 /gi|632163	-;-;35.7/6.96	61*/100*/387[#]
2402[b]	VS_2D_P1_369586000_04930	-	-	-	-	-	-
1507[b]	VS_2D_P1_410349000_04550	PEP4	Vacuolar aspartyl protease (proteinase A)	LC/MS-MS	gi|6325103	44.4/4.54	**10****
6526[b c]	VS_2D_P1_413809000_06920	TKL1	Transketolase 1	MALDI/TOF-TOF	gi|6325331	73.8/7.01	47.4*
4512	VS_2D_P1_414590000_05700	IMD 2/3/4	Inosine monophosphate dehydrogenase 2/3/4	MALDI/TOF-TOF	gi|6322012 / gi|6323464/ gi|6323585	56.5/8.59/; 56.5/7.38;	94*/100*/77*
8512	VS_2D_P1_430638000_01000	YBR026C/ETR1	2-enoyl thioester reductase	LC/MS-MS	gi|45269635	42.0/9/78	**1.94****
7512[b]	VS_2D_P1_440102000_01000	AAT2	Cytosolic aspartate aminotransferase,	LC/MS-MS	gi|37362677	46.0/8.5	**3.11****
6602[b c]	VS_2D_P1_467567000_06150	NPT1	Nicotinate phosphoribosyltransferase	LC/MS-MS	gi|6324783	49.0/6.92	**2.56****
2612[c]	VS_2D_P1_469722000_05100	SAM1	S-adenosylmethionine synthetase	LC/MS-MS	gi|6323209	41.8/4.9	**4.01****
6605[b c]	VS_2D_P1_473144000_06420	ERG13; PGK	3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; Phosphoglycerate Kinase	MALDI/TOF-TOF	gi|6323509/ gi|209595	55/8.23; 44.7/7.7	112[#]; 60[#]
3611[c]	VS_2D_P1_492318000_05390	KGD2	Dihydrolipoyl transsuccinylase	MALDI/TOF_TOF; LC/MS-MS	gi|6320352/ gi|171783	50.4/9.62	91.8*; **4.01****]
6703[b]	VS_2D_P1_544264000_06160	CYS4; AAT2	Cystathionine beta synthase; Alpha subunit of the F1 sector of mitochondrial F1F0 ATP synthase	LC/MS-MS	gi|88910013/ gi|6319370	56.0/6.6; 46.0/8.5	**25.57**3.1****
2102	VS_2D_P1_211318000_04930	AHP1	Alkyl hydroperoxide	MALDI/TOF-TOF	gi|6323138	19.1/4.87	97.48*

[a] Values taken from SGD; [b] Protein spots significant for quantile; [c] Protein spots significant for TVSC; * GPS Protein Confidence Interval score (MALDI); **Unused score of Paragon algorithm(LC/MS_MS) ; [#] MASCOT score (MALDI)

D1 Peroxiredoxins

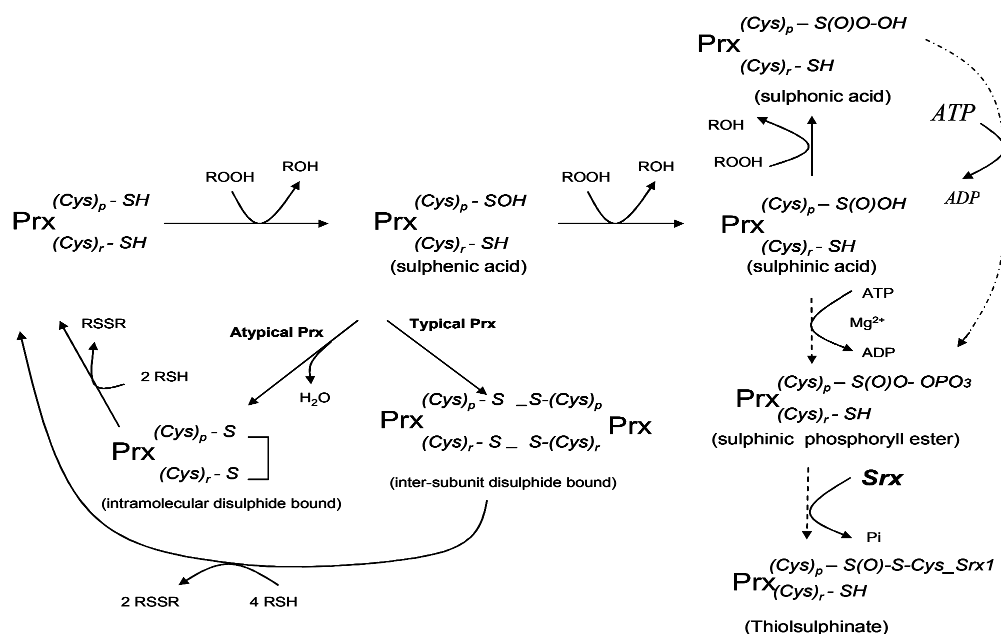
Peroxiredoxins, are a family of ubiquitous redox-active proteins, present in all organisms ranging from bacteria to humans, including anaerobic organisms [85]. Known for its reactive nitrogen and oxygen species scavenging properties, peroxiredoxins are characterized by a highly conserved cysteine (cys) residue that has a peroxidatic center (C_p) and one or two additional cys residues [85]. Resolving cysteine is a C terminal cysteine that is not conserved in all subfamilies of Prx but if present it can form an intra or inter molecular bond with the oxidized peroxidatic Cys. The peroxidase activity of the native protein (reduced form) is maintained by thioredoxins and/or glutaredoxins, using NADPH as reducing power. Use of glutathione (GSH) as a physiological reductant is still under debate but along with cyclophilin, glutathione-S-transferase and/or ascorbate it has been proposed to act as reductant for various categories of peroxiredoxns across different organisms [85]. Some peroxiredoxins can form oligomeric high molecular weight forms that have been described to display chaperone activity [90]. Based on the catalytic mechanism of Cys residues (Cp: peroxidatic and Cr: resolving) and the type of disulfide bond formed, peroxiredoxins can be categorized as: (1) typical 2-Cys Prx, where the peroxide-oxidized Cp (-SOH) reacts with a thiol group of a resolving Cys from another Prx to form an inter-subunit disulfide bond (Figure 1); (2) atypical 2-Cys Prx, where the oxidized Cp reacts with the Cr located within the same polypeptide to form an intramolecular disulfide bond (Figure 16); (3) 1-Cys Prx, which has only one N-terminal peroxidatic Cys with no resolving Cys present [85]. Cysteine residues of peroxiredoxins are capable of undergoing different oxidation states which has established its role as a regulatory component of stress response [87].

In yeast, five peroxiredoxins have been identified: the cytoplasmic forms Tsa1p (cTPxI), Tsa2p (cTPxII) and Ahp1p (cTPxIII), the mitochondrial Prx1p, and the nuclear Dot5p [91].

Oxidative stress induced by the aromatic peroxide cumene hydroperoxide (CHP) in yeast cells revealed, an interesting dynamics of

change occuring for the yeast peroxiredoxins Ahp1 and Tsa1, at the transcriptome and proteome levels.

Ahp1p is a thioredoxin peroxidase acting on alkyl hydroperoxides, such as *tert*-butyl hydroperoxide (t-BOOH) [92]. Ahp1p is a typical peroxiredoxin possessing three Cys residues. A mechanistic scheme for antioxidant role of Ahp1p was suggested to involve oxidation of the catalytic (peroxidatic) Cys^{62}, followed by a disulfide formation between Cys^{120} of the other monomer [93]. This disulfide can then be reduced by the thioredoxin/thioredoxin reductase (Trx/Trr) [92].Tsa1p is another typical peroxiredoxin, containing two Cys residues (Cys^{47} and Cys^{171}), involved in the yeast response to H_2O_2 [94] .



.

Figure 16: Hypothetical model of cysteine modifications of peroxiredoxins via Srx1

Dynamics of of Ahp1p and Tsa1p and its modified forms in response to oxidative stress

A comparative analysis of spot profiles between control and CHP-treated groups of gel images revealed 5 spots (fig. 17) for peroxiredoxin family of proteins, changing reproducibly and significantly (p<0.05) in CHP treatment (Table 3/ Fig. 18) group of gels. These spots were identified by PMF and MS/MS using MALDI –TOF mass spectrometry as different forms of alkyl

hydroperoxide reductase (Ahp1p) or thioredoxin peroxidase (Tsa1p), two of the five yeast peroxiredoxins (Table 4). The native proteins were located at a higher pH and the modified forms at lower pH. Despite different pH positions, spots 2102 and 1118 displayed similar mass profiles and were identified as Ahp1p (Fig 17). The identity of Ahp1^{red} and Ahp1^{ox} was ascertained by tandem mass spectrometry with sequence coverage of 45% and 75% respectively. Spot 1118 was found to be at acidic pI alongwith with another spot next to it (also identified as Ahp1p), indicating towards occurrence of protein modification.

Spots 2112, 2110 and 2103 with the same mass profile were identified as the three forms of Tsa1p. Proteins corresponding to spots 2103 and 2110 were more acidic and increased significantly with time (Fig. 17). The protein in

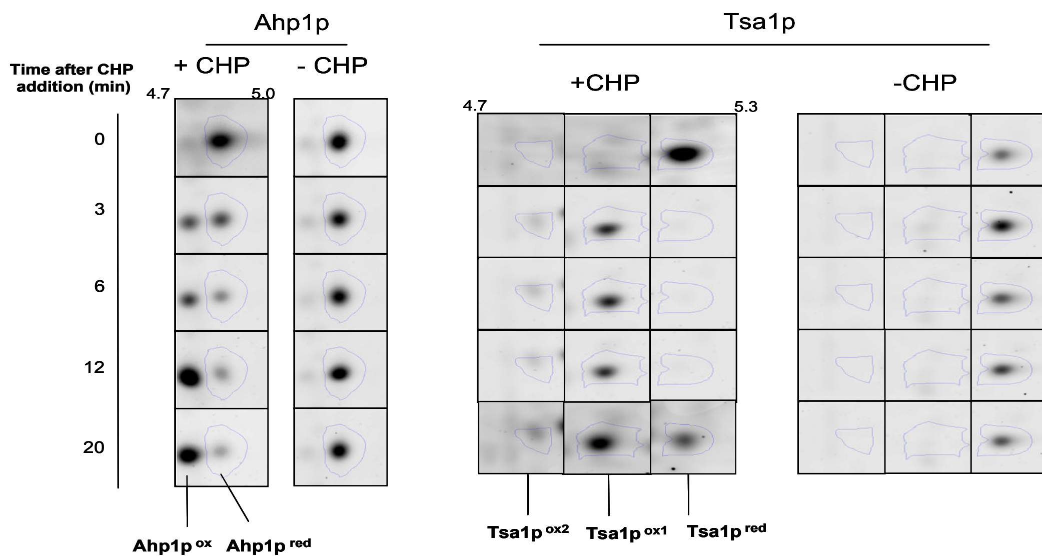


Figure 17: 2D-PAGE analysis of oxidized and reduced forms of Ahp1p and Tsa1p in yeast cells after exposure to CHP. CHP was added to the cultures immediately after sample collection at t=0 and a time-course was followed up to 20 min, as indicated. Results shown are one of the three biological replicates

spot 2112 decreased dramatically after t=0 and began re-appearing after 20 min. The protein in spot 2103 was observed for the first time in this study. A previous study where yeast cells were challenged with H_2O_2 did not detect this form of Tsa1p. Based on the pI profile and an increased level of expression observedover time, it is likely that proteins in spots 2110 and 2103 are oxidized forms of spot 2112, the reduced form of Tsa1p. This reduced form in spot 2112

seems to be quickly oxidized by CHP, as can also be observed in Figure 18, column B (note that in the control cultures – bold symbols in the same graph - there is no significant change in the levels of the reduced form of Tsa1p).

Expression of genes encoding yeast peroxiredoxins in CHP-induced stress

The transcript profiles for the two peroxiredoxins were analyzed in the same yeast samples used for protein extraction. Microarray data analysis revealed that AHP1 and TSA1 expressions did not change significantly ($p<0.05$) within 20 mins after addition of CHP to the cultures, although there is a clear pattern of increase in the later time points (12 and/or 20 min) (Figure 18, column A). This seems to indicate that the involvement of these proteins in the stress response occurs mainly at the post-transcriptional level, though some

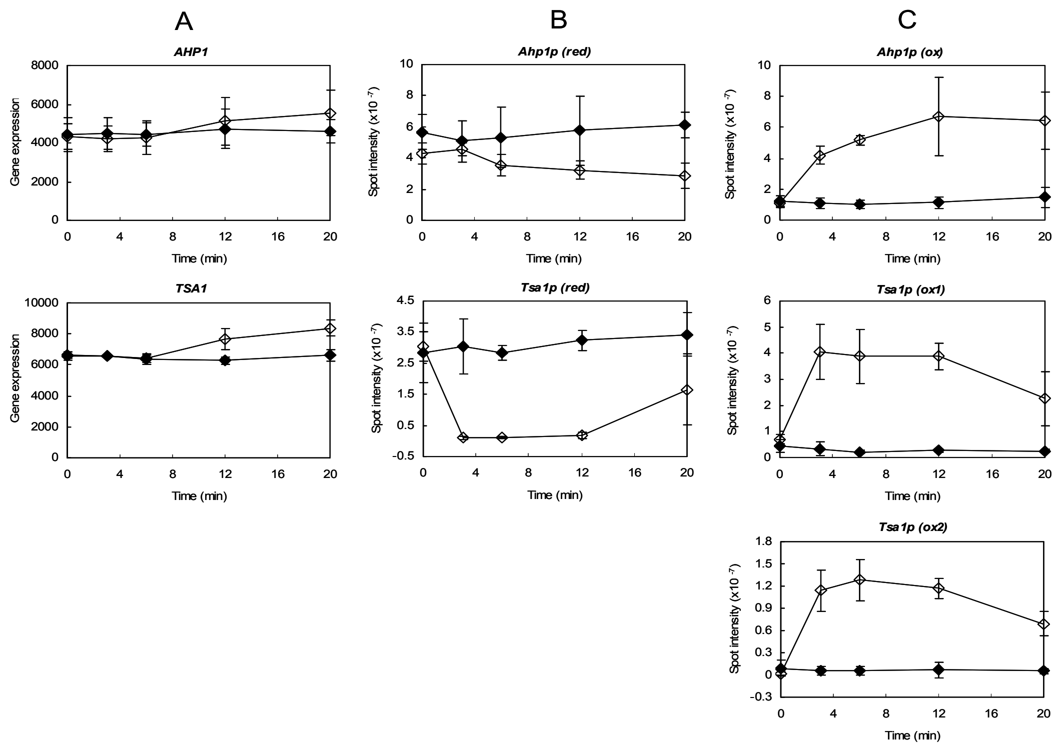


Figure 18: Dynamics of changes in the levels of Ahp1 and Tsa1 at the transcriptional (A) and proteome (B,C) levels. Bold symbols correspond to control (non-treated) cultures, while open symbols are used for CHP-treated cultures. Results shown are mean ± standard deviation of three biological replicates. Column (B) shows the time-course of changes for the oxidized forms of Ahp1p and Tsa1p, while (C) shows the same time-course for the oxidized forms (one Ahp1p^{ox}, and two Tsa1p^{ox}).

increase in the transcript levels may occur at later time points. Analysis of transcript profiles for the remaining three yeast peroxiredoxins, TSA2, PRX1

and DOT5 reveals that both TSA2 and PRX1 are significantly induced early (3 and/or 6 min), while the gene encoding the nuclear peroxiredoxin DOT5 is significantly up-regulated at 20 min. Data on protein expression profile for remaining peroxiredoxins: Tsa2p, Dot5p and Prx1p could not be obtained with this proteome study. Several possible explanation can be provides, i) since these three proteins are low abundant entities, use of a single stain 2D technique was not sufficient for its detection ii) Spot locations representative for these proteins were not present in the list of manually validated spots or iii) these proteins did not change significantly to show up in statistical analysis .

Discussion for Peroxredoxins:

The redox behavior of cysteine residues in proteins and peptides have gained rapid interest and have managed to change our perspective on the role of amino acids in redox sensing and cell signaling [87]. Different kinds of post-translational cysteine modifications that can occur are: thiyl radicals, disulfides, disulfide-S-oxides, sulfenic, sulfinic and sulfonic acids [87]. Oxidation of thiols have been suggested as a mechanism for oxidative stress (OS) sensing, with sulfinic acid being considered as an ideal sensor for OS, for quite some time. It is believed that extent of sulfinic acid formation is a good indicator of "total" oxidative stress for a given cell. Discovery of sulfiredoxin and its sulfinic acid reduction have not only established the role of both Prx and Srx in cell signaling and regulation, but have revealed interesting insights on cysteine residue redox mechanism (fig 1). The inability to directly reduce cysteine sulfinic acid by glutathione (*in vitro*) had suggested a phosphorylation step that serves as a good leaving moiety to be replaced by thiol(ate) [87]. Even though the importance of peroxiredoxins in controlling oxidative status of cells is widely accepted, the precise dynamic response of peroxiredoxin systems to CHP induced oxidative stress in yeast is not known. Using a proteomics approach, post translational modifications occurring for the cytosolic peroxiredxoins could be detected..

From our study, it appears that the lower isoelectric point (pI) form of Ahp1p is upregulated with CHP treatment of yeast cells indicating potential post-

translational modifications evident form mass spectrometry data (Fig 19). The oxidized form (acidic variant) increases significantly in intensity with time which is consistent with previous reports of Ahp1p inactivation correlated to its conversion to low pI form after t-BOOH treatment, but only *in vitro* [95]. Results show that during oxidative stress, Ahp1 is modified (addition of sulfonic acid group as seen in Fig .19) such that total protein conformation changes leading to separation of oxidized and native Ahp1p forms in the first dimension. Another interesting aspect observed for AHp1p, from two dimensional gel analyses, is the continuous presence of reduced Ahp1p observed suggesting de novo synthesis of the native, active Ahp1 protein.

TSA1 is one of the genes that is under Yap1 regulon, dependent upon Skn7 ability to co-operate with Yap1, and respond to oxidative stress. Most of the Yap1 regulated genes do not have YRE (Yap1 response element) [27], however, TSA1 is known to carry both Yap1- and Skn7- responsive element within a 200-bp proximal region of its promoter [96]. Tsa1 protein, generally known for its reducing properties and reacting with peroxides was observed as two spots in *C. albicans* hyphae, representing acidic and basic forms [97]. A proposed explanation was that conversion to hyphae form, occurs in parallel with Tsa1 modification into an oxidized form [97]. The oxidized (inactive) form of Tsa1 increased in comparison to its lower pI (reduced) form as a result of hyphal differentiation [97] In another study, characterization of Tsa1p in *C. albicans* revealed its indispensability for yeast-hyphae transition especially for cells growing under oxidative stress conditions (5mM H_2O_2). The increase in CaTsa1 was seen post-transcriptionally with protein being localized in the nuclei of hyphal cells [98]. Tsa1p was therefore designated as a moonlighting protein for its requirement in cell was composition of hypahl forms and peroxidase activity [99].

Detection of oxidized forms of Tsa1 as highly expressed proteins was interesting, as Tsa1 has been previously shown to be involved in yeast stress response [100]. Unlike Ahp1p, the reduced form of Tsa1p dramatically decreases in the first 4 minutes but reappears after 20 minutes. In parallel the two acidic forms increase upto 20 minutes and taper off around 20 minutes. Occurrence of sulfinic acid on proteins was earlier believed to be a random

event, until a proteomics approach by Aberosold group showed that presence of sulfinic acids is more controlled and widely seen *in vivo*. Initially the irreversible inactivation of cysteine residues by overoxidation was considered a dead end until Srx1 was identified as a repairing component, that would convert modified forms of Cys groups on proteins (eg TSa1) to their reduced forms [87]. The redox catalytic mechanism of peroxiredoxins is provided by the cysteine amino acid residue which can undergo different oxidation states, each transformation having a biological inference [87]. Our results are consistent with the effect of H_2O_2 (500uM) study observed on Tsa1p which demonstrated that the proportion of oxidized Tsa1 form increased at the expense of reduced Tsa1p and that the ratio of the reduced/oxidized Tsa1p had shifted back to what was observed in untreated yeast cells, after 30 minutes. Our results show similar dynamics, at qualitative level, for Tsa1p reduced/oxidized forms as early as 20 minutes in response to CHP treatment. Another interesting feature is the presence of a third variant observed for Tsa1p in CHP treated yeast cells, which has a very low spot intensity (Fig. 17). It would be interesting to explore both (acidic pI) variants of Tsa1p for known occurrence of post-translational modifications

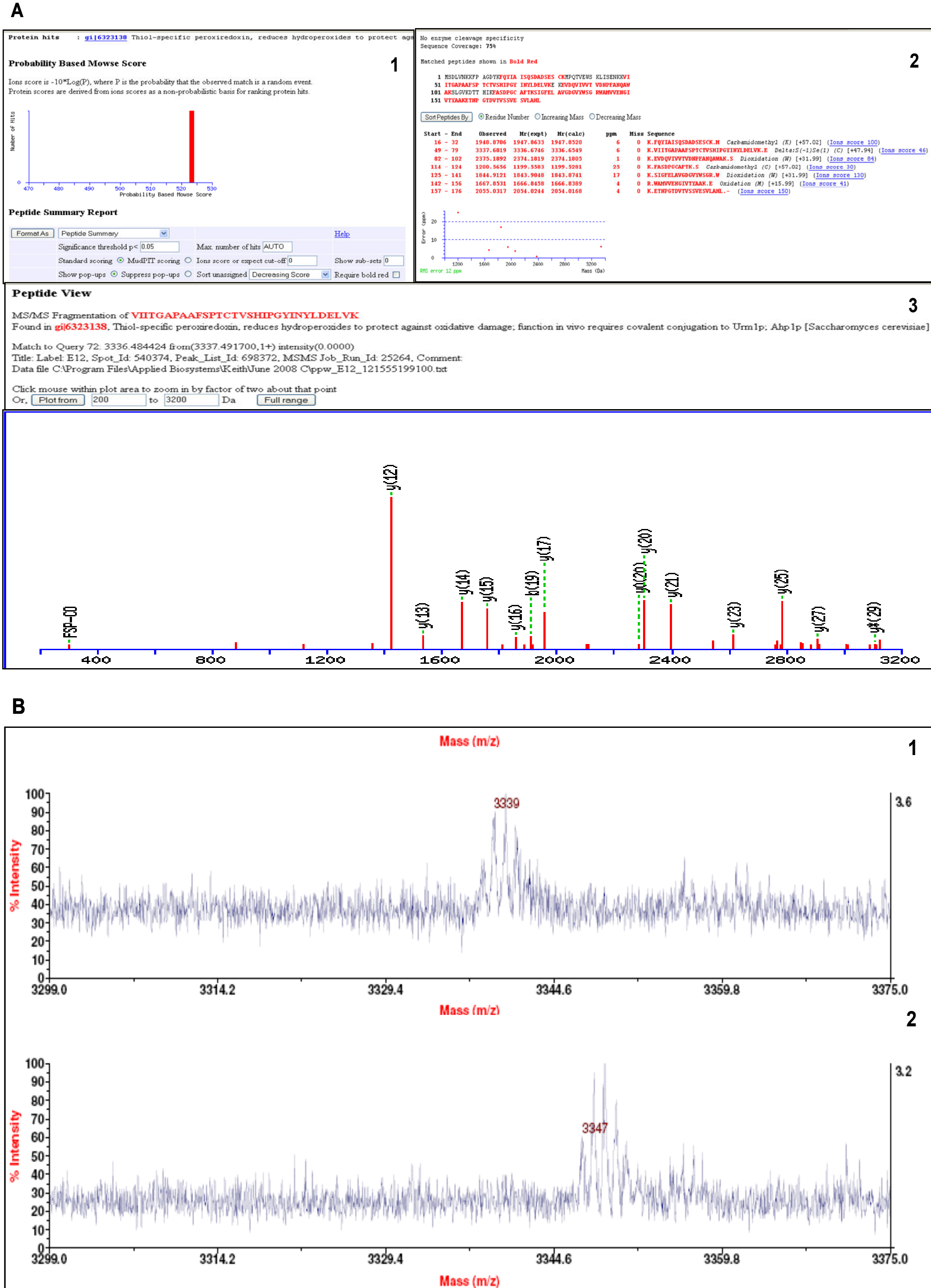


Figure 19: The MALDI-TOF mass spectra corresponding to A) AHP1 identification and B) two forms of Ahp1p are shown. A1) Database query result with significant identification made to Ahp1p; A2) Representative protein view with list of 8 peptides used towards Ahp1p identification; and A3) Representative product ion spectrum of *m/z* 3337±1 and its *b* and *y* ions. B1) Peak near 3338±1 corresponds to the peptide containing the active-site cysteine (modified with sulfonic acid) and peak near 3347 coressponds to alkylated cysteine (modified with iodoacetamide

D2 Other significant proteins:

TIF51A. The translation initiation factor eIF-51A,, also known as HYP2, is a highly conserved eukaryote protein, involved in the protein translation machinery [101]. eIF-51A is unique, as it is the only protein that contains the odd hypousine amino acid residue, suggesting an important role in cell metabolism signaling due to an existence of an unusual amino acid [101]. In yeast, eIF-51A is encoded by two homologous genes, TIF51A and TIF51B which are expressed under aerobic and anaerobic conditions, respectively. TIF51A gene give rise to two eIF-51A proteins (eIF-51Aa and eIF-51Ab) that are more acidic isoelectric variants, in comparison to TIF51B encoded proteins (more isolelectric basic variants). Previous studies have shown that both eIF-51Aa and eIF-51Ab separate distinctly at different isolelectric points [101].

Previous transcriptional and proteomic studies of lithium toxicity of yeast growing on galactose media, reveal downregulation of Tif51Ap as early as 40 minutes after addition of 10 mM LiCl. However, no significant change was observed at the TIF51A transcript level [33]. Likewise, protein expression studies of H_2O_2 (0.4mM) treated yeast have shown significant downregulation of Tif51Ap in treated yeast cells [1].

In our study, spot profile of 1016, identified as translation initiation factor eIF-5A revealed significnat increase in spot intensity, as early as 12 minutes after CHP treatment (Figure 20, column B). The transcript profiles for the TIF51A were analyzed in the same yeast samples used for protein extraction. Microarray data analysis revealed that TIF51A expression did not change significantly ($p<0.05$) within the 20 min after addition of CHP to the cultures. (Figure 20, column A).

Since no correlation could be determined between the transcript expression level and the total protein encoded, one plausible explanation for the differences observed for each level of data, is that spot 1016 could be a representative of the modified form of eiF-51A protein which seems to exclusively increase with CHP treatment, indicating towards post-translational regulation. However, an immunological approach would be needed in order to verify whether spot 1016 includes eIF-51Aa and/or eIF-51Ab variant. The current spot location matches with the theoretical Mr and pI values (17.1kDa/ 4.64). Previous studies show that

eIF-51Aa is a reversible post translational phosphorylated form of eIF-51Ab [101]. Since phosphorylation has been documented as an important mechanism for regulating gene expression related to translation apparatus, it is likely that spot 1016 could be the phosphorylated form of eIF-51Ab that increases over time with the stress response or is an occurrence of a different modification of eIF-51A protein, with increased accumulation with time. Further experimental data would be needed to provide such evidence

TDH1: Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), isozyme 1 is a tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate and is thus involved in glycolysis and gluconeogenesis [102]. The protein has been detected in the cytoplasm as well as in cell-wall. GAPDH, is encoded by three unlinked structural genes TDH1, TDH2 and TDH3 which are non-identical proteins. H_2O_2 treatment of fission yeast revealed that Tdh1p is transiently oxidized at Cys-152, which, if point mutated, leads to disruption of Tdh1p interaction with Mcs4 response regulator [102]. This was the first time when the role of Tdh1p in phosphor-relay signaling was recognized in response to stress studies. Tdh1's involvement has been reported in redox imbalance conditions in context with reductive stress, ethanol stress and osmotic shock with an induction of TDH1 expression at the transcript level [103].

Spot profile of 5301, identified as Tdh1p (Fig. 22 A) did not show any increase in protein expression with time but depicted a significant treatment effect between the control and CHP treated samples (Figure 20, column B). The spot location was found at the same Mr level as its theoretical molecular weight, but the experimental pI between 5.4-5.8 was much lower than its theoretical pI of 8.59 indicating towards a modified form of Tdh1. The transcript profiles for the TDH1 were analyzed in the same yeast samples with two probes (1 and 2). TDH1-probe 1 revealed significant change ($p<0.05$) with increase at 12 min after addition of CHP to the cultures (Figure 20, column A). This indicates transcriptional regulation of TDH1. However, TDH1-probe 2 did not show any significant change until 20 min after CHP addition (data not shown). TDH1 was also detected in another significant spot (6411) along with TDH2/3 with high confidence score. Protein expression pattern based on this spot showed a treatment effect at 12 minutes

exclusively but because the control samples were not stable through out 20 minutes, no conclusions can be derived from spot 6411 expression pattern.

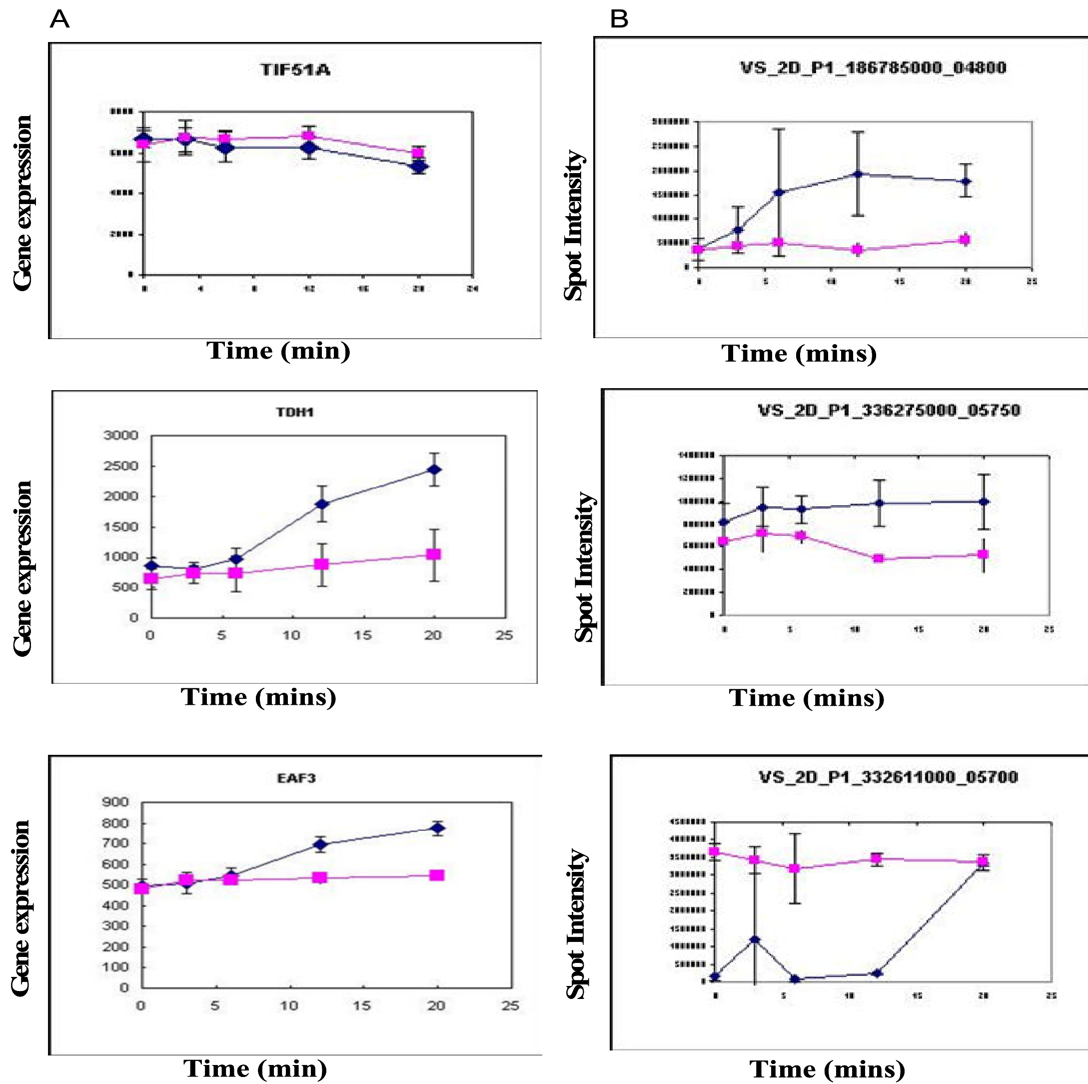


Figure 20: Dynamics of changes in the levels of TIF51A, TDH1 and EAF3 at the transcriptional (A) and proteome (B) level. Pink symbols correspond to control (non-treated) cultures, while blue diamond is used for CHP-treated cultures. Results shown are mean± standard deviation of three biological replicates

Even though it is difficult to determine the proportion of distribution of Tdh1p in between 5401 and 6411 spots due to the presence of highly similar protein isoforms, the role of Tdh1p in stress conditions cannot be ruled out.

EAF3: Is an Esa1p-associated factor which is also a component of the NuA4-histone acetyltransferase complex in yeast [104]. NuA4, is composed of 13 subunits, comprising of ESA1 as the main platform for complex formation. EAF3 is one of the remaining subunits of NuA4 that interact directly with Esa1 along with Arp4 - in vitro and is also found to be non-essential for cell cycle progression or catalytic activity [104]. NuA4 is one of the major HAT (histone acetyltransferases) complexes that acetylate histone (H4 specifically). Studies with eaf3Δ mutants have demonstrated altered genomic profiles of histone acetylayion, with increased acetylation observed at the coding region of H4 histones, which is reversed in WT strain [105]. Since post-translational modifications, such as phosphorylation, methylation and acetylation, of conserved amino acids in histone tail are known to modulate interaction potential of the tail domains, and hence influence folding and functional state of the chromatin fiber, role of Eaf3 in chromatin restructuring and functioning is clearly evident [105]. .

Both, protein profile of spot 4309 expression - identified as Eaf3 (Figure 20B and Fig22B), and gene expression (Figure 20, column A) increased significantly after 12 minutes of CHP addition. With no way of detrmining if spot 4309 is representative of total native Eaf3 protein or a modified form, an exact role of EAf3 protein in CHP induced stress study cannot be concluded. Nevertheless, given the fact that Eaf3 protein is being regulated at some level by stress conditions, and has the ability to influence folding of nucleosomal fiber, its possible role in oxidative stress needs further studies.

IMD2/3/4: Inosine monophosphate dehydrogenase, is an enzyme that catalyzes the first step of *de novo* guanine nucleotide synthesis [106]. Yeast has four homologs of IMPDH identified : namely *IMD1*, *IMD2*, *IMD3,* and *IMD4* genes [106]. Previous studies demonstrate that while IMD3 and IMD4 are constitutively expressed, IMD2 expression is repressed under limited nutrient conditions and that IMD1 is a non-functional protein [106]. Previous studies show that yeast IMD proteins share more than 80% amino acid identity with each other and 60% amino acid identity with human IMPDH's type I and type II proteins. Studies with extracellular guanine have shown that guanine represses IMD genes, encoding the yeast homologs of IMP dehydrogenase which is important as a feedback loop mechanism [106]. Human studies have shown conformational changes occurring

for IMD proteins, as a result of IMDH inhibition [107]. IMDH has been recognized to be critical for growth of different cell types including rapidly proliferating cells as well as lymphocytes. It has been an important regulator of cell growth and has thus been a potential target for immunosuppressive agents [107].

Protein profile of spot 4512 identified detection of three IMD proteins on one spot with protein confidence score of 94%, 100% and 77 % respectively. This could be due to the fact that either spot 4512 has all three proteins forms occurring or there could be only one protein present but since yeast IMD homologs share high percentage of amino acid similarity, other IMD proteins scored high on the protein identification part due to similar shared sequence. Microarary analysis reveals significant decrease in IMD3 and IMD4 gene expression in between 12-20 minutes, with no significant change observed for IMD2 gene. Going with the above possible scenario for protein existence, spot

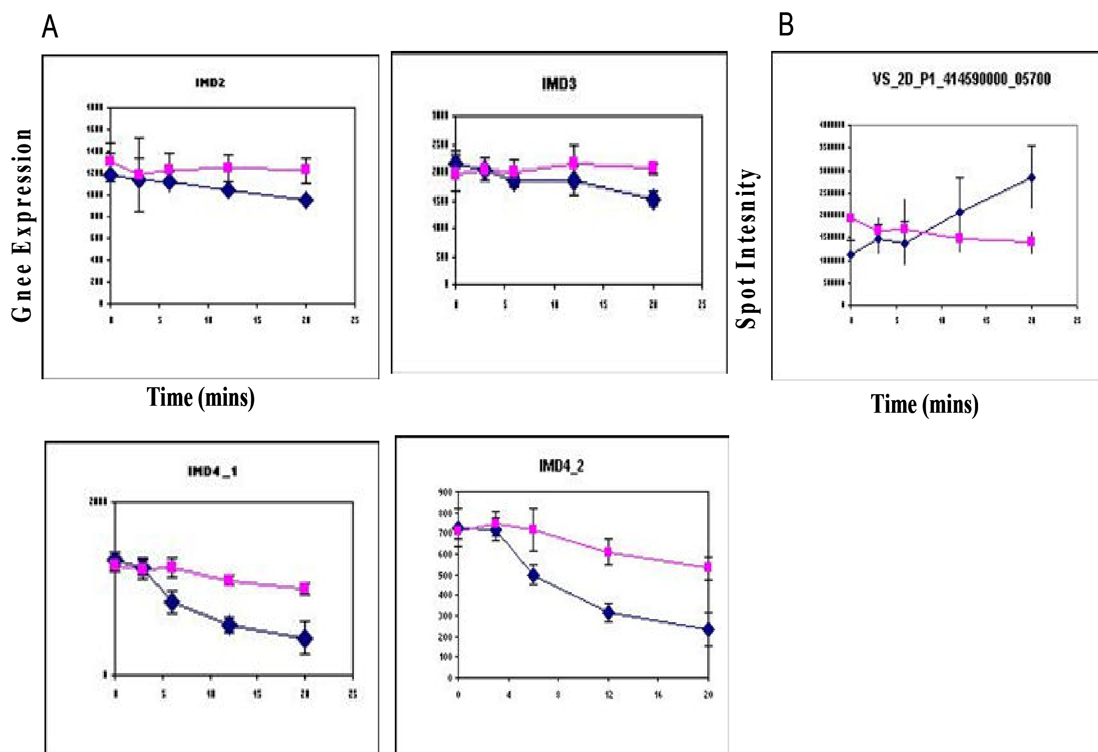


Figure 21: Dynamics of changes in the levels of IMD2/3/4 at the transcriptional (A) and proteome (B) level. Pink symbols correspond to control (non-treated) cultures, while blue diamond is used for CHP-treated cultures. Results shown are mean± standard deviation of three biological replicates.

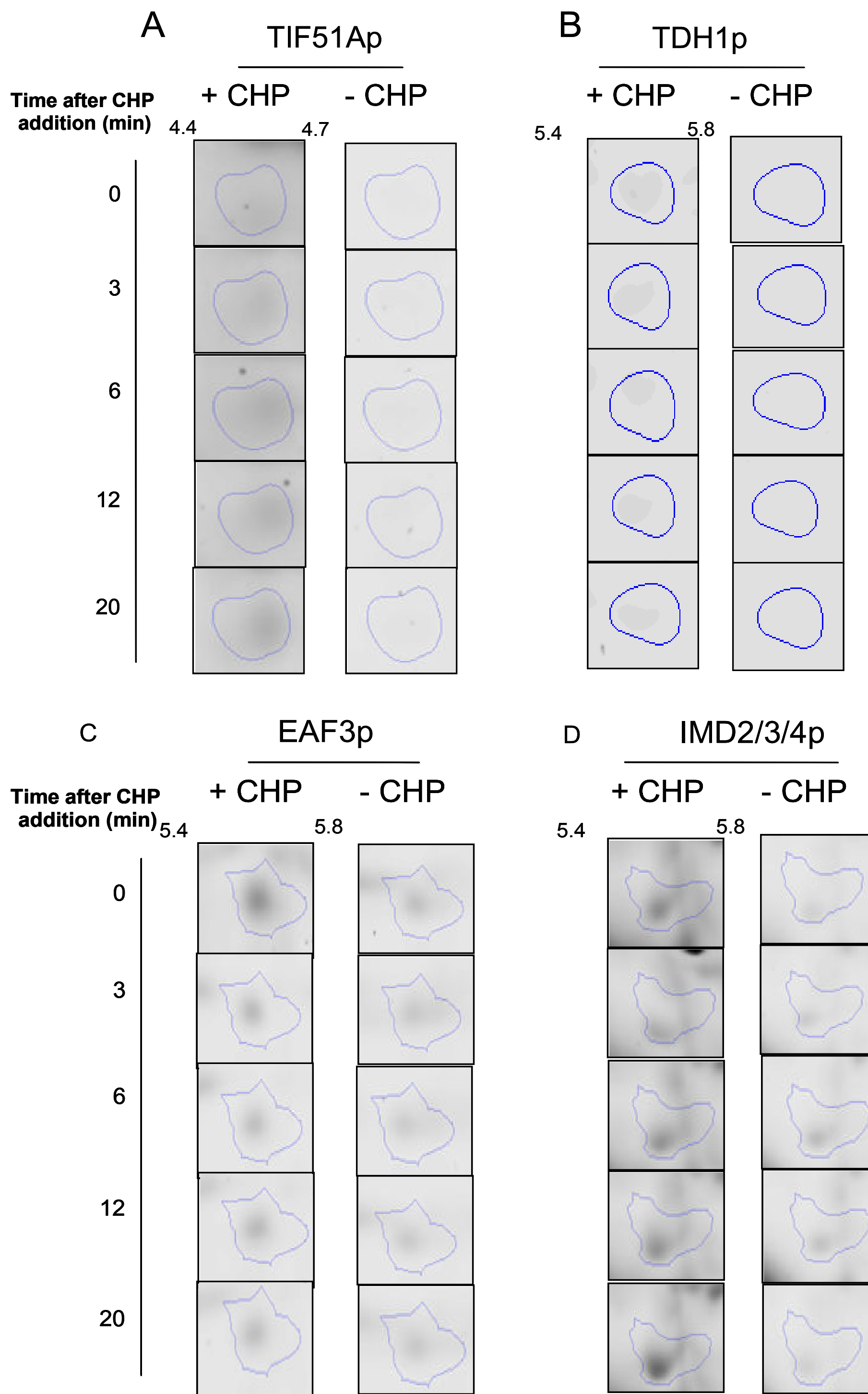


Figure 22: Spot profile of TIF51A (A) ; TDH1 (B); EAF3 (C) and IMD2/3/4 (D) in yeast cells in both control vs CHP treatment groups. CHP was added to the cultures imeediately aftre sample collection at t=0 and a time-course was followed up to 20 min, as indicated. Results shown are one of the three biological replicates.

4512 could be a representative of modified IMD2 protein that increase with CHP treatment over time without any changes observed for total protein production OR due to stress induced conditions a more complex regulation of IMD proteins could be occurring

D3 Conclusions

In an attempt to obtain an a holistic view of the yeast oxidative stress response, proteomics data was generated using two dimensional gel electrophoresis (2DE) technique, that would also be used towards development of mathematical tools for model building,. RIPA buffer followed by protein assay with a non-interfering kit (Calblochem) allowed the best global representation of the yeast proteome and a reliable estimate of total protein concentration, under given set of experimental conditions. Beads were effective in cell lyzing without losing too much protein at the solubilization step. A 24 cm gel, 12 % acrylamide gel with a pH range of 3-10 was efficient in resolving most proteins in comparison to 18 cm gels. Fluorescent stain, Flamingo Pink was used for detecting proteins spots on the gel due to its high sensitivity (as low as 1 ng), broad linear range, low background and mass spectrometry compatibility.

PDQuest was efficient in analyzing gel images generated for the yeast project but with some limitations. All gels including sample and reference gel were cropped to the same size (267.7mm x 222.7mm), to avoid introducing any technical variations. Spot detection parameters were kept same across all sample gels. A reference gel was generated with spot annotation for the *wildtype* strain (yeast) of experiments, to facilitate sample gel comparisons. Due to the inability of the software to handle more than 50 gels at a given time, image analysis in PDQuest was carried out such that each matchset included a reference gel along with the three biological replicates for that time point (0-20 min) and group (CTR vs CHP). All sample gels were made to warp to the reference gel using landmark protein spots. Automated spot matching generated some mismatch errors which were corrected and validated with manual matching of ~ 406 spots.

Quantile normalization was efficient in removing any experimental variations for the current 2DE based yeast protein expression data where all gels were made to have identical intensity distributions and lined up at the same median before statistical analysis. A 2- way ANOVA generated a list of 26 annotations that came out significant for time and/or treatment and/ or interaction effect.

Out of the 26 significant protein spots (for quantile method), based on FDR corrected p value <0.05, for a time course of 20 minutes between control and CHP treated groups, 16 proteins were identified with high confidence using MALDI-TOF/TOF and/or nano –LC/MS-MS analysis. Some of the identified proteins were discussed including peroxiredoxins, eTIF51a, EAF3, IMD2/3/4 and TDH1 that depicted transcriptional and/ or post-transcriptional regulations, as mentioned earlier.

Peroxiredoxins including Ahp1p and Tsa1p were found to be post translationally regulated. With increasing CHP induced stress conditions, Cys (62) residue of Ahp1 protein was found to be oxidatively modified to a sulfonic acid group (SOH) (confirmed by mass spectrometry). The modification of Ahp1p peroxidatic Cys residue observed in only treated groups, indicates that Ahp1 is involved in oxidative stress response and that any regulation that may have occurred was at the protein level. Likewise, for Tsa1 protein two variants (acidic pI) in addition to its reduced form were observed which suggest post translational regulation but have nor been validated with mass spectrometry data yet. Other processes affected by CHP treatment include guanine nucleotide synthesis, nucleosomal protein complex folding and unfolding, protein translational machinery etc.

In spite of the amount of time taken to profile hundreds of proteins separated on a 2D gel and considering the reproducibility related issues, 2DE provided the ability to visualize protein isoforms using isoelectric point and molecular weight parameters for a given set of proteins The recent introduction of Flamingo Pink as a fluorescent stain, has greatly improved detection of low amount proteins (nano grams) on a 2D gel., though it may be possible that some

of these spots had protein amounts well below some Mass Spec instrument identification threshold. Nano-LC/MSM-MS sensitivity was needed to identify proteins that could not be identified with MALDI-TOF.

CHAPTER 5

E. FUTURE STUDIES

Since the scientific direction of a study is pretty much driven by the approach used, with the continuing development observed in recent technologies, one may use whichever technology is available to acquire necessary information, provided one knows what to look for.

The proteomics approach that we used for the current project provides information on which proteins are affected and their possible roles in oxidative stress response. Studies were conducted at different levels (transcript, protein, metabolite) and data derived will be eventually integrated in order to acquire a comprehensive view of the stress response mechanism. However, just based on comparative analysis between transcriptomic and proteomic studies especially in case of yeast peroxiredxoins (in this project) we found that these redox-active proteins and their modified forms have an active role to play in redox regulated processes (to CHP induced stress). However, with many possibilities of protein modifications that may occur, modification dependency and it defined role is not well understood.

The current data provides information on which proteins are involved and at what level (transcriptional or post translational) are they regulated. Future studies with mass spectrometry based validation should clarify which protein modifications occur and their exact mechanism of regulation. It will be interesting to address whether other proteins identified (for this project) display a similar modification pattern as observed for peroxiredoxins. Previous studies focusing upon the redox modifications for some of these proteins identified can indicate towards some possible modifications that could have occurred for CHP induced stress response. For specific modifications such as oxidation, one can use flourophores tags that are specific for oxidized cysteines or use a different enzyme altogether for protein digestion eg. AspN enzyme cuts off at oxidized cysteines, which can be used in conjunction with trypsin.

For targeted protein profiling, where the researcher has an interest in specific proteins (based on microarray data) that cannot be detected by the use of a single stain experiment, for example Tsa2p and Prx1p an immunological approach can be used.

In this project, proteomic studies were also conducted on gpx1Δ and yap1 Δ strains, under the same experimental conditions (CHP) that have generated gel images available for data analysis. Image and statistical analysis of expression data from these strains would reveal possible involvement of YAP1 and GPX1 in influencing protein expression in response to above conditions. A comparison of proteins data between wt, gpx1Δ and yap1 Δ strains would reveal any differences that exist in stress response, due to the attributed mutations.

F. APPENDICES

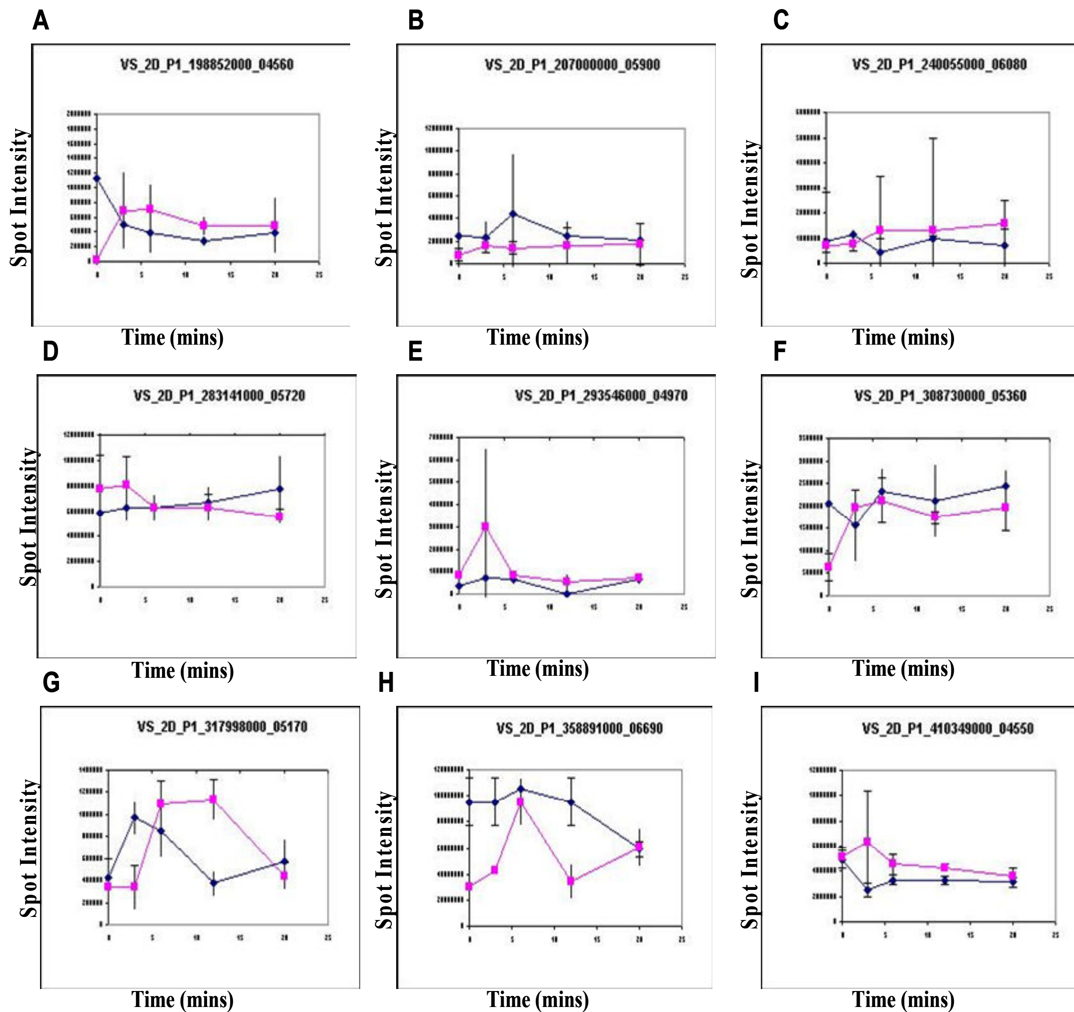


Figure 23: Time series of significant protein spots (pFDR <0.05). CHP sample is labeled in blue and the control samples are labeled in blue. Proteins identified for each spot include : A) RPN 13, B) TDH2, C) PRE8, D)TPI1, E)EGD2, F)HPT1, G)CAP1, H) TDH1/2/3 and I)PEP4

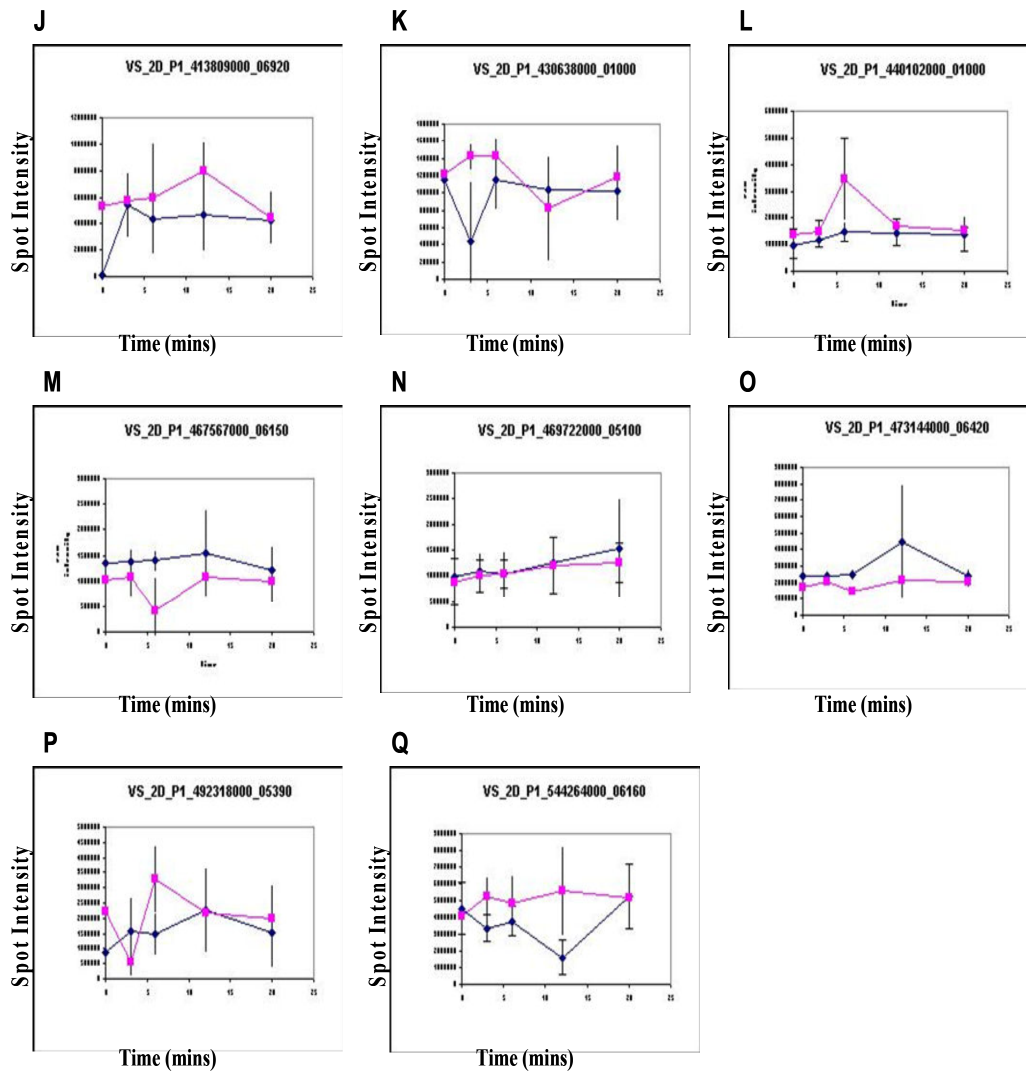


Figure 24: Time series of significant protein spots (pFDR <0.05). CHP sample is labeled in blue and the control samples are labeled in blue (continued). Proteins identified for each spot include: J) TKL1, K) YBR026C (ETR1), L)AAT2, M)NPT1, N)SAM1, O)ERG13/ PGK1, P)KGD2 and Q) CYS4 AND AAT2.

G. REFERENCES

1. Godon, C., Lagniel, G., Lee, J.,Buhler, J. M.,Kieffer, S.,Perrot, M.,Boucherie, H.,Toledano, M. B. and J. Labarre, J., *The H2O2 stimulon in Saccharomyces cerevisiae.* J Biol Chem, 1998. **273**(35): p. 22480-9.

2. Apel, K. and H. Hirt, *Reactive oxygen species: metabolism, oxidative stress, and signal transduction.* Annu Rev Plant Biol, 2004. **55**: p. 373-99.

3. Scandalios, J.G., *Oxidative stress responses--what have genome-scale studies taught us?* Genome Biol, 2002. **3**(7): p. REVIEWS1019.

4. Butterfield, D.A., Perluigi, M. and R. Sultana, *Oxidative stress in Alzheimer's disease brain: new insights from redox proteomics.* Eur J Pharmacol, 2006. **545**(1): p. 39-50.

5. Paul G. Winyard, C.J.M.a.C.J., *Oxidative activation of antioxidant defence.* Trends in Biochemical Sciences 2005. **30**(8): p. 453-461.

6. Michel B. Toledano, A.D., Benoit Biteau,Daniel Spector,Dulce Azevedo, *Oxidative stress responses in yeast*. Topics In Current Genomics. 2003. 241-303.

7. Inze, D.a.M.V.M., *Oxidative Stress in Plants.* 2002: p. 1-321.

8. Cabiscol, E., Tamarit, J. and J. Ros, *Oxidative stress in bacteria and protein damage by reactive oxygen species.* Internatl Microbiol, 2000. **3**: p. 3-8.

9. Fedoroff, N., *Redox regulatory mechanisms in cellular stress responses.* Ann Bot (Lond), 2006. **98**(2): p. 289-300.

10. Jamnik, P.a.P.R., *Methods for Monitoring Oxidative Stress Response in Yeasts.* Journal of Biochemical Molecuar Toxicology, 2005. **19**(4): p. 195-203.

11. Isabella Dalle-Donne, R.R., Ceciliani, F. ,Giustarini, R.C.D. and A. Milzani, *Proteins As Sensitive Biomarkers Of Human Conditions Associated With Oxidative Stress.* Wiley Interscience, 2006: p. 487-525.

12. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D. and A. Milzani, *Biomarkers of oxidative damage in human disease.* Clin Chem, 2006. **52**(4): p. 601-23.

13. Butterfield, D.A., Poon, H. F., St Clair, D., Keller, J. N., Pierce, W. M., Klein, J. B. and W. R. Markesbery, *Redox proteomics identification of oxidatively modified hippocampal proteins in mild cognitive impairment: insights into the development of Alzheimer's disease.* Neurobiol Dis, 2006. **22**(2): p. 223-32.

14. Patrono, C.a.G.A.F., *Oxidant stress in cardiovascular disease: an emerging modality or a disproved theory?* Ital Heart J, 2001. **2**(12): p. 865-6.

15. Sahaf, B., Heydari, K. and L. A. Herzenberg, *The extracellular microenvironment plays a key role in regulating the redox status of cell surface proteins in HIV-infected subjects.* Arch Biochem Biophys, 2005. **434**(1): p. 26-32.

16. Winzeler, E.A.a.R.W.D., *Functional analysis of the yeast genome.* Current Opinion in Genetics & Development 1997. **7**: p. 771-776.

17. Berlett, B.S.a.E.R.S., *Protein oxidation in aging, disease, and oxidative stress.* J Biol Chem, 1997. **272**(33): p. 20313-6.

18. Ayala, A., Parrado, J., Bougria, M. and A. Machado, *Effect of oxidative stress, produced by cumene hydroperoxide, on the various steps of protein synthesis. Modifications of elongation factor-2.* J Biol Chem, 1996. **271**(38): p. 23105-10.

19. Mutoh, N., Kawabata, M. and S. Kitajima, *Effects of Four Oxidants, Menadione, 1-Chloro-2,4-Dinitrobenzene,Hydrogen Peroxide and Cumene Hydroperoxide, on Fission Yeast Schizosaccharmoyces pombe.* J. Biochem. , 2005. **138**: p. 797-804.

20. Sies, H.a.K.H.S., *Hydroperoxide-metabolizing systems in rat liver.* Eur J Biochem, 1975. **57**(2): p. 503-12.

21. Jamieson, D.J., *Oxidative stress responses of the yeast Saccharomyces cerevisiae.* Yeast, 1998. **14**(16): p. 1511-27.

22. Seis, H., ed. *Oxidative Stress:oxidants and antioxidants*. ed. S. H. 1991, Academic Press Limited.

23. Carmel-Harel, O.a.G.S., *Roles of the glutathione- and thioredoxin-dependent reduction systems in the Escherichia coli and saccharomyces cerevisiae responses to oxidative stress.* Annu Rev Microbiol, 2000. **54**: p. 439-61.

24. Collinson, E.J.a.C.M.G., *Role of yeast glutaredoxins as glutathione S-transferases.* J Biol Chem, 2003. **278**(25): p. 22492-7.

25. Rodriguez-Manzaneque, M.T.T., Jordi, B., Ros, G. and J. E. Herrero, *Grx5 Is a Mitochondrial Glutaredoxin Required for the Activity of Iron/Sulfur Enzymes*. 2002. p. 1109-1121.

26. Thorpe, G.W., Fong, C. S., Alic, N., Higgins, V. J., I. W. Dawes, *Cells have distinct mechanisms to maintain protection against different reactive oxygen species: Oxidative-stress-response genes.* PNAS 2004. **vol. 101**(No. 17): p. 6564-6569.

27. Temple, M.D., Perrone, G. G. and I. W. Dawes, , *Complex cellular responses to reactive oxygen species.* Trends Cell Biol, 2005. **15**(6): p. 319-26.

28. Demasi, A.P.D., Pereira, A. G. , Netto, L. E.S., *Yeast oxidative stress response influences of cytosolic thioredoxin peroxidase I and of the mitochondrial functional state.* FEBS Journal 2006. **273**: p. 805-816.

29. Lee, J., et al., *Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast.* J Biol Chem, 1999. **274**(23): p. 16040-6.

30. Krems, B., C. Charizanis, and K.D. Entian, *The response regulator-like protein Pos9/Skn7 of Saccharomyces cerevisiae is involved in oxidative stress resistance.* Curr Genet, 1996. **29**(4): p. 327-34.

31. Gemma Bellı, M.a.M.M., JoseGarcı´a-Martı´nez, Jose´ E. Pe´rez-Ortı´n and Enrique Herrero, *Saccharomyces Cerevisiae Glutaredoxin 5-deficient Cells Subjected to Continuous Oxidizing Conditions Are Affected in the Expression of Specific Sets of Genes.* The Journal of Biological Chemistry, 2004. **279**(13): p. 12386–12395.

32. Flattery-O'Brien, J.A.a.I.W.D., *Hydrogen Peroxide Causes RAD9-dependent Cell Cycle Arrest in G2 in Saccharomyces cerevisiae whereas Menadione Causes G1 Arrest Independent of RAD9 Function.* The Journal Of Biological Chemsitry, 1998. **273**(15): p. 8564-8571.

33. Bro, C., Regenberg, B., Lagniel, G., Labarre, J., N. J. Montero-Lomelı´, *Transcriptional, Proteomic, and Metabolic Responses to Lithium in Galactose-grown Yeast Cells.* The Journal Of Biological Chemistry, 2003. **Vol. 278**(34): p. 32141-32149.

34. Nazifalic, T.F., Temple, M.D.,Gloeckner, G.,Higgins, V.J., Briza, P. and I. W. Dawes, *Genome-wide transcriptional responses to a lipid hydroperoxide: adaptation occurs without induction of oxidant defenses* Free Radical Biology & Medicine, 2004. **37**(1): p. 23-35.

35. Griffin, T.J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L. and R. Aebersold, *Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae.* Mol Cell Proteomics, 2002. **1**(4): p. 323-33.

36. Vido, K., Spector, D., Lagniel, G., Lopez, S., Toledano, M. B. and J. Labarre, *A Proteome Analysis of the Cadmium Response in Saccharomyces cerevisiae.* Journal of Biological Chemistry, 2001. **276**(11): p. 8469-8474.

37. Hu, Y., Wang, G., Chen, G. Y., Fu, X. and S. Q. Yao, *Proteome analysis of Saccharomyces cerevisiae under metal stress by two-dimensional differential gel electrophoresis.* Electrophoresis 2003. **24**: p. 1458-1470.

38. Haurie, V., F. Sagliocco, and H. Boucherie, *Dissecting regulatory networks by means of two-dimensional gel electrophoresis: Application to the study of the diauxic shift in the yeast Saccharomyces cerevisiae.* Proteomics 2004. **4**: p. 364-373.

39. Kolkman, A., Daran-Lapujade, P., Fullaondo,A., Olsthoorn, M., Pronk, J. T., Slijper, M. and A. R. Heck, *Proteome analysis of yeast response to various nutrient limitations.* Molecular Systems Biology 2006.

40. Toledano, M.B., Delaunay, A., Monceau, L. and F.Tacnet *Microbial H2O2 sensors as archetypical redox signaling modules.* Trends in Biochemical Sciences, 2004. **29**(7): p. 351-357.

41. Smith, A., M.P. Ward, and S. Garrett, *Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation.* Embo J, 1998. **17**(13): p. 3556-64.

42. Ikner, A.a.K.S., *Yeast signaling pathways in the oxidative stress response.* Mutation Research, 2005. **569**: p. 13-27.

43. Pitzschke, A. and H. Hirt, *Mitogen-activated protein kinases and reactive oxygen species signaling in plants.* Plant Physiol, 2006. **141**(2): p. 351-6.

44. Mishra, N.S., R. Tuteja, and N. Tuteja, *Signaling through MAP kinase networks in plants.* Arch Biochem Biophys, 2006. **452**(1): p. 55-68.

45. Samaj, J., F. Baluska, and H. Hirt, *From signal to cell polarity: mitogen-activated protein kinases as sensors and effectors of cytoskeleton dynamicity.* J Exp Bot, 2004. **55**(395): p. 189-98.

46. Sweetlove, L.J., Heazlewood, J. L., Herald, V., Holtzapffel, R., Day, D. A., Leaver, C. J. and A. H. Millar, *The impact of oxidative stress on Arabidopsis mitochondria.* Plant J, 2002. **32**(6): p. 891-904.

47. Hood, L., *Systems biology: integrating technology, biology, and computation.* Mech Ageing Dev, 2003. **124**(1): p. 9-16.

48. S.G.Oliver, D.B.K.a., *Here is the evidence, now what is the hypothesis?The complementary roles of inductive and hypothesis-driven science in the post-genomic era.* BioEssays, 2003. **26**: p. 99-105.

49. Souchelnytskyi, S., *Bridging proteomics and systems biology: What are the roads to be traveled?* Proteome 2005. **5**: p. 4123-4137.

50. Rabilloud, T., Heller, M., Gasnier, F., Luche, S., Rey, C., Aebersold, R., Benahmed, M., Louisot, P. and J. Lunardi, *Proteomics analysis of cellular*

response to oxidative stress. Evidence for in vivo overoxidation of peroxiredoxins at their active site. J Biol Chem, 2002. **277**(22): p. 19396-401.

51. Chambers, G., Lawrie, L., Cash, P. and G. I. Murray, *Proteomics:a new approach to the study of disease.* Journal of Pathology, 2000. **192**: p. 280-288.

52. Graves, P.R., T. A. J. Haystead, *Molecular Biologist's Guide to Proteomics.* Microbiology And Molceular Biology Reviews, 2002. **66**(1): p. 39-63.

53. Palcy, S. and E. Chevet, *Integrating forward and reverse proteomics to unravel protein function.* Proteomics, 2006. **6**(20): p. 5467-80.

54. Wells, D.A., D. A. Weil, *Directions in Automated Sample Preparation of Proteins.* Pharma Genomics, 2003.

55. Yarmush, M.A.a.A.J., *Advances in proteomic technologies.* Annu. Rev. Biomed eng., 2002. **4**: p. 347-372.

56. Lucia Monteoliva, J.P.A., *Differential proteomics:An overview of gel and nongel based approaches.* Breifings In Functional Genomics And Proteomics, 2004. **3**(3): p. 220-239.

57. Gulcicek, E.E., Colangelo, C. M., McMurray, W., Stone, K., Williams, K., Wu, T., Zhao, H., Spratt, H., Kurosky, A. and B. Wu, *Proteomics and the analysis of proteomic data: an overview of current protein-profiling technologies.* Curr Protoc Bioinformatics, 2005. **Chapter 13**: p. Unit 13 1.

58. Ruelle, V., Falisse-Poirrier, N., Elmoualij, B., Zorzi, D., Pierard, O., Heinen, E., De Pauw, E., Zorzi, W., *An immuno-PF2D-MS/MS proteomic approach for bacterial antigenic characterization: to Bacillus and beyond.* J Proteome Res, 2007. **6**(6): p. 2168-75.

59. Gustafsson, J.S., Blomberg, A. and M. Rudemo, *Warping two-dimensional electrophoresis gel images to correct for geometric distortions of the spot pattern images to correct for geometric distortions of the spot pattern.* 2001.

60. Veenstra, T.D.a.J.R.Y., *Proteomics for Biological Discovery*, ed. J.W.a. Sons. 2006: John Wiley and Sons.

61. Roe, M.R.a.T.J.G., *Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes.* Proteomics, 2006. **6**(17): p. 4678-87.

62. Matthiesen, R., *Introduction to Proteomics*, in *Mass Spectrometry Data Analysis in Proteomics*. 2007, Humana Press.

63. Eng, J.K., *Tutorial on tandem mas spectrometry database searching*. notes, Wiley Interscience. p. 1-7.

64. Veenstra, T.D., *Automation in Proteomics.* Proteomics for Biological Discovery, 2006.

65. O'Malley, M.A.a.J.D., *Fundamental issues in systems biology.* BioEssays 2005. **27**: p. 1270-1276.

66. Mustacchi, R., S. Hohmann, and J. Nielsen, *Yeast systems biology to unravel the network of life.* Yeast, 2006. **23**(3): p. 227-38.

67. Vo, T.D., Greenberg, H. J. and B. O. Palsson, *Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data.* J Biol Chem, 2004. **279**(38): p. 39532-40.

68. Oltvai, Z.N. and A.L. Barabasi, *Systems biology. Life's complexity pyramid.* Science, 2002. **298**(5594): p. 763-4.

69. Hood, L. and R.M. Perlmutter, *The impact of systems approaches on biological problems in drug discovery.* Nat Biotechnol, 2004. **22**(10): p. 1215-7.

70. Weinkauf, M., W. Hiddemann, and M. Dreyling, *Sample pooling in 2-D gel electrophoresis: a new approach to reduce nonspecific expression background.* Electrophoresis, 2006. **27**(22): p. 4555-8.

71. Herve Thiellement, M.Z., Catherine Damerval, Valerie Mechin, *Plant Proteomics: Methods and Protocols*. Methods in Molecular Biology, ed. J. Walker. Vol. 355. 2007.

72. Nishihara, J.C. and K.M. Champion, *Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain.* Electrophoresis 2002. **23**: p. 2203-2215.

73. Lopez, M.F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., W. F. Patton, *A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling.* Electrophoresis, 2000. **21**(17): p. 3673-83.

74. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. and W. Weiss, *The current state of two-dimensional electrophoresis with immobilized pH gradients.* Electrophoresis, 2000. **21**(6): p. 1037-53.

75. Åsa M. Wheelock and A.R. Buckpitt, *Software-induced variance in two-dimensional gel electrophoresis image analysis.* Electrophoresis 2005. **26**: p. 4508-4520.

76. Addario-Berry, L., *2D Gel Electrophoresis - An Overview*. 2002.

77. Dowsey, A.W., Dunn, M. J. and G. Z. Yang, *The role of bioinformatics in two-dimensional gel electrophoresis.* Proteomics, 2003. **3**(8): p. 1567-96.

78. Meleth, S., Deshane, J. and H. Kim, *The case for well-conducted experiments to validate statistical protocols for 2D gels: different pre-processing is equal to different lists of significant proteins.* BMC Biotechnology 2005. **5**(7): p. 1472-6750-5-7.

79. Wheelock, A.M. and A.R. Buckpitt, *Software-induced variance in two-dimensional gel electrophoresis image analysis.* Electrophoresis, 2005. **26**(23): p. 4508-20.

80. Almeida, J.S., Stanislaus, R., Krug, E. and J. M. Arthur, *Normalization and analysis of residual variation in two-dimensional gel electrophoresis for quantitative differential proteomics.* Proteomics, 2005. **5**(5): p. 1242-9.

81. Guelfi, K.J., Casey, T. M., Giles, J. J., Fournier, P. A. and P. G. Arthur, *A proteomic analysis of the acute effects of high-intensity exercise on skeletal muscle proteins in fasted rats.* Clin Exp Pharmacol Physiol, 2006. **33**(10): p. 952-7.

82. Chang, J., Van Remmen, H., Ward, W. F., Regnier, F. E., Richardson, A. and J. Cornell, *Processing of data generated by 2-dimensional gel electrophoresis for statistical analysis: Missing data, normalization, and statistics.* Journal of Proteome Research, 2004. **3**(6): p. 1210-1218.

83. Storey, J.D., *A direct approach to false discovery rates.* Journal of the Royal Statistical Society: Series B (Statistical Methodology),, 2002. **64**(3): p. 479-498.

84. Stein, R.C.a.M.J.Z., *The Application of 2D Gel-Based Proteomics Methods to the Study of Breast Cancer.* Journal of Mammary Gland Biology and Neoplasia, 2002. **7**(4): p. 385-393.

85. Flohe, L. and J.R. Harris, *Peroxiredoxin Systems*. Subcellular Biochemistry, ed. J.R. Harris. Vol. 44. 2007, New York: Springer.

86. Dalle-Donne, I., Scaloni,A. and Butterfield, D.A, *Redox Proteomics:From protein modifcations to cellular dysfunction and diseases*. 2006: Wiley Interscience.

87. Jacob, C., A.L. Holme, and F.H. Fry, *The sulfinic acid switch in proteins.* Org Biomol Chem, 2004. **2**(14): p. 1953-6.

88. Lushchak, V.I., *Budding yeast Saccharomyces cerevisiae as a model to study oxidative modification of proteins in eukaryotes.* Acta Biochim Pol, 2006. **53**(4): p. 679-84.

89. Cabiscol, E., Piulats, E., Echave, P., Herrero, E. and J. Ros, *Oxidative Stress Promotes Specific Protein Damage in Saccharomyces cerevisiae*. 2000. p. 27393-27398.

90. Inoue, Y., T. Matsuda, Sugiyama, K. S., Izawa, S. and A. Kimura, *Genetic Analysis of Glutathione Peroxidase in Oxidative Stress Response of Saccharomyces cerevisiae.* The Journal of Biological Chemistry, 1999. **Vol. 274**(No. 38): p. 27002-27009.

91. Park, S.G., Cha, M. K., Jeong, W. and I. H. Kim, *Distinct physiological functions of thiol peroxidase isoenzymes in Saccharomyces cerevisiae.* J Biol Chem, 2000. **275**(8): p. 5723-32.

92. Lee, J., Spector, D., Godon, C., Labarre, J. and M. B. Toledano, *A new antioxidant with alkyl hydroperoxide defense properties in yeast.* Journal of Biological Chemistry, 1999. **274**(8): p. 4537-4544.

93. Ogita, A., Hirooka, K., Yamamoto, Y., Tsutsui, N., Fujita, K., Taniguchi, M. and T. Tanaka, *Synergistic fungicidal activity of Cu(2+) and allicin, an allyl sulfur compound from garlic, and its relation to the role of alkyl hydroperoxide reductase 1 as a cell surface defense in Saccharomyces cerevisiae.* Toxicology, 2005. **215**(3): p. 205-13.

94. Demasi, A.P.D., G.A.G. Pereira, and L.E.S. Netto, *Cytosolic thioredoxin peroxidase I is essential for the antioxidant defense of yeast with dysfunctional mitochondria.* FEBS Letters, 2001. **509**(3): p. 430-434.

95. Prouzet-Mauleon, V., Monribot-Espagne, C., Boucherie, H., Lagniel, G., Lopez, S., Labarre, J., Garin, J. and G. J. Lauquin, *Identification in Saccharomyces cerevisiae of a new stable variant of alkyl hydroperoxide reductase 1 (Ahp1) induced by oxidative stress.* J Biol Chem, 2002. **277**(7): p. 4823-30.

96. Lee, J., Godon, C., Gilles, L., Daniel, S., Je´rome,G., Labarre, J. and M. B. Toledano, *Yap1 and Skn7 Control Two Specialized Oxidative Stress Response Regulons in Yeast.* The Journal of Biological Chemsitry, 1999. **274**(23): p. 16040-16046.

97. Choi, W., Yoo, Y. J., Kim, M., Shin, D., Jeon, H. B. and W. Choi, *Identification of proteins highly expressed in the hyphae of Candida albicans by two-dimensional electrophoresis.* Yeast, 2003. **20**(12): p. 1053-1060.

98. Shin, D.H., Jung, S., Park, S. J., Kim, Y. J., Ahn, J. M., Kim, W. and W. Choi, *Characterization of thiol-specific antioxidant 1 (TSAI) of Candida albicans.* Yeast, 2005. **22**(11): p. 907-918.

99. Urban, C., Xiong, X., Sohn, K., Schroppel, K., Brunner, H. and S. Rupp, *The moonlighting protein Tsa1p is implicated in oxidative stress response and in cell*

wall biogenesis in Candida albicans. Molecular Microbiology, 2005. **57**(5): p. 1318-1341.

100. Wong, C.M., Zhou, Y., Ng, R. W. M., Kung, H. F. and D. Y. Jin, *Cooperation of yeast peroxiredoxins Tsa1p and Tsa2p in the cellular defense against oxidative and nitrosative stress.* Journal of Biological Chemistry, 2002. **277**(7): p. 5385-5394.

101. Kang, H.A., Schwelberger, H. G. and J. W. Hershey, *Translation initiation factor eIF-5A, the hypusine-containing protein, is phosphorylated on serine in Saccharomyces cerevisiae.* J Biol Chem, 1993. **268**(20): p. 14750-6.

102. Valadi, H., Valadi, A., Ansell, R., Gustafsson, L., Adler, L., Norbeck, J. and A. Blomberg, *NADH-reductive stress in Saccharomyces cerevisiae induces the expression of the minor isoform of glyceraldehyde-3-phosphate dehydrogenase (TDH1).* Curr Genet, 2004. **45**(2): p. 90-5.

103. Morigasaki, S., et al., *Glycolytic enzyme GAPDH promotes peroxide stress signaling through multistep phosphorelay to a MAPK cascade.* Mol Cell, 2008. **30**(1): p. 108-13.

104. Eberharter, A.a.P.B.B., *Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics.* EMBO Rep, 2002. **3**(3): p. 224-9.

105. Reid, J.L., Moqtaderi, Z. and K. Struhl, *Eaf3 regulates the global pattern of histone acetylation in Saccharomyces cerevisiae.* Mol Cell Biol, 2004. **24**(2): p. 757-64.

106. Escobar-Henriques, M.a.B.D.-F., *Transcriptional regulation of the yeast gmp synthesis pathway by its end products.* J Biol Chem, 2001. **276**(2): p. 1523-30.

107. Ji, Y., Gu, J., Makhov, A. M., Griffith, J. D. and B. S. Mitchell, *Regulation of the interaction of inosine monophosphate dehydrogenase with mycophenolic Acid by GTP.* J Biol Chem, 2006. **281**(1): p. 206-12.