Multidimensional Separation of Intact Proteins for Differential Proteomics Employing Top-down and Bottom-up Proteomic Strategies

Brenna McJury Richardson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2010

Approved by:

Professor James W. Jorgenson

Professor J. Michael Ramsey

Professor Royce W. Murray

Professor Linda L. Spremulli

Professor Lee G. Pedersen

ABSTRACT

Brenna McJury Richardson: Multidimensional Separation of Intact Proteins for Differential Proteomics Employing Top-Down and Bottom-up Proteomic Strategies
(Under the direction of James W. Jorgenson)

Differential proteomics, sometimes referred to as differential expression proteomics, is a subset of proteomics that attempts to identify and quantify changes in protein expression between multiple samples. One of the major challenges associated with any proteomics experiment is sample complexity. The traditional method of analysis involves a gel-based separation of intact protein for a quantitative comparison followed by analysis by mass spectrometry to identify differentially expressed proteins. While this method does have the resolving power to accommodate the sample complexity encountered in proteomics, there has been a shift towards liquid chromatography-based separations to improve automation, reduce sample bias, and more easily couple the separation to mass spectrometry.

In Chapter 2, an on-line liquid chromatographic separation strategy was developed similar in nature to the traditional gel-based approach and was applied to the analysis of differential yeast samples. It involved the on-line 2D separation of intact protein followed by MS analysis and fraction collection. A digestion of select fractions was performed to identify differentially expressed protein in a bottom-up manner. The experiment is top-down as it uses the intact protein MS signal for quantification and determination of the intact protein molecular weight, but is bottom-up in that proteins are identified following enzymatic digestion. This methodology is applied in Chapter 6 to study the differential expression of

proteins in both wild-type and β-arrestin 1,2 double knockout mouse embryonic fibroblast cells.

Chapters 3, 4 and 5 focus on examining the correlation between experiments performed in either a top-down or bottom-up manner. A digestion of the complete sample set was performed in Chapter 3 in a conventional, bottom-up only experiment. The analysis discussed in Chapter 4 directly compares the differential expression of protein determined by a top-down experiment to the expression from a bottom-up experiment. To facilitate this comparison, intact proteins were initially fractionated and then split such that each fraction was analyzed by both proteomic methods. In Chapter 5, intact proteins were fractionated prior to digestions in an attempt to improve both identification and quantification of proteins. The differential expressions from all forms of analysis of the same sample set were analyzed and compared.

ACKNOWLEDGMENTS

There are many people who I would like to thank for making this dissertation possible. I would never have been able to complete it without the guidance of my adviser, help from my friends, and support from my family. I would first like to thank my advisor, Dr. Jorgenson. His guidance, caring, patience, and enthusiasm provided for an excellent research environment. I would also like to thank fellow group members, both past and present. I'm grateful to Charles Evans who patiently taught me many of the techniques I needed to carry out my research. Similarly, I would especially like to thank Jordan Stobaugh for his help in acquiring some of the data presented in the later chapters. Through our collaboration with Waters Corporation, there have been many people who have been instrumental in providing me with both basic research needs as well as thoughtful insight into my research. I would like to thank Keith Fadgen, Scott Berger, and Martha Stapels from Waters for their help.

With all of the stress and frustration that comes with research, I have found that it was the people that I worked with every day that helped keep me motivated. I would like to thank Christine Hebling for all of her thoughtfulness, scientific discussions, and friendship. Thank you to the rest of the group as well for helping me always stay positive and work through the difficult times.

On a personal level, I would like to thank my family – my mother, father, two brothers, and sister-in-law. Their undying support and encouragement throughout my graduate career and all other endeavors have helped me accomplish all that I have done.

Lastly, I would like to thank my husband, Kyle, for being by my side every step of the way. I never would have made it without you. I love you.

TABLE OF CONTENTS

LIST OF TABLES xiv

LIST OF FIGURES xvi

LIST OF ABBREVIATIONS xix

LIST OF SYMBOLS xxiii

CHAPTER 1: Introduction and Background for Multidimensional Separations and Differential Proteomics 1

1.1 Multidimensional separations for differential proteomics 1

1.1.1 Background and theory 1

1.1.2 Previous work in comprehensive multidimensional liquid chromatographic separations of intact proteins 3

1.2 Differential proteomics 4

1.2.1 Definition and challenges 4

1.2.2 Major categories of proteomic experiments 5

1.2.2.1 Bottom up proteomics 5

1.2.2.2 Top down proteomics 8

1.2.3 Conventional method for proteomics 9

1.2.4 The use of liquid chromatography for differential proteomics 11

1.3 Scope of dissertation 11

1.4 References 13

CHAPTER 2: Differential proteomic analysis of the soluble fraction of proteins produced by cell lysates of *S. cerevisiae* grown under different conditions using an on-line multidimensional separation strategy 15

2.1 Introduction 15

2.2 Experimental 16
2.2.1 Overview of experimental method 16
2.2.2 Reagents and mobile phases 17
2.2.3 Preparation of *S. cerevisiae* protein extracts 17
2.2.4 Bradford protein quantitation assay 18
2.2.5 On-line LC x LC instrumentation and run conditions 19
2.2.6 UV and MS detection and fraction collection 20
2.2.7 Intact protein data processing and selection of fractions to be analyzed 21
2.2.8 Lyophilization and tryptic digestion of select fractions 22
2.2.9 Capillary RPLC-MS/MS of tryptic peptides 23
2.2.10 Protein identification by database searching of MS/MS data 24
2.3 Results 25
2.3.1 Replicate analysis of yeast samples for evaluation of analytical reproducibility 25
2.3.2 Molecular weight distribution of intact protein masses from online LC-LC-MS analysis 27
2.3.3 Differential analysis of intact protein 2D chromatograms 28
2.3.4 Protein identifications based on LC-MS/MS data of protein digests 30
2.4 Discussion 31
2.4.1 Evaluation of this technique in terms of analytical merit 31
2.4.2 Statistical analysis of significance of difference 32
2.4.3 Biological relevance of differential proteins as compared with literature 32
2.5 Summary and Conclusions 36
2.6 Acknowledgement 37
2.7 References 38

2.8 Tables 40

2.9 Figures 46

CHAPTER 3: Traditional bottom-up analysis of S. cerevisiae cell lysates grown under various growth conditions without pre-fractionation 59

3.1 Introduction 59

3.1.1 Bottom-up proteomics for use in a differential proteomic experiment 59

3.1.2 MS^E peptide analysis 60

3.1.3 Label-free absolute quantitation of proteins based on MS intensity of peptides 62

3.2 Experimental 63

3.2.1 Outline for experimental method 63

3.2.2 Chemicals and *S. cerevisiae* sample preparation 64

3.2.3 Trypsin digestion 64

3.2.4 Instrumentation and run conditions 65

3.2.5 Protein identification and quantification using PLGS2.4 66

3.3 Results 66

3.3.1 Protein identifications based on MS^E data 66

3.3.2 Repeatability based on replicate analysis 68

3.3.3 Determination of differential proteins and significance of the difference 69

3.4 Discussion 72

3.4.1 Protein Identification 72

3.4.2 Evaluation of reproducibility of bottom-up analysis 74

3.4.3 Identification of differentially expressed proteins 74

3.4.4 Comparison of protein intensities and correlation between those identified in Chapter 2 75

3.5 Summary and Conclusions 76

3.6 References 78

3.7 Tables 80

3.8 Figures 85

CHAPTER 4: Off-line multidimensional analysis of intact proteins from S. cerevisiae cell lysates using deconvoluted intact protein MS intensity for differential comparison 98

4.1 Introduction 98

4.1.1 Background and previous work comparing top-down and bottom-up proteomic methodologies 98

4.1.2 Separations of intact proteins at ultra-high pressures 99

4.2 Experimental 100

4.2.1 Outline for experimental method 100

4.2.2 Reagent and sample preparation 101

4.2.3 Instrumentation and run conditions for anion exchange separation 101

4.2.4 Capillary UHPLC of intact protein fractions at elevated pressures 102

4.2.5 Mass spectrometric analysis of intact proteins 103

4.2.6 Digestion of intact proteins and analysis of resulting peptides by UPLC-MS^E 104

4.2.7 Comparison of quantitation based on both top-down and bottom-up data 105

4.3 Results 106

4.3.1 Anion exchange fractionation of intact proteins 106

4.3.2 Reversed-phase separation and analysis of intact proteins 106

4.3.3 Differential comparison of 2D intact protein chromatograms 107

4.3.4 Identification of differential proteins based on molecular weight 108

4.3.5 Protein identification and differential comparison based on peptide data 108

4.3.6 Comparison of relative protein abundance between both methods on a fraction-by-fraction basis 109

4.3.7 Comparison of relative protein abundance at the whole sample level....111

4.4 Discussion....111

4.4.1 Use of the short anion exchange column for fractionation....111

4.4.2 Reversed-phase separation of intact proteins after fractionation by anion exchange....113

4.4.3 Differential analysis of 2D chromatograms from TD analysis....114

4.4.4 Identification of proteins in the BU proteomic analysis....115

4.4.5 Comparison of fold changes observed for identified proteins selected as differentially expressed in the TD experiment....117

4.4.6 Comparison of protein fold changes determined from the TD data with those determined by BU proteomics on a fraction-by-fraction basis....120

4.5 Summary and Conclusions....122

4.6 References....125

4.7 Tables....126

4.8 Figures....132

CHAPTER 5: Bottom-up analysis of anion-exchange fractionated S. cerevisiae cell lysates grown under varying growth conditions 142

5.1 Introduction....142

5.2 Experimental....142

5.2.1 Overview of experimental method....142

5.2.2 Samples and reagents used....143

5.2.3 Anion-exchange separation of intact proteins....143

5.2.4 Fractionation and trypsin digestion....144

5.2.5 UPLC-MS^E of protein digests....145

5.2.6 Protein identification and quantitation based on MS/MS data 145

5.3 Results 146

5.3.1 Fractionation of intact proteins by anion exchange chromatography 146

5.3.2 Protein identifications based on MS/MS data 146

5.3.3 Differential protein identifications based on absolute quantitation from MS/MS data 147

5.4 Discussion 149

5.4.1 Intact Protein fractionation 149

5.4.2 Protein Identification Statistics 151

5.4.3 Analysis of differentially regulated proteins and significance of intensity differences 153

5.4.4 Comparison of results of fractionated bottom-up workflow to un-fractionated bottom-up analysis in Chapter 3 156

5.5 Summary and Conclusions 158

5.6 References 161

5.7 Tables 162

5.8 Figures 167

CHAPTER 6: Differential proteomic analysis of the soluble fraction of proteins produced by cell lysates of mouse embryonic fibroblast cells: both wild-type vs. β-arrestin 1, 2 double-knockout 173

6.1 Introduction 173

6.1.1 Beta-arrestin signaling 173

6.2 Experimental 175

6.2.1 Outline for experimental method 175

6.2.2 Preparation of mouse embryonic fibroblast cell lysates 176

6.2.3 Instrumentation and run conditions at the intact protein level 177

6.2.4 Digestion and LC-MS^E analysis of individual fractions 178

6.3 Results 179
6.3.1 Differential analysis of intact protein 2D chromatograms 179
6.3.2 Replicate analysis of βarr-KO cell lysates 180
6.3.3 Protein identifications based on MS^E data 180
6.4 Discussion 181
6.4.1 Selection of differential proteins 181
6.4.2 Identification of differential proteins 182
6.4.3 Determination of the significance of the intensity differences 183
6.4.4 General comparison of results with those obtained in the differential analysis of the Baker's yeast samples 185
6.4.5 Comparison of differential regulations with literature 187
6.5 Summary and conclusions 188
6.6 Future studies 189
6.7 References 193
6.8 Tables 194
6.9 Figures 197
APPENDIX 203

LIST OF TABLES

Table 2-1 : AXC gradient used for the on-line intact protein separation..... 40

Table 2-2 : RP gradient used for the second dimension of the on-line intact protein separation.. 40

Table 2-3 : Gradient profile for the capillary RPLC separation of the digested protein fractions. 41

Table 2-4 : Collision energy (CE) profile used in the MS/MS data-directed analysis of the digested protein fractions from the on-line 2D intact protein separation. 41

Table 2-5: AutoME processing parameters for deconvolution of protein mass spectra. 42

Table 2-6: Parameters used to process the MS/MS runs of the digested protein fractions using ProteinLynx Global Server 2.3 (Waters)..... 43

Table 2-7 : List of proteins determined to have significant intensity differences between the dextrose-grown and glycerol-grown yeast samples..... 44

Table 2-8: List of differential proteins from the comparison between yeast samples harvested at either the log phase or stationary phase of growth. 45

Table 3-1: RPLC gradient condition for the analysis of digested fraction from the anion exchange column..... 80

Table 3-2: PLGS 2.4 RC7 processing parameters used for raw data processing and database searching. 81

Table 3-3: Protein identification statistics of PLGS2.4 processing of traditional bottom-up analysis. 82

Table 3-4: Proteins determined to be differentially expressed in both the traditional bottom-up and combined top-down/bottom-up online analyses of the comparison between growth media. 83

Table 3-5: List of differential proteins that were identified in both the traditional bottom-up and combined top-down/bottom-up online analyses for the comparison between phases of growth. 84

Table 4-1: Gradient conditions for the anion exchange fractionation. 126

Table 4-2: Processing parameters used for AutoME deconvolution of intact protein mass spectra. 127

Table 4-3: Intact protein masses and deconvoluted MS intensities of differentially expressed proteins selected from the visual comparison of mass slices. 128

Table 4-4: Identification statistics from PLGS2.4 processing of the LC-MS^E data obtained in the bottom-up analysis of the anion exchange fractions. 129

Table 4-5: List of proteins found in both differential yeast samples in both the TD and BU analysis of anion exchange fractions that were up-regulated in the dextrose sample of the BU comparison.. 130

Table 4-6: List of proteins found in both differential yeast samples in both the TD and BU analysis of anion exchange fractions that were up-regulated in the glycerol sample of the BU comparison.. 131

Table 5-1: AXC gradient conditions for the fractionated bottom-up analysis of the differential yeast lysates. 162

Table 5-2: RPLC gradient condition for the analysis of digested fraction from the anion exchange column. 162

Table 5-3: PLGS 2.4 RC7 processing parameters used for raw data processing and database searching. 163

Table 5-4: Evaluation of protein identifications by PLGS2.4. 164

Table 5-5: Proteins up-regulated in the glycerol-grown sample that were also identified in the dextrose-grown and glycerol-grown samples in the completely bottom-up experiment from Chapter 3. 165

Table 5-6: Equivalent to Table 5-5 except with proteins up-regulated in the dextrose-grown sample. 166

Table 6-1: Intensity distribution of differentially expressed protein masses based on intact protein MS signal. 194

Table 6-2: Protein identification statistics from the analysis of 65 fractions containing differentially expressed protein masses. 194

Table 6-3: Differentially expressed proteins found to be up-regulated in the Wild-Type MEF cell lysate. 195

Table 6-4: Proteins determined to be more abundant in the β-arrestin 1,2 double knockout sample. 196

LIST OF FIGURES

Figure 2-1: Experimental Workflow for on-line differential analysis of intact proteins. 46

Figure 2-2 : Instrumentation diagram of the complete on-line 2D intact protein separation and RPLC-MS/MS of digested protein fractions. 47

Figure 2-3 : Illustration outlining the fluidic pathways made by switching the 10-port valve from A) Position 1 to B) Position 2. 48

Figure 2-4: Example of AutoME deconvolution of intact protein mass spectra. 49

Figure 2-5 : 2D chromatograms of Auto-ME deconvoluted intact protein mass spectra. 50

Figure 2-6 : Log-log intensity plot of the triplicate analysis of lysates of yeast cells grown on glycerol harvested at the log phase of growth. 51

Figure 2-7: Distribution of protein molecular weights detected in the online LC-LC-MS analysis of proteins.. 52

Figure 2-8 : 2D chromatograms of the on-line intact protein separations of the differential yeast samples. 53

Figure 2-9 : Representative comparison of mass slice chromatograms for visual identification of differential proteins. 54

Figure 2-10 : Addition of differential protein identified by visual inspection to the log-log plot of the replicate data.. 55

Figure 2-11: Metabolic pathways of S. cerevisiae. 56

Figure 2-12 : Differential proteins identified in the comparison between S. cerevisiae. 57

Figure 2-13 : Differential proteins identified based on changes in the growth phase at which cell harvesting was performed. 58

Figure 3-1: Workflow of traditional bottom-up proteomic analysis. 85

Figure 3-2: Distribution of the molecular weight of proteins detected. 86

Figure 3-3: Replication of proteins identified by PLGS2.4 for the triplicate analysis of each differential sample.. 87

Figure 3-4: Log/log plots of replicate analysis of each sample. 88

Figure 3-5: Log/Log plots of replicate data with differential comparisons appended.. 89

Figure 3-6: Log/Log intensity plot of differential sample comparison with error bars.. 90

Figure 3-7: Bar graph comparison of replicating proteins identified in both the glycerol, log phase sample and the dextrose, log phase sample. 91

Figure 3-8: Continuation of replicating proteins identified in both the glycerol, log phase sample and the dextrose, log phase sample. 92

Figure 3-9: Bar graphs of replicating proteins identified in both the dextrose, log phase and dextrose, stationary phase samples. 93

Figure 3-10: Continuation of replicating proteins identified in both dextrose, log phase and dextrose, stationary phase samples. 94

Figure 3-11: Replicating proteins identified in either the glycerol, log phase sample or the dextrose, log phase sample, but not both..... 95

Figure 3-12: Replicating protein identified in either the dextrose sample harvested at the log phase or the dextrose sample harvested at the stationary phase. 96

Figure 3-13: Distributions of %RSD of protein intensity by a bottom-up UPLC-MS^E analysis. 97

Figure 4-1: Instrumentation workflow for off-line combined top-down/bottom-up experiment..... 132

Figure 4-2: Instrument diagram for gradient UHPLC system. 133

Figure 4-3: Anion exchange fractionation of the intact proteins from cell lysates of Baker's yeast samples grown on different carbon sources. 134

Figure 4-4: Comparison of chromatograms from the RP-LC separation of intact proteins..... 135

Figure 4-5: Intact protein 2D chromatograms showing AutoME deconvoluted data. 136

Figure 4-6: Log/Log intensity scatter plot of deconvoluted intact protein intensity from the TD analysis of anion exchange fractions. 137

Figure 4-7: Log/Log plot of absolute quantitation of replicating proteins from PLGS2.4 processing of BUD data in dextrose-grown and glycerol-grown yeast cell lysates..... 138

Figure 4-8: Data processing strategy to compare differential expression of proteins from BU and TD data on a fraction-by-fraction basis. 139

Figure 4-9: Fraction-by-fraction comparison of fold changes determined from either the bottom-up (red) or top-down (blue) data. 140

Figure 4-10: Comparison of fold-changes of proteins that were considered significantly different from the mass slice comparison of the TD data. 141

Figure 5-1: Workflow diagram for fractionated bottom-up experiment including column and instrumentation information. 167

Figure 5-2: Anion fractionation of differential yeast cell lysates monitored by UV absorption at 280nm. 168

Figure 5-3: Identification of yeast proteins and intensity scatter results for the comparison of cell lines grown on different carbons sources. 169

Figure 5-4: Comparison of the yeast proteins identified in the differential yeast samples based on differences in growth cycle at the time of cell harvest. 170

Figure 5-5: Reproducibility of absolute quantitation of replicating proteins from Chapter 4. 171

Figure 5-6: Analysis of the difference between fold changes of proteins with significantly different expression in both the fractionated and un-fractionated BU analyses. 172

Figure 6-1: Signal transduction by seven transmembrane G protein coupled receptors. 197

Figure 6-2: 2D chromatograms of AutoME deconvoluted data from intact protein 2D-LC-MS including full molecular weight range of deconvolution. 198

Figure 6-3: 2D chromatograms of deconvoluted intact protein mass spectra for the replicate analysis of βarr-KO MEF cell lysates. 199

Figure 6-4: Log/Log intensity plots of the replicate analysis of βarr-KO MEF cell lysates. 200

Figure 6-5: Venn diagram illustrating the overlap of proteins selected as differentially expressed from the intact protein intensity comparison and those identified in selected fractions after digestion. 201

Figure 6-6: Log/Log intensity plot of the replicate analysis with confidence lines from the yeast differential analysis. 202

LIST OF ABBREVIATIONS

2D	two-dimensional
2-DE	two-dimensional polyacrylamide gel electrophoresis
3D	three-dimensional
Å	Angstroms
AMT	accurate mass and time
AutoME	automated maximum entropy de-convolution
AXC	anion exchange chromatography
BEH	bridged-ethyl hybrid particle
BSA	bovine serum albumin
BU	bottom-up
CapLC	Waters capillary LC system
CE	collision energy
CID	collision induced dissociation
cm	centimeter
Da	Dalton
DDA	data directed analysis
DIA	data-independent analysis
ECD	electron capture dissociation
EGFR	epidermal growth factor
ERK	extracellular receptor kinase
ESI	electrospray ionization

ETD	electron transfer dissociation
FC	fold change
fmol	femtomoles
FTICR	Fourier transform ion cyclotron resonance
g	Grams
GDP	guanosine-5’-diphosphate
GPCR	G protein coupled receptor
GTP	guanosine-5’-triphosphate
GPTMS	3-glycidoxypropyltrimethoxysilane
HPLC	high performance liquid chromatography
Hz	hertz
ID	inner diameter
kDa	kilodaltons
kV	kilovolts
LC	liquid chromatography
m/z	mass to charge ratio
MALDI	matrix assisted laser desorption ionization
MaxEnt	maximum entropy spectral de-convolution
MEF	mouse embryonic fibroblast
mg	milligram
min	minute
mL	milliliter
mm	millimeter

mM	millimolar
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MS^E	low/elevated energy mass spectrometry
MudPIT	multidimensional protein identification technology
MW	molecular weight
NADPH	nicotinamide adenine dinucleotide phosphate
nL	nanoliter
nm	nanometer
PDA	photodiode array
pI	isoelectric point
PLGS	ProteinLynx Global Server (Waters)
PMF	peptide mass fingerprinting
PPP	pentose-phosphate pathway
psi	pounds per square inch
PTM	post-translational modification
Q-TOF	quadrupole-time of flight
RP	reversed-phase
RSD	relative standard deviation
RT	retention time
sec	second
TFA	trifluoroacetic acid
TIC	total ion current

TOF	time of flight
UHP	ultra-high pressure
UHPLC	ultra-high pressure liquid chromatography
UPLC	ultra performance liquid chromatography
UV	ultraviolet
V	volts
v/v	volume to volume ratio
WT	wild-type
YPD	yeast extract-peptone-dextrose
YPG	yeast extract-peptone-glycerol
β-arr-KO	β-arrestin 1,2 double knockout
μg	micrograms
μL	microliter
μm	micrometer
μM	micromolar

LIST OF SYMBOLS

m	ion mass
M	molecular mass
n_c	peak capacity
t	time
z	charge

CHAPTER 1: Introduction and Background for Multidimensional Separations and Differential Proteomics

1.1 Multidimensional separations for differential proteomics

1.1.1 Background and theory

One of the main issues facing the field of chemical separations is the seemingly endless complexity of samples to be analyzed. An area currently being heavily researched that involves the analysis of highly complex samples includes the '-omics' analyses. Some examples include the comprehensive analysis of genes, as is the case in genomics, metabolites (metabolomics), and proteins (proteomics), to name a few. In proteomics, much of the research now involves analysis by mass spectrometry (MS) due to the high mass accuracy and structural information that can be gained. Although some mass spectrometers have mass resolution upwards of 100,000, as is the case with a Fourier-transform ion cyclotron resonance (FT-ICR) instrument, due to the complexity of the samples in this area, a separation is still usually needed prior to analysis by MS.

One way in which separation performance can be assessed is through the determination of peak capacity. Peak capacity is defined as the number of peaks that can fit in a separation space with a resolution of 1.0 between neighboring peaks.[1] This represents the maximum number of components that could be theoretically separated on a given column or gel within a specified gradient time or separation space, respectively. Conventional high pressure liquid chromatography (HPLC) methods with operating pressures below 6,000 psi

and columns packed with 5 μm particles have peak capacities around 200.[2] With the use of smaller particles, less than 2 μm in diameter, and pumps capable of supplying the necessary pressure to run these columns, the peak capacity can increase to roughly 600. [3, 4] In spite of these improvements, there still remains a substantial disparity between the complexity of a complete proteome and the resolving power of a single dimension separation. To address this shortcoming, advances have been made in the way of multidimensional separations.

The peak capacity of a two-dimensional (2D) separation is the linear combination of the peak capacities in both of the single dimensions. For example, with a peak capacity of 50 in the first dimension and 100 in the second dimension, the theoretically obtainable peak capacity for the multidimensional separation would be 5,000. However, this is only true under ideal conditions. To obtain the full advantage that a multidimensional separation can offer, two criteria must be fulfilled. First, the separation modes must be orthogonal, or dissimilar. Historically, this has been performed by using two different separation mechanisms, such as ion-exchange followed by reversed phase. More recent work has been performed involving the two-dimensional separation of peptides using a reversed phase separation in both dimensions.[5] The orthogonality is achieved through a change in pH of the mobile phase, which affects the polarity and therefore the retention times of the peptides.

The second criterion that must be met for the multiplicative nature of peak capacity to be realized involves the sampling of the first dimension. For a 2D separation performed in space, as is the case in 2D gel electrophoresis or 2D thin layer chromatography, this becomes a non-issue, since the first dimension is not technically sampled, but rather transferred or analyzed as a whole in the second dimension. Where this does become a major factor is in 2D in time techniques, in which 2D-LC separations falls. This is an issue for both on-line

and off-line analyses where fractions are either collected or transferred in a serial manner for separation in the second dimension. In the ideal case, all resolution gained from the separation in the first dimension would be retained in the transfer to the second dimension. However, complete retention of resolution is not feasible when collecting fractions and so, some will inevitably be lost. It has been suggested that in order to minimize this loss, fractions must be collected roughly three times across the 8σ base width of the peak in the first dimension.[6] If this guideline is not followed, then there is likely to be a significant deviation in peak capacity from what could be achieved if the first dimension were not under-sampled.

A 2D separation can be made to have much greater separation power than either single dimension on its own given that the conditions described above are met. This criterion is less likely to be met in the case of 2D in time techniques than it would be in a 2D in space technique. However, even if the first dimension of a 2D in time analysis is slightly under-sampled and therefore does not fulfill the multiplicative rule, it is still likely to produce an analysis with much greater resolving power than that of a single dimension separation.

1.1.2 Previous work in comprehensive multidimensional liquid chromatographic separations of intact proteins

Multidimensional separations for work involving intact proteins can be divided into two categories: gel-based, and non-gel based. The most widely used gel-based approach is two-dimensional gel electrophoresis (2-DE). As this also became the most widely used gel-based approach for differential proteomics, it is described in detail in section 1.2.3. With regard to LC-based separation strategies, the first truly comprehensive approach was published in 1990 by the Jorgenson lab.[7] The separation modes coupled in this experiment

included cation exchange chromatography in the first dimension followed by size exclusion chromatography in the second dimension. A modest value of 130 was reported for the combined 2D peak capacity. However, with the development of smaller particles and better-performing columns, much progress has been made since then. For example, in 2003, Yan and co-workers employed a chromatofocusing separation coupled to a reversed phase separation for the proteomic analysis of adenocarcinoma cell lysates. While protein identification was performed after digesting each fraction, monitoring by UV absorption revealed more than 1000 protein bands in the 2D chromatogram.[8] A separation coupling strong cation exchange with reversed phase chromatography for the analysis of intact proteins was reported in 2006.[9] While the authors did not report a specific peak capacity, after digestion of the approximately 150 fractions collected, over 1500 proteins were able to be identified directly by MALDI-MS or through LC coupled to MALDI-MS.

While there are many research groups focusing on the improvement of 2D separations of peptides for the purpose of bottom-up proteomics, there appear to be far less that are looking at 2D separations of intact proteins as a means for competing with two-dimensional gel electrophoresis (2-DE). The experiments in this dissertation are targeted at addressing this issue.

1.2 Differential proteomics

1.2.1 Definition and challenges

Differential proteomics, or differential expression proteomics, has evolved due to the need for a more targeted approach in the broader field of proteomics. The qualitative global proteomic analyses performed over a decade ago have evolved into more quantitative, reproducibility-driven experiments. Protein biomarker discovery is strong evidence of this trend as experiments are geared toward providing robust differential expression profiles

between samples that might indicate distinct biological or temporal states.[10] The challenges associated with a differential proteomics experiment are similar to those associated with a proteomics experiment with the major obstacle being sample complexity. For example, much of the work presented in this dissertation involves the analysis of cell lysates from *Saccharomyces cerevisiae*, or Baker's yeast. Even though this is the simplest eukaryote and one of the most heavily studied organisms, the protein database obtained from SwissProt still contains more than 6,500 protein entries. Adding to the number of proteins is the large range of expression levels present. With the understanding that there is a high likelihood that some of the more interesting proteins, i.e. those that change in abundance between two samples, will not be some of the most abundant, a technique with a large dynamic range is necessary.

1.2.2 Major categories of proteomic experiments

The majority of mass spectrometry-based proteomic analyses can be divided into two main experimental strategies: top-down (TD)[11, 12] and bottom-up (BU)[13] proteomics with BU experiments being the most widely used. Methods are categorized by the types of ions that are introduced into the mass spectrometer. If the ions are intact proteins upon introduction, then the analysis would fall into a TD experiment. However, if the proteins have been enzymatically digested into peptides prior to MS analysis, then it would be a BU experiment.

1.2.2.1 Bottom up proteomics

BU proteomic experiments, which typically involve a tryptic digest of intact proteins prior to analysis, have been readily used over the past two decades. There are many different subclasses of BU proteomics including multidimensional protein identification technology (MudPIT), accurate mass and time tag (AMT), and peptide mass fingerprinting (PMF).

In a PMF analysis, peptides are usually separated prior to analysis by mass spectrometry. The peptide mass list generated from the MS scan is then compared to a list of

calculated peptide masses from an '*in silico*', or theoretical, digestion of a set of proteins or gene sequences in a specified database for protein identification. For purified proteins or simple mixtures of proteins, a PMF analysis works well provided that several peptides unique to an identified protein are present. The success of a PMF experiment and the confidence in which a protein is identified is highly dependent on the calibration and mass accuracy of the mass spectrometer as well. An analysis performed solely by PMF becomes less feasible as the sample complexity increases. With more proteins present and larger databases to search, the possibility of multiple peptides having the same mass within the calibration error of the instrument increases tremendously.

Tandem mass spectrometry has been used to handle the sample complexity present in analyses of entire proteomes. Peptides are fragmented online via collision induced dissociation (CID), which cleaves along the peptide backbone, to give a series of fragment ions based on the position of the cleavage and whether the charge remains on the n-terminus or c-terminus. The most abundant fragmentation in a CID experiment occurs at the amide bond resulting in b (charge remains on n-terminus) and y (charge remains on c-terminus) ions. The fragment ions can be used to determine the sequence of a short stretch of amino acids, which is then searched against database. Primary sequence information gained in a tandem MS experiment decreases the occurrence of peptides with overlapping masses that is problematic in PMF searches, since overlapping peptides would also have to have the same sequence in order to be ambiguous.

A MudPIT analysis employs an on-line MS/MS analysis after a two-dimensional separation of the peptides on a single capillary column packed with strong cation exchange particles in the first half and reversed phase particles in the second. Elution of peptides off of

the first column and onto the second column is achieved by raising the ionic strength of the mobile phase in a single step. The elevated ionic strength elutes a set of peptides from the first half of the column and, with no organic mobile phase present, the peptides are essentially trapped at the head of the second half of the column. After a set period of time, the ionic strength is reduced and a RP gradient of increasing hydrophobicity is flowed through the column to elute proteins from the RP segment. Once the RP gradient is finished, the entire column is equilibrated at initial conditions. A second step of elevated ionic strength is performed, this time higher in ionic strength that the first step to elute a second set of peptides, which are subsequently eluted from the second half of the column by a second RP gradient. A series of ionic strength and RP gradient cycles are performed until all species are eluted off the column[14, 15]

Lastly, the AMT approach combines both single MS and tandem MS analyses into one. By using an instrument with high mass accuracy, such as a Fourier-transform ion cyclotron resonance (FT-ICR) MS, coupled to an LC separation with good retention time reproducibility, mass and time tags can be established.[16] In the initial experiment, LC-MS/MS data are acquired in a data-dependent manner. The sequences from short stretches of amino acids obtained from the fragmentation analysis are searched against a database to identify a protein. The chromatographic retention times of the peptides used to identify the protein as well as the peptide mass are recorded as AMT tags. In subsequent analyses, single MS analysis is performed and the presence of the protein is determined by the presence of a peptide mass at a retention time that both match, within a specified tolerance, an AMT tag from the MS/MS run. This type is most commonly employed for a higher throughput analysis where many similar samples will be analyzed for targeted proteins.

Some of the major advantages of a BU experiment include the relative ease of separations at the peptide level, the predictability of peptide fragmentation by collision induced dissociation, and the availability of commercially available software to perform database searching. For these reasons, it is far more widely used than TD proteomics, although this type of experiment does come with some limitations. For example, the sequence coverage in a bottom-up experiment can vary widely from protein to protein ranging anywhere from approximately 5-70%. The major disadvantage is the lack of determination of the molecular weight of the intact protein. Because many of the post-translational modifications (PTMs) may be lost during the MS/MS fragmentation at the peptide level, very little about a protein is known other than its presence or absence from a sample.

1.2.2.2 Top down proteomics

In comparison with bottom-up proteomics, top-down proteomics is a relatively immature field. It is based on the direct ionization of intact proteins and subsequent analysis by tandem mass spectrometry, usually in a data-dependent acquisition. Protein mixtures are usually separated to obtain a single protein prior to ionization. While analysis of intact proteins for structural information began with the invention of soft ionization techniques such as matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), it did not really take hold until the development of electron capture dissociation (ECD) in the McLafferty lab in 1998.[17] In an MS/MS experiment with ECD, low energy electrons are captured by multiply charged protein precursor ions to produce $[M + nH]^{(n-1)+\cdot}$ ions. Rapid fragmentation of this species occurs at the N-αC bond to produce c and z-type ions as opposed to the b and y-type produced in a CID experiment. Fragmentation occurs much more randomly in ECD than it does in CID and can therefore offer complementary

information to that of a CID experiment. Also, because it is a nonergodic fragmentation process, PTM's are more likely to be preserved in their original location and site-specific PTM information can be gained.[18] However, in order to obtain favorable signal-to-noise ratios in ECD experiments, several spectra must be averaged, which often precludes its online coupling to an LC separation. More recently, electron transfer dissociation (ETD) was developed which involves the transfer of an electron from an anion of low electron affinity to a multiply charged protein cation. Fragmentation occurs in the same manner as in ECD, but in an ion trap and on an LC timescale.[19]

Some advantages to a TD experiment include a greater possibility to increase the sequence coverage of a protein and determine more accurately the presence and position of many PTMs. However, there are still some significant limitations encountered with this type of analysis. Due to the complexity of the fragmentation spectra, this type of analysis is often limited to purified proteins or a small mixture. Lastly, data analysis for this analysis is more labor-intensive than that of a bottom-up experiment. There are several well-developed software packages that are commercially available to aid in the analysis of BU data. In contrast, TD data are often analyzed manually, which is one of the factors preventing its use as a high throughput method.

1.2.3 Conventional method for proteomics

The classical method for a differential proteomic experiment involves separated gel-based separations of the intact proteins from multiple samples for a quantitative comparison followed by digestion and analysis of peptides by mass spectrometry for identification. This method has shown the ability to resolve several thousand proteins in a single run.[20, 21] The typical separation method performed is two-dimensional gel electrophoresis (2-DE). Intact

proteins are separated by isoelectric focusing in the first dimension, which separates based on isoelectric point (pI). The gradient strip is then transferred to a sodium dodecyl sulfate polyacrylamide gel for an electrophoretic separation which roughly separates according to protein molecular weight.[22] After staining, images of the gels obtained from the separation of differential samples are processed through image analysis software in order to determine which protein spots vary in intensity between the samples. These spots are excised from the gel manually, or with the aid of a robotic device, digested and analyzed by MALDI- MS to gain accurate mass information for the resulting peptides. The accurate mass data are then searched against a protein database for tryptic peptides with identical masses as is done with a PMF analysis. To confirm the protein identity, a select few of the peptides used for identification are analyzed further by tandem mass spectrometry in order to gain more specific sequence information.

While 2-DE does have the peak capacity to accommodate the sample complexity encountered in proteomics, it does suffer from a few critical limitations. Namely, hydrophobic proteins may not enter the gel for separation, highly acidic or basic proteins are not as well resolved as those of intermediate pI, less abundant proteins may not be observed due to limitations with the various staining methods, and lastly, protein identification may be difficult and time consuming as spots must be excised from the gel, digested and often analyzed more than once to get a confident identification. Furthermore, only a small fraction of the peptides used to identify a protein undergo MS/MS analysis increasing the likelihood of false peptide hits, especially with overlapping gel spots.

1.2.4 The use of liquid chromatography for differential proteomics

Due to the limitations of gel-based differential proteomics described above, there has been a shift toward trying to do an equivalent analysis based on liquid chromatographic separations. The general trend in this type of experiment is to use an initial separation of the intact proteins for fractionation in order to simplify the mixtures prior to digestion.[10, 23-25] In some experiments the intact mass of the protein may be determined through analysis by mass spectrometry prior to digestion[23], although this is not often the case. By reducing the number of proteins present in a fraction, the number of peptides present in the sample is greatly reduced. The reduction is magnified due to the fact that digestion results in multiple peptides per protein.

1.3 Scope of dissertation

The experiments and data presented here represent a novel separation strategy employed to determine changes in protein expression between two samples. The classical approach to this is 2-DE, but, as described earlier, it suffers from some major limitations such as protein bias and limited dynamic range. Recent experiments aiming to replace 2-DE are based mainly on BU proteomic experiments in which both the protein identification and quantitation is performed at the peptide level. The molecular mass and, therefore, the protein isoform are never determined. The aim of this research is to present a method that combines some of the advantages from both TD and BU experiments while employing some of the latest technology. Relative quantitation is performed at the intact protein level, since that is the species actually present in the sample, while identification is achieved through the more reliable BU analysis after digestion. This method will be compared with some traditional BU experiments as well as those in which a separation of the intact protein is performed solely to simplify the mixture prior to digestion. The overlap between the differential comparisons of

protein abundance based on intact protein MS intensity and that based on peptide data will be compared to evaluate the extent of correlation.

1.4 References

(1) Giddings, J. C. *Analytical Chemistry* **1967**, *39*, 1027-1028.

(2) Wehr, T. *LC-GC North America* **2002**, *20*, 954-957.

(3) Mellors, J. S.; Jorgenson, J. W. *Analytical Chemistry* **2004**, *76*, 5441-5450.

(4) Wang, X.; Barber, W. E.; Carr, P. W. *Journal of Chromatography A* **2006**, *1107*, 139-151.

(5) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Journal of Separation Science* **2005**, *28*, 1694-1703.

(6) Murphy, R. E.; Schure, M. R.; Foley, J. P. *Analytical Chemistry* **1998**, *70*, 1585-1594.

(7) Bushey, M. M.; Jorgenson, J. W. *Analytical Chemistry* **1990**, *62*, 161-167.

(8) Yan, F.; Subramanian, B.; Nakeff, A.; Barder, T. J.; Parus, S. J.; Lubman, D. M. *Analytical Chemistry* **2003**, *75*, 2299-2308.

(9) Gao, M.; Zhang, J.; Deng, C.; Yang, P.; Zhang, X. *Journal of Proteome Research* **2006**, *5*, 2853-2860.

(10) Sutton, J.; Richmond, T.; Shi, X.; Athanas, M.; Ptak, C.; Gerszten, R.; Bonilla, L. *Proteomics: Clinical Applications* **2008**, *2*, 862-881.

(11) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. *Journal of the American Chemical Society* **2002**, *124*, 672-678.

(12) Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.

(13) Henzel, W.; Billeci, T.; Stults, J.; Wong, S.; Grimley, C.; Watanble, C. *Proceedings of the National Academy of the Sciences of the United States of America* **1993**, *90*, 5011-5015.

(14) Washburn, M. P.; Wolters, D. A.; Yates, J. R., III *Nature Biotechnology* **2001**, *19*, 242-247.

(15) Florens, L.; Washburn, M. P. *Methods in Molecular Biology* **2006**, *328*, 159-175.

(16) Bogdanov, B.; Smith, R. D. *Mass Spectrometry Reviews* **2005**, *24*, 168-200.

(17) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *13*, 3265-3266.

(18) Breuker, K.; Oh, H.; Lin, C.; Carpenter, B. K.; McLafferty, F. W. *Proceedings of the National Academy of the Sciences of the United States of America* **2004**, *101*, 14011-14016.

(19) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of the Sciences of the United States of America* **2004**, *101*, 9528-9533.

(20) Wehr, T. *LC-GC North America* **2001**, *19*, 702-711.

(21) Zhou, W.; Ryan, J. J.; Zhou, H. *Journal of Biological Chemistry* **2004**, *279*, 32262-32268.

(22) O'Farrell, P. H. *Journal of Biological Chemistry* **1975**, *250*, 4007-4021.

(23) Williams, T. L.; Monday, S. R.; Edelson-Mammel, S.; Buchanan, R.; Musser, S. M. *Proteomics* **2005**, *5*, 4161-4169.

(24) Wasinger, V. C.; Locke, V. L.; Raftery, M. J.; Larance, M.; Rothemund, D.; Liew, A.; Bate, I.; Guilhaus, M. *Proteomics* **2005**, *5*, 3397-3401.

(25) Webber, M. A.; Coldham, N. G.; Woodward, M. J.; Piddock, L. J. *Journal of Antimicrobial Chemotherapy* **2008**, *62*, 92-97.

CHAPTER 2: Differential proteomic analysis of the soluble fraction of proteins produced by cell lysates of *S. cerevisiae* grown under different conditions using an on-line multidimensional separation strategy

2.1 Introduction

As discussed in the first chapter, a switch from a gel-based differential proteomics approach to a liquid chromatography approach would be beneficial for multiple reasons. Namely, an LC-based approached would be more conducive to automation due to its inherent fluidic nature. In addition, the use of mass spectrometry to determine the intact molecular weight of the proteins during the two-dimensional (2D) separation would be a vast improvement over mass resolution achieved in the SDS-page separation of the gel-based approach. [1] The two main categories of proteomic analyses were also discussed: top-down and bottom-up proteomics. The work presented in this chapter combines the application of an LC-based approach to a differential proteomic analysis. The methodology is also a combination of both a top-down and a bottom-up analysis. After performing a number of off-line separations of intact proteins, it was determined that a multidimensional on-line separation of intact proteins by anion exchange followed by reversed phase would be the best option to obtain the intact protein molecular weight in a top-down manner. [2]

The purpose of the work presented in the chapter was to examine the capabilities of the instrumental method for a differential proteomic analysis; it will be applied later to a less well-understood sample set. When determining which type of samples to analyze, a logical

choice was *Saccharomyces cerevisiae* (baker's yeast) due to breadth of existing knowledge on this single cell eukaryote in the fields of both biochemistry and cellular biology.[3, 4] Additionally, the yeast cell lysates have a decreased sample complexity as compared to other eukaryotic species that will be studied in a later chapter. Differences in protein expression were induced in two separate ways: by varying the carbon source in the grown nutrient, and by varying at which phase during the growth cycle the cells were harvested.

2.2 Experimental

2.2.1 Overview of experimental method

The overall workflow for the complete experiment is shown in Figure 2-1. The instrumentation can be broken down into two separate processes; online LC-LC-MS of intact proteins and LC-MS/MS of peptides following digestion. The first separation includes the coupling of anion exchange chromatography in the first dimension to reversed phase LC in the second dimension. The majority of the effluent from the second dimension columns is collected by a fraction collector while a small amount is diverted to a mass spectrometer for determination of intact protein masses. All samples relevant to a single differential proteomic experiment are analyzed by online 2D-LC of intact proteins prior to moving on in the workflow. The resulting 2D chromatograms are compared to one another to highlight potential differences in protein expression between samples. Once a list of differential protein masses has been compiled, the corresponding fractions from the online separation are then lyophilized, digested, and run by LC-MS/MS for identification by a BU-type experiment. A comparison is then made between the differential protein mass list and the predicted molecular weights of proteins identified by database searching of the peptide MS/MS data. Two sets of three runs each were run during this experiment. The first

consisted of a triplicate analysis of the glycerol-grown yeast sample to assess analytical reproducibility and the second was a series of the three differential yeast samples.

2.2.2 Reagents and mobile phases

The chemicals used for the LC mobile phases were ammonium acetate, ammonium hydroxide, formic acid, LC-MS grade water and acetonitrile, all purchased from Fisher Scientific (Fair Lawn, NJ). Deionized water was purified using a Barnstead Nanopure System (Boston, MA). Chemicals used in the trypsin digestion were: ammonium bicarbonate, iodoacetamide and trifluoroacetic acid, also purchased from Sigma (St. Louis, MO), RapiGest SF, an acid-labile surfactant provided by Waters Corporation (Milford, MA), dithiothreitol (Research Products International, Mt. Prospect, IL), and TPCK-modified trypsin (Pierce, Rockford, IL).

Standard proteins, which were used for intensity adjustments, were Cytochrome C, myoglobin and bovine serum albumin, which were all purchased from Sigma Aldrich and used, as received, after dilution to the desired concentration.

2.2.3 Preparation of *S. cerevisiae* protein extracts

Cytosolic yeast protein samples were prepared from *S. cerevisiae* cultures by researchers at Waters Corporation (Milford, MA). Yeast cells were initially maintained on dextrose-containing agar plates purchased from Teknova (Hollister, CA). In order to induce changes in protein expression three select colonies were transferred to culture flasks, two of which contained YPD (dextrose carbon source) and one contained YPG (glycerol carbon source). After initial overnight incubation, the small scale cultures were used to inoculate prep scale cultures of each sample type. Similar growth conditions were maintained across all samples prior to harvest by centrifugation. The sample grown on glycerol was harvested

during the logarithmic phase of growth. While one of the dextrose-grown samples was also harvested at the logarithmic phase of growth, the other was harvested once the growth rate had reached the stationary phase. Cells were stored at -80°C prior to lysis.

The lysis conditions have been detailed previously by Millea, et al.[5] The final samples contained cytosolic yeast proteins in 25mM ammonium bicarbonate along with various phosphatase inhibitors.

2.2.4 Bradford protein quantitation assay

Total protein concentrations for the cytosolic yeast samples were determined using a Coomassie Plus Bradford assay kit (Pierce Biotechnology, Rockford, IL). Protein samples were diluted 50-fold in order to reduce the concentration to fall within the linear range of the standards prepared by following the standard test tube procedure outlined in the assay kit. A series of nine standard dilutions of bovine serum albumin were made for calibration. Fifty μL of each protein solution was transferred to a centrifuge tube and 1.5mL of the standard Coomassie reagent was added. Solutions incubated at room temperature for 10 minutes. Analysis was performed in triplicate for each sample and standard. Absorbance readings were taken at 595 nm. From this assay, it was determined that the concentration of the glycerol/log phase sample was 13.5 mg/ml, the dextrose/log phase sample was 11.0 mg/mL and the dextrose/stationary phase sample was 10.5 mg/mL. All solutions were diluted to a final concentration of 10 mg/mL prior to analysis. Cytochrome C and myoglobin were spiked into each sample at a concentration of 10 ng/mL each along with β-lactoglobulin at 20 ng/mL in order to account for fluctuations in instrument reproducibility from day to day.

2.2.5 On-line LC x LC instrumentation and run conditions

A diagram outlining the instrumentation for the on-line 2D LC separation is shown in Figure 2-2. The two separation modes used in the multidimensional separation were anion exchange in the first dimension followed by reversed phase in the second dimension. The first dimension column was a series of 3 custom-packed 6.6mm ID x 40 cm Omnifit glass column hardware (Sigma-Aldrich, St. Louis, MO) giving an effective column length of 111 cm. The packing material used in the anion exchange column was expelled from a Waters BioSuite Q preparative scale column and consisted of 13 μm polymeric particles with 1000 Å pores bonded with a strong anion exchange quaternary amine stationary phase. The initial injection of cell lysate onto the first dimension anion exchange column was performed using a Valco 6-port valve (VICI, Houston, TX) equipped with a 550 μL sample loop. The gradient for the first dimension was supplied by a Waters 600 quaternary gradient LC pump.

The outlet of the anion exchange column was connected to a Valco 10-port valve (VICI, Houston, TX), which directed anion exchange effluent onto one of the reversed phase columns. There were two identical columns used as the second dimension, which were Waters BioSuite pPhenyl columns with dimensions of 4.6 mm ID x 7.5 cm packed with 10 µm, 1000 Å pore polymeric particles bonded with a phenyl stationary phase. Figure 2-3 outlines the operation of the on-line set-up. When the valve was in position 1, proteins eluting from the anion exchange column were trapped at the head of reverse phase column B due to the absence of any organic modifier in the anion exchange mobile phase. The anion exchange mobile phase passed through the reverse-phase column, through a small piece of tubing and then back through the 10-port valve prior to being diverted to waste. Meanwhile, reverse phase column A received a gradient generated by a Hewlett-Packard 1050 HPLC

pump (Agilent, Santa Clara, CA) to elute proteins previously trapped on that column. After a set period of time, 30 minutes in this experiment, the 10-port valve was switched to divert effluent from the anion exchange column onto reverse phase column A. At this point, the proteins that had been previously trapped on reverse phase column B were eluted off the column by the gradient provided by the second HPLC pump. A full 2D run was completed by switching the 10-port valve multiple times while proteins continued to elute from the anion exchange column throughout the gradient supplied by the first LC pump.

The anion exchange gradient consisted of 10mM ammonium acetate, adjusted to pH 9.0 with ammonium hydroxide in mobile phase A. Mobile phase B was 750 mM ammonium acetate, also adjusted to pH 9.0 with ammonium hydroxide. The mobile phases for the reverse phase separations consisted of water in mobile phase A and acetonitrile in mobile phase B, both with 0.2% formic acid added. Table 2-1 and Table 2-2 outline the flow rates and gradient profiles used in the anion exchange and reversed phase separations, respectively.

2.2.6 UV and MS detection and fraction collection

The effluent from the second dimension RP columns was split such that 95% of the flow was directed to an Applied BioSystems 785A UV detector (Foster City, CA) set at 193nm and then on to a fraction collector (Waters Fraction Collector II). Sixteen 45-second fractions were collected from 2 minutes to 14 minutes during each of the reversed-phase runs for a total of 480 fractions per 2D run. The remaining 5% of the flow was directed to a Waters LCT mass spectrometer for intact mass analysis by ESI-MS. The mass spectrometer was operated in positive ion mode using a standard z-spray ion source. Calibration of the instrument was performed daily using a solution of sodium formate. For mass spectral

analysis, data were acquired at an acquisition rate of 1 Hz over the *m/z* range of 450-1600. Voltages applied to the electrospray needle and sample cone were +3000V and +35V, respectively. A nitrogen flow of 250 L/hr was used as the desolvation gas.

2.2.7 Intact protein data processing and selection of fractions to be analyzed

Intact protein MS chromatograms were processed by AutoME, a software program written by Ignatius Kass at Waters Corporation to perform maximum entropy (MaxEnt) processing of the mass spectra in an automated manner. AutoME divides the chromatogram into segments and combines all spectra within each segment before performing MaxEnt processing. Protein mass spectra obtained by electrospray ionization contain a series of multiply charged ions on a mass/charge ratio scale. MaxEnt is an iterative mathematical process designed to transform the multiply charged ion series to a single peak on a molecular mass scale in which all components have zero charge.[6] Probabilistic quantitation is also performed using the MaxEnt approach effectively combining intensities from all charge states to give a single intensity at the molecular mass. For reference, a mass spectrum from a peak separated in the 2D analysis of intact proteins is shown before and after AutoME deconvolution in Figure 2-4.The AutoME processing parameters are included in Table 2-5. Deconvoluted protein masses that are present as multiples of the main protein mass are artifacts of the deconvolution process. They are removed by AutoME prior to the generation of the deconvoluted protein mass lists through the de-harmonizing function, which was set to remove masses within 20 Da of any multiples of the main mass. The AutoME-processed mass spectra were used for the construction of the 2D plots.

2.2.8 Lyophilization and tryptic digestion of select fractions

Fractions determined to have differentially expressed proteins were thawed, transferred to 1.5 mL centrifuge tubes, and flash frozen prior to lyophilization. Lyophilization was performed using a SpeedVac Concentrator (Thermo-Electron, Bellefonte, PA), which was pumped down to pressures between 10^{-2} and 10^{-3} Torr using an Edwards double-stage rotary vacuum pump. (Wilmington, MA). Once the fractions were lyophilized to dryness, they were reconstituted in 25 μL of ammonium bicarbonate.

The standard digestion procedure provided with the RapiGest surfactant was used with modifications to account for the quantity of protein expected in each fraction as well as the desired volume after digestion. A 1mg vial of RapiGest was reconstituted with 150 µL ammonium bicarbonate. Three µL of RapiGest was added to each fraction. Samples were vortexed and incubated at 80°C for 15 minutes, vortexing periodically. Samples were centrifuged at 10,000 x g for 4 minutes to return condensate to the bottom of the vial. One µL of 100mM dithiothreitol was added, the solutions were vortexed, and allowed to incubate at 60°C for 30 minutes in order to reduce disulfide bonds. After centrifugation for 4 minutes at 10,000 x g, 1 µL of 200mM iodoacetamide was added to alkylate the sample and prevent disulfide bonds from re-forming. Solutions were incubated, in the dark, at room temperature prior to centrifuging again for 4 minutes at 10,000 x g. Finally, 10 µL of 1mg/mL trypsin in ammonium bicarbonate was added and digestion occurred overnight at 37°C.

Quenching of digestion was performed by the addition of 10 µL of 10% trifluoroacetic acid and 40 µL of water to reduce the pH of the solution below 2. Samples were incubated at 37°C for 30 minutes to allow for complete hydrolysis of the RapiGest

surfactant. After centrifugation at 12,000 x g for 10 minutes to remove precipitate, the supernatant was transferred to a sample vial for LC-MS/MS analysis.

2.2.9 Capillary RPLC-MS/MS of tryptic peptides

Bottom-up analysis of the selected digested fractions was performed using capillary RPLC-MS/MS. A Waters Integrated CapLC/PDA system equipped with an autosampler and a Waters Q-TOF Micro mass spectrometer were used for analysis. The PDA capabilities of the LC system were not utilized for this analysis. The capillary column was a 150 µm ID capillary packed with 1.5 µm bridged ethyl hybrid C_{18} particles and was packed in-house to a length of 8 cm. The system was operated using split flow in order to supply the low flow rates, roughly 500 nL/min, required by the column while shortening the gradient delay time. The splitter was located just prior to the inject valve in order to minimize the gradient delay volume without splitting the sample. The outlet of the column was connected to a 3cm x 20µm ID pigtail through a Teflon junction, which connected directly to the nano-electrospray source. Because a sintered particle frit was used for the end of the column, the protective polyimide coating was removed. This resulted in exposed silica leaving the end of the column very brittle. The Teflon junction and pigtail were used to preserve the column outlet as the Teflon junction produced less stress on the column the traditional ferrule-type fitting used in the electrospray interface would have produced. The small inner diameter of the pigtail was selected to reduce extra column band-broadening that may have been caused by a larger piece of capillary. The gradient profile and flow rate information is given in Table 2-3.

The nano-electrospray interface was equipped with a pulled, uncoated silica spray tip (New Objective, Woburn, MA), which had a 20µm ID pulled to 10 µm at the tip orifice. The

MS/MS analysis of the peptides was performed using a data-directed analysis method (DDA) on a Q-TOF Micro mass spectrometer (Waters). In a DDA method, there are two types of scans that take place, a survey scan and a product scan. In a survey scan, the quadrupole is set to allow all m/z values to pass through to the collision cell, which is set at a low voltage (+7V) so as not to fragment anything before entering the TOF analyzer. Once a survey scan is taken (400-1600 *m/z*), the two most intense ions above a user-defined threshold (25 counts) are selected, one at a time, to be subjected to an MS/MS analysis. A three-scan collision energy profile was used with each precursor ion selected, each with a different collision energy based on the precursor m/z as described in Table 2-4. Fragment ion scans were acquired from m/z 50-1800 at a scan rate of 1Hz. After the three product scans for each of the precursor ions selected have occurred, another survey scan at low collision energy is performed. In order to prevent a precursor ion from being selected multiple times for MS/MS analysis, after an MS/MS scan was taken for each peptide, the peptide *m/z* was added to a dynamic exclude list for one minute. This allowed MS/MS analysis of some of the lower intensity co-eluting peptides. Applied voltages for the capillary, sample cone and extraction cone were +2000V, +30V, and +2V, respectively. The rate of acquisition was 1 Hz for both survey and product scans.

2.2.10 Protein identification by database searching of MS/MS data

Identification of the proteins based on the MS/MS data was performed using the ProteinLynx Global Server 2.3 search platform (PLGS2.3, Waters Corporation), which allows for the automated processing of raw MassLynx data and searching of peptides against protein databases. PLGS processing is broken down into two steps. The first involves the background subtraction, de-isotoping, and centroiding of the raw data acquired by

MassLynx. Searching of the processed mass spectrum obtained in step one against a protein database occurs in the second step. The database that was used for identification was the entire yeast protein database obtained from the Swiss-Prot protein knowledgebase release 54.2 (http://us.expasy.org/sprot) with trypsin added as well. Appended to the end of the database was a 1x randomization of the yeast proteins in order to determine a false discovery rate. A number of parameters are defined in both processing steps and can be found in Table 2-6.

2.3 Results

2.3.1 Replicate analysis of yeast samples for evaluation of analytical reproducibility

In order to assess the reproducibility of the 2D intact protein separation, a series of three injections of the yeast sample grown on glycerol were performed. The full 2D chromatograms from these separations are shown in Figure 2-5. Upon visual inspection, the general pattern of peaks is almost identical. There are some slight variations in retention time likely due to the evaporation of ammonia from the mobile phase reservoirs throughout the runs, thereby reducing the mobile phase pH. These differences appear to affect all of the peaks within a run in a similar fashion. Because the determination of protein differences is performed visually, slight changes in retention time should not hinder this. Also apparent are slight differences in peak intensity, represented in false color in the plots. The peak intensities in each chromatogram have been adjusted based on the intensity of the cytochrome C that was spiked into the original sample. The purpose of the addition of cytochrome C was to account for day-to-day MS intensity fluctuations. Therefore, all changes in peak intensities between these replicates are due to differences imposed by the method itself.

In order to quantify the experimental error to determine if protein intensity differences are significant when comparing differential samples, a plot was made to visualize the differences in intensity between replicates. For this plot, the intensity of a single protein in one replicate was compared to the intensity of the identical mass in a second replicate. The data was plotted on a log scale to better visualize the large range of protein intensities. This plot can be seen in Figure 2-6 in which the intensity of a protein in one replicate is plotted against the intensity in another. Since there are three replicates, the data is plotted in the following manner: replicate 1 vs. replicate 2, replicate 2 vs. replicate 3, and replicate 3 vs. replicate 1. Also included in this plot is the line y=x, where all points would fall given ideal conditions. The data follow this line fairly well at high intensity, within 10-20%. However, as the intensity decreases, the data points deviate from linearity more and more so that near the detection limit, the differences reach a 100-fold change. The trend of increasing deviation from linearity with decreasing intensity closely follows a line with the equation y= mx + b in the linear scale. Experimentally, this equation takes into account a constant level of uncertainty, b, and a response factor uncertainty component, m, which is dependent on the signal intensity. Variations of this equation are plotted along with the replicate data in Figure 2-6 along with the mirror image of these lines across the line y=x. Along with the equation of each line, the legend also includes a percentage value in parentheses. This number corresponds to the percentage of replicate data points that fall within the two lines associated with that equation. To a first approximation, the value can be used as the percent certainty that the difference is real and not a product of the method.

It was arbitrarily decided that the curve with the equation $y = 1.2x + 10^{4.4}$, which contains 96.5% of the replicate data points, was an acceptable threshold for determining if a change in protein intensity between differential samples was significant.

2.3.2 Molecular weight distribution of intact protein masses from online LC-LC-MS analysis

One of the issues commonly seen with the analysis of intact proteins with ESI-MS is the bias that occurs towards proteins below a molecular weight of approximately 50 kDa. Proteins above this limit are less likely to be seen with ESI. To evaluate this phenomenon, a histogram demonstrating the molecular weight distribution of the detected protein masses was plotted with the actual molecular weight distribution of all yeast proteins as reported in the database. The histogram is included in Figure 2-7 A. In order to better see both distributions on the same scale, the number of counts for each molecular weight bin was divided by 5 for the distribution of all yeast proteins. As expected, there is a definite bias to proteins with lower molecular weights, but overall, the trends of the distributions are quite similar.

An artifact of the processing is also apparent in the molecular weight distribution and can be easily seen in Figure 2-7 B. The increase in the number of proteins detected with molecular weights greater than 50 kDa is actually the summation of low intensity noise present in almost every AutoME-deconvoluted mass spectrum. This high molecular weight noise does not affect the selection of protein masses with varying intensities due to the visual comparison that is performed on the mass slice chromatograms, which was described in section 2.2.7. In the 2D chromatograms, a low intensity background that is visible in mass

slices above 60 kDa is likely the source the majority of the masses in this range of the histogram.

2.3.3 Differential analysis of intact protein 2D chromatograms

2D chromatograms from the on-line separation of intact proteins of the differential yeast samples are shown in Figure 2-8. As was the case with the replicate analysis, there is great similarity between all three plots with regard to the general pattern of protein peaks. This is not unexpected, since the samples were all generated from the same organism. However, unlike the replicate data, there are clearly some strong differences in peak intensities. Also, there are a greater number of instances in which a peak is present in one sample and absent in another. Some intensity differences are easy to pick out from the full 2D chromatograms, but for a more thorough analysis of the data requiring a simplification of the plots was necessary.

All intact protein chromatograms in Figure 2-5 and Figure 2-8 contain proteins with molecular weights over the entire mass range of the AutoME output. Because the molecular weight of each peak in the chromatogram is known to within 2Da (the resolution defined by the AutoME parameters), the plot can be simplified so that only a limited range of masses are shown. Any range of masses could be chosen, but to limit the total number of slices generated while still decreasing the complexity, a mass range of 1 kDa was chosen. A side-by-side comparison of two 'mass-slice' chromatograms can be found in Figure 2-9. In this mass slice comparison all of the possibilities are highlighted: one where protein intensity does not appear to change significantly, one in which the intensity was greater in the glycerol sample, two where the intensity is greater in the dextrose sample, and one where a protein is present in the glycerol sample, but absent in the dextrose sample. After going through all of

the 1 kDa mass slices from 5 to 80 kDa and comparing both the glycerol/log phase sample to the dextrose/log phase sample and the dextrose/log phase sample to the dextrose/stationary phase sample, a total of 164 protein masses were selected for further analysis by LC-MS/MS of the digests for protein identification. Because some fractions contained multiple differential protein masses, it was only necessary to analyze 125 of the 1440 total fractions collected (<9%).

The use of the mass slice 2D chromatograms to visually pick out changes in protein intensity facilitated the selection of differential proteins. However, in order to determine if the protein difference was significant, a more mathematical approach was necessary. This was accomplished by creating log-log intensity plots identical to those used for the replicate analysis. The logical comparisons were between the glycerol/log phase sample and the dextrose/log phase sample and the dextrose/log phase sample and the dextrose/stationary phase sample. By plotting the differential protein data on the replicate data, the significance of the protein differences can be determined. This is shown in Figure 2-10. Fifty-two of the 76 differential proteins (68%) that were identified by the peptide LC-MS/MS were determined to be significantly different based on the 96.5% confidence threshold discussed in the previous section.

For proteins that were detected in both of the samples used in a comparison, determination of the significance is straight-forward as described above. However, for proteins that were only found in one sample, this becomes more challenging. The limit of detection for the online 2D separation of intact proteins was estimated to be roughly 500 counts. Therefore, the axes of the plots were set accordingly. Proteins that were only identified in one sample were plotted such that the intensity in the sample in which it was not

detected was set at 500 counts. This value should represent the greatest signal intensity that a protein could generate without being detected.

2.3.4 Protein identifications based on LC-MS/MS data of protein digests

When a differential protein was identified from the intact 2D LC separation, the corresponding fraction from the sample in which it was most intense was lyophilized and reconstituted for identification from the bottom-up portion of the analysis. Peptides were analyzed using capillary RPLC-MS/MS and MS/MS data was processed and searched against the SwissProt yeast proteins using ProteinLynx Global Server 2.3. In the 125 fractions that were found to contain potential differential proteins, 230 unique proteins were identified (766 proteins including redundant hits). Not all of the yeast proteins identified from the peptide data were differential proteins, but rather identification indicated the presence of a protein in a particular fractions. The proteins identified by PLGS were entered into the Compute pI/MW tool on the ExPASy website (http://us.expasy.org) in order to determine the theoretically processed molecular weight, which takes into account all known cleavages of the protein as well as other modifications. The molecular weights of the 'processed' proteins were then compared to the molecular weight of the differential protein as determined by the intact protein MS analysis. After comparing these two lists on a fraction-by-fraction basis, 76 of the proteins had similar masses. With 162 differential proteins selected from the intact protein data, approximately 53% were successfully identified. Fifty-two of the 76 differential proteins identified were also found to have a significant difference in expression at the 96.5% threshold. Of the 766 total proteins identified by PLGS from the MS/MS data, 44 were hits for proteins from the randomized database, which leads to a false discovery rate of 5.7%.

The proteins that were both successfully identified from the peptide LC-MS/MS data and had significant differences in expression are shown in Table 2-7 for the comparison between the glycerol-grown and dextrose-grown samples and in Table 2-8 for the comparison between the log-phase and stationary-phase samples. In addition to the protein description, the Swiss-Prot entry and the gene that encodes the protein are included. The number of peptide hits indicates the peptides that were used to identify the protein by PLGS and the intact mass is that which was calculated from the intact protein MS data by AutoME. The fold change was calculated by dividing the intensity of the protein in the sample in which it was most intense by the intensity in the sample in which it was least intense. For proteins that were detected in only one sample no fold change could be calculated. The percent probability signifies the outermost confidence line in Figure 2-10 that the protein fell outside.

2.4 Discussion

2.4.1 Evaluation of this technique in terms of analytical merit

A clear limitation of this technique is the fact that there was limited overlap between the differential proteins that were 'searched for' as a result of the intensity differences of the intact proteins and the identification of proteins at the peptide level. Even so, more than half (53%) of the target proteins were identified and this is in agreement with previously published literature attempting to correlate top-down and bottom-up analyses[5, 7, 8]. Some of the reasons that the overlap is so limited are the same for this technique as they are with many others. The major limiting factors include differences in selectivity of electrospray ionization of intact proteins and peptides, difficulty to detect low abundance proteins, and undocumented post-translational modifications on the intact proteins. The last limitation

could also be seen as an area for this technique to grow into. For example, if a specific PTM was known, two differential samples could be compared with an offset in the mass of one sample compared to the other corresponding to the mass difference imposed by a particular PTM. An intact protein analysis with high mass accuracy would be critical for that type of analysis. Due to the fact that undocumented PTM's, or unknown clips were excluded from the comparison in this technique, it is actually quite promising that the overlap was still greater than 50%.

2.4.2 Statistical analysis of significance of difference

For the purpose of this analysis, the statistical analysis was the best option for determining the significance of the changes in protein intensity. The replicate data of the glycerol/log phase sample appears to have adequately represented the scatter in the data caused by the method itself. Given that each intact protein 2D separation took 15 hours to complete and consumed greater than 2 mg total protein, it was not feasible to perform replicate analysis of each sample. Furthermore, in order to increase confidence that the differences in protein intensity are due to real changes caused by the different carbon sources or growth phases, an analysis of biological replicates would be ideal.

2.4.3 Biological relevance of differential proteins as compared with literature

Of the differentially expressed yeast proteins which were successfully identified, many are involved in metabolic pathways. It is logical to expect that such proteins would be well-represented in this analysis, since two of the samples were grown using different carbon sources, which would be anticipated to cause differences in metabolic processes within the yeast cells[9]. Likewise, the transition from logarithmic to stationary growth phase is

associated with a saturation of the growth medium with cells and a depletion of the readily available carbon source, which would also be expected to induce changes in metabolism.

Figure 2-11 illustrates the major metabolic pathways of yeast. The five major pathways are color-coded and include glycolysis, the pentose phosphate pathway (PPP), the citric acid cycle, the glyoxylate cycle, and anaerobic respiration. Figure 2-12 and Figure 2-13 highlight the metabolic proteins which were differentially expressed in the dextrose/glycerol and log/stationary comparisons, respectively. In Figure 2-12, it is particularly notable that numerous enzymes, including the malate dehydrogenases (MDHM and MDHC), citrate synthase (CISY1), isocitrate dehydrogenase (IDH2), and succinyl CoA ligase (SUCA), which are involved in the citric acid and glyoxylate cycles, are up-regulated in the yeast grown on glycerol media. Yeast is unable to derive energy from glycerol via fermentation to ethanol, as is possible with dextrose, and therefore must use the citric acid and glyoxylate cycles. It is therefore logical that enzymes involved in these cycles would be present in larger quantities in the glycerol sample that in dextrose. The fermentative enzymes, alcohol dehydrogenase (ADH1) and aldehyde dehydrogenase (ALDH4), were also found to be up-regulated in the glycerol-grown yeast sample. Initially, this appears to contradict the evidence that yeast cannot ferment glycerol to produce ethanol. However, because the glycerol-grown sample was initially grown on dextrose, it is possible that these enzymes were up-regulated in order to use the ethanol produced by the initial fermentation of dextrose as a secondary carbon source in addition to glycerol[10]. In the comparison between growth phases, it would be expected for proteins involved in glycolysis to be up-regulated in the log phase due to the fact that all of the growth nutrient would have been consumed in the

stationary phase. This is indeed the case for the glycolytic enzymes phosphoglycerate mutase (PMG1) and phosphoglycerol kinase (PGK).[11, 12]

Figure 2-12 also illustrates the up-regulation of proteins involved in the pentose phosphate pathway (PPP) in the glycerol-grown sample, namely transketolase (TKT1), transaldolase (TAL1), and phosphoglucose isomerase (G6PI). It has been shown that a metabolic shift from oxidative to fermentative growth causes a redirection from the PPP towards glycolysis[13]. A shift towards the PPP during the stationary phase has also been noted, explaining the up-regulation of G6PI and TAL1 in the stationary phase yeast sample as shown in Figure 2-13[14]. Some metabolic proteins are up-regulated in the dextrose sample. For example, the protein PMG1 is known to be unnecessary from growth on glycerol, which explains the fact that it was detected as being more abundant in the dextrose sample[15].

Aside from proteins directly involved in metabolic pathways of yeast, several other proteins which were detected as being differentially expressed can be associated with biochemical changes caused by growing yeast under different conditions. For instance, mitochondrial matrix factor 1 (MMF1) and its homologue HMF1, were detected as being up-regulated in the glycerol sample. These proteins are known to be essential for the growth of yeast on non-fermentable carbon sources, such as glycerol[16]. Likewise, mutants lacking aspartate aminotransferase (AATC), which is involved in amino acid biosynthesis, NADPH dehydrogenase (OYE2), and ribonucleoside diphosphate reductase (RIR4) exhibit growth defects when grown on non-fermentable carbon sources indicating that they are necessary from growth on these media[17]. The protein alanine: glyoxylate aminotransferase 1 (AGX1), which is involved in the biosynthesis of glycine, was detected as being more abundant in the

glycerol sample. This observation correlates with the fact that expression of this protein is known to be repressed when yeast is grown on a dextrose-containing medium[18].

Related to the metabolic pathways and as a direct result of the increased use of the glyoxylate cycle, serine hydroxymethyltransferase (GLYM) was up-regulated in the glycerol-grown yeast sample. GLYM uses the components of the glyoxylate cycle for the biosynthesis of serine[19]. The dextrose-grown sample also exhibited an up-regulated protein involved in amino acid biosynthesis, cystathionine gamma lyase (CYS3). A CYS3 mutant has been shown to have reduced fitness in a dextrose medium[20]. A final example is NADPH dependent methylglyoxal reductase (GRE2), which catalyzes the reduction of methyl glyoxal, a cytotoxic compound. This protein was found to be up-regulated in the stationary phase-harvested sample, which is logical given that accumulation of waste products is likely to occur when the growth medium reaches saturation with yeast cells[21]. The same logic can also be used to explain the up-regulation of the mitochondrial manganese superoxide dismutase (SODM) in the stationary phase[22].

Many of the proteins that were differentially expressed in the comparison between the log phase and stationary phase dextrose samples can be attributed to the fact that in the stationary phase, cells are no longer growing or replicating[23]. This holds true for guanylate kinase (KGUA), which is involved in the synthesis of purine and pyrimidine nucleosides, acetyl-CoA binding protein (ACBP), and tubulin specific chaperone A (TBCA), which is involved in microtubule formation[24, 25]. The 14-3-3 proteins BMH1 and BMH2 are involved in vesicular transport and cell cycle regulation, which would be expected to be up-regulated in cells that are actively growing and replicating as is shown by their up-regulation in the log phase yeast samples[26]. In the stationary phase of yeast, not only is the carbon source

completely consumed, but the ethanol produced by initial fermentation may also be consumed as a secondary carbon source. The uncharacterized protein MRP8 has been shown to be up-regulated during ethanol stress. Due to the lack of ethanol in the stationary phase, it would follow that the protein would be potentially up-regulated in the log phase, which is in agreement with what was found here[27].

Roughly one quarter of the proteins which were detected as being differentially expressed in the samples have no obvious link to changes which would be expected to occur due to the differences in the growth medium or a shift from logarithmic growth to stationary phase. The expression of these proteins could be related to metabolic changes in an indirect manner. Alternatively, it is possible that some instances in which a protein was detected as being marginally up or down regulated were due to random variation between samples rather than significant biochemical changes. Inclusion of biological replicates in future studies would improve certainty in interpretation of the data. Nevertheless, the fact that numerous well-documented changes in protein expression were detected using this method demonstrates its ability to reveal useful information when used for differential analysis of cell lysates containing water-soluble proteins.

2.5 Summary and Conclusions

A multidimensional liquid chromatography separation strategy was developed for the identification and relative quantification of differential proteins in complex mixtures. This method is based on the improved separation of proteins prior to enzymatic digestion to produce less complicated peptide mixtures that can be used to simplify the identification. Complex protein mixtures were separated using an online two-dimensional technique, which included an anion exchange separation followed by a reverse phase separation. Detection

and analysis by mass spectrometry was performed for intact protein molecular weight information. Two-dimensional differential chromatograms were constructed to identify areas in the chromatogram that contained proteins with varying expression levels. After enzymatic digestion of the proteins of interest, the resulting peptides were analyzed by tandem mass spectrometry and identified using the ProteinLynx Global Server 2.0 algorithm. Using this strategy, differential yeast proteome samples were analyzed to identify proteins that vary in intensity between samples. Through the use of a replicate analysis of a single sample the statistical significances of the differential proteins were determined. A comparison of the differential proteins to the expected differences based on the literature was made; a large percentage of the differential proteins identified at a 96.5% confidence level were validated by the literature. Overall, while the LC-based approach is still time consuming, it maintains an advantage over the gel-based approach in that the intact molecular weights of the proteins as they exist in the cell are available for further comparison and analysis. This feature was not explored fully in this experiment, but may be essential in future, more targeted analyses.

2.6 Acknowledgement

I would like to acknowledge the guidance and work with Dr. Charles Evans, a fellow graduate student with whom I worked closely with on this experiment.

2.7 References

(1) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Beker, C. H. *Analytical Chemistry* **2003**, *75*, 4818-4826.

(2) Evans, C. R., University of North Carolina at Chapel Hill, Chapel Hill, 2007.

(3) Ghaemmaghami, S.; Huh, W.-K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. *Nature Biotechnology* **2003**, *425*, 737-741.

(4) Kolkman, A.; Daran-Lapujade, P.; Fullaondo, A.; Olsthoorn, M. M.; Pronk, J. T.; Slijper, M.; Heck, A. J. *Molecular Systems Biology* **2006**, *2*, 1-16.

(5) Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(6) Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J. *Rapid Communications in Mass Spectrometry* **1992**, *6*, 707-711.

(7) Hamler, R. L.; Zhu, K.; Buchanan, N. S.; Kreunin, P.; Kachman, M. T.; Miller, F. R.; Lubman, D. M. *Proteomics* **2004**, *4*, 562-577.

(8) Kreunin, P.; Urquidi, V.; Lubman, D. M.; Goodison, S. *Proteomics* **2004**, *4*, 2754-2765.

(9) Fraenkel, D. G. In *The Molecular Biology of the Yeast Saccharomyces*; Strathern, J., Jones, E. W., Broach, J. R., Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1982, pp 1-37.

(10) Schuller, H.-J. *Curr Genet FIELD Full Journal Title:Current genetics* **2003**, *43*, 139-160.

(11) Van Doorn, J.; Valkenburg, J. A. C.; Scholte, M. E.; Oehlen, L. J. W. M.; Van Driel, R.; Postma, P. W.; Nanninga, N.; Van Dam, K. *Journal of Bacteriology* **1988**, *170*, 4808-4815.

(12) Ono, B.; Tanaka, K.; Naito, K.; Heike, C.; Shinoda, S.; Yamamoto, S.; Ohmori, S.; Oshima, T.; Tohe, A. *Journal of Bacteriology* **1992**, *174*, 3339-3347.

(13) Frick, O.; Wittmann, C. *Microbial Cell Factories* **2005**, *4*, No pp. given.

(14) Lee, J.; Godon, C.; Lagniel, G.; Spector, D.; Garin, J.; Labarre, J.; Toledano, M. B. *Journal of Biological Chemistry* **1999**, *274*, 16040-16046.

(15) Rodicio, R.; Heinisch, J. J.; Hollenberg, C. P. *Gene* **1993**, *125*, 125-133.

(16) Kim, J.-M.; Yoshikawa, H.; Shirahige, K. *Genes to Cells* **2001**, *6*, 507-517.

(17) Steinmetz, L. M.; Scharfe, C.; Deutschbauer, A. M.; Mokranjac, D.; Herman, Z. S.; Jones, T.; Chu, A. M.; Giaever, G.; Prokisch, H.; Oefner, P. J.; Davis, R. W. *Nature Genetics* **2002**, *31*, 400-404.

(18) Schlosser, T.; Gatgens, C.; Weber, U.; Stahmann, K.-P. *Yeast* **2003**, *21*, 63-73.

(19) Ulane, R.; Ogur, M. *Journal of Bacteriology* **1972**, *109*, 34-43.

(20) Deutschbauer, A. M.; Jaramillo, D. F.; Proctor, M.; Kumm, J.; Hillenmeyer, M. E.; Davis, R. W.; Nislow, C.; Giaever, G. *Genetics* **2005**, *169*, 1915-1925.

(21) Garay-Arroyo, A.; Covarrubias, A. A. *Yeast* **1999**, *15*, 879-892.

(22) Ditlow, C.; Johansen, J. T. *Carlsberg Research Communications* **1982**, *47*, 71-79.

(23) Werner-Washburne, M.; Braun, E.; Johnston, G. C.; Singer, R. A. *Microbiological Reviews* **1993**, *57*, 383-401.

(24) Konrad, M. *Journal of Biological Chemistry* **1992**, *267*, 25652-25655.

(25) Lopez-Fanarraga, M.; Avila, J.; Guasch, A.; Coll, M.; Zabala, J. C. *J Struct Biol FIELD Full Journal Title:Journal of structural biology* **2001**, *135*, 219-229.

(26) Fu, H.; Subramanian, R. R.; Masters, S. C. *Annu Rev Pharmacol Toxicol FIELD Full Journal Title:Annual review of pharmacology and toxicology* **2000**, *40*, 617-647.

(27) Zhou, W.; Ryan, J. J.; Zhou, H. *Journal of Biological Chemistry* **2004**, *279*, 32262-32268.

2.8 Tables

Time (min)	% mobile phase B
0	0
20	7
600	67
675	67
680	0

Table 2-1 : AXC gradient used for the on-line intact protein separation. Mobile phase A consisted of 10 mM ammonium acetate adjusted to pH 9 with ammonium hydroxide. Mobile phase B consisted of 750 mM ammonium acetate adjusted to pH 9 with ammonium hydroxide. The flow rate for the separation was 0.2 mL/min.

Time (min)	% mobile phase B
0	5
2	20
5	30
15	45
15.5	90
16	90
16.5	5

Table 2-2 : RP gradient used for the second dimension of the on-line intact protein separation. Mobile phase A consisted of water with 0.2% formic acid and mobile phase B consisted of acetonitrile with 0.2% formic acid. The flow rate for the separation was 1 mL/min.

Time (min)	% mobile phase B
0	1
20	35
20.5	65
22	65
22.5	1

Table 2-3 : Gradient profile for the capillary RPLC separation of the digested protein fractions. Mobile phase A consisted of water with 0.2% formic acid and mobile phase B consisted of acetonitrile with 0.2% formic acid. The flow rate for the gradient was 10 µL/min which was split down to 500nL/min prior to the injection valve.

From Mass	To Mass	CE 1	CE 2	CE 3
400	500	26	22	31
500	600	28	24	33
600	700	30	26	35
700	800	32	28	37
800	900	35	30	39
900	1600	37	32	42

Table 2-4 : Collision energy (CE) profile used in the MS/MS data-directed analysis of the digested protein fractions from the on-line 2D intact protein separation.

Processing	MaxEnt 1
Chromatogram	
Scan interval	5
Scan range	100-800
TIC threshold	20,000 counts
Raw data processing	
Background subtraction	None
Smooth	None
Center	None
MaxEnt 1 parameters	
Input mass range	600 – 1600 m/z
Output mass range	5000 – 80,000 Da
Mass resolution	2 Da
Damage model	Gaussian: $w_{1/2}$ = 0.75 Da; 33% either side
Completion option	Maximum 50 iterations
Post processing	
Subtraction	None
Lockmass calibration	None
De-harmonization	20 Da

Table 2-5: AutoME processing parameters for deconvolution of protein mass spectra.

Processing Parameters	**Value**
Mass Accuracy	No lockspray calibration
Noise reduction (electrospray survey and MS/MS)	Adaptive
Perform de-isotoping	Yes
Deisotoping type	Slow (MaxEnt 3)
Maximum iterations	40
Workflow Template	**Value**
Search engine type	PLGS
Databanks	Yeast proteins and trypsin with a 1x randomization
Species	None specified
Peptide tolerance	100 ppm
Fragment tolerance	0.1 Da
Estimated calibration error	0.005 Da
Molecular weight range	0-100,000 Da
pI range	0-14
Minimum peptides to match	1
Maximum hits	20
Primary digest reagent	Trypsin
Secondary digest reagent	None
Missed cleavages	2
Fixed modifications	None
Variable modifications	Oxidation M, Carbamidomethyl C
Validate	Yes
Filter	None

Table 2-6: Parameters used to process the MS/MS runs of the digested protein fractions using ProteinLynx Global Server 2.3 (Waters).

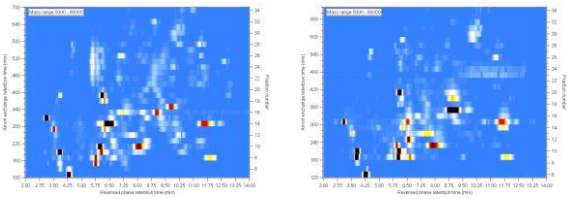
Swiss-Prot Name	Ordered Locus name	Description	Pep. Hits	Intact Mass	Up-reg in	Fold change	% Prob. Diff.
TPIS	YDR050C	Triosephosphate isomerase	4	26666.3	Dex	1.7	99.5
YCO26	YCL026C-B	Uncharacterized protein YLC026C B	2	20906.2	Dex	9.3	99.5
MRS1	YIR021W	RNA splicing protein MRS1	1	41078.3	Dex	14	99.5
PMG1	YKL152C	Phosphoglycerate mutase 1	8	27480.7	Dex	2.0	98.7
CYS3	YAL012W	Cystathionine gamma lyase	5	42414.4	Dex	-	98.7
GLRX1	YCLO35C	Glutaredoxin 1	4	12245.7	Dex	22	98.7
BMH1	YER177W	Protein BMH1	10	30004.2	Dex	1.8	98.0
SODC	YJR104V	Superoxide dismutase Cu Zn	1	15722.9	Dex	1.9	96.5
OYE2	YHR179W	NADPH dehydrogenase 2	7	44884.2	Dex	9.6	96.5
TRX2	YGR209C	Thioredoxin 2	3	11202.1	Dex	-	96.5
ACBP	YGR037C	Acyl CoA binding protein	1	9930.17	Dex	1.5	96.5
MDHM	YKL085W	Malate dehydrogenase mitochondrial	5	33836.2	Gly	6.3	99.5
ADH1	YOL086C	Alcohol dehydrogenase	1	36646.2	Gly	-	99.5
ALDH4	YOR374W	Aldehyde dehydrogenase	15	53979.1	Gly	-	99.5
SUCA	YOR142W	Succinyl CoA ligase	4	33222.7	Gly	15.4	98.7
IPB2	YNL015W	Protease B inhibitors 2 and 1	2	8459.6	Gly	8.6	98.7
AATC	YLR027C	Aspartate aminotransferase	5	45893.4	Gly	7.5	98.0
TKT1	YPR074C	Transketolase 1	8	73677.7	Gly	-	98.0
YM71	YMR226C	Unchar. Oxidoreductase YMR226C	6	29071.6	Gly	5.3	98.0
TAL1	YLR354C	Transaldolase	6	36951.3	Gly	11	98.0
CISY1	YNR001C	Citrate synthase mitochondrial	7	49221.9	Gly	36	98.0
MDHC	YOL126C	Malate dehydrogenase cytoplasmic	2	40604.2	Gly	-	98.0
HMF1	YER057C	Protein HMF1	1	13775.5	Gly	4.7	98.0
AGX1	YFL030W	Alanine glyoxylate aminotransferase	6	41778.4	Gly	-	96.5
GLYM	YBR263W	Serine hydroxymethyltransferase	3	51605.7	Gly	-	96.5
IDH2	YOR136W	Isocitrate dehydrogenase NAD	3	37802.1	Gly	-	96.5
PROF	YOR122C	Profilin	1	13589.5	Gly	12	96.5
G6PI	YBR196C	Glucose 6 phosphate isomerase	12	61216.9	Gly	1.3	96.5
RIR4	YGR180C	Ribonuleoside diphosphate reductase	4	40130.7	Gly	5.1	96.5
BMH2	YDR099W	Protein BMH2	10	30974.2	Gly	1.6	96.5

Table 2-7 : List of proteins determined to have significant intensity differences between the dextrose-grown and glycerol-grown yeast samples. **Swiss-Prot Name**: Protein entry in SwissProt database followed by '_YEAST'. **Ordered Locus Name**: Predicted gene that encodes the protein sequence. **Description**: Brief description of the protein. **Pep. Hits**: Number of peptides hits used to identify the protein in PLGS2.3. **Intact Mass**: AutoME-deconvoluted molecular weight of the intact protein. **Up-reg in**: The sample in which the protein was most intense. **Fold-change**: The degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense. The absence of a fold change signifies the protein was only present in one sample. **%Prob. Diff**: Confidence of the significance of the difference as determined through the analysis of replicate data

Swiss-Prot Name	Ordered Locus name	Description	Pep. Hits	Intact Mass	Up-reg in	Fold change	% Prob. Diff.
BMH2	YDR099W	Protein BMH2	15	30968.7	Log	-	99.5
PMG1	YKL152C	Phosphoglycerate mutase 1	8	27477.9	Log	2	99.5
KGUA	YDR454C	Guanylate kinase	1	20906.2	Log	7.1	99.5
PGK	YCR012W	Phosphoglycerate kinase	16	44652.7	Log	-	99.5
G3P3	YGR192C	Glyceraldehyde 3 phosphate dehydrogenase 3	11	32617.8	Log	122.6	99.5
HSP12	YFL014W	12 kDa heat shock protein	5	11604.6	Log	373.1	99.5
CH10	YOR020C	10 kDa heat shock protein mitochondrial	3	11283.9	Log	29.5	98.7
ACBP	YGR037C	Acyl CoA binding protein	1	9930.17	Log	2.5	98.7
BMH1	YDR099W	Protein BMH1	10	30004.2	Log	-	98.7
CYS3	YAL012W	Cystathionine gamma lyase	5	42414.4	Log	18.5	98.0
SODC	YJR104C	Superoxide dismutase Cu Zn	1	15722.9	Log	1.8	96.5
TBCA	YOR265W	Tubulin specific chaperone A	3	12248.4	Log	-	96.5
PNC1	YGL037C	Nicotinamidase	2	24993.9	Log	17.9	96.5
MRP8	YKL142W	Uncharacterized protein MRP8	6	25004.8	Log	-	96.5
G6PI	YBR196C	Glucose 6 phosphate isomerase	24	61202.9	Stat	1.7	99.5
DHOM	YJR139C	Homoserine dehydrogenase	10	38404.9	Stat	2.2	99.5
COFI	YLL050C	Cofilin	12	15809.9	Stat	1.9	98.7
YL364	YLR364W	Glutaredoxin like protein YPR364W	2	12066.0	Stat	2.0	98.7
MRS1	YIR021W	Mitochondrial RNA splicing protein MRS1	1	41070.4	Stat	1.4	98.0
G3P1	YJL052W	Glyceraldehyde 3 phosphate dehydrogenase 1	11	35742.1	Stat	-	98.0
SODM	YHR008C	Superoxide dismutase Mn	4	23083.0	Stat	3.7	98.0
TAL1	YLR354C	Transaldolase	5	36942.8	Stat	7.4	96.5

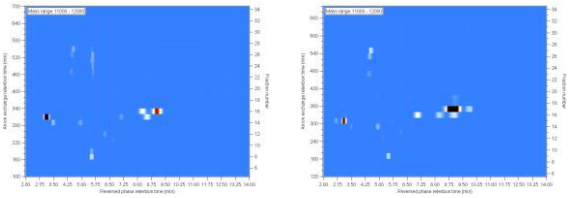
Table 2-8: List of differential proteins from the comparison between yeast samples harvested at either the log phase or stationary phase of growth. Only proteins with differences significant at greater than 96.5% are included. **Swiss-Prot Name**: Protein entry in SwissProt database followed by '_YEAST'. **Ordered Locus Name**: Predicted gene that encodes the protein sequence. **Description**: Brief description of the protein. **Pep. Hits**: Number of peptides hits used to identify the protein in PLGS2.3. **Intact Mass**: AutoME-deconvoluted molecular weight of the intact protein. **Up-reg in**: The sample in which the protein was most intense. **Fold-change**: The degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense. The absence of a fold change signifies the protein was only present in one sample. **%Prob. Diff**: Confidence of the significance of the difference as determined through the analysis of replicate data

2.9 Figures



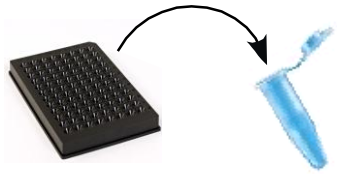
Sample A Sample B

2DLC-MS of intact proteins for all differential samples

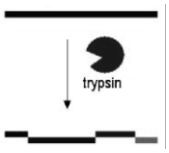


Sample A 11 - 12 kDa Sample B 11 - 12 kDa

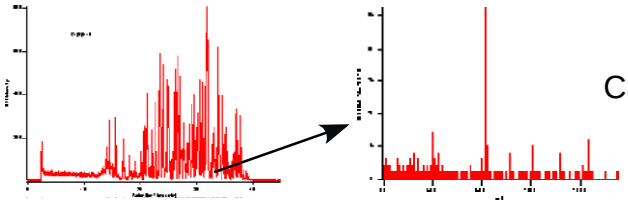
Visual comparion of mass slice chromatograms to identify differences in protein intensity



Selection of fractions

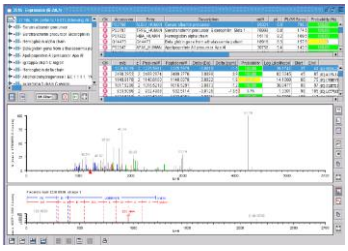


Trypsin digestion



LC chromatogram MS/MS spectrum

Capillary RPLC-MS/MS of peptides



Database searching for protein identification

Figure 2-1: Experimental workflow for on-line differential analysis of intact proteins.

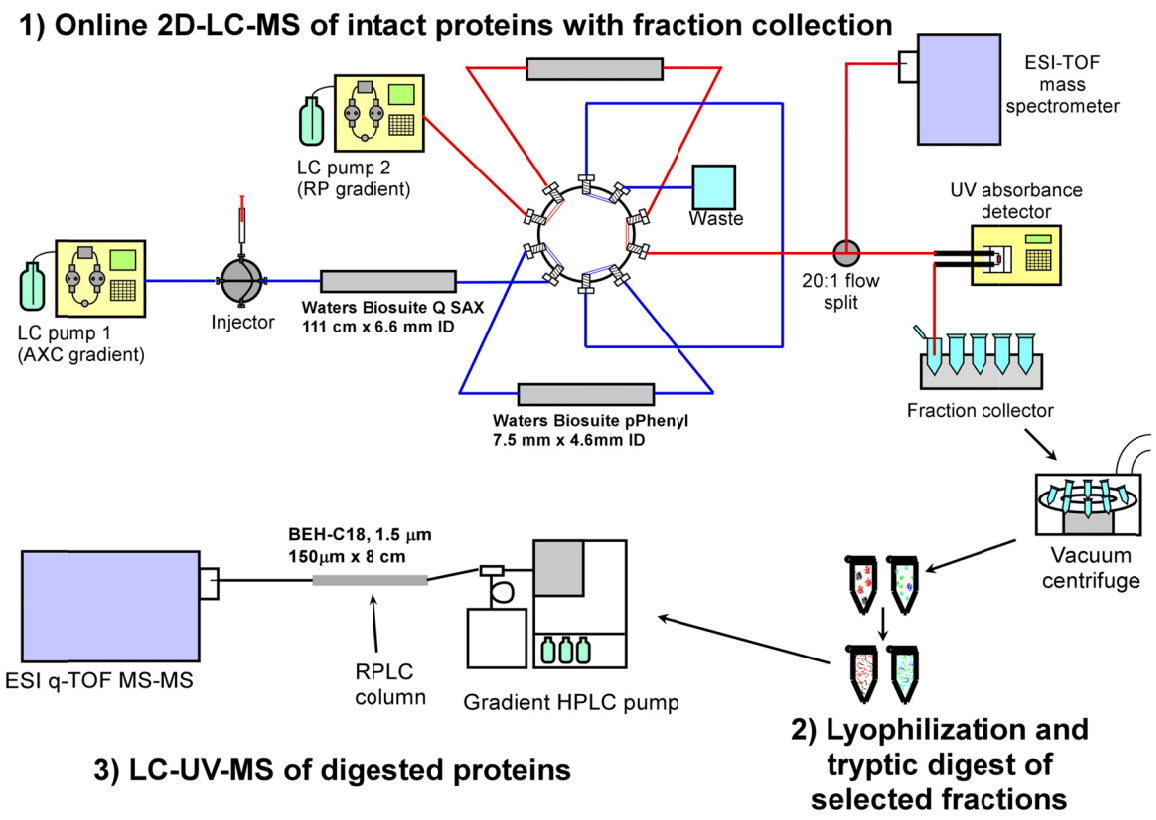


Figure 2-2 : Instrumentation diagram of the complete on-line 2D intact protein separation and RPLC-MS/MS of digested protein fractions.

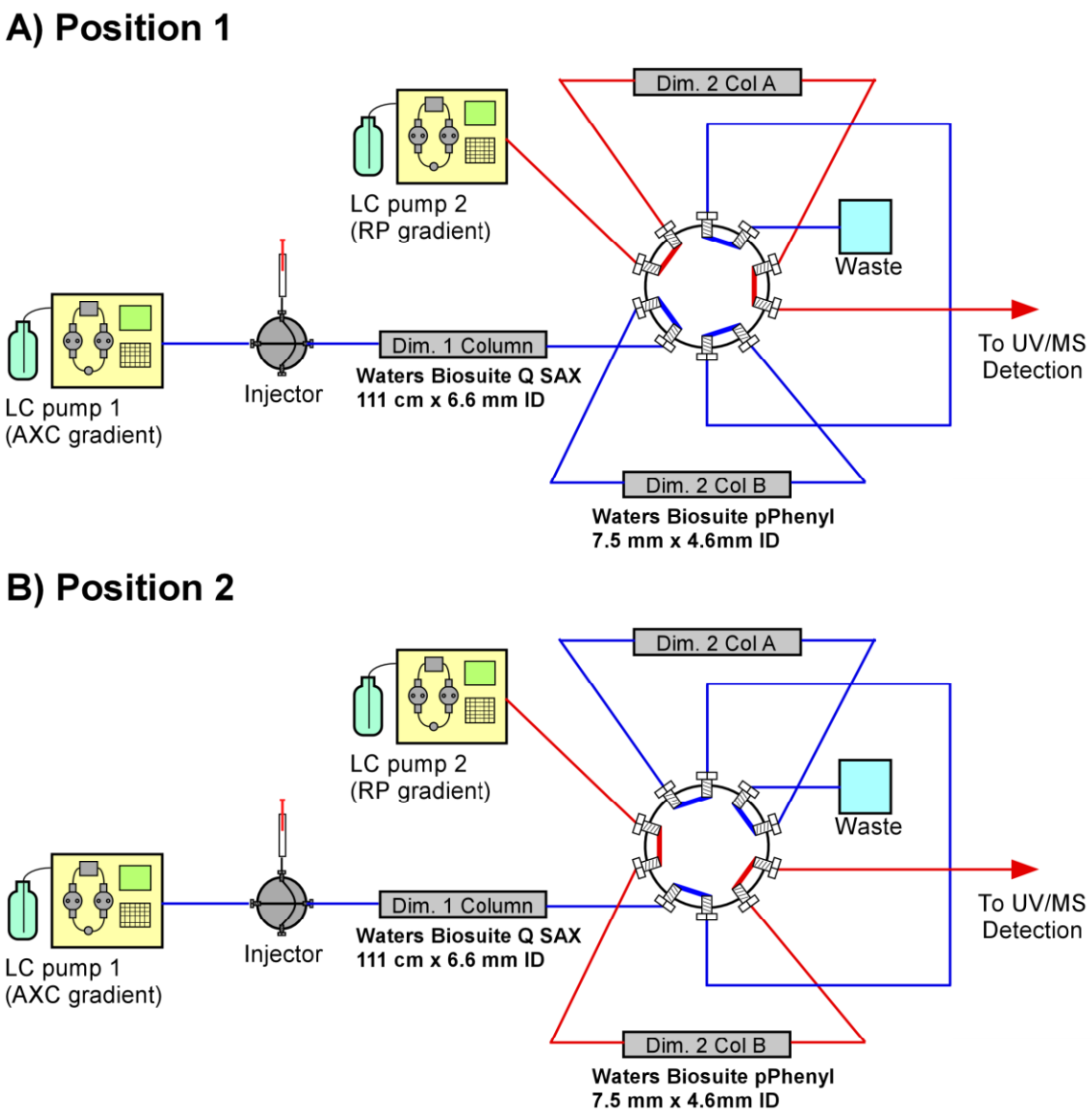


Figure 2-3 : Illustration outlining the fluidic pathways made by switching the 10-port valve from A) Position 1 to B) Position 2.

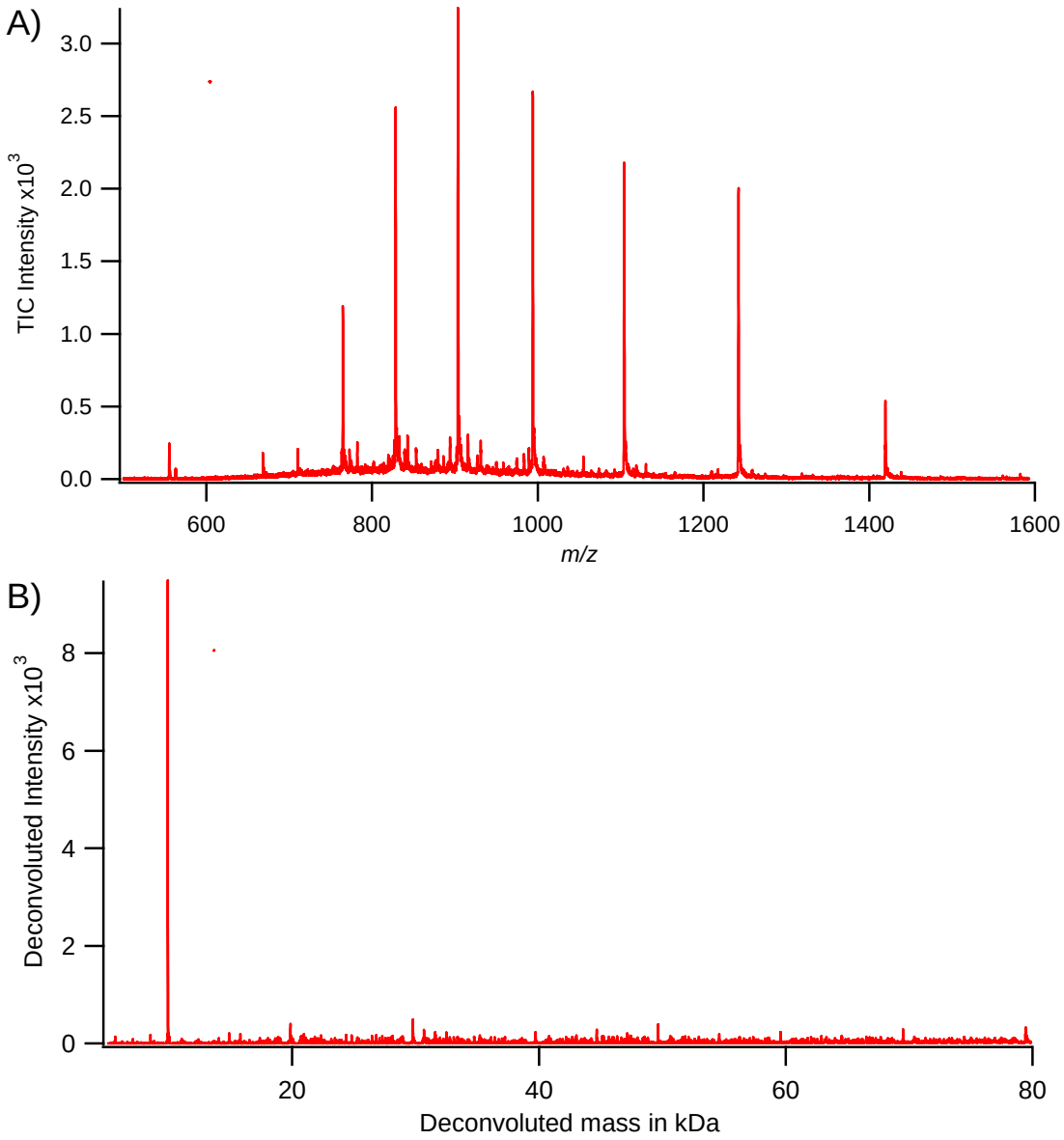
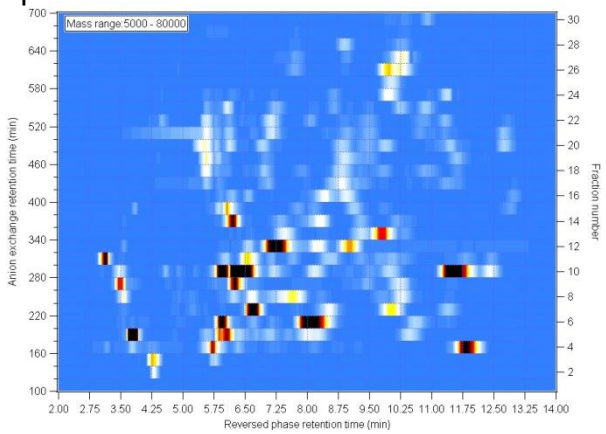
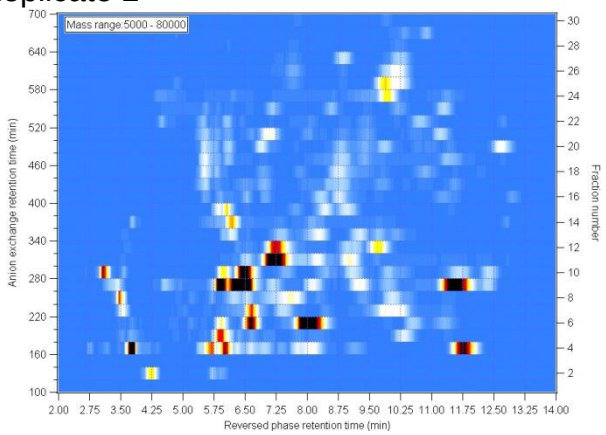


Figure 2-4: Example of AutoME deconvolution of intact protein mass spectra. A) Raw mass spectrum obtained from the online LC-LC-MS analysis of intact proteins. B) Deconvoluted masses spectrum illustrating the summation of the protein charge envelope into a single protein mass.

Replicate 1



Replicate 2



Replicate 3

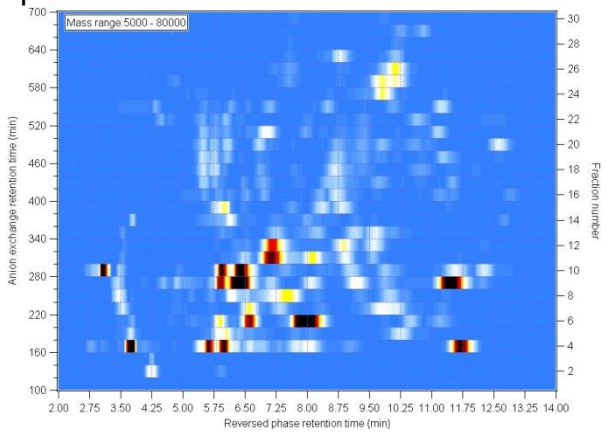


Figure 2-5 : 2D chromatograms of Auto-ME deconvoluted intact protein mass spectra. Each replicate injection was a 2.25 mg of the glycerol-grown yeast sample.

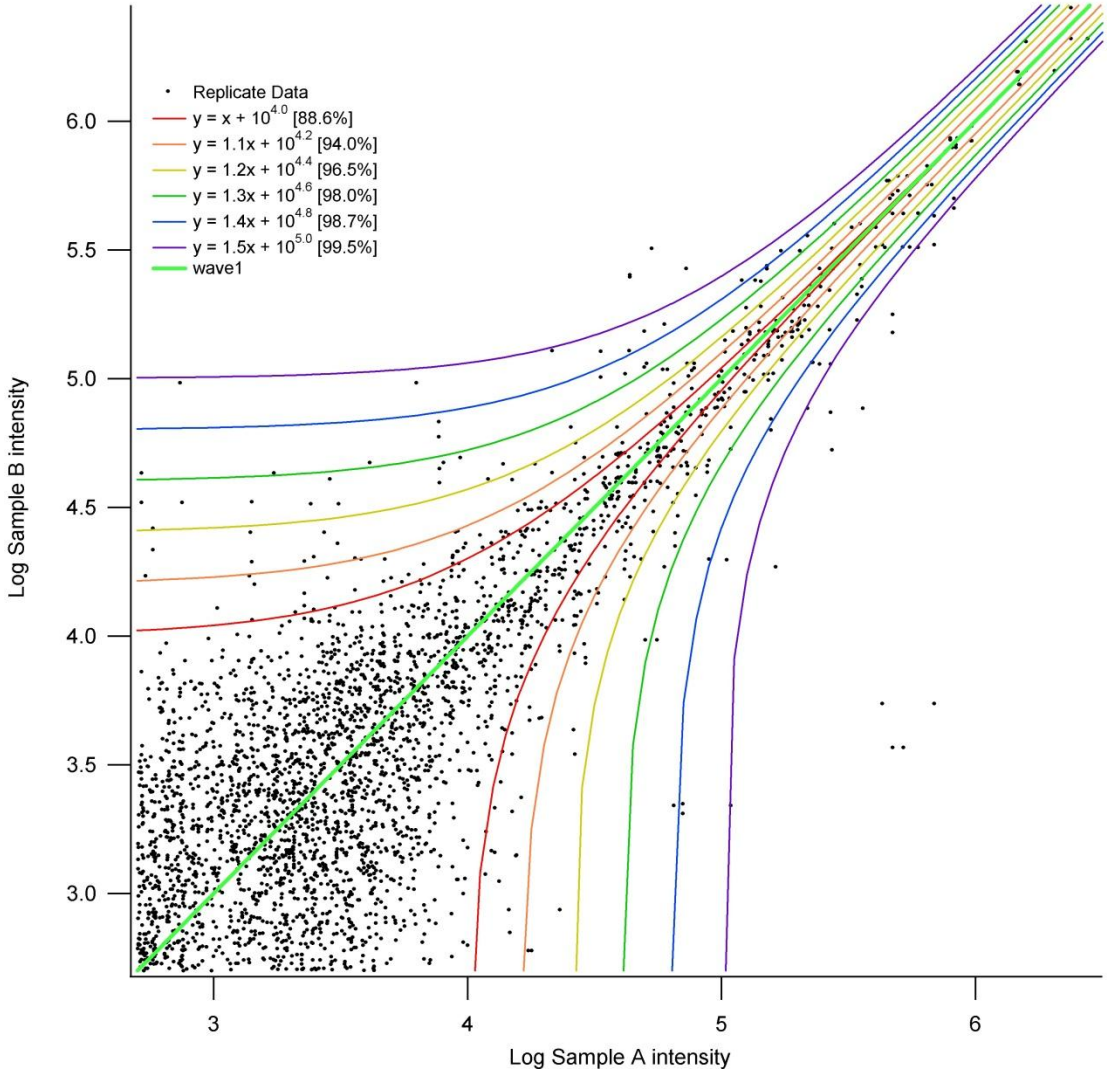


Figure 2-6 : Log-log intensity plot of the triplicate analysis of lysates of yeast cells grown on glycerol harvested at the log phase of growth. The lines indicated confidence intervals based on the percentage of data points encapsulated by the curves.

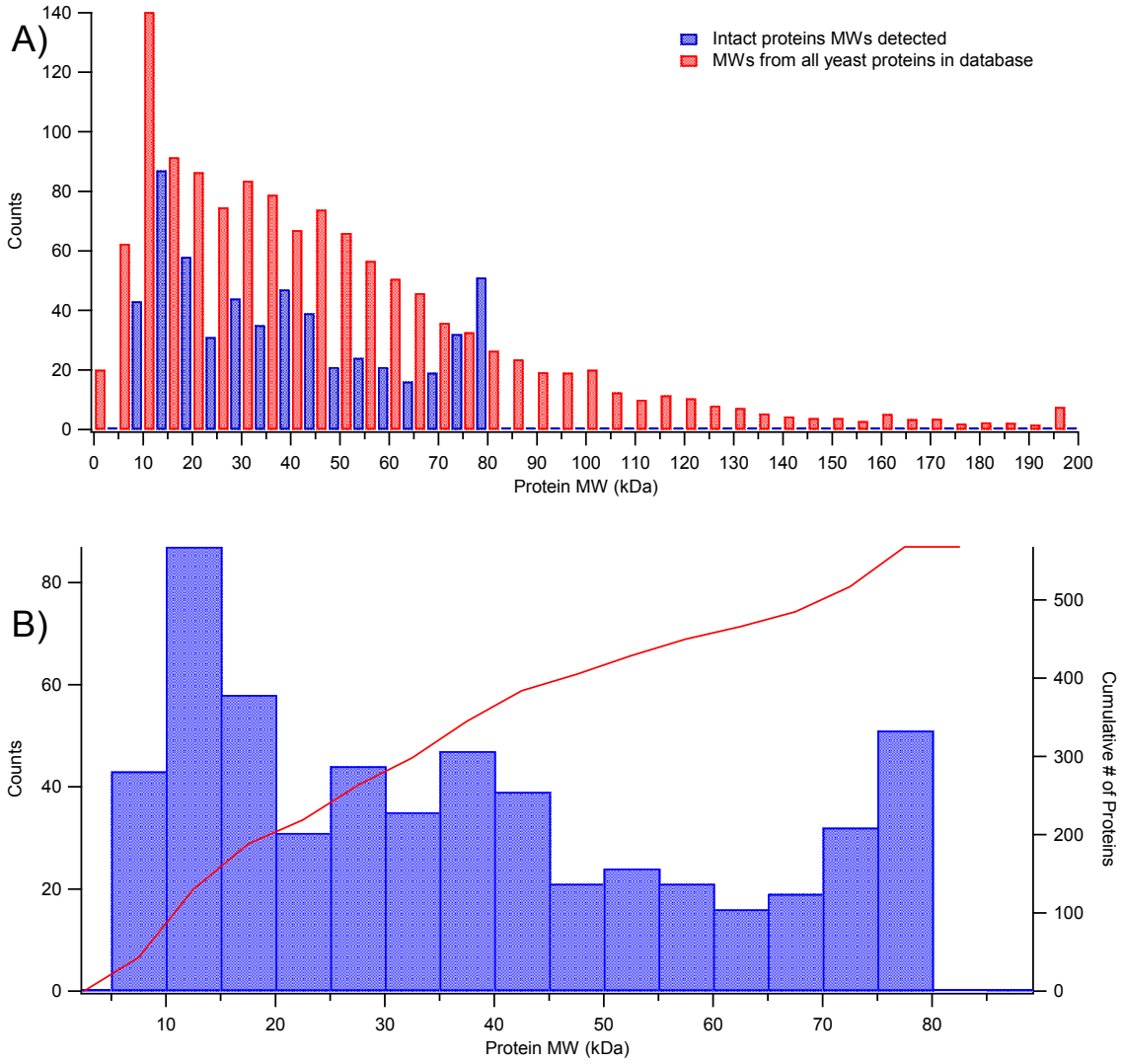


Figure 2-7: Distribution of protein molecular weights detected in the online LC-LC-MS analysis of proteins. A) Molecular weight distribution as it compares to the distribution of all of the yeast proteins in the database. B) Molecular weight distribution with the cumulative number of detected molecular weights appended to the plot.

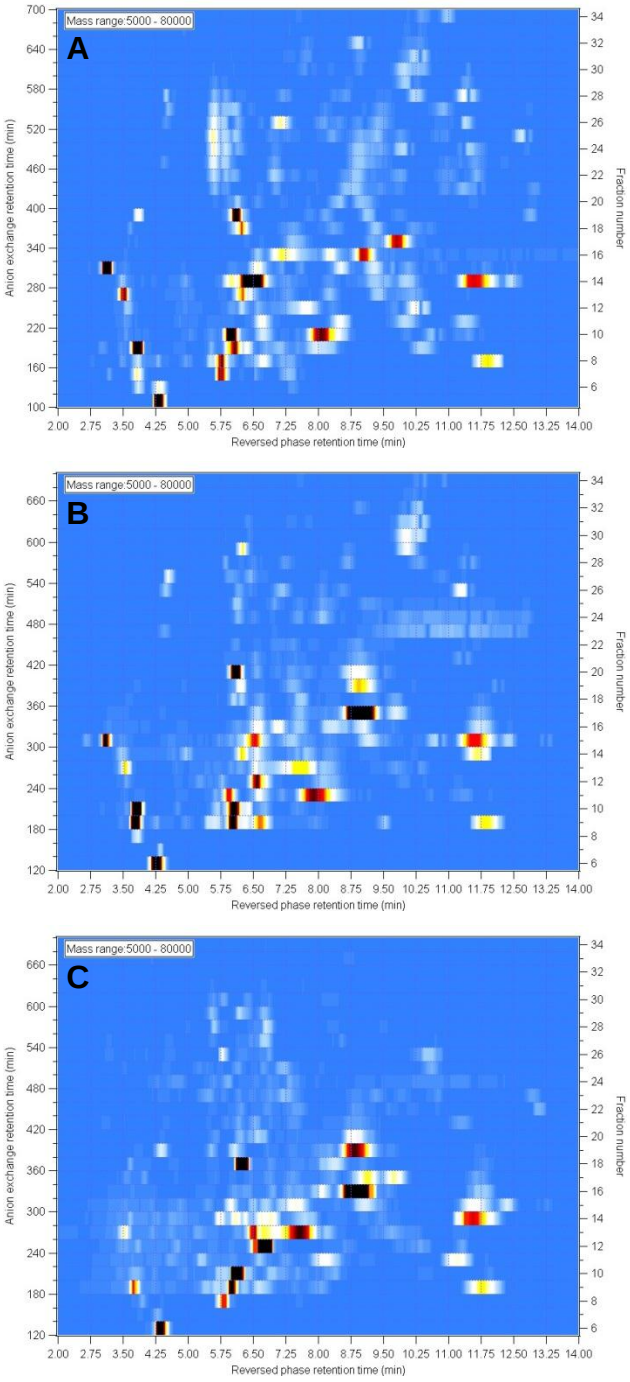
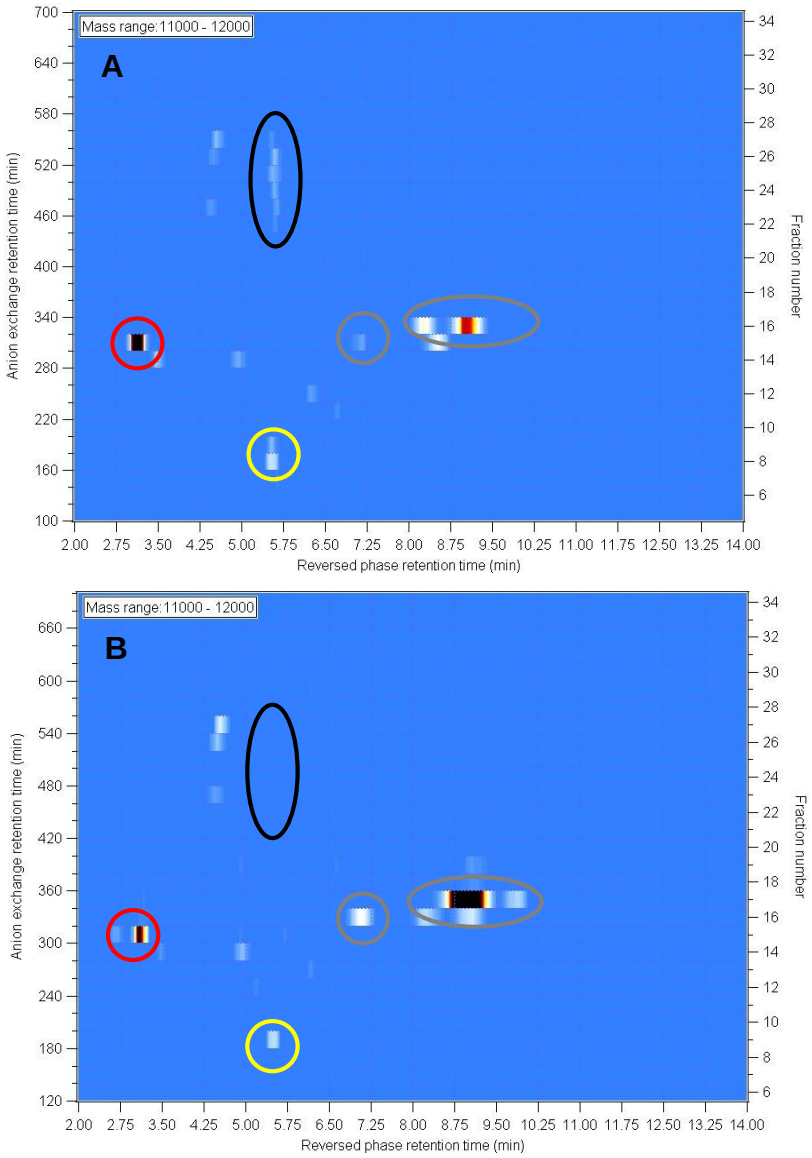


Figure 2-8 : 2D chromatograms of the on-line intact protein separations of the differential yeast samples. A) Glycerol-grown harvested during the logarithmic growth phase. B) Dextrose-grown harvested during the logarithmic growth phase. C) Dextrose-grown harvested during the stationary growth phase. All injections were 2.25mg of total protein as determined by the Bradford assay with a BSA standard.



Key:

- Consistent Intensity
- Up-regulated in Glycerol sample
- Down-regulated in Glycerol sample
- Present in only one sample

Figure 2-9 : Representative comparison of mass slice chromatograms for visual identification of differential proteins. The mass range of proteins included in the plots was limited to 11 to 12 kDa. A) Glycerol/log phase sample. B) Dextrose/log phase sample.

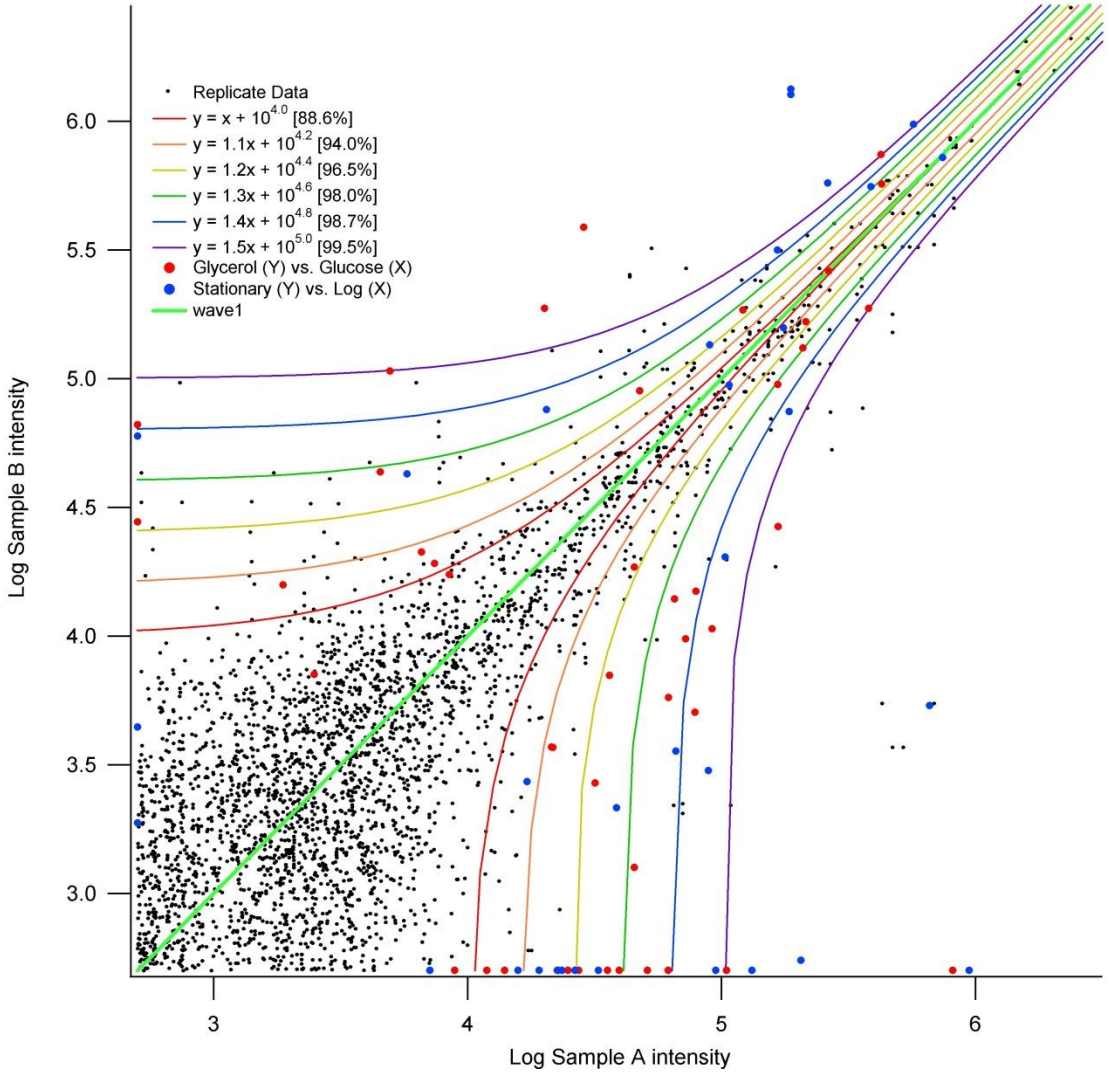


Figure 2-10 : Addition of differential proteins identified by visual inspection to the log-log plot of the replicate data. Differential proteins which were detected in only one of the samples were plotted at the detection limit (500) of the sample in which it was not detected.

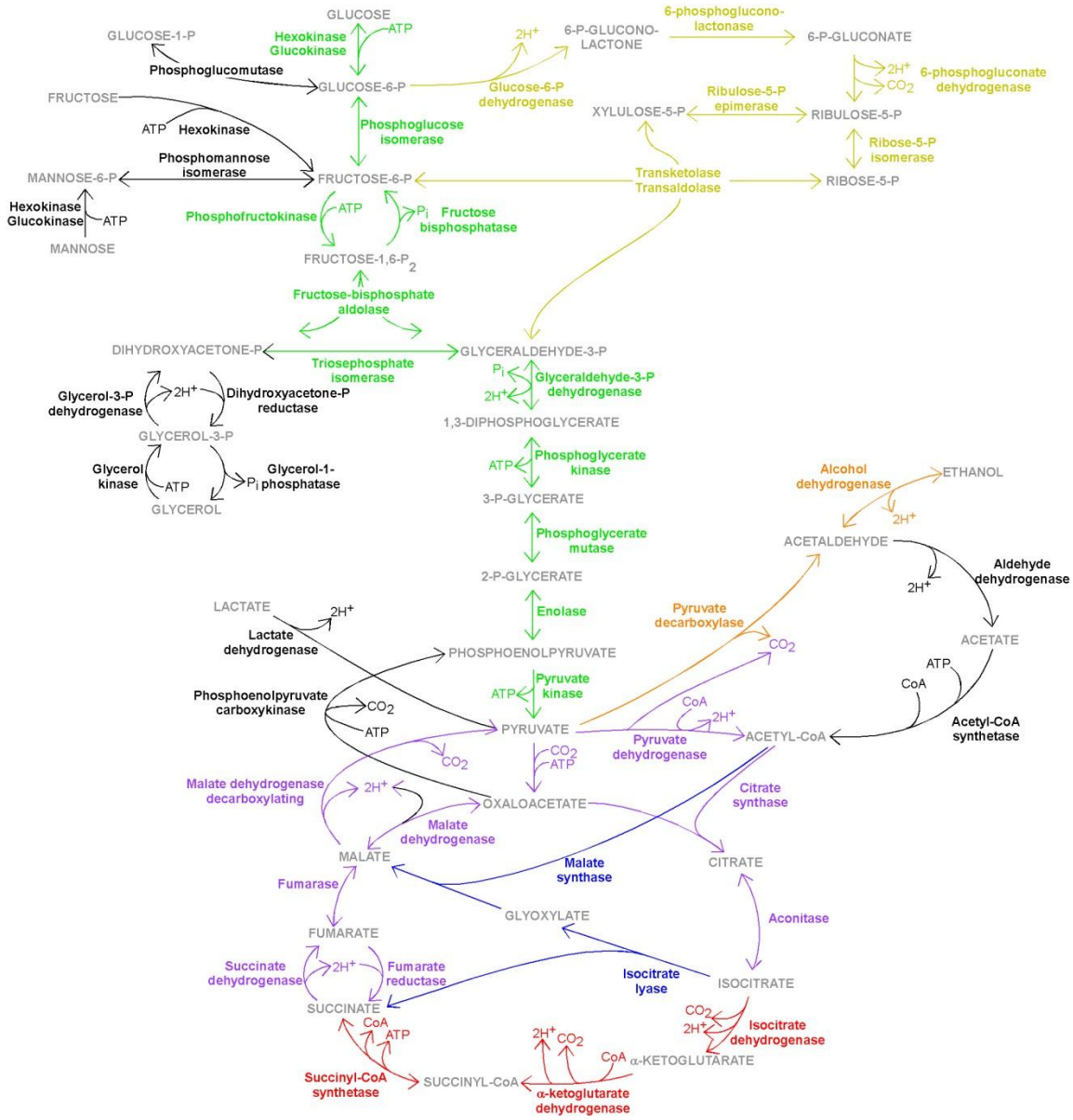


Figure 2-11: Metabolic pathways of *S. cerevisiae*. The following pathways are highlighted: Glycolysis in green, the pentose phosphate pathway in yellow, anaerobic respiration in orange, the citric acid cycle in red, and the glyoxylate cycle in blue. Proteins shared between the citric acid and glyoxylate cycles are highlighted in purple.

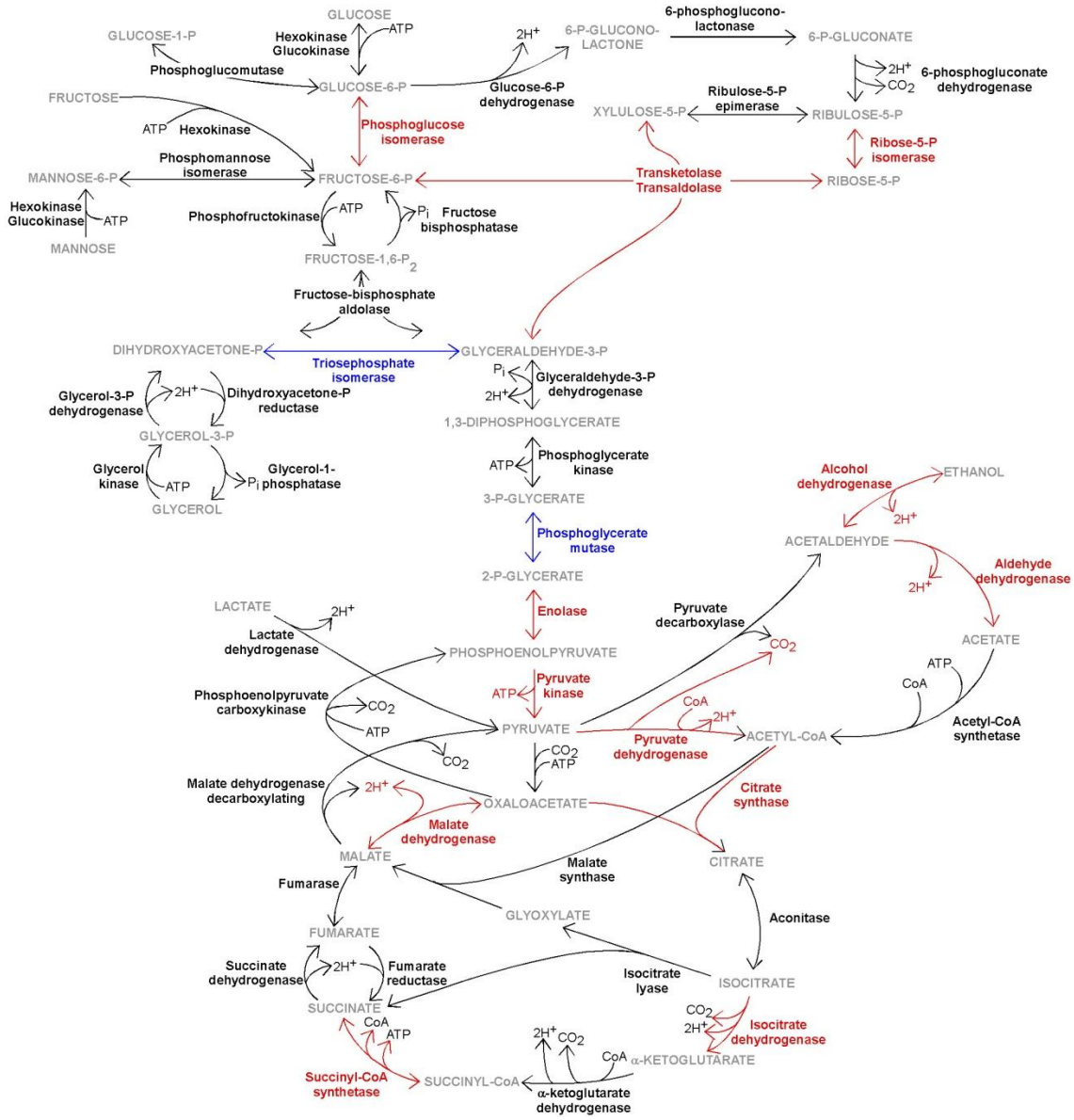


Figure 2-12 : Differential proteins identified in the comparison between *S. cerevisiae* cell lysates. Proteins determined to be up-regulated in the glycerol grown sample are highlighted in red, while those found to be up-regulated in dextrose-grown cells are highlighted in blue.

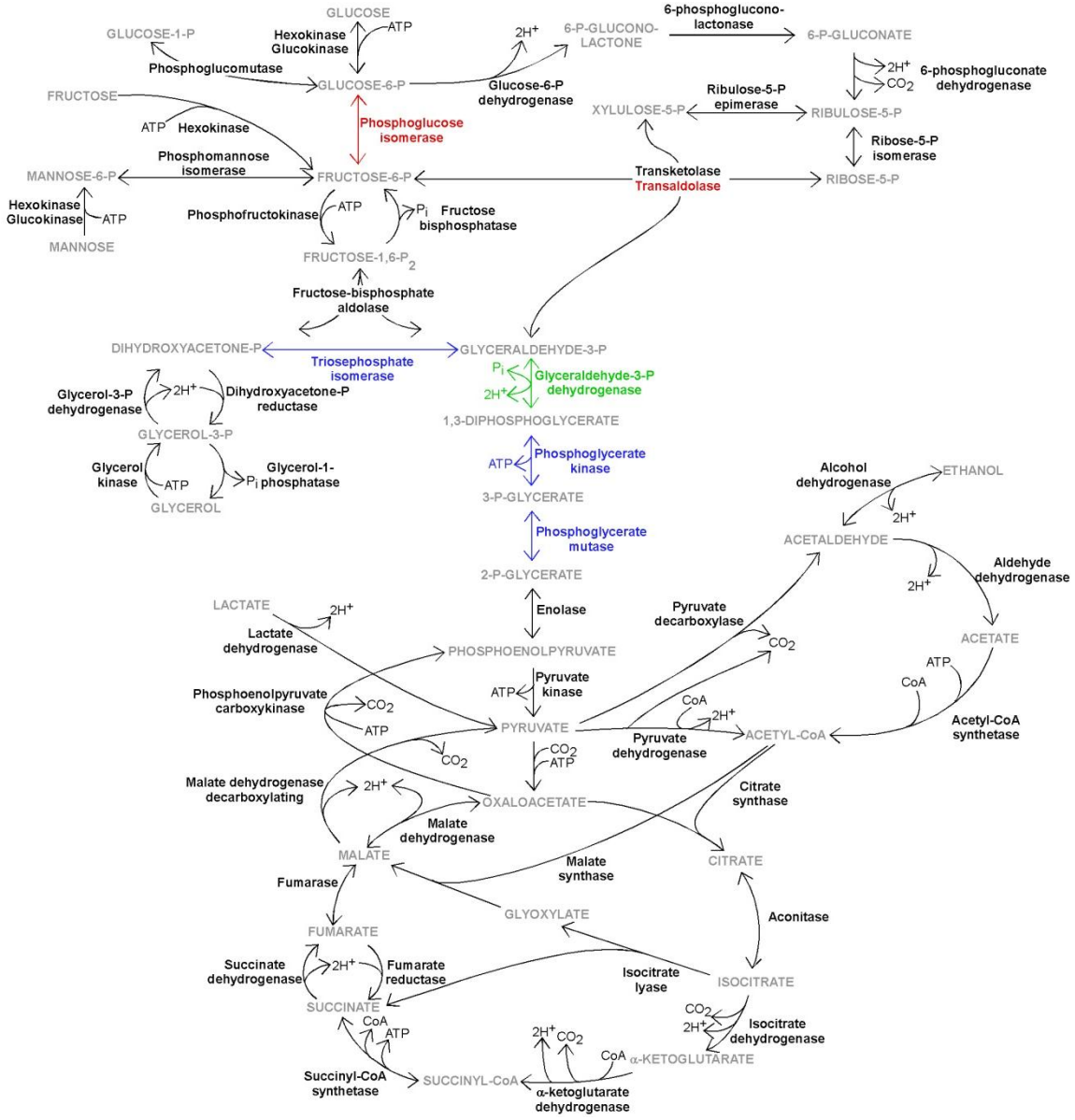


Figure 2-13 : Differential proteins identified based on changes in the growth phase at which cell harvesting was performed. Protein highlighted in red were significantly more abundant in yeast cells harvested at the stationary growth phase, while proteins highlighted in blue were up-regulated in the sample harvested during the logarithmic growth phase. For one enzyme, highlighted in green, there were two isozymes identified; one was up-regulated in the log phase sample and one was up-regulated in the stationary phase sample.

.

CHAPTER 3: Traditional bottom-up analysis of S. cerevisiae cell lysates grown under various growth conditions without pre-fractionation

3.1 Introduction

3.1.1 Bottom-up proteomics for use in a differential proteomic experiment

Bottom-up (BU) proteomic analysis is currently the most widely-used approach to mass spectrometry (MS) based experiments. It involves the introduction of enzymatically digested proteins into the mass spectrometer as opposed to top-down (TD) proteomics in which intact proteins are analyzed directly by MS. Peptides are more easily separated than intact proteins and tandem mass spectrometry data is commonly used to unambiguously identify proteins based on database searching. Historically, it has been favored over top-down proteomics due to the following shortcomings of a TD approach: (1) the complexity of the fragmented intact protein spectra limit analysis to isolated proteins; (2) the required high resolution instruments (Fourier transform- ion cyclotron resonance (FT-ICR), hybrid ion trap FT-ICR, or hybrid ion trap-orbitrap) are costly; (3) it tends to be biased towards proteins less than 50 kDa; (4) the favored dissociation techniques (electron-transfer dissociation, ETD, and electron-capture dissociation, ECD) are slow processes not necessarily amenable to on-line coupling to a liquid chromatography (LC) separation; (5) the mechanism of dissociation is not as well understood as it is in BU proteomics limiting the availability of searchable databases and bioinformatic tools.[1] Conversely, the main drawback of a BU experiment is that the molecular weight of the intact species is never determined, limiting the identification

of post-translational modifications that are often lost during the collision-induced dissociation (CID) of the peptide fragments. In spite of this, both the availability of advanced LC/MS instrumentation and the ever-improving development of software have continued to make BU proteomics the method of choice.

3.1.2 MS^E peptide analysis

In BU proteomics, there are two main tandem mass spectrometric methods employed in label-free experiments. These include a data dependent acquisition, DDA, and a data independent acquisition, DIA. In the experiment described in Chapter 2, MS/MS data acquisition was performed in a data dependent manner through a repeating cycle of one survey scan followed by a series of fragment ion scans of selected *m/z* values at higher collision energies. As discussed previously, only the two most abundant peptides in a survey scan are subsequently fragmented in the product ion scans before another survey scan is performed. This serial nature of this analysis increases the likelihood of missing less abundant peptides or peptides that elute during the higher collision energy scans. Also, there is no guarantee that fragmentation will occur at or near the chromatographic apex, the point at which the intensity of fragment ions would likely be highest. In order to limit the probability of missing peptides and increase sequence coverage, it is common practice for samples to be run multiple times and the data merged prior to processing. This can be time consuming since the multiple runs performed are not merely replicate analyses that can be used to evaluate reproducibility but rather complementary analyses that, when combined, give a more comprehensive analysis.

Recently, a new form of MS/MS data independent acquisition has been designed, MS^E data acquisition.[2] This form of analysis is parallel in nature as opposed to the serial

DDA experiments in which intense peptides are sequentially fragmented and analyzed. In an experiment of this form, accurate mass LC-MS data are collected in alternating low collision energy (MS) and elevated collision energy (MS^E) scans. Analysis is performed on a quadrupole time-of-flight (Q-TOF) mass spectrometer with the quadrupole essentially acting as another ion guide, passing all ions through to the collision cell in both MS and MS^E scans. The fragmentation of all species present in a low energy scan produces highly complex fragmentation spectra, which then have to be processed post-acquisition to assign fragment ions to the appropriate precursor ions. A novel processing algorithm patented by Waters Corporation in 2005 has made this type of analysis possible.[2, 3]

A detailed progression of the processing algorithms can be found in a series of patents filed by Waters from 2005 to 2009.[4-8] Briefly, fragment ions are assigned to precursor ions based on both intensity and peak apex retention time. The apex of the chromatographic peak of the fragment ions must match that of the precursor ion within the time error associated with the duty cycle of the MS runs. For example, if the scan time of each MS or MS^E scan is 0.7 sec, with a 0.1 sec interscan delay, the duty cycle of the MS acquisition would be 1.6 sec, or 0.027 min, so the chromatographic apices must not differ by more than 0.027 min. With regard to ion assignment based on intensity, fragmentation of the most intense precursor ion should give some of the most intense fragment ions and low intensity precursor ion fragmentation should give fragment ions of lower intensity. Likewise, highly abundant fragment ions would not be assigned to precursor ions of low abundance and vice versa. This form of analysis allows for the simultaneous acquisition of both quantitative and qualitative data.

3.1.3 Label-free absolute quantitation of proteins based on MS intensity of peptides

Numerous studies have been performed for differential analyses employing an LC separation followed by MS detection. In these studies, relative quantitation is performed primarily through the use of stable-isotope labeling of one sample or the use of isotopically-labeled internal standards.[9-12] Frequently, one of two differential samples is treated with the isotope labeling reagents while the other is untreated. The two samples are mixed and relative quantitation is contingent upon the observation of mass shifts caused by the labeling. Drawbacks to this approach include the reliability of labeling to completion as well as the doubling of sample complexity when the two samples are added together.

In the past decade, there has been a growing trend towards the use of label-free quantitation strategies. For a semi-quantitative analysis, the spectral counting method is most commonly used. This method is based on a the use of a DDA in which 1-5 of the most abundant precursor ions are selected from the MS survey scan for subsequent fragmentation. There are no limits set to the number of times that a precursor ion may be selected. Therefore, the intensity of a peptide, represented by its precursor ion in the mixture, is directly related to the number of times that it was selected for MS/MS analysis. The spectral counts for all peptides of a given protein are then averaged to give a protein abundance index. A major limitation of this strategy is the lack of error rates in these index values. Many times, improperly identified peptides as well as peptides found in multiple proteins, such as in highly conserved regions, are given equal weight to the overall protein abundance. There are steps being taken to remedy this shortcoming, such as weighting the spectral counts of a peptide based on its identification score from the database search however it is still primarily used for semi-quantitative studies.[13, 14]

The second type of label-free quantitation is based on the measurements of changes in peptide ion intensity between samples. Work published by Silva and co-workers addressed the shortcomings historically associated with the use of MS signal intensity for quantitation including the non-linearity of signal response and the effects that ion suppression may have on the response.[2] Known concentrations of standard protein digest mixtures were spiked into a human serum digest and the intensity of multiple peptides were measured. A response curve for this data resulted in a straight line across the 2 orders of dynamic range investigated in the study with an R^2 value of 0.9995. Intensity values for all peptides used to identify a protein are summed to give a value of protein abundance. A standard of a tryptically digested protein is spiked in for use as an internal standard. Calculated protein intensities in each sample are normalized to the sum of all of the peptides used for identification of the standard protein. A necessity for this type of analysis is the presence of an ample number of peptide hits per protein and a comprehensive analysis of a peak as it elutes, usually 10 MS cycles per peak. This would not be possible with a DDA where cycle times are several seconds each, so only a DIA such as MS^E can be used. The work presented in this chapter focuses on this approach to quantitation utilizing a data independent acquisition.

3.2 Experimental

3.2.1 Outline for experimental method

The workflow for the experiment can be found in Figure 3-1. The soluble fractions of cell lysates from differential samples of *S. cerevisiae* were digested with trypsin without any further fractionation. Digested samples were analyzed in triplicate by capillary RPLC-MS^E. Raw data were processed using ProteinLynx Global Server 2.4, Release Change 7 (PLGS2.4) for identification and absolute quantitation. Proteins identified in more than one replicate

were subject to a quantitative analysis, whereas those that did not replicate could only be analyzed qualitatively.

3.2.2 Chemicals and *S. cerevisiae* sample preparation

Most of the reagents used in this experiment were identical to those used in the previous chapter. Optima grade water and acetonitrile and LC-MS grade formic acid and ammonium bicarbonate were purchased from Fisher Scientific (Fair Lawn, NJ). Sequencing grade modified trypsin by Promega (Sigma-Aldrich, St. Louis, MO) was used in place of the Pierce trypsin used previously. The MassPrep BSA digest used for absolute quantification and the acid-labile surfactant, RapiGest SF, were provided by Waters Corporation (Milford, MA). The *S. cerevisiae* samples used in this experiment were the same as used in the previous chapter. The complete preparation procedure can be found in section 2.2.3.

3.2.3 Trypsin digestion

All solutions were prepared using 50 mM ammonium bicarbonate. For each sample, an aliquot from the stock was diluted to 50 µL to give a final concentration of 1 µg/µL total protein. Reconstituted RapiGest SF was added to give a final surfactant concentration of 0.1%. After vortexing, solutions were incubated for 15 minutes at 80°C with periodic vortexing. Solutions were centrifuged at 10,000 x g for 4 minutes to return condensate to the bottom of the centrifuge tube. To reduce disulfide bonds, 100 mM dithiothreitol was added to give a final concentration of 10 mM. The reduction was allowed to take place at 60°C for 30 minutes. After centrifugation, 200 mM iodoacetamide was added to give a final concentration of 20 mM and prevent reformation of disulfide linkages. Incubation took place in the dark at room temperature for 30 minutes. Samples were then centrifuged again and trypsin was added at a 1:50 enzyme: protein ratio. Digestion occurred overnight at 37°C.

Quenching of the digestion and hydrolysis of the surfactant was performed by the addition of TFA to a final concentration of 1%. Samples were incubated at 60°C for 2 hours. After centrifugation at 14,000 x g for 20 minutes, the supernatant was transferred to sample vials for analysis. MassPrep BSA digest was added at a concentration of 25 fmol/µL thereby injecting 100 fmol on column for a 4µL injection.

3.2.4 Instrumentation and run conditions

Samples were analyzed in triplicate by capillary ultra-high performance LC-MS^E (UPLC-MS^E) using a nanoAcquity LC system and a QTOF Premier mass spectrometer, both from Waters (Milford, MA). Mobile phase A consisted of water with 0.1% formic acid and mobile phase B consisted of acetonitrile with 0.1% formic acid. The system was run in trapping mode, with a 20 mm x 180 µm ID trap column packed with 5 µm C_{18} Symmetry particles. Loading of the sample onto the trap column was performed by flowing 0.5% mobile phase B at 15 µl/min through the inject valve and onto the trap column with the outlet of the trap column diverted to waste. Once the sample was properly loaded, the flow was decreased to 300 nL/min and the mobile phase composition changed to 5% B. The trapping valve was switched at this point such that the outlet of the trap column was sent directly to the analytical column. Separations were performed on an analytical column with dimensions of 250 mm x 75 µm ID packed with 1.7 µm BEH-C_{18} particles. Connection with the mass spectrometer was achieved by connecting a 20 µm ID capillary from the outlet of the column directly to a nanoflow lockspray-ESI source fitted with a 20 µm ID tip pulled down to 10 µm at the orifice. The standard solution for the reference channel of the nano-lockspray was 200 fmol/µL glu-fibrinopeptide, which was delivered by an auxiliary pump on the nanoAcquity system at 0.5µL/min. The full gradient profile for the separation is specified in Table 3-1.

The QTOF Premier was operated in positive ESI mode with the lockspray enabled and the reflectron operated in V-mode. The capillary, sample cone, and extraction cone were held at +3000V, +35V, and +4V, respectively throughout the analysis. In order to prevent water cluster formation, the source temperature was held at 100°C. The scan time for the low energy, elevated energy, and reference scans was 0.7 seconds across the *m/z* range from 50 to 1990. The collision energy was held at 5V for the low energy scans, 6V for the reference scans, and ramped from 15V-40V during the elevated energy scan. The reference channel was sampled every 30 seconds to allow for dynamic calibration.

3.2.5 Protein identification and quantification using PLGS2.4

The raw data acquired on the Q-TOF Premier for each fraction was processed using the latest version of ProteinLynx Global Server, PLGS2.4 RC7 (Waters, Milford, MA). Processed mass spectra were searched against a SwissProt database composed of the non-redundant yeast proteins, porcine trypsin, bovine serum albumin, and five human keratin proteins that have been the most abundant contaminants in previous experiments. A 1x randomized database was generated within PLGS2.4 and added to the end of the yeast database. For quantitation purposes, the known concentration of the standard BSA digest was added to the workflow. This allowed for absolute quantification of yeast proteins based on peptide intensity. A complete listing of the values for processing the raw data and searching against the database can be found in Table 3-2.

3.3 Results

3.3.1 Protein identifications based on MS^E data

The raw data obtained from the LC-MS^E analysis of the differential yeast samples was processed directly by PLGS2.4. With the addition of the lockspray source a dynamic

mass calibration was performed to account for ambient temperature fluctuations. Data processing was performed on the calibrated mass spectra. Table 3-3 describes the protein identification statistics as output from PLGS2.4 for all three differential samples. In the glycerol sample, there were a total of 776 yeast protein identifications across all three replicate injections. In PLGS2.4, the results were automatically curated to have values of 2, 1, or 0, which correspond to identification probabilities of 95%, 50%, and 'not probable', respectively. For the 776 proteins identifications in the glycerol sample, 694 were identified with 95% probability, 65 were identified at 50% probability, and 17 were likely false identifications. After removal of redundant proteins across the replicates, there were a total of 365 unique proteins identified. In order for an identification to be considered 'real', it had to fulfill one of two requirements: it had to have been identified at 95% confidence in at least one sample, or, if it was only identified at 50% confidence, then it had to replicate in at least two of the three analyses. After removal of proteins that did not fulfill either of these criteria, there were 302 unique proteins identified, with 241 of them replicating. On average, proteins were identified with approximately 15 peptides. The median value of peptide hits was slightly lower at 10. The corresponding values for the dextrose-grown log phase sample and dextrose-grown stationary phase sample can also be found in Table 3-3.

The distribution of the intact molecular weights of the proteins identified in the bottom-up analysis across all 3 differential samples is shown in Figure 3-2. This distribution will be compared to the distribution obtained from the on-line top-down analysis of intact proteins in Figure 2-5 from Chapter 2 later in this chapter.

3.3.2 Repeatability based on replicate analysis

Each sample, while only digested once, was run in triplicate by LC-MS. The overlap of proteins identified in multiple runs can be seen in Figure 3-3. All of the non-replicating proteins not identified with a 95% confidence have been removed; only curated protein hits are present in this figure. On average, only 8% of all of proteins identified in each sample were identified in only one replicate with a confidence of 95%.

The Venn diagrams illustrate the high reproducibility of the bottom-up analysis with regard to protein identification, but do not lend any insight into the reproducibility of the absolute quantitation, a critical aspect of a differential analysis. In order to assess quantitative reproducibility, a log/log plot of protein intensities across replicates was constructed just as was done previously in section 2.3.1. The replicate log/log plots of each sample are shown in Figure 3-4. Only proteins identified in at least 2 replicates are included in the plots. Overall, the replicate points fall closely along the line y=x. There does appear to be a similar increase in scatter at the lower intensities, but due to the decrease in data points as compared to the replicate data taken in the previous chapter, it was not feasible to include error curves. Figure 3-5 shows all of the replicate data points from the differential samples in the comparison between the glycerol-grown sample and dextrose-grown sample in A, and the comparison of the log phase sample versus the stationary phase sample in B. In both plots, replicate data is plotted in gray and the differential comparison is in red. There is not an immediately apparent cut-off that could be used to determine which differences should be considered significant. However, with the addition of replicate analysis for each sample, comparisons can be made on a protein-by-protein basis. If the intensity of a single protein is

consistent within the replicate runs, then it should be easier to compare the intensity of the protein to that of the identical protein in another sample.

3.3.3 Determination of differential proteins and significance of the difference

Two different methods were investigated in order to determine proteins that were significantly different in intensity. The first was an extension of the use of a Log/Log intensity plot used in Chapter 2. From the replicate analysis, error bars were added in order to see if a difference in intensity was indeed significant. Error bars, normally plotted at +/- one standard deviation in the linear scale, are plotted as relative errors so as not to give a skewed view of the true error in the measurement. The derivation of this as an acceptable form of error representation is as follows. The intensity value plotted, z, is different than the measured quantity, y.

$$z = \log (y)$$

The error, dz, of z would then be

$$dz = d[\log (y)]$$

Based on the assumption that errors are relatively small, the differential analysis gives

$$dz = d[\log(y)] = \frac{1}{2.303}\frac{dy}{y} \approx 0.434\frac{dy}{y}$$

Therefore, dz can be expressed as a relative error in y:

$$dz \approx 0.434\frac{dy}{y}$$

This can be seen in Figure 3-6. From this plot, it is easy to pick out some proteins that are clearly differential because they do not fall near the line y=x and have small error bars. Some of these points have been highlighted in the figure. The protein identity of each of the data points is known, so correlating a point on the plot to a yeast protein is straightforward.

Difficulty arises, however, when trying to determine if differences are significant when the average intensities of a protein in each sample are close, or when proteins are identified in the replicate analysis in only one of the samples.

In order to facilitate the selection of differentially expressed proteins without imposing any bias that might arise from the use of logarithmic plots, the data were plotted in bar graph format with both the averages and standard deviations of the protein intensity in fmols plotted in the linear scale. Figure 3-7 and Figure 3-8 show the differences in intensity between the proteins identified in at least two out of three replicates that were identified in both the glycerol-grown and dextrose-grown samples. Proteins with non-overlapping error bars can be easily selected as having significant differences. A list of all of the differentially expressed proteins in the comparison of growth media can be found in the Appendix. A shortened list containing only those proteins that were found to be differential in this experiment as well as the combined online top-down/bottom-up analysis described in the previous chapter is included in Table 3-4. All headings are identical to those described in the previous chapter with the exception of the last 4 columns. The 'Bottom-up' columns correspond to the fold change and sample in which the protein was most intense as determined by the data presented in this experiment. The 'On-line Top Down' columns refer to the corresponding fold change and sample in which a protein was up-regulated as determined by the differences in intensity from the intact protein MS intensity in the previous chapter. Figure 3-9, Figure 3-10, and Table 3-5 show the equivalent data and comparisons between the yeast samples harvested during different stages of growth.

For proteins identified by replicate analysis in both samples used for comparison, the determination of whether or not the intensity difference is significant is made easy due to the

addition of error bars. However, when proteins either do not replicate, or are identified in only one of the samples, the distinction of significance is less clear. In the case where a protein is identified in multiple replicates of one sample, but either never identified or identified in only one replicate in the other sample, the data can be plotted in bar graph format, just as was done when a protein replicated in both samples. Figure 3-11 and Figure 3-12 show these proteins in the glycerol/dextrose and log/stationary phase comparisons, respectively. If the protein was identified at all in the other sample, the intensity is added to the plot. A protein is considered to be a potential differential protein if the error bars do not span below zero and, in the event that a protein was identified in just one replicate in the comparative sample, the intensity of that protein must not fall within the error of the intensity of the protein in the replicating sample. Through this analysis, 111 out of the 116 proteins that replicated only in the glycerol-grown sample are considered to be differentially expressed. Likewise, 58 of the 60 proteins that replicated only in the dextrose-grown sample are differential. In the log/stationary phase comparison, 63 out of 67 proteins that replicated only in the log phase sample are significant, and 24 out of 27 proteins that replicated only in the stationary phase sample are significant.

A final case of protein identification occurs when a protein is only identified in one replicate at a confidence of 95% in one or both samples used for comparison. Due to the fact that the protein was identified at that high of a confidence level, it is safe to conclude that the protein is indeed present. However, there is no error associated with the intensity; it was only used in a qualitative aspect. This is supported by the lower number of peptides used to identify non-replicating proteins when compared to that of proteins that did replicate (median

value of 6 peptides/protein as compared to 12 for the replicating proteins). Because of this, it is likely that there would be greater variation in the intensity values.

3.4 Discussion

3.4.1 Protein Identification

As stated previously, peptide tandem MS was performed using a data independent acquisition, MS^E, instead of a data dependent acquisition as was used in the previous chapter. Fragment ions are paired with precursor ions in the data processing step as opposed to during the acquisition by limiting the fragmentation to that of a single ion as is done in DDA. As a result, multiple precursor ions can be fragmented simultaneously, limiting the need for multiple runs to adequately sample the data. This is demonstrated in the number of peptides that are used to identify each protein. Looking just at the differential proteins identified in the on-line analysis in Chapter 2, the average number of peptides used to identify each protein across all three samples is just 5.8 compared to 16.4 peptide hits per protein in the bottom-up MS^E data. Furthermore, because the proteins were selected as differential from the top-down data, which is likely less sensitive than the bottom-up and so is representative of some of the more abundant proteins, a value of 5.8 actually overestimates the average peptide hits per protein when taking into account all protein hits. Looking back at all of the proteins identified in just the fractions containing potential differentially expressed proteins, on average only 3.2 peptides were used to identify all proteins. That is a drastic difference from the data presented in this chapter with greater than 15 peptide hits per protein. Presumably, if the fractions analyzed in the previous chapter had been run more than once, the data could have been merged into one file prior to data processing and database searching in order to improve the total number of precursor ions, or peptides, selected for

fragmentation. Replicate analysis by DDA MS would still likely fall short in terms of peptides/protein than a single MS^E analysis, due to the limitations addressed in the introduction.

Figure 3-2 shows the molecular weight distribution of the proteins identified in the bottom-up only experiment and Figure 2-5 shows the same distribution for the proteins detected in the on-line top-down analysis. Comparing these two histograms highlights one of the major limitations of intact protein electrospray-MS in that electrospray efficiency tends to decrease for proteins above 50 kDa. It is worth pointing out that the scaling of the x-axis is significantly different between the two plots. In the TD histogram, the maximum molecular weight is 80 kDa, whereas in the BU plot, the range extends all the way out to 230 kDa. This is due to the absence of reliably detecting intact proteins above 80 kDa in previous experiments in our lab, so the AutoME processing parameters were set to only deconvolute spectra up to this threshold. There is no such need for setting a threshold of this type with digested proteins, since the majority of proteins, regardless of size, give similar peptide molecular weight profiles and are not subject to a bias from electrospray ionization. To examine the skew of the data, the cumulative number of proteins in each bin is plotted on the axis to the right. If the mass range at which 50% of the proteins have been identified is compared between the two methods, the bottom-up data would have a slightly higher median molecular weight of 45 kDa than that of the top-down data with a median value of 35 kDa. Admittedly, this a modest difference compared to what might be expected, however it should also be noted that the intact protein deconvolution tends to have a higher background at higher molecular weights. The upward trend of protein counts above 60 kDa is likely due to increased noise at the higher molecular weight range instead of actual protein signal.

Therefore, it is likely that the median molecular weight range is less than the 35 kDa reported here.

3.4.2 Evaluation of reproducibility of bottom-up analysis

The reproducibility of protein identification is shown in Figure 3-3. Roughly 69% of all proteins across all three samples were identified in more than one replicate. This value includes those identified at all probability thresholds. If all non-replicating proteins with probabilities less than 95% are removed, this percentage increases to 82%. Removal of these proteins was deemed acceptable due the description of a '0' in PLGS as being 'not probable' in the scoring scheme.

There was also high reproducibility in the absolute quantitation data; average and median percent relative standard deviation (%RSD) were 21.5 and 12.5%, respectively, for replicating proteins across all three samples. This is along the same lines as the 15 and 20% that has been reported in this type of absolute quantitation based on MS^E analysis.[3, 15] The full distribution of %RSD across all three samples can be found in Figure 3-13. The distribution is heavily skewed towards low %RSD values, which is advantageous when trying to compare intensities between differential samples. Smaller %RSD values result in greater confidence of small changes in protein intensity between differential samples.

3.4.3 Identification of differentially expressed proteins

The identification of differentially expressed proteins was straightforward in the cases where a protein was identified in replicate analyses in both samples. If the difference in intensity was great enough to prevent overlap between the error bars, then it was considered significant. However, if a protein was only identified in the replicate analysis of one sample, and either not identified in the other, or only identified in one replicate, the determination of

it being a differential protein is not as certain. As the intensity decreases down to near zero, it becomes less clear whether the protein was present and just not detected in the other sample, or was indeed below the level of quantitation.

3.4.4 Comparison of protein intensities and correlation between those identified in Chapter 2

The overlap between the differentially expressed proteins found in the data presented in this chapter compared with the differential proteins from the previous chapter is shown in Table 3-4 and Table 3-5. Proteins that are not highlighted are identified as being up-regulated in the same sample between the glycerol-grown and dextrose-grown comparison, however were only detected in one of the samples in one or both experiments and so a comparison of the fold change could not be determined. Proteins highlighted in green are those with similar fold changes, differing by a factor less than 1.5, between the two methods and show a correlation between bottom-up based differences and top-down based changes in intensity.

The proteins highlighted in yellow and red are of the greatest interest when comparing between the two data sets. Yellow highlighting signifies proteins that are shown to be up-regulated in the same sample between the two methods, but by different amounts. Red indicates that the sample in which a protein was determined to be up-regulated was not consistent between the two analysis methods. For example, alcohol dehydrogenase, ADH1, was found to be up-regulated in the dextrose-grown sample in the bottom-up analysis, but was more abundant in the glycerol-grown samples based on the top-down data. One reason that the two methods may not give correlating fold changes is due to the post translational processing that occurs in the form of post-translational modifications (PTMs). For example,

in a bottom-up experiment of this type with no pre-fractionation, all forms of a single protein will be summed into one intensity value, thus negating any information that may be gained by looking at changes in the cellular processing of a protein between two conditions. When a separation is performed on the intact proteins prior to digestion, different forms a protein may be separated from one another if the post-translational modification is one that elicits a change in the chemical property probed by the separation mechanism. An example of this would be the phosphorylation of a protein changing its charge and therefore its retention time in an anion exchange separation.

3.5 Summary and Conclusions

The data presented in this chapter was obtained through a traditional bottom-up proteomic experiment. The soluble fractions of baker's yeast cell lysates were enzymatically digested prior to any further separation or fractionation. Protein digests were analyzed in triplicate by UHPLC-MS^E. The transition from a DDA acquisition to MS^E, a DIA, allowed for the parallel analysis of MS/MS peptide fragmentation as opposed to the serial analysis of a DDA acquisition. Thus, the average number of peptides used to identify each protein was almost three times greater in the MS^E analysis. This also facilitated the use of MS^E as a tool for absolute quantitation, which was critical for the differential analysis shown here.

Overall, as compared to the on-line 2D-LC data presented in the previous chapter, there was good alignment between the comparisons. Thirty-nine of 43 differentially expressed proteins identified in both methods are shown to be up-regulated in the same sample, with six of them having almost identical fold changes. The proteins with large differences in fold change and even opposing values are also interesting. One possibility for this difference is the grouping of all forms of a single protein into one intensity value in the BU data, while

they may be separated by LC prior to MS analysis in the TD-based differential analysis yielding two different intensities for the multiple species. Even if they are not separated by LC, analysis by MS gives an added dimension of separation by separating by the intact protein MW. This is not meant to demonstrate whether one method is better than the other, rather that complementary information can be gained from both techniques.

3.6 References

(1) Wehr, T. *LC-GC North America* **2006**, *24*, 1006-1008.

(2) Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M. V.; Kass, I. J.; Li, G.-Z.; McKenna, T.; Nold, M.; Richardson, K.; Young, P.; Geromanos, S. *Analytical Chemistry* **2005**, *77*, 2187-2200.

(3) Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M. V.; Li, G.-Z.; Richardson, K.; Wall, D.; Geromanos, S. J. *Molecular and Cellular Proteomics* **2006**, *5*, 589-607.

(4) Geromanos, S. J.; Silva, J. C.; Li, G.-Z.; Gorenstein, M. V.; (Waters Investment Limited, USA). WO, 2005, pp 49.

(5) Gorenstein, M. V.; Geromanos, S. J.; Silva, J. C.; Li, G.-Z.; (Waters Investments Limited, USA). WO, 2006, pp 94.

(6) Geromanos, S.; Silva, J. C.; Vissers, H.; Li, G.-Z.; (Waters Investments Limited, USA). WO, 2006, pp 32.

(7) Geromanos, S. J.; Li, G.-Z.; Silva, J. C.; Gorenstein, M. V.; Vissers, H.; (Waters Investments Limited, USA). WO, 2007, pp 35.

(8) Gorenstein, M. V.; Stapels, M. D.; Geromanos, S.; Golick, D.; Silva, J. C.; Li, G.-Z.; (Waters Technologies Corporation, USA). WO, 2009, pp 91.

(9) Goodchild, A.; Raftery, M.; Saunders, N. F. W.; Guilhaus, M.; Cavicchioli, R. *Journal of Proteome Research* **2005**, *4*, 473-480.

(10) Steen, H.; Pandey, A. *Trends in Biotechnology* **2002**, *20*, 361-364.

(11) Ong, S. E.; Blagoev, B.; Kractchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Molecular and Cellular Proteomics* **2002**, *1*, 376-386.

(12) Mann, M. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 952-958.

(13) Mueller, L. N.; Brusniak, M.-Y.; Mani, D. R.; Aebersold, R. *Journal of Proteome Research* **2008**, *7*, 51-61.

(14) Allet, N.; Barrillat, N.; Baussant, T.; Boiteau, C.; Botti, P.; Bougueleret, L.; Budin, N.; Canet, D.; Carraud, S.; Chiappe, D.; Christmann, N.; Colinge, J.; Cusin, I.; Dafflon, N.; Depresle, B.; Fasso, I.; Frauchiger, P.; Gaertner, H.; Gleizes, A.; Gonzalez-Couto, E.; Jeandenans, C.; Karmime, A.; Kowall, T.; Lagache, S.; Mahe, E.; Masselot, A.; Mattou, H.; Moniatte, M.; Niknejad, A.; Paolini, M.; Perret, F.; Pinaud, N.; Ranno, F.; Raimondi, S.; Reffas, S.; Regamey, P.-O.; Rey, P.-A.; Rodriguez-Tome, P.; Rose, K.; Rossellat, G.; Saudrais, C.; Schmidt, C.; Villain, M.; Zwahlen, C. *Proteomics* **2004**, *4*, 2333-2351.

(15) Cheng, F.-y.; Blackburn, K.; Lin, Y.-m.; Goshe, M. B.; Williamson, J. D. *Journal of Proteome Research* **2009**, *8*, 82-93.

3.7 Tables

Time (min)	% Mobile phase B
0	5
60	40
65	85
70	85
73	5

Table 3-1: RPLC gradient condition for the analysis of digested fraction from the anion exchange column. Mobile phase A was water with 0.1% formic acid and mobile phase B was acetonitrile with 0.1% formic acid. The flow rate was 300 nL/min.

Processing Parameters	**Value**
Chromatographic peak width	Automatic
MS TOF resolution	Automatic
Lock Mass for charge 1	-
Lock Mass for charge 2	785.8426 Da/e
Lock mass window	0.25 Da
Low energy threshold	200 counts
Elevated energy threshold	75 counts
Retention time window	Automatic
Elution start time	-
Elution stop time	-
Intensity threshold	1500 counts
Workflow Template	**Value**
Search engine type	PLGS
Databank	Yeast proteins with trypsin, BSA, and 5 human keratin proteins with a 1x randomization
Peptide tolerance	Automatic
Fragment tolerance	Automatic
Minimum ions per peptide	3
Minimum ions per protein	7
Minimum peptides per protein	1
Maximum protein mass	250,000 Da
Primary digest reagent	Trypsin
Secondary digest reagent	None
Missed cleavages	1
Fixed modifications	None
Variable modifications	Acetyl N-term, Carbamidomethyl C, Deamidation N, Oxidation M
False positive rate	4
Calibration protein	P02769 (BSA)
Protein concentration on column	100 fmol

Table 3-2: PLGS 2.4 RC7 processing parameters used for raw data processing and database searching.

	Glycerol/Log	Dextrose/Log	Dextrose/Stat
Total proteins ID's	776	560	449
ID confidence (2/1/0)	694/65/17	516/41/3	402/43/4
Unique proteins	365	248	197
Non-replicating 1's and 0's removed	302	221	166
Avg. peptides/protein	14.4	18.6	17.0
Median peptides/protein	10	12	10
Replicating proteins	241	181	140

Table 3-3: Protein identification statistics of PLGS2.4 processing of traditional bottom-up analysis. **Total protein ID's**: the total number of proteins identified across all three replicates. **ID confidence**: the confidence level output by PGLS2.4 based on the individual peptide sequencing assignments. **Unique protein**: the number of protein hits with the redundant protein identifications across replicates removed. **Non-replicating 1's and 0's removed**: number of proteins identified not including proteins with confidence values of 0 and 1 that did not replicate removed. **Avg. peptides/protein**: the average number of peptide hits used to identify each protein in a sample. **Median peptides/protein**: the median number of peptides used for each protein identification. **Replicating proteins**: the number of proteins that were identified in at least two of the three replicates.

Swiss Prot Name	Description	Intact Mass	Bottom up		On-line, Top-Down	
			Up-Reg. In	Fold Change	Up-Reg. In	Fold Change
AGX1	Alanine glyoxylate aminotransferase	41778	GLY	N/A	GLY	N/A
ALDH4	Potassium activated aldehyde dehydrogenase	56688	GLY	28	GLY	N/A
CISY1	Citrate synthase, mit.	53327	GLY	4.8	GLY	36
COFI	Cofilin	15890	GLY	1.3	GLY	1.3
EF1A	Elongation factor 1 alpha	50001	GLY	2.6	GLY	N/A
FKBP	FK506 binding protein	12150	GLY	1.4	GLY	2.9
HMF1	Protein HMF1	13776	GLY	N/A	GLY	4.7
IDH2	Isocitrate dehydrogenase 2	37802	GLY	N/A	GLY	N/A
IPB2	Protease B inhibitors 2 and 1	8460	GLY	N/A	GLY	8.6
MDHC	Malate dehydrogenase cyt.	40604	GLY	N/A	GLY	N/A
MDHM	Malate dehydrogenase mit.	35627	GLY	6.5	GLY	6.3
RIR4	Ribonucleoside diphosphate reductase	40028	GLY	2.5	GLY	5.1
SUCA	Succinyl CoA ligase	35010	GLY	7.5	GLY	15.4
TAL1	Transaldolase	37013	GLY	1.7	GLY	11
TKT1	Transketolase	73759	GLY	1.4	GLY	N/A
ADH1	Alcohol dehydrogenase	36799	DEX	2.8	GLY	N/A
CYS3	Cystathionine gamma lyase	42516	DEX	2.1	DEX	N/A
ENO1	Enolase 1	46773	DEX	3.4	GLY	N/A
GLRX1	Glutaredoxin 1	12246	DEX	N/A	DEX	22
OYE2	NADPH dehydrogenase 2	44884	DEX	N/A	DEX	9.6
PGK	Phosphoglycerate kinase	44710	DEX	1.7	DEX	2.0
SODC	Superoxide dismutase Cu Zn	15844	DEX	2.0	DEX	1.9
TPIS	Triosephosphate isomerase	26778	DEX	1.8	DEX	1.7
TRX2	Thioredoxin 2	1196	DEX	3.8	DEX	N/A
TRXB1	Thioredoxin reductase	34216	DEX	6.9	DEX	3.2

Table 3-4: Proteins determined to be differentially expressed in both the traditional bottom-up and combined top-down/bottom-up online analyses of the comparison between growth media. Proteins in green were consistently up-regulated in the same yeast sample and at roughly the same fold-change. Protein in yellow were up-regulated in the same sample, but at differing extents. Proteins in red were found to be up-regulated in opposite yeast sample. Un-highlighted proteins were only identified in one of the samples used for comparison and therefore, a fold change could not be computed nor compared.

Swiss Prot Name	Description	Intact Mass	Bottom-up		Online, Top Down	
			Up. Reg In	Fold Change	Up. Reg In	Fold Change
ACBP	Acyl-CoA-binding protein	10055	LOG	1.4	LOG	2.5
BMH1	Protein BMH1	30072	LOG	1.9	LOG	N/A
CH10	10kDa heat shock protein	11284	LOG	N/A	LOG	30
CYS3	Cystathionine gamma lyase	42414	LOG	N/A	LOG	19
G3P3	Glyceraldehyde 3 phosphate dehydrogenase 3	35742	LOG	2.2	LOG	123
HSP12	12kDa heat shock protein	11605	LOG	N/A	LOG	373
KGUA	Guanylate kinase	20906	LOG	N/A	LOG	7.1
MRP8	Uncharacterized protein MRP8	25005	LOG	N/A	LOG	N/A
PGK	Phosphoglycerate kinase	44710	LOG	3.3	LOG	N/A
PNC1	Nicotinamidase	24977	LOG	1.9	LOG	N/A
THRC	Threonine synthase	54738	LOG	1.9	LOG	N/A
G3P1	Glyceraldehyde 3 phosphate dehydrogenase 1	35727	STAT	3.0	STAT	N/A
G6PI	Glucose-6-phosphate isomerase	61261	STAT	1.4	STAT	1.7
GRE2	NADPH-dependent methylglyoxal reductase	38145	STAT	1.9	STAT	N/A
HBN1	Putative nitroreductase	20980	STAT	5.4	STAT	7.1
HSP31	Probable chaperone protein	25654	STAT	1.4	STAT	N/A
PMG1	Phosphoglycerate mutase 1	27591	STAT	1.4	LOG	2.0
SODC	Superoxide dismutase Cu Zn	15844	STAT	1.1	LOG	1.8

Table 3-5: List of differential proteins that were identified in both the traditional bottom-up and combined top-down/bottom-up online analyses for the comparison between phases of growth. Proteins in green were consistently up-regulated in the same yeast sample and at roughly the same fold-change. Protein in yellow were up-regulated in the same sample, but at differing extents. Proteins in red were found to be up-regulated in opposite yeast sample. Un-highlighted proteins were only identified in one of the samples used for comparison and therefore, a fold change could not be computed nor compared.

3.8 Figures

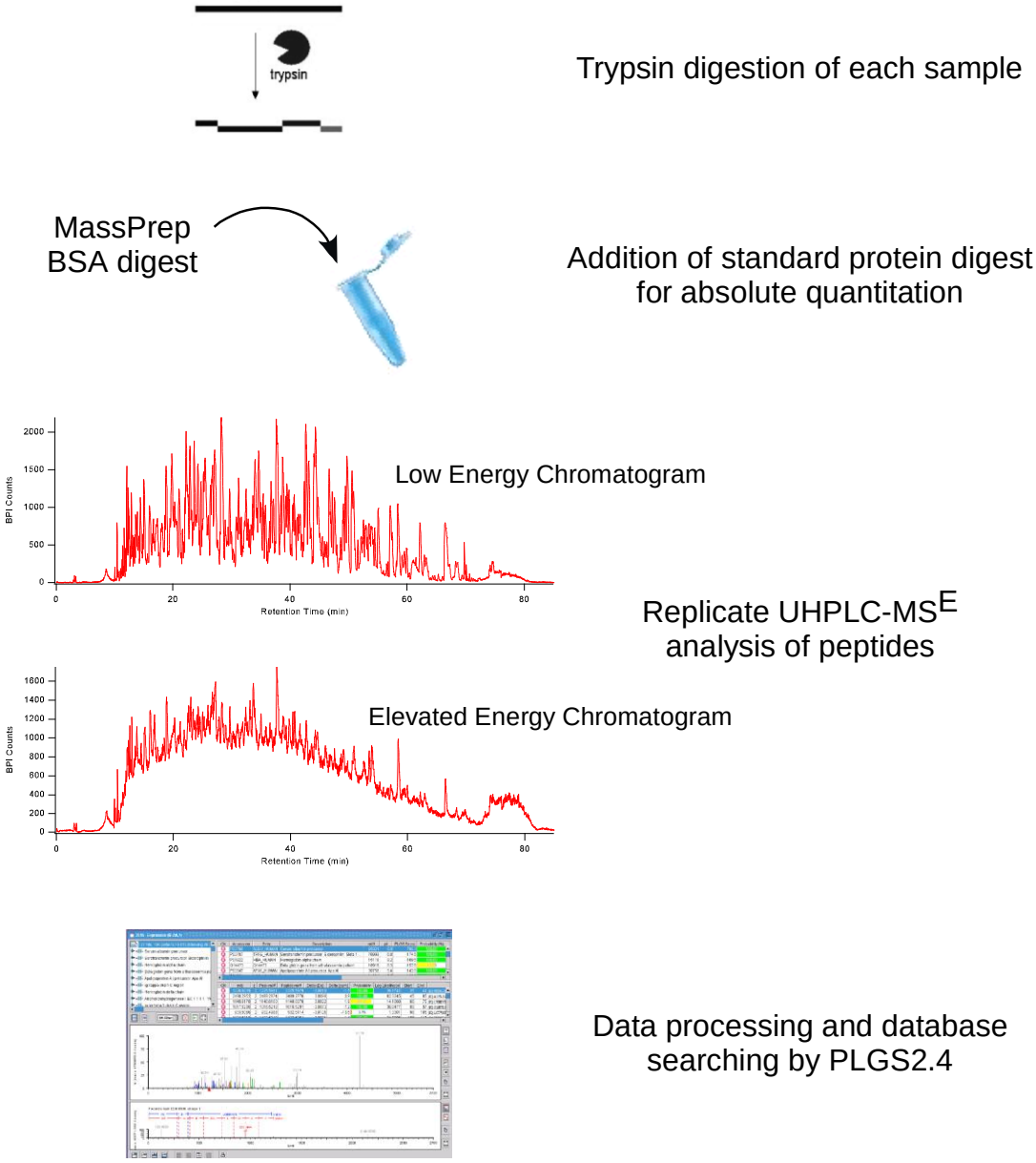


Figure 3-1: Workflow of traditional bottom-up proteomic analysis.

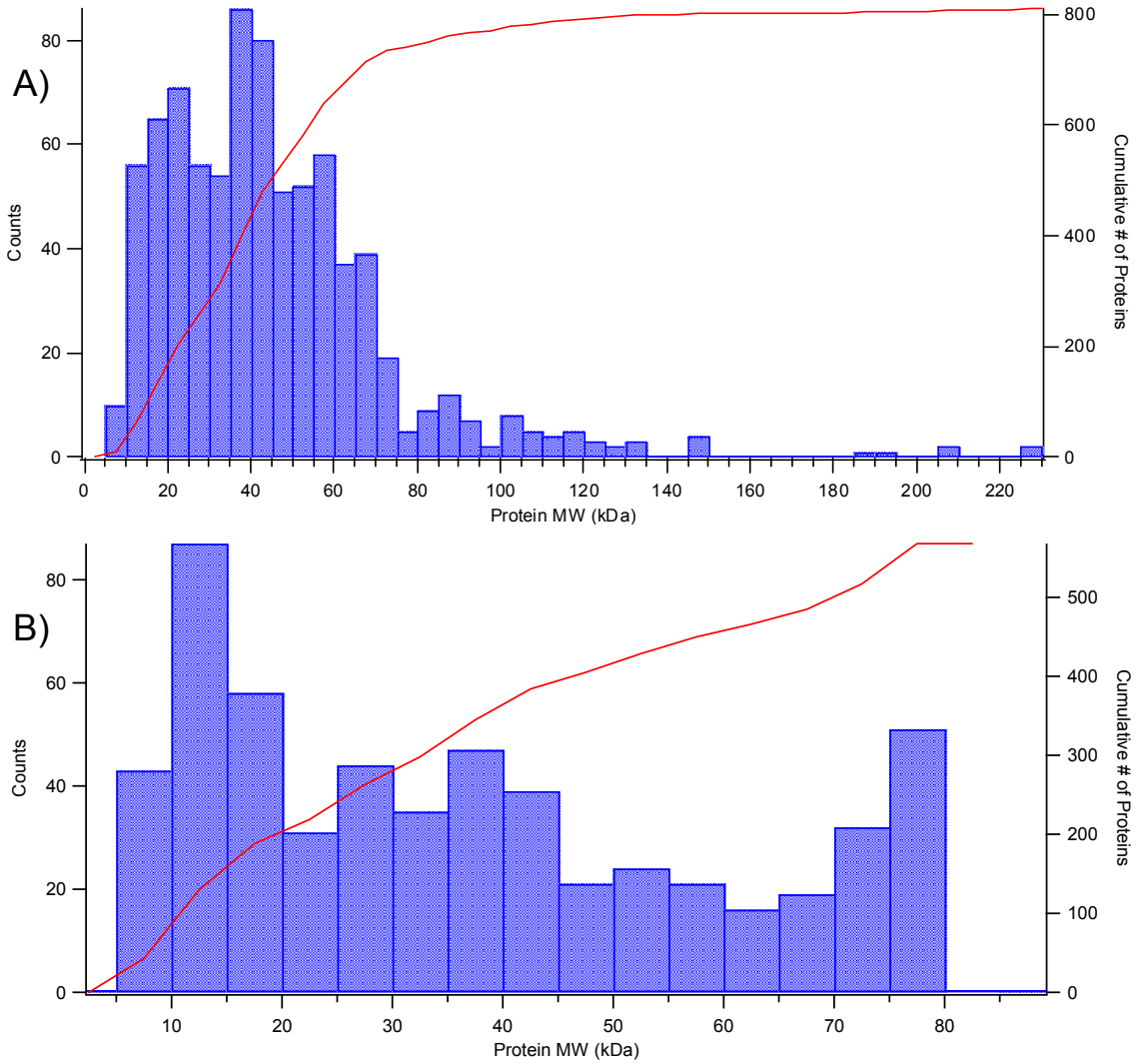


Figure 3-2: Distribution of the molecular weight of proteins detected. A) Proteins identified in the experiment presented in this chapter through a completely BU analysis of differential yeast samples. B) Distribution of protein masses detected from the online LC-LC-MS analysis of intact proteins performed in the previous chapter. Figure 2–7B is reprinted here for comparison

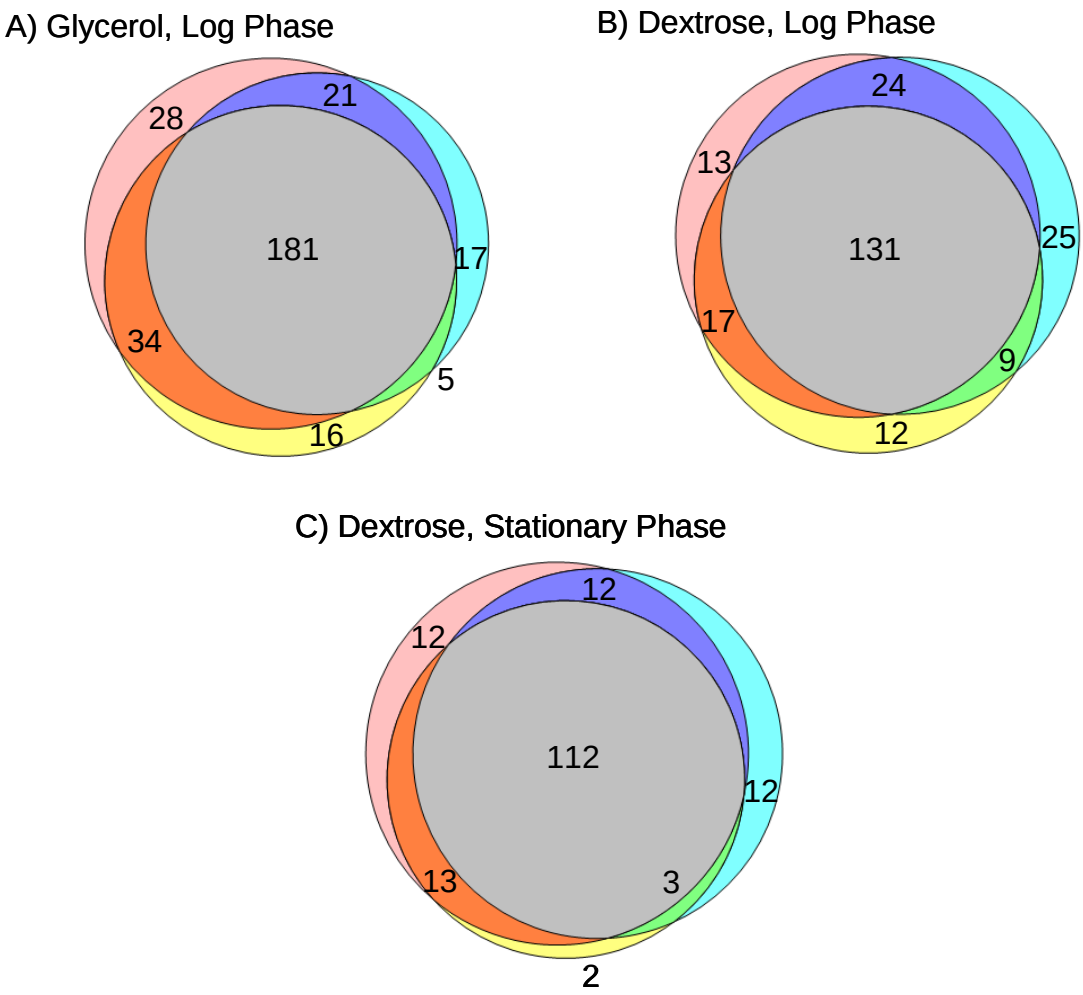


Figure 3-3: Replication of proteins identified by PLGS2.4 for the triplicate analysis of each differential sample. All replicating proteins and only non-replicating proteins identified at a 95% confidence are included.

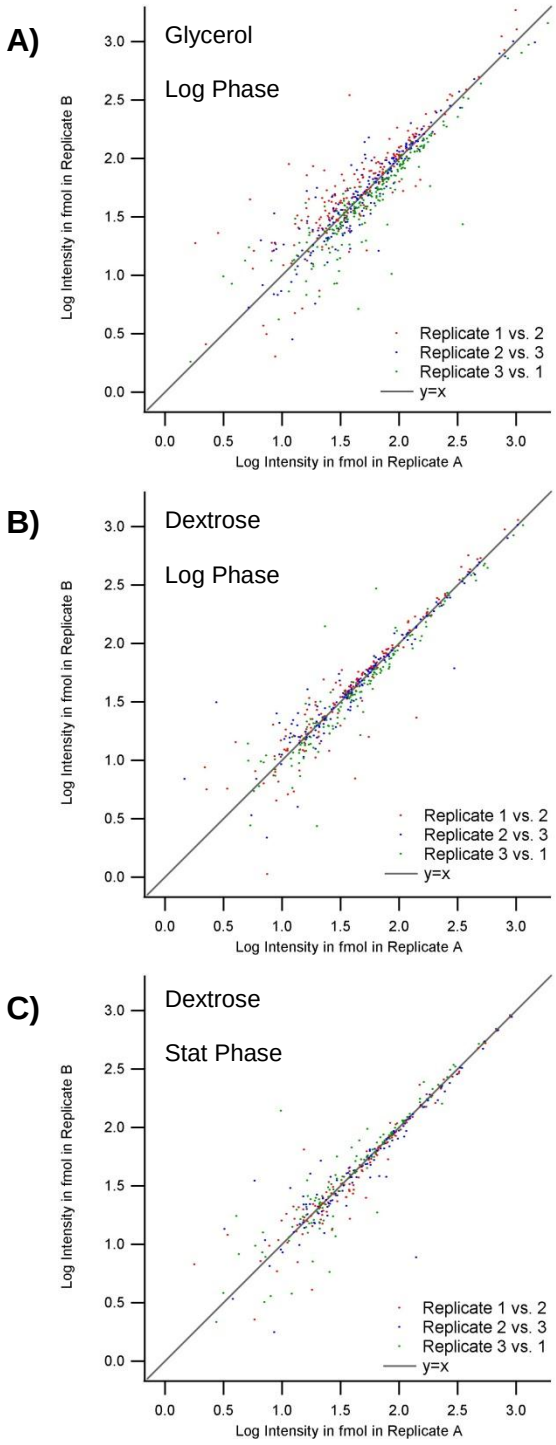


Figure 3-4: Log/log plots of replicate analysis of each sample. A) Bottom-up replicate analysis of glycerol/log phase sample. B) Replicate analysis of dextrose/log phase sample. C) Replicate analysis of dextrose/stationary phase sample.

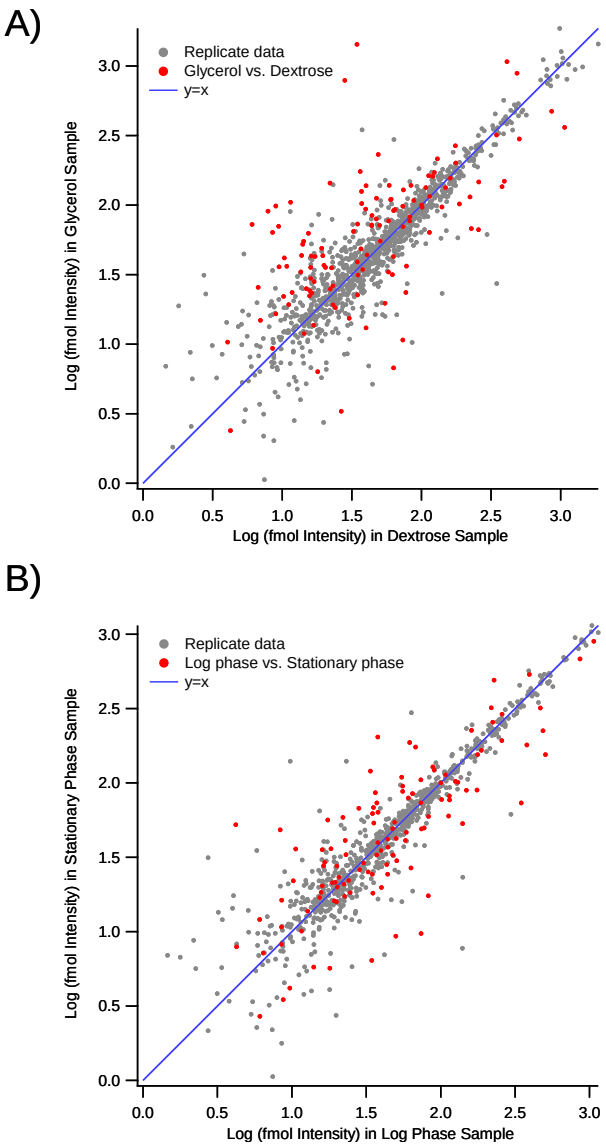


Figure 3-5: Log/Log plots of replicate data with differential comparisons appended. Only proteins identified in both differential samples are included. Replicate data and comparative data are given on a single plot for the comparison between the growth nutrient samples in A, and growth phase at the time of harvest in B.

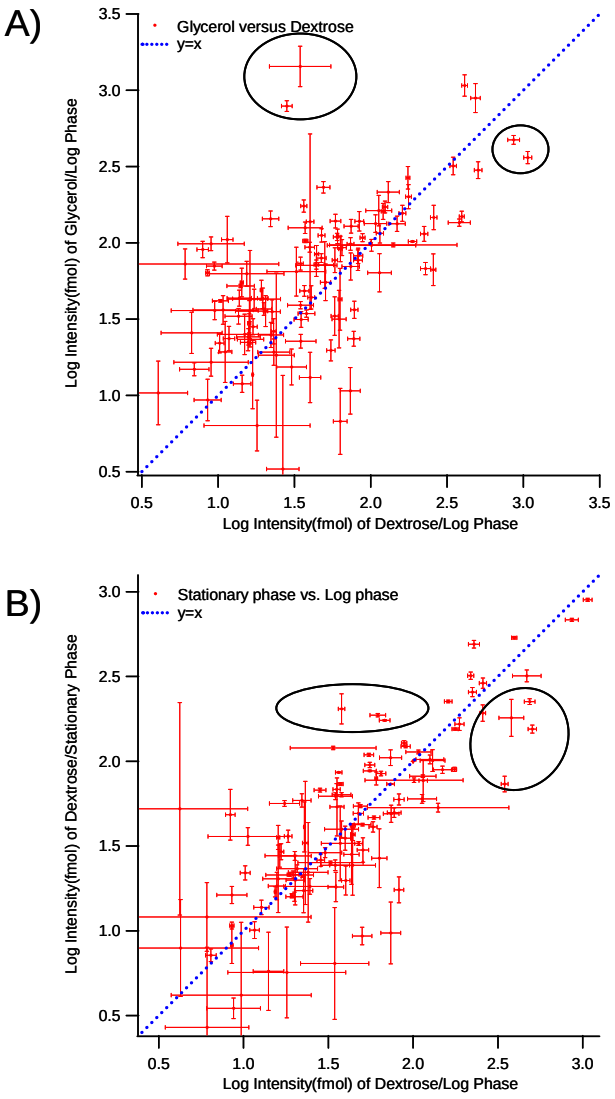


Figure 3-6: Log/Log intensity plot of differential sample comparison with error bars. Highlighted data points are representative points that are straightforward to designate as having significant differences.

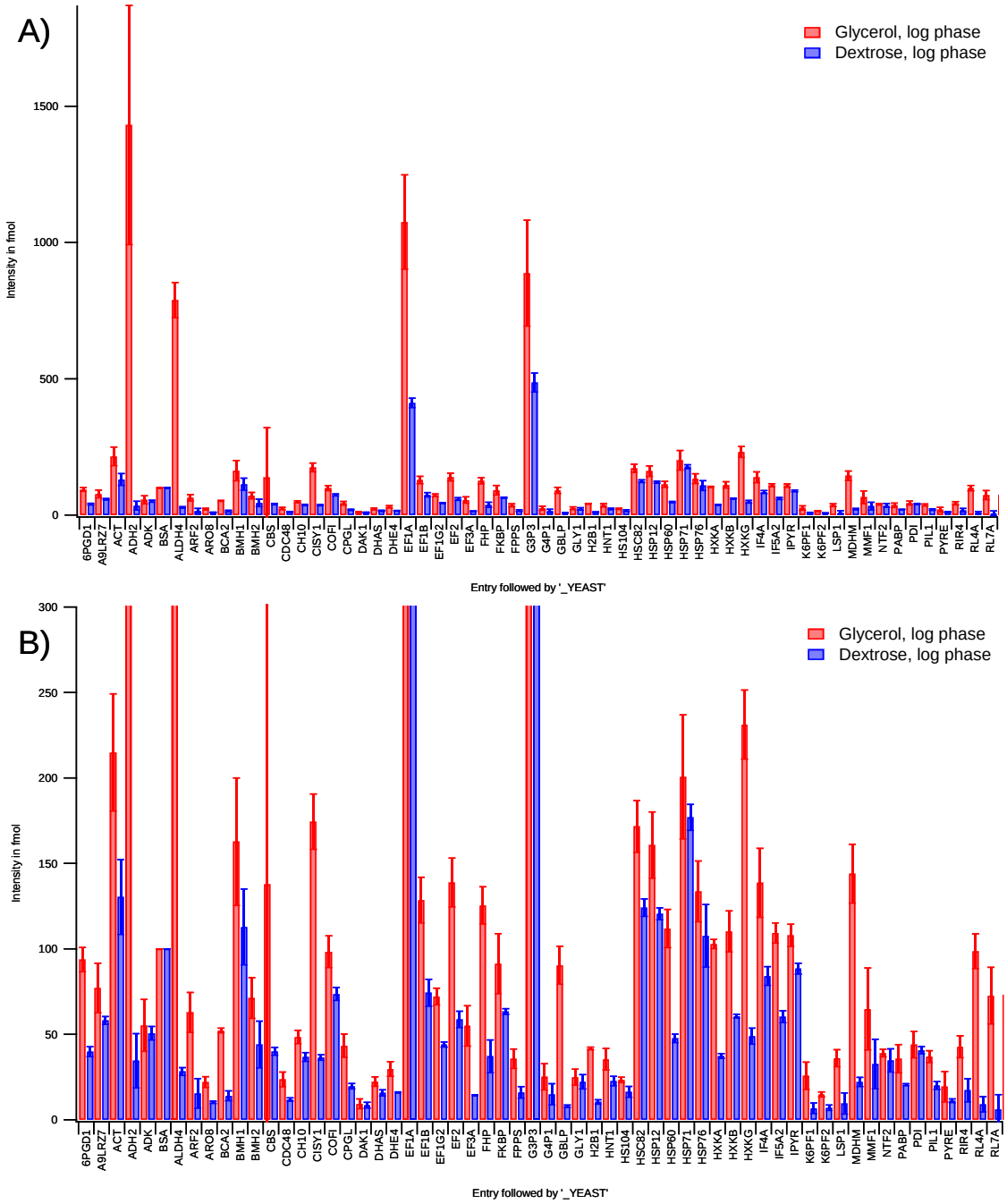


Figure 3-7: Bar graph comparison of replicating proteins identified in both the glycerol, log phase sample and the dextrose, log phase sample. A) Protein intensity in fmol full scale. B) Zoomed in view of protein intensity to better illustrate differences of proteins of lower intensities.

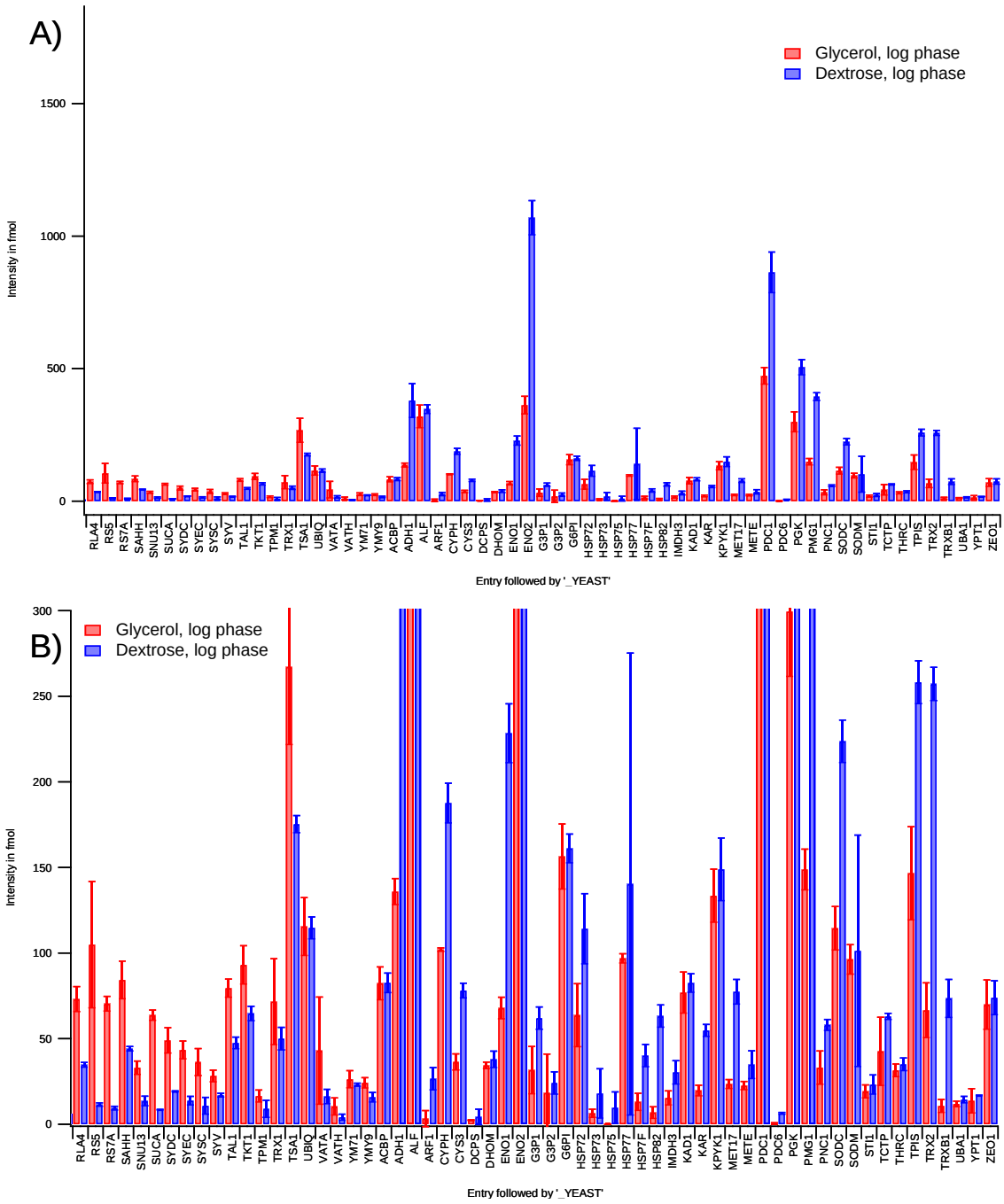


Figure 3-8: Continuation of replicating proteins identified in both the glycerol, log phase sample and the dextrose, log phase sample. The y-axis scaling is the same as in the previous figure. A) Protein intensity in fmol full scale. B) Zoomed in view of protein intensity to better illustrate differences of proteins of lower intensities.

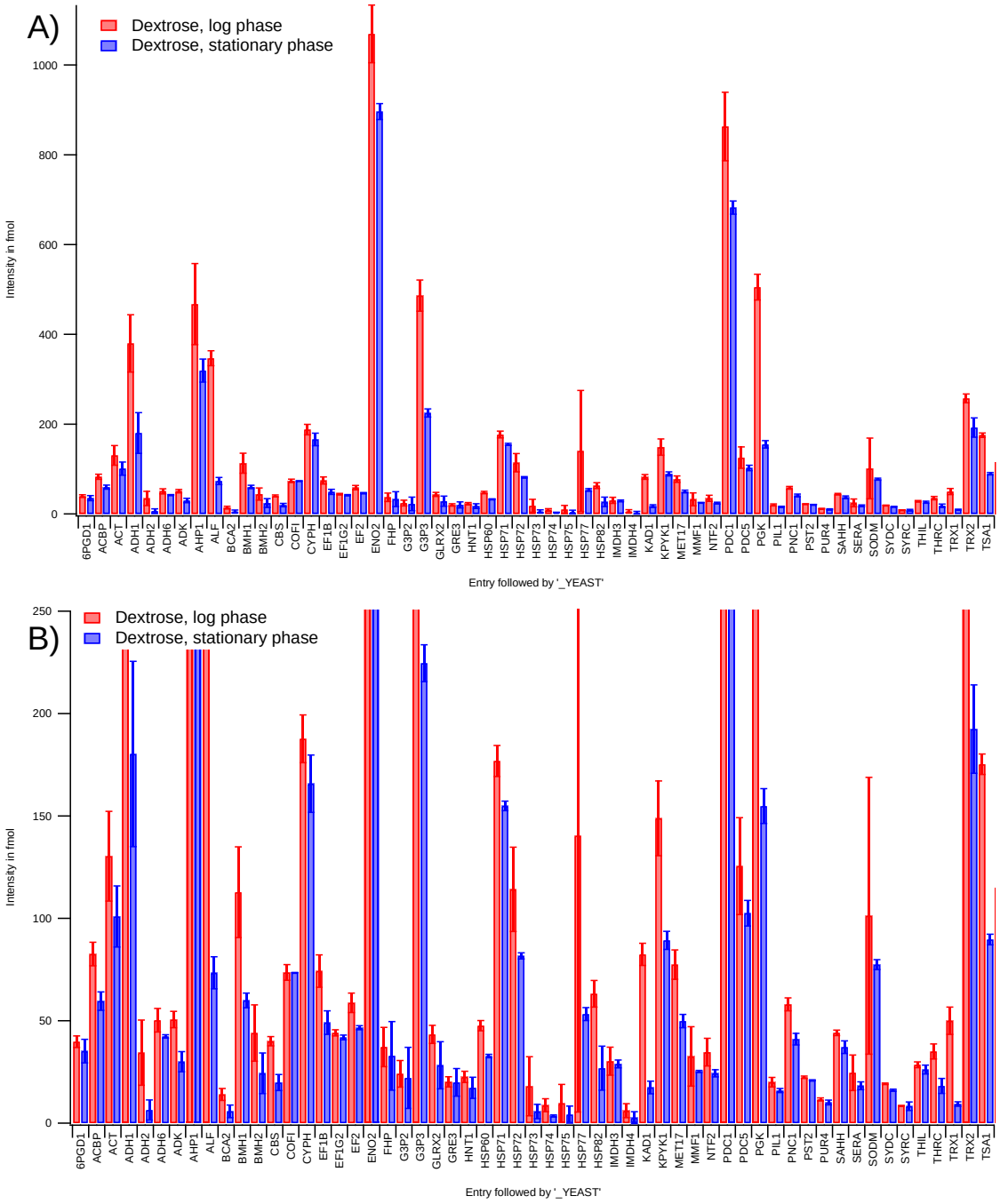


Figure 3-9: Bar graphs of replicating proteins identified in both the dextrose, log phase and dextrose, stationary phase samples. A) Protein intensity in fmol full scale. B) Zoomed in view of protein intensity to better illustrate differences of proteins of lower intensities.

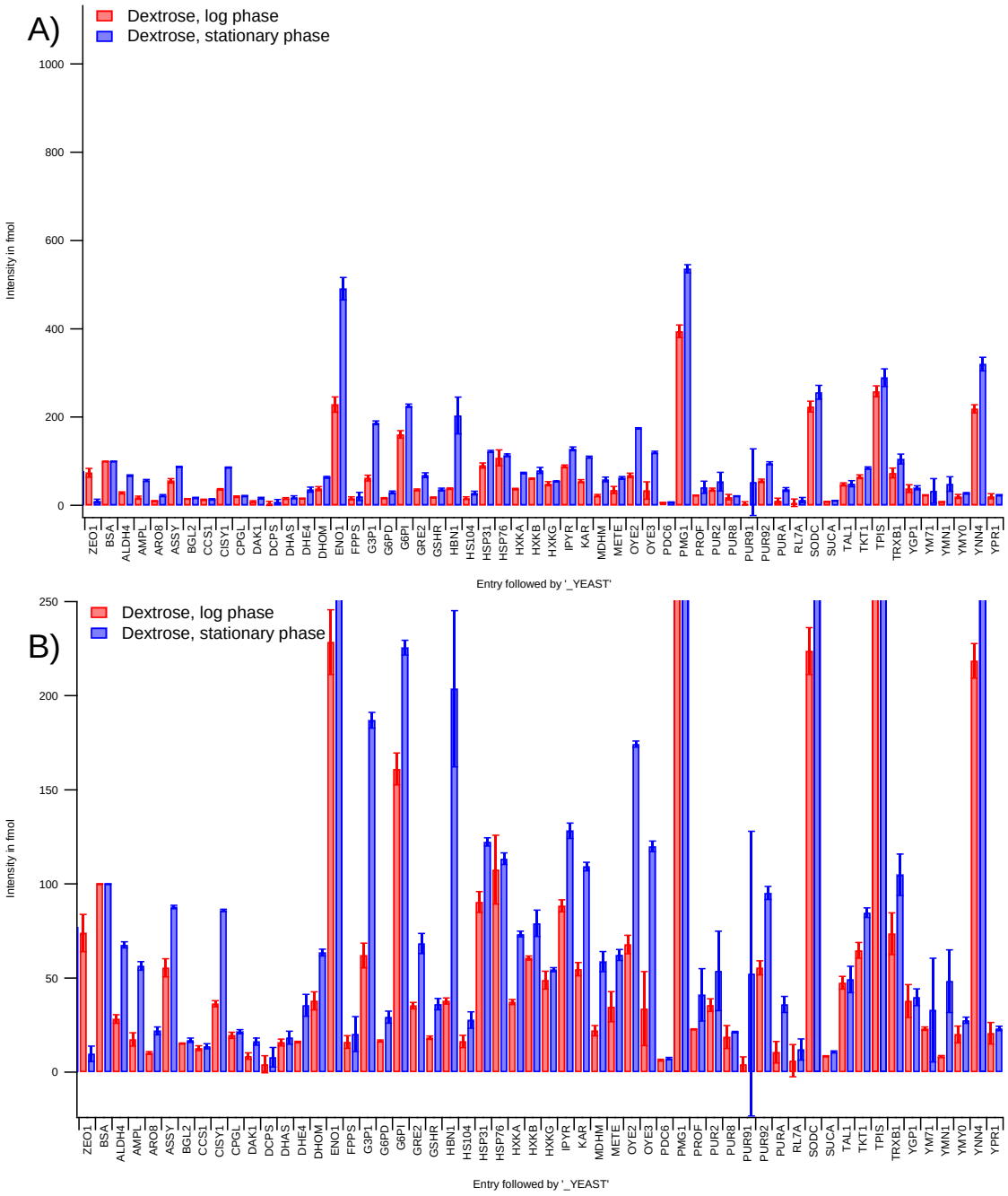


Figure 3-10: Continuation of replicating proteins identified in both dextrose, log phase and dextrose, stationary phase samples. A) Protein intensity in fmol full scale. B) Zoomed in view of protein intensity to better illustrate differences of proteins of lower intensities.

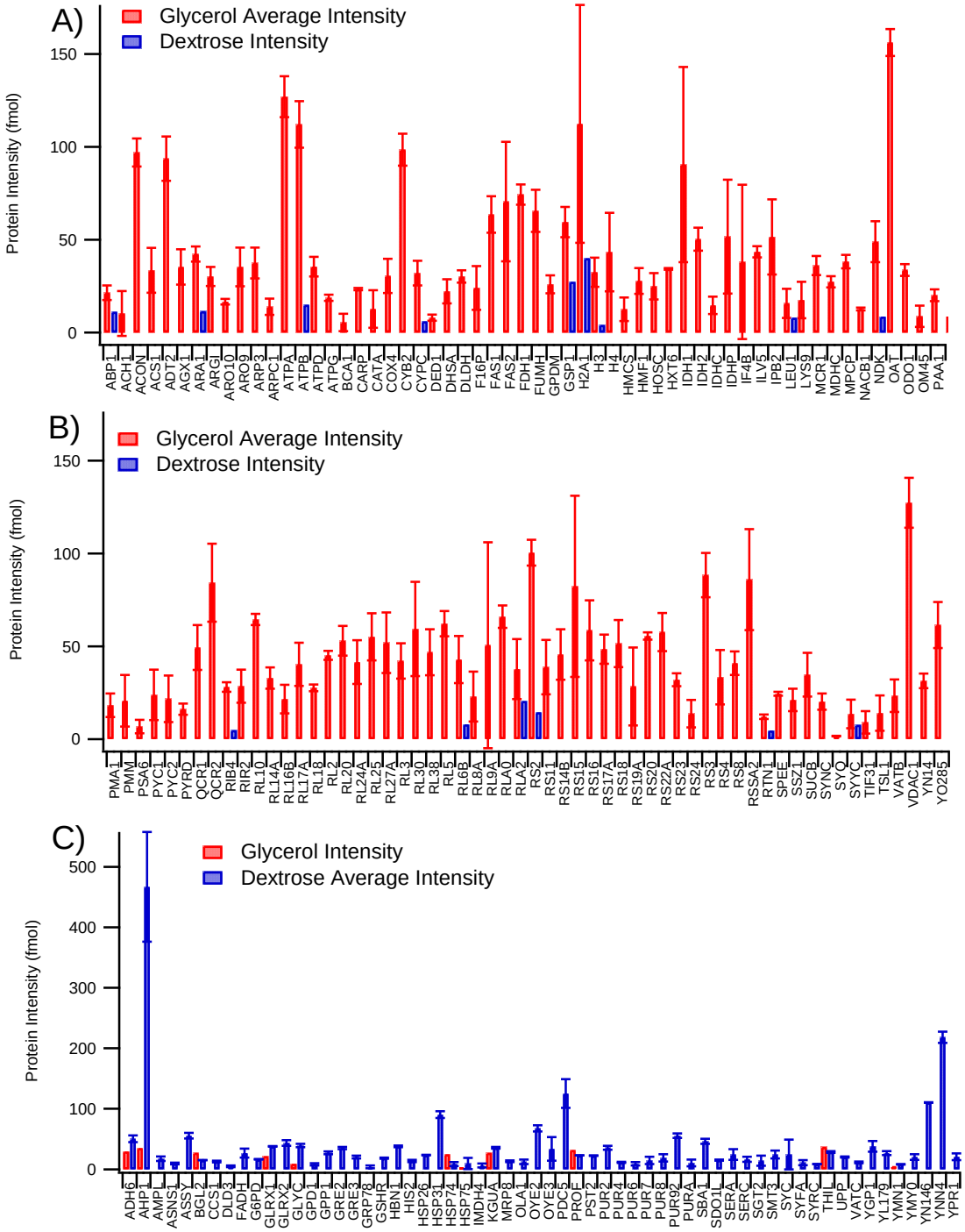


Figure 3-11: Replicating proteins identified in either the glycerol, log phase sample or the dextrose, log phase sample, but not both. A and B include the proteins which replicated in the glycerol, log phase sample only. C includes the proteins which replicated in only the dextrose, log phase sample.

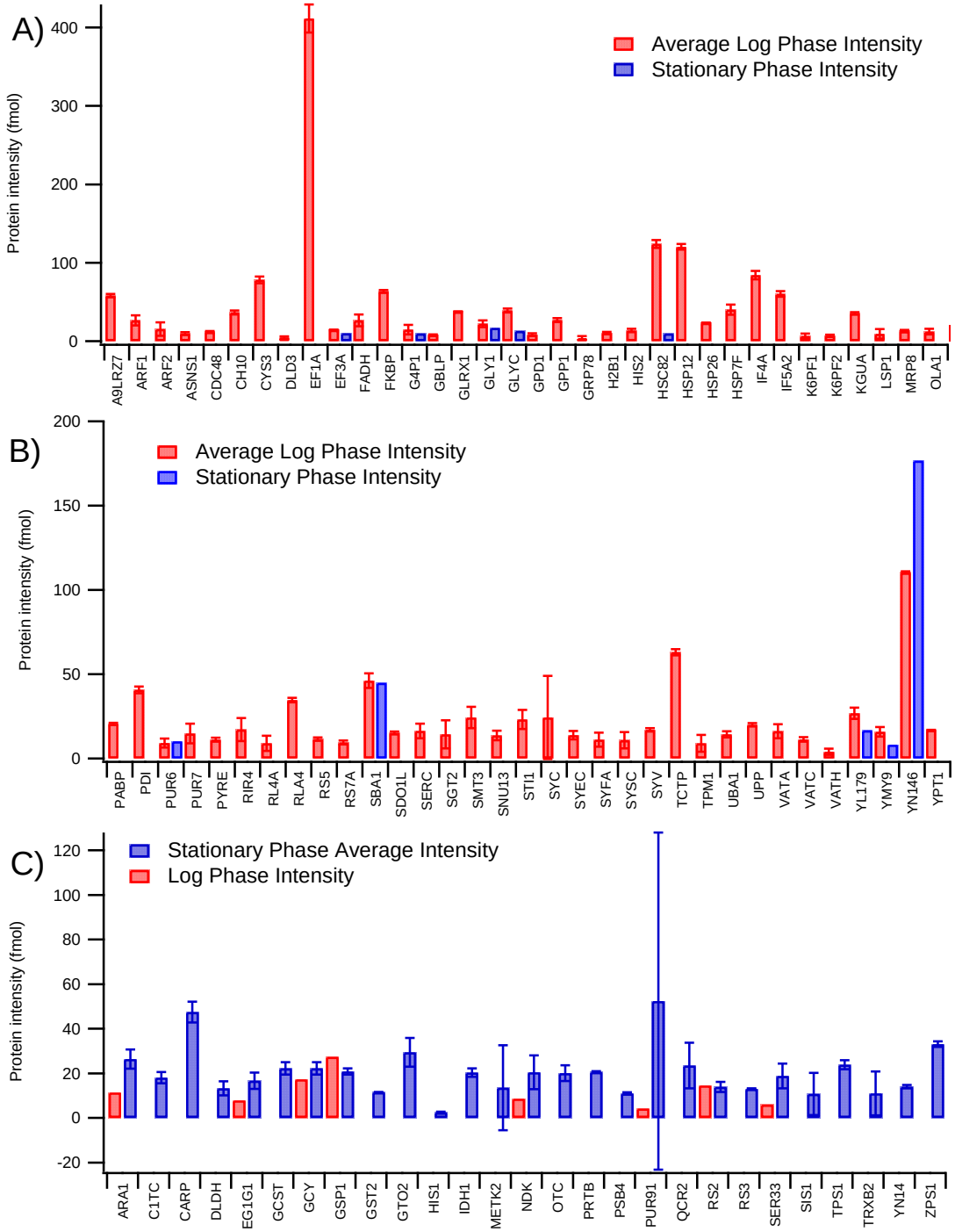


Figure 3-12: Replicating protein identified in either the dextrose sample harvested at the log phase or the dextrose sample harvested at the stationary phase. A and B contain proteins which replicated only in the log phase. C contains proteins which replicated only in the stationary phase.

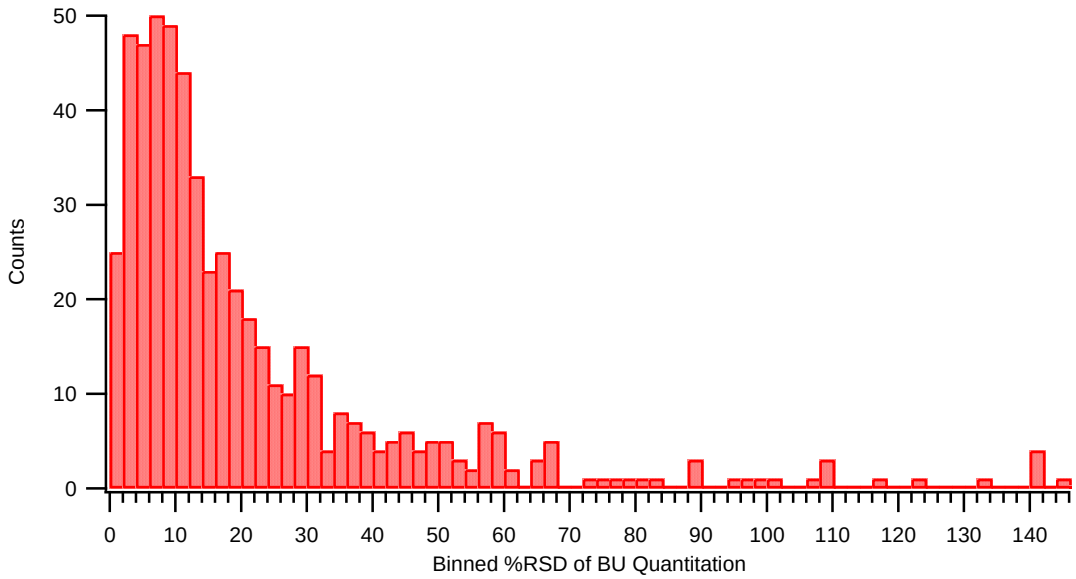


Figure 3-13: Distributions of %RSD of protein intensity by a bottom-up UPLC-MS^{E} analysis. The plot includes errors from all three differential yeast samples analyzed.

CHAPTER 4: Off-line multidimensional analysis of intact proteins from S. cerevisiae cell lysates using deconvoluted intact protein MS intensity for differential comparison

4.1 Introduction

4.1.1 Background and previous work comparing top-down and bottom-up proteomic methodologies

As addressed in Chapter 1, there are two main analytical strategies in mass spectrometry (MS) based proteomics, top-down (TD) and bottom-up (BU). While TD proteomics is increasing in popularity due to advances in ionization techniques and data processing, by far the majority of proteomic analysis is performed in a BU manner. Several studies have emerged that are similar in nature to the combined approach described in Chapter 2 whereby the mass of a protein is determined after some form of separation at the intact protein level followed by digestion and analysis of the resulting peptides by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Protein identification is performed by searching the BU data against a protein database and masses from the database are compared to those determined from the intact protein MS experiment. The first studies employing this strategy used it strictly as a protein mapping approach in order to identify as many proteins as possible in a single sample.[1-5] The overlap between protein masses from the intact MS data to the masses from the proteins identified from the database searching was generally low, around 30%, even when additional post-translational processing was factored in. While these studies were integral for the determination of some post-translational

modifications (PTMs), later work applied experiments utilizing a hybrid approach to differential proteomics. Differential expression of proteins in two or more samples was determined from a TD analysis based on protein MS intensity and compared to a list of protein identifications generated from digestion of a fraction containing the differential protein.[6, 7] These studies, however, do not address how the differential expression might change if analysis was performed as a completely BU experiment. Essentially, the question that needs to be addressed is how well does a fold change based on intact protein MS signal intensity compare to a fold change based on peptide MS intensity following digestion.

With the increase in TD proteomics, namely in the realm of differential proteomics, it is essential to examine if the differential expression determined by this type of proteomics gives similar or contradicting results to those obtained from a BU study of the same samples. The experiments presented in this chapter are aimed at addressing this issue.

4.1.2 Separations of intact proteins at ultra-high pressures

Historically, intact proteins have behaved poorly in reversed-phase liquid chromatography (RPLC) separations. There have been numerous reports of issues such as limited resolution, poor peak shape, and protein carryover or 'ghosting'. Previous work in the Jorgenson Lab has shown an improvement in the behavior of intact proteins when separated at pressures greater than 20,000 psi.[8] When comparing capillary columns operated at conventional pressures, 2,200 psi, and ultra-high pressures (UHP), 23,000 psi, protein recovery at UHP was 60% greater than that observed at conventional pressures. Furthermore, while the main goal of the study was to investigate protein carryover and ghosting, an improvement in peak shape and resolution of adjacent protein peaks was also observed.

The work presented in this chapter is aimed at investigating the similarities between a differential expression study based on both TD and BU proteomic analyses. Experiments performed in a TD nature utilized the UHPLC separation of intact proteins in order to improve the peak capacity of the RP separation. Fewer co-eluting peaks would then allow for a reduction in overlap of protein charge envelopes in the mass spectrum and result in improved mass spectral deconvolution.

4.2 Experimental

4.2.1 Outline for experimental method

The experimental method for the analysis presented in this chapter included both a top-down analysis and a bottom-up analysis as illustrated in Figure 4-1. The difference between the methodology presented here and that presented in Chapter 2 is that here, the analysis was performed in an off-line fashion. Also, instead of performing two dimensions of separation at the intact protein level prior to digestion, the intact proteins were only separated by anion exchange prior to digestion. After vacuum centrifugation and reconstitution, each anion exchange fraction was split in half; one half was analyzed in a top-down manner and the other was analyzed in a bottom-up manner. For the bottom-up analysis, after digestion with trypsin, samples were analyzed by LC-MS^E and the data processing with ProteinLynx Global Server. For the top-down analysis, proteins are analyzed by UHPLC-MS using a custom LC set-up capable of delivering flow at pressures up to 40,000 psi. The abundance comparison between differential samples was made with both sets of data, the TD and the BU. Once all analysis of the individual halves was complete, the expression changes that were found in each half were compared.

4.2.2 Reagent and sample preparation

Optima grade water and acetonitrile and LC-MS grade formic acid, ammonium bicarbonate, and ammonium acetate were purchased from Fisher Scientific (Fair Lawn, NJ). Sequencing grade modified trypsin by Promega, iodoacetamide, and dithiothreitol were purchased from Sigma-Aldrich (St. Louis, MO). The digestion standard, MassPREP BSA digest, and the acid-labile surfactant, RapiGest SF, were provided by Waters Corporation (Milford, MA). The soluble fraction from Baker's yeast cell lysates were the same as have been used in previous chapters and were also provided by Waters. While all three samples were analyzed by anion exchange, fractions were collected and analyzed only for the samples grown on different carbon sources and harvested at the log phase of growth.

4.2.3 Instrumentation and run conditions for anion exchange separation

The anion exchange separation was performed on a commercially available Waters BioSuite Q strong anion exchange column with dimensions of 75mm x 7.5mm ID. The packing material was almost identical to that used in the long anion exchange column from Chapter 2. The only difference was that the particle diameter was slightly smaller at 10 μm as opposed to a 13 μm diameter for the long anion exchange column. Gradients were delivered using a Waters 600 quaternary gradient LC pump at a flow rate of 0.5mL/min. The mobile phase composition of both mobile phases A and B was ammonium acetate adjusted to pH 9.0 with ammonium hydroxide. The ammonium acetate buffer concentration of A was 10 mM and the concentration in B was 750 mM. The full gradient profile can be found in

Table 4-1. Monitoring of the separation was performed by UV detection on a Waters 2487 dual wavelength detector set at a wavelength of 280 nm. One-mL fractions were collected every two minutes beginning two minutes after injection through to 72 minutes

after injection for a total of thirty-five fractions collected per sample. The dextrose-grown stationary phase sample was analyzed first without fraction collection followed by the glycerol-grown and dextrose-grown log phase samples with fraction collection.

Fractions were lyophilized by vacuum centrifugation and reconstituted in 50 μL of 50 mM ammonium bicarbonate. Twenty-five μL was withdrawn from each centrifuge tube and placed in a sample vial for intact protein analysis. An additional 25 μL of DI water was added to each sample vial to dilute the buffer concentration and increase the volume. The increase in volume was necessary to allow for at least two possible injections of each fraction in the event that instrumentation difficulties arose.

4.2.4 Capillary UHPLC of intact protein fractions at elevated pressures

The instrumentation of the gradient UHPLC system has been described previously.[8-10] A diagram of this system is provided in Figure 4-2. Briefly, the complete system is composed of multiple commercially available components in addition to a custom-built hydraulic amplifier, which was designed and fabricated by Waters Corporation for the production of gradients at pressures up to 40,000 psi. The commercially available components include a Waters integrated CapLC-PDA system, which is comprised of an autosampler, CapLC pump, and a PDA detector (not used), and a Waters 1525 binary gradient pump. The custom-built pump provided by Waters is a hydraulic amplifier that receives working fluid from the 1525 binary gradient pump. Brake fluid was used as a non-compressible working fluid.

In the operation of the gradient UHPLC system, the integrated CapLC system is used to both load a gradient onto the gradient storage loop and perform the sample injection in an automated manner. The status of all valving and direction of flow during this process is

illustrated in Figure 4-2 A. The gradient is loaded first in reverse as a pre-column conditioning method in the software, followed by a sample plug from the injection cycle. While there is no valve preventing the sample from entering either the flow splitter or the analytical column, the flow restriction provided by these paths is roughly 200 times and 8,000 times, respectively, that of the gradient storage loop. Once the gradient and sample have been loaded past the 4-port micro-volume cross, the flow from the CapLC is diverted to waste and both of the pin valves are closed. The hydraulic amplifier flow is initiated and the sample and gradient are forced onto the column in that order as shown in Figure 4-2 B. After the run is finished the vent valve is opened to release the pressure slowly while still isolating the CapLC from the high pressures. Once the hydraulic amplifier pressure drops below 1,000 psi, the CapLC vent valve is opened and the system is ready for another gradient loading and injection cycle.

The capillary column used for the separation was 15 cm x 50 μm ID packed with 1.5 μm bridged ethyl hybrid (BEH) C18 particles from Waters. Mobile phase A was comprised of 0.2% (v/v) formic acid in water and mobile phase B was acetonitrile, also with 0.2% (v/v) formic acid. The gradient was loaded such that the hydraulic amplifier would produce a 30 minute gradient from 20 to 60% B at a flow rate of 4 μL/min. The flow was split at a 10:1 ratio such that only 400 nL/min flowed through the column.

4.2.5 Mass spectrometric analysis of intact proteins

The outlet of the column is connected to a pulled glass capillary spray tip (New Objective, Woburn, MA) to which 2 kV is applied only when the hydraulic amplifier is flowing to preserve the integrity of the tip and achieve electrospray ionization for analysis on a Micromass LCT mass spectrometer scanning from *m/z* values of 450-1600. Data are

acquired from the time the hydraulic amplifier flow is initiated until the system is depressurized. Analysis of the mass spectra is performed in an automated fashion using AutoME just as was done in Chapter 2. The processing parameters for the maximum entropy deconvolution are shown in

Table 4-2. The AutoME output is loaded into Igor Pro graphing and analysis software (Wavemetrics, Lake Oswego, OR) for further summing and duplicate removal across scan ranges and fractions. The output from all of the processing is a mass list with a corresponding intensity and retention time of all of the detected proteins. Intensity values from the protein lists generated from each of the yeast samples are compared within specified retention time (3 min), fraction range (3 fractions) and molecular weight tolerances (10 Da) to determine differentially expressed proteins.

4.2.6 Digestion of intact proteins and analysis of resulting peptides by UPLC-MS^E

The digestion conditions were identical to those used in the previous chapter with minor differences to account for the different amounts of protein likely to be present in each fraction. The amount of trypsin to add was determined by using a 25:1 protein: enzyme ratio with a protein amount corresponding to what would be present in each fraction assuming equal distribution of protein across all fractions. This is not what actually occurs, but has served as a good approximation in previous experiments. LC-MS^E was performed on a Waters nanoAcquity LC system coupled to a Q-TOF Premier mass spectrometer, also from Waters. The LC system was run in trapping mode with the separation occurring on a 25 cm x 75 μm ID analytical column packed with 1.7 μm BEH-C18 particles. Mobile phases A and B consisted of 0.1% (v/v) formic acid in water and 0.1% (v/v) formic acid in acetonitrile, respectively. A 60 minute gradient program was used ranging from 5 to 40% mobile phase B

before ramping up to 85 % B for a short period to clean off the column. Mass spectrometric analysis was performed in a data-independent manner using MS^E. Data were processed using ProteinLynx Global Server 2.4 RC7 with the same parameters as described in the previous chapter in order to gain both identification and absolute quantification of the proteins. Only replicating proteins were used for further analysis so that the significance of fold changes in the comparison between differential samples could be determined. The output from PLGS was loaded into excel to sum duplicates across fractions. The list of protein identity and absolute quantitation was compared to an equivalent list generated from another sample to identify differences in protein expression. All fractions were analyzed in duplicate as a measure for both fractionation and system reproducibility.

4.2.7 Comparison of quantitation based on both top-down and bottom-up data

The comparison of quantitative data obtained from the two proteomic strategies was done in two different ways. The first was on a fraction-by-fraction basis. The sample complexity of each fraction should be much less than the sample as a whole and should be easier to compare across methods. Because only the bottom-up analysis gives absolute quantitation data, relative changes in intensity across the differential samples must be compared. For example, the fold change of protein 'A' across fraction 'X' in the bottom-up comparison would be compared to the fold change of a protein with a similar intact mass in the same fraction. The second method of comparison was done with the summation of intensities of identical proteins in the bottom-up analysis and in the top-down analysis whereby protein intensities of equivalent molecular weight (± 10 Da) and similar retention time (± 3 min.) were summed.

4.3 Results

4.3.1 Anion exchange fractionation of intact proteins

UV chromatograms of the anion exchange separations are shown in Figure 4-3. Vertical lines designate the fractionation that was performed on the samples. For this analysis, fractions were collected only for the glycerol-grown yeast sample and the dextrose-grown yeast sample that were harvested at the log phase. The lysate of cells harvested at the stationary phase was analyzed initially to determine the range over which fractions should be collected for subsequent analyses. The UV trace of all three samples is shown in Figure 4-3 A. In this plot, the three chromatograms are overlaid in order to examine overall changes in UV intensity. In Figure 4-3 B, the samples harvested at the log phase are plotted with an offset to better visualize the individual chromatograms. As stated in the experimental section, fractions were collected at 2 minute intervals starting 2 minutes after injection and continuing to 72 minutes for a total of 35 fractions per sample.

4.3.2 Reversed-phase separation and analysis of intact proteins

After lyophilization and reconstitution of anion exchange fractions, half of each fraction was analyzed by UHPLC-MS to determine the intact molecular weights of proteins in the complex mixtures. The separation was performed at ultra-high pressures (23 kpsi) on capillary columns (50 μm ID) as opposed to standard HPLC operating pressures of less than 6 kpsi on standard bore stainless steel columns (4.6 mm ID). For comparison, two chromatograms are shown in Figure 4-4, one from a standard HPLC intact protein separation and one from a UHPLC intact protein separation. In both cases, the sample analyzed was of a fraction from an anion exchange separation. The HPLC separation was performed online, so the injection was a continuous 30 minute stream of the effluent from the long anion

exchange column. This in contrast to the UHPLC separation in which anion exchange fractions were collected, lyophilized, reconstituted, and injected onto the capillary column using an autosampler equipped with an injection loop. The asterisk in each chromatogram denotes the same protein peak in each analysis.

4.3.3 Differential comparison of 2D intact protein chromatograms

LC-MS data from the analysis of the intact portion of each protein fraction were processed with AutoME and subjected to the same data analysis strategies as described in Chapter 2 for the on-line 2D separation of intact proteins. The 2D chromatograms reconstructed from the UHPLC separations of individual anion exchange fractions are shown in Figure 4-5 with the dextrose-grown yeast shown in A and the glycerol-grown yeast shown in B. The maximum intensity for a deconvoluted protein peak was over 300,000 counts; however these chromatograms are plotted at a z-axis maximum intensity of 50,000 in order to be able to see some of the less intense peaks.

A mass list was generated of proteins that appeared to change in intensity based on color changes seen in the comparison of 1 kDa mass slices. This list of protein masses is included in Table 4-3. The fraction in which the protein was most intense in each sample is included, along with the deconvoluted mass. Intensity values are a summation of the intensity in the fraction indicated and the intensities in the two neighboring fractions. A fold-change was determined by dividing the greater of the two intensities by the lesser of the two. To determine the significance of the protein intensity differences, a log/log intensity scatter plot was constructed of the differential proteins from the mass slice comparison with the confidence lines from Chapter 2 included. This is shown in Figure 4-6. The percent probability that the intensity difference is significant based on these error curves is included

in the last column of the data. From Chapter 2, the confidence threshold at which intensity differences were considered significant was arbitrarily set at 96.5%. Using the same threshold, three masses that appeared to show a significant change in color in the mass slice chromatograms fell inside the 96.5% confidence lines.

4.3.4 Identification of differential proteins based on molecular weight

The identification of proteins determined to have differential expressions based on the intact protein MS data was performed at the peptide level after digestion by LC-MS^E. The remaining half of each anion exchange fraction was used for this purpose. After analyzing each sample and processing the data with PLGS2.4, masses of the identified proteins were compared with those determined to have differential expressions from the TD data. The masses used from the BU data were the average masses of the protein, as opposed to monoisotopic, and included the removal of the N-terminal methionine, if known to occur, and any other reviewed modifications. If the two protein masses differed by less than 100 Da, they were considered the same, although the average difference was less than 25 Da with a median difference under 5 Da. Of the 39 masses determined to differ in intensity between the two samples, 22, or 56%, were successfully identified. Two out of the three highlighted proteins in Table 4-3 that had intensity differences that were not deemed to be significant were also not identified in the BU analysis of the same fraction.

4.3.5 Protein identification and differential comparison based on peptide data

Half of each anion exchange fraction was digested and subjected to analysis by LC-MS^E to identify as many proteins as possible in each sample. Every fraction was analyzed regardless of whether or not a potential differentially expressed protein mass was detected in that fraction from the intact protein analysis. A small amount of a MassPREP BSA digest

was added to each digested fraction as a means to obtain absolute quantitation of the proteins present in each fraction and sample. Each fraction was analyzed twice to determine the reproducibility of the absolute quantitation. A log/log plot was created to compare the changes in quantitation in each of the replicate analyses. This is shown in Figure 4-7 A along with a comparison in the replication from the un-fractionated BU experiment discussed in Chapter 3 in Figure 4-7 B. Statistics of the BU analysis are given in Table 4-4.

The initial runs of each sample were acquired over a time period of one week. The replicate runs were acquired over the same period of time roughly three weeks later. For the purpose of the statistics shown in the table, the initial runs of the 35 fractions from the dextrose-grown yeast sample are grouped to give the data under the 'Dextrose 1' column and the replicated runs of the fraction were grouped separately under the 'Dextrose 2' column. The same grouping was performed on the glycerol-grown fractions. The total number of proteins identified includes redundant proteins found in more than one fraction. The total peptide mass injected was calculated by summing the amounts output by PLGS2.4 across all 35 fractions. The mass accounted for was determined by multiplying the total mass injected by 57 to account for a 1 μL injection of the 57 μL of sample present after digestion. On average, 177 μg of the 1.11 mg of total protein was accounted for (only half of each fraction from the original 2.22 mg injection was digested).

4.3.6 Comparison of relative protein abundance between both methods on a fraction-by-fraction basis

A flow diagram for the data processing used to compare fold changes from the TD and BU data for each fraction individually is shown in Figure 4-8. In order to determine protein abundance in each fraction from the BU data, the replicates of each fraction from

both samples` were compared to determine the average abundance in fmol of a protein in each fraction. Proteins that replicated between the two runs of each fraction from the dextrose-grown yeast sample were then compared on a fraction-by-fraction basis to the proteins that replicated in the glycerol-grown yeast sample. A list was generated that compared the average protein abundance in each fraction to determine a fold change.

Due to the fact that the TD data does not have a unique identifier associated with each protein, such as is the case with the SwissProt name for the BU data analysis, windows needed to be used to sum intensities of masses within a specified range. A retention time window was also set to insure that only intensities from identical proteins would be compared. To generate protein lists for each sample, the AutoME-processed data was analyzed such that the intensities of protein masses less than 10 Da apart and eluted within a retention time window of 3 minutes would be summed. This was performed individually for each anion exchange fraction.

Overall, there were 67 proteins that were identified in both BU replicates of both differential yeast samples that had masses matching those from the AutoME deconvolution of the intact protein MS data. A list of these proteins with the corresponding fold changes determined form each analysis method is included in Table 4-5 (proteins with greater abundance in dextrose in the BU analysis) and Table 4-6 (proteins more abundant in the glycerol sample from the BU data). To facilitate the comparison of fold changes between the two methods, a bar graph was constructed where the fold changes of each protein in both forms of analysis were plotted side by side. This is shown in Figure 4-9. Fold changes in which the proteins were more abundant in the dextrose sample are shown as positive values and those determined to be more abundant in the glycerol sample were plotted as negative

values. To see the differences of the lower fold changes, three protein fold changes were allowed to go off scale. The true fold change for these proteins is written in the plot area.

4.3.7 Comparison of relative protein abundance at the whole sample level

The plots described in the previous section include no designation of whether or not the fold change of the differential proteins would have been considered significant in either comparison. Figure 4-10 includes only those proteins that were both considered significantly different in the TD comparison and identified in both samples of the BU analysis. All but one of these proteins was also considered significantly different in the summation of protein abundance across three fractions in the bottom-up data. Also, only one protein did not have significant intensity differences when total abundance across all 35 fractions was summed. Fold changes that were not considered significant are designated with an asterisk in the figure.

4.4 Discussion

4.4.1 Use of the short anion exchange column for fractionation

The initial separation of the differential yeast cell lysates was performed using a different anion exchange column than the one used in the online separation from Chapter 2. In an online separation, the first dimension is often operated at a decreased flow rate in order to allow time for the second dimension separation to occur. The longer anion exchange column held an advantage of the short column when operated in this manner since it showed less decrease in the performance than did the short column when operated a lower flow rates. This is very important for separation performed in an online manner since, ideally, all resolution gained in the first dimension would be preserved in the transfer to the second dimension. Contrastingly, in theory, offline separations are not limited by the time needed to

perform the second dimension separation. Practically speaking, however, there is a time frame in which it would be preferred to be able complete an experiment. Because the combined TD/BU offline separation performed here required that each sample be analyzed at least twice (a TD LC-MS analysis and a BU LC-MS^E analysis), a realistic number of fractions to be collected was around 30.

In order to have a faster, more reproducible anion exchange separation, a shorter column was used. By using the shorter column in an offline separation, a flow rate near the optimum for the column could be used as opposed to slowing down the separation to wait for the second dimension as was done for the online separation. Also, the gradient was programmed such that an equivalent number of column volumes would be used as were used with the long column. The separation could be performed much more rapidly since the volume of the column was smaller, and a faster flow rate could be used. Using this strategy, all three differential yeast samples were separated by ion exchange in a single day versus one sample per day with the long column online separation.

One of the drawbacks of using a shorter column is that there is usually some loss in efficiency. However, the long anion exchange column was under-sampled so the slightly better efficiency offered was lost in the fractionation step. In looking at the anion exchange fractionation shown in Figure 4-3, peaks generally eluted over 3 fractions or less, which is closer to the recommended sampling rate for multidimensional separation of three times across the 8 σ width at the base of the peak.[11]

Initially, the dextrose-grown yeast cell lysate that was harvested at the stationary phase was analyzed to verify that thirty-five 2 min fractions would be sufficient to collect all proteins eluting from the column. The gradient described in Table 4-1 was used for this

initial analysis and was also used for the subsequent analyses of the differential yeast samples. In Figure 4-3 B, the two yeast samples that differed in preparation by growth nutrient are plotted with an offset to see if a similar peak pattern was present as was seen in the 2D chromatograms from the online experiment. Looking at just one dimension of separation, while there are some similarities between the samples, the extent is not as readily apparent as it was in the 2D chromatograms in Chapter 2 or even in Figure 4-5 of the intact protein separation from this analysis.

4.4.2 Reversed-phase separation of intact proteins after fractionation by anion exchange

The advantages of using a UHPLC separation for intact proteins over a standard HPLC separation were discussed in section 4.1.2 and include improved protein recovery, decreased protein carryover, and improved peak shape. Due to pressure limitations of commercially available valves and pumping systems, it was not feasible to couple this to an online separation strategy. For the offline separation, however, the application of a UHPLC separation was attractive. Two chromatograms are shown in Figure 4-4, the first from the online 2D separation of the dextrose-grown yeast sample and the second from the second dimension UHPLC separation in the offline analysis of the same sample. The chromatogram shown in A is the 14^{th} anion exchange fraction from the online separation and in B, it is the 16^{th} fraction from the offline separation. These fractions were selected for comparison due to the presence of the same protein peak in both fractions. The protein peak with an intact mass of 9,930 Da is designated with an asterisk in both chromatograms. The peak width of the protein in A was 30 sec and in B it was 20 sec. These peak widths correspond to peak

capacities of roughly 30 for the HPLC separation and 60 for the UHPLC separation taking into account the amount of time over which peaks actually eluted during the gradient.

Protein carryover observed in subsequent blank injections that is often seen with the analysis of proteins on silica-based reversed phase columns was not detected with these UHPLC separations. For the online separation, the polymer-based columns were selected due to the requirement for no protein carryover, since the use of blank gradients to clean off columns between each run was not feasible. The low pressure limitation of the polymeric particles (450 psi) limited the speed at which the separation could be performed. Through the use of ultra-high pressures (23 kpsi), protein carryover was eliminated and fractions were able to be analyzed without the need for blank runs in between to clean off the column. If carryover was present and blank gradients were needed, the analysis time would have doubled.

4.4.3 Differential analysis of 2D chromatograms from TD analysis

2D chromatograms were constructed from the UHPLC-MS analysis of the intact portion of the anion exchange fractions. These 2D chromatograms are analogous to those shown in Figure 2-4 from the online separation. A visual comparison of the mass slice chromatograms was performed to identify identical masses (+/- 10 Da) with different intensities in the two samples by picking out peaks that changed color. Comparison of the dextrose-grown sample to the glycerol-grown sample resulted in the selection of 39 masses with changing intensities with 13 more intense in the glycerol-grown sample and 26 more intense in the dextrose-grown sample. The list of differential protein masses is included in Table 4-3. Over half, 22, of the differential protein masses eluted from the anion exchange

column over the fraction range from 10 to 15 indicating that it may have been better to run a shallower gradient over this elution time.

The determination of whether or not the difference in intensity was significant was performed using the replicate data from Chapter 2. This analysis was performed using the same mass spectrometer, LCT classic, and thus would have the same instrumental variability. The error curves discussed in section 2.3.2 were used to determine a cutoff for whether or not a difference was significant. These are plotted in Figure 4-6 with the intensities of the differential protein masses selected in this experiment. Unlike in Chapter 2, all but one of the masses had intensity differences that were significant above the 96.5% threshold. Even the three proteins that fell inside the 96.5% confidence lines, highlighted in the table, had a difference that fell outside the 94.0% confidence line.

One reason for this improved rate of the selection of significantly different protein masses through the visual comparison of mass slices is related to the purpose of this experiment, which is to examine the relationship between protein fold changes identified in a TD experiment and those calculated in a BU experiment. For this reason, differential protein masses were only selected from the mass slices if there was a corresponding peak in the same mass range and location in the separation space of both samples, since only relative changes in intensity could be compared. This improvement would likely not be seen if proteins present in only one sample were included in the list as they were usually lower in intensity and therefore likely contained more error in the measurement.

4.4.4 Identification of proteins in the BU proteomic analysis

After removing an aliquot of the reconstituted anion exchange fraction for TD analysis, the remainder was digested and analyzed by LC-MS^E with the addition of a small

amount of a standard BSA digest to allow for absolute quantitation. The protein identification statistics are included in Table 4-4 The total number of proteins identified and number of unique proteins identified is higher than seen for the un-fractionated analysis discussed in Chapter 3. This was to be expected since the individual anion exchange fractions should have been greatly simplified compared to the original sample. The numbers of replicating proteins were roughly twice that observed in Chapter 3, even with just two replicates analyzed for each sample as opposed to three in the un-fractionated analysis. The average number of fractions in which a protein was identified was 2, which indicates that the sampling of the anion exchange separation was less than recommended. One potential advantage to under-sampling would be that proteins would be more abundant in each fraction and therefore be more likely to be detected.

Based on the addition of a standard protein digest, the average amount of protein present in the original samples was 177 µg. The amount in the original sample was calculated based on 1.75% of each anion exchange fraction being injected for LC-MS^E analysis. This is just over 15% of the actual 1.11 mg that was actually in the sample based on the Bradford assay of the sample prior to fractionation. With an average of 531 unique protein identifications, which equates to approximately 8% of the total 6,500 entries in the yeast database, it is likely that the remainder of the protein is spread across many of the unidentified proteins. These proteins were not identified either because they were present below the limit of detection for this method or because they were overshadowed by peptides from the more abundant proteins.

There was a concern of whether or not fractionation of the sample would increase the error in absolute quantitation in comparison to analysis of the samples without fractionation.

To investigate this, the replicate data from this experiment were plotted along with the replicate data from the experiment in Chapter 3 shown in Figure 4-7 B. Visual inspection of the scatter plot does not indicate that there is a significant decrease in the reproducibility of absolute quantitation upon fractionation of a sample. The calculation of correlation coefficient confirms this with a value of 0.94 for the fractionated analysis and 0.91 for the un-fractionated experiment.

4.4.5 Comparison of fold changes observed for identified proteins selected as differentially expressed in the TD experiment

In order to generate protein mass lists from the AutoME deconvoluted data in the TD analysis, thresholds had to be set to determine which protein masses would be considered equal and subsequently summed. Unlike in a BU experiment where the identity of each protein is known and can be used as a unique identifier to compare multiple lists, the TD data, which includes deconvolution of 10 sec increments in each RP run to obtain a protein mass and an associated error close to 150 ppm, comparison of protein intensities between samples was not so straightforward. With regard to the summation of intensities within each fraction, a retention time window was set to 3 minutes. It is unlikely that a protein would be spread across this broad a retention time window given that most of the peaks were less than 1 minute wide. It is even more unlikely that a second protein with a deconvoluted mass within 10 Da of the first protein would be present within this window. The determination of the number of anion exchange fractions that a protein would elute into was based on the peak widths from the anion exchange fractionation. The threshold was set such that a 3 fraction window was used to sum intensities across fractions. This should be sufficient to encompass

the entire protein peak from the 2D separation space. Protein intensities and fold changes of the TD data in Table 4-3 and Figure 4-10 are the result of this thresholding.

In order to include a relevant comparison of the fold changes calculated after mass slice comparison in the TD analysis, the absolute protein abundance determined by PLGS2.4 searching of BU data of each fraction was summed in the same manner using the values from the same 3 fractions used in the TD analysis. In Figure 4-10, this fold change is reported in blue. As an additional source for comparison, the fold change calculated using the total protein abundance in each sample without the 3 fraction limitation, is also included as the green bars. Also, the error associated with the replicate analysis of the BU fractions was used to determine whether or not the fold change was significant. Fold changes where the error bars associated with the abundance of a protein in both samples of the BU analysis overlapped are denoted with an asterisk and are not significant.

Overall, the fold changes from the TD analysis are greater in magnitude than those from either BU processing strategy. The average fold change for the comparison of intact protein MS intensity was 12.3 regardless of which sample the protein was most abundant. This is in contrast to the average fold changes from the summation of 3 fractions of BU data and total protein abundance across all fractions in the BU analysis which were 3.9 and 2.8, respectively. The smaller average fold change for the total protein abundance in the BU analysis is not surprising since all isoforms of a given protein would be combined into one intensity value. If there were a differential regulation of the downstream processing of the protein, this would be overlooked by an analysis of this type. Therefore, it is more likely that fold changes calculated from the summation of protein abundance in 3 anion exchange

fractions would follow the same trend of the fold changes from the TD data than would the total protein abundance across all fractions in which it was present.

Looking only at the first two types of analysis in Figure 4-10, there should be better alignment as compared to the total BU analysis. Overall, 20 of the 21 proteins identified as significantly from the TD data were also determined to be significantly different in the 3 fraction BU analysis. Fifteen of these 20 proteins were also up-regulated in the same yeast sample. The five proteins that had conflicting fold changes based on the different analyses also tended to have a greater difference in the theoretically computed MW and the experimentally determined intact MW. The average difference between the two was 48.6 Da for proteins with opposite direction of fold change in contrast to 17.7 Da for proteins that were determined to be up-regulated in the same sample across both methods. This may be an indication that the protein identification of the differentially expressed intact mass was incorrect and the comparison is actually being made for different proteins.

Overall, 75% of the differentially expressed proteins identified in the TD experiment were determined to be up-regulated in the same yeast sample in the BU experiment; however, there was generally a difference in magnitude. Nevertheless, the fold changes were considered significant in both forms of analysis and individual TD and BU experiments would have produced the same outcome for the generation of the list of differential proteins. For proteins that had significant fold changes in both analyses, but were up-regulated in opposite yeast samples, the overall increase in the difference between the theoretically computed MW and the experiment MW indicates that the correlation of identified proteins in the BU analysis and the masses from the TD analysis may not have been correct.

4.4.6 Comparison of protein fold changes determined from the TD data with those determined by BU proteomics on a fraction-by-fraction basis

In the previous section, only proteins that had significant changes in intensity in the TD experiment were analyzed. Data processing involving the summation of protein intensities across multiple anion exchange fractions to account for splitting of a protein peak across multiple fractions was performed in order to look at the change in the expression of a protein across the entire sample. While this approach is valid and necessary to identify differentially expressed proteins, it still involved looking at the 2D separation as a whole. In this chapter, where the goal was to determine the correlation of fold changes between TD and BU data rather than to identify differential proteins, this was not necessary. The anion exchange fractionation was meant to decrease the sample complexity and potentially separate out protein isoforms that may not have been separated in a complete BU analysis, such as the one performed in Chapter 3. Therefore, for the purpose of looking at protein fold changes in both types of analyses, it was useful to compare the changes on a fraction-by-fraction basis.

Essentially, each anion exchange fraction was treated as a separate sample. For example, the intensity of a protein in fraction X of one sample was only compared to the intensity of the same protein in fraction X of another sample. This comparison was performed for both the TD and BU experiments with similar mass and retention thresholds used for the summation of TD protein intensities. There was no bias as to whether or not changes in protein abundance were significant or not. The list of all of the proteins identified and the fraction in which each was identified is shown in Table 4-5 and Table 4-6 and a bar graph comparison of fold changes is shown in Figure 4-9. While there were 67 proteins that were identified in both replicates of both samples in the BU data and both samples in the TD

data, there are redundant proteins within that list. This is due to the comparison being made at the anion exchange fraction level. Proteins split over more than one fraction are treated separately in this comparison strategy.

As plotted, proteins on the left were consistently up-regulated in the dextrose-grown sample and those on the right were consistently up-regulated in the glycerol-grown sample. There is a greater similarity in the magnitude of the fold changes for protein up-regulated in dextrose; the median factor of the difference in fold changes was 1.2 for the proteins up-regulated in dextrose and 2.1 for the proteins up-regulated in glycerol. The proteins up-regulated in glycerol in both analyses showed the same trend as was mentioned in the previous section, where the magnitude of the difference is greater in the TD comparison than it is in the comparison of BU abundance.

Proteins centrally located in the graph are those which were up-regulated in conflicting samples between the two forms of analysis. These account for 17 of the 67 proteins in the plot, or 25% of the proteins identified in all analyses. This follows the same trend as was seen in the comparison of only the differentially expressed proteins from the TD in the previous section. While not as different as was seen in the previous section, proteins with conflicting fold changes were seen to have a slightly higher difference in the computed MW as compared the experimental MW with an average difference of 18.8 Da as opposed to 15.4 Da for proteins with consistent up-regulation. This difference is relatively small compared to previous section so a higher instance of misidentification is likely not the cause.

A potential cause for the discrepancy of which sample a protein was up-regulated in could be the uncertainty of the intensity or abundance measurements. For example, in the BU analysis, there was greater occurrence of protein fold changes that were not statistically

significant in the proteins that had conflicting directions of differential expression. Five proteins located in the center of the plot had intensity differences that were not considered significant based on the replicate analysis. Only six of the proteins that were up-regulated consistently in the same sample in both the TD and BU analyses had fold changes that were not statistically significant. Proteins that were up-regulated in opposing samples in two types of analyses are indicated with an asterisk in Figure 4-9.

4.5 Summary and Conclusions

BU proteomic experiments have been the most widely used in the field of differential proteomics for the past two decades. However, there has been a recent surge in the use of TD proteomic strategies due to improved instrumentation and data analysis software. Even with the prevalence of both methodologies, little work has been done to compare the results obtained with both techniques when performed on the same set of samples. An anion exchange separation of intact proteins was performed on both the glycerol-grown and dextrose-grown yeast cell lysates to simplify the complex mixtures. Fractions were split in half such that each fraction could be analyzed separately both in a TD analysis of the intact proteins by UHPLC-MS and in a BU analysis of the digested proteins by LC-MS^E.

For the analysis of the intact protein half of the fraction, a UHPLC separation was performed due to the improved recovery of proteins seen at these pressures along with an improvement in peak shape. Both of these phenomena were observed in this experiment when compared to the RP intact protein separations performed as part of the on-line separation of intact proteins. The end result of these improvements was a doubling of the peak capacity and the absence of the need to run blanks in between each sample injection to clean off the column. As a result, the potential analysis time was cut in half.

The comparison of protein differential expression was made in two ways. The first involved the initial selection of differentially expressed protein masses as determined through the comparison of mass slice chromatograms generated from the TD experiment. A fold change was calculated by summing the intensity values from the AutoME deconvolution of protein with a mass range of 10 Da, and anion exchange fraction of +/- one, and a retention time window of 3 minutes. The fold change determined from this analysis was then compared to the fold change observed when the total protein abundance from the BU data was used and when just the selected three fractions were used. Regarding the fold changes that were determined for the differentially expressed proteins that were identified in the TD analysis, it appears that there is a good correlation with the fold changes calculated from the BU analysis. While the magnitude of the fold change quite often differed, 75% of the differentially expressed proteins were found to be most abundant in the same differential yeast sample and had changes that were significant in both cases. Those that differed in the direction of differential expression were more apt to have been proteins that had greater differences between the experimental and computed MWs.

Analyzing the data on a fraction-by-fraction basis offered the opportunity to remove the bias of whether or not a protein was differentially expressed and just look at the differences in relative protein abundance between the two methods. Through this comparison there were a total of 67 instances in which a protein mass was detected by TD analysis in the same fraction in both samples and matched the mass (within 100 Da) of a protein identified in both replicates of both samples in the BU experiment. The percentage of these proteins that were up-regulated in the same sample in both the TD and BU analyses was 75%, the same as was observed looking only at the differentially expressed proteins

from the TD data. A greater percentage of BU fold changes were considered insignificant for proteins that had opposing fold changes compared to those based on TD data than did proteins found to be up-regulated in the same sample in both types of analysis. This indicated that some opposing fold changes were a result of error in the quantitation rather than differences TD and BU proteomics.

Overall, there was a good correlation in the selection of differentially expressed proteins when comparing the fold changes observed in TD and BU experiments. While the magnitude was often different, the fold changes were considered significant in both cases and followed the same pattern of up-regulation between the two samples.

An experiment that would more accurately evaluate the correlation between TD and BU proteomics involves the introduction of standard proteins to the complex mixtures. Multiple standard proteins that would elute at various times in the anion exchange separation would be spiked into two cell lysates. Proteins would be spiked in at various known levels to each sample. Concentrations of the standard proteins would be different in each cell lysate to cause a specific ratio in abundance between samples. For example, one protein would be spiked into sample A at an amount four times greater than it was spiked in sample B. An equivalent hybrid TD/BU offline analysis would be performed, analyzing the same amount of cell lysate, and the fold change of the standard proteins determined both by TD and BU to see which analysis more closely matched the actual fold change based on the amounts initially spiked into the samples.

4.6 References

(1) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Jr. *Journal of Proteome Research* **2002**, *1*, 239-252.

(2) Galasinski, S. C.; Resing, K. A.; Ahn, N. G. *Methods* **2003**, *31*, 3-11.

(3) Zhu, K.; Kim, J.; Yoo, C.; Miller, F. R.; Lubman, D. M. *Analytical Chemistry* **2003**, *75*, 6209-6217.

(4) Zhu, K.; Miller, F. R.; Barder, T. J.; Lubman, D. M. *Journal of Mass Spectrometry* **2001**, *39*, 770-780.

(5) Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(6) Lubman, D. M.; Kachman, M. T.; Wang, H.; Gong, S.; Yan, F.; Hamler, R. L.; O'Neil, K. A.; Zhu, K.; Buchanan, N. S.; Barder, T. J. *Journal of Chromatography, B: Analytical Technologies in the Biomedical and Life Sciences* **2002**, *782*, 183-196.

(7) Kreunin, P.; Urquidi, V.; Lubman, D. M.; Goodison, S. *Proteomics* **2004**, *4*, 2754-2765.

(8) Eschelbach, J. W.; Jorgenson, J. W. *Analytical Chemistry* **2006**, *78*, 1697-1706.

(9) Evans, C. R., University of North Carolina at Chapel Hill, Chapel Hill, 2007.

(10) Eschelbach, J. W., University of North Carolina - Chapel Hill, Chapel Hill, 2006.

(11) Murphy, R. E.; Schure, M. R.; Foley, J. P. *Analytical Chemistry* **1998**, *70*, 1585-1594.

4.7 Tables

Time (minutes)	% Mobile Phase B
0	0
2	7
60	67
65	100
85	100
90	0

Table 4-1: Gradient conditions for the anion exchange fractionation. The pump was operated at a flow rate of 0.5 mL/min. Mobile phases A and B both consisted of ammonium acetate buffer adjusted to pH 9.0 with ammonium hydroxide. The buffer concentrations for mobile phases A and B were 10 mM and 750 mM, respectively.

AutoME Processing Parameter	Value
Processing type	MaxEnt 1
Scans to sum	10
Scans to process	270-3540
Threshold to process	20,000 counts
Subtract/smooth/center	No raw data processing
Input mass range	600-1600 m/z
Output mass range	5000-80,000 Da
Mass resolution	2 Da
Damage model	Gaussian; $w_{1/2}$ = 0.75 Da
Minimum	33% left and 33% right
Maximum iterations	30
Post-process subtraction	Not performed
Sum based on	Area
De-harmonizing	20 Da

Table 4-2: Processing parameters used for AutoME deconvolution of intact protein mass spectra.

Fraction		Deconvoluted Mass		Intensity		N-fold Change	% Prob
Dextrose	Glycerol	Dextrose	Glycerol	Dextrose	Glycerol		
9	10	33228.8	33226	512	26700	51	96.5
17	16	10971.1	10972.6	566	22700	40	94.0
3	3	13944.2	13944.2	3830	39600	10	96.5
10	10	58148.3	58148.5	4620	30500	6.6	94
12	12	35618.8	35619.4	14000	70000	5.0	98
12	12	11116.5	11117.5	14300	63200	4.4	98
28	27	23471.8	23472.4	15000	66600	4.5	98
10	10	16045.7	16045.8	15700	118000	7.5	98.7
11	11	24538.3	24538.6	16200	66000	4.1	98
11	10	24170.3	24169.9	16400	125000	7.6	99.5
16	15	54135	54134.8	16800	142000	8.5	99.5
11	11	11377.9	11343.6	16800	43600	2.6	94
24	24	30614.4	30616.1	17400	866000	5.0	98
14	14	23087.8	23087.7	20500	441000	22	99.5
12	12	11087.1	11088.1	22800	55300	2.4	96.5
11	11	33868.2	33836.5	26800	365000	14	99.5
10	10	8460.0	8460.0	27400	128000	4.7	98.7
13	14	30889.8	30880.3	31000	2120	15	96.5
3	2	38424.6	38422.8	31000	84300	2.7	98
14	15	27499.8	27494.4	39800	719	55	96.5
21	21	12249.3	12248.8	39900	1340	30	96.5
15	15	9948.1	9947.63	40900	12700	3.2	96.5
28	27	23585	23584.9	41200	183000	4.4	99.5
8	8	17304.6	17304.4	45900	84700	1.8	96.5
10	10	11855.7	11855.9	49400	97600	2.0	96.5
15	15	41084.3	41084.7	59200	3010	19	95
14	14	15815.5	15815.3	59300	112000	1.9	96.5
23	23	31279.7	31279.7	67500	167000	2.5	98.7
32	32	22619.6	22622.4	67800	138000	2.0	98
3	2	8575.27	8575.09	86500	3660	24	98.7
10	10	20907.8	20908	111000	33400	3.3	98.7
12	12	11072.1	11072.4	111000	41100	2.7	98
12	12	11621	11621.5	117000	243000	2.1	98.7
12	12	27499.1	27489.8	133000	11400	12	99.5
2	2	8559.46	8559.59	162000	103000	1.6	96.5
15	14	9932.38	9932.14	164000	10700	15	99.5
12	12	26668.5	26669.2	209000	69100	3.0	99.5
10	10	15724	15723.7	300000	186000	1.6	98
9	9	44657.2	44657.4	578000	829000	1.4	98

Table 4-3: Intact protein masses and deconvoluted MS intensities of differentially expressed proteins selected from the visual comparison of mass slices. **Fraction**: the anion exchange fraction in which the mass was most intense. **Deconvoluted Mass**: mass from the AutoME-deconvolution of the protein charge envelope. **Intensity**: intensity of protein after deconvolution. **N-fold change**: Fold change of mass calculated such that it was greater than 1.0. Red shading indicates that the protein intensity would not be considered significant based on replicate data from Chapter 2.

	Dextrose 1	Dextrose 2	Glycerol 1	Glycerol 2
Total proteins	2,329	1,946	2,444	2,060
Avg. pep/prot.	9.9	9.9	9.3	9.4
Med. pep/prot.	8	8	7	7
Total mass inj.	3.2 μg	3.0 μg	3.1 μg	3.0 μg
Mass acc. for	184μg	173μg	180 μg	171 μg
Unique prot.	514	440	605	566
Avg frac/prot	4.5	4.4	4.0	3.6
Med frac/prot	2	2	2	2
Replicating	365		450	
Found in both	303			

Table 4-4: Identification statistics from PLGS2.4 processing of the LC-MS^E data obtained in the bottom-up analysis of the anion exchange fractions.

Frxn	Entry	Comp MW	Exp. MW	Bottom-up Data				Top-down Data			
				Fmol Dex	Fmol Gly	FC	Up reg in	Dex Int	Gly Int	FC	Up reg in
14	ACBP	9930.2	9931.6	95.3	25.6	3.7	D	26900	9950	2.7	D
16	ACBP	9930.2	9931.6	48.1	15.0	3.2	D	1750	672	2.6	D
9	CYPH	17259.5	17304.5	137	52.0	2.6	D	13500	12200	1.1	D
11	ENO1	46670.9	46694.2	102	42.0	2.4	D	2450	1190	2.1	D
14	ENO1	46783.0	46793.6	66.8	25.1	2.7	D	2900	1630	1.8	D
11	ENO2	46783.0	46792.7	67.0	27.3	2.5	D	1560	928	1.7	D
12	ENO2	46783.0	46795.8	658	281	2.3	D	51500	33000	1.5	D
16	MET17	48540.5	48550.9	148	141	1.1	D	27500	22000	1.2	D
12	OYE2	44879.5	44861.6	210	33.2	6.3	D	1660	929	1.8	D
13	OYE2	44879.5	44878.8	93.5	14.9	6.3	D	1700	907	1.9	D
8	PGK	44607.2	44657.6	747	409	1.8	D	121000	54400	2.2	D
9	PMG1	27477.4	27485.8	63.2	12.5	5.1	D	11000	2800	4.0	D
10	PMG1	27477.4	27486.9	661	382	1.7	D	529000	365000	1.4	D
12	PMG1	27477.4	27499.1	901	101	8.9	D	57300	9040	6.3	D
9	SODC	15723.4	15724.2	165	86.8	1.9	D	64800	35600	1.8	D
10	SODC	15723.4	15724.3	756	477	1.6	D	206000	135000	1.5	D
11	SODC	15723.4	15724.2	267	110	2.4	D	31000	19300	1.6	D
11	TPIS	26664.3	26668.3	987	501	2.0	D	109000	46100	2.4	D
12	TPIS	26664.3	26668.5	562	88.8	6.3	D	164000	60600	2.7	D
13	TPIS	26664.3	26669.6	322	34.4	9.4	D	43400	9610	4.5	D
12	TRX2	11072.7	11072.1	812	307	2.6	D	84500	30000	2.8	D
13	TRX2	11072.7	11072.6	542	200	2.7	D	26300	11100	2.4	D
16	TRXB1	34106.9	34111.6	60.0	19.5	3.1	D	5130	1330	3.9	D
23	TSA1	21458.5	21425.5	63.2	37.2	1.7	D	2610	1860	1.4	D
3	UBIQ	8556.8	8559.4	395	7.8	50.5	D	34400	9660	3.6	D
3	CYPC	17567.0	17573.1	63.8	5.5	11.6	D	1370	17700	12.9	G
3	CYPH	17260.0	17306.5	590	6.7	88.3	D	76700	86800	1.1	G
8	CYPH	17260.0	17304.6	315	283	1.1	D	32400	72600	2.2	G
15	CYS3	42410.9	42420.3	304	197	1.5	D	7930	18900	2.4	G
3	DHOM	38502.1	38424.6	83.6	6.5	12.9	D	12900	32100	2.5	G
12	ENO1	46670.9	46693.1	89.0	30.7	2.9	D	1790	2590	1.5	G
12	G4P1	42084.0	42069.9	20.8	16.9	1.2	D	631	949	1.5	G
15	HXKB	53811.3	53821.3	86.6	60.9	1.4	D	564	4970	8.8	G
9	PGK	44607.2	44657.2	1500	1252	1.2	D	343000	600000	1.8	G
11	PMG1	27477.4	27486.1	1050	637	1.6	D	99500	248000	2.5	G
14	SODC	23083.0	23087.8	89.5	43.3	2.1	D	9120	344000	37.7	G
15	SODM	23083.0	23088.4	102	86.1	1.2	D	7700	876000	11.4	G
16	SODM	23083.0	23087.3	45.6	14.7	3.1	D	1840	5010	2.7	G
28	TCTP	18741.1	18741.7	147	143	1.0	D	17400	29800	1.7	G
12	TKT1	73674.5	73674.4	36.5	23.8	1.5	D	1530	3620	2.4	G
10	TPIS	26664.3	26668.2	1130	1000	1.1	D	1340000	1580000	1.2	G

Table 4-5: List of proteins found in both differential yeast samples in both the TD and BU analysis of anion exchange fractions that were up-regulated in the dextrose sample of the BU comparison. **Frxn**: anion exchange fraction. **Entry**: SwissProt entry followed by _YEAST. **Comp MW**: average molecular weight including reviewed modifications. **Exp. MW**: AutoME deconvoluted molecular weight from TD data. **FC**: Fold change.

Frxn	Entry	Comp MW	Exp. MW	Bottom-up Data				Top-down Data			
				Fmol Dex	Fmol Gly	FC	Up reg in	Dex Int	Gly Int	FC	Up reg in
15	ACBP	9930.2	9932.4	497	592	1.2	G	136000	703	190	D
11	CH10	11372.3	11344.9	38.7	62.4	1.6	G	16800	43600	2.6	G
11	CISY1	49216.3	49223.8	50.0	321	6.4	G	742	6030	8.1	G
8	CYC1	12050.8	12071.0	11.0	40.2	3.6	G	631	4770	7.6	G
8	DHOM	38502.1	38419.7	28.2	42.9	1.5	G	12500	18200	1.5	G
12	G3P3	35615.5	35618.8	801	823	1.0	G	9870	51000	5.2	G
13	G3P3	35615.5	35621.3	265	292	1.1	G	4150	15900	3.8	G
11	GLRX2	11833.6	11833.9	157	160	1.0	G	20800	34300	1.6	G
31	IF5A2	16983.1	16951.3	26.3	31.6	1.2	G	1270	3200	2.5	G
10	IPB2	8458.6	8460.0	73.8	400	5.4	G	27400	105000	3.8	G
11	MDHM	33832.8	33838.2	100	724	7.2	G	24800	279000	11	G
12	MDHM	33832.8	33836.8	83.9	361	4.3	G	2020	64000	32	G
13	MDHM	33832.8	33842.7	38.1	112	2.9	G	774	16600	21	G
8	MMF1	13939.9	13941.6	72.3	444	6.1	G	760	18700	25	G
9	MMF1	13939.9	13941.8	40.4	51.7	1.3	G	775	2880	3.7	G
19	SMT3	11261.7	11252.8	89.8	93.4	1.0	G	2620	2890	1.1	G
13	SODM	23083.0	23085.6	48.1	57.5	1.2	G	3680	8790	2.4	G
16	SYV	119994.9	119932	21.0	121	5.8	G	723	1600	2.2	G
10	TAL1	36905.	36953.3	94.7	263	2.8	G	1070	8160	7.6	G
9	TKT1	73674.5	73725.5	41.4	89.7	2.2	G	3250	3770	1.2	G
27	TPM1	23540.7	23586.1	40.4	90.0	2.2	G	15800	128000	8.1	G
28	TPM1	23540.7	23585	51.3	52.6	1.0	G	20400	54300	2.7	G
12	TRX1	11103.8	11116.5	258	451	1.7	G	14300	58800	4.1	G
7	UBIQ	8556.8	8559.5	23.2	26.9	1.2	G	1260	2120	1.7	G
8	UBIQ	8556.8	8558.3	208	368	1.8	G	21200	32500	1.5	G
10	YMN1	39976.1	40001.8	25.0	30.3	1.2	G	1520	1880	1.2	G

Table 4-6: List of proteins found in both differential yeast samples in both the TD and BU analysis of anion exchange fractions that were up-regulated in the glycerol sample of the BU comparison. **Frxn**: anion exchange fraction. **Entry**: SwissProt entry followed by _YEAST. **Comp MW**: average molecular weight including reviewed modifications. **Exp. MW**: AutoME deconvoluted molecular weight from TD data. **FC**: Fold change.

4.8 Figures

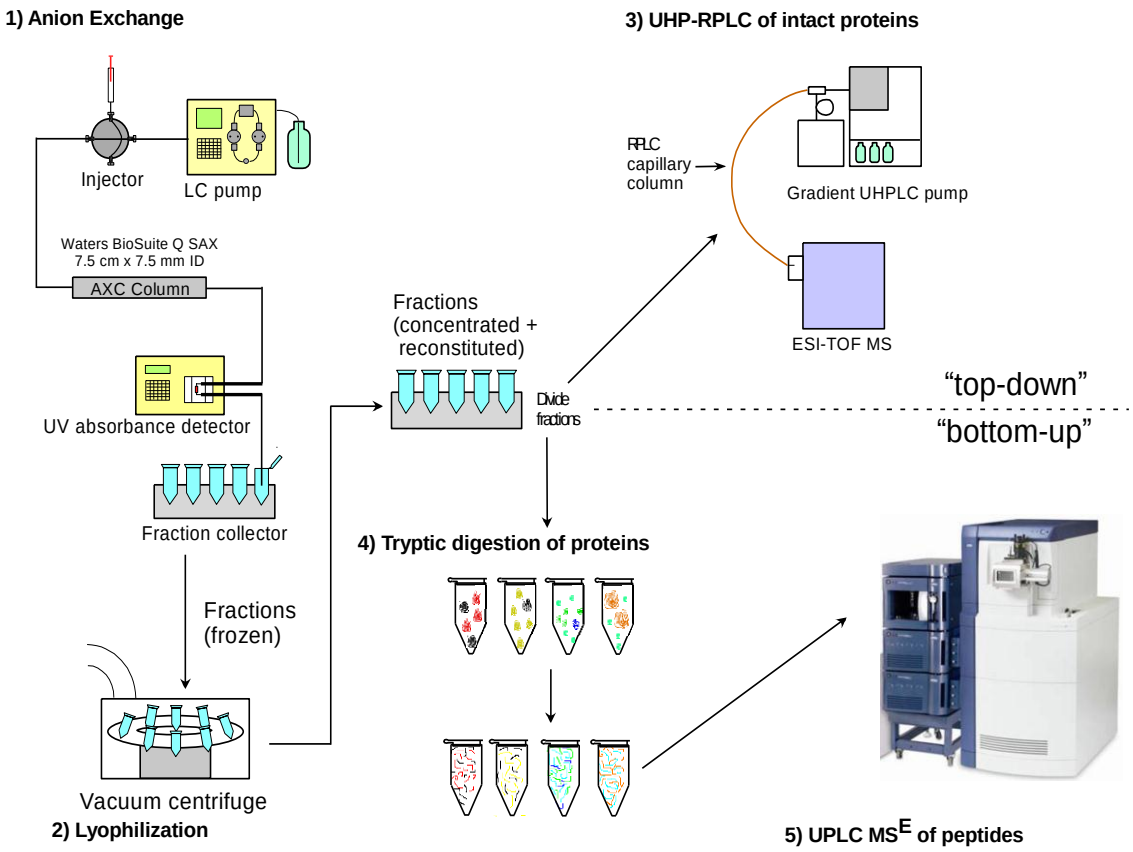


Figure 4-1: Instrumentation workflow for off-line combined top-down/bottom-up experiment. 1) Soluble fractions of cell lysates are fractionated by anion exchange. 2) Fractions are lyophilized to remove volatile anion exchange buffer and reconstituted in a smaller volume before being split in half. 3) One half of each reconstituted fraction is run on a gradient UHPLC-MS instrument for intact mass measurement and relative quantitation. 4) The remaining half of each fraction is tryptically digested. 5) Identification and quantitation is made at the peptide level by UPLC-MS^E.

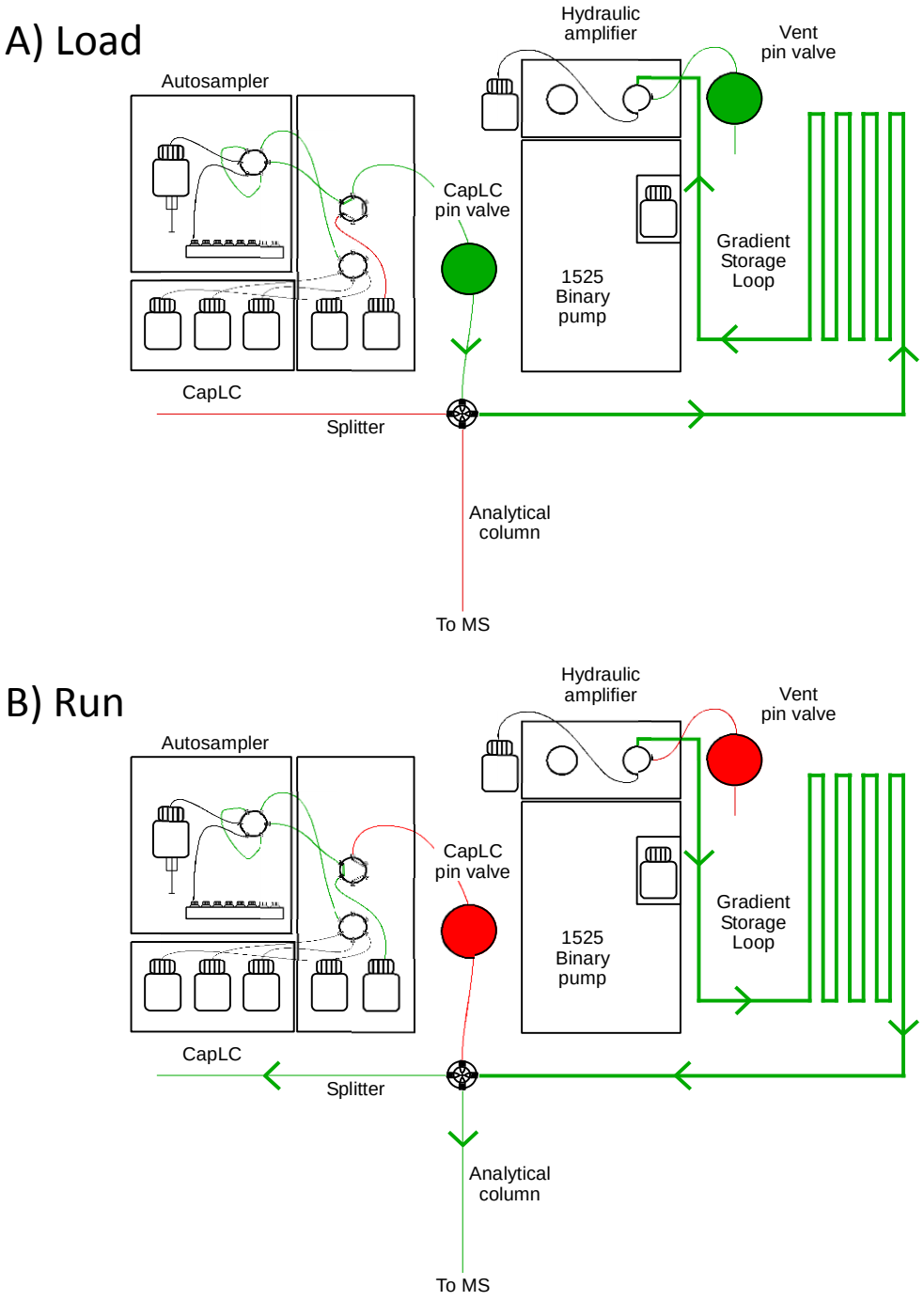


Figure 4-2: Instrument diagram for gradient UHPLC system. A) With both pin valves open and the hydraulic amplifier not pumping, the gradient and sample are loaded onto the gradient storage loop by the CapLC. B) Closing both valves and initiating flow from the hydraulic amplifier pushed the sample and gradient off of the storage loop and into the analytical column.

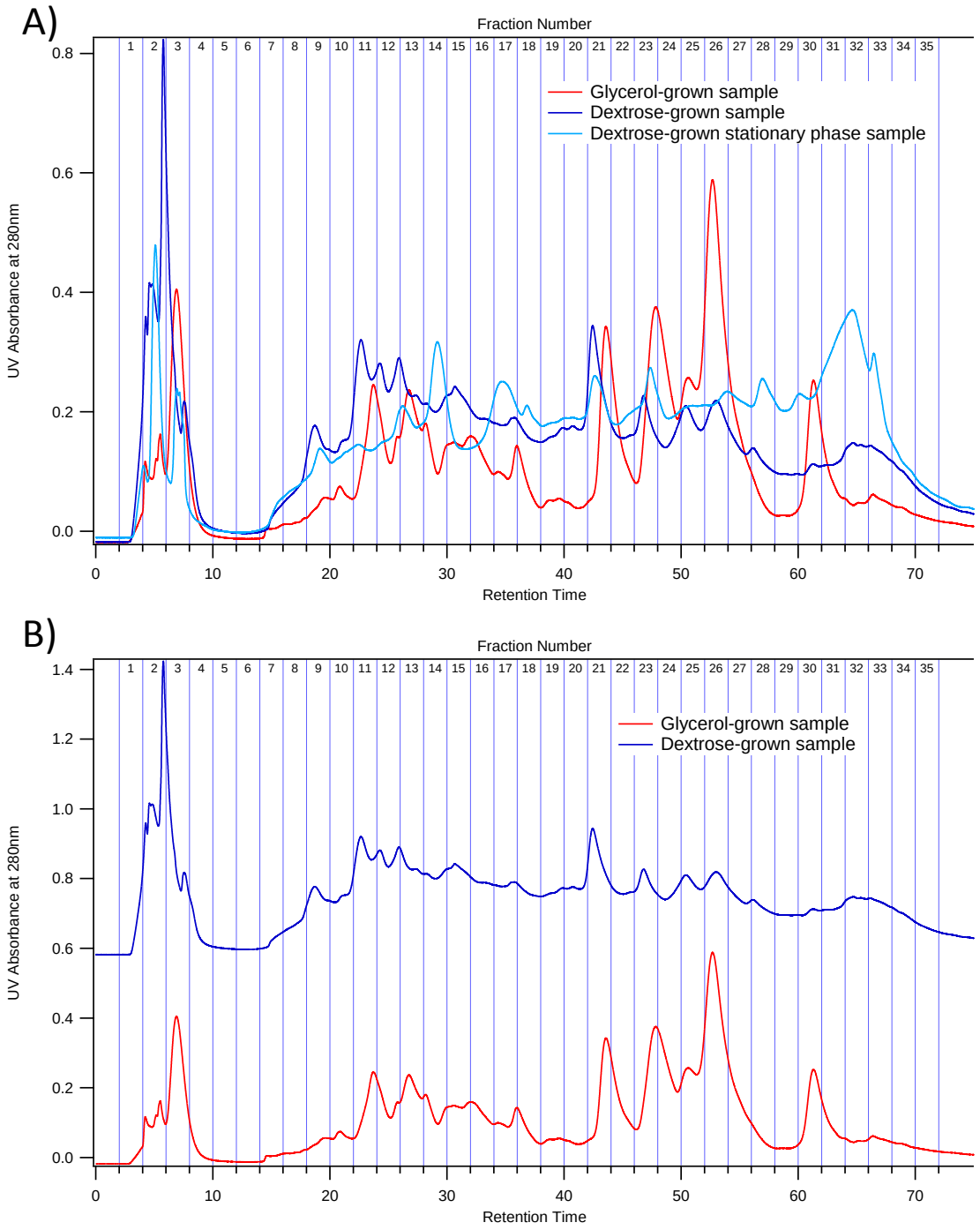


Figure 4-3: Anion exchange fractionation of the intact proteins from cell lysates of Baker's yeast samples grown on different carbon sources. Fraction collection is denoted by the vertical lines. A) Overlaid chromatograms from all three differential yeast samples. B) Chromatograms of the glycerol, log phase and dextrose, log phase samples plotted with an offset for comparison.

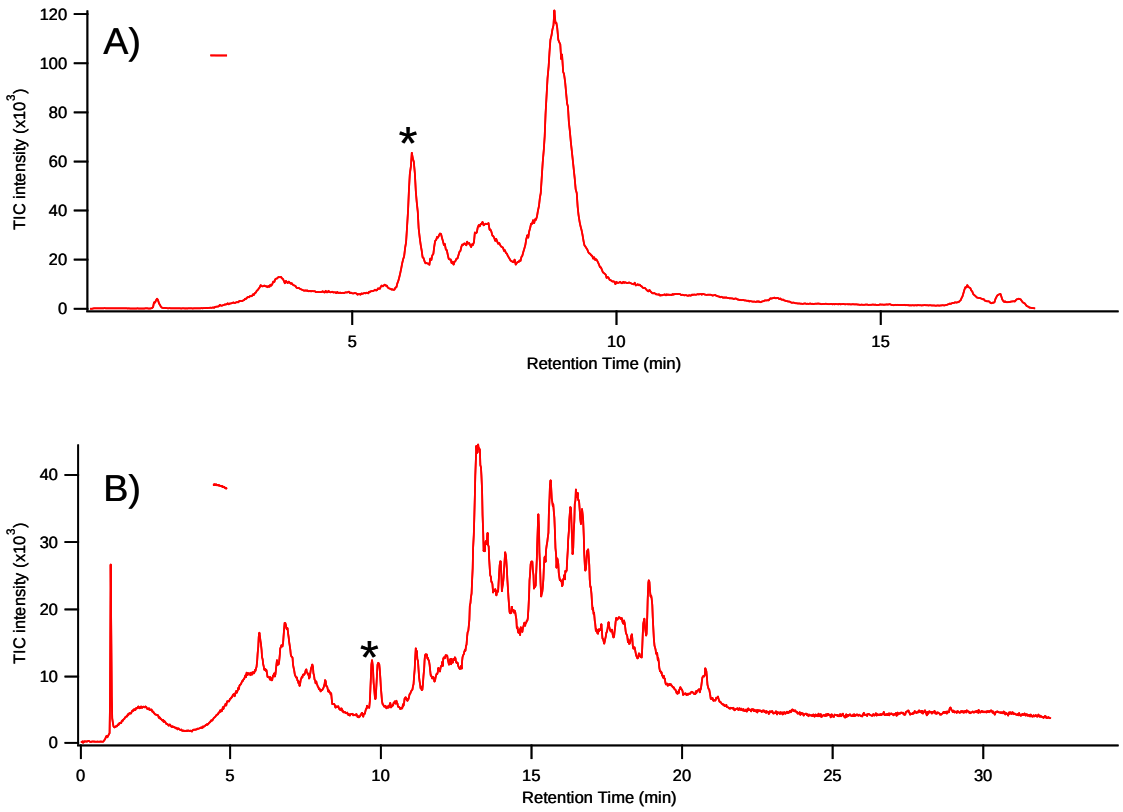


Figure 4-4: Comparison of chromatograms from the RP-LC separation of intact proteins. A) RP-LC separation of the 16th fraction taken from the long anion exchange column during the online 2DLC analyses of the dextrose-grown yeast cell lysate reported in Chapter 2. B) UHPLC separation of the 14th fraction collected from the short anion exchange column in the offline separation described in this chapter. The highlighted peak in each chromatogram represents the same protein.

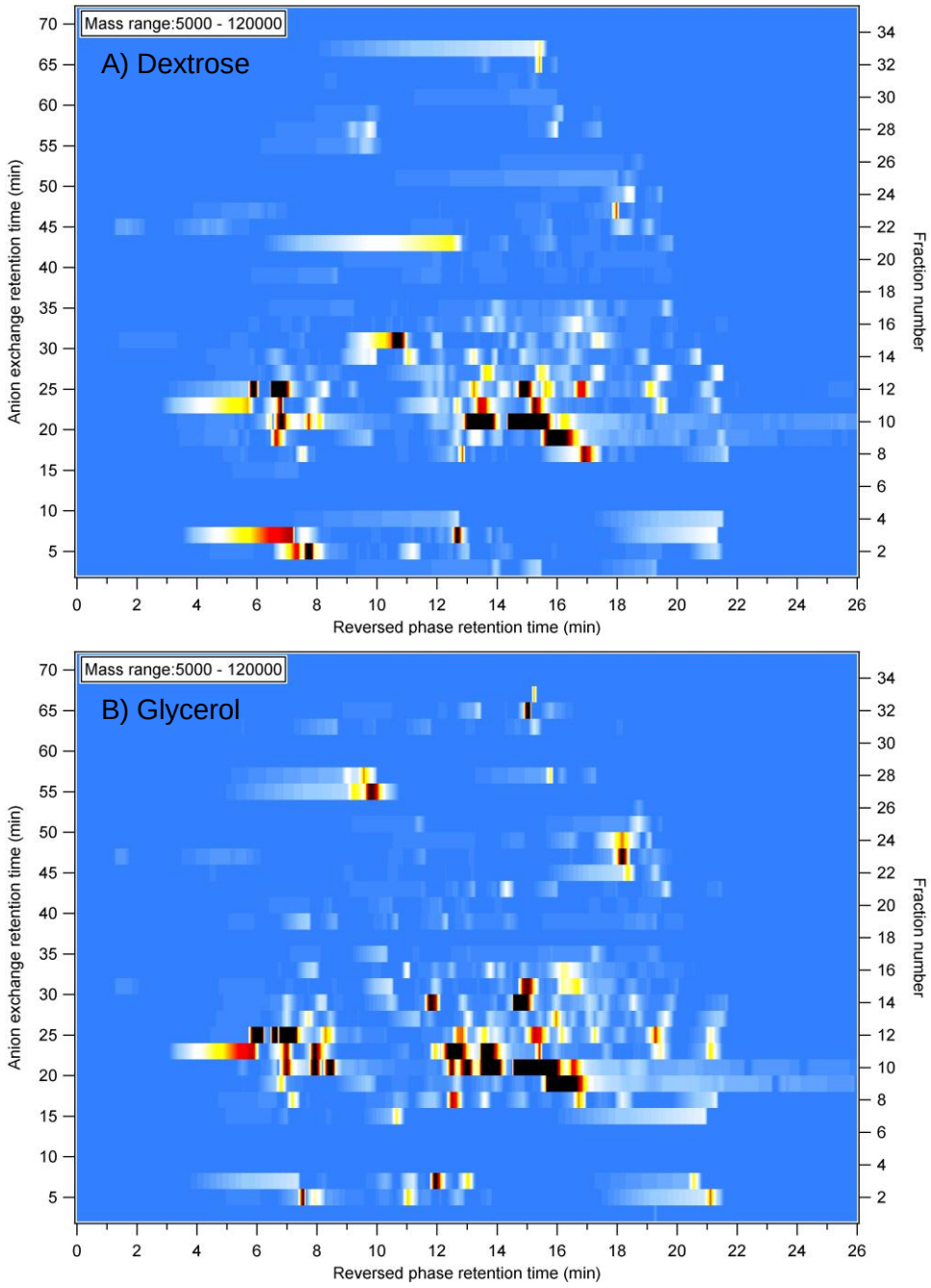


Figure 4-5: Intact protein 2D chromatograms showing AutoME deconvoluted data. Protein intensities are plotted in false color. A) 2.22 mg injection of dextrose-grown yeast cell lysate. B) 2.22 mg injection of glycerol-grown yeast cell lysate.

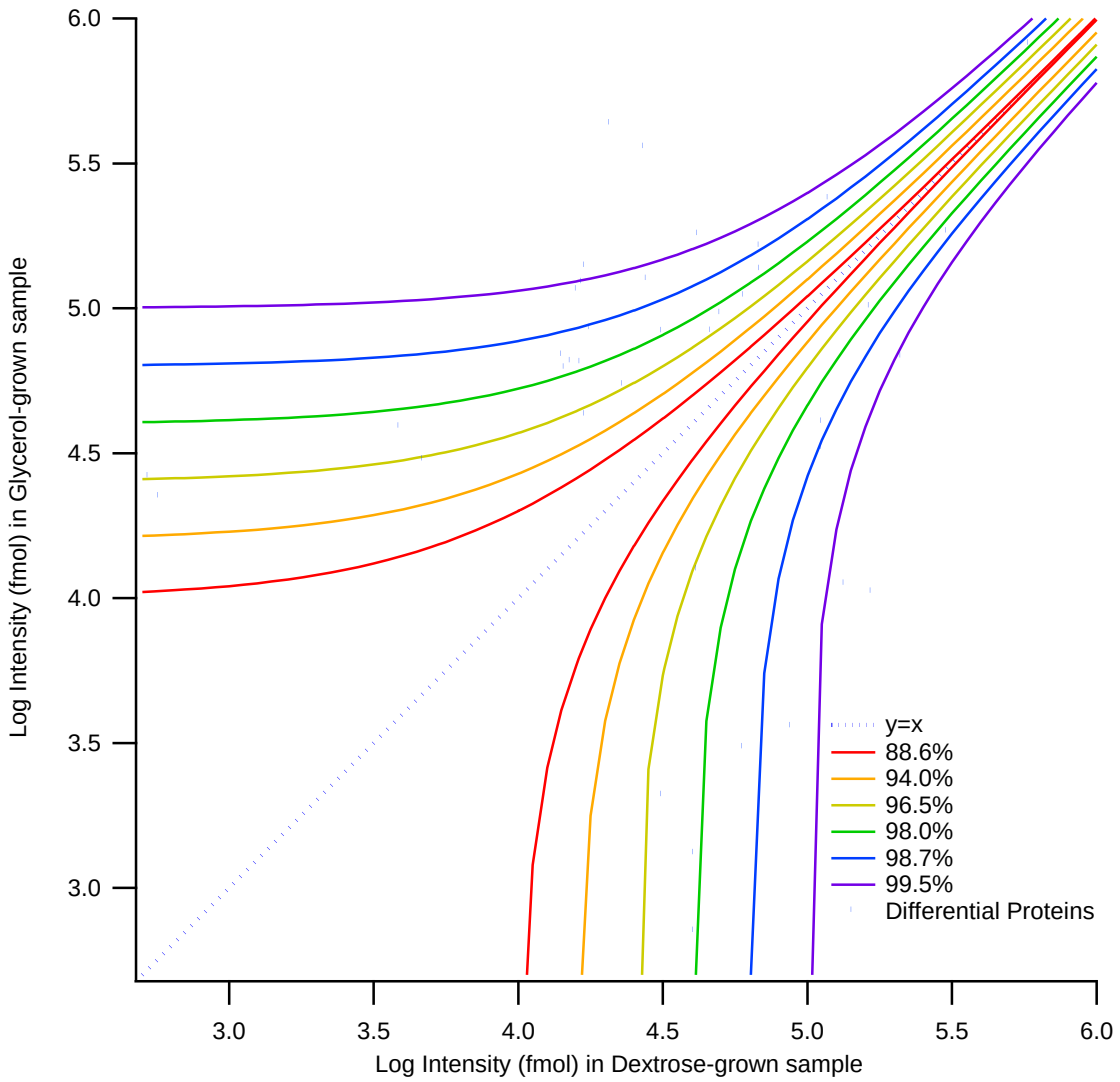


Figure 4-6: Log/Log intensity scatter plot of deconvoluted intact protein intensity from the TD analysis of anion exchange fractions. Confidence curves constructed in Chapter 2 based on replicate data from the online intact protein separation are included.

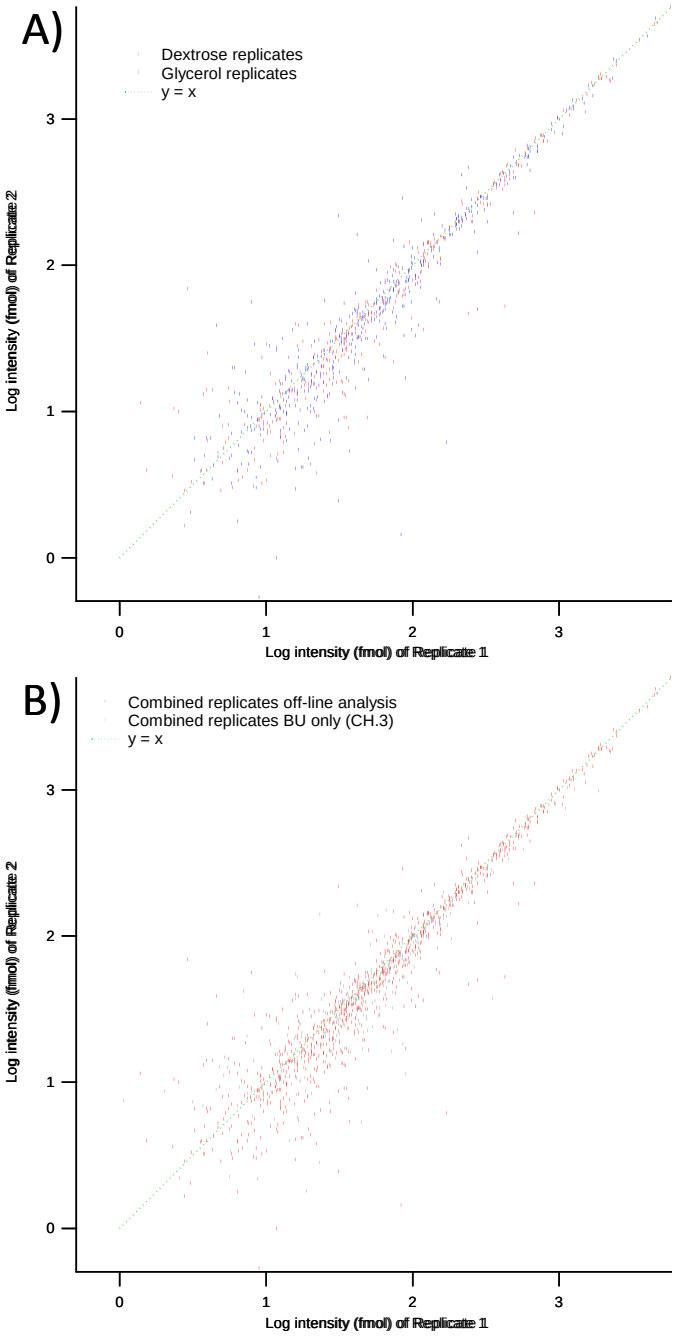


Figure 4-7: Log/Log plot of absolute quantitation of replicating proteins from PLGS2.4 processing of BUD data in dextrose-grown and glycerol-grown yeast cell lysates. A) Intensity comparison of proteins found in both replicates of either sample. B) Comparison of the replicate data from both samples from the BU half of the analysis in this chapter, plotted in red, versus the replicate analyses performed in Chapter 3, plotted in gray.

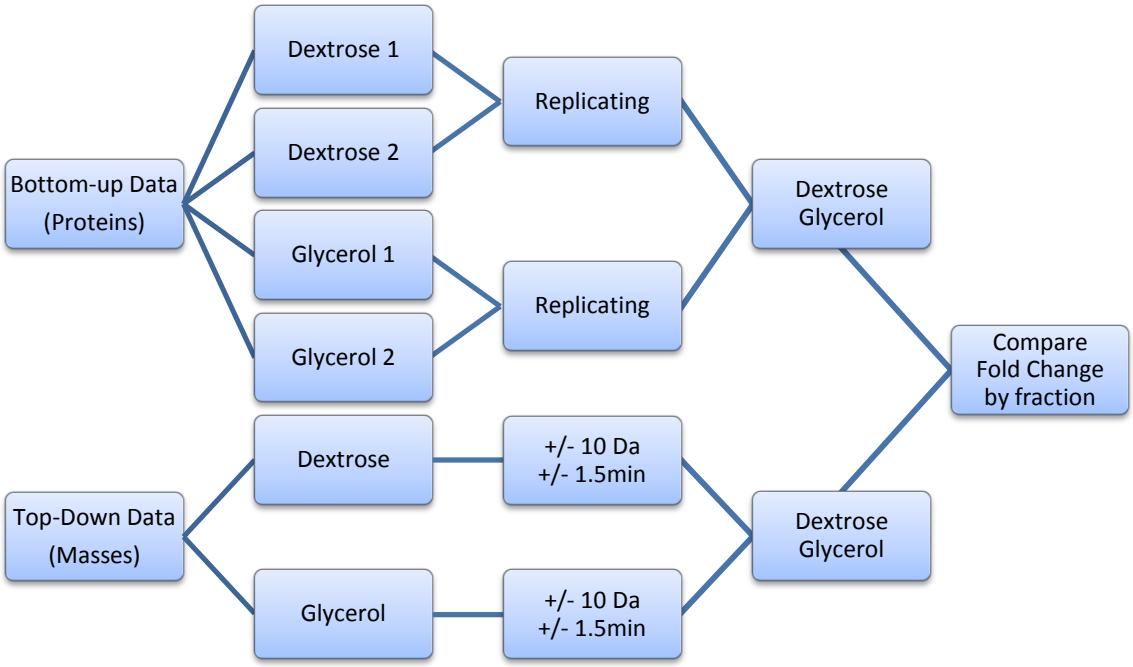


Figure 4-8: Data processing strategy to compare differential expression of proteins from BU and TD data on a fraction-by-fraction basis.

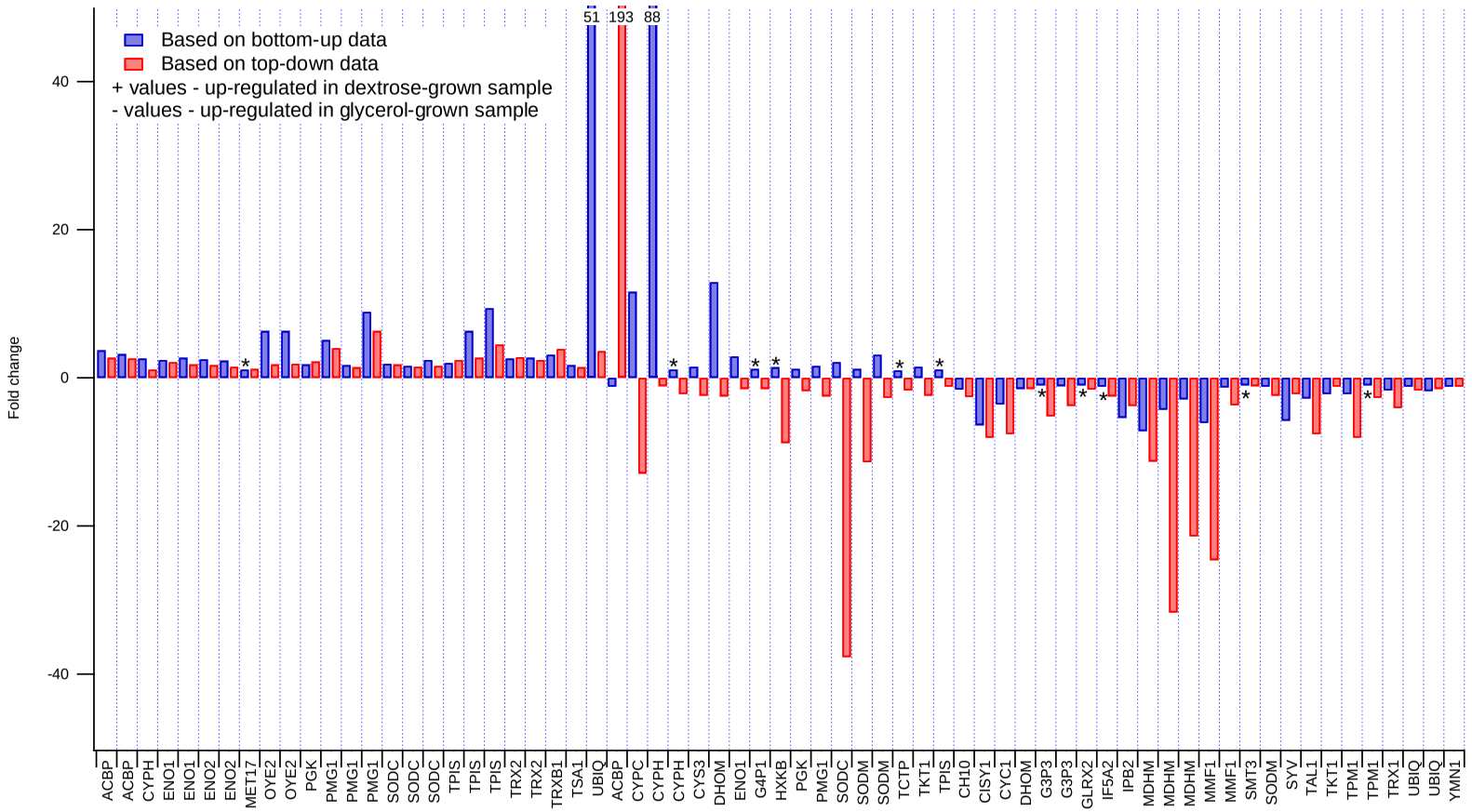


Figure 4-9: Fraction-by-fraction comparison of fold changes determined from either the bottom-up (red) or top-down (blue) data. Positive values indicate the protein was more abundant in the dextrose-grown sample and negative values indicate higher abundance in the glycerol-grown sample. The actual values for the three cases that were off scale are included in the plot at the maximum intensity plotted.

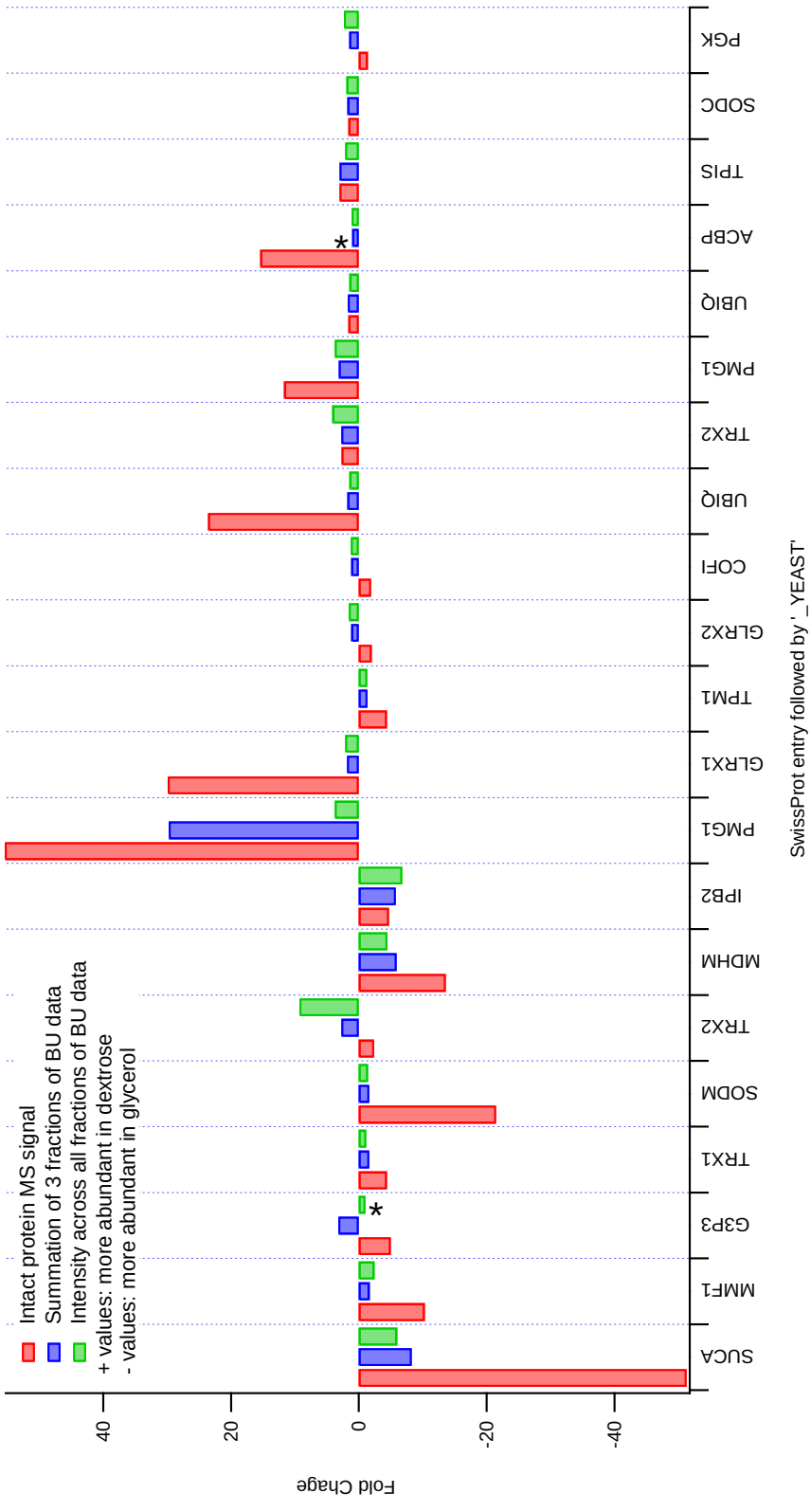


Figure 4-10: Comparison of fold-changes of proteins that were considered significantly different from the mass slice comparison of the TD data. Red bars indicate the fold change determined from the AutoME deconvolution of the intact protein mass spectra across three fractions. Blue bars indicate the summation of protein abundance across the same three fractions in the BU analyses. Green bars indicated the fold change across all 35 fractions of the BU data. Only proteins that appeared to have significant changes in intensity from the mass slice chromatograms are included.

CHAPTER 5: Bottom-up analysis of anion-exchange fractionated S. cerevisiae cell lysates grown under varying growth conditions

5.1 Introduction

The overall trend for differential proteomics includes fractionation of the complex mixture prior to digestion for the purpose of simplifying the resultant peptide mixtures.[1-3] The previous experiments described in Chapters 2 and 4 were aimed at combining this approach with mass spectrometric analysis of the intact proteins to get molecular weight information. However, this is not standard, but rather an added dimension to the experimental method. The work presented in this chapter will follow what is commonly reported in literature in which an MS analysis of the intact proteins is not performed, but rather a separation of the proteins followed by digestion and MS analysis of the resulting peptides. Proteins are fractionated to a greater extent than described in earlier chapters. The purpose of this was to evaluate the increase in proteins identified upon further fractionation. Also, because the data is fully reliant on the identification and quantification of the proteins based on peptide intensity following digestion, it would be advantageous to further decrease the sample complexity prior to digestion.

5.2 Experimental

5.2.1 Overview of experimental method

The overall workflow of this technique is shown in Figure 5-1. The first dimension of separation for this off-line bottom-up analysis was the same as was used in Chapter 2.

Intact proteins were separated initially by charge on a meter-long anion exchange column and detected by UV absorbance. The effluent of the anion exchange column was directed towards a fraction collector. All fractions collected were lyophilized and reconstituted prior to a trypsin digestion. The protein digests were spiked with a standard tryptic digest of BSA and analyzed by UPLC-MS^E. Quantitation and identification were performed simultaneously using ProteinLynx Global Server 2.4 RC7. Differential proteins were assigned based on the changes in absolute protein abundance as determined from ProteinLynx Global Server processing of the digested fractions.

5.2.2 Samples and reagents used

The majority of the reagents used were identical to those described previously in section 2.2.2. The cytosolic fractions from yeast cell lysates that were analyzed in the previous chapters were the same samples used for this analysis. These included yeast grown on glycerol, harvested at the log phase; yeast grown on dextrose, harvested at the log phase; and yeast grown on dextrose, harvested at the stationary phase. A detailed description of the sample preparation can be found in section 2.2.3. A tryptic digest standard of bovine serum albumin (BSA) was used as the standard for quantitative data (MassPREP BSA from Waters, Milford, MA).

5.2.3 Anion-exchange separation of intact proteins

Yeast cell lysates were diluted from the stock solutions to 10 mg/mL total protein concentration as determined by the Bradford protein assay with BSA as the standard protein. Injections of 225 µL (2.25 mg total protein) of each sample were performed. A Waters 600 pump (Milford, MA) provided the gradient which consisted of 10mM ammonium acetate in mobile phase A and 750 mM ammonium acetate in mobile phase B, both adjusted to pH 9.0

with ammonium hydroxide. Because this separation was not performed on-line and is therefore not influenced by the time required for a second dimension separation, the gradient was condensed down to six hours from ten hours, and the flow rate increased to provide an equivalent gradient volume. The exact gradient conditions can be found in Table 5-1. Detection was performed by a Waters 2487 dual wavelength detector (Milford, MA) set at 280nm.

5.2.4 Fractionation and trypsin digestion

Fractions were collected by a Waters Fraction Collector II every 4.75 min starting 40 minutes after the injection through to 420 minutes for a total of 80 fractions. After lyophilization to dryness by vacuum centrifugation, proteins were reconstituted with 50 µL of 50 mM ammonium bicarbonate and only half of each fraction was used for digestion. The digestion protocol was similar to that described in section 2.2.7 except for the amount of trypsin added and the quenching step following overnight digestion. The amount of trypsin that was added gave a 28:1 protein: enzyme ratio. This ratio is based on the assumption that the 2.25 mg of protein injected would be spread equally over all 80 fractions. The amount of protein actually present in each fraction varied considerably, which is why such a high ratio of trypsin was used. It was presumed that it would be better to have the autolysis of trypsin rather than undigested protein. Trifluoroacetic acid (TFA) was added to bring the acid concentration up to 1% in each fraction. Fractions were incubated for 2 hours at 60°C to allow for proper hydrolysis of the RapiGest surfactant. After digestion, fractions were centrifuged at 14,000 x g for 20 minutes to completely remove precipitate from the hydrolysis of the surfactant. Seventy-five µL of each fraction were transferred to sample vials for analysis. A 1 nmol vial of MassPREP BSA standard digest was reconstituted with 1

mL of 30% acetonitrile in water with 0.1% formic acid. A small amount of the BSA digest standard (0.77 μL) was added to each fraction such that the final concentration of BSA was 10 fmol/µL. Fractions were frozen at -20°C prior to analysis.

5.2.5 UPLC-MS^E of protein digests

Analysis of the digested proteins was performed using a Waters nanoAcquity UPLC system coupled to a QTOF Premier mass spectrometer, also from Waters (Milford, MA). The LC conditions as well as the MS^E parameters were identical to those described in detail in section 3.2.4. Briefly, the LC system was operated in trapping mode with a 20 mm x 180 μm ID trap column packed with 5 μm C_{18} Symmetry particles from waters. The analytical column was a Waters nanoAcquity BEH-C_{18} capillary column with dimensions of 250 mm x 75 μm ID packed with 1.7 μm particles.

5.2.6 Protein identification and quantitation based on MS/MS data

The raw data from the peptide separations was processed by PLGS2.4. The full list of processing parameters is shown in Table 5-3. The main difference between the parameters used for processing in Chapter 3 is the amount of BSA standard digest injected on-column. In Chapter 3, BSA digest was added to a final concentration of 50 fmol/µL and 100 fmol of BSA digest was injected on-column by performing a 2 µL injection. For this experiment, 4 µL sample injections were performed after spiking in BSA digest standard to a final concentration of 10 fmol/µL. This provided 40 fmol injection of the standard BSA digest.

Unique protein lists for each sample were generating by removing proteins that were identified across several fractions. The intensities from each fraction were summed in order to allow for quantitation across the entire 80 fractions. The number of fractions that each protein was identified in was also determined. After preparing unique proteins lists for each

sample, two comparisons were made: one between proteins identified in the glycerol-grown log phase sample and those identified in the dextrose-grown log phase sample, and one between the proteins identified in the dextrose-grown log phase sample and those identified in the dextrose-grown stationary phase sample. A fold change was then calculated from the summed intensities in each sample.

5.3 Results

5.3.1 Fractionation of intact proteins by anion exchange chromatography

The soluble fraction of cell lysates from Baker's yeast cells were analyzed directly by anion exchange chromatography. Effluent from the column was directed through a UV detector set at 193 nm for monitoring purposes prior to fraction collection. The resulting UV chromatograms for each differential sample are shown in Figure 5-2. The samples were analyzed on consecutive days in the following order: (1) glycerol-grown/log phase, (2) dextrose-grown/log phase, (3) dextrose-grown/stationary phase. Blank gradients were run in between each fraction in order to properly clean off the column prior to the next sample injection.

5.3.2 Protein identifications based on MS/MS data

Raw data acquired by UPLC-MS^E on the nanoAcquity-QTOF premier instrument were processed using PLGS2.4. The protein list output from the processing of each fraction was combined in order to remove duplicates between fractions from the same sample prior to making a comparison between samples to identify differential proteins. Overall, there were 4,191 yeast protein identifications in the 80 glycerol/log phase fractions, 4,646 identifications in the dextrose/log phase fractions, and 3,519 identifications in the dextrose/stationary phase fractions. Removing the duplicate proteins across fractions from the same sample gives 701

unique proteins in the glycerol/log phase sample, 604 proteins in the dextrose/log phase sample, and 504 proteins in the dextrose/stationary phase sample. Proteins were generally identified with multiple peptide hits. The average peptide hits used to identify a protein in the glycerol/log phase, dextrose/log phase, and dextrose stationary phase samples were 15.9, 17.5 and 14.4, respectively. These parameters, along with more detailed analysis of the protein identifications can be found in Table 5-4. The total mass of protein accounted for is also included for each set of fractions from each sample. This value was calculated by summing the absolute quantitation of each protein in every fraction while accounting for the percentage of the total volume of each fraction that was injected. Four µL out of the 80 µL remaining at the conclusion of the digestion procedure were injected for LC-MS^E analysis, thus the total quantity of proteins in each fraction from PLGS2.4 processing was multiplied by 20 to account for all of the protein in a given fraction. The total mass of intact protein injected onto the anion exchange column for each sample was the same based on the protein concentration determined from the Bradford assay. One would expect that the mass accounted for would also be equal. Another way to look at the consistency of mass injected across samples is to look at the slope of the best fit line for the log/log scatter plot. The scatter plots are included in Figure 5-3 and Figure 5-4. Ideally, the value for the y-intercept, b, would be zero, and the slope of the line, m, would be 1.

5.3.3 Differential protein identifications based on absolute quantitation from MS/MS data

After the protein identifications were summed to removed duplicates, the resulting protein list for the glycerol/log phase sample was compared the list for the dextrose/log phase sample. From the 481 unique proteins identified in the glycerol sample and the 390 proteins

identified in the dextrose sample, 385 proteins were found in both of the samples. The overlap between the two samples is shown in Figure 5-3 along with a log/log plot of the intensities of the proteins that were identified in both samples. Figure 5-4 shows the comparison between the two dextrose-grown samples harvested at different stages of growth. The line y=x and the replicate data from the experiment discussed in Chapter 4 are plotted on the log/log plots as well for reference.

In order to determine whether fold changes were consistent, the replicate analysis from Chapter 4 was used to infer the relative error in the quantitation measurements performed in this chapter. In Figure 5-5, the data points in red are based on this analysis. The error associated with the absolute quantitation of all proteins that were identified in both replicates was calculated for abundances across all thirty-five fractions. The average relative difference in the concentration determined through analysis by PLGS was plotted against the greatest number of peptides that was used to identify the protein between the two replicate analyses. This was calculated as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left|A_1 - A_2\right|}{\left(\frac{A_1 + A_2}{2}\right)}$$

Where A_1 is the abundance of a protein in one replicate and A_2 is the abundance of the same protein in another replicate. Relative differences calculated in this manner were then averaged for all proteins identified with the same number of peptides. This average is plotted in Figure 5-5.

As expected, there is an inverse relationship between error in quantitation and the number of peptides used for identification. Initially, the data were divided up into groups of 10 peptides/protein (1-10, 11-20, 21-30, etc.). A threshold was set at the median of the

average relative errors for proteins identified with a number of peptides within that range. This threshold was used to infer the relative error in the quantification measurements performed in this chapter. For example, in the replicate data plotted over the range of 11 to 20 peptides/proteins, the median relative error was 12%. Therefore, all proteins identified with 11 to 20 peptides in the fractionated bottom-up analysis presented in this chapter were assumed to have a relative error in the absolute quantitation of 12%. There was a large span of average relative error in the range from 1 to 10 peptides, so it was split into 5 peptide spans.

5.4 Discussion

5.4.1 Intact Protein fractionation

As stated previously, the anion exchange fractionation of the proteins is shown in Figure 5-2. Upon inspection of the UV chromatograms, it appears that there was a significant increase in the background signal with each sample injection. Considering the UV detector was auto-zeroed prior to each injection, it would have been expected for the traces to align somewhat better overall. The injection amounts for this analysis were based on a Bradford assay, so roughly the same mass of total protein was injected each time. One possibility for the increased background would be protein ghosting on the column. Because the UV detector was auto-zeroed prior to each injection, this forced an initially equivalent background betweens runs, but if ghosting occurred, it would cause an increase in overall signal as the gradient progressed and eluted carry-over proteins. If this were the case, one would expect that the amount of protein in each fraction of the dextrose/stationary phase sample would be much greater than that found in the first sample analyzed, which was the glycerol/log phase sample. When looking at the identification statistics, the number of

proteins identified with each successive injection does not exhibit an upward trend. However, what is more indicative of potential ghosting effects is the increase in the overall protein abundance detected. Looking only at the proteins that were identified above 95% confidence, the total amount injected across all 80 fractions for the glycerol/log phase, dextrose/log phase, and dextrose/stationary phase samples gives 203, 244, and 302 μg, respectively. From the Bradford assay and the dilution from digestion, this should have been closer to 1.125 mg for all three samples. The fact that the amount of protein accounted for in the analysis is only ~20% of the total is not a cause for concern given that less that 10% of the yeast proteins were identified. It is somewhat concerning, however, that this value increases with each injection.

For the purposes of the differential analysis, ghosting would only cause a problem if there was a significant skew in the log/log plots indicating that a majority of the proteins were found to be up-regulated in the sample fractionated second in the comparisons. For example, in the comparison between growth phases at the time of cell harvest, the ghosting would appear to have affected the comparison if the stationary phase proteins were generally found to be up-regulated in comparison with the log phase proteins since the stationary phase sample was run last. Figure 5-4 shows a line of best fit to the comparison data. As mentioned previously, a slope of 1 would be desirable to ensure that the same amount of protein was analyzed in both cases. For the growth phase comparison, the slope was less than 1 at 0.72. Because the slope is less than one, it indicates that overall, the log phase proteins are slightly more intense than the stationary phase proteins. This is in contrast to what we might expect based on the higher UV absorbance seen in the anion exchange separation of the stationary phase sample and the greater amount of total protein detected as

well. While this slope is still not ideal, for the purposes of the differential comparison, it discredits protein ghosting as the main factor affecting the differences in protein expression.

Also noticeable in Figure 5-2 is the retention time shift that occurred between the runs. Each of the samples generates the same general peak pattern, which makes it easy to pick out these shifts. For example, the peak with a retention time of 200 minutes in the glycerol/log phase sample elutes at 192 minutes in the dextrose/log phase sample, and even earlier still at 185 minutes in the dextrose/stationary phase sample. The drift in retention can be attributed to the change in pH of the mobile phase. The ammonium acetate buffer was selected for the anion exchange mobile phase due to its volatility which allowed for fairly rapid lyophilization and removal of the buffer from the system prior to digestion and LC-MS/MS analysis. The volatility that is a benefit for interfacing with a mass spectrometric analysis is a detriment to the reproducibility of the chromatographic retention times due to the evaporation of ammonia, reducing the mobile phase pH. For this experiment, the identified proteins and their intensities were summed across all 80 fractions, so exact retention time reproducibility was not as important as it might be when attempting to quantify based on intact protein MS intensity. If the comparisons had been made on a fraction-by-fraction basis as was performed in Chapter 4, this would have been prohibitive.

5.4.2 Protein Identification Statistics

Parameters describing the identification of proteins based on PLGS2.4 searching of the MS^E data are included in Table 5-4. The 'Total protein hits' in the first row corresponds to the number of total protein identifications from all eighty fractions of a sample without the removal of duplicates. After the removal of proteins identified in multiple fractions within a sample and summing the intensities of each protein, the number of unique proteins identified

at a confidence of 95% or greater resulted in 481 proteins in the glycerol-grown yeast sample, 390 proteins in the log phase dextrose-grown sample, and 385 proteins in the stationary phase dextrose-grown sample. Corresponding values for the completely bottom-up analysis without fractionation discussed in Chapter 3 were 302, 221, and 166, respectively. Through the fractionation of the intact proteins from the cell lysate prior to digestion, between 1.6 and 2.3 times the number of proteins were identified at 95% confidence. The average number of peptides used to identify each protein was slightly lower, approximately 10 peptides per protein across all three samples as compared to 15 peptides per protein in the un-fractionated analysis. A reduction in this value is not surprising given that the additional proteins identified in the fractionated approach are likely some of the less abundant proteins that were overshadowed by peptides from more abundant proteins in Chapter 3.

A statistic that was initially surprising was the average number of fractions in which a protein was identified in. This value ranged from just under 7 for the glycerol-grown sample up to almost 10 for the dextrose-grown log phase sample. The number of fractions containing a specific protein should follow the chromatographic efficiency obtained in the anion exchange separation. From the UV chromatograms, it appeared that proteins would have only been spread across 3 or 4 fractions at most. Upon closer inspection of the data and calculation of the median number of fractions in which a protein was identified, which was 2 or 3, it was determined that a small number of proteins that were identified in more that 60 of the 80 fractions were skewing the average. The majority of the proteins identified in almost all fractions were heat shock proteins. A likely cause for the widespread identification is the high sequence homology present in the heat shock protein (HSP) family. For example,

proteins that are members of the HSP70 family are shown to have 50-96% nucleotide identity in the genes and eight of the HSPs identified in more than 30 fractions are part of this family.[4] Based on the peptides identified, there may have been an equal probability that multiple heat shock proteins were in a given sample. PLGS2.4 outputs all of these proteins, resulting in the assignment of a single peptide to multiple proteins.

5.4.3 Analysis of differentially regulated proteins and significance of intensity differences

The individual protein lists from the processing of MS^E data from the analysis of the anion exchange fractions of a given sample were combined to give an overall list of proteins identified in the entire sample. The intensities of proteins identified in multiple fractions were summed such that overall intensities could be compared between samples. The lists of proteins identified in the glycerol-grown and dextrose-grown samples were compared to generate a list of proteins identified in both. The overlap of proteins identified at 95% confidence in both samples is shown in Figure 5-3a with a log/log plot of the abundance in fmol of a protein in one sample versus the abundance in the other. Figure 5-4 includes the same analysis performed for the comparison in growth phase using the dextrose-grown sample harvested at the log phase and the dextrose-grown sample harvested at the stationary phase. The log/log plots in both of these figures indicate that it would be difficult to determine whether or not the intensity of a protein is significantly different in a given comparison based on these plots alone.

In chapters 3 and 4, the absence of a clear cut-off in the log/log comparison was resolved by performing replicate analysis of the digested samples. However, due to the number of fractions that were collected in this experiment, it was not feasible to run them in

triplicate eliminating the opportunity for a replicate analysis. Instead, the replicate analysis performed in the experiment described in Chapter 4 was used to infer the relative error in the determination of protein abundance in the heavily fractionated experiment described in this chapter. Figure 5-5 shows the analysis of the replicate data and how it was used to determine the error of abundance measurements for the fractionated BU analysis described here. The replicate data from Chapter 4 was used as opposed to that obtained in the total BU analysis from Chapter 3 due to the greater similarity in the experimental methods. The overall abundance of proteins in a given sample determined in Chapter 4 was the result of the summation of the quantitative analysis performed on multiple anion exchange fractions as opposed to the abundance determined from the analysis of the sample as a whole as was described in Chapter 3.

To calculate the average relative error in protein abundance, proteins that replicated in the analysis from the previous chapter were grouped based on the greatest number of peptides that were used for its identification. Relative errors of protein abundance from proteins identified with the same number of peptides were averaged. This average was then plotted versus the number of peptides and is shown in Figure 5-5. A clear trend is observed in that there is an increase in the relative error as the number of peptides used to identify a protein decreases. The application of this analysis to infer the relative error in the quantitation of proteins in the sample after summing 80 fractions is described in section 5.3.3. After arbitrarily segmenting the data into 10-peptide segments, the median average error observed across the segment was used as the relative error in the quantitation of proteins with an equivalent number of peptides used for identification.

Bar graphs of total protein abundance were plotted for individual proteins identified in both the glycerol-grown and dextrose-grown samples harvested at the log phase. Using the rationale outline above for the determination of relative error, error bars were plotted. When there was no overlap of error bars for the intensity of a single protein, the difference was considered significant. Overall, 231 proteins identified in both samples exhibited significantly different expression. Considering that there were 298 proteins identified in both samples, 78% had significant changes in intensity. A similar trend was observed for the comparison between the two samples that differed in growth phase with 190, or 82%, of the 233 proteins identified in both samples having significant changes in abundance. The percentage of proteins identified in both samples that had significant changes in intensity is quite high. For comparison, the percentage of proteins with significant differential expression from the top down analysis in Chapter 2 was approximately 20%. Looking at the scatter in the differential log/log plots in Figure 5-3B and Figure 5-4B, there are a large number of points that differ from the majority of the replicate data points. The increase in fractionation and the resulting increase in sample handling performed in this analysis could have resulted in more variability in the abundance measurements than was predicted based on the replicate data from a less fractionated sample. This would have the potential to cause an increase in the number of proteins considered to have significant differential expression. The complete lists of proteins having significant difference in expression in either comparison are included in the Appendix.

5.4.4 Comparison of results of fractionated bottom-up workflow to un-fractionated bottom-up analysis in Chapter 3

The differential expressions of proteins determined after fractionation of the cell lysates as presented in this chapter were compared to those determined from the BU analysis of the same proteins through the analysis of the samples without any pre-fractionation that was described in Chapter 3. Only proteins that were identified in both differential yeast samples in an analysis were included since relative changes in abundance from the two experimental methods were to be compared. For the comparison between the two samples that differed in preparation by the carbon source on which they were grown, the differential proteins that were identified in the fractionated and un-fractionated analyses are included in Table 5-5 and 5-6.

All of the proteins included in these tables were determined to have significant differences in abundance in the fractionated BU analysis. Of the 231 proteins that were mentioned earlier, 76 were also identified in both samples of the un-fractionated BU analysis. Only 53, however, were determined to also have significant changes in the un-fractionated analysis. Proteins with changes in abundance that were not significantly different in the un-fractionated samples are included in the tables in red text. The majority of the significantly different proteins were up-regulated in the same yeast sample in both experiments. Forty-one of the 53 proteins, or 77%, were consistently up-regulated in the same sample.

With regards to the degree to which proteins were differentially expressed, however, there were some differences. The histogram in Figure 5-6 illustrates the occurrence of the ratios of protein fold changes determined by dividing the larger of the fold changes by the small to give a ratio greater than 1. Only proteins that were determined to have significant

differences in both the fractionated and un-fractionated analyses are included in the plot. The majority of the proteins, 63%, had fold changes that differed by a factor less than two, which shows consistent regulation in terms of both the sample in which the protein was up-regulated and the extent by which it was up-regulated. Proteins that were found to be most abundant in opposing samples between the two methods were more concerning. Four of these proteins in listed Table 5-6 that had differences in abundance that were significant in both methods had fold changes of less than 1.5 in one of the methods. With a fold change so close to 1, variability in abundance could easily alter the determination of the sample in which the protein was most abundant. An increase in the number of replicates analyzed may resolve this by more accurately calculating the error in the quantitation. It could also be the result of using the replicate data from a previous experiment to infer the relative error in of protein abundance in the fractionated BU approach, since replicates of the actual fractions collected were not analyzed.

One change in the data that was expected was the increase in both the number of proteins identified and the number of differentially expressed proteins identified. The anion exchange fractionation allowed for the simplification of the samples prior to digestion, reducing the occurrence of peptides from more abundant proteins drowning out signal of the less abundant proteins. For example, if many of the peptides from a given protein co-eluted with peptides from a more abundant protein in the un-fractionated analysis, these proteins could be split into different anion exchange fractions allowing for the independent MS analysis of the two protein digests. In the un-fractionated BU analysis, there were 302 and 221 unique proteins identified in the glycerol-grown and dextrose-grown samples, respectively. After comparing the two proteins lists, there were 121 proteins identified in

both samples with 80 having significantly different expression. In comparison, there were 481 and 390 identified in the same two samples in the fractionated BU experiment, with an overlap of 298 proteins. A total of 231, or close to three times more proteins than the un-fractionated analysis, were determined to have significant changes in protein abundance.

Although the improvement in differential protein identifications is large, it did come at the expense of a much larger increase in the analysis time. The un-fractionated BU experiment was performed over the course of 48 hours including digestion and triplicate analysis of all three differential yeast samples. For the fractionated experiment, the anion exchange fractionation was conducted over the course of three days, one day for each sample, followed by roughly one month of analyzing the digested fractions by LC-MS^E. The end result was a single analysis of each fraction from all three differential samples. While more differential proteins were identified, the increase did not scale in proportion to the increase in the amount of time required to perform the experiment.

5.5 Summary and Conclusions

The soluble fraction of yeast cell lysates that were grown on either dextrose or glycerol and harvested at either the logarithmic or stationary phase of growth were analyzed in a BU fashion after anion exchange fractionation. Eighty fractions were collected across the entire anion exchange separation of each differential yeast sample. After digestion with trypsin, a standard of a protein digest was spiked in and each fraction was analyzed by LC-MS^E to both identify and quantify the proteins contained in each fraction. Individual protein intensities were summed across all fractions of a given sample to give the total abundance of each protein in the sample. The summed intensities were used for the comparison of protein expression between yeast samples grown on different carbon sources or harvested during

different phases of growth. To determine the significance of the changes in abundance, replicate data from Chapter 4 was used to infer error in the measurement. The determination of error was performed by using the average relative error in the quantitation of proteins that were identified with a similar number of peptides.

The purpose of this experiment was to look at the increase in the number of proteins and differential proteins that could be identified by simplifying the cell lysate prior to digestion. Fifty-one proteins were determined to have significant intensity differences in both the un-fractionated and fractionated BU analyses. The majority of the proteins exhibited fold changes that were within a factor of 2 when compared between the two experiments which demonstrated good correlation between them. Overall, three times as many proteins with significantly different levels of expression between the glycerol-grown and dextrose-grown samples were identified in the fractionated analysis than were identified in the un-fractionated analysis. While this increase is substantial, it does not compare to the increase in time necessary to perform the experiments. To get this 3x increase in the differentially expressed proteins that were identified, it required a greater than 15x increase in the amount of time required to acquire all of the data. Because of this significant increase in time, it was not feasible to run triplicate or even duplicate analysis of all of the samples.

In conclusion, while this analysis did prove that by further fractionating a cell lysate, greater detail can be learned about the composition of the samples, it also demonstrated that there is a point of diminishing returns of the time invested in the analysis. If fewer fractions were collected, there would have still been an increase in the identification of differentially expressed proteins. It is unclear how many additional differential proteins were identified

due to fractionating the sample 80 times that would not have been identified if the sample were only fractionated 40 times.

5.6 References

(1) Kreunin, P.; Urquidi, V.; Lubman, D. M.; Goodison, S. *Proteomics* **2004**, *4*, 2754-2765.

(2) Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(3) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Jr. *Journal of Proteome Research* **2002**, *1*, 239-252.

(4) Lindquist, S. *Annual Reviews of Genetics* **1988**, *22*, 631-677.

5.7 Tables

Time (min)	% Mobile phase B
0	0
12	7
360	67
405	67
410	0

Table 5-1: AXC gradient conditions for the fractionated bottom-up analysis of the differential yeast lysates. Mobile phase A contained 10mM ammonium acetate, pH 9.0. Mobile phase B contained 750mM ammonium acetate, pH 9.0. The flow rate of the separation was 0.39 mL/min.

Time (min)	% Mobile phase B
0	5
60	40
65	85
70	85
73	5

Table 5-2: RPLC gradient condition for the analysis of digested fraction from the anion exchange column. Mobile phase A was water with 0.1% formic acid and mobile phase B was acetonitrile with 0.1% formic acid. The flow rate was 300 nL/min.

Processing Parameters	**Value**
Chromatographic peak width	Automatic
MS TOF resolution	Automatic
Lock Mass for charge 1	-
Lock Mass for charge 2	785.8426 Da/e
Lock mass window	0.25 Da
Low energy threshold	200 counts
Elevated energy threshold	75 counts
Retention time window	Automatic
Elution start time	-
Elution stop time	-
Intensity threshold	1500 counts
Workflow Template	**Value**
Search engine type	PLGS
Databank	Yeast proteins with trypsin, BSA, and 5 human keratin proteins with a 1x randomization
Peptide tolerance	Automatic
Fragment tolerance	Automatic
Minimum ions per peptide	3
Minimum ions per protein	7
Minimum peptides per protein	1
Maximum protein mass	250,000 Da
Primary digest reagent	Trypsin
Secondary digest reagent	None
Missed cleavages	1
Fixed modifications	None
Variable modifications	Acetyl N-term, Carbamidomethyl C, Deamidation N, Oxidation M
False positive rate	4
Calibration protein	P02769 (BSA)
Protein concentration on column	40 fmol

Table 5-3: PLGS 2.4 RC7 processing parameters used for raw data processing and database searching.

	Glycerol/Log	Dextrose/Log	Dextrose/Stat
Total protein hits	4032	4487	3362
ID confidence (2/1/0)	3300/576/156	3802/577/108	3034/303/25
Unique proteins	702	605	504
Unique proteins at 95%	481	391	385
Avg. peptides/protein	10.1	10.8	9.6
Median hits/protein	8	9	8
Avg. fractions/protein	6.9	9.7	7.9
Median fractions/prot.	3	3	2
Total mass acct. for	203 μg	244 μg	302 μg

Table 5-4: Evaluation of protein identifications by PLGS2.4. **Total protein hits** signifies the total number of yeast proteins identified across all 80 fractions in each sample. **ID Confidence** denotes the confidence level as output from PLGS ('2' indicates 95% confidence;'1' indicates 50%; '0' indicates 'not probable). **Unique proteins** are the proteins remaining after all duplicates are removed regardless of confidence level. **Unique proteins at 95%** includes only proteins identified with a confidence level of 95%. **Avg. peptides/protein** and **Median hits/protein** indicate the average or median number of peptides used to identify a protein. **Avg. fractions/protein** and **Median fractions/prot.** specify the average or median number of fractions in which a protein was identified. **Total mass acct. for** is the total amount of digested protein injected onto the column multiplied by 20 to account for the analysis of 4 µL of the 80 µL of sample present at the conclusion of the digestions

Swiss Prot Name	Description	Fractionated		Un-fractionated		
		Pep Hits	FC	Pep Hits	FC	Up-reg in
FHP	Flavohemoprotein	5	16.5	18	3.4	Gly
ADH2	Alcohol dehydrogenase 2	22	11.5	28	41.5	Gly
ALDH4	Potassium activated aldehyde dehydrogenase	39	10.9	49	27.9	Gly
H2B1	Histone H2B 1	3	8.9	6	4.0	Gly
RIR4	60S ribosomal protein L4 A	24	7.7	10	2.5	Gly
GBLP	Guanine nucleotide binding protein beta	5	6.5	12	11.4	Gly
SUCA	Succinyl CoA ligase ADP forming sub alpha	27	4.3	10	7.5	Gly
MDHM	Malate dehydrogenase mitochondrial	29	3.2	20	6.5	Gly
ARO8	Aromatic amino acid aminotransferase 1	18	3.1	10	2.2	Gly
PABP	Polyadenylate binding protein	10	3.0	15	1.7	Gly
BCA2	Branched chain amino acid aminotransferase	7	2.7	10	3.7	Gly
CISY1	Citrate synthase mitochondrial	16	2.7	19	4.8	Gly
SYDC	Aspartyl tRNA synthetase cytoplasmic	22	2.6	13	2.6	Gly
G3P1	Glyceraldehyde 3 phosphate dehydrogenase 1	14	2.1	21	2.0	Dex
SYV	Valyl tRNA synthetase mitochondrial	40	1.9	25	1.7	Gly
CH10	10 kDa heat shock protein mitochondrial	10	1.9	4	1.3	Gly
HS104	Heat shock protein 104	12	1.8	16	1.4	Gly
G3P3	Glyceraldehyde 3 phosphate dehydrogenase 3	30	1.7	31	1.8	Gly
EF2	Elongation factor 2	33	1.6	40	2.4	Gly
HSP12	12 kDa heat shock protein	17	1.5	9	1.3	Gly
DHE4	NADP specific glutamate dehydrogenase 1	23	1.4	15	1.9	Gly
FKBP	FK-506 binding protein 1	14	1.3	11	1.4	Gly
EF1A	Elongation factor 1 alpha	18	1.3	24	2.6	Gly
TAL1	Transaldolase	25	1.3	16	1.7	Gly
MMF1	Protein MMF1 mitochondrial	11	16.1	5	2.0	Gly
DCPS	Scavenger mRNA decapping enzyme DcpS	17	2.5	9	1.8	Dex
TRX1	Thioredoxin 1	11	2.4	7	1.4	Gly
HSP77	Heat shock protein homolog SSE1	29	2.4	26	1.4	Dex
SODM	Superoxide dismutase Mn	4	2.4	6	1.1	Dex
ZEO1	Protein ZEO1	14	2.2	12	1.1	Dex
VATA	V type proton ATPase catalytic subunit A	21	2.1	20	2.7	Gly
UBIQ	Ubiquitin	6	1.7	6	1.0	Gly

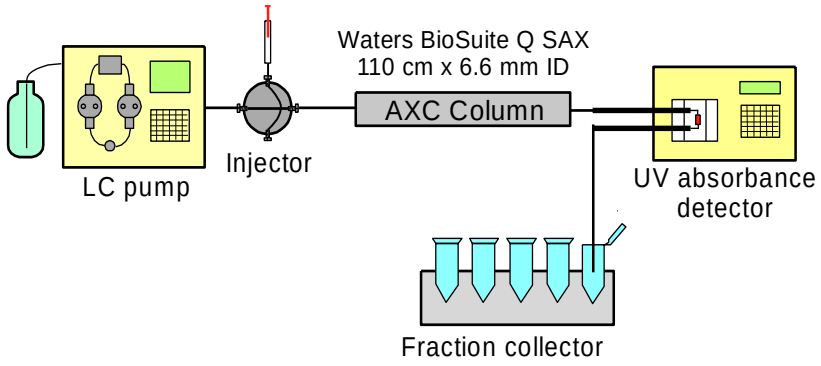
Table 5-5: Proteins up-regulated in the glycerol-grown sample that were also identified in the dextrose-grown and glycerol-grown samples in the completely bottom-up experiment from Chapter 3. The fractionated peptide hits and fold changes (FC) are from the experiment described in this chapter and the un-fractionated columns correspond to data presented in Chapter 3. Proteins written in red text were identify in both BU samples, but were not determine to have significantly different abundance. Shaded proteins were up-regulated in different samples in the two analyses.

Swiss Prot Name	Description	Fractionated		Un-fractionated		
		Pep Hits	FC	Pep Hits	FC	Up-reg in
PDC1	Pyruvate decarboxylase isozyme 1	26	15.2	45	1.8	Dex
IF5A2	Eukaryotic translation initiation factor 5A 2	2	14.8	9	1.8	Gly
ADH1	Alcohol dehydrogenase 1	19	14.7	25	2.8	Dex
ENO2	Enolase 2	35	7.0	50	3.0	Dex
IMDH3	Probable inosine 5 monophosphate dehydrog.	19	6.2	18	2.0	Dex
CPGL	Glutamate carboxypeptidase like protein	6	5.9	17	2.2	Gly
CYS3	Cystathionine gamma lyase	17	4.8	17	2.1	Dex
TRX2	Thioredoxin 2	11	4.0	9	3.9	Dex
RLA4	60S acidic ribosomal protein P2 beta	3	4.0	3	2.1	Gly
HSP72	Heat shock protein SSA2	56	3.7	47	1.8	Dex
ACT	Actin	7	3.4	21	1.6	Gly
ENO1	Enolase 1	58	3.1	34	3.4	Dex
PMG1	Phosphoglycerate mutase 1	25	2.8	17	2.7	Dex
PGK	Phosphoglycerate kinase	37	2.5	40	1.7	Dex
TPIS	Triosephosphate isomerase	20	2.5	19	1.8	Dex
TRXB1	Thioredoxin reductase 1	11	2.4	12	6.9	Dex
HSP82	ATP dependent molecular chaperone HSP82	44	2.1	35	9.3	Dex
CYPH	Peptidyl prolyl cis trans isomerase	12	2.0	8	1.8	Dex
METE	5 methyltetrahydropteroyltriglutamate homoC	47	1.9	24	1.5	Dex
IPYR	Inorganic pyrophosphatase	12	1.9	13	1.2	Gly
HXKB	Hexokinase 2	6	1.8	27	1.8	Gly
PNC1	Nicotinamidase	4	1.7	19	1.8	Dex
DHAS	Aspartate semialdehyde dehydrogenase	13	1.7	5	1.4	Gly
SODC	Superoxide dismutase Cu Zn	14	1.7	14	2.0	Dex
MET17	Protein MET17	22	1.7	18	3.3	Dex
HSP60	Heat shock protein 60 mitochondrial	33	1.6	23	2.4	Gly
COFI	Cofilin	5	1.5	7	1.3	Gly
HSC82	ATP dependent molecular chaperone HSC82	54	1.2	35	1.4	Gly
IF4A	ATP dependent RNA helicase eIF4A	29	1.2	19	1.6	Gly
ADK	Adenosine kinase	8	21.9	14	1.1	Gly
HSP73	Heat shock protein SSA3	19	3.3	24	2.8	Dex
KPYK1	Pyruvate kinase 1	25	2.6	29	1.1	Dex
UBA1	Ubiquitin activating enzyme E1 1	11	2.3	16	1.2	Dex
ALF	Fructose bisphosphate aldolase	14	2.2	24	1.1	Dex
CBS	Cystathionine beta synthase	30	1.9	15	3.4	Gly
NTF2	Nuclear transport factor 2	6	1.8	7	1.1	Gly
DHOM	Homoserine dehydrogenase	23	1.7	13	1.1	Dex
STI1	Heat shock protein STI1	19	1.7	16	1.2	Dex
HSP76	Heat shock protein SSB2	34	1.7	32	1.2	Gly
G3P2	Glyceraldehyde 3 phosphate dehydrogenase 2	23	1.6	25	1.3	Dex
G6PI	Glucose 6 phosphate isomerase	25	1.3	28	1.0	Dex
KAD1	Adenylate kinase 1	21	1.3	15	1.1	Dex
BMH1	Protein BMH1	22	1.2	15	1.4	Gly

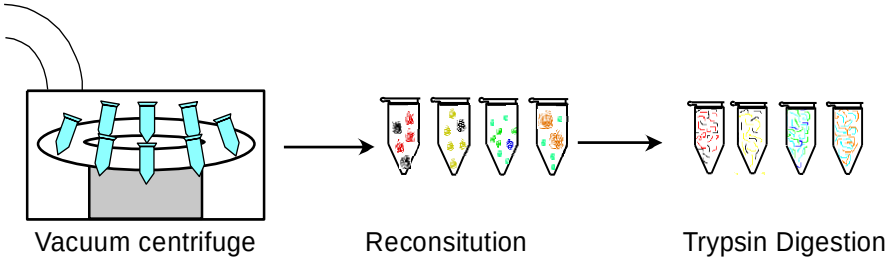
Table 5-6: Equivalent to Table 5-5 except with proteins up-regulated in the dextrose-grown sample.

5.8 Figures

1) Anion Exchange



2) Lyophilization and Digestion



3) UPLC-MS^E of peptides

Waters nanoAcquity LC system
Waters Q-Tof Premier MS

4) Data processing with PLGS

Figure 5-1: Workflow diagram for fractionated bottom-up experiment including column and instrumentation information.

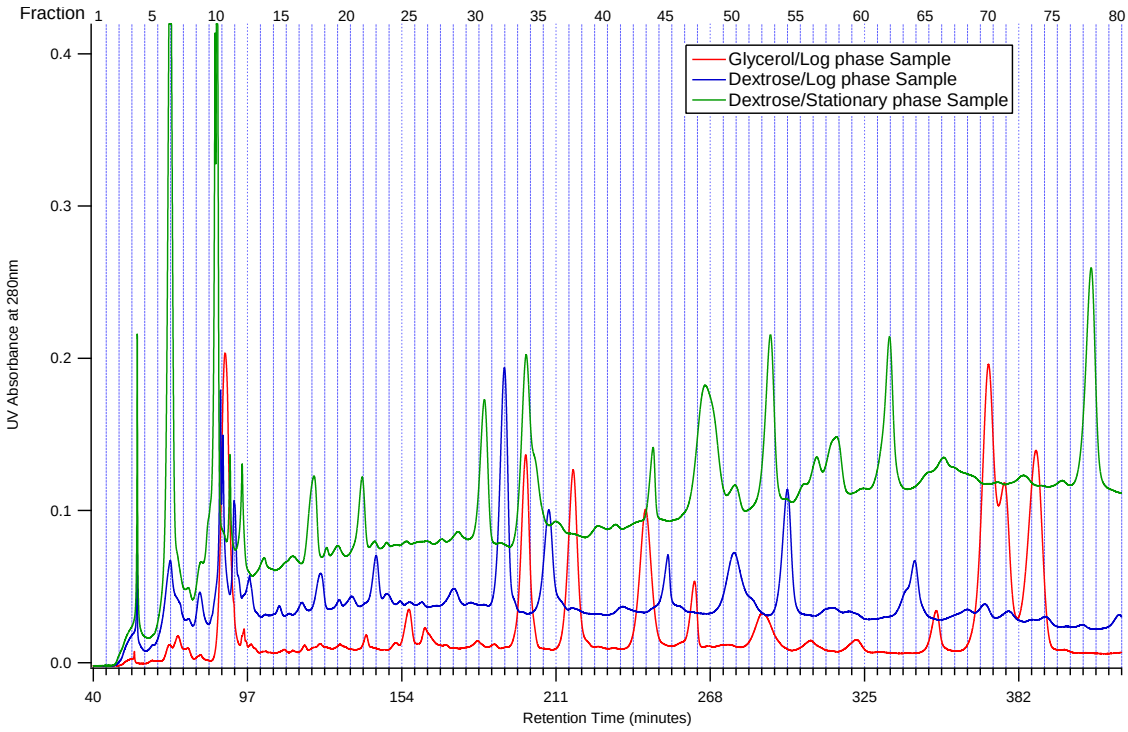


Figure 5-2: Anion fractionation of differential yeast cell lysates monitored by UV absorption at 280nm.

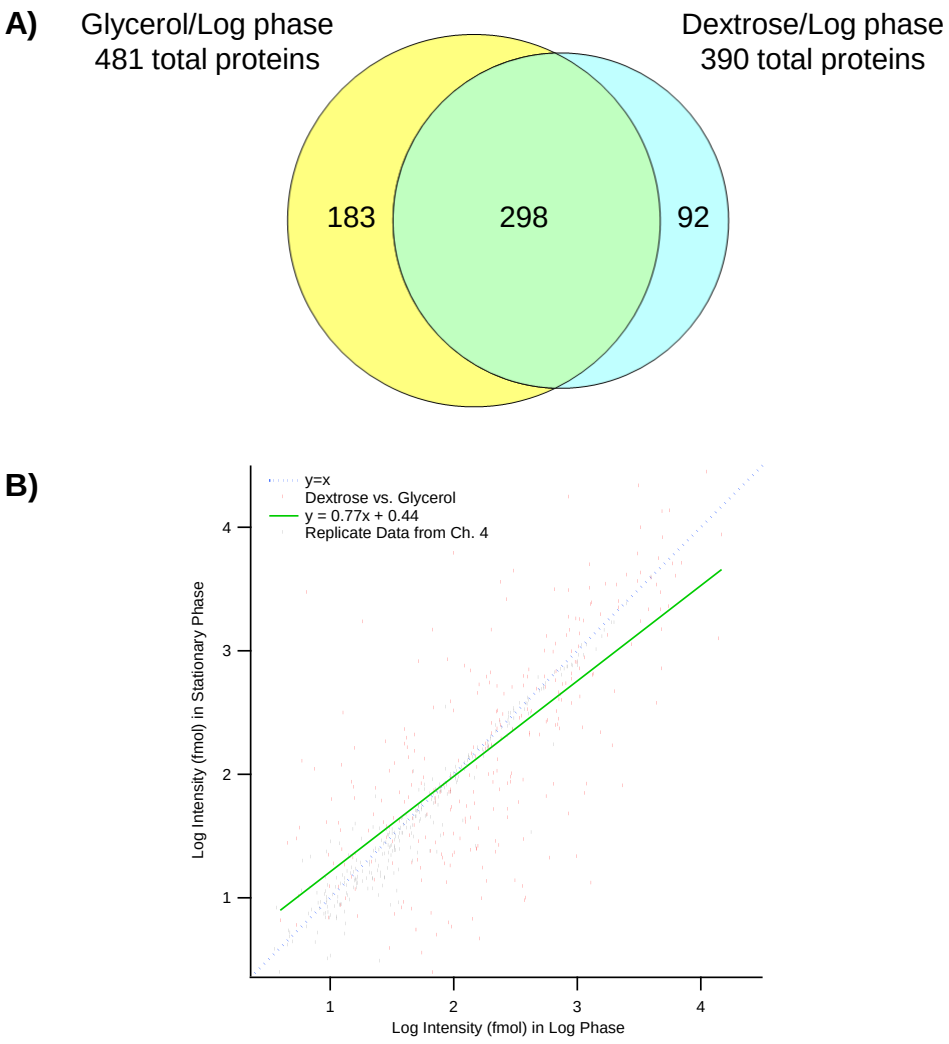


Figure 5-3: Identification of yeast proteins and intensity scatter results for the comparison of cell lines grown on different carbons sources. A) Venn diagram illustrating the overlap between yeast proteins identified in each sample. B) Log/Log intensity plot of the absolute intensity in fmol of the proteins identified in both the glycerol-grown and glucose-grown samples.

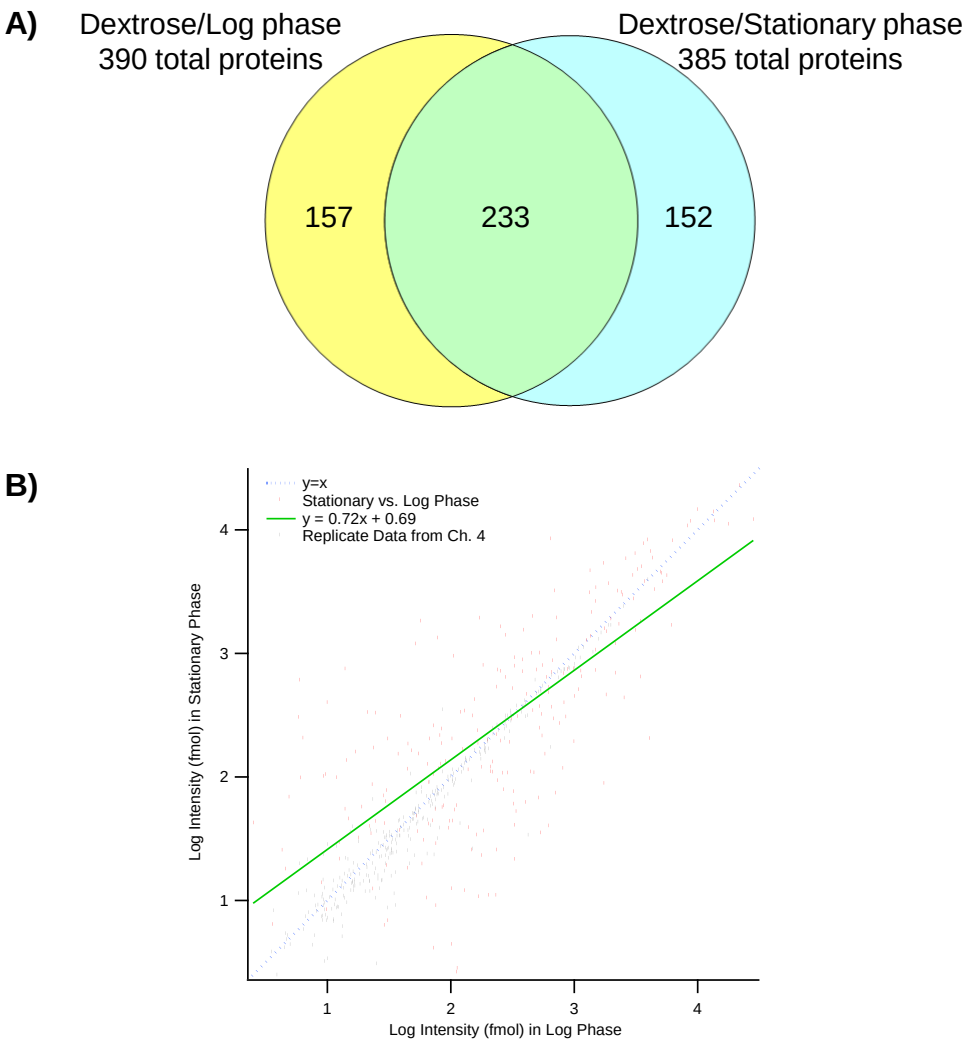


Figure 5-4: Comparison of the yeast proteins identified in the differential yeast samples based on differences in growth cycle at the time of cell harvest. A) Venn diagram indicating the degree of overlap between yeast proteins identified in the log phase sample and those identified in the stationary phase sample. B) Log/Log intensity plot of the absolute intensity in fmol of the proteins identified in both the log phase and stationary phase samples.

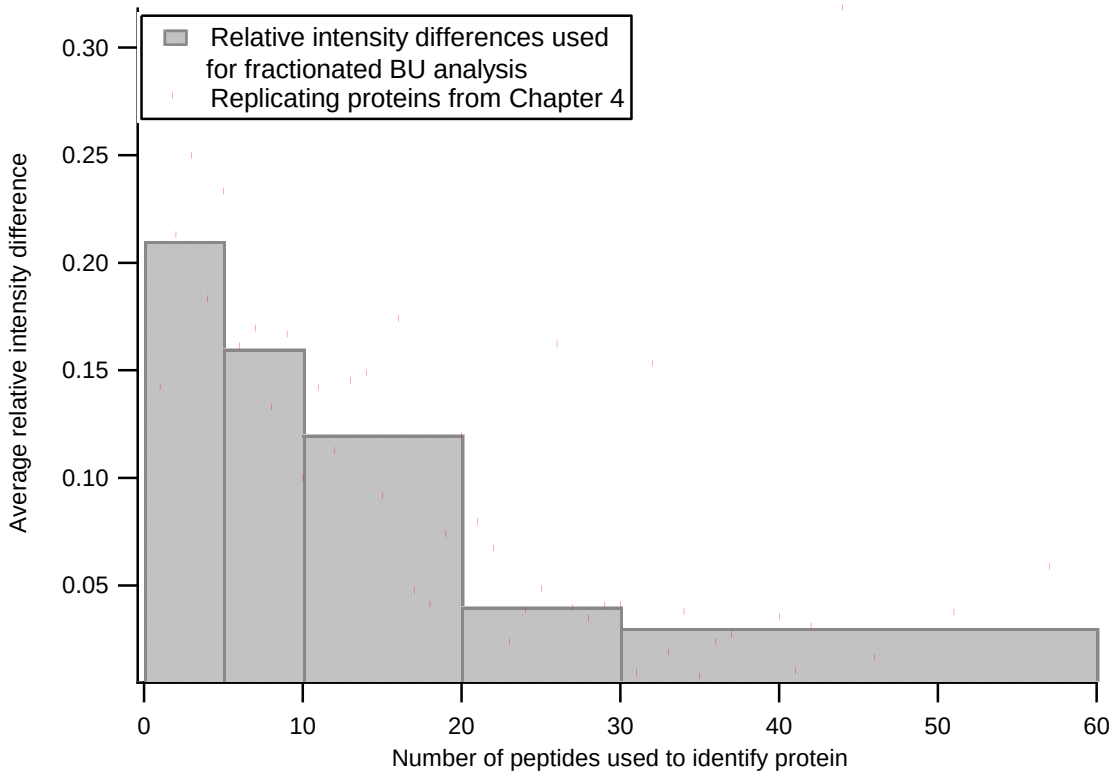


Figure 5-5: Reproducibility of absolute quantitation of replicating proteins from Chapter 4. The average relative difference in intensity measurement between replicates is plotted against the maximum number of peptides used to identify each protein. Gray bars indicate the error in intensity that was used to infer the potential error in the quantitation of proteins determined in this chapter based on the number of peptide hits used to identify a particular protein. This was used to decipher whether or not differences in intensity were significant.

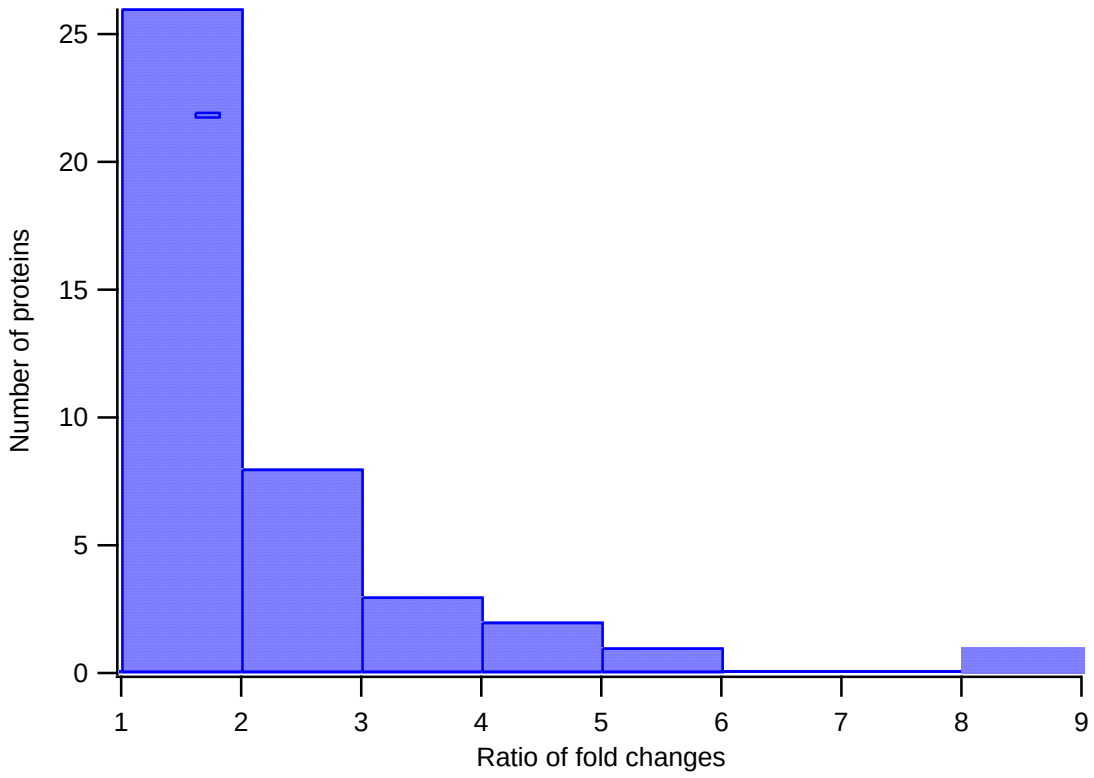


Figure 5-6: Analysis of the difference between fold changes of proteins with significantly different expression in both the fractionated and un-fractionated BU analyses.

CHAPTER 6: Differential proteomic analysis of the soluble fraction of proteins produced by cell lysates of mouse embryonic fibroblast cells: both wild-type vs. β-arrestin 1, 2 double-knockout

6.1 Introduction

Previous chapters have used Baker's yeast samples as a model system for the characterization of the hybrid top-down/bottom-up proteomics methodology due to its well-characterized biology. These samples have proven the validity of this approach as a platform for differential expression proteomics. This concluding chapter aims to apply this analysis to a less well-understood sample set to advance the knowledge of a more complex biological system.

6.1.1 Beta-arrestin signaling

Cellular signaling that involves the sensing of molecules outside of a cell that elicit a change at the cell wall or within the cell itself is of high biological and therapeutic importance. The 7-transmembrane spanning G-protein coupled receptors (GPCR) are the largest family of cell surface receptors and roughly 40% of all drugs in clinical use target GPCRs in some way.[1] The classical pathway of GPCR signaling, illustrated in Figure 6-1 A, involves an activation based on ligand binding which causes the conversion of guanosine-5'-triphosphate (GTP) for guanosine-5'-diphosphate (GDP) on heterotrimeric G-proteins. This activation results in the dissociation of the G-protein into two subunits, which in turn regulate the activity of enzymatic effectors to produce secondary messengers that regulate activity of

intermediary metabolic enzymes. Desensitization of the GCPR occurs through phosphorylation by a G protein receptor kinase (GPK) and the subsequent binding to arrestin to uncouple the G protein from GPCR and sterically inhibit further activation via ligand binding.[2, 3]

Recently, a second type of GPCR-related signaling has been discovered.[2, 4] In this cell signaling pathway, GPCR operates in a G protein-independent manner. β-arrestin signaling was initially found to regulate the endocytosis of activated GPCR, but has now been linked as a recruiter of many enzymes to agonist-occupied GPCRs including the Src family tyrosine kinases, the E3 ubiquitin kinase, diacylglycerol kinase, and a serine/threonine protein phosphatase.[5] An increasing number of signaling pathways are being uncovered as the beta-arrestin signaling pathway after GPCR activation becomes better understood. An example of β-arrestin-dependent signaling as it applies to heart failure is illustrated in Figure 6-1 B.

The samples studied here have the larger goal of understanding the signaling processes involved in heart failure. A heavily studied pathway involving a beta-arrestin dependent signaling in this model is the activation of the extracellular regulated kinase (ERK). Classical activation of this pathway is via agonist stimulation of the epidermal growth factor receptor (EGFR). The beta-arrestin dependent activation of this signaling cascade has been shown to have a cardioprotective effect in mice when chronically stimulated with catecholamine, whereas this effect is not seen with G protein-dependent signaling.[6, 7] Studies have further demonstrated that the classical G protein-dependent signaling is detrimental in heart failure causing hypertrophy, the thickening of the ventricular walls that decrease the capacity of the heart, and bradycardia, a slowing of the heart rate.[8]

One of the pathways known to cause cardioprotective remodeling as a result of β-arrestin-mediated signaling is shown in Figure 6-1 B. Src is recruited by β-arrestin after agonist activation of the GPCR, specifically the β-2-adrenergic receptor. Src-dependent matrix-metalloproteinase (MMP) then mediates the shedding of Heparin-Binding epidermal growth factor (HB-EGF). The resulting cascade promotes cardioprotective promoting mitogenic and anti-apoptotic effects.[7]

In order to add to the information gained through the study of the β-arrestin interactome[9] and the continuing work using genetic mutations[10, 11], both wild-type and β-arrestin 1,2 double knockout mouse embryonic fibroblast cells were studied. An on-line 2D-LC separation of the intact proteins with fraction collection was performed on each sample. 2D chromatograms of the deconvoluted mass spectra were compared to identify protein masses with varying intensities between the two samples. Subsequent digestion and LC-MS^E analysis were performed to identify the differentially expressed proteins.

6.2 Experimental

6.2.1 Outline for experimental method

The experimental workflow for this method is identical to that described in section 2.2.1. Briefly, differential samples were individually analyzed by online 2D-LC-MS. Protein masses were deconvoluted in an automated fashion and the resulting deconvoluted 2D chromatograms were compared on a mass slice basis to select masses with differential expression. Fractions containing those masses were tryptically digested and analyzed by LC-MS^E. Theoretically processed masses of the identified proteins from the database searching of the peptide data were compared to the differentially expressed intact protein masses to identify a differential protein.

6.2.2 Preparation of mouse embryonic fibroblast cell lysates

MEF cell lysates were provided by Kevin Xiao from the Lefkowitz lab at Duke University. β-arrestin 1,2 double knockout mouse embryos were prepared as described previously.[12] Mouse embryonic fibroblast cells (MEF) were prepared and cultured according to the 3T3 protocol of Todar and Green.[13] Cell harvesting and lysis was performed by harvesting the cells in a phosphate buffered saline solution, centrifuging, and removing the supernatant to form a cell paste. Re-suspension buffer was composed of 25 mM ammonium bicarbonate with the following phosphatase inhibitors: 1 mM potassium fluoride, 1 mM sodium pervanadate, 1 μM microcystin, and 10 nM calyculin A. Protease inhibitors from Roche were added as well at a concentration of 1 tablet/10 mL. Cells were re-suspended in 1 mL of the buffer described above and subjected to three freeze/thaw cycles. An additional 9 mL of buffer was added and samples were dounced for 20 strokes. Sonication was performed for a total of 1 minute by sonicating for 10 seconds followed by a cooling time of 20 seconds. Centrifugation at 15,000 x g for 30 minutes was performed in order to remove cell debris and insoluble protein. Filtration was performed with a 0.2 μm filter to further clean up the sample. Cell lysates were frozen at -80° C for storage prior to analysis. Upon thawing the samples immediately before analysis, further centrifugation was performed to remove proteins that did not go back into solution after freezing and thawing.

A Bradford protein assay was performed just as was done in Chapter 2 to determine the total protein concentration in both the wild-type and β-arrestin 1,2 double knockout samples. Bovine serum albumin was used as the calibration protein. Total protein concentrations for the wild-type lysate and double knockout lysate were 2.88 ± 0.22 mg/mL and 2.35 ± 0.08 mg/mL, respectively.

6.2.3 Instrumentation and run conditions at the intact protein level

The instrumentation for the online 2D-LC of intact proteins was similar to that described in Chapter 2. The anion exchange column was a quaternary amine strong anion exchange column with dimensions of 110 cm x 6.6 mm ID packed with 13 μm polymeric particles. Anion exchange mobile phase consisted of ammonium acetate buffer adjusted to pH 9.0 with ammonium hydroxide. The buffer concentration was 25 mM ammonium acetate in mobile phase A and 750 mM in mobile phase B. The gradient conditions for the first dimension separation were the same as used in Chapter 2. New reversed phase columns were used with the same dimensions (4.6 mm x 7.5 cm) and packing material (10 μm polymeric particles bonded with a phenyl stationary phase) as well. The RP gradient was also identical to that used previously ranging from 5 – 40% mobile phase B over 20 minutes with a step up to 20% B at 2 minutes and was operated at a flow rate of 0.5 mL/min. The effluent from the reversed-phase column was split at a 9:1 ratio such that only 55 μL/min was sent to the mass spectrometer for intact protein mass analyses with the remainder sent to a UV detector, a Waters 2487 dual wavelength detector set to 280 nm, followed fraction collection. Fractions were collected every minute from four minutes to 20 minutes for a total of sixteen fractions per RP run. With thirty RP runs performed throughout the anion exchange separation, this resulted in a total number of 480 fractions collected per sample. The initial injection contained 1.3 mg of total protein for both samples.

MS detection was performed on a Waters Q-TOF Premier (Q-TOF P) instrument (Milford, MA) set to acquire MS scans only (MS/MS analysis was not performed on the intact proteins). A standard flow electrospray ionization lockspray source was used to interface the LC to the MS. The reference channel was sampled once every thirty seconds to

perform dynamic calibration and contained a solution of glu-fibrinopeptide. The capillary, extraction cone, and sample cone voltages were set at +3000 V, +4 V, and +35 V, respectively. The desolvation gas flow rate was 350 L/hr at 300°C and the source temperature was set at 100°C to improve desolvation of ions. Continuum data were acquired over the range 400 – 1600 *m/z* over a scan duration of 1 sec with a 0.1 sec interscan delay.

Intact protein mass spectra were deconvoluted using automated maximum entropy processing (AutoME) as described previously. Some parameters were adjusted to account for a more sensitive mass spectrometer. Three separations were performed at the intact protein level; two replicate analyses of a β-arrestin 1, 2 double knockout MEF cell lysate and one analysis of a wild-type MEF cell lysate.

6.2.4 Digestion and LC-MS^E analysis of individual fractions

The digestion was performed in a similar fashion to that described in section 3.2.3. Briefly, fractions were lyophilized to dryness and reconstituted in 50 mM ammonium bicarbonate with 0.1% RapiGest SF. After reduction of disulfide linkages with dithiothreitol and alkylation with iodoacetamide, digestion was initiated by the addition of trypsin and allowed to continue overnight at 37°C. Trypsin was added at a ratio of 25:1 assuming equal distribution of proteins over the 480 fractions collected in an online 2D analysis. This ratio was chosen to reduce the likelihood of having undigested protein with the understanding that this would cause the presence of peptides resulting from the autolysis of trypsin.

LC-MS^E analysis of the digested fractions was performed on a Waters nanoAcquity coupled to a Q-TOF P mass spectrometer, also from Waters. LC run conditions and MS voltages were identical to those described in section 3.2.4. Processing of the raw data and database searching were performed by ProteinLynx Global Server 2.4 RC7 (PLGS2.4). The

processing parameters were similar to those described in Table 3-2 with the exception of the database. The database was composed of the reviewed *Mus musculus* protein entries in the UniProt knowledgebase release 15.7.[14] Porcine trypsin and bovine serum albumin were also added to the database. A 1x randomization of the complete database was appended to the end to set the false discovery rate.

6.3 Results

6.3.1 Differential analysis of intact protein 2D chromatograms

Both the wild-type (WT) and *β*-arrestin 1, 2 double knockout (*β*arr-KO) samples were analyzed by online 2D-LC-MS for intact protein molecular weight. After processing by AutoME deconvolution, 2D chromatograms were constructed from the individual second dimension runs for comparison. While the comparisons were actually made on a mass-slice basis, as reported earlier, the 2D chromatograms for both samples that include the entire mass range of deconvoluted molecular weights is shown in Figure 6-2 for reference. In this plot, the intensity of the protein peaks are plotted in false color, with the maximum of the color scale set at 100,000 counts. The small number of peaks that appear black in color are above this threshold. For the purposes of showing all of the proteins in one plot, the threshold was set below the actual maximum to allow for some of the lower intensity proteins to be seen.

When looking at the mass slice comparisons to identify differential proteins, three different intensity maxima were used to attempt to reduce potential bias that could be caused by intensity values at the limits of the color spectrum. The mass slices of each 1 kDa molecular weight range were plotted at maximum intensity levels of 200,000, 50,000, and 20,000 counts. The number of proteins determined to have changes in intensity between samples based on the changes in color in the mass slice chromatograms is shown in Table

6-1. A total of 65 fractions were selected for digestion in order to identify 68 masses that showed different intensities in the comparison of mass slice chromatograms.

6.3.2 Replicate analysis of *β*arr-KO cell lysates

The intact protein 2D chromatograms for replicate injections of the double knockout cell lysates are shown in Figure 6-3. Overall, the pattern of peaks is very similar between the two analyses. In order to more readily see the differences between intensity values for proteins between the two analyses, a log/log intensity plot was made. This is shown in Figure 6-4a. Also included in the figure is a log/log plot in which the replicate data from this analysis is overlaid atop the replicate data obtained from multiple injections of the glycerol-grown yeast cell lysate as described in Chapter 2. The intensity scatter of this replicate data appears to be slightly lower than what was found in Chapter 2. To quantify this difference, a correlation coefficient was calculated for the replicate analysis of yeast samples acquired on the LCT mass spectrometer and for the replicate analysis of MEF samples acquired on the Q-TOF P instrument. The correlation for the yeast replicate analysis, plotted in gray, was 0.82, whereas for the MEF replicates plotted in red, the correlation coefficient was 0.88, indicating slightly less scatter.

6.3.3 Protein identifications based on MS^E data

The identification of proteins within each fraction containing a differentially expressed protein mass was performed by PLGS2.4. Overall, across all 65 fractions analyzed, a total of 1,153 protein identifications were made. This value includes all proteins found at an identification probability of 50% or greater as determined by PLGS2.4. As this data was used merely as a qualitative analysis of which proteins were present, replication was not necessary for either the 50% or the 95% confidence levels. From previous experiments, such

as those described in Chapter 3, proteins with an identification probability value of '0', or 'not likely', rarely replicated and were therefore removed from further processing. The average and median number of peptides used to identify each protein was 7.2 and 5, respectively. After removing proteins that were identified in multiple fractions, the number of unique protein identifications was 318. These values are listed in Table 6-2. Of the proteins identified in all 65 fractions, 39 had masses similar, meaning it differed by less than 1 kDa, to the 68 differentially expressed masses from the top-down analysis. A complete list of the identified differential proteins found to be up-regulated in each sample can be found in Table 6-3 and Table 6-4 and the overlap of proteins identified and those that were searched for as being differential is shown in Figure 6-5.

6.4 Discussion

6.4.1 Selection of differential proteins

The selection of differentially expressed proteins was initially performed based on changes in color of the intact protein MS data after deconvolution by AutoME. This was performed on a mass-slice basis in order to simplify the 2D chromatograms and facilitate the selection process. As described in section 6.3.1, three different maxima were used for the false color scale. The purpose of this was to allow for changes in the most intense peaks to be seen at that highest intensity maximum, 200,000 counts. By lowering this maximum and allowing high intensity peaks to go off-scale, changes in intensity of proteins of lower abundance could be seen. The total numbers of differentially expressed proteins selected at each intensity threshold level are listed in Table 6-1. For example, in the WT MEF sample, 19 proteins were seen to have changes in color (intensity) when the maximum of the color scale was set to 200,000 counts. By lowering this threshold to 50,000 counts, an additional 16 proteins were found to have changes in color and nine more were found at a maximum of

20,000 counts. Varying the maximum of the color scale allowed for the selection of almost three times more differential proteins than if the plots were set to a maximum corresponding to the most intense protein.

6.4.2 Identification of differential proteins

Overall, the theoretically processed masses of the 39 proteins identified from the bottom-up analyses matched the protein masses that were found to be differentially expressed in the top-down analysis to within 1 kDa. Of those 39 matches, ten included some ambiguity meaning that likely only 34 proteins were identified. This ambiguity can arise from two sources. The first, which accounts for four out the five occurrences, happens when two proteins identified from the bottom-up analysis have similar theoretically processed molecular weights both within 1 kDa of the mass determined experimentally in the top-down analysis. One of the two proteins is likely to be the true protein, but without further analysis, it is not clear which one. The other case that occurred just once was due to the fact that two differentially expressed protein masses from the top-down analysis had the greatest abundance in the same fraction and were within 100 Da of each other. After digesting and analyzing the fraction by LC-MS^E, only one protein was identified with a processed mass within 100 Da of either protein that was searched for. Therefore, only one of the two differential masses was identified, but it remains unclear which one. In Table 6-3, proteins from both cases are listed with superscripts identifying which pairs are ambiguous. Assuming that one of the protein identifications is correct in each ambiguous pair, 34 of the 68 proteins were identified, which gives 50% success rate for the identification of differentially expressed protein masses.

As noted in the description of the AutoME processing parameters, the background ion counts for the intact protein analysis performed on the Q-TOF P instrument was higher than that seen before on the LCT. To account for this, the threshold for the total number of ion counts in a summation of ten 1 sec scans was increased from 20,000 counts to 100,000. This threshold was set by summing the appropriate number of scans during a time in the LC run in which no proteins were eluting. For the purpose of AutoME processing, this worked fairly well, with a relatively small number of 10 sec scans converging on noise alone. However, this did still occur. After looking through the differential protein masses that were not identified from the bottom-up analysis, six of the unidentified masses were cases in which AutoME converged on noise and were not actual protein masses. Removing these masses from the total number of proteins searched for increases the identification percentage from 50% to 55%, which is slightly better than what was achieved in the experiment from Chapter 2.

6.4.3 Determination of the significance of the intensity differences

A duplicate analysis was performed on the *β*arr-KO sample to identify the extent of differences in intensity resulting from the methodology rather than between the differential samples themselves. AutoME-deconvoluted masses that were found in both replicates were plotted on a log scale in Figure 6-4a and overlaid with the replicate yeast lysate data from Chapter 2 Figure 6-4b. As described earlier, the correlation coefficient of the *β*arr-KO analysis was slightly better than that obtained with the yeast data, 0.88 versus 0.82. This could be the result of using the Q-TOF P instead of the LCT, which has two times the resolving power. It could also be an artifact from the lower number of data points in the analysis performed on the Q-TOF P, 250 as opposed to 1,000 for the comparison between

two replicates. Also, due to the increased background in the raw data as well as the deconvolution of the Q-TOF P data, the limit of detection was slightly greater and thus may have led to an apparent decrease in the intensity scatter near the limit of detection. Lastly, since the difference was so small, it could have just been due variability in the 2D separation after more than two years of use.

Nevertheless, due to the difference being so slight, confidence curves from the yeast replicate data were plotted with the *β*arr-KO replicate analysis and the differential proteins from the comparison with a WT cell lysate in Figure 6-6 to determine significance. As described in Chapter 2, these confidence curves in the log scale represent straight lines when plotted in the linear scale. The percentages noted in the legend signify the percentage of points from the yeast replicate analysis that fall between the confidence lines. The corresponding percentages for the replicate *β*arr-KO analysis are included in parentheses for comparison. Because the scatter in the data between the two replicate analyses is relatively close, the percentage value used for the confidence threshold is based on the yeast replicate data due to the greater number of data points. Setting 96.5% as the threshold would make 14 of the differentially expressed proteins significant, whereas lowering this to the 94% cut-off would make 19 of the 34 differentially expressed proteins significant. The 94% threshold was arbitrarily set as the confidence limit for labeling proteins as significantly different due to the slightly greater correlation from the *β*arr-KO replicate data as opposed to the yeast replicate data in which a 96.5% threshold was used. The percent confidence value for the significance of the intensity difference is noted in the last column in Table 6-3 and Table 6-4.

6.4.4 General comparison of results with those obtained in the differential analysis of the Baker's yeast samples

For the intact protein separation, there were fewer protein masses detected in this experiment than were found in the differential analysis of yeast cell lysates. One explanation for this could be in the sample itself. When analyzing the yeast samples, the lysates could be injected directly onto the column without the need for filtering after thawing. However, with the MEF lysates, proteins appeared to be less well-behaved and did not react well to the freezing and thawing performed to preserve the sample during the time between cell harvest and intact protein analysis. The result was a thick, murky, lysate that needed further centrifugation and filtration. This may have resulted in the removal of some proteins that were soluble when the lysate was initially prepared. This loss of protein is definitely a concern and will be remedied by improved coordination with the lab at Duke to analyze the lysates directly after harvest preventing the need for freezing. There were also fewer differential proteins which may have been the result of fewer proteins detected overall or that the changes induced by the removal of the β-arrestins were less that those created by changing carbon source or growth phase as was done in the yeast sample preparation. It is probably a combination of the two.

In terms of protein identification statistics from the LC-MS^E analysis, the average and median numbers of peptides used to identify each protein were 7.2 and 5. These values are slightly lower than those used to identify proteins in the bottom-up only analysis reported in Chapter 3 where, on average, 16.4 peptides were used to identify each protein. However, when looking at the average and median number of proteins used to identify the proteins that were similar in molecular weight to those determined to be differential from the TD data, the

values increase to 14.5 and 12, respectively. These values are more in line with what was seen in Chapter 3. A possible explanation for the lower number of peptides used to identify the proteins overall could be that in Chapter 3, only the most abundant proteins were identified. For example, in the glycerol-grown log phase sample, the 302 most abundant proteins were probably identified. This is in contrast to the data presented in this chapter where 480 fractions were collected from a single analysis. Because only fractions containing proteins that changed in intensity were digested and analyzed further, some of the most abundant proteins were never digested and identified. Also, due to the fractionation and simplification of the protein mixtures, some proteins of lower abundance that would have been overshadowed by more abundant proteins without fractionation were able to be identified.

As stated earlier, there was a substantial increase in the average and median number peptide hits per protein when only looking at those proteins identified as differential proteins. This increase is due in large part to the increase in scatter of the proteins at lower intensities. In order for a protein intensity difference to be significant it was a large change at a low intensity, with smaller and smaller changes being significant as the intensity increased. Looking back at the data, few, if any, of the lower abundance proteins that were identified actually had significant changes in intensity. Because, in general, more peptides are used to identify a more abundant protein, the increase in peptides/protein when limiting only to those used to identify differential proteins is expected.

In comparison to the data-directed acquisition performed for the tandem MS analysis of the peptides in Chapter 2, the peptide hits per protein almost doubled. This demonstrates the clear advantage of the MS^E acquisition over a more traditional DDA MS analysis. As

stated in Chapter 3, the more parallel nature of MS^E in contrast to the serial DDA analysis allows for more peptides to be analyzed by MS/MS while at the chromatographic apex resulting in greater coverage of each protein.

6.4.5 Comparison of differential regulations with literature

In 2007, the Yates and Lefkowitz labs published a paper reporting the β-arrestin interactome, which includes the proteins found to interact with β-arrestin.[9] The experiment was performed using MS-based proteomics approaches including both the multidimensional protein identification technique (MudPIT) and a gel-based LC-MS/MS approach. Over 300 proteins were found to potentially interact with β-arrestin 1 and/or 2. To isolate interacting proteins, β-arrestin complexes were immunoprecipitated from HEK293 cells overexpressing β-arrestin with a C-terminal FLAG epitope. The list of proteins determined to be differentially regulated in this experiment was compared to the list of interacting proteins from that paper. There was no overlap in protein identifications between the two experiments. However, the authors did note the presence of metabolic enzymes in the interactome, which was not initially expected. Most of the metabolic enzymes reported in the interactome were involved in the glycolysis pathway and it was hypothesized that the β-arrestins may scaffold the glycolytic enzymes to facilitate energy production. This may be an explanation for the up-regulation of phosphoglycerate mutase, PGAM1, in the WT sample.

General signaling proteins were also found to interact with the β-arrestins including annexin II (ANXA2). In the combined TD/BU differential proteomics experiment, annexin V (ANXA5) was found to be up-regulated in the WT sample. While they serve different purposes, ANXA2 as a calcium-regulated membrane-binding protein and ANXA5 as an

anticoagulant protein, they are both annexins that interact with the phospholipids of the cell membrane and are in some way involved in extocytosis of which the β-arrestins are also known to be involved.[14] The remaining proteins do not have immediately apparent relationships to β-arrestins and further analysis and experiments must be performed to both improve the understanding of and verify the differential regulation.

6.5 Summary and conclusions

An LC-based multidimensional separation was performed on intact proteins from lysates of both wild-type and β-arrestin 1, 2 double-knockout mouse embryonic fibroblast cells. The methodology used in this experiment was similar to that used in Chapter 2 with the major exception being the separation and analysis of the digested fractions. A data-independent acquisition, MS^E, was used instead of a data-directed tandem MS acquisition which allowed for greater coverage of the peptides present in the sample and a greater number of proteins identified in each fraction. Using this technique and matching the masses of proteins identified from the LC-MS^E analysis of the peptides with the experimental masses of intact proteins obtained from the 2DLC-MS analysis, 34 of the 68 differentially expressed intact protein masses were able to be identified. Using a confidence threshold of 94% to determine the significance of the difference, 19 of the 34 identified were significantly different.

In order to gain a better understanding of the biological significance of the differential proteins, the list was compared to the recently reported β-arrestin interactome. Although 337 proteins were identified as interacting with either β-arrestin 1 or 2, there was no overlap with the differential proteins identified here. There were some similarities in the types of proteins identified in that some enzymes involved in glycolysis were in the interactome and a glycolytic enzyme, PGAM1, was identified as more abundant in the WT sample. However,

it is still unclear how β-arrestin is involved in metabolism. Overall, with the information obtained solely from this experiment, little insight is gained into the β-arrestin-mediated signaling in the heart, which is the end goal. The hope, however, is that this information may be used in the future to supplement and support other experimental data to improve the understanding of this cardioprotective signaling pathway.

6.6 Future studies

Of the methods that have been discussed in this dissertation, currently only the hybrid TD/BU approach with an on-line 2D separation of intact proteins has been performed. From the other approaches that have been presented for the analysis of differential yeast cell lysates, it is clear that to complete analysis of these samples, it would be beneficial to perform a completely BU analysis. One of the limitations posed by the MEF cell lysates was the insolubility of some proteins after even one freeze/thaw cycle. An un-fractionated bottom-up analysis would not be limited by this and could therefore have the potential to analyze a larger set of proteins. Some of the lower abundant proteins may be lost in this analysis, but it may offer a complementary subset of differential proteins that may be of interest. This type of analysis would not be difficult to implement as it is the least time consuming of all that have been discussed here.

Another type of analysis that was not performed on the yeast cell lysates and has not been discussed yet is a western blot. In a western blot analysis, proteins are separated by gel electrophoresis in a native or denaturing environment. In native conditions proteins are separated by their 3D structure, whereas in denaturing conditions, they are separated by the length of the polypeptide. Proteins are transferred to a membrane and detected using antibodies, monoclonal or polyclonal, specific to a single protein. This type of analysis would not be realistic without prior information on the samples. For the MEF cells, this

could be used to confirm the differential expression determined by the hybrid TD/BU approach in this chapter. Some of the proteins have antibodies readily available and would therefore be used first to see if the intensity changes seen in this experiment correlated with those of the antibody-based western blot analysis. With a polyclonal antibody, this would have the same effect as an un-fractionated bottom-up analysis, where different isoforms of a protein would be created equal. Though more expensive and therefore probably not feasible for this analysis, a monoclonal antibody offers the opportunity to probe the different isoforms provided that the modification occurred at the epitope of the protein. If more site-specific information were known about post translational modifications on each protein this would be an expensive, though feasible experiment.

With the depth of data that is acquired at both the intact protein and peptide level, there are innumerable ways to look at the data. One way that was not probed was to look for repeating mass shifts in the intact molecular weight of the protein equal to those expected for various post-translational modifications. For example, looking for occurrences of phosphorylation, offsets could be made every 80 Da. This would be interesting when looking at the 2D chromatograms because a shift in the anion exchange retention time would also be expected due to the addition of a negative charge on each phosphorylation site. The easiest way to view this type of shifting in mass would be to plot mass versus anion exchange or reversed phase retention time. For phosphorylation, for example, a diagonal pattern of peaks should be seen corresponding to the shifts in both mass and anion exchange retention time.

It also may prove to be beneficial to attempt to reduce the sample consumption of the analysis. Currently, the lowest amount of total protein that has been injected onto the anion

exchange column is 1.3 mg, which is a prohibitive amount of protein for many analyses. The driving factor for the need for that much protein initially was due to the limitations imposed by the sensitivity of a DDA acquisition and that the protein injected had the potential to be spread over 480 fractions. It was not likely that a peptide would be fragmented at its chromatographic apex due to the cycle of a single precursor scan followed by multiple fragment ion scans. Therefore, the analysis benefitted from overloading the capillary columns to broaden the peaks in an attempt to ensure that a sufficient amount of peptide would continue to elute from the column during the fragmentation scans. With the advent of MS^E fragmentation, this is not as critical, thus the amount of digested protein needed for proper sequencing is less. By reducing the inner diameter of both dimensions of the intact protein 2D separation and coupling the second dimension to a nanoflow electrospray as opposed to the standard spray the sensitivity of the intact protein MS signal may not be detrimentally affected.

Lastly, with regard to these samples in particular, it would be interesting to look at the differential expression of knock-outs of β-arrestin 1 and 2 separately. Recent literature has reported that these two arrestins may serve divergent or opposing roles.[15, 16] β-arrestins 1 and 2 are the two most abundant non-retinal β-arrestins and have been known to be involved in the activation of extracellular signal-regulated kinase ½ (ERK1/2). These two studies investigated the role of these two β-arrestins on the desensitization and internalization of protease-activated-receptors 1 and 2. Their findings suggest that the arrestins can promote their desensitizing effects differently and can mediate internalization and downstream signaling of a receptor in different ways. Therefore, it may be interesting to look at

individual knockouts of the arrestins in addition to the wild-type and double knockout cell lysates.

6.7 References

(1) Filmore, D. *Modern Drug Discovery* **2004**, *7*, 24-26, 28.

(2) Lefkowitz, R. J.; Shenoy, S. K. *Science* **2005**, *308*, 512-517.

(3) Moore, C. A. C.; Milano, S. K.; Benovic, J. L. *Annual review of physiology* **2007**, *69*, 451-482.

(4) Lefkowitz, R. J.; Rajagopal, K.; Whalen, E. J. *Molecular Cell* **2006**, *24*, 643-652.

(5) Luttrell, L. M.; Gesty-Palmer, D. *Pharmacological Reviews* **2010**, *62*, 305-330.

(6) Noma, T.; Lemaire, A.; Prasad, S. V. N.; Barki-Harrington, L.; Tilley, D. G.; Chen, J.; Le Corvoisier, P.; Violin, J. D.; Wei, H.; Lefkowitz, R. J.; Rockman, H. A. *Journal of Clinical Investigation* **2007**, *117*, 2445-2458.

(7) Patel, P. A.; Tilley, D. G.; Rockman, H. A. **2008**, *72*, 1725-1729.

(8) Zhai, P.; Myamamoto, M.; Galeotti, J.; Liu, J.; Masurekar, M.; Thaisz, J.; Irie, K.; Holle, E.; Yu, X.; Kupershmidt, S.; Roden, D.; Wagner, T.; Yatani, A.; Vatner, D. E.; Vatner, S. F.; Sadoshima, J. *Journal of Clinical Investigation* **2005**, *115*, 3045-3056.

(9) Xiao, K.; McClatchy, D. B.; Shukla, A. K.; Zhao, Y.; Chen, M.; Shenoy, S. K.; Yates, J. R.; Lefkowitz, R. J. *Proceedings of the National Academy of the Sciences of the United States of America* **2007**, *104*, 12011-12016.

(10) Tilley, D. G.; Kim, I.-M.; Patel, P. A.; Violin, J. D.; Rockman, H. A. *Journal of Biological Chemistry* **2009**, *284*, 20375-20386.

(11) Mangmool, S.; Shukla, A. K.; Rockman, H. A. *Journal of Cell Biology* **2010**, *189*, 573-587.

(12) Kohout, T. A.; Lin, F.-T.; Perry, S. J.; Conner, D. A.; Lefkowitz, R. J. *Proceedings of the National Academy of the Sciences of the United States of America* **2001**, *98*, 1601-1606.

(13) Todaro, G. J.; Green, H. *Journal of Cell Biology* **1963**, *17*, 299-313.

(14) Jain, E.; Bairoch, A.; Duvaud, S.; Phan, I.; Redaschi, N.; Suzek, B. E.; Martin, M. J.; McGarvey, P.; Gasteiger, E. *BMC Bioinformatics* **2009**, *10*, 136.

(15) Kuo, F. T.; Lu, T. L.; Fu, H. W. *Cellular Signalling* **2006**, *18*, 1914-1923.

(16) Kumar, P.; Lau, C. S.; Mathur, M.; Wang, P.; DeFea, K. A. *American Journal of Physiology - Cell Physiology* **2007**, *293*, C346-C357.

6.8 Tables

Plotted Max	Up in WT	Up in βarr-KO	Total
200,000	19	4	23
50,000	16	17	35
20,000	9	3	12
All	44	24	68

Table 6-1: Intensity distribution of differentially expressed protein masses based on intact protein MS signal. **Plotted Max**-maximum of the intensity plotted in false color. **Up in WT**-number of proteins more abundant in WT sample. **Up in βarr-KO**- number of proteins more intense in double knockout sample. **Total-** combined differential proteins across both samples at given intensity maxima.

Differentially expressed protein masses	68
Protein hits from LC-MS^E analysis	1153
Unique protein identifications	318
Average peptides/protein	7.2
Median peptides/protein	5
Average peptides/differential protein	14.5
Median peptides/differential protein	12

Table 6-2: Protein identification statistics from the analysis of 65 fractions containing differentially expressed protein masses.

Swiss-Prot Name	Description	Pep. Hits	Exp. Intact Mass	Theo. Intact Mass	Fold change	% Prob. Diff.
AATC	Aspartate aminotransferase cytoplasmic	24	46102.1	46100.39	2.8	94.0
AK1A1	Alcohol dehydrogenase	3	36450.5	36455.73	61.3	96.5
ANXA5	Annexin A5	32	35664.3	35752.44	n/a	99.5
CH10	10 kDa heat shock protein mitochondrial	10	10873.3	10831.51	1.4	98.0
COF1	Cofilin 1	16	18466.7	18428.35	1.3	94.0
FKB1A	Peptidyl prolyl cis trans isomerase FKBP1A	13	11791.4	11791.44	1.4	96.5
GSTM1[a]	Glutathione S transferase Mu 1	25	26022.7	25838.8	3.7	88.6
GSTM2[a]	Glutathione S transferase Mu 2	18	26022.7	25585.51	3.7	88.6
GSTP1	Glutathione S transferase P 1	22	23479.6	23477.99	2.2	-
LEG3	Galectin 3	11	27326.2	27383.68	n/a	88.6
NQO1	NAD(P)H dehydrogenase quinone 1	14	30871.1	30828.48	n/a	96.5
PA1B3	Platelet activating factor acetylhydrolase IB subunit gamma	8	25764.7	25853.4	17.3	-
PGAM1	Phosphoglycerate mutase 1	26	28742.9	28700.79	1.8	98.0
PRDX1	Peroxiredoxin 1	9	22628.9	22176.5	4.2	-
PROF1	Profilin 1	12	14868.1	14826.02	9.1	99.5
RSU1[c]	Ras suppressor protein 1	12	31372.9	31419.08	3.7	88.6
RSU1[c]	Ras suppressor protein 1	12	31555.4	31419.08	2.8	-
S10A6	Protein S100 A6	8	9961.33	10050.62	2.2	99.5
SERC	Phosphoserine aminotransferase	9	40516.5	40472.58	n/a	-
SODM	Superoxide dismutase Mn mitochondrial	12	22222.6	22222.14	14.9	98.0
SUMO2	Small ubiquitin related modifier 2	4	10519.5	10608.91	3.0	88.6
TAGL2	Transgelin 2	4	22628.9	22264.24	4.2	-
TBCA	Tubulin specific chaperone A	7	12668.6	12626.48	6.1	96.5
TPM1[b]	Tropomyosin alpha 1 chain	30	32751.4	32680.56	6.0	96.5
TPM2[b]	Tropomyosin beta chain	19	32751.4	32836.7	6.0	96.5

Table 6-3: Differentially expressed proteins found to be up-regulated in the Wild-Type MEF cell lysate. **Swiss-Prot Name**: Protein entry in SwissProt database followed by '_MOUSE'. **Description**: Brief description of the protein. **Pep. Hits**: Number of peptides hits used to identify the protein in PLGS2.3. **Exp. Intact Mass**: AutoME-deconvoluted molecular weight of the intact protein. **Theo. Intact Mass**: Intact mass of protein including all annotated PTMs. **Fold-change**: The degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense. The absence of a fold change signifies the protein was only present in one sample. **%Prob. Diff.**: Confidence of the significance of the difference as determined through the analysis of replicate data.

[a.b] Indicate the pairs of proteins in which both were within 100 Da of the differentially expressed mass from the TD data.

[c] Indicates that two differentially expressed masses were found in the TD down data and only one protein with a similar mass was identified in the BU data and was within 100 Da of both of the differentially expressed masses.

Swiss-Prot Name	Description	Pep. Hits	Exp. Intact Mass	Theo. Intact Mass	Fold change	% Prob Diff.
ACBP	Acyl CoA binding protein	11	9911.1	9869.24	1.8	
FABPH	Fatty acid binding protein heart	3	14729.5	14687.66	n/a	
G6PI	Glucose 6 phosphate isomerase	44	62679.3	62635.73	1.3	
GDIR1	Rho GDP dissociation inhibitor 1	22	23318.6	23276.22	1.5	
LDHA	L lactate dehydrogenase A chain	27	36408.7	36367.34	n/a	
LEG1	Galectin 1	11	14771	14734.66	2.3	
MIF	Macrophage migration inhibitory factor	5	12526.3	12373.07	1.6	
NDKB[a]	Nucleoside diphosphate kinase B	10	17457.4	17231.87	2.2	
NENF	Neudesin	11	15628	15612.35	1.4	
PARK7	Protein DJ 1 OS Mus musculus	16	19928.5	20021.31	4.4	
PPIA[a]	Peptidyl prolyl cis trans isomerase A	17	17457.4	17840.15	2.2	
PRDX1[b]	Peroxiredoxin 1	12	22117.8	22176.5	2.4	
PRDX5	Peroxiredoxin 5 mitochondrial	18	16881.6	17014.79	6.6	
SODM[b]	Superoxide dismutase Mn mitochondrial	7	22117.8	22222.14	2.4	

Table 6-4: Proteins determined to be more abundant in the β-arrestin 1,2 double knockout sample. **Swiss-Prot Name**: Protein entry in SwissProt database followed by '_MOUSE'. **Description**: Brief description of the protein. **Pep. Hits**: Number of peptides hits used to identify the protein in PLGS2.3. **Exp. Intact Mass**: AutoME-deconvoluted molecular weight of the intact protein. **Theo. Intact Mass**: Intact mass of protein including all annotated PTMs. **Fold-change**: The degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense. The absence of a fold change signifies the protein was only present in one sample. **%Prob. Diff.**: Confidence of the significance of the difference as determined through the analysis of replicate data.
[a.b] Indicate the pairs of proteins in which both were within 100 Da of the differentially expressed mass from the TD data.

6.9 Figures

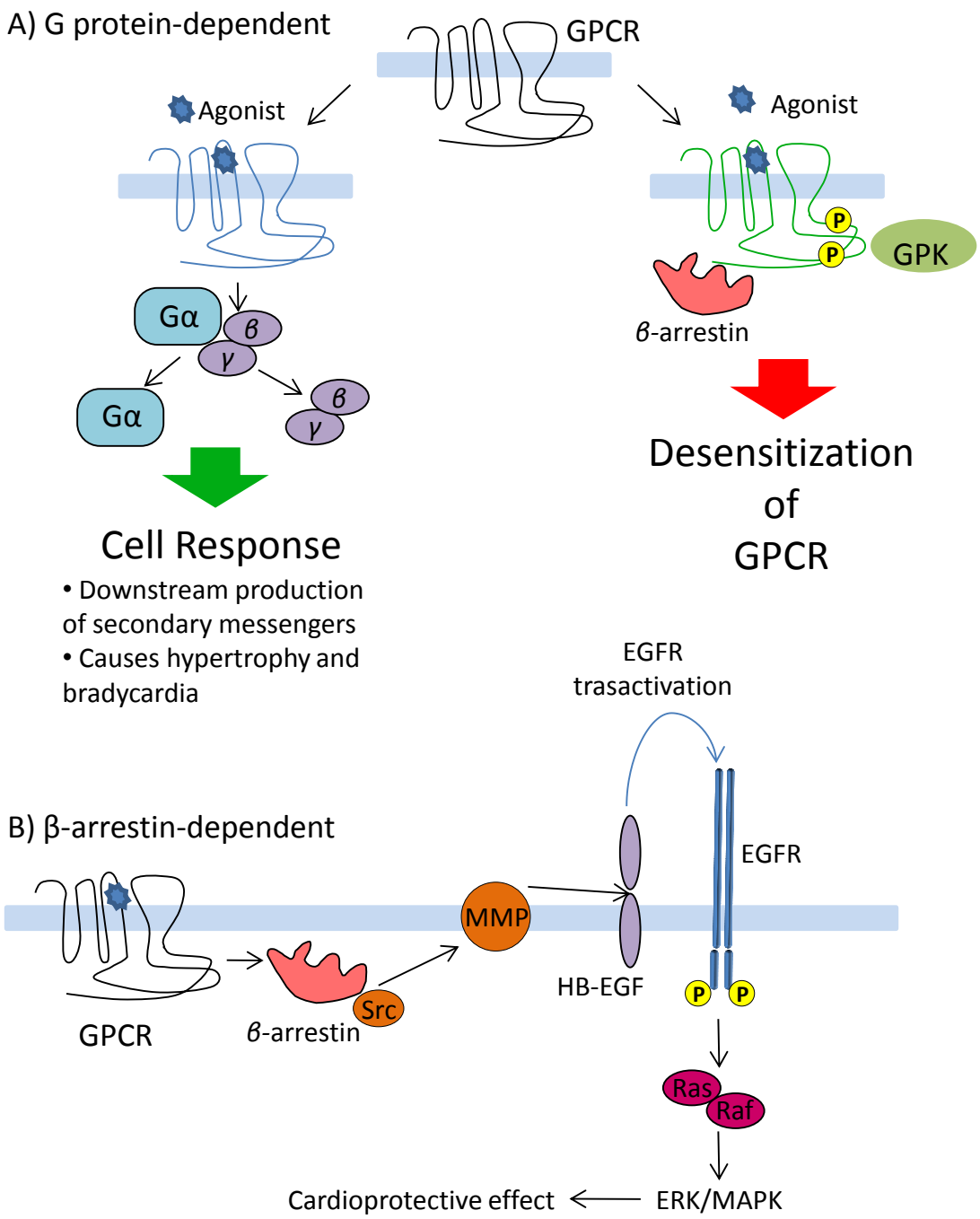


Figure 6-1: Signal transduction by seven transmembrane G protein coupled receptors. A) Classical paradigm involving stimulates G protein signaling and is desensitized by phosphorylation and β-arrestin recruitment. B) New paradigm in which β-arrestins can act as signal transducers themselves. A cardioprotective signaling cascade is shown.

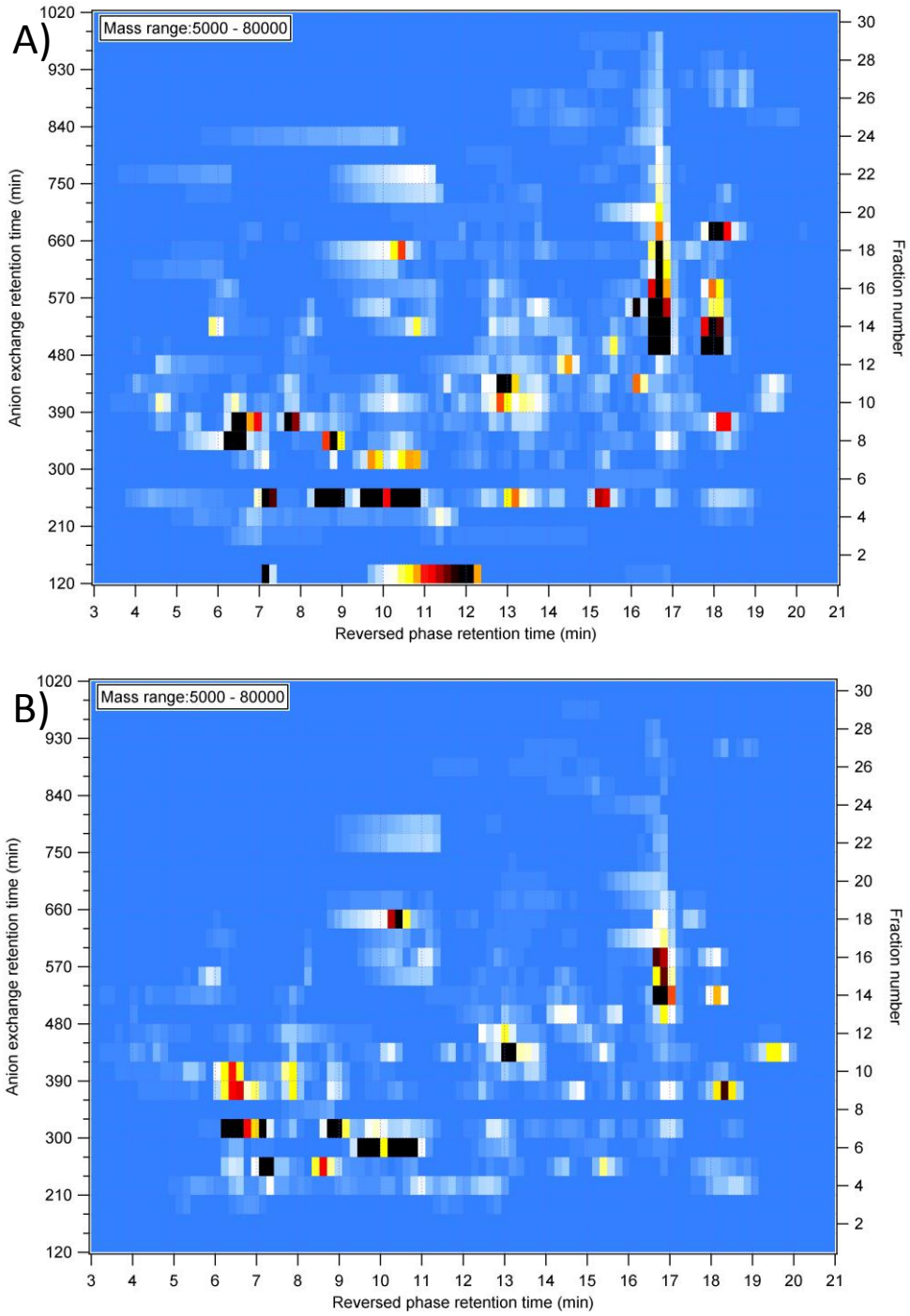


Figure 6-2: 2D chromatograms of AutoME deconvoluted data from intact protein 2D-LC-MS including full molecular weight range of deconvolution. A) Wild-type MEF cells. B) *β*-arrestin 1,2 double knockout MEF cells.

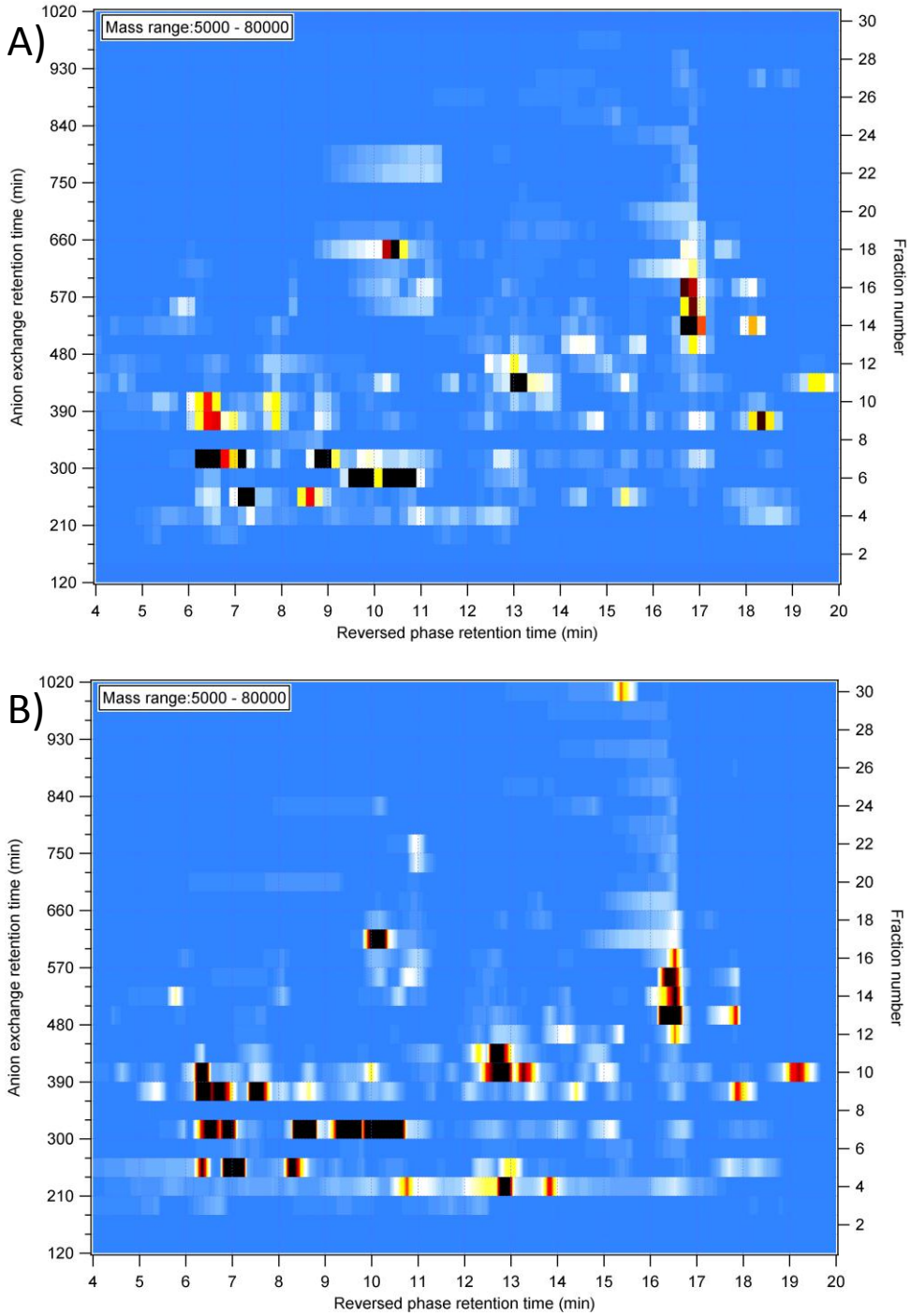


Figure 6-3: 2D chromatograms of deconvoluted intact protein mass spectra for the replicate analysis of *β*arr-KO MEF cell lysates.

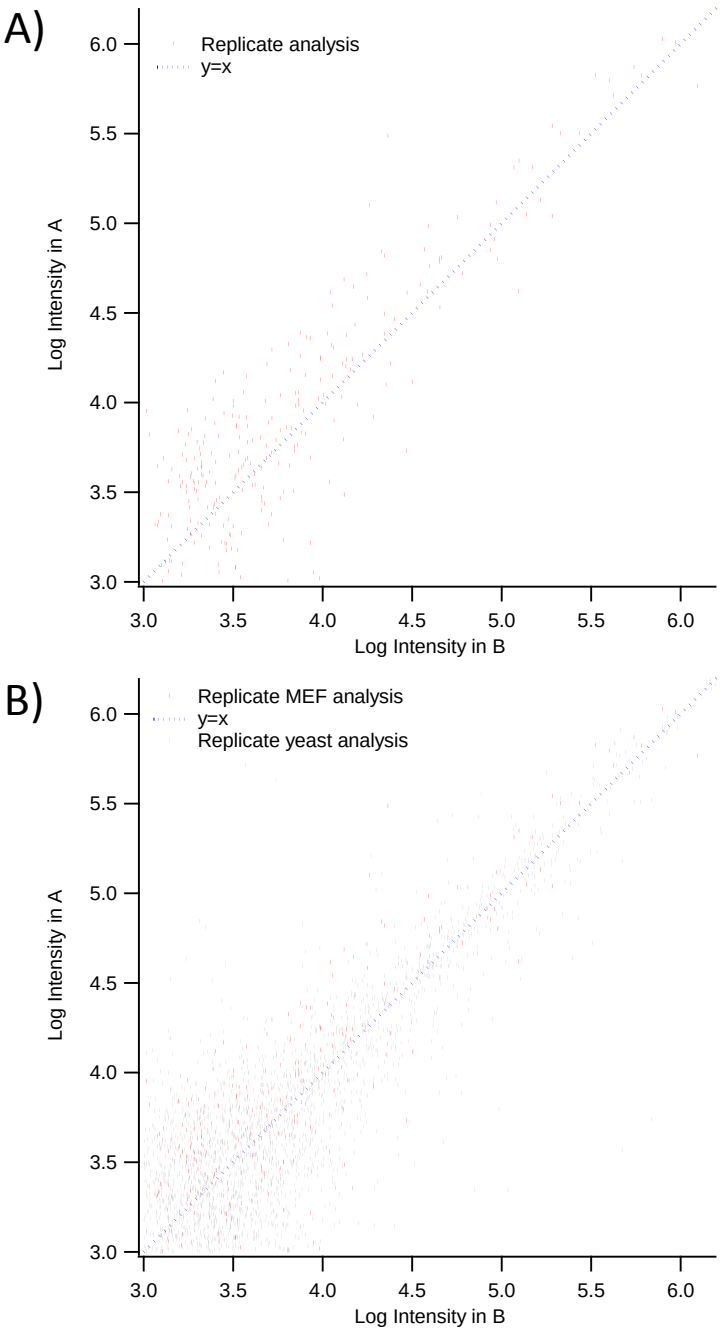


Figure 6-4: Log/Log intensity plots of the replicate analysis of *β*arr-KO MEF cell lysates. A) Replicate data from *β*arr-KO MEF samples only. B) *β*arr-KO MEF replicates overlaid with the replicate analysis of Baker's yeast cell lysates from Chapter 2.

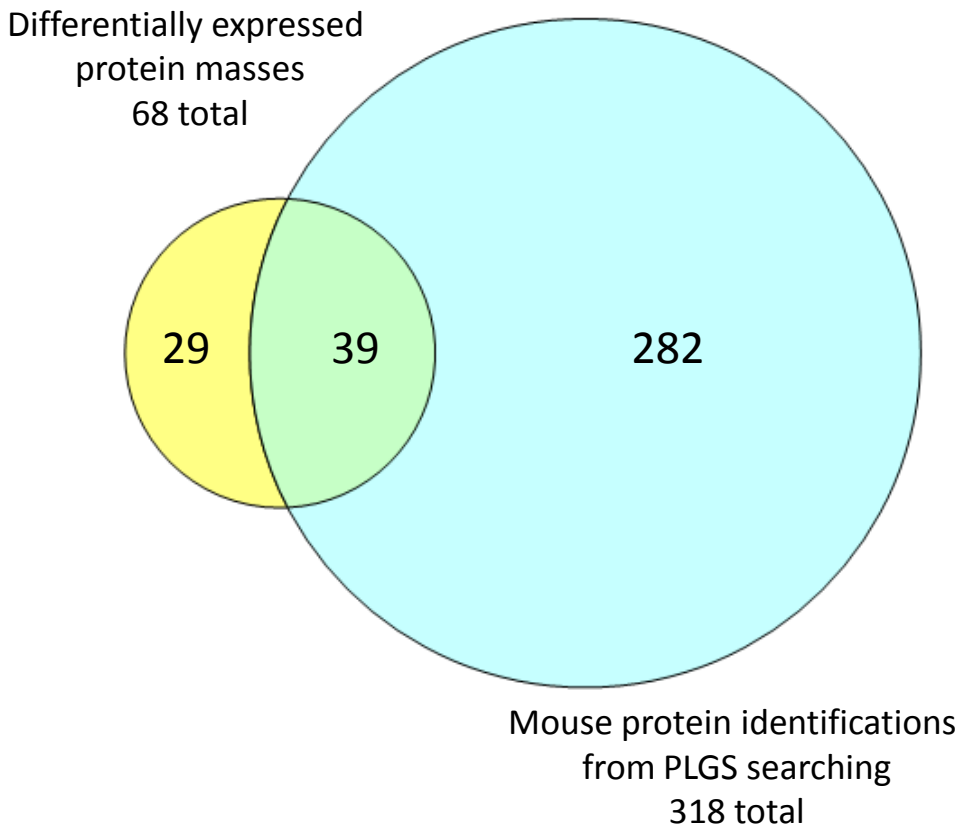


Figure 6-5: Venn diagram illustrating the overlap of proteins selected as differentially expressed from the intact protein intensity comparison and those identified in selected fractions after digestion.

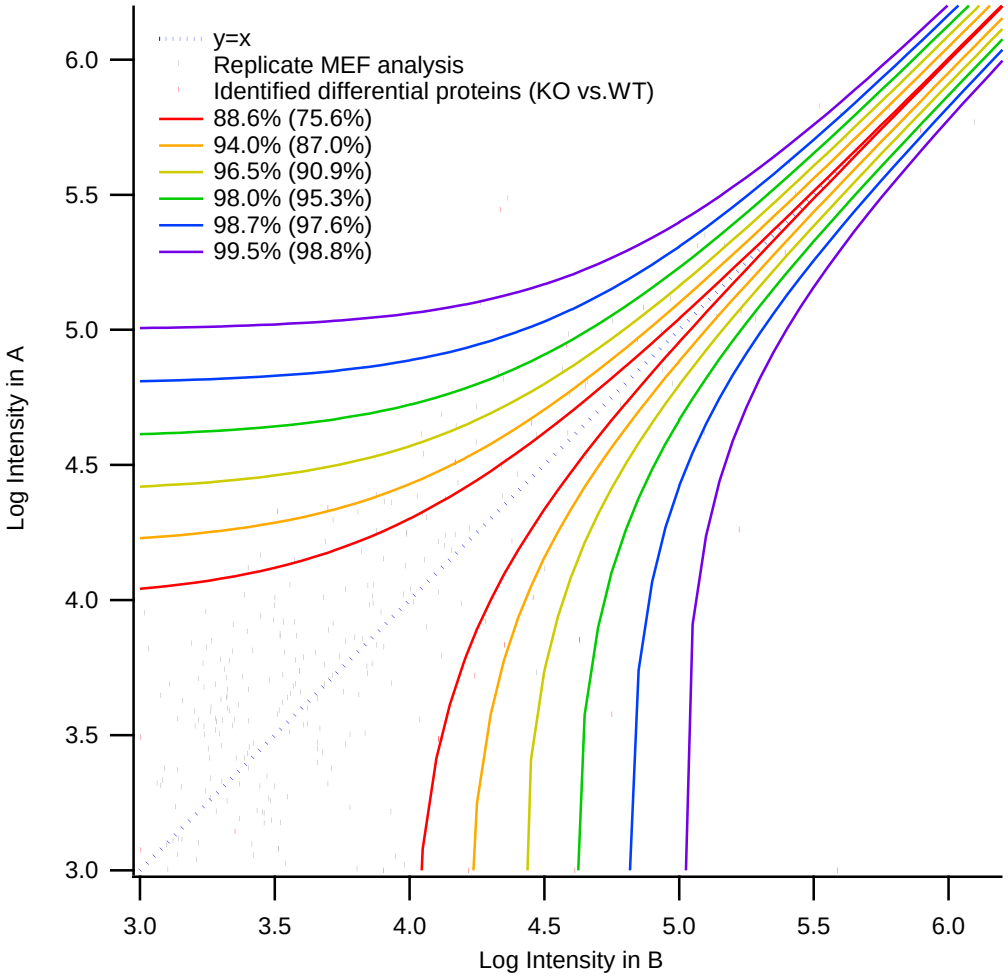


Figure 6-6: Log/Log intensity plot of the replicate analysis with confidence lines from the yeast differential analysis. The percentages noted in the graph legend designate the percentage of replicate points held within the corresponding confidence line for the yeast replicate analysis in Chapter 2.

APPENDIX

The following tables include all of the identified proteins that were determined to have significantly different expression in the bottom-up analyses performed in Chapters 3 and 5.

Table A-1: List of proteins determined to be significantly different in the comparison between the glycerol-grown and dextrose-grown samples both harvested at the log phase. Data were taken from the analysis described in Chapter 3. **Swiss Prot Name**: protein entry in the SwissProt database followed by '_YEAST'. **Ordered Locus Name**: predicted gene that encodes the protein sequence. **Description**: Brief description of the protein. **Pep. Hits**: number of peptide hits used to identify the protein in PLGS2.3. **Intact Mass**: AutoME-deconvoluted molecular weight of the protein. **Up-reg in**: sample in which the protein was most intense. **Fold Change BU**: the degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense for the BU analysis. **Fold Change TD**: If a protein mass matched that of a protein that had a significant difference in Chapter 2, the corresponding fold change is included. A value of N/A for the fold change indicates the protein was only present in one sample. This table continues on the next page.

Swiss-Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up-reg in	Fold Change BU	Fold Change TD
6PGD1	YHR183W	6 phosphogluconate dehydrogenase	24/15	53509	GLY	2.4	
A9LFZ7	ALD6	Cytosolic Aldehyde dehydrogenase	34/35	54379	GLY	1.3	
ACT	YFL039C	Actin	32/23	41662	GLY	1.6	
ADH2	YMR303C	Alcohol dehydrogenase 2	57/24	36708	GLY	42	
ALDH4	YOR374W	Potassium activated aldehyde dehydrogenase	100/24	56688	GLY	28	G – N/A
ARF2	YDL137W	ADP ribosylation factor 2	10/9	20644	GLY	4.1	
ARO8	YGL202W	Aromatic amino acid aminotransferase 1	11/8	56142	GLY	2.1	
BCA2	YJR148W	Branched chain amino acid aminotransferase	10/9	41598	GLY	3.7	
CDC48	YDL126C	Cell division control protein 48	18/14	91938	GLY	2.0	
CH10	YOR020C	10 kDa heat shock protein	5/8	11365	GLY	1.3	
CISY1	YNR001C	Citrate synthase mit.	38/12	53327	GLY	4.8	G – 36
COFI	YLL050C	Cofilin	10/16	15890	GLY	1.3	G – 1.3
DUG1	YFR044C	Cys-Gly metallodipeptidase	17/11	52837	GLY	2.2	
DHAS	YDR158W	Aspartate semialdehyde dehydrogenase	5/7	39518	GLY	1.4	
DHE4	YOR375C	NADP specific glutamate dehydrogenase	19/13	49538	GLY	1.9	
EF1A	YPR080W	Elongation factor 1 alpha	63/47	50001	GLY	2.6	G – N/A
EF1B	YAL003W	Elongation factor 1 beta	14/10	22613	GLY	1.7	
EF1G2	YKL081W	Elongation factor 1 gamma 2	9/14	46490	GLY	1.6	
EF2	YOR133W	Elongation factor 2	51/54	93230	GLY	2.4	
EF3A	YLR249W	Elongation factor 3A	29/15	115919	GLY	3.8	
FHP	YGR234W	Flavohemoprotein	28/13	44618	GLY	3.4	
FKBP	YNL135C	FK506 binding protein	11/9	12150	GLY	1.4	G – 2.9
FPPS	YJL167W	Farnesyl pyrophosphate synthetase	16/8	40457	GLY	2.2	
G3P3	YGR192C	Glyceraldehyde 3 phosphate dehydrogenase 3	58/58	35724	GLY	1.8	

Swiss-Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up-reg in	Fold Change BU	Fold Change TD
GBLP	YMR116C	Guanine nucleotide binding protein	15/8	34783	GLY	1	
H2B1	YDR224C	Histone H2B 1	9/3	14243	GLY	4.0	
HNT1	YDL125C	Hit family protein 1	7/3	17668	GLY	1.6	
HS104	YLL026W	Heat shock protein 104	18/16	101972	GLY	1.4	
HSC82	YMR186W	ATP dependent chaperone	58/69	80849	GLY	1.4	
HSP12	YFL014W	12 kDa heat shock protein	11/21	11685	GLY	1.3	
HSP60	YLR259C	Heat shock protein 60	34/25	60714	GLY	2.4	
HXKB	YGL253W	Hexokinase 1	32/30	53908	GLY	1.8	
HXKG	YCL040W	Hexokinase 2	47/28	55342	GLY	4.7	
IF4A	YKR059W	ATP dependent RNA helicase	25/23	44669	GLY	1.6	
IF5A2	YEL034W	Eukaryotic translate initiation factor 5A	12/15	17103	GLY	1.8	
IPYR	YBR011C	Inorganic pyrophosphatase	17/20	32279	GLY	1.2	
K6PF1	YGR240C	6 phosphofructokinase alpha	15/17	107903	GLY	3.8	
K6PF2	YMR205C	6 phosphofructokinase beta	13/16	104552	GLY	2.1	
LSP1	YPL004C	Sphingolipid long chain base responsive protein	20/10	38047	GLY	3.8	
MDHM	YKL085W	Malate dehydrogenase mit.	29/9	35627	GLY	6.5	G – 6.3
PABP	YER165W	Polyadenylate binding protein	18/12	64304	GLY	1.7	
RIR4	YGR180C	Ribonucleoside diphosphate reductase	13/9	40028	GLY	2.5	G – 5.1
RL4A	YBR031W	60S ribosomal protein L4	10/4	39068	GLY	11	
RL7A	YGL076C	60S ribosomal protein L7	13/4	27621	GLY	12	
RS5	YJR123W	40S ribosomal protein S5	12/4	25023	GLY	9.1	
RS7A	YOR096W	40S ribosomal protein S7	12/3	21608	GLY	7.5	
SAHH	YER043C	Adenosylhomocysteinase	25/27	49094	GLY	1.9	
SNU13	YEL026W	13 kDa ribonucleoprotein associated protein	5/3	13560	GLY	2.4	
SUCA	YOR142W	Succinyl CoA ligase	13/8	35010	GLY	7.5	G – 15.4
SYDC	YLL018C	Aspartyl tRNA synthetase	14/10	63476	GLY	2.5	
SYEC	YGL245W	Glutamyl tRNA synthetase	17/11	80791	GLY	3.2	
SYSC	YDR023W	Seryl tRNA synthetase	16/7	53276	GLY	3.3	
SYV	YGR094W	Valyl tRNA synthetase	27/21	125690	GLY	1.7	
TAL1	YLR354C	Transaldolase	21/26	37013	GLY	1.7	G – 11
TKT1	YPR074C	Transketolase	26/19	73759	GLY	1.4	G – N/A
TSA1	YML028W	Peroxiredoxin TSA1	31/26	21576	GLY	1.5	
YMY9	YMR099C	UPF0010 protein	10/11	33934	GLY	1.5	
ADH1	YOL086C	Alcohol dehydrogenase	41/47	36799	DEX	2.8	G – N/A
ARF1	YLD192W	ADP ribosylation factor 1	9/8	20516	DEX	8.1	
CYPH	YDR155C	Peptidyl prolyl cis trans isomerase	10/23	17379	DEX	1.8	
CYS3	YAL012W	Cystathionine gamma lyase	7/19	42516	DEX	2.1	D – N/A
ENO1	YGR254W	Enolase 1	42/92	46773	DEX	3.4	G – N/A
ENO2	YHR174W	Enolase 2	62/142	46885	DEX	3.0	
G3P1	YJL052W	Glyceraldehyde 3 phosphate dehydrogenase 1	18/27	35727	DEX	2.0	
HSP72	YLL024C	Heat shock protein SSA2	65/89	69427	DEX	1.8	
HSP75	YDL229W	Heat shock protein SSB1	36/47	66560	DEX	0	
HSP7F	YPL106C	Heat shock protein homolog SSE1	21/32	77318	DEX	3.0	
HSP82	YPL240C	ATP dependent molecular chaperone	42/57	81356	DEX	9.3	
IMDH3	YLR432W	Inosine 5 monophosphate dehydrogenase	12/21	56548	DEX	2.0	
KAR	YDL124W	NADPH dependent alpha keto amide reductase	9/22	35538	DEX	2.8	
MET17	YLR303W	Protein MET17	14/26	48641	DEX	3.3	
PDC1	YLR044C	Pyruvate decarboxylase isozymes 1	58/106	61456	DEX	1.8	
PDC6	YGR087C	Pyruvate decarboxylase isozymes 3	10/24	61542	DEX	1.7	
PGK	YCR012W	Phosphoglycerate kinase	69/114	44710	DEX	1.7	D – 2.0
PMG1	YKL152C	Phosphoglycerate mutase 1	24/39	27591	DEX	2.7	
PNC1	YGL037C	Nicotinamidase	5/14	24977	DEX	1.8	
SODC	YJR104V	Superoxide dismutase Cu Zn	24/34	15844	DEX	2.0	D – 1.9
TPIS	YDR050C	Triosephosphate isomerase	36/36	26778	DEX	1.8	D – 1.7
TRX2	YGR209C	Thioredoxin 2	9/15	11196	DEX	3.8	D – N/A
TRXB1	YDR353W	Thioredoxin reductase	7/15	34216	DEX	6.9	D – 3.2

Table A-2: List of proteins determined to be significantly different in the comparison between the log phase and stationary phase harvests of yeast samples grown on dextrose. Data were taken from the analysis described in Chapter 3. **Swiss Prot Name**: protein entry in the SwissProt database followed by '_YEAST'. **Ordered Locus Name**: predicted gene that encodes the protein sequence. **Description**: Brief description of the protein. **Pep. Hits**: number of peptide hits used to identify the protein in PLGS2.3. **Intact Mass**: AutoME-deconvoluted molecular weight of the protein. **Up-reg in**: sample in which the protein was most intense. **Fold Change BU**: the degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense for the BU analysis. **Fold Change TD**: If a protein mass matched that of a protein that had a significant difference in Chapter 2, the corresponding fold change is include. A value of N/A for the fold change indicates the protein was only present in one sample. This table continues on the next page.

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change BU	Fold change TD
ACBP	YGR037C	Acyl-CoA-binding protein	5/3	10055	Log	1.4	L – 2.5
ADH1	YOL086C	Alcohol dehydrogenase	47/33	36799	Log	2.1	
ADH2	YMR303C	Alcohol dehydrogenase 2	19/15	36708	Log	5.4	
ADK	YJR105W	Adensosine kinase	19/15	36349	Log	1.7	
AHP1	YLR109W	Peroxiredoxin type-2	19/16	19102	Log	1.5	
ALF	YKL060C	Fructose-bisphosphate aldolase	60/27	39595	Log	4.7	
BCA2	YJR148W	Branched chain amino acid aminotransferase	9/5	41598	Log	2.4	
BMH1	YER177W	Protein BMH1	21/17	30072	Log	1.9	L – N/A
CBS	YGR155W	Cystathionine beta-synthase	32/15	55987	Log	2.0	
EF1B	YAL003W	Elongation factor 1 beta	10/4	22613	Log	1.5	
EF2	YOR133W	Elongation factor 2	54/31	93230	Log	1.3	
ENO2	YHR174W	Enolase 2	142/125	46885	Log	1.2	
G3P3	YGR192C	Glyceraldehyde 3 phosphate dehydrogenase 3	58/43	35724	Log	2.2	L – 123
HSP60	YLR259C	Heat shock protein 60	25/13	60714	Log	1.5	
HSP71	YAL005C	Heat shock protein SSA1	92/83	69614	Log	1.1	
HSP72	YLL024C	Heat shock protein SSA2	89/73	69427	Log	1.4	
HSP74	YER103W	Heat shock protein SSA4	47/37	69608	Log	2.5	
HSP82	YPL240C	ATP dependent molecular chaperone	57/14	81356	Log	2.4	
KAD1	YDR226W	Adenylate kinase 1	23/5	24239	Log	4.7	L – 1.1
KPYK1	YAL038W	Pyruvate kinase 1	51/27	54510	Log	1.7	
MET17	YLR303W	Protein MET17	26/21	48641	Log	1.6	
PDC1	YLR044C	Pyruvate decarboxylase isozymes 1	106/83	61456	Log	1.3	
PGK	YCR012W	Phosphoglycerate kinase	114/51	44710	Log	3.3	L – N/A
PNC1	YGL037C	Nicotinamidase	14/10	24977	Log	1.4	L – 18
SAHH	YER043C	Adenosylhomocysteinase	27/19	49094	Log	1.2	
SYDC	YLL018C	Aspartyl tRNA synthetase	10/7	63476	Log	1.2	
THRC	YCR053W	Threonine synthase	19/9	57438	Log	1.9	L – N/A
TRX1	YLR043C	Thioredoxin 1	13/7	11227	Log	5.4	
TRX2	YGR209C	Thioredoxin 2	15/11	11196	Log	1.3	
TSA1	YML028W	Peroxiredoxin TSA1	26/24	21576	Log	2.0	
ZEO1	YOL109W	Protein ZEO1	14/6	12581	Log	7.6	
ALDH4	YOR374W	Potassium activated aldehyde dehydrogenase	24/31	56688	Stat	2.4	
AMPL	YKL103C	Vacuolar aminopeptidase 1	14/20	57057	Stat	3.2	
ARO8	YGL202W	Aromatic amino acid aminotransferase 1	8/8	56142	Stat	2.1	
ASSY	YOL058W	Argininosuccinate synthase	23/23	46910	Stat	1.6	
CISY1	YNR001C	Citrate synthase mit.	12/28	53327	Stat	2.4	
DAK1	YML070W	Dihidroxyacetone kinase	10/15	62167	Stat	1.9	

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change BU	Fold change TD
DHE4	YOR375C	NADP specific glutamate dehydrogenase	13/14	49538	Stat	2.2	
ENO1	YGR254W	Enolase 1	92/93	46773	Stat	2.1	
G3P1	YJL052W	Glyceraldehyde 3 phosphate dehydrogenase 1	27/37	35727	Stat	3.0	S – N/A
G6PD	YNL241C	Glucose-6-phosphate 1-dehydrogenase	20/15	27485	Stat	1.8	
G6PI	YBR196C	Glucose-6-phosphate isomerase	51/61	61261	Stat	1.4	S – 1.7
GRE2	YOL151W	NADPH-dependent methylglyoxal reductase	18/23	38145	Stat	1.9	S – N/A
GSHR	YPL091W	Glutathione reductase	9/12	53407	Stat	2.0	
HBN1	YCL026-C-B	Putative nitroreductase	12/24	20980	Stat	5.4	S – 7.1
HS104	YLL026W	Heat shock protein 104	16/19	101972	Stat	1.7	
HSP31	YDR533C	Probable chaperone protein	16/17	25654	Stat	1.4	S – N/A
HXKA	YFR053C	Hexokinase-1	23/33	53704	Stat	2.0	
HXKB	YGL253W	Hexokinase 1	30/27	53908	Stat	1.3	
IPYR	YBR011C	Inorganic pyrophosphatase	20/22	32279	Stat	1.5	
KAR	YDL124W	NADPH dependent alpha keto amide reductase	22/25	35539	Stat	2.0	
MDHM	YKL085W	Malate dehydrogenase mit.	9/21	35627	Stat	2.7	
METE	YER091C	5-methyltetrahydropteroyltriglutamate – homocysteine methyltransferase	32/38	85806	Stat	1.8	
OYE2	YHR179W	NADPH dehydrogenase 2	30/43	44982	Stat	2.6	
OYE3	YPL171C	NADPH dehydrogenase 3	23/40	44892	Stat	3.6	
PMG1	YKL152C	Phosphoglycerate mutase 1	39/49	27591	Stat	1.4	L – 2.0
PROF	YOR122C	Profilin	9/6	13668	Stat	1.8	
PUR92	YMR120C	Bifuntional purine biosynthesis protein ADE17	36/48	65222	Stat	1.7	
PURA	YNL220W	Adenylosuccinate synthetase	11/20	48249	Stat	3.4	
SODC	YJR104V	Superoxide dismutase Cu Zn	34/34	15844	Stat	1.1	L – 1.8
SUCA	YOR142W	Succinyl CoA ligase	8/3	35010	Stat	1.3	
TKT1	YPR074C	Transketolase	19/24	73759	Stat	1.3	
TRXB1	YDR353W	Thioredoxin reductase	15/15	34216	Stat	1.4	
YMN1	YML131W	Uncharacterized membrane protein	8/14	39951	Stat	5.8	
YMY0	YMR090W	UPF0659 protein	9/11	24866	Stat	1.4	
YNN4	YNL134C	Uncharacterized protein YNL134C	40/48	41138	Stat	1.5	

Table A-3: List of proteins determined to be significantly different in the comparison between the glycerol-grown and dextrose-grown samples both harvested at the log phase. Data were taken from the analysis described in Chapter 5. **Swiss Prot Name**: protein entry in the SwissProt database followed by '_YEAST'. **Ordered Locus Name**: predicted gene that encodes the protein sequence. **Description**: Brief description of the protein. **Pep. Hits**: number of peptide hits used to identify the protein in PLGS2.3. **Intact Mass**: Average molecular weight of the protein after loss of initiating Methionine if known. **Up-reg in**: sample in which the protein was most intense. **Fold Change frac**: the degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense for the fractionated BU analysis. **Fold Change un-frac**: If the protein was also found to be differential expressed in the un-fractionated BU analysis, the corresponding fold change is included here. This table continues on the next page.

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
3HAO	YJR025C	3 hydroxyanthranilate 3 4 dioxygenase	4	20222	GLY	3.7	
AATC	YLR027C	Aspartate aminotransferase cytoplasmic	19	46028	GLY	2.4	
ABP1	YCR088W	Actin binding protein	7	65536	GLY	3.9	
ACON	YLR304C	Aconitate hydratase mitochondrial	32	85314	GLY	28.6	
ACPM	YKL192C	Acyl carrier protein mitochondrial	4	13934	GLY	8.6	
ADH2	YMR303C	Alcohol dehydrogenase 2	22	36708	GLY	11.5	G - 42
AGX1	YFL030W	Alanine glyoxylate aminotransferase 1	18	41880	GLY	45.2	
AIP1	YMR092C	Actin interacting protein 1	16	67283	GLY	4.0	
ALDH4	YOR374W	Potassium activated aldehyde dehydrogenase	39	56688	GLY	10.9	G - 28
AP2B	YJR005W	AP 2 complex subunit beta	5	80402	GLY	16.9	
ARA1	YBR149W	D arabinose dehydrogenase NAD P	19	38859	GLY	2.0	
ARO8	YGL202W	Aromatic amino acid aminotransferase 1	18	56142	GLY	3.1	G - 2.1
AROC	YGL148W	Chorismate synthase	13	40812	GLY	3.3	
AROG	YBR249C	Phospho 2 dehydro 3 deoxyheptonate aldolase	7	39724	GLY	9.7	
ARP2	YDL029W	Actin related protein 2	6	44045	GLY	6.1	
ARP3	YJR065C	Actin related protein 3	6	49510	GLY	6.1	
ARPC2	YNR035C	Actin related protein 2 3 complex subunit 2	7	39541	GLY	13.4	
ATIF	YDL181W	ATPase inhibitor mitochondrial	7	9864	GLY	3.2	
ATPA	YBL099W	ATP synthase subunit alpha mitochondrial	14	58572	GLY	11.1	
ATPB	YJR121W	ATP synthase subunit beta mitochondrial	36	54760	GLY	6.7	
ATPO	YDR298C	ATP synthase subunit 5 mitochondrial	7	22800	GLY	5.6	
ATX1	YNL259C	Metal homeostasis factor ATX1	2	8215	GLY	15.4	
BCA2	YJR148W	Branched chain amino acid aminotransferase	9	41598	GLY	2.7	G - 3.7
CAPZB	YIL034C	F-acting-capping protein subunit beta	3	32629	GLY	1.7	
CCPR	YKR066C	Cytochrome c peroxidase mitochondrial	13	40327	GLY	1.1	
CH10	YOR020C	10 kDa heat shock protein mitochondrial	10	11365	GLY	1.9	G - 1.3
CISY1	YNR001C	Citrate synthase mitochondrial	16	53327	GLY	2.7	G - 4.8
CKS1	YBR135W	Cyclin dependent kinases regulatory subunit	3	17783	GLY	4.5	
COX12	YLR038C	Cytochrome c oxidase subunit 6B	4	9781	GLY	4.3	
COX4	YGL187C	Cytochrome c oxidase subunit 4 mit.	3	17131	GLY	17.5	
COX6	YHR051W	Cytochrome c oxidase subunit 6 mit.	10	17330	GLY	2.5	
CUE5	YOR042W	Ubiquitin binding protein CUE5	6	46841	GLY	3.0	
CYPC	YML078W	Peptidyl prolyl cis trans isomerase C mit.	19	19906	GLY	4.6	
DCPS	YLR270W	Scavenger mRNA decapping enzyme DcpS	17	40743	GLY	2.5	
DCS2	YOR173W	Protein DCS2	10	40914	GLY	2.3	
DHE4	YOR375C	NADP specific glutamate dehydrogenase 1	23	49538	GLY	1.4	G - 1.9
DLDH	YFL018C	Dihydrolipoyl dehydrogenase mitochondrial	40	53976	GLY	8.4	
DLHH	YDL086W	Putative carboxymethylenebutenolidase	3	30817	GLY	2.1	
ECM33	YBR078W	Cell wall protein ECM33	4	43741	GLY	6.1	
EF1A	YPR080W	Elongation factor 1 alfpha	18	50033	GLY	1.3	G – 2.6
EF1G2	YPL048W	Elongation factor 1 gamma 2	6	46490	GLY	1.4	G - 1.6
EF2	YOR133W	Elongation factor 2	33	93230	GLY	1.6	G - 2.4
ELP2	YGR200C	Elongator complex protein 2	8	89354	GLY	1.4	

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
ERG19	YOL110W	Diphosphomevalonate decarboxylase	13	44088	GLY	1.7	
ESS1	YJR017C	Peptidyl prolyl cis trans isomerase ESS1	12	19392	GLY	1.9	
ETR1	YBR026C	Enoyl acyl carrier protein reductase NADPH	15	42040	GLY	37.6	
FAT2	YBR222C	Peroxisomal coenzyme A synthetase	11	60450	GLY	1.8	
FHP	YGR234W	Flavohemoprotein	3	44618	GLY	16.5	G - 3.4
FIMB	YDR129C	Fimbrin	19	71728	GLY	1.5	
FKBP	YNL135C	FK-506 binding protein 1	14	12158	GLY	1.3	G – 1.4
FMP21	YBR269C	Protein FMP21 mitochondrial	3	15494	GLY	2.7	
FUMH	YPL262W	Fumarate hydratase mitochondrial	21	53118	GLY	96.8	
G3P1	YJL052W	Glyceraldehyde 3 phosphate dehydrogenase 1	14	35727	GLY	2.1	D - 2.0
G3P3	YGR192C	Glyceraldehyde 3 phosphate dehydrogenase 3	30	35724	GLY	1.7	G - 1.8
GBLP	YMR116C	Guanine nucleotide binding protein subunit	5	34783	GLY	6.5	G - 1.0
GCY	YOR120W	Protein GCY	30	35057	GLY	1.5	
GDIR	YDL135C	Rho GDP dissociation inhibitor	8	23124	GLY	3.1	
H2B1	YDR224C	Histone H2B 1	1	14243	GLY	8.9	G - 4.0
HS104	YLL026W	Heat shock protein 104	12	101972	GLY	1.8	G - 1.4
HSP12	YFL014W	12 kDa heat shock protein	17	11685	GLY	1.5	G - 1.3
HSP77	YJR045C	Heat shock protein SSC1 mitochondrial	29	70584	GLY	2.4	
IDH1	YNL037C	Isocitrate dehydrogenase NAD subunit 1 mito	18	39299	GLY	28.2	
IDH2	YOR136W	Isocitrate dehydrogenase NAD subunit 2 mito	12	39714	GLY	38.7	
IF4B	YPR163C	Eukaryotic translation initiation factor 4B	10	48493	GLY	4.2	
IF6	YPR016C	Eukaryotic translation initiation factor 6	10	26441	GLY	5.1	
IPB2	YNL015W	Protease B inhibitors 2 and 1	13	8584	GLY	7.0	
LEU1	YNL104C	2 isopropylmalate synthase	6	68366	GLY	9.2	
MBF1	YOR298C	Multiprotein bridging factor 1	5	16393	GLY	1.9	
MDHM	YKL085W	Malate dehydrogenase mitochondrial	29	35627	GLY	3.2	G - 6.5
MDHP	YDL078C	Malate dehydrogenase peroxisomal	18	37162	GLY	26.8	
MLC1	YGL106W	Myosin light chain 1	14	16434	GLY	1.6	
MMF1	YIL051C	Protein MMF1 mitochondrial	11	15898	GLY	16.1	
MNP1	YBL068W	54S ribosomal protein L12 mitochondrial	7	20637	GLY	5.1	
MPG1	YDL055C	Mannose 1 phosphate guanyltransferase	5	39541	GLY	1.7	
MPI	YER003C	Mannose 6 phosphate isomerase	15	48158	GLY	3.9	
MRP8	YKL142W	Uncharacterized protein MRP8	27	25081	GLY	1.2	
NACA	YHR193C	Nascent polypeptide associated complex sub.	6	18697	GLY	1.9	
NACB1	YPL037C	Nascent polypeptide associated complex sub.	5	17009	GLY	12.2	
NDK	YKL067W	Nucleoside diphosphate kinase	12	17155	GLY	1.4	
NHP6A	YPR052C	Non histone chromosomal protein 6A	6	10795	GLY	20.9	
ODPB	YBR221C	Pyruvate dehydrogenase E1 component sub.	9	40028	GLY	2.9	
PABP	YER165W	Polyadenylate binding protein cytoplasmic	3	64304	GLY	3.0	G - 1.7
PFD6	YLR200W	Prefoldin subunit 6	6	13275	GLY	3.1	
PHO85	YPL031C	Cyclin dependent protein kinase PHO85	6	34883	GLY	7.6	
PRX1	YBL064C	Mitochondrial peroxiredoxin PRX1	11	29477	GLY	3.0	
PSA4	YGR135W	Proteasome component Y13	6	28696	GLY	6.0	
PUR91	YLR028C	Bifunctional purine biosynthesis ADE16	21	65241	GLY	1.1	
PYRD	YKL216W	Dihydroorotate dehydrogenase	8	34778	GLY	4.4	
QCR2	YPR191W	Cytochrome b c1 complex subunit 2 mit.	21	40453	GLY	14.1	
RIB4	YOL143C	6 7 dimethyl 8 ribityllumazine synthase	5	18543	GLY	4.3	
RIR4	YGR180C	Ribonucleoside diphosphate reductase	24	40028	GLY	7.7	G - 2.5
RL12	YEL054C	60S ribosomal protein L12	3	17811	GLY	2.5	
RL30	YGL030W	60S ribosomal protein L30	3	11408	GLY	2.3	
RLA0	YLR340W	60S acidic ribosomal protein P0	4	33745	GLY	22.5	
RPAB4	YHR143W	DNA directed RNA polymerases I II and III	1	7711	GLY	2.4	
RS12	YOR369C	40S ribosomal protein S12	7	15462	GLY	3.0	
RS20	YHL015W	40S ribosomal protein S20	2	13898	GLY	28.7	
RS21A	YKR057W	40S ribosomal protein S21 A	9	9739	GLY	8.7	
RS21B	YJL136C	40S ribosomal protein S21 B	7	9753	GLY	2.8	
RS27A	YKL156W	40S ribosomal protein S27 A	4	8873	GLY	2.0	
SCW4	YGR279C	Probable family 17 glucosidase SCW4	8	40148	GLY	7.5	
SEC23	YPR181C	Protein transport protein SEC23	15	85330	GLY	4.4	
SGS1	YMR190C	ATP dependent helicase SGS1	12	163734	GLY	4.5	
SODM	YHR008C	Superoxide dismutase Mn mitochondrial	4	25758	GLY	2.4	
SUCA	YOR142W	Succinyl CoA ligase ADP forming sub.	27	35010	GLY	4.3	G - 7.5
SYDC	YLL018C	Aspartyl tRNA synthetase cytoplasmic	22	63476	GLY	2.5	G - 2.5

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
SYV	YGR094W	Valyl tRNA synthetase mitochondrial	40	125690	GLY	1.9	G - 1.7
TAL1	YLR354C	Transaldolase	25	37013	GLY	1.3	G - 1.7
TBCA	YOR265W	Tubulin specific chaperone A	7	12371	GLY	2.1	
THIL	YPL028W	Acetyl CoA acetyltransferase	15	41702	GLY	1.4	
TIM10	YHR005C	Mit. import inner membrane translocase	6	10297	GLY	2.5	
TIM13	YGR181W	Mit. import inner membrane translocase	8	11278	GLY	34.2	
TIM9	YEL020W	Mit. import inner membrane translocase	5	10194	GLY	2.4	
TRX1	YLR043C	Thioredoxin 1	11	11227	GLY	2.4	
UBIQ	YIL148W	Ubiquitin	6	8551	GLY	1.7	
UGPA1	YKL035W	UTP glucose 1 phosphate uridylyltransferase	22	55953	GLY	1.7	
ULS1	YOR191W	ATP dependent helicase ULS1	19	184290	GLY	14.3	
VATA	YKL080W	V type proton ATPase catalytic subunit A	21	118562	GLY	2.1	
VATC	YKL080W	V type proton ATPase subunit C	24	44161	GLY	2.0	
VATE	YOR332W	V type proton ATPase subunit E	10	26455	GLY	2.5	
WTM1	YOR230W	Transcriptional modulator WTM1	27	48353	GLY	1.5	
YBM6	YBR016W	Uncharacterized protein YBR016W	3	14607	GLY	5.5	
YCP4	YCR004C	Flavoprotein like protein YCP4	12	26333	GLY	5.6	
YMH9	YML079W	Uncharacterized protein YML079W	6	22447	GLY	1.8	
YN14	YNL274C	Putative 2 hydroxyacid dehydrogenase	4	38807	GLY	14.4	
YO285	YOR285W	Putative thiosulfate sulfurtransferase	2	15403	GLY	4.0	
YP067	YPL067C	Uncharacterized protein YPL067C	7	22761	GLY	3.5	
YP225	YPL225W	UPF0368 protein YPL225W	13	17433	GLY	2.0	
ZEO1	YOL109W	Protein ZEO1	14	12581	GLY	2.2	
ACE2	YLR131C	Metallothionein expression activator	9	86580	DEX	1.7	
ACT	YFL039C	Actin	7	41662	DEX	3.4	
ADH1	YOL086C	Alcohol dehydrogenase 1	19	36799	DEX	14.7	D - 2.8
ADH6	YMR318C	NADP dependent alcohol dehydrogenase 6	8	39591	DEX	3.3	
ADK	YJR105W	Adenosine kinase	8	36349	DEX	21.9	
AHP1	YLR109W	Peroxiredoxin type 2	12	19102	DEX	20.8	
ALF	YKL060C	Fructose bisphosphate aldolase	14	39595	DEX	2.2	
AMPL	YKL103C	Vacuolar aminopeptidase 1	10	57057	DEX	2.9	
BMH1	YER177W	Protein BMH1	22	29960	DEX	1.2	G – 1.4
CALM	YBR109C	Calmodulin	10	16124	DEX	4.5	
CBS	YGR155W	Cystathionine beta synthase	30	55987	DEX	1.9	
CCS1	YMR038C	Superoxide dismutase 1 copper chaperone	6	27312	DEX	9.7	
COFI	YLL050C	Cofilin	5	15900	DEX	1.5	G – 1.3
CPYI	YLR178C	Carboxypeptidase Y inhibitor	3	24357	DEX	2.0	
CYPD	YDR304C	Peptidyl prolyl cis trans isomerase D	12	25310	DEX	1.5	
CYPH	YDR155C	Peptidyl prolyl cis trans isomerase	12	17379	DEX	2.0	
CYS3	YAL012W	Cystathionine gamma lyase	17	42516	DEX	4.8	D - 2.1
DHAS	YDR158W	Aspartate semialdehyde dehydrogenase	13	39518	DEX	1.7	G - 1.4
DHOM	YJR139C	Homoserine dehydrogenase	23	38478	DEX	1.7	
DOHH	YJR070C	Deoxyhypusine hydroxylase	5	36142	DEX	1.7	
DUG1	YFR044C	Glutamate carboxypeptidase like protein	6	52837	DEX	5.9	G - 2.2
ENO1	YGR254W	Enolase 1	28	46773	DEX	3.1	D - 3.4
ENO2	YHR174W	Enolase 2	35	46885	DEX	7.0	D - 3.0
FADH	YDL168W	S hydroxymethyl glutathione dehydrogenase	19	41015	DEX	2.3	
FAS2	YPL231W	Fatty acid synthase subunit alpha	26	206816	DEX	5.7	
FES1	YBR101C	Hsp70 nucleotide exchange factor FES1	6	32604	DEX	1.8	
FMP31	YOR286W	Putative thiosulfate sulfurtransferase FMP31	5	16686	DEX	2.1	
FRDS	YEL047C	Fumarate reductase	8	50812	DEX	6.2	
FSH1	YHR049W	Family of serine hydrolases 1	12	27322	DEX	6.6	
G3P2	YJR009C	Glyceraldehyde 3 phosphate dehydrogenase 2	23	35824	DEX	1.6	
G6PD	YNL241C	Glucose 6 phosphate 1 dehydrogenase	24	57485	DEX	3.4	
G6PI	YBR196C	Glucose 6 phosphate isomerase	25	61261	DEX	1.3	
GLRX1	YCL035C	Glutaredoxin 1	12	12372	DEX	7.4	
GLRX2	YDR513W	Glutaredoxin 2	12	12380	DEX	1.3	
GSHR	YPL091W	Glutathione reductase	22	53407	DEX	2.9	
GTO2	YGR154C	Glutathione S transferase omega like 2	16	43246	DEX	4.1	
HBN1	YCL026C	Putative nitroreductase HBN1	8	20980	DEX	5.1	
HCH1	YNL281W	Hsp90 co chaperone HCH1	7	17235	DEX	2.3	
HIS7	YOR202W	Imidazoleglycerol phosphate dehydratase	5	23818	DEX	4.8	
HOSC	YDL182W	Homocitrate synthase cytosolic isozyme	9	47069	DEX	3.6	

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
HPRT	YDR138W	Hypoxanthine guanine phosphoribosyltransfe.	8	25174	DEX	15.9	
HSC82	YMR186W	ATP dependent molecular chaperone HSC82	54	80849	DEX	1.2	G - 1.4
HSP31	YDR533C	Probable chaperone protein HSP31	13	25654	DEX	463.9	
HSP60	YLR259C	Heat shock protein 60 mitochondrial	33	60714	DEX	1.6	G - 2.4
HSP72	YLL024C	Heat shock protein SSA2	26	69427	DEX	3.7	D - 1.8
HSP73	YBL075C	Heat shock protein SSA3	19	70503	DEX	3.3	
HSP74	YER103W	Heat shock protein SSA4	15	69608	DEX	4.4	
HSP75	YDL229W	Heat shock protein SSB1	31	66560	DEX	1.6	D - 1.0
HSP76	YNL209W	Heat shock protein SSB2	34	66553	DEX	1.7	
HSP79	YBR169C	Heat shock protein homolog SSE2	29	77572	DEX	3.2	
HSP7Q	YLR369W	Heat shock protein SSQ1 mitochondrial	5	72320	DEX	1.7	
HSP82	YPL240C	ATP dependent molecular chaperone HSP82	44	81356	DEX	2.1	D - 9.3
HXKB	YGL253W	Hexokinase 2	6	53908	DEX	1.8	
IDI1	YPL117C	Isopentenyl diphosphate Delta isomerase	1	33330	DEX	2.9	
IF4A	YKR059W	ATP dependent RNA helicase eIF4A	29	44669	DEX	1.2	G - 1.6
IF5A2	YEL034W	Eukaryotic translation initiation factor 5A 2	2	17103	DEX	14.8	G - 1.8
IMDH2	YHR216W	Inosine 5 monophosphate dehydrogenase	15	56493	DEX	4.1	
IMDH3	YLR432W	Probable inosine 5 monophosphate dehydrog.	19	56548	DEX	6.2	D - 2.0
IMDH4	YML056C	Probable inosine 5 monophosphate dehydrog.	9	56357	DEX	5.1	
IPYR	YBR011C	Inorganic pyrophosphatase	12	32279	DEX	1.9	
KAD1	YDR226W	Adenylate kinase 1	21	24239	DEX	1.3	
KES1	YPL145C	Protein KES1	10	49461	DEX	1.5	
KPYK1	YAL038W	Pyruvate kinase 1	25	54510	DEX	2.6	
MAS5	YNL064W	Mitochondrial protein import protein MAS5	7	44642	DEX	3.2	
MET17	YLR303W	Protein MET17	22	48641	DEX	1.7	D - 3.3
METE	YER091C	5 methyltetrahydropteroyltriglutamate homoC	47	85806	DEX	1.9	
METK1	YLR180W	S adenosylmethionine synthetase 1	10	41792	DEX	24.9	
METK2	YDR502C	S adenosylmethionine synthetase 2	10	42229	DEX	10.3	
NIF3	YGL221C	NGG1 interacting factor 3	4	31868	DEX	6.6	
NPT1	YOR209C	Nicotinate phosphoribosyltransferase	10	48987	DEX	2.3	
NTF2	YER009W	Nuclear transport factor 2	6	14444	DEX	1.8	
OYE2	YHR179W	NADPH dehydrogenase 2	25	44982	DEX	9.8	
OYE3	YPL171C	NADPH dehydrogenase 3	21	44892	DEX	94.8	
PCNA	YBR088C	Proliferating cell nuclear antigen	7	28897	DEX	1.5	
PDC1	YLR044C	Pyruvate decarboxylase isozyme 1	26	61456	DEX	15.2	D - 1.8
PGK	YCR012W	Phosphoglycerate kinase	37	44710	DEX	2.5	D - 1.7
PMG1	YKL152C	Phosphoglycerate mutase 1	25	27591	DEX	2.8	D - 2.7
PNC1	YGL037C	Nicotinamidase	4	24977	DEX	1.7	D - 1.8
PROF	YOR122C	Profilin	7	13677	DEX	1.4	
PRP5	YBR237W	Pre mRNA processing ATP dep. RNA helic.	9	96299	DEX	4.4	
PSB2	YER012W	Proteasome component C11	5	22502	DEX	2.0	
PUR6	YOR128C	Phosphoribosylaminoimidazole carboxylase	11	62299	DEX	3.7	
PUR92	YMR120C	Bifunctional purine biosynthesisADE17	40	65222	DEX	22.0	
RIR2	YJL026W	Ribonucleoside diphosphate reductase	4	46118	DEX	1.7	
RLA4	YDR382W	60S acidic ribosomal protein P2 beta	3	11043	DEX	4.0	
SDO1L	YHR087W	SDO1 like protein YHR087W	9	12002	DEX	3.3	
SEC31	YDL195W	Protein transport protein SEC31	9	138617	DEX	1.5	
SGT2	YOR007C	Small glutamine rich tetratricopeptide repeat	12	37195	DEX	1.7	
SODC	YJR104C	Superoxide dismutase Cu Zn	14	15844	DEX	1.7	
SRO9	YCL037C	RNA binding protein SRO9	4	48030	DEX	6.2	
STI1	YOR027W	Heat shock protein STI1	19	66224	DEX	1.7	
SYC	YNL247W	Cysteinyl tRNA synthetase	9	87475	DEX	1.4	
TPIS	YDR050C	Triosephosphate isomerase	20	26778	DEX	2.5	D - 1.8
TRX2	YGR209C	Thioredoxin 2	11	11196	DEX	4.0	D - 3.8
TRXB1	YDR353W	Thioredoxin reductase 1	11	34216	DEX	2.4	D - 6.9
UBA1	YKL210W	Ubiquitin activating enzyme E1 1	11	114194	DEX	2.3	
YB085	YBR085C	Uncharacterized protein YBR085C A	4	9398	DEX	5.8	
YGP1	YNL160W	Protein YGP1	6	37305	DEX	9.6	
YHU6	YGR149C	Uncharacterized protein YHR146W	5	51084	DEX	3.5	
YL301	YLR301W	Uncharacterized protein YLR301W	6	27483	DEX	1.6	
YN034	YNR034W	Uncharacterized protein YNR034W A	6	10774	DEX	5.1	

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
YNN4	YNL134C	Uncharacterized protein YNL134C	17	41138	DEX	62.4	
YPD1	YDL235C	Phosphorelay intermediate protein YPD1	5	19156	DEX	2.3	
YPR1	YDR368W	Putative reductase 1	14	34733	DEX	1.8	

Table A-4: List of proteins determined to be significantly different in the comparison between the log phase and stationary phase harvests of yeast samples grown on dextrose. Data were taken from the analysis described in Chapter 5. **Swiss Prot Name**: protein entry in the SwissProt database followed by '_YEAST'. **Ordered Locus Name**: predicted gene that encodes the protein sequence. **Description**: Brief description of the protein. **Pep. Hits**: number of peptide hits used to identify the protein in PLGS2.3. **Intact Mass**: Average molecular weight of the protein after loss of initiating Methionine if known. **Up-reg in**: sample in which the protein was most intense. **Fold Change frac**: the degree to which a protein was up-regulated expressed as multiples of the intensity of the protein in the sample in which it was least intense for the fractionated BU analysis. **Fold Change un-frac**: If the protein was also found to be differential expressed in the un-fractionated BU analysis, the corresponding fold change is included here. This table continues on the next page.

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
AATC	YLR027C	Aspartate aminotransferase cytoplasmic	12	46028	LOG	1.9	
ADH1	YOL086C	Alcohol dehydrogenase 1	19	36692	LOG	1.3	L – 2.1
ADH3	YMR083W	Alcohol dehydrogenase 3 mitochondrial	4	40344	LOG	4.5	
ADH6	YMR318C	NADP dependent alcohol dehydrogenase 6	8	39591	LOG	15.3	
AHP1	YLR109W	Peroxiredoxin type 2	12	19102	LOG	1.6	L - 1.5
ARF2	YDL137W	ADP ribosylation factor 2	5	20644	LOG	2.4	
CALM	YBR109C	Calmodulin	10	16124	LOG	7.6	
CCPR	YKR066C	Cytochrome c peroxidase mitochondrial	7	40327	LOG	4.7	
CH10	YOR020C	10 kDa heat shock protein mitochondrial	11	11365	LOG	19.4	
CPGL	YFR044C	Cys-Gly metallodipeptidase DUG1	6	52871	LOG	1.5	S – 1.1
CYPC	YML078W	Peptidyl prolyl cis trans isomerase C mitocho	11	19906	LOG	2.5	
CYPD	YDR304C	Peptidyl prolyl cis trans isomerase D	12	25310	LOG	15.1	
CYPH	YDR155C	Peptidyl prolyl cis trans isomerase	12	17379	LOG	3.6	
CYS3	YAL012W	Cystathionine gamma lyase	17	42516	LOG	2.0	
DCPS	YLR270W	Scavenger mRNA decapping enzyme DcpS	9	40743	LOG	2.9	
DHAS	YDR158W	Aspartate semialdehyde dehydrogenase	13	39518	LOG	7.5	
DHOM	YJR139C	Homoserine dehydrogenase	23	38478	LOG	1.8	
DUT	YBR252W	Deoxyuridine 5 triphosphate nucleotidohydro.	7	15297	LOG	1.5	
EF1A	YPR080W	Elongation factor 1 alpha	19	50001	LOG	1.3	
ESS1	YJR017C	Peptidyl prolyl cis trans isomerase ESS1	5	19392	LOG	8.0	
EST1	YLR233C	Telomere elongation protein EST1	5	81747	LOG	8.3	
FADH	YDL168W	S hydroxymethyl glutathione dehydrogenase	19	41015	LOG	1.4	
FAT2	YBR222C	Peroxisomal coenzyme A synthetase	6	60450	LOG	4.5	
FIMB	YDR129C	Fimbrin	19	71773	LOG	1.3	
FKBP	YNL135C	FK506 binding protein 1	15	12150	LOG	5.3	
FKBP2	YDR519W	FK506 binding protein 2	8	14477	LOG	8.0	
FRDS	YEL047C	Fumarate reductase	8	50812	LOG	6.0	
FSH1	YHR049W	Family of serine hydrolases 1	12	27322	LOG	5.0	
G4P1	YGL105W	GU4 nucleic binding protein 1	12	42057	LOG	5.7	
GCY	YOR120W	Protein GCY	30	35057	LOG	2.5	
GDIR	YDL135C	Rho GDP dissociation inhibitor	2	23124	LOG	2.6	
GLRX1	YCL035C	Glutaredoxin 1	12	12372	LOG	1.4	
GPX3	YIR037W	Peroxiredoxin HYR1	7	18629	LOG	20.8	
GRP78	YJL034W	78 kDa glucose regulated protein homolog	11	74422	LOG	1.6	
HOSC	YDL182W	Homocitrate synthase cytosolic isozyme	9	47069	LOG	2.0	
HPRT	YDR399W	Hypoxanthine guanine phosphoribosyltransf.	8	25174	LOG	2.0	
HSC82	YMR186W	ATP dependent molecular chaperone HSC82	54	80849	LOG	1.3	
HSP12	YFL014W	12 kDa heat shock protein	18	11685	LOG	19.6	
HSP31	YDR533C	Probable chaperone protein HSP31	13	25654	LOG	1.1	S - 1.4
HSP60	YLR259C	Heat shock protein 60 mitochondrial	33	60714	LOG	1.1	L - 1.5
HSP74	YER103W	Heat shock protein SSA4	15	69608	LOG	1.7	L - 2.5
HSP76	YNL209W	Heat shock protein SSB2	34	66553	LOG	1.9	

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
HSP77	YJR045C	Heat shock protein SSC1 mitochondrial	27	70584	LOG	1.3	
HSP7E	YEL030W	Heat shock protein SSC3 mitochondrial	11	70041	LOG	10.6	
IF4A	YKR059W	ATP dependent RNA helicase eIF4A	29	44669	LOG	5.7	
IF5A2	YEL034W	Eukaryotic translation initiation factor 5A 2	2	17103	LOG	1.9	
IMDH3	YLR432W	Probable inosine 5 monophosphate dehydrog.	19	56548	LOG	2.8	
KES1	YPL145C	Protein KES1	10	49461	LOG	4.2	
KGUA	YDR454C	Guanylate kinase	15	20624	LOG	6.8	
KPYK2	YOR347C	Pyruvate kinase 2	6	55160	LOG	18.1	
MET4	YNL103W	Transcriptional activator of sulfur metabolism	3	74328	LOG	5.3	
METE	YER091C	5 methyltetrahydropteroyltriglutamate homoC	47	85806	LOG	1.2	S - 1.8
MLC1	YGL106W	Myosin light chain 1	6	16434	LOG	4.0	
MMF1	YIL051C	Protein MMF1 mitochondrial	8	15898	LOG	1.6	
MRP8	YKL142W	Uncharacterized protein MRP8	23	25081	LOG	3.7	
NACA	YHR193C	Nascent polypeptide associated complex sub.	6	18697	LOG	6.4	
NDK	YKL067W	Nucleoside diphosphate kinase	10	17155	LOG	1.7	
NET1	YJL076W	Nucleolar protein NET1	15	128453	LOG	22.5	
PCNA	YBR088C	Proliferating cell nuclear antigen	7	28897	LOG	1.6	
PDX3	YBR035C	Pyridoxamine 5 phosphate oxidase	7	26891	LOG	1.6	
PGK	YCR012W	Phosphoglycerate kinase	37	44710	LOG	2.3	L - 3.3
PNC1	YGL037C	Nicotinamidase	4	24977	LOG	3.9	L - 1.4
PROF	YOR122C	Profilin	7	13668	LOG	2.1	S - 1.8
PRX1	YBL064C	Mitochondrial peroxiredoxin PRX1	7	29477	LOG	2.5	
PUR7	YAR015W	Phosphoribosylaminoimidazole succinocarb.	6	34510	LOG	39.3	
PUR91	YLR028C	Bifunctional purine biosynthesis ADE16	24	65241	LOG	15.4	
PYRE	YML106W	Orotate phosphoribosyltransferase 1	9	24649	LOG	6.4	
RIB4	YOL143C	6 7 dimethyl 8 ribityllumazine synthase	2	18543	LOG	1.9	
RS27A	YKL156W	40S ribosomal protein S27 A	5	8873	LOG	2.3	
SDO1L	YHR087W	SDO1 like protein YHR087W	9	12002	LOG	15.0	
SEC14	YMR079W	SEC14 cytosolic factor	15	34878	LOG	2.2	
SGT2	YOR007C	Small glutamine rich tetratricopeptide repeat	12	37195	LOG	9.5	
SMT3	YDR510W	Ubiquitin like protein SMT3	5	11589	LOG	2.3	
SNU13	YEL026W	13 kDa ribonucleoprotein associated protein	4	13560	LOG	42.1	
STI1	YOR027W	Heat shock protein STI1	19	66224	LOG	2.0	
SUCA	YOR142W	Succinyl CoA ligase ADP forming sub.	19	35010	LOG	4.3	S - 1.3
SYEC	YGL245W	Glutamyl tRNA synthetase cytoplasmic	5	80791	LOG	1.9	
SYV	YGR094W	Valyl tRNA synthetase mitochondrial	31	125690	LOG	1.6	
THIL	YPL028W	Acetyl CoA acetyltransferase	15	41702	LOG	1.8	
TKT1	YPR074C	Transketolase 1	19	73759	LOG	5.1	S - 1.3
TRX1	YLR043C	Thioredoxin 1	10	11227	LOG	1.6	L - 5.4
TSA2	YDR453C	Peroxiredoxin TSA2	2	21601	LOG	3.5	
UBC1	YDR177W	Ubiquitin conjugating enzyme E2 24 kDa	7	24163	LOG	2.6	
UBIQ	YIL148W	Ubiquitin	13	8551	LOG	3.1	
UMPK	YKL024C	Uridylate kinase	13	22918	LOG	17.4	
VATA	YKL080W	V type proton ATPase catalytic subunit A	23	118562	LOG	3.3	
YB085	YBR085C	Uncharacterized protein YBR085C A	4	9398	LOG	9.0	
YGP1	YNL160W	Protein YGP1	6	37305	LOG	2.8	
YK23	YKR043C	Uncharacterized protein YKR043C	8	31002	LOG	2.0	
YM71	YMR226C	Uncharacterized oxidoreductase YMR226C	17	29140	LOG	2.1	
YO021	YOR021C	Uncharacterized protein YOR021C	10	24744	LOG	2.4	
YPR1	YDR368W	Putative reductase 1	14	34733	LOG	2.2	
ZEO1	YOL109W	Protein ZEO1	11	12581	LOG	1.8	L - 4.6
6PGD1	YHR183W	6 phosphogluconate dehydrogenase decarbox.	5	53509	STAT	40.7	
ABP1	YCR088W	Actin binding protein	1	65536	STAT	54.6	
ACON	YLR304C	Aconitate hydratase mitochondrial	16	85314	STAT	2.7	
ACT	YFL039C	Actin	6	41662	STAT	17.4	
ADH2	YMR303C	Alcohol dehydrogenase 2	12	36708	STAT	1.9	L - 5.4
ADK	YJR105W	Adenosine kinase	7	36349	STAT	2.7	L - 1.7
AGX1	YFL030W	Alanine glyoxylate aminotransferase 1	5	41880	STAT	1.9	
ALF	YKL060C	Fructose bisphosphate aldolase	13	39595	STAT	2.0	L - 4.7

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
AMPL	YKL103C	Vacuolar aminopeptidase 1	14	57057	STAT	4.4	S - 3.2
ATPA	YBL099W	ATP synthase subunit alpha mitochondrial	14	58572	STAT	11.4	
ATPB	YJR121W	ATP synthase subunit beta mitochondrial	10	54760	STAT	2.7	
BMH2	YDR099W	Protein BMH2	20	31042	STAT	1.4	
CAJ1	YER048C	Protein CAJ1	2	44826	STAT	3.8	
CCS1	YMR038C	Superoxide dismutase 1 copper chaperone	7	27312	STAT	2.8	
CISY1	YNR001C	Citrate synthase mitochondrial	11	53327	STAT	2.9	S - 2.4
DAK1	YML070W	Dihydroxyacetone kinase 1	17	62167	STAT	5.1	S - 1.9
DHE4	YOR375C	NADP specific glutamate dehydrogenase 1	16	49538	STAT	1.8	S - 2.2
DLDH	YFL018C	Dihydrolipoyl dehydrogenase mitochondrial	5	53976	STAT	3.1	
DLHH	YDL086W	Putative carboxymethylenebutenolidase	7	30817	STAT	2.3	
ECM33	YBR078W	Cell wall protein ECM33	2	43741	STAT	4.2	
EF1G1	YPL048W	Elongation factor 1 gamma 1	14	47058	STAT	2.4	
EF1G2	YKL081W	Elongation factor 1 gamma 2	5	46490	STAT	8.5	
EF2	YOR133W	Elongation factor 2	16	93230	STAT	1.5	L - 1.3
ENO1	YGR254W	Enolase 1	18	46773	STAT	1.5	S - 2.1
ESA1	YOR244W	Histone acetyltransferase ESA1	3	52579	STAT	9.6	
FHP	YGR234W	Flavohemoprotein	5	44618	STAT	52.4	
G3P1	YJL052W	Glyceraldehyde 3 phosphate dehydrogenase 1	15	35727	STAT	13.4	S - 3.0
G3P2	YJR009C	Glyceraldehyde 3 phosphate dehydrogenase 2	16	35824	STAT	3.5	
G3P3	YGR192C	Glyceraldehyde 3 phosphate dehydrogenase 3	15	35724	STAT	1.4	L - 2.2
G6PD	YNL241C	Glucose 6 phosphate 1 dehydrogenase	15	57485	STAT	1.3	S - 1.8
G6PI	YBR196C	Glucose 6 phosphate isomerase	24	61261	STAT	1.4	S - 1.4
GGA2	YHR108W	ADP ribosylation factor binding GGA2	10	64307	STAT	3.0	
GLRX2	YDR2153W	Glutaredoxin 2 mitochondrial	10	15851	STAT	1.6	
GLYC	YLR058C	Serine hydroxymethyltransferase cytosolic	8	52185	STAT	2.1	
GRE3	YHR104W	NADPH dependent aldose reductase GRE3	6	37095	STAT	6.0	
GSHR	YPL091W	Glutathione reductase	22	53407	STAT	1.4	S - 2.0
GTO2	YGR154C	Glutathione S transferase omega like 2	17	43246	STAT	4.0	
HBN1	YCL026C	Putative nitroreductase HBN1	9	20980	STAT	4.7	S - 5.4
HS104	YLL026W	Heat shock protein 104	15	101972	STAT	23.3	S - 1.7
HSP26	YBR072W	Heat shock protein 26	11	23865	STAT	1.9	
HSP71	YAL005C	Heat shock protein SSA1	30	69614	STAT	2.1	L - 1.1
HSP72	YLL024C	Heat shock protein SSA2	26	69427	STAT	2.1	L - 1.4
HSP73	YBL075C	Heat shock protein SSA3	15	70503	STAT	2.8	
HSP7F	YPL106C	Heat shock protein homolog SSE1	8	77318	STAT	2.3	
HSP82	YPL240C	ATP dependent molecular chaperone HSP82	30	81356	STAT	1.5	L - 2.4
HXKB	YGL253W	Hexokinase 2	9	53908	STAT	2.2	S - 1.3
HXKG	YCL040W	Glucokinase 1	6	55342	STAT	16.4	
IDH1	YNL037C	Isocitrate dehydrogenase NAD subunit 1 mito	11	39299	STAT	15.8	
IDH2	YOR136W	Isocitrate dehydrogenase NAD subunit 2 mito	4	39714	STAT	11.0	
IF4B	YPR163C	Eukaryotic translation initiation factor 4B	6	48493	STAT	11.0	
IMDH2	YHR216W	Inosine 5 monophosphate dehydrogenase	10	56493	STAT	2.0	
IMDH4	YML056C	Probable inosine 5 monophosphate dehydrog.	8	56357	STAT	1.5	
KAR	YNL188W	NADPH dependent alpha keto amide reduct.	8	35538	STAT	9.2	S - 2.0
KPYK1	YAL038W	Pyruvate kinase 1	11	54510	STAT	1.5	L - 1.7
LEU1	YNL104C	2 isopropylmalate synthase	9	68366	STAT	7.3	
MAS5	YNL064W	Mitochondrial protein import protein MAS5	4	44642	STAT	2.8	
MBF1	YOR298C	Multiprotein bridging factor 1	5	16393	STAT	6.6	
MDHM	YKL085W	Malate dehydrogenase mitochondrial	23	35627	STAT	2.3	S - 2.7
MDHP	YDL078C	Malate dehydrogenase peroxisomal	13	37162	STAT	17.1	
MET17	YLR303W	Protein MET17	13	48641	STAT	1.6	L - 1.6
METK1	YLR180W	S adenosylmethionine synthetase 1	8	41792	STAT	1.7	
MPG1	YDL055C	Mannose 1 phosphate guanyltransferase	2	39541	STAT	2.8	
NACB1	YPL037C	Nascent polypeptide associated complex	5	17009	STAT	2.6	
ODPB	YBR221C	Pyruvate dehydrogenase E1 component sub.	7	40028	STAT	2.0	
OMS1	YDR316W	Methyltransferase OMS1 mitochondrial	2	55554	STAT	32.7	
OYE2	YHR179W	NADPH dehydrogenase 2	13	44982	STAT	2.3	S - 2.6
OYE3	YPL171C	NADPH dehydrogenase 3	21	44892	STAT	3.1	S - 3.6

Swiss Prot Name	Ordered Locus Name	Description	Pep. Hits	Intact Mass	Up reg. in	Fold change frac.	Fold change un-frac
OYE3	YPL171C	NADPH dehydrogenase 3	21	44892	STAT	3.1	S - 3.6
PABP	YER165W	Polyadenylate binding protein cytoplasmic	9	64304	STAT	2.6	
PDC1	YLR044C	Pyruvate decarboxylase isozyme 1	24	61456	STAT	2.2	L - 1.3
PDI	YCL043C	Protein disulfide isomerase	20	58190	STAT	13.4	
PMM	YFL045C	Phosphomannomutase	6	29044	STAT	9.5	
PPID	YLR216C	Peptidyl prolyl cis trans isomerase CPR6	12	42045	STAT	8.1	
PUR2	YGL234W	Bifunctional purine biosynthetic ADE5	13	86014	STAT	3.1	
PUR92	YMR120C	Bifunctional purine biosynthesis ADE17	26	65222	STAT	1.8	S - 1.7
PURA	YNL220W	Adenylosuccinate synthetase	6	48249	STAT	103.9	S - 3.4
QCR2	YPR191W	Cytochrome b c1 complex subunit 2 mit.	11	40453	STAT	2.9	
RAD7	YJR052W	DNA repair protein RAD7	4	63737	STAT	4.1	
RIB3	YOL143C	3 4 dihydroxy 2 butanone 4 phosphate synth.	1	22553	STAT	3.4	
RIR4	YGR180C	Ribonucleoside diphosphate reductase	6	40028	STAT	2.2	
RLA0	YLR340W	60S acidic ribosomal protein P0	4	33745	STAT	4.0	
RS3	YNL178W	40S ribosomal protein S3	1	26486	STAT	3.6	
RTN1	YDR233C	Reticulon like protein 1	2	32895	STAT	2.2	
SAHH	YER043C	Adenosylhomocysteinase	4	49094	STAT	14.2	L - 1.2
SIS1	YNL007C	Protein SIS1	12	37566	STAT	1.5	
SODM	YHR008C	Superoxide dismutase Mn mitochondrial	5	25758	STAT	2.4	
SUB2	YDL084W	ATP dependent RNA helicase SUB2	4	50277	STAT	2.4	
SYLC	YPL160W	Leucyl tRNA synthetase cytoplasmic	12	124062	STAT	1.5	
TMA7	YLR262C	Translation machinery associated protein 7	2	6936	STAT	2.5	
TRXB1	YDR353W	Thioredoxin reductase 1	14	34216	STAT	2.4	S - 1.4
VTI1	YMR197C	t SNARE VTI1	3	24652	STAT	3.1	
YCP4	YCR004C	Flavoprotein like protein YCP4	4	26333	STAT	1.8	
YCZ2	YCR102C	Uncharacterized protein YCR102C	2	40096	STAT	1.8	
YMN1	YML131W	Uncharacterized membrane YML131W	13	39951	STAT	6.6	S - 5.8
YMY0	YMR090W	UPF0659 protein YMR090W	3	24866	STAT	4.4	S - 1.4
YMY9	YMR099C	UPF0010 protein YMR099C	3	33934	STAT	3.9	
YN14	YNL274C	Putative 2 hydroxyacid dehydrogenase	2	38807	STAT	32.9	
YNN4	YNL134C	Uncharacterized protein YNL134C	19	41164	STAT	1.3	S – 1.5