

Global Analysis of Protein Folding Thermodynamics for Disease State Characterization
and Biomarker Discovery

by

Jagat Adhikari

Department of Biochemistry
Duke University

Date: _____

Approved:

Michael C. Fitzgerald, Supervisor

Kenneth Kreuzer

Terrence G. Oas

Jiyong Hong

Seok-Yong Lee

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2015

ABSTRACT

Global Analysis of Protein Folding Thermodynamics for Disease State Characterization
and Biomarker Discovery

by

Jagat Adhikari

Department of Biochemistry
Duke University

Date: _____

Approved:

Michael C. Fitzgerald, Supervisor

Kenneth Kreuzer

Terrence G. Oas

Jiyong Hong

Seok-Yong Lee

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School of
Duke University

2015

Copyright by
Jagat Adhikari
2015

Abstract

Protein biomarkers can facilitate the diagnosis of many diseases such as cancer and they can be important for the development of effective therapeutic interventions. Current large-scale biomarker discovery and disease state characterization studies have largely focused on the global analysis of gene and protein expression levels, which are not directly tied to function. Moreover, functionally significant proteins with similar expression levels go undetected in the current paradigm of using gene and protein expression level analyses for protein biomarker discovery. Protein-ligand interactions play an important role in biological processes. A number of diseases such as cancer are reported to have altered protein interaction networks. Current understanding of biophysical properties and consequences of altered protein interaction network in disease state is limited due to the lack of reproducible and high-throughput methods to make such measurements. Thermodynamic stability measurements can report on a wide range of biologically significant phenomena (e.g., point mutations, post-translational modifications, and new or altered binding interactions with cellular ligands) associated with proteins in different disease states. Investigated here is the use of thermodynamic stability measurements to probe the altered interaction networks and functions of proteins in disease states. This thesis outlines the development and application of mass spectrometry based methods for making proteome-wide thermodynamic measurements

of protein stability in multifactorial complex diseases such as cancer. Initial work involved the development of SILAC-SPROX and SILAC-PP approaches for thermodynamic stability measurements in proof-of-concept studies with two test ligands, CsA and a non-hydrolyzable adenosine triphosphate (ATP) analogue, adenylyl imidodiphosphate (AMP-PNP). In these proof-of-principle studies, known direct binding target of CsA, cyclophilin A, was successfully identified and quantified. Similarly a number of known and previously unknown ATP binding proteins were also detected and quantified using these SILAC-based energetics approaches.

Subsequent studies in this thesis involved thermodynamic stability measurements of proteins in the breast cancer cell line models to differentiate disease states. Using the SILAC-SPROX, ~800 proteins were assayed for changes in their protein folding behavior in three different cell line models of breast cancer including the MCF-10A, MCF-7, and MDA-MB-231 cell lines. Approximately, 10-12% of the assayed proteins in the comparative analyses performed here exhibited differential stability in cell lysates prepared from the different cell lines. Thermodynamic profiling differences of 28 proteins identified with SILAC-SPROX strategy in MCF-10A versus MCF-7 cell line comparison were also confirmed with SILAC-PP technique. The thermodynamic analyses performed here enabled the non-tumorigenic MCF-10A breast cell line to be differentiated from the MCF-7 and MDA-MB-231 breast cancer cell lines. Differentiation of the less invasive MCF-7 breast cancer cell line from the more highly invasive MDA-

MB-231 breast cancer cell line was also possible using thermodynamic stability measurements. The differentially stabilized protein hits in these studies encompassed those with a wide range of functions and protein expression levels, and they included a significant fraction (~45%) with similar expression levels in the cell line comparisons. These proteins created novel molecular signatures to differentiate the cancer cell lines studied here. Our results suggest that protein folding and stability measurements complement the current paradigm of expression level analyses for biomarker discovery and help elucidate the molecular basis of disease.

Dedication

This work is dedicated to my beloved parents who nurtured me to be who I am.

Contents

Abstract	iv
List of Tables	xiii
List of Figures	xv
List of Abbreviations	xx
Acknowledgements	xxii
1. Background and Significance	1
1.1 Global Disease Burden and Early Diagnosis	1
1.2 Biomarker Discovery and Disease Characterization Studies	2
1.3 Current Strategies for Global Disease State Characterization.....	4
1.3.1 Gene Expression Profiling.....	4
1.3.2 Protein Expression Profiling	5
1.3.2.1 Protein Microarrays	6
1.3.2.2 Mass Spectrometry based Protein Expression Analysis.....	7
1.3.3 Metabolomics	9
1.4 Strategies beyond Expression Levels for Biomarker Analysis.....	10
1.4.1 Protein Interactome	10
1.4.2 Activity Based Protein Profiling.....	11
1.5 Motivation and Significance of the Present Study	12
1.5.1 Protein Folding Thermodynamics in Disease states	12
1.5.2 Approaches for Protein Folding and Thermodynamics Measurements	13

1.5.3 Stability of Proteins from Rates of Oxidation (SPROX)	17
1.5.3.1 SPROX with Colorectal Cancer Tissue Lysates	19
1.5.4 Pulse Proteolysis.....	21
1.6 SILAC-based Quantitation Strategy	22
1.7 Cell Line Models of Breast Cancer	24
2. Proteomic Characterization of Therapeutic Response in Patient-Derived Tumor Explants of Colorectal Cancer in Murine Models.	26
2.1 Introduction.....	26
2.2 Experimental Procedures	29
2.2.1 Comparative Analyses Schematic.....	29
2.2.2 Tissue Lysate Preparation	30
2.2.3 iTRAQ-SPROX Protocol	32
2.2.4 Methionine-containing Peptide Enrichment	33
2.2.5 LC-MS/MS Analysis.....	34
2.2.6 iTRAQ-SPROX Data Analysis	35
2.3 Results and Discussion	38
2.3.1 General Strategy	38
2.3.2 Global Thermodynamic Profiles of Proteins in the Lysate from CRC Explants	39
2.3.2.1 Oxaliplatin Treated versus Not Treated Samples SPROX Analysis	39
2.3.3.2 Oxaliplatin Resistant versus Sensitive CRC Tumors	43
2.4 Conclusions	47
3. The Development of SILAC-SPROX in Solution Protocol	48

3.1 Introduction.....	48
3.2 Experimental	52
3.2.1 Yeast Cell Culture.....	52
3.2.2 Yeast Cell Lysate Preparation.....	53
3.2.3 SILAC-SPROX Analysis	54
3.2.4 Proteomic Sample Preparation.....	56
3.2.5 LC-MS/MS Analyses	57
3.2.5.1 Instrument Parameters.....	57
3.2.5.2 Spectrum Mill Parameters	58
3.3 Quantitative Proteomic Data Analysis.....	59
3.4 Results and Discussion	62
3.4.1 General Strategy.	62
3.4.2 Identification of Cyclosporine A Binding Molecular Targets.	65
3.4.3 False Positive Rate Determination	68
3.5 Conclusions	69
4. The Development of a SILAC-Pulse Proteolysis Approach	71
4.1 Introduction.....	71
4.2 Experimental	74
4.2.1 Yeast Cell Lysate Preparation.....	74
4.2.2 SILAC-Pulse Proteolysis Analysis	76
4.2.3 LC-MS/MS Analyses	79
4.3 Data Analysis	81

4.3.1	Quantitation of Ligand Binding	85
4.4	Results and Discussion	86
4.4.1	SILAC-Pulse Proteolysis Protocol	86
4.4.2	Control Experiment: False Positive Rate Determination	90
4.4.3	SILAC-PP: A Tool for Protein Target Discovery	93
4.4.4	SILAC-PP: A Tool for Protein-Target Validation	101
4.5	Conclusions	106
5.	Global Thermodynamic Profiling of Proteins in Breast Cancer Cell lines	108
5.1	Introduction	108
5.2	Experimental	111
5.2.1	Cell Culture and Cell Lysate Preparations	111
5.2.2	SILAC-SPROX Analyses	113
5.2.3	SILAC-Pulse Proteolysis Analysis	115
5.2.4	Proteomic Sample Preparation	117
5.2.5	LC-MS/MS Analyses of Proteomic Samples	118
5.2.6	Proteomic Data Analysis	120
5.2.7	Protease Activity Assays	124
5.3	Results and Discussion	125
5.3.1	Global Analysis of Protein Folding Thermodynamics using SILAC-SPROX Strategy	125
5.3.2	Hit Proteins Map to a Wide Range of Protein Expression Levels	130
5.3.3	Protein Signatures for Disease State Differentiation	132

5.3.4 Classification of Proteins with Altered Thermodynamic Stability.....	139
5.3.5 Cross-Validation and Discovery using SILAC-PP Protocol.....	142
5.3.6 Correlating Changes in Thermodynamic Stability with Changes in Function	147
5.4 Conclusion.....	149
6. Conclusions and Future directions.....	150
Appendix A: SILAC-Pulse Proteolysis Protocol.....	155
A.1 CsA-Binding Experimental Parameters	155
A.2 ATP-Binding Experimental Parameters.....	156
A.3 Control Experiment Experimental Parameters	157
A.4 SDS-PAGE Gels Generated in the SILAC-PP Experiments (A-E).	158
Appendix B: Disease State Analysis with Breast Cancer Cell Lines.....	161
B.1 Peptides and Proteins with Altered Thermodynamics in MCF-7 vs MCF-10A..	161
B.2 Peptides and Proteins with Altered Thermodynamics in MCF-7 vs MDA-MB-231	170
B.3 Bioinformatics Analyses of Peptides and Proteins with Altered Thermodynamics	177
B.4 Peptides and Proteins with Altered Thermodynamics in MCF-7 versus MCF-10A using SILAC-PP.	182
References	185
Biography	207

List of Tables

Table 1: Proteomic coverage for the oxaliplatin treated (OX) and not treated (NT) experiment with iTRAQ-SPROX protocol in the first comparative analyses.....	39
Table 2: The peptides and proteins identified with altered thermodynamic stability between not treated (NT) and oxaliplatin treated (OX) CRC samples.....	42
Table 3: Proteomic coverage for the oxaliplatin treated sensitive and resistant experiment with iTRAQ-SPROX protocol in the second comparative analyses.....	43
Table 4: The peptides and proteins identified with altered thermodynamic stability between sensitive and resistant CRC samples.....	44
Table 5: Proteomic results summary in the Control and CsA-binding experiments with the proteins in a yeast cell lysate.....	65
Table 6: Peptides and proteins identified to interact with CsA.....	66
Table 7: Summary of the peptide and protein coverage with the protein hits observed in the SILAC-Pulse-proteolysis experiments performed in this work.....	92
Table 8: False Positives from SILAC-PP Control Experiment.	93
Table 9: ATP-binding proteins identified in SILAC-PP experiments.....	97
Table 10: CsA-binding proteins identified using SILAC-PP protocol.....	100
Table 11: Protein targets with previously unknown ATP binding properties that were identified and cross-validated with SILAC-PP and SILAC-SPROX.....	104
Table 12: Proteomic Coverage of peptides and proteins assayed for thermodynamic stability changes in the MCF-7 Vs MCF-10A and MCF-7 vs MDA-MB-231 cell line comparisons.....	127
Table 13: Protein hits detected in both of the cell line comparisons, the MCF-7 versus MCF-10A, and the MCF-7 versus MDA-MB-231.	134
Table 14: Proteomic Coverage obtained in the MCF-7 versus MCF-10A comparative study using SILAC-PP protocol.....	143

Table 15: Proteins with the altered thermodynamics in MCF-7 versus MCF-10A comparison identified in both SILAC-PP and SILAC-SPROX experiments..... 144

List of Figures

- Figure 1: Summary of the comparative analyses performed using iTRAQ-SPROX strategies on tumor tissue lysate derived from colorectal cancer patient explants. NT and OX represent not treated and oxaliplatin treated mice respectively. 31
- Figure 2: General scheme of the iTRAQ-SPROX protocol(63, 90). 37
- Figure 3: Distribution of the normalized reporter ion intensities for the lowest denaturant, 113 tag (filled squares) and highest denaturant, 121 tag (filled circles) in each of the three experiments. Shown are the distribution for not treated (NT) on the left and oxaliplatin treated (OX) on the right. For each distributions, experiment 1 (solid red lines), experiment 2 (dashed blue lines) and experiment 3 (dotted black lines) are shown. 40
- Figure 4: Representative iTRAQ-SPROX data sets from two proteins that showed difference in thermodynamic stability between oxaliplatin resistant (blue bars) and sensitive (red bars) CRC samples. Shown are the data sets for Pyruvate kinase isozymes M1/M2 (FGVEQDVDMVFASFIR) (A) and Transgelin-2 (QMEQISQFLQAAER) (B). In each case, the red and blue arrows represent the $C_{1/2}$ values for the peptide in sensitive and resistant samples. The dotted line represents the transition base line value for each of the experiments. 46
- Figure 5: Global analysis of the L/H ratios observed in the Control and CsA-Binding experiments. Shown are the data for all the identified peptides that were successfully quantified in the Control Experiment (A), CsA Binding Experiment 1 (B), and CsA Binding Experiment 2 (C). The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value. 61
- Figure 6: The SILAC-SPROX protocol. 62
- Figure 7: Expected Result (A) where there is no interaction between the protein and the ligand, SPROX behavior of the protein in the presence (solid line) and absence (dotted line) of the ligand resulting a straight line (B) if the ligand stabilize the proteins, the SPROX curve is shifted to the right resulting in a dip for the oxidized Met peptides and a peak for the un-oxidized Met peptide. 64
- Figure 8: Representative SILAC-SPROX data obtained on proteins identified in the CsA-binding experiments. Shown are results from CsA-binding experiment 2 (A) non-hit

protein, Glyceraldehyde-3-phosphate dehydrogenase 3 peptides, YAGEVSHDDKHIIVDGK (+2) non-methionine (closed circles), the oxidized and non-oxidized versions (+3 charge state) of a methionine-containing peptide VINDAFGIEEGLMTTVHSLTATQK (open and closed squares, respectively), (B) known CsA-binding protein, Cyclophilin A peptides, including the +3 charge state of the non-methionine containing peptide HVVFGEVVDGYDIVK (closed circles) and the oxidized and non-oxidized versions (albeit the +2 and +3 charge states, respectively) of a methionine-containing peptide VIPDFMLQGGDFTAGNGTGGK (open and closed squares, respectively). 67

Figure 9: Global analysis of the L/H ratios determined for peptides identified and successfully quantified in (A) Control Experiment 1, and (B) Control Experiment 2, (C) the ATP Experiment 1 and (D) the ATP Experiment 2. The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value. The median $\text{Log}_2(\text{L}/\text{H})$ values in (A) and (B) were 0.03 and -0.29, respectively. The median $\text{Log}_2(\text{L}/\text{H})$ values in (C) and (D) were -0.11 and -0.02, respectively. 83

Figure 10: Global analysis of the L/H ratios observed in the CsA experiments in SILAC-PP. Shown are the distributions of the $\text{Log}_2(\text{L}/\text{H})$ for all the identified peptides that were successfully quantified in (A) the CsA Experiment 1, (B) the CsA Experiment 2, (C) the CsA Experiment 3, and (D) the CsA Experiment 4. The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value. The median $\text{Log}_2(\text{L}/\text{H})$ values in (A), (B), (C), and (D) were 0.04, 0.23, 0.67, and -0.11, respectively..... 84

Figure 11: SILAC-pulse proteolysis protocol. 87

Figure 12: Expected Result for the intact protein readout (A) No interaction between the protein and the ligand, behavior of the protein in the presence (solid line) and absence (dotted line) of the ligand resulting a straight line (B) ligand induced stabilization of the proteins causes the pulse proteolysis curve to shift to the right resulting in a dip and (C) ligand induced destabilization that shifts the curve to the left resulting in a peak. 89

Figure 13: SDS-gel image for SILAC-PP control experiment 2. The gel regions excised in this experiment are indicated with rectangular boxes..... 91

Figure 14: SDS-PAGE gel image ATP-binding experiment 2. The gel regions excised in this experiment are indicated with rectangular boxes..... 94

Figure 15: SILAC-PP data for the ATP-dependent heat shock protein, HSC82p, obtained from intact protein band, experiment 1 (solid squares) and experiment 2 (solid circles). Also shown are the data from control experiment 1 (open squares) and experiment 2 (open circles). The $\Delta C^{1/2}$ values extracted was 2.5 and 2.6 M for ATP-binding experiment 1 and 2 respectively..... 96

Figure 16: SDS-gel image for CsA-binding experiment 2. The gel regions excised in this experiment are indicated with rectangular boxes. 98

Figure 17: SILAC-PP data for the Cyclophilin A, CypA, obtained from intact protein band, solid green, red, black and blue squares in CsA-binding experiments 1, 2, 3 and 4 respectively. Empty and solid circles represent data on the intact CypA protein from control experiments 1 and 2 respectively. The $\Delta C^{1/2}$ values extracted were 4, 4, 3.4, and 4.8 M for the CsA-binding experiment 1, 2, 3 and 4 respectively. 99

Figure 18: SILAC-PP data for the Phosphoglycerate Mutase, GPM1, obtained from intact protein band in ATP-binding experiment 1 (solid squares) and experiment 2 (solid circles) respectively. Also shown are the intact protein data for GPM1 from control experiment 1 (empty squares) and experiment 2 (empty circles). The $\Delta C^{1/2}$ values extracted was >1.5 M and 2.6 M for ATP-binding experiments 1 and 2 respectively. 102

Figure 19: SILAC-PP data for the Pyruvate Kinase, obtained from intact protein band, solid green, red, black and blue squares in CsA-binding experiments 1, 2, 3 and 4 respectively and empty and solid circles represent data on the intact protein from control experiments 1 and 2 respectively. 105

Figure 20: SDS-PAGE gel image for the MCF-7 versus MCF-10A comparative study using SILAC-PP strategy. The rectangular boxes in the gel mark the gel bands excised for the in-gel digestion. 116

Figure 21: Global distribution of H/L ratios for all the non-methionine containing peptides identified in each experiments using SILAC-SPROX in (A) MCF-7 versus MCF-10A and (B) MCF-7 versus MDA-MB-231. Shown are the data from biological replicates 1 (dotted line), 2 (dashed line) and 3 (solid line) in each comparative studies. The arrows mark 5th and 95th percentile values in each experiment..... 122

Figure 22: Schematic representation of the SILAC-SPROX experimental workflow used in this study. 126

Figure 23: Representative SILAC-SPROX data obtained in the two cell line comparisons in this work. (A) Data obtained on peptides from calpain small subunit 1 in the MCF-7 versus MCF-10A comparison including the wild-type (filled circles) and doubly oxidized (filled squares) forms of the methionine-containing peptide

SMVAVMDSDTTGK (detected in biological replicates 3 and 2, respectively), (B) Data obtained on peptides from splicing factor 3B subunit 3 in biological replicate 2 of the MCF-7 versus MCF-10A comparison including the wild-type form of the methionine-containing peptide AVMISAIEK (filled circles), (C) Data obtained on peptides from peptidyl-prolyl cis-trans isomerase protein in biological replicate 3 of the MCF7 versus MDA-MB-231 comparison including the oxidized form of the methionine-containing peptide (ac)MVNPTVFFDIAVDGEPLGR (closed squares). In each case the median data for all the non-methionine-containing peptides from respective proteins are also shown (open squares). Shown on the bottom panels are schematic representations of the expected unfolding curves that produced the observed SILAC-SPROX behavior in each example. 129

Figure 24: Subset of protein hits detected in the comparative analyses performed here. Shown in (A), (B), and (C) are the gene names of proteins that were detected as hits in both cell line comparisons, only the MCF-7 versus MCF-10A comparison, and only the MCF-7 versus MDA-MB-231 comparison, respectively. The numbers in the boxes represent the protein expression level data (i.e., $\text{Log}_2(\text{H/L})$ values) either measured in this work or previously reported in the literature for each protein hits. The literature values are from reference (12). The protein hits highlighted here are those that were assayed with the same peptide probe in both cell line comparisons..... 132

Figure 25: Representative SILAC-SPROX behaviors of four MCF-7 specific proteins. Shown are the data in the MCF-7 versus MDA-MB-231 cell line comparison (empty squares) and the MCF-7 versus MCF-10A cell line comparison (filled circles) from biological replicate 3 and biological replicate 2 respectively. The solid and dotted lines in each plot indicate the H/L ratio determined for each protein in the MCF-7 versus MDA-MB-231 and MCF-7 vs MCF-10A cell line comparisons, respectively..... 136

Figure 26: The SILAC-SPROX plots of the HEAM(ox)ITDLEER peptide from myosin-9 in the MCF-7 versus MDA-MB-231 (empty squares) and the MCF-7 versus MCF-10A (filled circles) cell line comparisons from biological replicate 3 and biological replicate 2 respectively. The solid and dotted lines in each plot indicate the H/L ratio determined for each protein in the MCF-7 versus MDA-MB-231 and MCF-7 vs MCF-10A cell line comparisons, respectively..... 138

Figure 27: Heat map showing the distribution of protein classes observed in the protein hits identified from the thermodynamic and protein expression level analyses performed in this work. The numbers in the boxes represent the percentage of protein hits from each class in each experiment..... 140

Figure 28: SILAC-PP workflow for MCF-7 versus MCF-10A comparison.....	142
Figure 29: Data confirming the altered thermodynamic stability for the protein, Glyceraldehyde-3-phosphate dehydrogenase, using (A) SILAC-PP and (B) SILAC-SPROX. The data in (A) is for the intact protein and (B) is from the un-oxidized peptide VIISAPSADAPMFVMGVNHEK.....	145
Figure 30: Data showing the altered thermodynamic stability behavior of the catalytic subunit from (A) calpain-1 and (B) calpain-2 obtained from 50-75kDa gel band region of MCF-7 versus MCF-10A SILAC-PP experiment.....	146
Figure 31: Catalytic activity data for calpain protease in MCF-7 lysate and MCF-10A lysate (A) and for cathepsin D protease in MCF-7, MCF-10A, and MDA-MB-231 cell lysates (B). Error bars represent +/- one standard deviation from duplicate measurements.....	147

List of Abbreviations

SPROX	Stability of Proteins from Rates of Oxidation
PP	Pulse Proteolysis
SILAC	Stable Isotope Labeling by Amino acids in Cell culture
SILAC-SPROX	SILAC quantitation coupled with SPROX reaction
SILAC-PP	SILAC quantitation coupled with Pulse Proteolysis reaction
CsA	Cyclosporine A
MCF	Michigan Cancer Foundation
H/D	Hydrogen/Deuterium
SUPREX	Stability of Unpurified Proteins from Rates of H/D Exchange
LC-MS	Liquid Chromatography Mass Spectrometry
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
2D	Two Dimensional
SDS PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TCA	Trichloroacetic Acid
TEAB	Triethyl Ammonium Bicarbonate
TCEP-HCl	Tris(2-carboxyethyl)phosphine Hydrochloride
MMTS	Methyl Methane Thiosulfonate
Q-TOF	Quadrupole Time of Flight

MS	Mass Spectrometry
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
TMT	Tandem Mass Tags
EDTA	Ethylenediaminetetraacetic acid
GdmCl	Guanidinium chloride
iTRAQ-SPROX	iTRAQ quantitation coupled with SPROX reaction
CRC	Colorectal Cancer
ETD	Electron-transfer dissociation
HCD	Higher energy collision dissociation
FDR	False Discovery Rate
$C^{1/2}_{\text{SPROX}}$	Midpoint of transition in SPROX
ACN	Acetonitrile
HPLC	High performance liquid chromatography
UPLC	Ultra-High performance liquid chromatography
MCF-7	Estrogen positive breast cancer cell line
MCF-10A	Normal control breast cell line
MDA-MB-231	Estrogen negative breast cancer cell line

Acknowledgements

Graduate school has provided me the opportunity to work and explore the field of scientific research. My experiences throughout the PhD journey have augmented my passion for learning and made me more mature both in terms of scientific knowledge and personal growth. I am indebted to my advisor, Professor Michael C. Fitzgerald for guidance, motivation, optimism, and friendship. This thesis would not have been possible without his mentorship, confidence, and encouragement.

I would also like to thank the members of my committee Professors Kenneth Kreuzer, Terry G. Oas, Jiyong Hong, and Seok-Yong Lee for offering valuable advice, monitoring my progress and guiding my research in the right direction.

I am indebted to our collaborator Professor David Hsu and former member of his laboratory Dr. Joshua Uronis for providing the tissue cell lysates for experiments with CRC samples. Also, I am thankful to members of Professor Edward Patz's laboratory specially Dr. Michael J. Campa for his help during cell line cultures.

I would also like to acknowledge both current and past Fitzgerald lab members. In particular, I am grateful to past members particularly Dr. Duc T. Tran for her help and guidance in the lab and also for collaboration in successful development of SILAC-SPROX strategy, Dr. Erin Strickland for initial days of mentoring in the lab and Dr. Graham M. West for help with running the samples from disease state analysis. Additionally, Ariel Geer, it has been great working with you from the beginning, and

thank you for your friendship and support during our years together in the lab. I am thankful to all the present members Yingrong Xu, Julia H. Roberts, Ryenne N. Ogburn, Lorrain Xin, Fang Liu and He Meng. Thank you for making my stay in the laboratory so much more than a scientific exercise.

My family has been incredibly supportive throughout my graduate years as well. I am thankful to my parents, my siblings Sarita, Shova, Namita, Bijaya, and my wife Ritu for all the love, support and encouragement. I would also like to thank my friends for the support, encouragement and insightful discussions that we shared during my stay here.

1. Background and Significance

1.1 Global Disease Burden and Early Diagnosis

The world health organization (WHO) reported 8.2 million deaths due to cancer in the year 2012 worldwide(1). In the US alone, it is estimated that over 580,000 people will succumb to cancer-related deaths in 2015(2). In recent years, the global trend in mortality has shifted from communicable infectious diseases to non-communicable and complex diseases such as cardiovascular diseases, cancer and neurodegenerative diseases (Alzheimer, Parkinson, amyotrophic lateral sclerosis and others)(3). The incidence and mortality rates for complex diseases such as cancer over the last couple of decades have not decreased despite tremendous research efforts that have been fueled by recent advances in molecular biology, genomics, and proteomics(4). The global health burden in terms of quality of life and economic cost is estimated to grow even more in the years to come. Thus, there is an on-going need for novel methods to identify diseases at an early stage and reduce the global disease burden.

The concept of early detection and targeted treatment strategies has attracted scientists and medical experts globally for decades(4). The completion of the human genome project in 2001 and subsequent sequencing of a number of other pathogens marked the successes of the genomic era. Subsequently, genomic studies led to novel and exciting research avenues in the pursuit of early diagnosis and personalized

medicine to deal with complex diseases. Indeed, early detection and therapeutic intervention dramatically increases the overall survival rate and reduces complications associated with late stage treatments in most ailments associated with cancer and aging related diseases. For instance changes in the brains of Alzheimer's patient are believed to start as early as 20 years before the advanced symptoms appear(5). Similarly, there is a more than 90% five-year-survival-rate for patients diagnosed with breast cancer at an early stage(6). Unfortunately, because there is a lack of proper early detection methods, 85% of the lung cancer, 60% of the colorectal, and 39% of the breast cancer patients in the US are diagnosed at late stages making medical intervention very challenging(7). Early detection thus has the potential of not only saving lives but also ameliorating the sufferings of patients for better quality of life. This has prompted research efforts directed towards comprehensive understanding of human genome, transcriptome and more recently proteome and metabolome to better understand disease pathogenesis for early diagnosis and to create novel biomarkers and therapeutic targets.

1.2 Biomarker Discovery and Disease Characterization Studies

Biomarkers are unique molecules such as DNA, mRNA, proteins, metabolites or other molecular features that act as signatures in a disease state(8). Molecular profiling is an important step for characterization and management of complex diseases that show multifactorial causes. High throughput gene and protein expression level profiling

strategies using microarrays and mass spectrometry are widely used in the current paradigm of disease state characterization efforts(9-13). The goal of biomarker discovery effort is to identify molecular signatures with high sensitivity and specificity so as to apply preventative and curative measures. But very few biomarkers reported so far have these properties required for large-scale screening and are approved for clinical application(14). Prostate specific antigen (PSA) is an example of one of the few biomarkers that shows required specificity and sensitivity for population screening at a relatively early stage(15). The human epidermal growth receptor 2 (HER2) has shown prognostic capabilities in breast cancer therapy(16). The trastuzumab antibody drug targets the HER2 receptor in a cancer patient and has exhibited great response in metastatic HER2 positive breast cancer patients(17). The large majority of other reported markers have failed to translate into useful clinical markers. Therefore a number of strategies have been employed for specific biomarker discoveries and evaluation of treatment efficiencies. These mainly involve gene and protein expression profiling studies to understand the differences between a normal and a disease state. Some of these prominent strategies for disease state characterization and biomarker discovery are discussed below.

1.3 Current Strategies for Global Disease State Characterization

1.3.1 Gene Expression Profiling

DNA microarrays have played a large role in expanding the scope of gene expression profiling in translational research(18). DNA microarrays allow DNA and/or RNA hybridization analysis to be performed in high-throughput assays. In a typical gene-profiling analysis, the DNA or RNA is isolated from tissues or cell samples and fluorescently labeled for quantitation. These labeled transcripts are hybridized to a specific probe (cDNA, and oligonucleotides) and quantified in a genome wide analysis(19). The Serial Analysis of Gene Expression (SAGE) technique is a variant of the DNA microarray that allows identification and quantitation of both known and novel transcripts(20). Disease-associated changes are studied by comparing the expression pattern of certain genes in a normal state as compared to a diseased one. A number of studies have utilized microarray technology for studying the aging process in mice, aging related diseases in humans, and a number of cancers(21-24). These large scale gene-expression profiling analysis have helped uncover potential diagnostic, prognostic and therapeutic targets in different diseases(25-27).

A large number of the gene expression profiling based studies have focused on understanding cancer related changes. Gene expression profiling has not entered into routine clinical application despite studies reporting potential diagnostic and predictive

signatures. One of the challenges associated with clinical performance of these reported markers from gene profiling studies has been the issue of reproducibility in clinical settings. The panel of markers reported with microarray display a weak correlation to prognostic outcome in many cases(28). This weak correlation suggests that it is important to characterize the phenotypic effects of genetic alterations identified in different biological states by gene expression level studies to better understand the biology of pathogenesis. Nevertheless, gene expression profiling has provided insights into molecular basis of a number of diseases and shows the potential of providing complementary information when applied in tandem with other large-scale analysis of diseases.

1.3.2 Protein Expression Profiling

Proteins are the functional components within the organisms. Posttranslational modifications such as phosphorylation, acetylation, methylation and glycosylation impact the functions of proteins to regulate cell differentiation and/or proliferation. DNA microarray techniques for gene expression level analysis can only provide a partial picture of biological processes as they do not correlate well with the protein expression levels due to posttranscriptional and posttranslational modifications events that regulate protein abundance and activity(29). A better correlation can be achieved by measuring protein expression level changes associated with functional alteration of proteins in

different biological states and in diseased pathways. Thus measuring protein expression level is expected to reflect the actual state of the cell. Two approaches are commonly in use for global analysis of protein expression level in different biological states: protein microarrays and mass spectrometry based approaches.

1.3.2.1 Protein Microarrays

The concept and design of protein microarrays are similar to the DNA microarrays(30). The protein arrays consist of immobilized protein baits that include an antibody, a cell lysate or a recombinant protein or peptide in the spots. The detection of the target proteins in the array is achieved by the use of tagged antibody, ligand or cell lysate probe. Finally, the amount of the tagged molecule bound to the bait molecule is measured to quantitate the signal intensity. Mainly two formats of protein arrays, forward-phase protein microarrays (FPPM)(31), and reverse-phase protein microarrays (RPPM)(32) are available; the former of the two is more widely used. As the name suggests, the array spot in the FPPM contains antibody or similar probe that is used to capture the target protein from a cell lysate or other samples whereas the reverse is true for the RPPM. In RPPM the cell lysate or serum sample is plated onto the microarray slide and an antibody probe is used to detect the expression level of the target protein. These protein microarrays have been used in biomarker discovery and validation and in disease diagnostics(11, 33). Protein microarrays have been used to examine protein

expression levels and subsequent identification of biomarker candidates in cancer patients(34). One of the drawbacks of the protein microarray technique has been the potential cross-reactivity of the antibodies used in these analyses leading to non-specific binding(35). Therefore, despite tremendous progress in the protein microarray field there still is a huge challenge finding biomarker that is applicable in clinical setting.

1.3.2.2 Mass Spectrometry based Protein Expression Analysis

Mass spectrometry-based techniques for protein expression level analyses have attracted scientist working in the field of biomarker discovery as it allowed high throughput identification and quantitation of proteins in complex mixtures. Traditionally, the proteins were run on a two-dimensional gel electrophoresis (2DE), separated based on the isoelectric point, molecular mass, solubility and relative abundance, and spots with differentially expressed proteins were excised and identified using mass spectrometry(36). The 2DE technique is biased towards highly abundant proteins masking the low abundant proteins(37). While it is very cost effective way to measure protein abundance, it mostly excludes analysis of proteins highly basic or acidic and it required high levels of expertise to obtain reproducible gels. Recent improvements on 2DE and the development of two-dimensional fluorescence-difference gel electrophoresis (2D-DIGE) helped to address some of the shortcomings associated

with the reproducibility of 2D gel results but still lacks the desired proteomic coverage and quantitative abilities needed for high-throughput proteomic analysis(37).

Introduction of several chemical and metabolic stable isotope tagging strategy with high throughput mass spectrometry analysis made quantitation of the proteins in biological states more reliable(38-40). This high throughput quantitative protein identification capability of mass spectrometry has been widely utilized in disease state analysis. More recently, directed and targeted proteomic strategies and the use of online and offline fractionation strategies coupled to mass spectrometry have dramatically improved the sensitivity of the technique. These technical advances in proteomic protocols enable a more robust large scale protein expression level analysis in a number of diseased proteomes. Thus, a large number of candidate protein markers have been identified using mass spectrometry based expression level profiling assays(12, 41). Additionally, mass spectrometry-based protein expression level profiling methods have also been used to identify novel potential markers of prognostic and pharmacological importance(12). These examples highlight the potential of proteome wide expression level analyses for the detection, classification and progression profiling of complex disease states. However, despite the large number of reported candidate markers that have come out of genomic and proteomic studies, only a very few have been approved by the US Food and Drug Administration(42). This discrepancy between identified

markers and the approved candidates emphasize the need to probe more biologically significant and functionally relevant dimension of the proteome to obtain robust molecular targets that can perform better in clinical validation studies.

1.3.3 Metabolomics

In the pursuit of systems level understanding of disease states, metabolomics has also evolved as a valuable tool in recent years(43). Metabolomics involves quantitative detection of small molecule metabolites in body fluids (i.e., cell, tissue, organs or biological fluids) for diagnosis, prognosis and therapeutic development. Like genomics, and proteomics a recurring theme in metabolomics is to develop the approach amenable for identification of biomarkers for early diagnosis and treatment of human diseases. Urine and blood are the two most routinely studied bio-fluids considered to be the mine for metabolites and thus for metabolite based biomarkers. Global analyses of metabolites have utilized nuclear magnetic resonance (NMR) and mass spectrometry (MS) based approaches in understanding metabolite status(43, 44). Metabolic profiling has been applied to explore biomarkers for breast, lung, pancreatic and a number of other cancers(45). Similarly investigations of metabolite status for cardiovascular disease and diabetes have also been performed. It is an attractive methodology in disease profiling as it utilizes samples (biological fluids such as urine and blood) that can be acquired through non-invasive or minimally invasive routes. As with other 'omics

approaches, a number of biomarkers have been identified and reported as significant in many diseases using metabolomics which are yet to realize true potential as clinical markers.

1.4 Strategies beyond Expression Levels for Biomarker Analysis

1.4.1 Protein Interactome

The biomarker discovery strategies shifted from genes to proteins and from one specific molecule to a panel of multiple biomarkers. We now know that it is not only about expression levels of genes and proteins; instead the emerging phenotypes in complex diseases are dependent on the dynamics of interaction between proteins (i.e., protein-protein interactions). A full understanding of the molecular basis of disease progression requires knowledge of protein-protein and protein-ligand interaction networks. These protein interaction datasets provide information about both direct physical interactions between proteins and indirect functional interactions in a biological state. Comprehensive mapping of the interaction networks in healthy and disease processes will greatly enhance effective diagnosis and prognosis of diseases. Current methods for large scale and high throughput analyses to detect protein-protein interactions include the yeast two-hybrid assay(46, 47) and protein complex purification techniques coupled with mass spectrometry (48, 49). However, drawbacks such as limitation to binary interactions in yeast two-hybrid assay(50, 51), high false positives

and false negatives(52, 53) and lack of quantitative information(54, 55) in both yeast two-hybrid and affinity capture mass spectrometry complicate their application in disease related changes. Thus, there are many significant challenges that still remain to develop accurate models of protein interaction maps in both healthy and diseased states.

1.4.2 Activity Based Protein Profiling

Proteins in a disease state often exhibit changes in activities(56). Often these changes are the consequence of posttranslational and reversible modifications that regulate different functions *in vivo*(57). The protein expression levels only provide an indirect estimate of protein activity and function. An effort towards portraying a more direct functional state of the protein has been achieved through Activity-Based Protein Profiling (ABPP) technique(58). Initial ABPP experiments employed gel based analytical methods to detect difference in catalytic activity of enzymes in different biological states. However, this reliance on gel-based strategies for protein resolution, detection and quantitation, ultimately limited the complexity of protein samples that could be interrogated in ABPP methodology. Currently, ABPP methodology has been interfaced with multidimensional protein identification using mass spectrometry platforms and takes advantage of the capabilities of modern LC-MS/MS systems to resolve, detect and quantify the protein components of complex biological mixtures(58). Jessani, N. et al used ABPP technique for global analysis of the activity, subcellular localization and

glycosylation state for the serine hydrolase superfamily across normal and disease tissue(56). The authors used ABPP approach to identify post-translationally regulated protein, active urokinase, in their assay of a human breast and melanoma cell line. They further identified a cluster of proteases, lipases, and esterases to distinguish cancer cell lines based on the tissue of origin. This study illustrated the potential of the activity-based approach in normal and disease state analysis that goes beyond the current expression based analysis. Nevertheless, it requires specific probe for a class or family of enzymes or proteins and relies on the analysis of one specific class of enzymes or proteins at a time. This underscores the need for novel and complementary approaches beyond expression level profiling in disease state analysis that are general irrespective of protein class or family.

1.5 Motivation and Significance of the Present Study

1.5.1 Protein Folding Thermodynamics in Disease states

Protein folding and thermodynamic stability measurements can provide novel information for detecting changes in disease states. Thermodynamic stability of protein plays an important role in the regulation and function of many cellular processes. Mutations, post-translational modifications, and altered binding interactions change the thermodynamic stability of a protein leading to detrimental consequences. Mutations in tumor suppressor protein 53 (p53) are responsible for the inactivation of the tumor

suppressor pathways in approximately fifty percent of human cancers(59). A third of these mutations inactivating the p53 protein have been shown to change the thermodynamic stability of the protein(59, 60). Similarly, post-translational modifications of proteins have also been implicated in many human disorders such as rheumatoid arthritis and Alzheimer's disease(61). Glutathionylation of (Cu/Zn) superoxide dismutase provides an example of a post-translational modulation of thermodynamic stability with potential link to amyotrophic lateral sclerosis(62). These thermodynamic analyses of disease-related proteins performed, to date, have largely involved purified protein systems that were already found to be associated with disease states. However, such thermodynamic analyses have not been used for the global analysis of disease states. This is in large part because methods for making protein folding and thermodynamic stability measurements on the proteomic scale have only recently been developed(63-73).

1.5.2 Approaches for Protein Folding and Thermodynamics Measurements

The calorimetric(54, 74, 75) and spectroscopic(55) techniques have been traditionally used for making quantitative thermodynamic measurements of protein folding and ligand binding. Aromatic residues such as tryptophan or tyrosine have intrinsic fluorescence properties. Fluorescence spectroscopy and circular dichroism are the most frequently reported spectroscopic approaches utilized to measure protein

folding and thermodynamics measurements. The fluorescence properties of tryptophan and tyrosine are sensitive to changes in the folding/unfolding equilibrium of proteins in solution. Such changes can be measured as a function of pH, temperature or chemical denaturation unfolding conditions by measuring fluorescence intensity of tryptophan and tyrosine probes. Often these methods require mutagenesis to incorporate the tryptophan or tyrosine residues in the absence of a native fluorophore in the proteins of interest. Such mutagenesis can perturb the structure of the protein and change observed native stability. Circular dichroism (CD) spectroscopy monitors the difference in absorption of the left and right circularly polarized light by the chromophores. It measures how secondary structure changes with the change in the environment such as chemical denaturant, pH or temperature. The intrinsic backbone circular dichroism can be monitored with the far UV (180-260 nm wavelengths) and the CD bands from aromatic chromophores can be followed using near UV (250-350 nm wavelengths) measurements. The calorimetric approach involves thermal denaturation of proteins with increasing temperature and measured with isothermal calorimetry (ITC) or differential scanning calorimetry (DSC). This method measures the change in heat capacity of proteins in the folded and unfolded states. Both the spectroscopic and calorimetric approaches have experimental limitations (e.g., requirements of large amounts of purified proteins) that preclude their use in proteome-wide studies.

Several mass spectrometry-based techniques utilizing amide H/D exchange and various oxidative foot-printing assays have also been established for the characterization of protein folding and ligand binding interactions(76-79). The protein folding and thermodynamic stability measurement approaches reported so far have been used to characterize the binding of target proteins to specific ligands(80-82). Such energetics-based approaches are attractive for protein-ligand binding analyses because they can be quantitative and general with respect to ligand class. They also can be performed directly in solution and do not require modification and/or immobilization of the ligand. While oxidative foot-printing strategies(76, 78, 83) can provide similar information as continuous H/D exchange(77, 84) and SUPREX (stability of unpurified protein from rates of H/D exchange) (79, 80), the amino acid modifications generated in oxidative foot-printing are stable and irreversible compared to H/D exchange strategies. However, these methods have also been limited in their throughput and are not amenable to large-scale analyses on the proteomic scale.

Recently, several different protease protection assays(67, 69-71) and thermal shift assays(72) have been developed for the detection of protein-ligand binding interactions on the proteomic scale. The cellular thermal shift assay (CETSA) (72, 85) measures the ligand induced thermal stabilization of the target protein. This approach has been successfully applied to detect ligand induced change in melting point of proteins upon

treatment of cells with the ligand. Similarly, Limited Proteolysis (LiP)(86) and the Drug Affinity Responsive Target Stability (DARTS)(71) approaches monitor the differences in conformation of proteins in native environment in the presence or absence of an external perturbation such as the ligand or other cellular changes. The results in the DARTS experiment are visualized using gel electrophoresis readout and the proteins in gel bands with differential intensities are identified using mass spectrometry. Only very recently, LiP coupled with the single reaction monitoring (SRM)(67) and CETSA coupled with TMT10plex strategy(68) enabling mass spectrometry utilization for high-throughput multiplexing capabilities were reported. These approaches confer several advantages in terms of unbiased detection in the context of complex matrices without a need for derivatization of the ligand or probe. However, while the LiP and DARTS approach rely on proteolytic susceptibility under native conditions, this proteolytic digestion can also occur as a result of local fluctuation rather than global conformational changes and may not reflect an accurate measure of protein stability. On the other hand, the CETSA provides an attractive approach for the *in vivo* drug-mode-of action and ligand binding changes but the non-equilibrium measurements in thermal shift assay can complicate the determination of unfolding enthalpy and heat capacity measurements and also is challenging to apply in the case of certain protein classes (i.e., heat shock proteins) that are sensitive to temperature changes. Two energetics-based

approaches, the SPROX⁽⁷³⁾ and pulse proteolysis⁽⁷⁰⁾ techniques, have shown promise for measurements of protein folding thermodynamics in protein-ligand binding analyses on the proteomic scale.

The work described in this dissertation involves the development of SILAC-based SPROX and pulse proteolysis techniques for proteome wide thermodynamic profiling studies in disease states. The basic principles of these energetics based approaches, are reviewed below as is a summary of the status of these techniques prior to the work described in this thesis.

1.5.3 Stability of Proteins from Rates of Oxidation (SPROX)

SPROX is a mass spectrometry-based chemical modification technique that uses the denaturant dependence of a hydrogen peroxide-mediated oxidation reaction of methionine side chains in proteins to measure the thermodynamic properties of protein-folding and ligand binding interactions. The inherent chemical stability of the chemical modification products in SPROX makes it compatible with shotgun proteomic analyses using bottom-up strategies.

The protocol involves the dilution of a protein sample into different denaturant-containing buffers followed by selective oxidation of the thioether group in the side chain of the methionine residue. The reaction conditions are tuned such that the pseudo-first order oxidation reaction of an unprotected methionine residue occurs for

approximately three half-lives. The oxidation reaction is quenched at the end of the three half-lives with excess methionine solution or catalase. The extent of oxidation in different denaturant containing buffers is measured using mass spectrometric readouts. SPROX has been interfaced with LC-MS based platforms to achieve identification and quantitation of protein-ligand binding interactions using quantitative bottom up proteomics with isobaric mass tags at the lysate level(63, 64, 73). In this strategy, a cell lysate is subjected to two SPROX analyses, one with and the other without the ligand. The peptides are labeled with isotope mass tags (iTRAQ or TMT) after SPROX reaction and subjected to LC-MS/MS analysis. The oxidation states of methionine residues in the samples with and without the ligand are quantified as a function of an increasing denaturant concentration. Since the oxidation reaction is performed on the intact proteins prior to protease digestion, the oxidation state of the detected methionine containing peptide provides thermodynamic information on the protein from which it was derived.

Several proteomic applications of proteome-wide SPROX analyses have demonstrated that the SPROX method is capable of detecting both the direct and indirect targets of protein-ligand binding interaction(64, 73). The altered thermodynamic stability of a protein in a disease state due to changed protein interaction networks can include both direct and indirect protein interactions arising from a change in cellular

ligands including proteins. As such the ability of SPROX to detect both direct and indirect binding interactions is an important asset to the proposed application in differential disease state analyses.

The iTRAQ-SPROX protocol used for proteome-wide SPROX analyses relies on the detection and quantitation of methionine-containing peptide in the bottom-up proteomics readout. The frequency of methionine residues in proteins, approximately 2.5%, is relatively low (87). However, an analysis of the E. coli genome revealed that a large majority of proteins have at least one methionine residue in their structure other than the initiator methionine(88). A majority of methionine residues in proteins are expected to be globally protected in the protein's native structure due to the hydrophobic nature of the methionine side chain(89). This is important because the SPROX oxidation behavior of even a single globally or sub-globally protected methionine residue in a protein can report on the thermodynamic stability of the entire protein or protein domain to which it maps. Thus, the scope of SPROX is not fundamentally limited by the relatively low frequency of methionine residues in proteins.

1.5.3.1 SPROX with Colorectal Cancer Tissue Lysates

The first disease state characterization study was attempted with the existing iTRAQ-SPROX analysis protocol(63). This study involved thermodynamic profiling of

proteins from tissue lysates extracted from tumors of colorectal cancer patient xenografted on the mice models treated with or without drug exhibiting different therapeutic response. The experimental details and the findings from this experiment are described in chapter 2 of this dissertation. The iTRAQ-SPROX experiment with colorectal cancer samples exposed several limitations associated with the existing protocol for disease state analysis that spurred the subsequent development and application of the Stable Isotope Labeling with Amino acid in Cell culture (SILAC)-based protocols described in the rest of the dissertation. A large fraction (~30%) of the data acquired in the iTRAQ-SPROX experiments was of poor quality which compromised the ability to define the potential protein targets. The availability and use of iTRAQ-8plex allowed only eight denaturant points and it created a practical limitation on the sensitivity of the approach. This could have a big impact in protein target identification. Additionally, this protocol also has a relatively high false discovery rate in the large scale proteomic analysis⁽⁹⁰⁾. Therefore, it was necessary to address such issues to effectively use SPROX to measure the changes in the disease state analysis presented here. As part of this work, SILAC-based quantitation strategy was utilized as an alternate quantitation strategy with SPROX to increase the quality of the data and reduce the uncertainty of the experiment. Concurrently, use of complementary approaches such as pulse proteolysis was also

explored together with SPROX to increase the proteomic coverage and as a validation tool required for disease state analysis.

1.5.4 Pulse Proteolysis

Pulse proteolysis is another energetics-based technique with the potential to provide information complementary to SPROX. Park *et al* developed the pulse proteolysis(69, 91) method based on the principle of differential proteolytic digestion(92, 93) of the folded and unfolded conformations of proteins. In this approach, the proteins are subjected to a brief pulse of digestion using a nonspecific protease such as thermolysin in the presence of a denaturant. Pulse proteolysis is tuned such that the protease in the assay selectively cleaves unfolded proteins in an equilibrium mixture of native and unfolded proteins, unless the native protein has intrinsically unstructured regions susceptible to proteolysis. Pulse proteolysis coupled with 2D gel electrophoresis has been applied to detect protein-ligand binding interactions and was extended to study thermodynamic stability of protein-ligand interactions at the cell lysate level(69, 91). The quantitation in these previous experiments was achieved using densitometry measurements of the stained gel band intensity to obtain the fraction folded in a particular denaturant concentration. Pulse proteolysis coupled to SDS PAGE has also been reported(94). In these experiments, the complexity of the lysate was reduced by using column chromatography to elute protein fractions before running them on SDS

PAGE. While this approach has been successfully applied to detect protein-ligand interactions in relatively small proteomes, it is tedious, time consuming, and will likely be difficult to apply to complex proteomes. Therefore, a SILAC-based quantitation strategy coupled with pulse proteolysis and SDS PAGE to detect and validate molecular targets using mass spectrometry was investigated as part of this work.

1.6 SILAC-based Quantitation Strategy

In an effort to improve the quantitative proteomics readout in the SPROX and Pulse proteolysis techniques described above an alternative Stable Isotope Labeling by Amino acids in Cell culture (SILAC)-based approach was explored (38). Proteome-based expression profiling studies using SILAC have been widely used in functional and quantitative proteomics to understand the basic physiological processes and biological pathways involved with aging and disease(13, 95, 96). In SILAC cells are grown in media containing either heavy or light labeled essential amino acids that are encoded into the proteome when proteins are synthesized. In a typical experiment, an equal amount of both the light and heavy labeled proteins are mixed together, digested with a protease, and subjected to mass spectrometry analysis. The ratio of peak intensities of the heavy and light peptide provides quantitation in SILAC. Since the proteins are labeled, this allows the samples to be combined earlier in the experimental protocol. The combined samples can then be subjected to proteomic sample preparations. This has

several potential advantages over iTRAQ quantitation: i) it helps to lower the uncertainty associated with quantitation as samples are combined earlier in the proteomic sample preparation, ii) it allows for incorporation of more than eight (in iTRAQ) denaturant points to increase sensitivity of the approach, and iii) SILAC quantitation relies on precursor ion (MS1), and therefore eliminates potential problems associated with chimeric peptides in product ion (MS2) quantitation, and iv) it enables the concurrent analysis of expression level profiling and thermodynamic profiling of proteins in different biological state, offering orthogonality to existing norms in the disease state analysis.

Similarly, a SILAC-based quantitation strategy coupled to pulse proteolysis offers several advantages to traditional pulse proteolysis. These include: i) it negates the need for extensive fractionation of sample prior to reaction so that both direct and indirect binding interactions can be probed, ii) it enables the development of quantitative bottom up proteomics using mass spectrometry in pulse proteolysis, and iii) it creates an opportunity to compare and validate results from other large scale proteomic methods such as SPROX.

As part of the work described in this thesis two approaches that utilize SILAC quantitation strategy were developed. The Stable Isotope Labeling with Amino acid in Cell culture (SILAC) quantitation strategy coupled to Stability of Proteins from Rates of

Oxidation (SILAC-SPROX) or Pulse Proteolysis (SILAC-PP) were used in a proof-of-principle studies with two different ligand systems, Cyclosporine A (CsA) and ATP analogue AMP-PNP, involving protein from yeast cell lysates. This technique development is described in the chapters 3, and 4 of this dissertation. These strategies were ultimately used to study protein folding and thermodynamic stability changes in cell line models of breast cancer.

1.7 Cell Line Models of Breast Cancer

The disease states studied in this work included three different cell culture models of breast cancer. The optimized strategies developed and described in Chapters 3 and 4 of this dissertation were applied to measure thermodynamic stability of proteins in breast cancer cell lines. The studies described here involved comparative analyses of protein stabilities in the MCF-10A cell line, a commonly used “normal control” in breast cancer studies, and the MCF-7 and MDA-MB-231 cell lines that have different molecular features and display different degrees of invasiveness. These cell culture models of breast cancer have been well studied using gene and protein expression level studies, and a number of differentially expressed proteins have been identified in these cell lines(12, 97-102). The wide use of these cell lines for genomic and proteomic studies make them especially suitable for this first in class study(103). These cell lines are also compatible with the proposed quantitation strategy as they have shown successful

adaptation in the SILAC media(12). All these characteristics make these cell lines a good initial system for study in the research presented here. The work involving these cell line models and applications of the two SILAC-based strategies to study the changes in protein stability profile is described in detail in chapter 5 of this dissertation.

2. Proteomic Characterization of Therapeutic Response in Patient-Derived Tumor Explants of Colorectal Cancer in Murine Models.

2.1 Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer death in the United States. It is estimated that around 133,000 people will be diagnosed and 50,000 will die due to CRC in the year 2015(2). Around 60% of the CRC patients are diagnosed at the advanced stages and the treatment regimen involves radiation chemotherapy and drugs to contain the tumor post-surgery(104). Recent technological advances and the introduction of cytotoxic drugs such as oxaliplatin, irinotecan and fluorouracil have helped improve the overall treatment efficacy of advanced CRC patients in recent years. Several targeted antibody based therapeutics such as cetuximab against epidermal growth factor receptor (EGFR) and bevacizumab, another monoclonal antibody against vascular endothelial growth factor have also been introduced for cancer therapy(105, 106). In multiple randomized trials investigating the strength of the adjuvant regimen, oxaliplatin together with fluorouracil performed the best(107). Therefore, oxaliplatin has been the focus of a number of subsequent adjuvant therapy research and applications in metastatic CRC patients. Oxaliplatin is a cytotoxic platinum analogue that is currently used as part of the adjuvant therapy in metastatic colorectal cancer. Oxaliplatin imparts cytotoxicity by forming DNA adducts, and it is

most efficacious when combined with fluorouracil and leucovorin against late stage colon cancer(108). However, not all CRC patients benefit equally from chemotherapy and other adjuvant therapies. There have been incidences of adverse effects stemming from these therapeutic interventions. Oxaliplatin treatment has been shown to cause cumulative neuropathy that is irreversible in some cancer patients(109). It is therefore important to elucidate the molecular characteristics associated with patient tumors for better adjuvant therapy outcomes and reduce toxicity.

Several genomic and proteomic studies probing specific pathways and gene and protein expression profiles have been carried out to better understand the disease progression and potential biomarker discovery in the context of CRC tumors (25, 110, 111). Currently CRC and other cancer lack early diagnostic markers and large majority of patient are diagnosed at advanced stages of cancer. Many patients with advanced CRC initially respond to treatment regimen but experience disease relapse after some time due to development of drug resistant tumors. There are no effective predictive markers to disease resistance and it is frequently difficult to determine which drug treatments will be most efficacious for those CRC patients. The discovery of molecular markers to better classify different tumors has the potential to inform the outcome of specific drug treatments. The clinical utility of anti-cancer drugs, such as oxaliplatin could also benefit from a better understanding of their mode-of-action. We hypothesized

that the thermodynamic stabilities of proteins in tissue cell lysates from patient derived tumors could be used for protein biomarker discovery to understand disease progression and predict therapeutic outcomes. Here we utilize the iTRAQ-SPROX methodology to search for such protein biomarkers.

The iTRAQ/TMT-SPROX strategy has been successfully utilized in a number of *in vitro* drug mode of action studies involving yeast cell lysates(64, 73). For example, the iTRAQ/TMT-SPROX strategy has been used to probe various known and novel targets of drugs such as cyclosporine and resveratrol(64, 73). These and other SPROX studies of protein-ligand binding interactions have involved an *in vitro* analysis of cell lysates with and without the target ligand spiked into the sample. The sensitivity and applicability of this approach to *in vivo* drug treated diseased samples have not been performed. The studies on murine models help to identify targets of the therapeutic molecule and/or metabolite of the drug and help elucidate both direct and off-target effects of the therapy under consideration *in vivo*.

Described in this chapter are the first efforts to apply thermodynamic measurements of protein folding and stability to the characterization of disease-related states. Here, the iTRAQ-SPROX strategy is used to generate thermodynamic stability profiles of proteins derived from a series of patient-derived tumor explants propagated in mouse models of colorectal cancer to interrogate disease related changes for disease

state characterization and biomarker discovery efforts. In collaboration with Professor David Hsu's Laboratory in the Oncology Department in the Duke University Medical Center, thermodynamic stability measurements were performed on eight different tumor tissue lysates, which were prepared by the Hsu lab. These lysates were utilized in this study to interrogate, i) the proteomic signatures of sensitive and resistant tumors post oxaliplatin treatment, ii) elucidating thermodynamic profiles of mode-of-action in oxaliplatin treated (OX) and not treated (NT) xeno-graft samples from oxaliplatin resistant or sensitive CRC patients.

2.2 Experimental Procedures

2.2.1 Comparative Analyses Schematic

Figure 1 shows the summary of the two comparative analyses performed in this work. First comparison involved tumors from CRC patient resistant (CRC_183) or sensitive (CRC_102) to oxaliplatin therapy. The tumor explants from these patients were xenografted in two different mice and were either administered with oxaliplatin (OX) or not (NT). In the case of CRC_102 explants tumors from multiple generations (CRC_102_2 and CRC_102_5) were used for analyses. The second comparative analysis involved the two CRC patient tumor explants, sensitive (CRC_096) and resistant (CRC_240) propagated in two mice and oxaliplatin was injected in both mice. The oxaliplatin elicited different therapeutic response, shrinkage (CRC_096) and no

shrinkage (CRC_240) of tumors in mice models consistent with the observation in patients.

2.2.2 Tissue Lysate Preparation

The tissue lysate were prepared in the Hsu lab. In all cases, the patient derived explant tumor tissues were lysed in a lysing Matrix A tube (MP Biomedicals) with 400 μ L 20 mM Phosphate buffer (pH 7.4) and 1 X protease inhibitor and a cocktail of protease inhibitors that included 1 mM AEBSF, 500 μ M Bestatin, 15 μ M E-64, 20 μ M Leupeptin and 10 μ M Pepstatin A (Thermo pierce). A MP Fast-Prep-24-Instrument homogenized the tissues, for a minute. After homogenization, samples were centrifuged at 4°C and 14,000 x g for 10 minutes to pellet down the cellular debris out of the supernatant. The supernatant was centrifuged with the same setting one more time to precipitate any remaining cellular debris. The protein concentrations of all lysates were measured using a Bradford assay. The total concentration of protein in each tissue lysate was typically between 10 and 28 mg/ml.

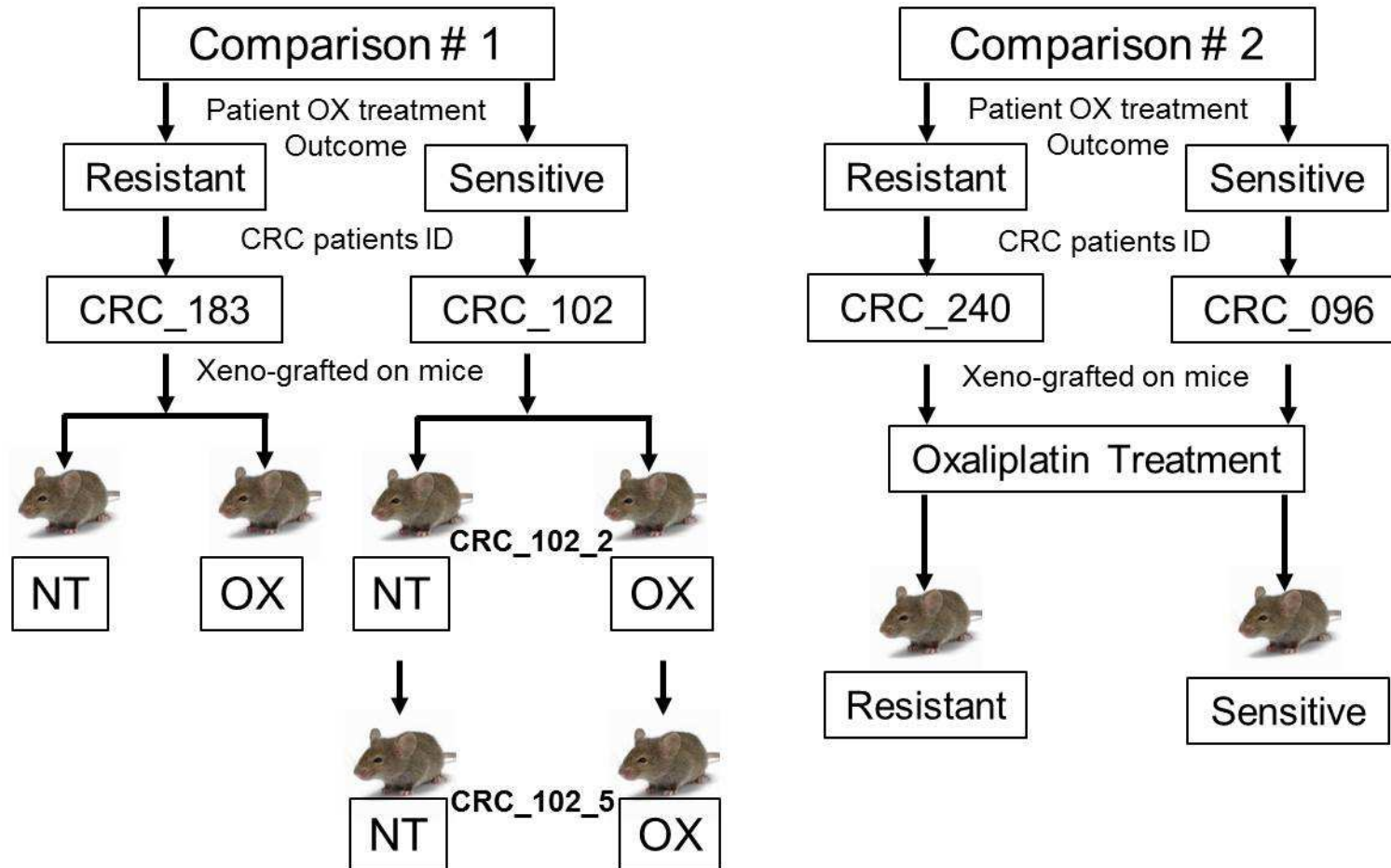


Figure 1: Summary of the comparative analyses performed using iTRAQ-SPROX strategies on tumor tissue lysate derived from colorectal cancer patient explants. NT and OX represent not treated and oxaliplatin treated mice respectively.

2.2.3 iTRAQ-SPROX Protocol

The lysate concentrations were normalized to the same initial concentration of 6 mg/ml for the first comparative analysis involving the NT and OX samples and of 10 mg/ml for the second comparative analysis involving the oxaliplatin sensitive and resistant explant samples. A 20 μ L portion of each tissue lysate sample was then diluted into a series of eight 75 μ L aliquot of denaturant-containing buffer consisting of 20 mM phosphate and concentrations of urea ranging from 0 M to 5.9 M. The final urea concentrations in the reaction mixture for the NT versus OX comparison were 0, 1, 2.1, 2.8, 3.1, 3.5, 4, and 5.4 M for the CRC_183 sample analysis, and 0, 1, 2.1, 2.7, 3.1, 3.4, 4, and 5.3 M for the CRC_102 sample analyses. The final urea concentration was 0, 2, 2.6, 3.3, 3.9, 4.5, 5.2, and 5.9 M for the samples in the second comparative analysis involving the resistant and sensitive tumor samples. Oxidation of the methionine side chain was initiated by adding 5 μ L of 9.8 M (30% w/v) H_2O_2 to each sample. After 6 minutes, oxidation reaction was quenched with 950 μ L of a 300 mM saturated methionine solution prepared in water. An aliquot of 100% (w/v) trichloroacetic acid (TCA) was added to each SPROX sample to a final concentration of 16% and incubated overnight on ice to precipitate proteins. The samples were centrifuged at $8000 \times g$ for 30 min at 4°C and the resulting pellets were washed with ice cold ethanol three times. These protein

pellets were redissolved in 0.5 M triethylammonium bicarbonate buffer (TEAB) (pH 8.5) containing 0.1% SDS, disulfide bonds of the proteins in each sample were reduced with tris (2-carboxyethyl) phosphine hydrochloride (TCEP HCl, Thermo Scientific) and cysteines were modified with S-methyl methanethiosulfonate (MMTS) before digestion with trypsin. Peptide samples were labeled with iTRAQ 8-plex isobaric mass tags (AB Sciex) as per the previously reported protocol(63).

2.2.4 Methionine-containing Peptide Enrichment

The iTRAQ labeled samples were combined within a set, OX or NT, and Sensitive or Resistant, using 40 μ L of each of the iTRAQ labeled samples. An 80 μ L of the combined sample was removed from each set and desalted using C18 resin (The Nest Group) according to the manufacturer's protocol. The remaining sample volume was reduced to 50 μ L using a SpeedVac Concentrator (Thermo pierce). A Pi³ Methionine Selective Resin (The Nest Group) was used according to the established protocol(63) to selectively isolate un-oxidized methionine-containing peptides except the capture step was carried out for three hours and wash step after capture with β ME extended to an hour. The methionine enriched samples were desalted using C18 resin (The Nest Group) before they were subjected to an LC-MS/MS analysis.

2.2.5 LC-MS/MS Analysis.

All the NT and OX CRC samples were analyzed on an Agilent 6520 Q-TOF mass spectrometer. The Q-TOF utilized a Chip Cube Interface, and the enrichment column consisted of 150 mm × 75 μm Zorbax 300SB-C18 5 μm packing. A 40 μL aliquot of each nonenriched and methionine enriched sample was injected onto the column. The condition for the LC-MS/MS was similar to previously reported protocol(63, 90). Data were collected in a data dependent acquisition mode. The NT and OX non-enriched samples of the CRC_183 samples were each run three times and the methionine-enriched samples were run six times. The NT and OX non-enriched and enriched samples of the CRC-102 samples were each run three times.

Similarly, six runs of non-enriched and five runs of methionine enriched samples from resistant and sensitive CRC samples (CRC_240 and CRC_096 , respectively) were ran on QTOF in the similar manner as described above. The CRC_240 and CRC_096 non-enriched samples were also analyzed on an Orbitrap Elite ETD mass spectrometer equipped with an EASY-nLC system at the Proteomics Resource, Fred Hutchinson Cancer Research Center. The LC-MS/MS acquisition parameter and the set up for the mass spectrometry analysis were similar to previously described protocol(90). Product ion scans, at a resolution of 15,000, were collected for the top 10 most intense peaks in a

given precursor scan, at a resolution of 60,000, with an intensity threshold of 5000. Collision induced dissociation was achieved using HCD with normalized collision energy of 40% and an HCD activation time of 0.1 ms.

2.2.6 iTRAQ-SPROX Data Analysis

The peak lists from the .d files acquired in QTOF and .raw files in orbitrap elite runs were extracted with Agilent's spectrum mill software version B.04.00. In each case, the iTRAQ 8-plex at the N-terminus and lysine residues as well as modification of cysteines by MMTS was used as fixed modifications and the oxidation of methionine and de-amidination of asparagine and glutamine in the peptide sequence as variable modifications. The spectrum mill output was exported to excel and the data acquired for NT and OX samples and sensitive and resistant samples were analyzed following previously reported procedures (63) and using excel. Briefly, iTRAQ intensities for methionine-containing peptides were subjected to two normalizations. In the first normalization (N1), the raw intensity for each iTRAQ tag was divided by the average value of all iTRAQ intensities for that particular peptide. In the second normalization (N2), the methionine-containing peptides were normalized with the non-methionine-containing peptides. Average N1 values for the non-methionine-containing peptides were filtered to have raw intensity sums greater than 1000, no zero intensities, and FDR

1% or less. The N1 values for the methionine peptides were then divided by the average of the N1 values for the filtered non-methionine-containing peptides to generate N2 values for the methionine peptides, which were used to generate normalized reporter ion intensities in each iTRAQ-SPROX data sets. Only peptides present in both NT and OX or in both Sensitive and Resistant samples were used to assess protein folding thermodynamics.

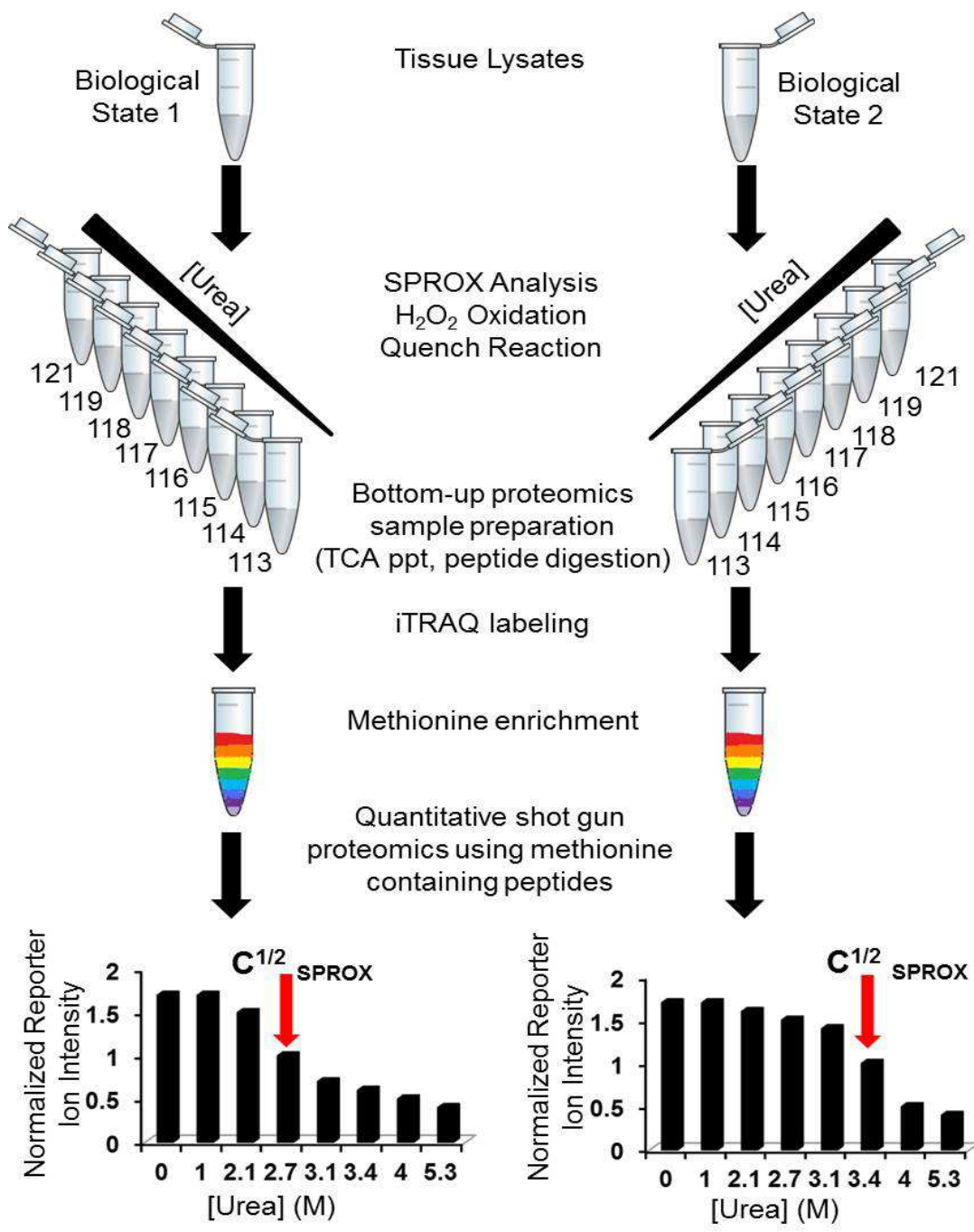


Figure 2: General scheme of the iTRAQ-SPROX protocol(63, 90).

2.3 Results and Discussion

2.3.1 General Strategy

The protein folding thermodynamic changes in CRC samples were determined using the iTRAQ-SPROX protocol outlined in Figure 2(63). As shown in Figure 2, equal amounts of tissue lysates from either biological state 1 or 2 (i.e. Oxaliplatin treated (OX) versus not treated (NT) or Sensitive versus Resistant) were diluted into increasing urea-containing buffers and were allowed to reach equilibrium in urea containing buffers before subjecting it to SPROX analysis in parallel. The reaction was quenched with excess methionine and proteins were subjected to bottom-up proteomic sample preparation protocol. The resulting peptides were labeled with isobaric mass tags using iTRAQ 8 plex for quantitation. The peptides in the lowest denaturant concentration were labeled with the 113 reporter tag and in succession the sample from highest denaturant was labeled with the 121 tag. These labeled samples were then combined into a single sample allowing for multiplexing in subsequent LC-MS/MS runs. A portion of the labeled sample underwent enrichment step for wild-type methionine-containing peptides. Samples from both the non-enriched and methionine-enriched were then subjected to methionine peptide detection and quantitation using mass spectrometry. Data sets of normalized reporter intensities of iTRAQ tags from same methionine-

containing peptides from different samples were visually inspected to identify those with a significant change in the $C_{1/2}$ value (>1 M [Urea]).

2.3.2 Global Thermodynamic Profiles of Proteins in the Lysate from CRC Explants

2.3.2.1 Oxaliplatin Treated versus Not Treated Samples SPROX Analysis

The iTRAQ-SPROX data generated on the CRC explant samples (i.e., the CRC_183 and CRC_102 samples) treated with and without oxaliplatin were compared in the first comparative analysis (see Figure 1). Two biological replicates of the CRC_102 sample were performed. A summary of the proteomic coverage for the samples in this first set of comparative analyses is shown in Table 1.

Table 1: Proteomic coverage for the oxaliplatin treated (OX) and not treated (NT) experiment with iTRAQ-SPROX protocol in the first comparative analyses.

Experiment(s)	Not Treated (NT)		Oxaliplatin Treated (OX)		Matched Peptides (Proteins)
	Unique Peptides (Proteins) Identified				
	All	Methionine containing	All	Methionine containing	
1 (CRC_183)	420(193)	248(149)	401(203)	238(147)	166(94)
2 (CRC_102_2)	417(240)	266(180)	517(265)	197 (128)	177(105)
3 (CRC_102_5)	301(180)	244(164)	362(176)	324(204)	161(98)

In each of these comparative analyses around 170 peptides from about 100 proteins were assayed for potential change in protein folding thermodynamics. The total

number of proteins and peptides included those that were identified with global FDR 1% or less. A global analysis of the distribution of normalized reporter ion intensities for methionine-containing peptides from the lowest denaturant containing buffer labeled with 113 iTRAQ tag and the highest denaturant containing buffer sample labeled with 121 reporter tag for each of the three experiments are shown in Figure 3.

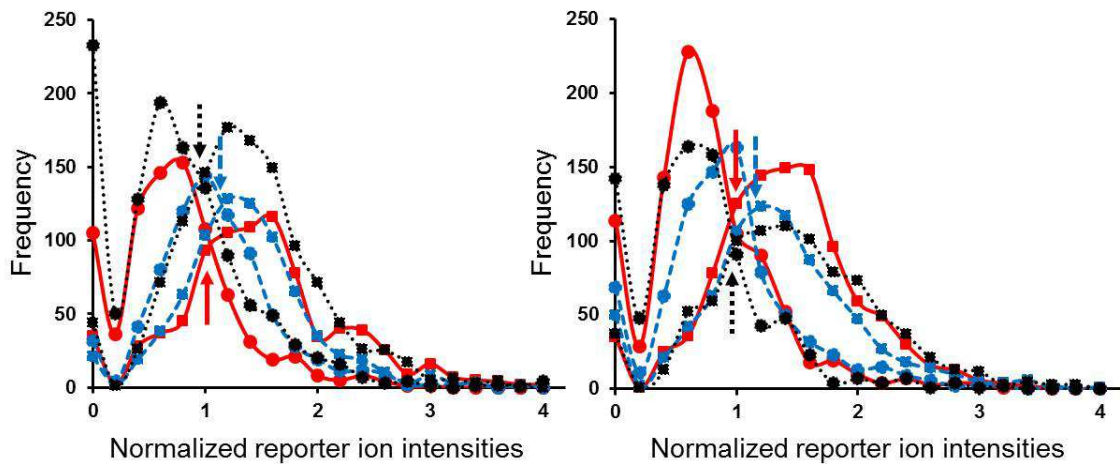


Figure 3: Distribution of the normalized reporter ion intensities for the lowest denaturant, 113 tag (filled squares) and highest denaturant, 121 tag (filled circles) in each of the three experiments. Shown are the distribution for not treated (NT) on the left and oxaliplatin treated (OX) on the right. For each distributions, experiment 1 (solid red lines), experiment 2 (dashed blue lines) and experiment 3 (dotted black lines) are shown.

Only the wild type methionine containing peptides were used to create these distributions. The point of intersection for these distributions occurred around the normalized reporter ion intensity of close to 1.0 in each case. The intersection points are indicated by arrows in Figure 3. This normalized reporter ion intensity is used to distinguish between the pre-transition (above 1.0) and post-transition (below 1.0) regions in the data sets. $C^{1/2}_{\text{SPROX}}$ values were assigned to each of the sample set via visual inspection using the established guidelines reported previously(63). The peptides and corresponding protein with the change in $C^{1/2}_{\text{SPROX}}$ 1 M urea or greater were considered significant in each analysis and were included in the initial hit list. A total of 10 unique peptides from 9 proteins were identified with significant $C^{1/2}_{\text{SPROX}}$ value shifts (Table 2). There were no overlapping peptides identified with altered stability between the different samples. However, one protein (elongation factor 2) showed stabilizing behavior in two experiments (Table 2) albeit with two different peptides. It is interesting to note that these two experiments involved tumors from patient exhibiting two different responses to oxaliplatin drug. More experiments are needed to establish potential difference in the proteins arising from changes in the domains represented by these peptides leading to different response to the drug.

Table 2: The peptides and proteins identified with altered thermodynamic stability between not treated (NT) and oxaliplatin treated (OX) CRC samples.

Exp.	Proteins	Peptides	$\Delta C1/2$ (M)
#1 (CRC 183)	Alpha-enolase	(K)LMIEMDGTENK(S)	1.2
	Elongation factor 2	(R)IMGPNYTPGK(K)	1
#2 (CRC 102_2)	Fructose-bisphosphate aldolase A*	(K)IGEHTPSALAIMENANVLAR(Y)	1.3
	Keratin, type II cytoskeletal 8*	(R)SLDMDSIIAEVK(A)	-2.4
	Putative heat shock protein HSP 90-alpha A4	(R)TLTIVDTGIGMTK(A)	1.3
	Elongation factor 2	(K)GPLMMYISK(M)	1.3
#3 (CRC 102_5)	Proteasome subunit alpha type-3	(R)VFQVEYAMK(A)	1.7
	Protein disulfide-isomerase A4	(K)TFDSIVMDPK(K)	1.3
	Triosephosphate isomerase	(K)VTNGAFTGEISPGMIK(D)	>2.9
	L-lactate dehydrogenase A	(R)VHPVSTMIK(G)	1.1

*Proteins previously implicated in colorectal cancer((112, 113)

The remaining eight peptides were either not assayed or had poor quality data making it difficult to confirm the result. All but one peptide and corresponding protein showed higher thermodynamic stability in the sample treated with oxaliplatin compared with the sample from not treated tumor derived tissue lysates. Some of the proteins with altered stability have also been shown to have altered expression level using cDNA macroarray and mass spectrometry based methods (see Table 2) (112, 113).

2.3.3.2 Oxaliplatin Resistant versus Sensitive CRC Tumors

The protocol described in Figure 2 above was also utilized to assess the altered thermodynamic stability in two additional patient derived tumor tissue lysates including one (CRC_096) that was sensitive to oxaliplatin treatment and one (CRC_240) that was not sensitive to the drug. The proteins in the tissue lysates from oxaliplatin sensitive and resistant tumors were analyzed for potential stability differences. Table 3 shows the proteomic summary of the sensitive versus resistant comparison.

Table 3: Proteomic coverage for the oxaliplatin treated sensitive and resistant experiment with iTRAQ-SPROX protocol in the second comparative analyses.

Experiment(s)	Sensitive (CRC_096)		Resistant (CRC_240)		Matched Peptides (Proteins)
	Unique Peptides (Proteins) Identified				
	All	Methionine containing	All	Methionine-containing	
QTOF	511(253)	160(107)	908(430)	547(316)	134(68)
Orbitrap	668(395)	144(111)	1150(516)	226(159)	96(56)

All the peptides were identified with global false discovery rate of 1% or less in both the Q-TOF and Orbitrap mass spectrometry platform analysis. A total of 134 peptides from 68 proteins and 96 peptides from 56 proteins were matched between the sensitive and resistant sample analysis in the QTOF and orbitrap platform respectively.

Table 4 shows the list of proteins and peptides identified with altered thermodynamic stability in the QTOF and Orbitrap analysis of the sensitive versus resistant samples.

Table 4: The peptides and proteins identified with altered thermodynamic stability between sensitive and resistant CRC samples.

Experiment(s)	Protein(s)	Peptide(s)	$\Delta C_{1/2}$ (M)
QTOF	Heat shock cognate 71 kDa17	(K)NSLESYAFNMK(A)	1.25
Orbitrap	Pyruvate kinase isozymes M1/M218	(K)FGVEQDVDMVFASFIR(K)	1.25
	Transgelin-219	(K)QMEQISQFLQAAER(Y)	-1.95
	Glyceraldehyde-3- phosphate dehydrogenase	(K)VIHDNFGIVEGLMTTVHAI TATQK(T)	1.2

Only one peptide NSLESYAFNMK from Heat shock cognate 71 kDa protein was identified with significant change in thermodynamic stability in QTOF analysis. Unfortunately, this NSLESYAFNMK peptide was not assayed in the orbitrap mass spectrometry analysis.

In the Orbitrap analysis, three proteins with a peptide probe each were identified with significant shift in the midpoint of transition via visual inspection of the matched data. Of the three hits in Orbitrap analysis, only VIHDNFGIVEGLMTTVHAI TATQK peptide from Glyceraldehyde-3-phosphate dehydrogenase was assayed in both QTOF and Orbitrap analysis but failed to replicate

similar behavior in QTOF analysis. Remaining two hits were not assayed in the QTOF analysis. The representative iTRAQ-SPROX data sets for these two proteins, Pyruvate kinase isozymes M1/M2 (FGVEQDVDMVFASFIR) and Transgelin-2 (QMEQISQFLQAAER) are shown in Figure 4. All the hit proteins and peptides showed positive shift in the transition midpoint of the un-folding curve in the sensitive tumor except transgelin-2 peptide that showed decreased stability. All three of the proteins identified here have been previously reported to be associated with colorectal cancer(114-116). One of the proteins, pyruvate kinase, has been shown to be involved in oxaliplatin response and oxaliplatin resistant was correlated to decreased expression level of the protein in colorectal cancer tissues and cell line samples(110).

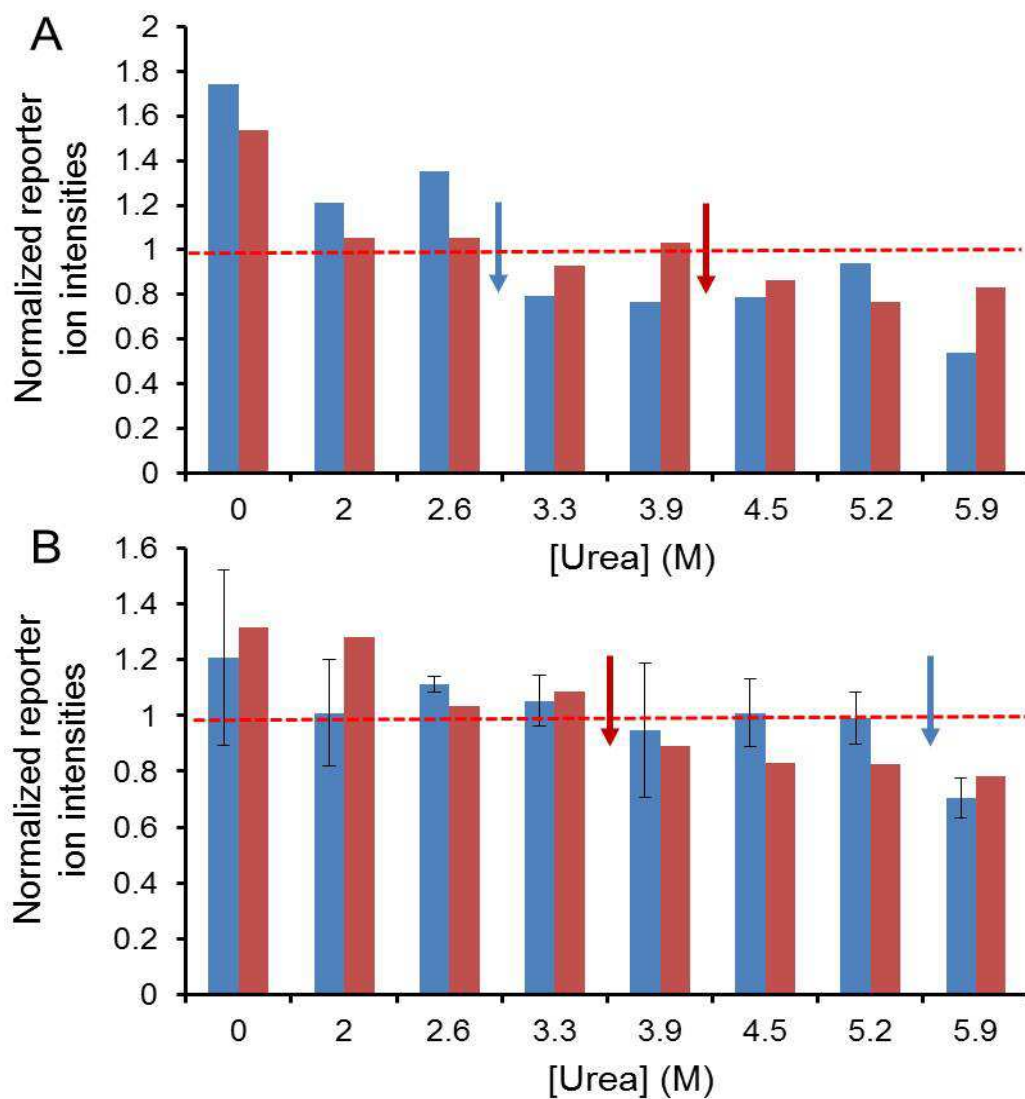


Figure 4: Representative iTRAQ-SPROX data sets from two proteins that showed difference in thermodynamic stability between oxaliplatin resistant (blue bars) and sensitive (red bars) CRC samples. Shown are the data sets for Pyruvate kinase isozymes M1/M2 (FGVEQDVDMVFASFIR) (A) and Transgelin-2 (QMEQISQFLQAAER) (B). In each case, the red and blue arrows represent the $C_{1/2}$ values for the peptide in sensitive and resistant samples. The dotted line represents the transition base line value for each of the experiments.

2.4 Conclusions

This study provided a first insight into the potential of SPROX methodology to identify protein targets in complex disease related changes such as shown here with colorectal cancer tissue lysates. The SPROX methodology identified proteins with altered stability, some of which have been previously reported to be involved in colorectal cancer using differential expression level and other studies. However, the sensitivity and specificity of the existing iTRAQ-SPROX methodology is limited by the low proteomic coverage, high false discovery rate, and the poor quality of reporter ion intensities. Moreover, more biological and technical replicate measurements will be needed to confirm the findings reported here. The subsequent chapters 3 and 4 in this dissertation describe development of SILAC-based energetics approaches in the context of protein-ligand binding to address aforementioned challenges using endogenous yeast cell lysates and subsequent application of these approaches to study breast cancer related changes.

3. The Development of SILAC-SPROX in Solution Protocol

3.1 Introduction

The iTRAQ-SPROX experiment with colorectal cancer samples described in the preceding chapter demonstrated the potential of the SPROX analysis to perform unbiased detection of molecular targets in disease state characterization and biomarker discovery. Yet, it also revealed several limitations associated with the existing protocol for such an analysis. A significant fraction of the data acquired in the iTRAQ-SPROX experiments with tumor tissue samples was populated with more than two outlier data points that compromised the ability to define the potential protein targets. As shown in Figure 2, in the iTRAQ-SPROX protocol the samples in eight different denaturant concentrations for each of the experimental and control samples are processed in parallel and involve a number of proteomics sample preparation steps in each analysis (i.e., TCA precipitation, re-dissolution, reduction and modifications, trypsin digestion and desalting steps). These sample handling steps potentially add up to a large technical (between the reporter tags) and experimental (between the control and the experimental sample) errors. Indeed, the reported technical and experimental errors associated with iTRAQ quantitation is +/-11% and +/-23% respectively (117), and this increases to as high as 30-40% in iTRAQ-SPROX strategies. Furthermore, the commercial availability of

iTRAQ-8plex allowed labeling of peptides from eight different denaturant samples with eight reporter ions enabling multiplexing in iTRAQ-SPROX strategy but there still exists a practical limitation on the sensitivity of the approach to transition midpoint changes ($\Delta C^{1/2}_{\text{SPROX}}$) greater than or equal to 0.5 M in GdmCl or 1 M in urea between the samples. This is because the sensitivity of the iTRAQ-SPROX to altered protein folding thermodynamics relies on the precision with which $\Delta C^{1/2}_{\text{SPROX}}$ values can be determined for a given protein. It has been possible to determine $C^{1/2}_{\text{SPROX}}$ values to within ~0.2 M GdmCl (or ~0.4 M urea) using the current spacing of the eight different denaturant points in proteomic applications of existing iTRAQ-SPROX protocol. Additionally, this protocol also has a relatively high false discovery rate in the large scale proteomic analysis(64, 90). In the case of product ion (MS2) quantitation, as in iTRAQ, such high false positive rate potentially stem from the analysis of chimeric peptides produced when sequenced peptides come from simultaneous fragmentations of two precursor ions with similar m/z during tandem mass spectrometry data acquisition(118). It is necessary to address such issues related to iTRAQ-SPROX strategy to effectively utilize SPROX to measure the changes in the disease state analysis presented in this dissertation.

As part of this work, we incorporated an in vivo metabolic labeling of proteins termed Stable Isotope Labeling by Amino acids in Cell culture (SILAC)⁽³⁸⁾-based quantitation strategy with SPROX to increase the quality of the data and reduce the uncertainty of the experiment. As described in Chapter 1, SILAC quantitation has several advantages over iTRAQ quantitation: i) lower uncertainty associated with quantitation as samples are combined earlier in the proteomic sample preparation, ii) increased sensitivity as more than eight (in iTRAQ8plex) denaturant points can be used, and iii) quantitation based on the precursor ion (MS1) in SILAC, therefore does away with problems associated with chimeric peptides in product ion (MS2) quantitation, and iv) most importantly the non-methionine peptides identified in the experiment can be used to calculate the abundance of the proteins and thus enables the concurrent analysis of expression level profiling and thermodynamic profiling of proteins in different biological state, offering orthogonality to existing expression level based strategies in the disease state analysis.

Described in this chapter is the work involving development of a SILAC-based quantitative mass spectrometry-based proteomics method for the large-scale thermodynamic analysis of proteins. The methodology utilizes the Stability of Proteins from Rates of Oxidation (SPROX) technique in combination with a stable isotope

labeling with amino acids in cell culture (SILAC) approach to compare the equilibrium unfolding/folding properties of proteins. The proof-of-concept study was first carried out with proteins in the endogenous yeast cell lysate in the absence and in the presence of target ligand. The methodology, which is general with respect to ligand, enables measurement of the ligand-induced stability changes associated with protein-ligand binding interactions. The capabilities of the methodology were demonstrated and benchmarked with the binding of proteins in the endogenous proteins in a yeast cell lysate to cyclosporine A (an immunosuppressant with well-characterized protein targets). The false positive rate of ligand binding using SILAC-SPROX was also determined with a control experiment involving no ligand. Beside the application to disease state analysis presented in Chapter 5, this strategy has promise for use in future drug-mode-of action studies and protein-ligand interaction studies. Indeed, SILAC-SPROX strategy has been used to identify protein targets of adenylyl imidodiphosphate (AMP-PNP), a non-hydrolyzable analogue of the ubiquitous enzyme co-factor, adenosine triphosphate (ATP), and the results have been published in *Molecular & Cellular Proteomics Journal* along with the work described in this chapter¹.

¹ Tran, D. T., Adhikari, J., and Fitzgerald, M. C. (2014) SILAC-Based Strategy for Proteome-Wide Thermodynamic Analysis of Protein-Ligand Binding Interactions, *Molecular & Cellular Proteomics*.

3.2 Experimental

3.2.1 Yeast Cell Culture

The SILAC labeled yeast cells were cultured as described previously (119). Briefly, a glycerol stock of the yeast deletion strain BY4739 (Open Biosystems, Lafayette, CO), an auxotroph for lysine, was streaked on a petri-dish containing synthetic complete (SC) media with Difco™-agar and L-lysine. The Agar-media for the plate comprised of 1.7 g of yeast nitrogen base (Amresco), 5 g of ammonium sulfate (Sigma-aldrich), 20 g of Difco™-agar (BD, USA), 2 g of glucose (Sigma), 1.92 g of synthetic drop out mix – lysine (Sunrise Science Product, San Diego, CA), and 30 mg of L-Lysine all in 1 L media. The SC-media for the subsequent culture was comprised of 1.7 g of yeast nitrogen base, 5 g of ammonium Sulfate, 20 g of glucose, 1.92 g of synthetic drop out mix – lysine (Sunrise Science Product, San Diego, CA). After 3 days of growth at 30°C, an isolated colony was selected and inoculated into 10 mL of SC-media containing 30 mg/L of light L-lysine (Sigma Aldrich), hereafter referred to as light SC-media. The 10 mL culture was incubated overnight at 30°C. A 100 µL portion of the cell culture was transferred into 100 mL of the light SC-media. Simultaneously, another 100 µl of the cell culture was transferred into 100 ml of SC-media containing 30 mg/L of heavy Lysine (13C6, 15N2) (Sigma Aldrich and Cambridge isotopes laboratories, MA), hereafter referred to as

heavy SC-media. The 100 mL culture was incubated overnight at 30°C. The Light and Heavy cell cultures (100 µL portions) were transferred two more times into 100 mL portion of the corresponding light and heavy SC-media and each time grown overnight at 30°C, before they were ultimately transferred to 1 L of the corresponding light and heavy SC-media and grown overnight at 30°C. Portions of the 1 L overnight culture, 250 mL each, were centrifuged at 1000 x g to generate light and heavy lysine labeled cell pellets. The cell pellets were stored at -20°C.

3.2.2 Yeast Cell Lysate Preparation

For both the CsA-binding experiments and the control experiment, one cell pellet each from cultures with either heavy labeled lysine or normal “light” labeled lysine was taken. Cells were lysed in approximately 500µl of 20mM phosphate buffer with 1 X protease inhibitor and a cocktail of protease inhibitors that included 1 mM AEBSE, 500µM Bestatin, 15 µM E-64, 20 µM Leupeptin and 10 µM Pepstatin A (Thermo pierce) to generate heavy and light cell lysates. The cells were lysed in a Disruptor Genie (Scientific Industries) with glass beads 15 times with 25-30s cycle of mechanical disruption and 1 min of cooling on ice. After lysis the tubes containing the lysate were centrifuged at 15000 X g for 10 min at 4°C. Protein concentrations were determined to be 18 mg/ml (heavy) and 24 mg/ml (light) for the first and 9.0 mg/ml (heavy) and 9.1 mg/ml

(light) in the second CsA-binding experiment and 6.9 mg/ml (light) and 10.5 mg/ml (heavy) in the control experiment as determined using Bradford reagent protein assay. The protein concentrations in the light and heavy lysates were normalized to the lower concentration prior to the SILAC-SPROX analysis in each experiment.

3.2.3 SILAC-SPROX Analysis

The stock solutions of the denaturant-containing buffers used in this experiment were prepared in 20mM phosphate buffer (pH 7.4) with varying concentration of guanidinium chloride (GdmCl) (EMD chemical,inc.) ranging from 0 to 8 M. Previously described method was used to calculate the specific concentrations of GdmCl in each denaturant-containing buffers by measuring the refractive index of the buffer using a Bausch & Lomb refractometer (Rochester, NY)(120). For each of the 3 SILAC-SPROX analyses performed here, a pair of heavy and light labeled yeast cell lysates with matched protein concentrations was prepared and used to generate (-) and (+) ligand stock solutions for each of the CsA-binding experiments and equal volume of water was added to both (-) ligand samples for the control experiment. In the two CsA-binding experiments 9 volumes of the heavy and light yeast cell lysates were mixed with 1 volume of DMSO and with 1 volume 10 mM CsA in DMSO to generate the (-) and (+) ligand stock solutions, respectively. In the Control Experiment, 9 volumes of the heavy

and light yeast cell lysates were mixed with 1 volume of water to generate the two (-) ligand samples analyzed in the Control experiment.

In control experiment (both (-) ligand heavy and light samples) and CsA-binding experiments (the heavy (-) CsA and light (+) CsA), lysates were incubated at the room temperature for ~45 min. After incubation, 20ul of the lysate-drug or lysate-DMSO samples were aliquoted into two sets of tubes with 25ul of the denaturant containing buffers ranging from 0 to 3.4 M final concentration. The final concentrations in the 12 denaturant-containing buffers were 0, 0.6, 1.1, 1.4, 1.6, 1.9, 2.1, 2.4, 2.6, 2.9, 3.1, 3.4 M. The light and heavy samples with and without the drug or with water in control experiment were equilibrated an additional ~45 min in the denaturant containing buffers before initiating the SPROX reaction. Following incubation, 5 μ l of the 0.98 M hydrogen peroxide solution was added to each sample to start the methionine oxidation reaction. The oxidation reaction was allowed to proceed for 3 min and was quenched with 760 μ l of 375mM L-methionine solution. Following quenching with excess methionine solution, the samples from the same denaturant containing buffers but from with (light) and without (heavy) ligand in the CsA-binding experiment or without ligand light and heavy samples in the control experiment were combined before subjecting to subsequent proteomic procedures.

3.2.4 Proteomic Sample Preparation

In each analysis, the proteins in the samples were precipitated by adding 324 μ l of 100% aqueous solution of trichloroacetic acid (TCA) (1g/ml) to a final concentration of 20% (w/v). The samples were then vortexed to mix the TCA and centrifuged briefly. Following centrifugation, samples were incubated on ice and kept at 4°C overnight for further TCA precipitation. The following day, the samples were vortexed for few seconds and centrifuged at 8000 X g for 30 minutes at 4°C, discarded the supernatants, and the pellets were washed three times with 500 μ l of ice cold ethanol. Added 500 μ l of cold acetone and centrifuged for 30 minutes at 8000 X g. Next, added 500 μ l of acetone and stored at -20°C overnight. Next day, the tubes containing protein pellets were centrifuged at 8000 X g for half an hour, discarded the supernatant and the residual acetone was evaporated off using a savant speed vac concentrator for 3 minutes.

The dried protein pellets from the SILAC-SPROX analyses were dissolved in 60 μ l (CsA-Binding Experiment 1) or 30 μ l (Control Experiment and CsA-Binding Experiment 2) of 0.5M triethyl ammonium bicarbonate (TEAB) containing 3 μ l (Experiment 1) or 1.5 μ l (CsA-Binding Experiment 2 and Experiment 2) of a 2% stock solution of SDS. The disulfide bonds in each protein sample were reduced upon addition of 5 μ l of 50 mM TCEP and treatment for 1 hour at 60°C. The protein samples

were each reacted with 2.5 μl of 200 mM methyl methane thiosulfonate (MMTS) for 10 minutes at room temperature to block cysteine side chains. Ultimately, 3 μl of a 1 $\mu\text{g}/\mu\text{l}$ trypsin solution was added to the protein sample in each tube and the samples were incubated overnight (~15 hours) at 37°C before 5 μl of 10% trifluoroacetic acid (TFA) was added to quench the trypsin digestion. The samples were desalted using C18 resin (The Nest Group, Southborough, MA) according to the manufacturer's protocol, and an estimated 6 μg of total peptide from each sample, was subjected to an LC-MS/MS analysis.

3.2.5 LC-MS/MS Analyses

3.2.5.1 Instrument Parameters

The 12 samples generated in the CsA binding and control experiment were subjected to an LC-MS/MS analysis using an Agilent 6520 Q-TOF mass spectrometer equipped with a Chip Cube interface. The HPLC chip used in the SILAC-SPROX experiment involved a 160 nl enrichment column and a 75 μm X 150 mm analytical column with Zorbax 300SB-C18 5 μm packing material. Samples were loaded onto the column and peptides were eluted using linear gradient from 0 to 3% Solvent B (ACN with 0.1% formic acid (FA)) in Solvent A (H₂O with 0.1% FA) for first five minutes, 3 to 15% Solvent B in A for next two and half minutes, 15 to 45% Solvent B in A for another

78 minutes, and 45 to 100% Solvent B in A for next 10 minutes. The column was flushed with 100% Solvent B for the next 10 minutes and equilibrated with 3% Solvent B for additional 10 minutes and waited for 15 more minutes before injecting the next sample. The flow rate was 0.4 μ l/min. The capillary voltage ranged from 1800 to 1850 V. The flow rate of the drying gas was set at 6 L/min at 350°C. The skimmer and fragmentor were set at 65 and 175 V, respectively. The collision energy was as determined by the equation 3.5 V/100 Da with a -4.8 V offset. The inclusion window width for precursor ions was 4 m/z. The precursor purity stringency and purity cutoff were set to 100% and 30% respectively. The scan rate was three scans per second in the mass spectra and two scans per second in the product ion spectra. In every cycle, four precursors were selected for fragmentation. All the spectra were collected in both centroid and profile mode.

3.2.5.2 Spectrum Mill Parameters

The LC-MS/MS data generated from 12 runs on the Agilent Q-TOF system, on Agilent's Spectrum Mill MS Proteomics Workbench software, Rev B.04.00.122 PreRelease was used to extract peak lists from the LC-MS/MS data generated on the Agilent Q-TOF system and to search the data against the 6619 *Saccharomyces cerevisiae* proteins in the UniProtKB/Swiss-Prot database (uniprot_sprot_fasta.gz Release 2012_01/ modified on 2/22/12 downloaded on 2/28/2012). The search parameters used in the

Control Experiment and the CsA-Binding Experiments 1 and 2 were: *Saccharomyces cerevisiae* as the organism, SILAC heavy labeled lysine-8 and cysteine modification with MMTS as a fixed modification, oxidation of methionine and deamidation of asparagine (N) as variable (0-1) modifications, the precursor and product mass ion tolerances as 20 and 50 ppm (respectively), the trypsin as protease with a maximum of three missed cleavages, and the maximum product ion charge as 7. The L/H ratios of all peptides output from the Spectrum Mill searches, regardless of score, were used in the subsequent data analysis steps. Our justification for using all peptide identifications in subsequent data analysis steps was that true false positives would be best filtered out at later steps (e.g., would not be identified in samples from multiple denaturant concentrations, and or yield meaningful L/H ratios).

3.3 Quantitative Proteomic Data Analysis

The protein and peptide identifications obtained from the LC-MS/MS data searches were filtered to contain lysine residues. Only the identified peptides (and corresponding protein identifications) with non-zero and positive ratios were used in the subsequent data analysis steps. If multiple L/H ratios were calculated for a given peptide sequence and charge state at a specific denaturant concentration (e.g., both the Light and Heavy version of a peptide and charge state was identified in the LC-MS/MS

analysis and/or a peptide and charge state was identified in multiple product ion scans) they were averaged to give a single L/H ratio for the peptide and charge state at that denaturant concentration. Ultimately, these average L/H ratios were used to generate SILAC-SPROX data sets. Methionine-containing peptides that were identified in the protein samples derived from 4 or more different denaturant-containing buffers were assayed for potential change in thermodynamic stability due to ligand binding.

Using Microsoft Excel, peptides were further filtered to identify those with significantly altered L/H ratios at 2 or more consecutive denaturant concentrations. Significantly altered L/H ratios in this work were taken to be those that were greater than 1.7-fold difference from the baseline value observed for each peptide. The selection of this 1.7-fold cut-off value was based on a global analysis of the L/H ratios. Shown are the data for all the identified peptides that were successfully quantified in the Control experiment and the CsA-binding experiments (Figures 5). The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value. In all the binding experiments performed here $\geq 90\%$ of the measured L/H ratios were within 1.7-fold of the median value. Thus, the estimated p-values associated with the hits in this study were generally less than 0.01.

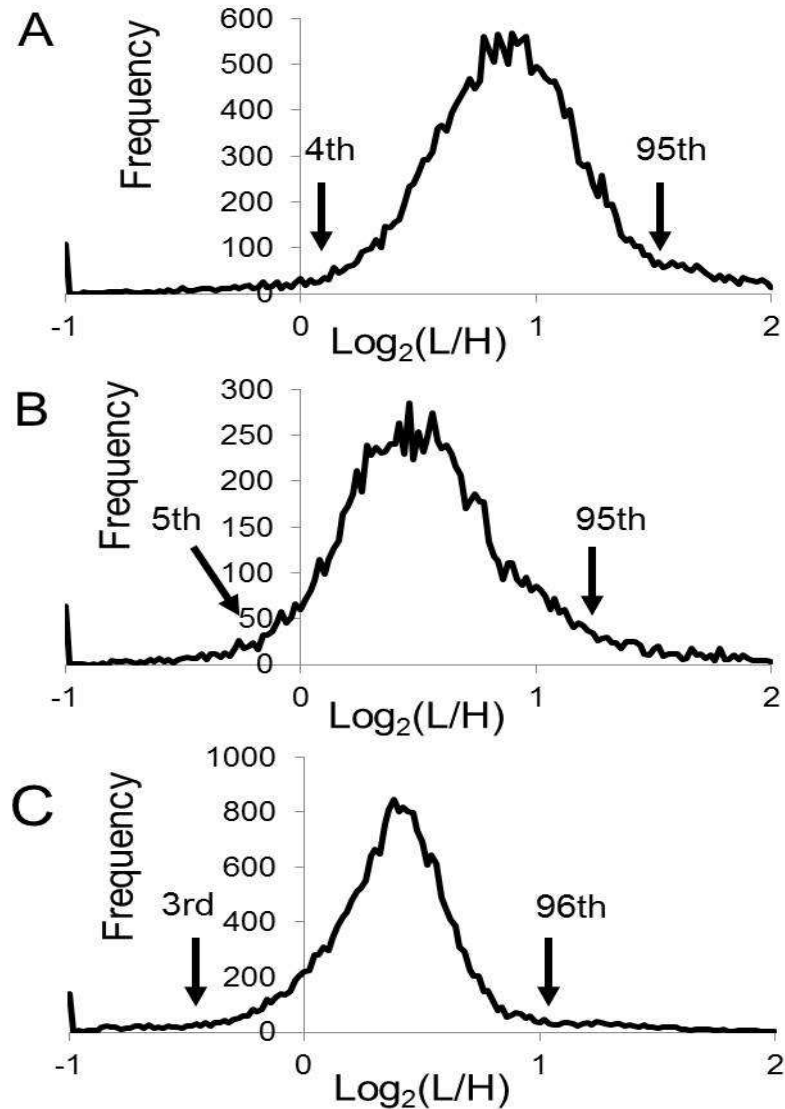


Figure 5: Global analysis of the L/H ratios observed in the Control and CsA-Binding experiments. Shown are the data for all the identified peptides that were successfully quantified in the Control Experiment (A), CsA Binding Experiment 1 (B), and CsA Binding Experiment 2 (C). The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value.

3.4 Results and Discussion

3.4.1 General Strategy.

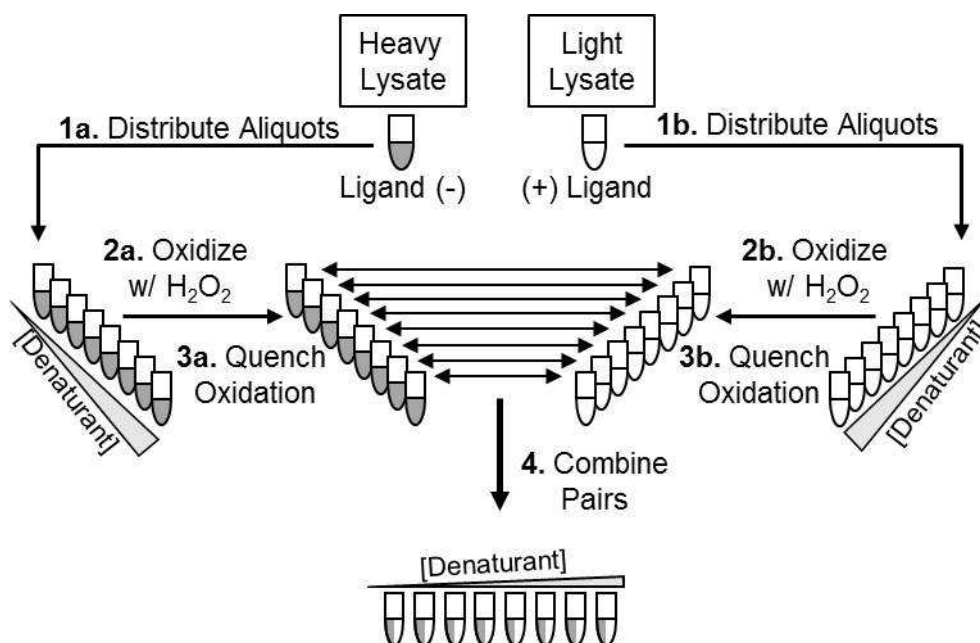


Figure 6: The SILAC-SPROX protocol.

The SILAC-SPROX protocol developed in this work is outlined in Figure 6. The strategy incorporates the SILAC metabolic labeling of proteins. As shown in Figure 6, equal amounts of light and heavy cell lysates were diluted into increasing denaturant containing buffers and were allowed to reach equilibrium in denaturant containing buffers before subjecting it to SPROX analysis. The H₂O₂ oxidation reaction was quenched with excess methionine. The heavy and light lysate samples within the same

denaturant containing buffers are combined before subjecting to bottom-up proteomic sample preparation procedures. The peptides in each of the twelve denaturants are subjected to shot-gun bottom-up proteomic analysis. The SILAC-SPROX strategy described here distinguishes itself from iTRAQ-SPROX primarily in two ways. The first is the incorporation of the metabolic labeling that allows samples to be combined immediately after the SPROX analysis. The light and heavy samples now undergo subsequent proteomic sample preparation steps potentially reducing random errors associated with separate sample handling. The second advantage is the ability to perform precursor ion (MS1) quantitation to eliminate potential chimeric peptide interference associated with product ion (MS2) quantitation.

The L/H ratios of the peptides identified in the bottom-up proteomics analysis are evaluated as a function of the denaturant concentration at which the protein oxidation reaction was performed in the SPROX analysis. The non-methionine containing peptides will have L/H ratios that are close to 1 and constant with respect to the denaturant concentration at which the protein oxidation reaction was performed. These L/H ratios from the non-methionine containing peptides are useful to confirm protein specific concentration (expression level) in each experiment. Methionine-containing peptides derived from proteins with no ligand interaction will also be close

to 1 and have constant L/H ratio at all denaturant concentrations, as their SPROX behavior in the presence and in the absence of ligand will be identical (Figure 7A). Methionine-containing peptides derived from protein hits that interact with the test ligand will have altered (e.g., high or low) L/H ratio at intermediate denaturant concentrations, as a result of their different SPROX behavior in the presence and absence of ligand (Figures 7B and 7C).

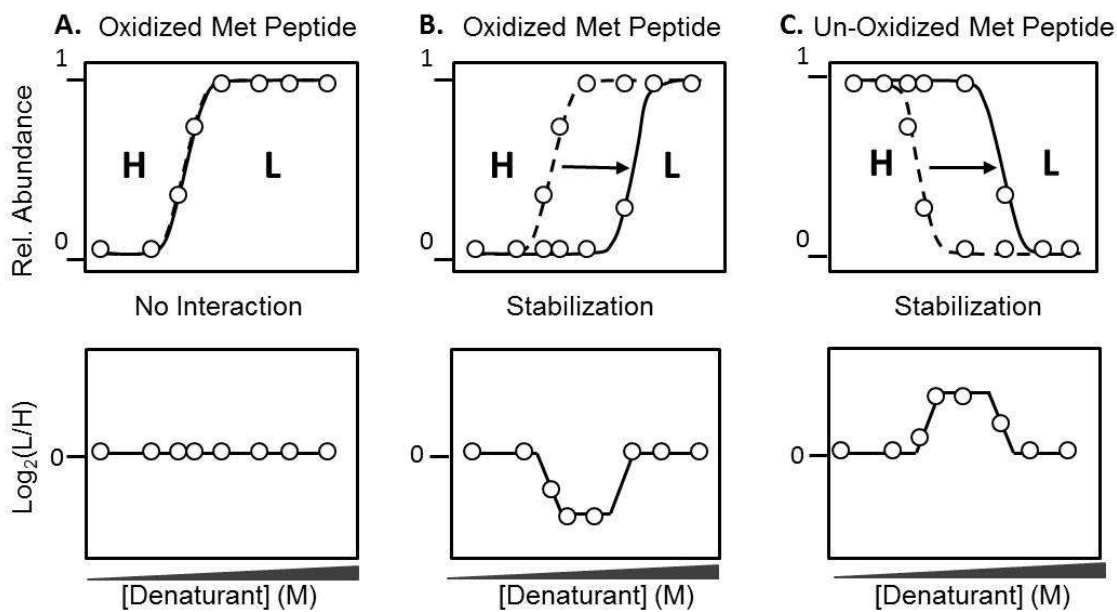


Figure 7: Expected Result (A) where there is no interaction between the protein and the ligand, SPROX behavior of the protein in the presence (solid line) and absence (dotted line) of the ligand resulting a straight line (B) if the ligand stabilize the proteins, the SPROX curve is shifted to the right resulting in a dip for the oxidized Met peptides and a peak for the un-oxidized Met peptide.

3.4.2 Identification of Cyclosporine A Binding Molecular Targets.

The experimental protocol outlined in Figure 6 was used in a proof-of-principle study to detect and quantify the direct-binding interaction between CsA and cyclophilin A in an endogenous yeast cell lysate. Besides the biological and pharmacological implications of CsA drug, the primary motivation of using this model system was to assay the ability of the protocol developed here to detect the known interactions between the drug and the yeast cell lysate. The peptidyl-prolyl cis-trans isomerase, cyclophilin A, a known direct binding target of the drug, CsA, served as the positive control in this experiment. In the CsA binding study the combined sample pairs generated in Step 4 of the protocol (see Figure 6) were submitted to a bottom-up shotgun proteomics analysis using the LC-MS/MS readout.

Table 5: Proteomic results summary in the Control and CsA-binding experiments with the proteins in a yeast cell lysate.

Experiment(s)	Total peptides (Proteins) Identified		
	Lys-Containing	Lys- and Met-Containing	Assayed for Binding
Control (No Ligand)	1471 (546)	415 (203)	119 (58)
CsA Binding Exp. 1	982 (429)	298 (144)	69 (25)
CsA Binding Exp.2	1149 (493)	328 (165)	104 (48)

Summarized in Table 5 are the proteomics results obtained from the LC-MS/MS analyses of the combined sample pairs generated in two replicate CsA-binding

experiments and in a control experiment in which no ligand was used. Approximately one-third of the total lysine containing peptides identified in the assay contained at least one methionine residue in its sequence (Table 5). A total of 69 peptides from 25 different proteins and 104 peptides from 48 proteins were assayed in the CsA binding experiment 1 and 2, respectively, and four peptide hits from two different proteins were identified.

Table 6: Peptides and proteins identified to interact with CsA.

Protein(s)	Peptide(s)	Charge state(s)	Exp.#
Peptidyl-prolyl cis-trans isomerase, Cyclophilin A	VIPDFMLQGGDFTAGN(a)GTGGK	+2,+3	1, 2
	VIPDFM(ox)LQGGDFTAGN(a)GTGGK	+2	2
Elongation factor 1- α	FVPSKPMCVEAFSEYPPLGR	+3	1
	SVEM(ox)HHEQLEQGVPGDNVGFNVK	+4	1

Representative data from the SILAC-SPROX experiments are shown in Figure 8. One non-methionine containing peptide, YAGEVSHDDKHIIVDGK, and both the oxidized and non-oxidized version of the methionine-containing peptide, VINDAFGIEEGLMTTVHSLTATQK from Glyceraldehyde-3-phosphate dehydrogenase 3 that showed no interaction with the drug CsA are shown in Figure 8A.

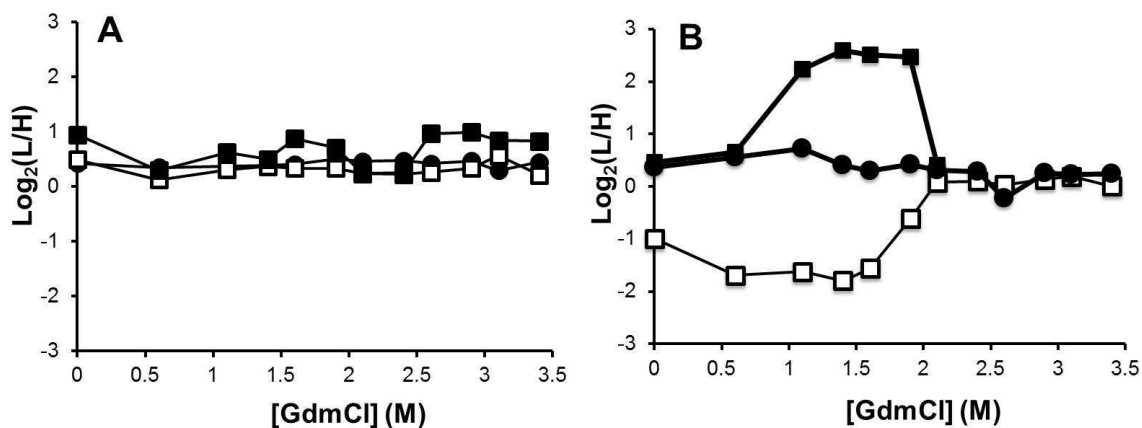


Figure 8: Representative SILAC-SPROX data obtained on proteins identified in the CsA-binding experiments. Shown are results from CsA-binding experiment 2 (A) non-hit protein, Glyceraldehyde-3-phosphate dehydrogenase 3 peptides, YAGEVSHDDKHIIVDGK (+2) non-methionine (closed circles), the oxidized and non-oxidized versions (+3 charge state) of a methionine-containing peptide VINDAFGIEEGLMTTVHSLTATQK (open and closed squares, respectively), (B) known CsA-binding protein, Cyclophilin A peptides, including the +3 charge state of the non-methionine containing peptide HVVFGEVVDGYDIVK (closed circles) and the oxidized and non-oxidized versions (albeit the +2 and +3 charge states, respectively) of a methionine-containing peptide VIPDFMLQGGDFTAGNGTGGK (open and closed squares, respectively).

The only hit protein identified in both of the CsA binding experiments was cyclophilin A. The one lysine- and methionine-containing peptide from cyclophilin A, VIPDFMLQGGDFTAGNGTGGK, that was assayed in the analyses yielded SILAC-SPROX data consistent with the known CsA-induced stabilization of this protein (Figure 8B). Both the oxidized and non-oxidized version of the VIPDFMLQGGDFTAGNGTGGK

peptide were found to have significantly altered L/H ratios (i.e., >1.7-fold change as described in the Experimental) at 4 different denaturant concentrations (Figure 8B). Two peptides, FVPSKPMCVEAFSEYPPLGR and SVEM(Ox)HHEQLEQGVPGDNLVGFNVK, from elongation factor 1 α (EF1 α) were also identified as hits with two significantly altered L/H ratios, but only in Experiment 1. The peptide, FVPSKPMCVEAFSEYPPLGR, was inconclusive because of the missing data points in the second replicate analysis. The other peptide from EF1 α , SVEM(Ox)HHEQLEQGVPGDNLVGFNVK, showed no change in the light to heavy ratio in the second CsA-binding experiment. More replicate analyses are required to confirm the interaction of these two peptides from elongation factor 1 α with CsA. The detection of cyclophilin A, the well-known target of CsA demonstrates that the protocol developed here can be used to detect and quantify molecular targets of various ligands in the context of a cell lysate.

3.4.3 False Positive Rate Determination

The proteomics results of the LC-MS/MS analyses obtained in a control experiment in which no ligand was used are shown in Table 5. Four of the 119 lysine- and methionine-containing peptides assayed in the control experiment were identified as hits (i.e., had significantly altered L/H ratios at 2 or more denaturant concentrations). This implies a false positive rate of 3.4%. The other way to assess the false positive rate

in SILAC-SPROX experiments is to ask what fraction of hit peptides would be selected from all the peptides assayed in the solution-based experiments. Such an analysis produced 13 peptide hit or (2.2%) of the 591 peptides assayed in the solution-based control experiment performed on the Q-TOF instrument. It is also possible to calculate the false positive rate by examining the data obtained on non-methionine peptides from the samples with the ligand, CsA. These non-methionine containing peptides should not exhibit significant change in L/H ratios in CsA binding experiments. Such an analysis revealed 1 of the 282 in experiment 1 and 3 out of 374 non-methionine peptides in experiment 2 showed significant change and were deemed false positives. In such analysis of CsA binding experiments using non-methionine containing peptides, the false positive rate appeared close to the false discovery rate set for the peptide identification used during bioinformatics search. Additionally, we note that false positives appear to be largely the result of random error as all false positives detected in the two solution-phase CsA binding experiments were unique.

3.5 Conclusions

The SILAC-SPROX protocol described here successfully identified known direct binding target of CsA drug, cyclophilin A protein in complex biological mixture of yeast endogenous cell lysate. The cyclophilin A binding to CsA was detected in two biological

replicate analyses performed as part of this work. The SILAC-based quantitation coupled with SPROX allowed isotopic labeling of proteins (metabolic labeling) and thus samples could be combined immediately after the SPROX analysis. The quantitation in SILAC is performed with precursor ions (MS1). This eliminated the several drawbacks in iTRAQ-SPROX protocol arising from peptide level labeling of the sample and product ion (MS2) quantitation. These improvements slightly lowered the false positive rates in SILAC-SPROX protocol to 0.4-3.4% from 1-5% previously reported for iTRAQ-SPROX experiments (64, 90). Thus, the challenges still remain as how to determine true-positives from false positives. The results of this study and others suggest that biological replicates can be used to filter out some false positives but not all. Employing a different energetics based approaches such as pulse proteolysis, have the potential to substantiate the findings of one another and to help determine which protein hits are false positives. Thus, the application of multiple energetics-based approaches can provide a much needed means by which to both corroborate protein hits and eliminate false positives in a proteome-wide ligand binding experiments and make these approaches applicable in disease state analysis. Next chapter outlines such efforts in the development of complementary energetics based approach, SILAC-pulse proteolysis.

4. The Development of a SILAC-Pulse Proteolysis Approach

4.1 Introduction

A current challenge in interpreting the results of proteome-wide protein-ligand binding analyses using existing approaches, such as SPROX and Pulse Proteolysis (PP), is the differentiation of true positives from false positives. In SPROX experiments the false positive rate of ligand discovery has been reported to be in the range of 1 to 5% (64, 65, 90). One motivation for developing the SILAC-based pulse proteolysis (SILAC-PP) protocol described here was to create a method by which results of protein-ligand binding studies (e.g., SPROX results) in complex mixtures could be cross-validated. This requires the ability to perform the pulse-proteolysis experiment directly on the proteins in complex biological mixtures, without a pre-fractionation step. As outlined earlier (see chapter 1), the proteome-wide applications of the PP technique reported (69, 94) thus far relied on the pre-fractionation of the complex lysates before running on the one- and two-dimensional gels for densitometry-based differential analyses of gel-band intensities to identify the protein targets of test ligands. Unfortunately, fractionation of the sample prior to PP limits the detection of indirect binding interactions as the full complement of proteins is not actually present in the binding assay. Indirect binding

interactions can result when the direct binding of ligand to one protein either induces or disrupts the binding of additional proteins to the protein-ligand complexes.

The incorporation of SILAC quantitation into the PP technique makes the PP approach amenable to complex mixture analyses without the need for pre-fractionation. The SILAC-PP protocol also enables the use of bottom-up, shotgun proteomics methods to identify the protein targets of ligands using the PP technique. The SILAC-PP technique described here can also be used to validate results obtained using other proteome-wide ligand binding assays such as those involving affinity capture mass spectrometry, drug affinity responsive target stability (DARTS), SPROX, or the cellular thermal shift assay (71, 73, 85, 121, 122). SILAC-PP and SPROX both report on the chemical-denaturant-induced equilibrium unfolding properties of proteins, albeit with different probes. This makes the cross validation of results obtained using these two techniques especially attractive.

The capabilities of the SILAC-PP strategy are demonstrated here in two proof-of-principle experiments in which the endogenous proteins in a yeast cell lysate are analyzed for binding to two different ligands, including a ubiquitous biological ligand (adenosine triphosphate (ATP)) and a well-characterized drug (cyclosporine A (CsA)). In the ATP binding experiments, which employed a nonhydrolyzable ATP analogue

(adenylyl imidodiphosphate (AMP-PNP)), the strategy successfully characterized the binding properties of many known ATP binding proteins. Proteins with previously unknown ATP binding properties were also detected. The detection and quantification of the well-characterized, tight-binding interaction between CsA and cyclophilin A (CypA) in four replicate experiments demonstrated the robustness of the new methodology. Furthermore, the utility of SILAC-PP as a cross-validation tool is demonstrated here by comparing the results obtained on the model systems in this study with previously reported SPROX results on the same model systems. As part of this work, a control experiment, in which no ligand was used in the analysis, was also performed in order to define the false positive rate of ligand discovery using the technique. The approach shows promise for future studies of proteome-wide protein-protein and/or ligand binding studies including diseased proteome. This work has been published in the *Journal of American Society for Mass Spectrometry*².

² Adhikari, J., and Fitzgerald, M. C. (2014) SILAC-Pulse Proteolysis: A Mass Spectrometry-Based Method for Discovery and Cross-Validation in Proteome-Wide Studies of Ligand Binding, *Journal of the American Society for Mass Spectrometry* 25, 2073-2083.

4.2 Experimental

4.2.1 Yeast Cell Lysate Preparation

The SILAC labeled yeast cells were cultured as described previously (119). Briefly, a glycerol stock of the yeast deletion strain BY4739 (Open Biosystems, Lafayette, CO), an auxotroph for lysine, was streaked on a petri-dish containing synthetic complete (SC) media with Difco™-agar and L-lysine. The Agar-media for the plate comprised of 1.7 g of yeast nitrogen base (Amresco), 5 g of ammonium sulfate (Sigma-aldrich), 20 g of Difco™-agar (BD, USA), 2 g of glucose (Sigma), 1.92 g of synthetic drop out mix – lysine (Sunrise Science Product, San Diego, CA), and 30 mg of L-Lysine all in 1 L media. The SC-media for the subsequent culture was comprised of 1.7 g of yeast nitrogen base, 5 g of ammonium Sulfate, 20 g of glucose, 1.92 g of synthetic drop out mix – lysine (Sunrise Science Product, San Diego, CA). After 3 days of growth at 30 °C, an isolated colony was selected and inoculated into 10 mL of SC-media containing 30 mg/L of light L-lysine (Sigma Aldrich), hereafter referred to as light SC-media. The 10 mL culture was incubated overnight at 30 °C. A 100 µL portion of the cell culture was transferred into 100 mL of the light SC-media. Simultaneously, another 100 µl of the cell culture was transferred into 100 ml of SC-media containing 30 mg/L of heavy Lysine (¹³C₆, ¹⁵N₂) (Sigma Aldrich and Cambridge isotopes laboratories, MA), hereafter referred to as

heavy SC-media. The 100 mL culture was incubated overnight at 30 °C. The Light and Heavy cell cultures (100 µL portions) were transferred two more times into 100 mL portion of the corresponding light and heavy SC-media and each time grown overnight at 30 °C, before they were ultimately transferred to 1 L of the corresponding light and heavy SC-media and grown overnight at 30 °C. Portions of the 1 L overnight culture, 250 mL each, were centrifuged at 1000 x g to generate light and heavy lysine labeled cell pellets. The cell pellets were stored at -20 °C.

The light and heavy lysine-labeled yeast cell lysates used in the ligand binding studies were prepared using the above cell pellets. To each pellet was added 400 µL of 100 mM Tris-HCl buffer (pH 8.0) containing 50 mM NaCl, 10 mM CaCl₂, and a cocktail of protease inhibitors that included 1 mM AEBSF, 500µM Bestatin, 15 µM E-64, 20 µM Leupeptin and 10 µM Pepstatin A (Thermo pierce) to generate heavy and light cell lysates. The cells in each pellet were lysed by mechanical disruption using a disruptor genie (Scientific Industries) and 425-600 µm diameter glass beads. A total of 15 cycles consisting of mechanical disruption for 25 s and cooling on ice for 1 min were used to lyse the cells in each experiment. Ultimately, the lysed cells were centrifuged at 14-15000 X g for 15 min at 4°C, and the supernatants were used in the SILAC-PP analyses. The concentrations of total protein in the supernatants were determined using the Bradford

assay. The total protein concentration in the supernatants typically ranged from 6-8 mg/ml (except for in ATP-binding Experiment 1 and CsA-binding Experiment 2 where the total proteins concentrations were 18.1mg/ml and 3.5mg/ml, respectively). The protein concentrations in the light and heavy labeled lysate sample pairs were normalized by diluting the higher concentration lysate with an appropriate amount of lysis buffer before proceeding to the SILAC-PP analysis. The exact protein concentrations and experimental conditions used in each experiment described here are summarized in Appendix A.1-A.3.

4.2.2 SILAC-Pulse Proteolysis Analysis

In each of the experiments in this study the ligand of interest was added to the heavy cell lysate. In the ATP binding experiments, 8 volumes of each of the heavy and the light yeast cell lysates were mixed with 1 volume of a 1 M MgCl₂ (Sigma) solution followed by the addition of 1 volume of lysis buffer to the (-) ligand sample and 1 volume of an 823 mM adenylyl imidodiphosphate (AMP-PNP) (Sigma) solution to the (+) ligand samples. In the CsA binding experiments, 9 volumes of the heavy yeast cell lysate were mixed with 1 volume of DMSO and 9 volumes of the light yeast cell lysates were mixed with 1 volume of a 10 mM CsA (Sigma) solution prepared in DMSO in order to generate stock solutions of the (-) and (+) ligand samples, respectively.

Similarly, 9 volumes of the heavy and light yeast cell lysate were each mixed with 1 volume of lysis buffer solution to generate the control samples. The stock solutions of the (-) and (+) ligand samples were equilibrated for 1-2 hours on ice or at room temperature. More detailed information about the conditions and the amounts of reagents used in each experiment are summarized in Appendix A.1-A.3.

The PP reaction conditions were similar to those previously described (91, 94). Briefly, 25 μ l aliquots of the (-) and (+) ligand samples were each combined with 75 μ l of denaturant-containing buffer stock solutions, which were prepared in 100 mM Tris-HCl buffer (pH 8.0) with urea concentrations ranging from 0 to 9 M. The exact concentration of urea in each denaturant-containing buffer stock solution was determined from a refractive index measurement of the buffer as described previously (123). The final urea concentrations in the denaturant-containing buffers in these experiments ranged from 0 to 6.9 M. The samples in the denaturant-containing buffers were equilibrated for ~18 hours at room temperature. The proteolysis reaction was initiated by adding thermolysin protease from *Bacillus thermoproteolyticus rokko* (Sigma) to the protein and protein-ligand samples in each denaturant-containing buffer. The protease to protein ratio was approximately 1:10 (w/w), except in the first ATP-binding experiment in which the ratio was 1:35 (w/w). The proteolysis reactions were allowed to proceed for 1 min

before each reaction was quenched with the addition of an EDTA solution (pH 7.9-8.0) (Thermo scientific and Boston Bioproducts, Boston, MA) (see Appendix A.1-A.3).

Equal volumes of the heavy and light samples with or without the ligand in buffers containing the same denaturant concentration were combined. Aliquots (50-200 μ L) of the combined sample pairs were mixed with 10-40 μ L of a 6X stock solution of Laemmli sample buffer, which was comprised of 375 mM Tris•HCl buffer (pH 6.8) containing 6% SDS, 50% Glycerol and 0.045% Bromophenol (Boston Bioproducts, Boston, MA) and 6-7 % of the 14.3 M β -mercaptoethanol (BME) solution (Aldrich). The samples were mixed, centrifuged at 2000 \times g for 20 seconds, and heated at 95°C for ~5 mins. Approximately, 20-45 μ l, or ~50 μ g of total protein, of the protein samples in the denaturant-containing buffers, were loaded on to a 4-20% midsize polyacrylamide gel (BioRad Criterion, Hercules, CA). An aliquot of the Precision Plus standard marker (BioRad, Hercules, CA) was also loaded in the wells at the two ends of the gel.

The SDS-PAGE gels were fixed upon treatment with a solution of 25% Isopropanol, 10% acetic acid and 65% water for approximately 25 minutes. The gels were stained overnight with a staining solution that contained 0.01% R-250 (Bio-Rad) in 10% acetic acid (EMD Chemicals Inc.). Excess dye was removed by treatment of the gels with a 10% acetic acid solution. The protein bands across the twelve lanes (ten in ATP

experiment 1) in each of the gels were excised in specific molecular weight regions and processed using a standard in-gel digestion protocol as previously described (124). Trypsin was used as the protease, even though the yeast cells were only labeled with heavy lysine. Our use of trypsin generated a number of non-lysine containing peptides that could not be used in the quantitative LC-MS/MS analyses (see below). This ultimately limited the number of proteins that were effectively assayed in our experiments (see below). In theory, the proteomic coverage obtained here could have been increased if an enzyme like Lys-C was used or the yeast cells were labeled with both heavy lysine and arginine. Trypsin and lysine labeling were used in the current proof-of-principle work in order to minimize the cost of this proof-of-principle experiment. The molecular weight regions of the gel bands excised in the ATP binding experiments were 25-34 kDa and 70-85 kDa. The gel bands excised in CsA binding experiment were in the 15-20 kDa and 50-60 kDa regions. The gel bands excised in the Control experiment were in the 15-20 kDa, 25-34 kDa, 50-60 kDa and 70-85 kDa regions.

4.2.3 LC-MS/MS Analyses

The peptide samples extracted from the gel-bands generated in the ligand binding studies and in the control experiments were each subjected to an LC-MS/MS analysis using an Agilent 6520 Q-TOF mass spectrometer equipped with a Chip Cube

interface. The HPLC chip used in the SILAC-PP experiment was a ProtID-Chip43(II) that consisted of a 40 nl enrichment column and a 75 μm X 43 mm analytical column packed with Zorbax 300SB-C18 5 μm packing material. Approximately 1-5 μg of the samples were loaded onto the column and the tryptic peptides were eluted using a linear gradient from 3 to 5% solvent B (ACN with 0.1% formic acid (FA)) for first two minutes, 5 to 15% B for next two minutes, 15 to 60% B for another 18 minutes, 60 to 90% B for next 3 minutes. The column was flushed with 100% solvent B for next two minutes and equilibrated with 5% solvent B for additional three minutes. The flow rate was 0.4 $\mu\text{l}/\text{min}$. The capillary voltage ranged from 1725 to 1925 V. The flow rate of the drying gas was set at 4 L/min at 350°C. The skimmer and fragmentor were set at 65 and 175 V, respectively. The collision energy was as determined by the equation 3.5 V/100 Da with a -4.8 V offset. The inclusion window width for precursor ions was 4 m/z. The scan rate was three scans per second in the precursor mass spectra and two scans per second in the product ion spectra. In every cycle, four precursors were selected for fragmentation. All the spectra were collected in the profile mode.

The LC-MS/MS data generated in the ATP- and CsA-binding experiments and in the control experiments described here were searched using Agilent's Spectrum Mill MS Proteomics Workbench software, B.04.00. The search parameters used for peptide and

protein identifications included: SILAC heavy labeled lysine 8 mix and cysteine modification with carbamidomethylation were set as fixed modifications. Oxidation of methionine and deamidation of asparagine (N) were set as variable modifications. The precursor and product ion mass tolerances were 20 and 50 ppm, respectively. The maximum product ion charge was set to 7, and the protein cleavage chemistry was set for trypsin with a maximum of three missed cleavages. All the mass spectral data searches used the 6619 *Saccharomyces cerevisiae* proteins in the UniProtKB/Swiss-Prot database (uniprot_sprot_fasta.gz Release 2012_01/ modified on 2/22/12 downloaded on 2/28/2012). The identified peptides and proteins with non-zero positive light to heavy ratios were used in the subsequent data analysis steps. In cases where the proteins were identified with multiple accession numbers, the first accession number reported was used in the analysis.

4.3 Data Analysis

The L/H ratios generated for all the peptides in a given experiment were used to determine a median L/H ratio for each experiment (Figures 9 and 10). The median L/H ratios for all the experiments were between 0.8 and 1.2, with the exception of CsA Binding Experiment 3 where the median L/H ratio was 1.6 (Figures 9 and 10). The L/H ratios of all peptides from a given protein identified in each gel-band (i.e., at each

denaturant concentration) were averaged to generate a single L/H ratio for each protein at each denaturant concentration. Ultimately, these average L/H ratios were used to generate SILAC-PP data sets (i.e., L/H ratio versus [Denaturant] plots). Only those proteins that were identified in at least half the total number of denaturant concentrations used in an experiment were assayed for potential changes in thermodynamic stability due to ligand binding and for any false positives in the control experiments. The hit proteins in each experiment were initially identified as those proteins with >1.7 fold deviations from the median L/H ratio, at two or more consecutive denaturant concentrations.

A global analysis of the L/H ratios obtained in the two control experiments revealed that ~90% of the ratios were within 1.7-fold of the median values (see Figure 9A and B). Based on this global analysis and the above requirement that hit peptides must have significantly altered L/H ratios at two or more consecutive denaturant concentrations, the estimated p-value associated with each hit peptide in this work was <0.01. Ultimately, the SILAC-pulse proteolysis data sets from the initial list of hit proteins were visually inspected to confirm that each protein hit was identified with at least two or more peptides in one or more of the gel-bands. Protein hits that were identified with only one unique peptide sequence were removed from the final hit list.

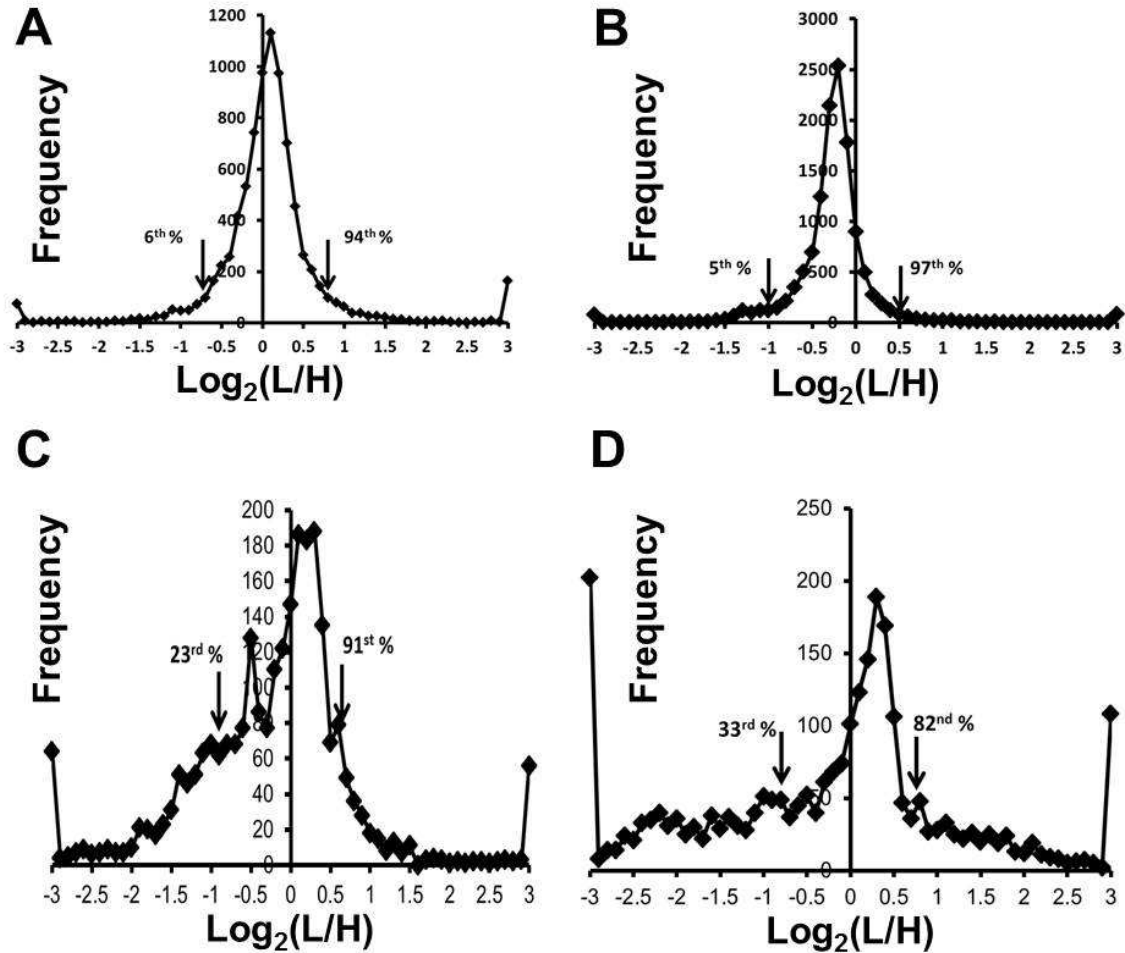


Figure 9: Global analysis of the L/H ratios determined for peptides identified and successfully quantified in (A) Control Experiment 1, and (B) Control Experiment 2, (C) the ATP Experiment 1 and (D) the ATP Experiment 2. The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value. The median $\text{Log}_2(\text{L}/\text{H})$ values in (A) and (B) were 0.03 and -0.29, respectively. The median $\text{Log}_2(\text{L}/\text{H})$ values in (C) and (D) were -0.11 and -0.02, respectively.

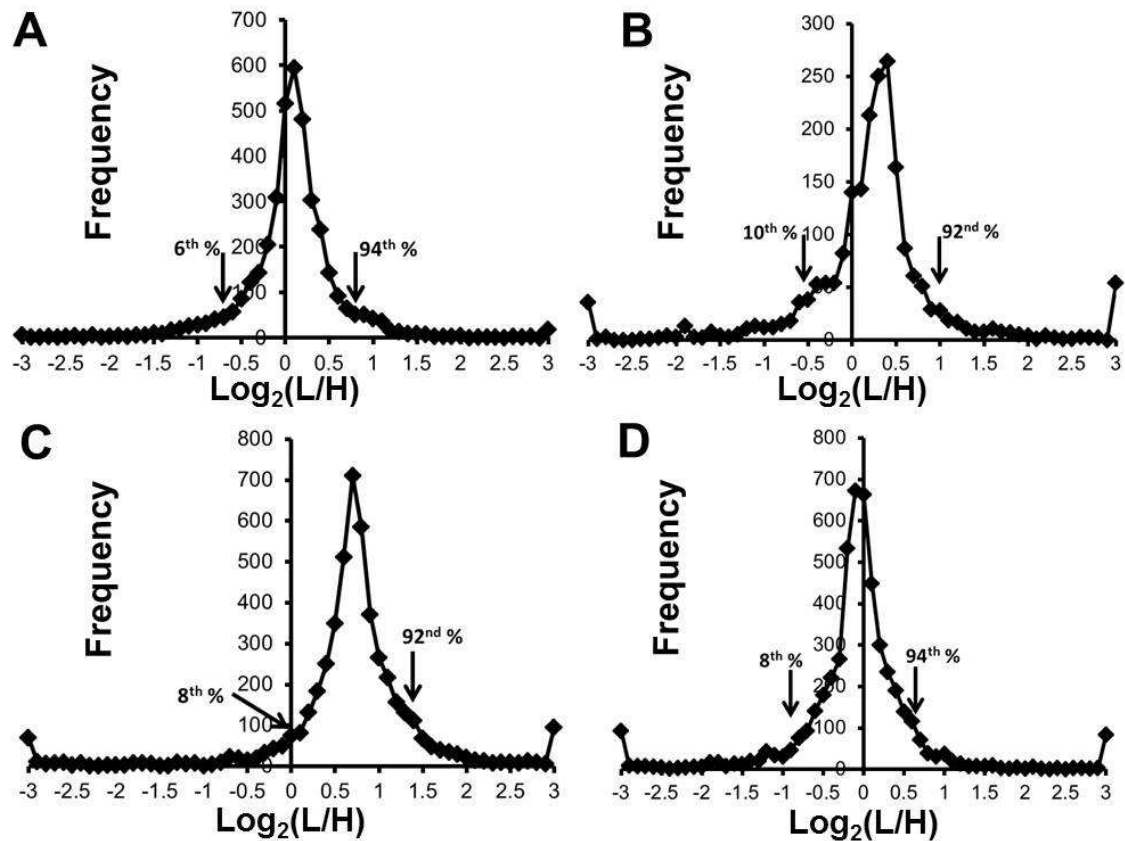


Figure 10: Global analysis of the L/H ratios observed in the CsA experiments in SILAC-PP. Shown are the distributions of the $\text{Log}_2(\text{L}/\text{H})$ for all the identified peptides that were successfully quantified in (A) the CsA Experiment 1, (B) the CsA Experiment 2, (C) the CsA Experiment 3, and (D) the CsA Experiment 4. The arrows indicate the $\text{Log}_2(\text{L}/\text{H})$ values that are 1.7-fold away from the median. Noted above each arrow is the percentile associated with the indicated $\text{Log}_2(\text{L}/\text{H})$ value. The median $\text{Log}_2(\text{L}/\text{H})$ values in (A), (B), (C), and (D) were 0.04, 0.23, 0.67, and -0.11, respectively.

4.3.1 Quantitation of Ligand Binding

Dissociation constants (K_d values) were calculated in an analogous fashion to that which we have described in SILAC-SPROX (65). Briefly, the L/H versus [urea] plot obtained for each hit proteins was visually inspected to generate a $\Delta C^{1/2}$ value (see below). The $\Delta C^{1/2}$ value was used in Equation 1 to calculate the change in the folding free energy, $\Delta\Delta G_f$.

$$\Delta\Delta G_f = -m \times \Delta C_{SILAC-PP}^{1/2} \quad \text{Equation 1}$$

In Equation 1, m is the change in the protein folding free energy with respect to change in denaturant concentrations. The m -value used in the cyclophilin A-CsA K_d value calculations was $1.675 \text{ kcal mol}^{-1}\text{M}^{-1}$, which is half the average of two experimentally determined m -values (3.0 and $3.7 \text{ kcal mol}^{-1}\text{M}^{-1}$) that were previously reported in chemical denaturant-induced equilibrium unfolding experiments on CypA using GdmCl (125, 126). This estimate is based on the data in Myers *et. al.* in which it was determined that m -values in chemical denaturation experiments using urea are generally half those in experiments using GdmCl (127). An estimated m -value of $1.3 \text{ kcal mol}^{-1}\text{M}^{-1}$ was used for all other proteins. This estimate was based on the assumption that the average protein folding domain contains ~ 100 amino acids and that the

magnitude of the average m -value in urea is 0.013 kcal mol⁻¹ M⁻¹ per amino acid residue. This per residue value was empirically derived by Myers et. al (127).

Ultimately, $\Delta\Delta G_f$ values were used in Equation 2 to calculate a K_d value for each protein that yielded L/H versus [urea] data consistent with a ligand-induced stabilization.

$$K_d = \frac{[L]}{(e^{-\Delta\Delta G_f/nRT} - 1)} \quad \text{Equation 2}$$

In Equation 2, [L] is the molar concentration of the free ligand, n is the number of independent binding sites, R is the universal gas constant, and T is the temperature in Kelvin. In all K_d value calculations the number of binding sites was assumed to be 1, and the total ligand concentration (20.6 mM and 250 μ M in the ATP and CsA binding experiments, respectively) was used as the free ligand concentration.

4.4 Results and Discussion

4.4.1 SILAC-Pulse Proteolysis Protocol

The SILAC-PP protocol developed here (Figure 11) involves the use of SILAC labeled cell lysates with light and heavy amino acids. The ligand of interest is spiked into either the heavy or the light labeled lysate. In this study the ligand was spiked into

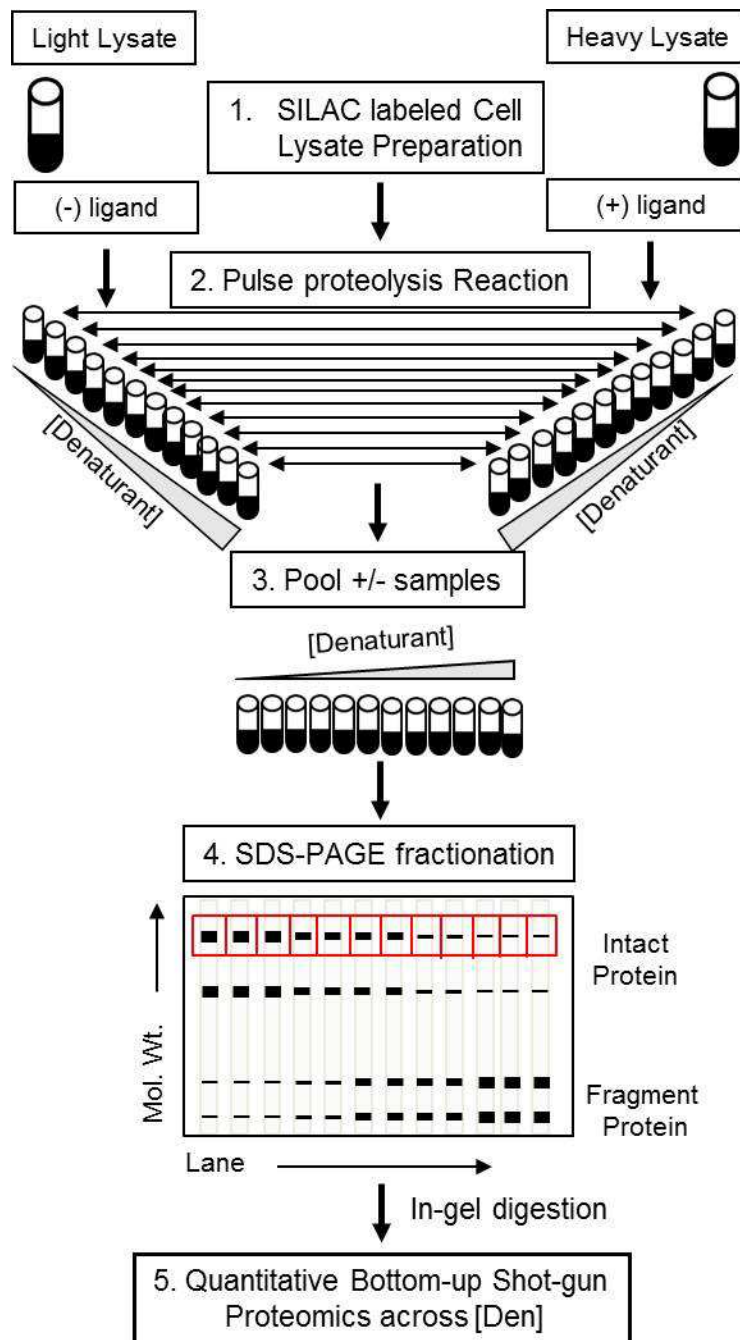


Figure 11: SILAC-pulse proteolysis protocol.

the heavy lysate. The light and heavy labeled lysate samples are each distributed into a series of denaturant containing buffers. The protein and protein-ligand samples in the denaturant-containing buffers are incubated overnight, and treated with thermolysin for a minute. While the catalytic activity of thermolysin is diminished at higher urea concentrations, the enzyme concentration and reaction time used in the experiment are generally sufficient to digest unfolded proteins(91). Thus, the extent of proteolytic cleavage in each denaturant-containing buffer is largely determined by the fraction of unfolded protein. The proteolysis reaction with themolysin is quenched upon addition of excess EDTA solution. The protein and protein ligand samples in the same denaturant concentrations are combined together and each of the combined samples is fractionated using SDS-PAGE (Figure 11). The gel is sectioned into a series of gel-bands (Figure 11), and a quantitative bottom-up shotgun proteomics analysis is used to identify and quantify the relative amounts of protein from the light and heavy cell lysate in each gel-band. Ultimately, the L/H ratios of the proteins identified in a given row of the gel (i.e., specific molecular weight range) are analyzed to determine which proteins have a $\Delta C^{1/2}$ value indicating an interaction with the ligand (Figure 12).

Proteins that do not interact with ligand will be digested with the thermolysin to the same degree in both the light and heavy labeled lysates at each denaturant

concentration. Therefore, the L/H ratios recorded for these non-interacting proteins will be ~1 and unchanged for the samples from the different denaturant concentrations (see Figure 12A).

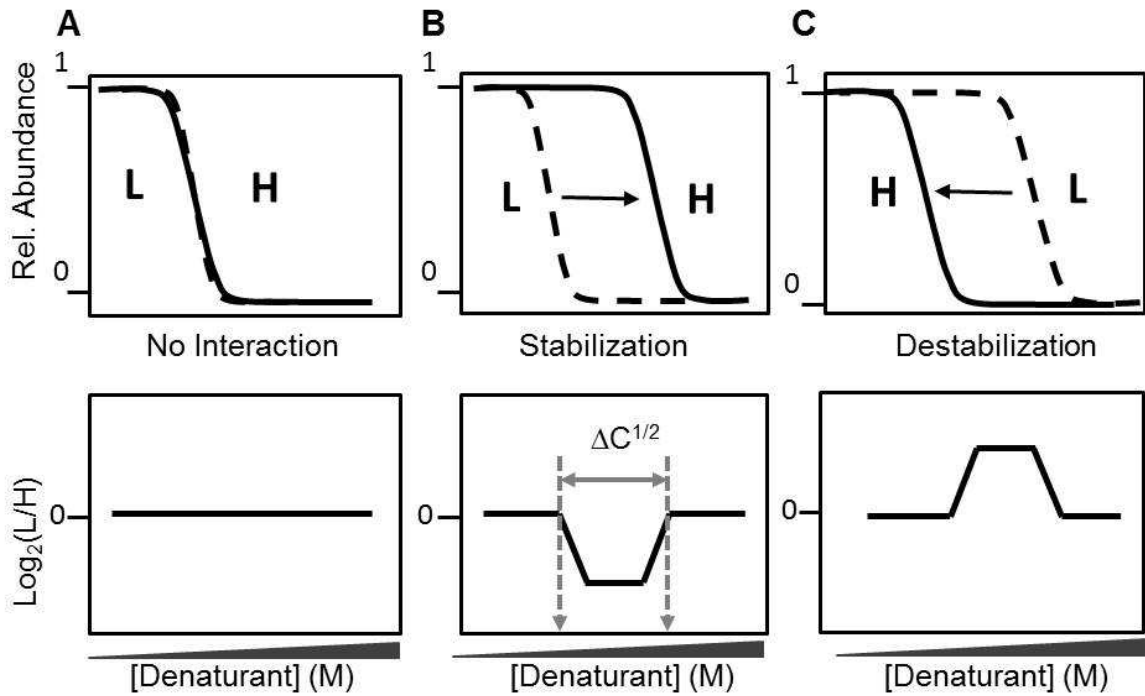


Figure 12: Expected Result for the intact protein readout (A) No interaction between the protein and the ligand, behavior of the protein in the presence (solid line) and absence (dotted line) of the ligand resulting a straight line (B) ligand induced stabilization of the proteins causes the pulse proteolysis curve to shift to the right resulting in a dip and (C) ligand induced destabilization that shifts the curve to the left resulting in a peak.

Proteins that interact with the ligand will have an altered resistance to proteolysis depending on the nature of the interaction (e.g., whether or not the ligand

binding event is stabilizing or destabilizing). The L/H ratios recorded for all the peptides derived from the intact protein bands of such interacting proteins will depart from 1 at intermediate denaturant concentrations where the chemical denaturation curves of the (-) and the (+) ligand samples differ (see Figure 12B and C). In cases where there are allosteric effects associated with ligand binding the chemical denaturant dependence of the L/H ratios may be more complicated. In general, the L/H ratios of all the peptides derived from the intact protein band should have the same chemical denaturant dependence, regardless of the complexity. The observed L/H ratios in SILAC-PP should also report on the net stabilization of the protein in the presence of ligand. Some allosteric effects (e.g., the case where a ligand binds with high affinity to a specific domain, but the overall stabilization of the protein is small) may obscure protein-ligand binding detection by SILAC-PP.

4.4.2 Control Experiment: False Positive Rate Determination

A control experiment in which no ligand was spiked into either the light or the heavy lysate was performed in order to assess the false positive rate of hit protein discovery using the SILAC-PP. Shown in Figures 13 and Appendix A.4A are the four molecular weight regions of the gel-bands excised in both the biological replicate of the control experiment performed here. Gel-bands in similar molecular weight regions were

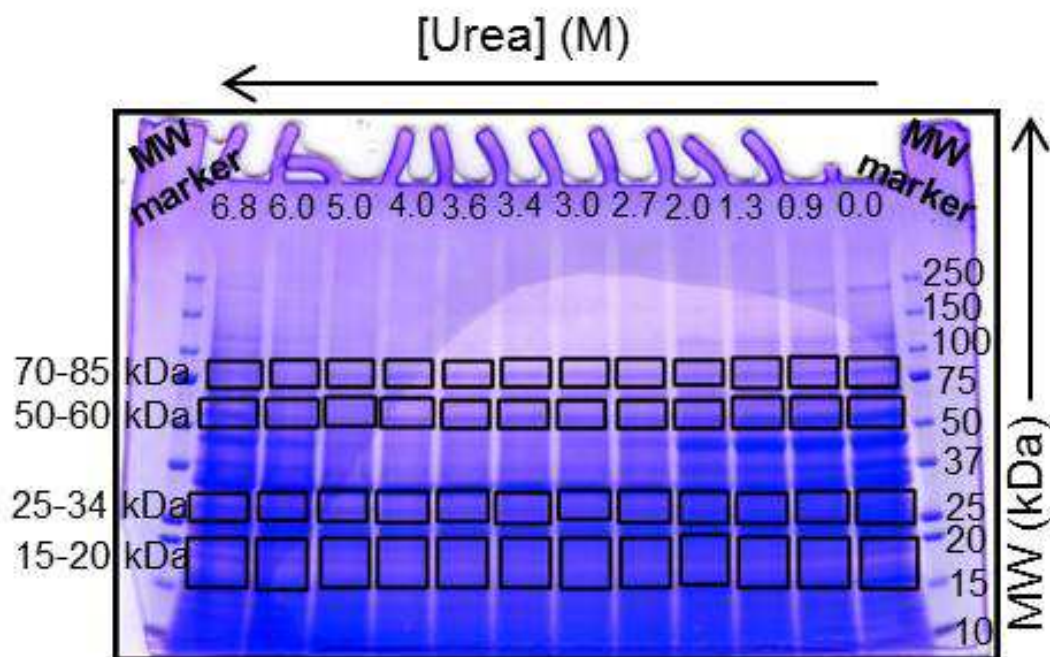


Figure 13: SDS-gel image for SILAC-PP control experiment 2. The gel regions excised in this experiment are indicated with rectangular boxes.

excised in each biological replicate of the control experiment. The proteomic results obtained in the two biological replicates of the control experiment are summarized in Table 7. A total of 4 out of the 110 proteins effectively assayed in Control Experiment 1 and 3 out of the 141 proteins effectively assayed in Control Experiment 2 (Table 8) yielded L/H ratios consistent with hit behavior in our analysis. These results suggest the false positive rate of hit protein discovery using SILAC-PP is in the range of 2.1 to 3.6%. That is, for every 100 proteins assayed (i.e., successfully identified and effectively quantified in samples from 4 or more denaturant-containing buffers) 2-3 proteins will

appear as false positives. Also, there was no overlap of the hit proteins in the two biological replicates. This suggests that the false positives result from random errors associated with SILAC quantitation.

Table 7: Summary of the peptide and protein coverage with the protein hits observed in the SILAC-Pulse-proteolysis experiments performed in this work.

Experiment(s)	Total Peptides (Proteins) Identified		Total Peptides (Proteins) Assayed ¹		Total Protein Hits Identified	
	17kDa	54kDa	17kDa	54kDa	17kDa	54kDa
CsA Binding						
1	263 (134)	205 (82)	170 (52)	125 (16)	2	1
2	142 (95)	154 (67)	57 (21)	79 (7)	2	2
3	529 (295)	197 (81)	266 (64)	89 (6)	3	0
4	336 (196)	410 (199)	141 (35)	157 (16)	2	0
Control Exp.1	274 (154)	197 (95)	148 (46)	105 (13)	2	0
Control Exp. 2	402 (203)	388 (143)	242 (63)	194 (19)	1	1
ATP Binding	25kDa	80kDa	25kDa	80kDa	25kDa	80kDa
1	429 (166)	119 (65)	239 (40)	38 (8) ²	18	4
2	535 (266)	210 (99)	174 (23)	58 (7)	18	5
Control Exp.1	366 (173)	307 (124)	203 (39)	144 (12)	1	1
Control Exp. 2	432(177)	354 (108)	259 (38)	238 (21)	1	0

¹Includes the proteins that were successfully identified and quantified in samples from at least half of the denaturant concentrations. ²Includes proteins that were identified and quantified in samples from 4 or more of the 10 denaturant concentrations.

Table 8: False Positives from SILAC-PP Control Experiment.

Protein(s)	Protein MW	Fragment /Intact	Exp.# (Gel bands)
40S ribosomal protein S17-A	15845.1	Intact	1 (17kDa)
60S ribosomal protein L26-A	14233.4	Intact	1 (17kDa)
6-phosphogluconolactonase 3	28012.2	Intact	1 (25kDa)
Elongation factor 3A	116847	Fragment	1 (54kDa)
UPF0592 protein YDL073W	114802	Fragment	2 (17kDa)
GAPDH 3	35860.1	Fragment	2 (25kDa)
Pyruvate decarboxylase isozyme 1	61722.5	Fragment/Intact?	2 (54kDa)

4.4.3 SILAC-PP: A Tool for Protein Target Discovery

The SILAC-PP strategy was used in a proof-of-principle study to identify yeast protein targets of ATP using AMP-PNP, a nonhydrolyzable ATP analogue and the well-studied immunosuppressant drug, CsA, with its known direct binding target, CypA, in the context of the yeast cell lysate from *S. cerevisiae*. The goal of this study was to demonstrate the protocol's ability to discover and cross validate (see below) the protein targets of selected ligands, and not to perform an exhaustive search for new protein targets. Thus, only specific molecular weight regions of the gel bands were excised. In the ATP experiments, two molecular weight regions of the gel, 25-34 kDa and 70-85 kDa, were analyzed (Figure 14 and Appendix A.4 E). The ~80 kDa size range was chosen to specifically test the protocol's ability to detect the already well-known ATP binding interaction with ATP-dependent heat shock protein, HSC82p. The ~25 kDa size range

was specifically chosen to cross-validate the recently reported ATP binding properties of phosphoglycerate mutase (65, 94)

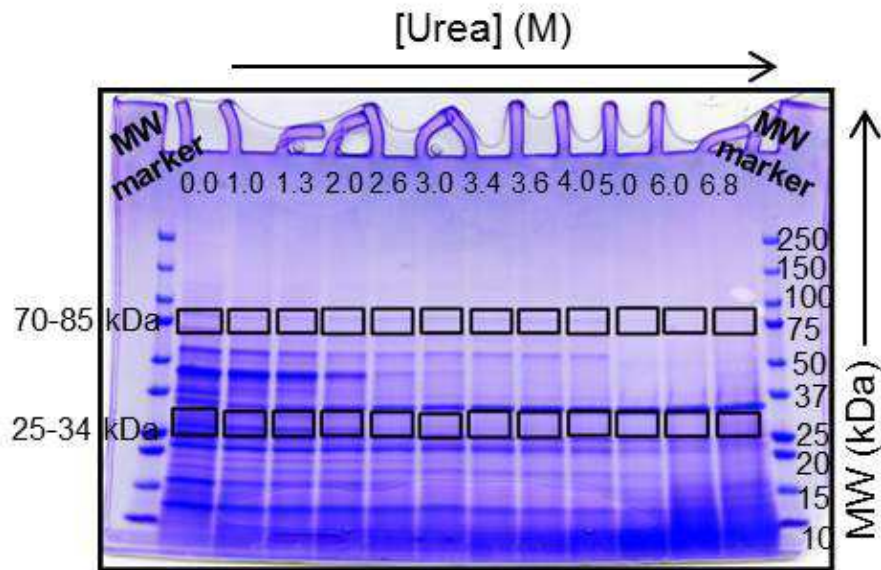


Figure 14: SDS-PAGE gel image ATP-binding experiment 2. The gel regions excised in this experiment are indicated with rectangular boxes.

The described protocol successfully identified a total of 33 unique protein hits to have ATP-binding interactions in the two gel regions that were analyzed (see Table 7 and Table 9). Ten of the protein hits were consistently detected in both of the ATP-binding experiments performed here. Nine of these 10 consistently identified protein targets are known to bind ATP based on a GO-term analysis and/or the results of other experiments (Table 9).

The 23 protein hits that were not consistently detected as hits in both of the ATP-binding experiments performed here using SILAC-PP were only assayed in one experiment (i.e., they were not identified in enough denaturant concentrations in the other experiment). Of these 23 protein hits, 12 are known ATP binding proteins based on GO-term analyses and/or the results of other experimental approaches (see Table 9) (65, 94, 122). Together a total of 21 of the 33 proteins are known to bind to ATP based on GO-term analyses and/or the results of previous studies.

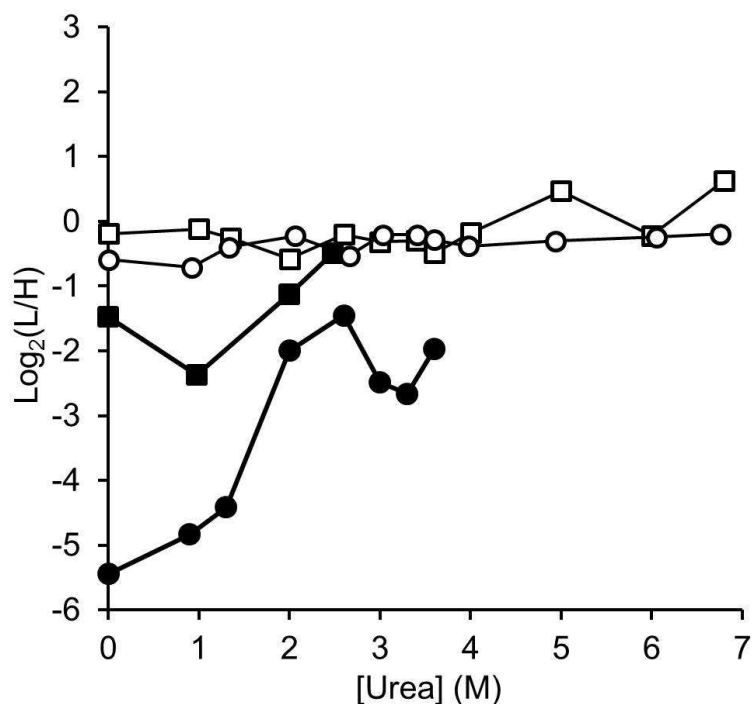


Figure 15: SILAC-PP data for the ATP-dependent heat shock protein, HSC82p, obtained from intact protein band, experiment 1 (solid squares) and experiment 2 (solid circles). Also shown are the data from control experiment 1 (open squares) and experiment 2 (open circles). The $\Delta C^{1/2}$ values extracted was 2.5 and 2.6 M for ATP-binding experiment 1 and 2 respectively.

SILAC-PP results obtained on one of the proteins known to bind ATP, the ATP-dependent heat shock protein HSC82p, are shown in Figure 15. The K_d values for ATP binding to HSC82p were calculated using the observed $\Delta C^{1/2}$ value (see Figure 15) and Equations 1 and 2 above. The K_d values calculated for the intact HSC82p protein band in this study, 89 μ M and 69 μ M in ATP-binding Experiment 1 and 2 (respectively), were in reasonable agreement with the previously reported value of 132 \pm 47 μ M (128).

Table 9: ATP-binding proteins identified in SILAC-PP experiments.

Protein(s)	Exp# (Gel Bands)	Binds ATP?*
V-type proton ATPase catalytic subunit A	1,2 (25kDa)	1
Tryptophan synthase	2 (25kDa)	-
Transketolase 1	1,2 (80kDa)	2
Transaldolase	1 (25kDa)	-
Pyruvate kinase 1	1 (80kDa), 2(25kDa)	1,4
Pyruvate decarboxylase isozyme 1	2 (25kDa)	4
Protein NUF1	1 (25kDa)	-
Protein BMH1	2 (25kDa)	2
Phosphoglycerate mutase 1	1,2 (25kDa)	3,4
Kinetochores-associated protein DSN1	2 (25kDa)	-
Heat shock protein SSA1	2 (25kDa)	1,4
Glycyl-tRNA synthetase 1	2 (80kDa)	1,4
Glyceraldehyde-3-phosphate dehydrogenase 1	1,2 (25kDa)	3,4
Glucose-6-phosphate isomerase	2 (25kDa)	-
Enolase 1*	1,2 (25kDa)	4
Elongation factor 3A	1 (80kDa)	1,4
Elongation factor 2	1,2(25kDa),2(80kDa)	3,4
C-1-tetrahydrofolate synthase, cytoplasmic	2 (80kDa)	1
ATP-dependent molecular chaperone HSC82	1,2(80kDa),2(25kDa)	1,4
Alcohol dehydrogenase 1*	1,2 (25kDa)	4
Adenosylhomocysteinase	1 (25kDa)	4
Actin	2 (25kDa)	1,4
Acetyl-CoA acetyltransferase	2 (25kDa)	-
60S ribosomal protein L7-A	1 (25kDa)	-
60S ribosomal protein L4-A	1 (25kDa)	-
60S ribosomal protein L3*	2 (25kDa)	4
60S ribosomal protein L2	1 (25kDa)	-
60S ribosomal protein L15-A	1 (25kDa)	-
40S ribosomal protein S5	1 (25kDa)	-
40S ribosomal protein S4	1,2 (25kDa)	-
40S ribosomal protein S3*	1 (25kDa)	4
40S ribosomal protein S2*	1 (25kDa)	4
40S ribosomal protein S1-A*	1 (25kDa)	4

** (1) Known to have GO-Term "ATP binding", (2) Reference (122), (3) Reference (70), (4) Reference(65) *Overlapped with results from SILAC-SPROX(65) that are novel ATP-binding, previously not reported.

In the ATP binding SILAC-PP experiments we also investigated the 22 proteins that were assayed and determined not to be hits. Three of these 22 are known to bind ATP based on the GO term Analysis. All three were assayed in only one experiment in which they were only identified in bands much lower than their intact MW.

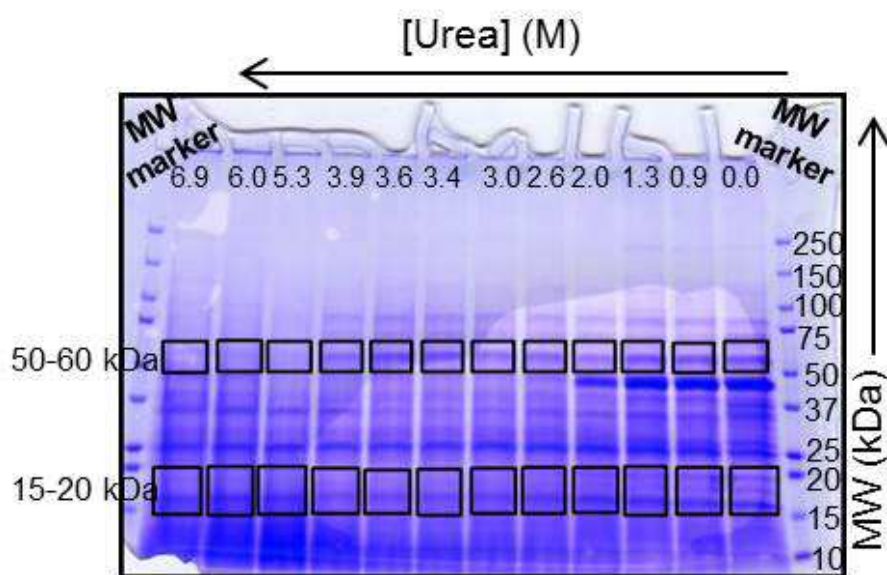


Figure 16: SDS-gel image for CsA-binding experiment 2. The gel regions excised in this experiment are indicated with rectangular boxes.

The CsA-binding experiments also involved the analysis of two molecular weight regions of the gel, the 15-20 kDa and 50-60 kDa regions (Figure 16 and Appendix A.4 B-D). The lower molecular weight region was chosen to include CypA, which has a well-known binding interaction with CsA. The higher molecular weight region was chosen to include pyruvate kinase, which was previously identified as a CsA protein

target in an iTRAQ-SPROX experiment (73). A total of eight protein hits were identified in the four different SILAC-PP experiments involving CsA and the yeast cell lysate (Table10). Only 1 of the 8 proteins, CypA, was consistently identified as a hit in all four experiments (Figure 17).

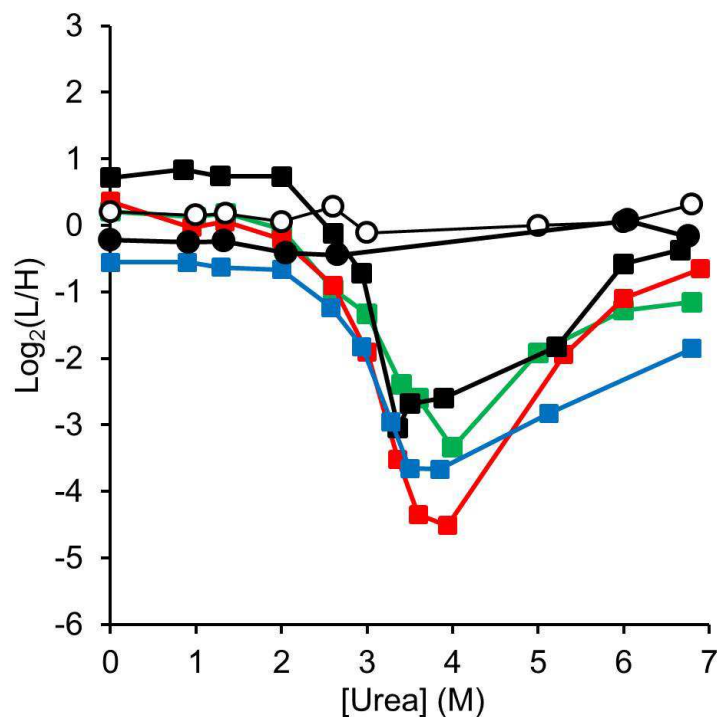


Figure 17: SILAC-PP data for the Cyclophilin A, CypA, obtained from intact protein band, solid green, red, black and blue squares in CsA-binding experiments 1, 2, 3 and 4 respectively. Empty and solid circles represent data on the intact CypA protein from control experiments 1 and 2 respectively. The $\Delta C^{1/2}$ values extracted were 4, 4, 3.4, and 4.8 M for the CsA-binding experiment 1, 2, 3 and 4 respectively.

The K_d values for CsA binding to CypA were calculated using the observed $\Delta C^{1/2}$ values (see Figure 17) and Equations 1 and 2 above. The calculated K_d values in the four different biological replicate analyses were in the range of 0.3 to 16 nM. The average value of 6 nM was within 5-fold of the low end of the range of K_d values previously reported for the CsA-CypA complex, which range from 30 to 200 nM (129-131). Our K_d value calculations employed an estimated m -value (see **Quantitation of Ligand Binding** section in **Experimental**). We note that just a 30% decrease in the m -value used in our calculations would increase the average K_d value for the CsA-CypA complex in this work to 130 nM, which is well within the previously reported range.

Table 10: CsA-binding proteins identified using SILAC-PP protocol.

Protein(s)	Protein MW	Fragment /Intact	Exp# (Gel Bands)
Peptidyl-prolyl cis-trans isomerase, Cyclophilin A	17504.5	Intact	1,2,3,4 (17kDa)
Threonine synthase	57587.3	Intact	2 (54kDa)
Hexokinase-2	54169.7	Intact	1 (54kDa)
Glucose-6-phosphate isomerase	61298.1	Intact/fragment?	2 (54kDa)
60S ribosomal protein L26-A	14233.4	Intact	2,4 (17kDa)
60S ribosomal protein L25	15757.3	Intact	3 (17kDa)
40S ribosomal protein S17-A	15845.1	Intact	3 (17kDa)
(DL)-glycerol-3-phosphatase 1	28117.5	Fragment	1 (17kDa)

Seven protein hits identified in the CsA-binding experiments performed here were not consistently identified as hits in all four CsA-binding experiments (Table10).

These seven protein hits included 6 proteins that were only identified as hits in one experiment and 1 protein that was identified as a hit in two experiments. Five of these 7 proteins actually yielded inconsistent data (i.e., displayed hit behavior in one experiment and not in others). The other two proteins, which only appeared as a hit in one experiment, were only effectively assayed in one experiment. The biological replicate analyses with SILAC-PP consistently identified the true positive, cyclophilin A, and also helped identify the five proteins with inconsistent data that are likely false positives.

4.4.4 SILAC-PP: A Tool for Protein-Target Validation

The ability of SILAC-PP to cross-validate results from SPROX experiments (and *vice versa*) was evaluated by comparing the SILAC-PP data generated in this work to SPROX data we have previously reported (65, 73). A comparison of SILAC-PP and SILAC-SPROX ATP-binding results successfully cross-validated 17 of the 33 protein targets identified here (65). Nine of these cross-validated ATP-binding hits have previously known ATP-binding properties based on GO-term analyses and the result of other experiments (see Table 9 above).

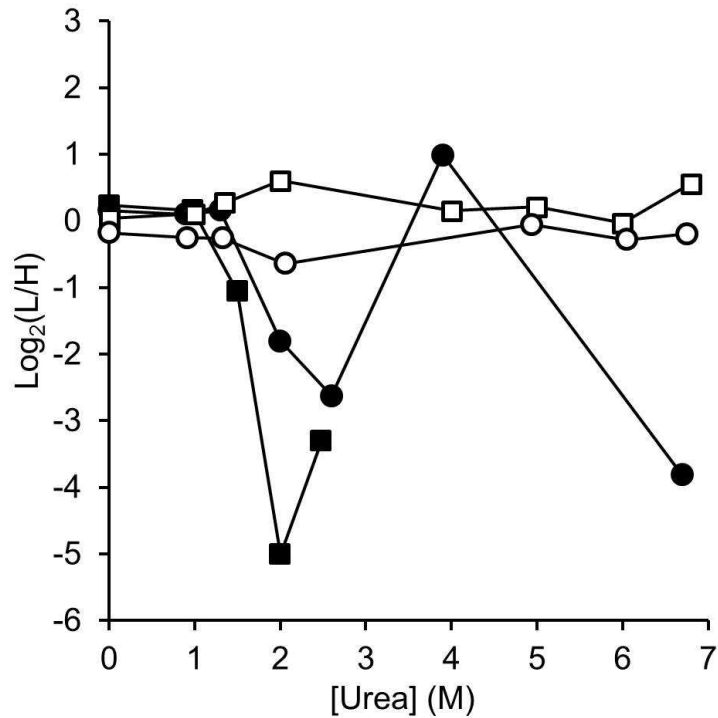


Figure 18: SILAC-PP data for the Phosphoglycerate Mutase, GPM1, obtained from intact protein band in ATP-binding experiment 1 (solid squares) and experiment 2 (solid circles) respectively. Also shown are the intact protein data for GPM1 from control experiment 1 (empty squares) and experiment 2 (empty circles). The $\Delta C^{1/2}$ values extracted was >1.5 M and 2.6 M for ATP-binding experiments 1 and 2 respectively.

For example, we corroborated the ATP binding properties of phosphoglycerate mutase (Figure 18), which was recently discovered to have ATP binding properties in a previous PP study in *E. coli* (94) and in a SPROX experiment involving proteins in yeast cell lysate (65). Eight of the 17 cross-validated proteins in this study were only previously identified as ATP binding proteins in an earlier SILAC-SPROX experiment.

These newly validated ATP-binding proteins are summarized in Table 11. The SILAC-PP and SPROX experiments utilize two different reaction probes (thermolysin versus oxidation) and mass spectrometry readouts (e.g., quantitation at the peptide versus protein level). Thus, it is unlikely that these 8 protein targets are false positives.

These eight newly validated ATP-binding proteins are reported to have NAD⁺, RNA, Mg, and other metal-binding properties based on a GO-term analysis (see Table 11). It is not surprising to find that some of the cross-validated protein hits in our ATP-binding study also bind NAD as these two ligands share a common adenine moiety. It is also interesting that several known Mg binding proteins were cross-validated as hits in our ATP-binding study, as Mg is a common cofactor involved in ATP-binding (132). Similarly, a number of metal binding proteins are known to have ATPase activity and thus bind ATP (133). There is also increasing evidence that many RNA binding proteins have roles in other cellular processes and require ATP for function (134, 135). Several of the novel ATP binding proteins cross-validated here have known RNA binding properties (see Table 11), and are likely involved in protein-protein interactions and/or cellular functions that require ATP.

Table 11: Protein targets with previously unknown ATP binding properties that were identified and cross-validated with SILAC-PP and SILAC-SPROX.

Protein	Known Ligand*
Pyruvate decarboxylase isozyme 1	Thiamine pyrophosphate, Mg, Metal-binding
Enolase 1	Mg, Metal-binding
Alcohol dehydrogenase 1	NAD ⁺ , zinc, Metal-binding
Adenosylhomocysteinase	NAD ⁺ binding
40S ribosomal protein S3	RNA-binding
40S ribosomal protein S2	RNA-binding
40S ribosomal protein S1-A	None
60S ribosomal protein L3	None

*Based on GO term analyses using the Universal Protein Resource (www.uniprot.org) (136).

A total of 16 of the 33 ATP binding proteins identified here by SILAC-PP were not identified as hits in an earlier SILAC-SPROX experiment [13]. Cross-examination of the data revealed that 4 of these 16 proteins were not assayed in the earlier SILAC-SPROX experiment. However, the remaining 12 protein targets were assayed but did not appear as hit in the SILAC-SPROX experiment. Interestingly, 4 of these 12 proteins have been previously reported to bind ATP, either through GO-term analyses or affinity capture mass spectrometry experiments (see Table 9).

One goal of the CsA-binding experiments using SILAC-PP was to cross-validate the CsA-binding behavior of pyruvate kinase, which was identified as a protein target of CsA in an earlier iTRAQ-SPROX experiment (73). Thermodynamic analysis of the CsA binding properties of a purified sample of pyruvate kinase did not indicate a direct

binding interaction between CsA and pyruvate kinase (137). This result suggested that pyruvate kinase was either a false positive in our earlier iTRAQ-SPROX study or that it was an indirect target of CsA (e.g., involved in protein-protein interactions induced by the presence of CsA). The SILAC-PP protocol developed in this work made it possible to perform a pulse proteolysis experiment with CsA and a yeast cell lysate. This enabled a direct comparison to the previous iTRAQ-SPROX experiment with CsA and a yeast cell lysate in which pyruvate kinase was identified as a target.

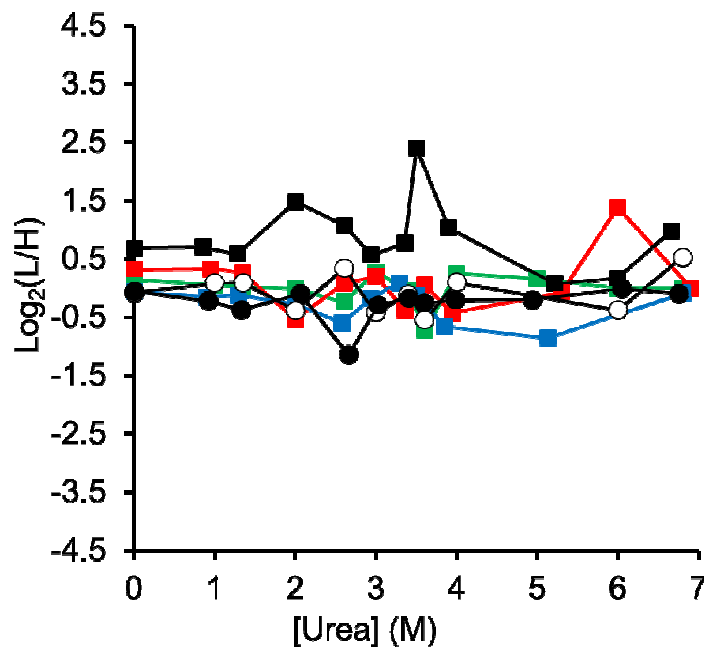


Figure 19: SILAC-PP data for the Pyruvate Kinase, obtained from intact protein band, solid green, red, black and blue squares in CsA-binding experiments 1, 2, 3 and 4 respectively and empty and solid circles represent data on the intact protein from control experiments 1 and 2 respectively.

The denaturant-dependence observed for the L/H ratios measured for pyruvate kinase (Figure 19) in the SILAC-PP experiment reported here were consistent with the behavior expected for a non-hit protein. These results suggest that pyruvate kinase was likely a false positive in the earlier iTRAQ-SPROX experiment (73). This appearance of pyruvate kinase as a false positive in the iTRAQ-SPROX experiment is likely due to the relative precision of the iTRAQ readout. Indeed, we have recently shown that the false positive rate of hit peptide discovery in iTRAQ-SPROX is on the order of 1-5%(64, 90).

4.5 Conclusions

The SILAC-PP protocol described here enables the PP technique to be interfaced with bottom-up, shotgun proteomics strategies. The modified PP protocol reported here is particularly advantageous for protein-ligand discovery because it allows the proteins in the whole cell lysates to be efficiently assayed for binding to target ligands without the need for pre-fractionation. Compared to existing PP protocols, the SILAC-PP protocol is more directly comparable to other experimental approaches for proteome-wide analyses of protein-ligand binding interactions (e.g., DARTS, affinity capture mass spectrometry methods, the cellular thermal shift assay, and SPROX)(73, 85, 121, 122). This makes the SILAC-PP technique a useful tool for cross-validating the findings from large-scale protein-ligand binding experiments using these other techniques. The

SILAC-PP results presented here for two ligands in multiple biological replicates help establish both the robustness of the method's discovery capability and the ability to cross-validate the target proteins from other similar approaches. The results of this study demonstrate that the application of SILAC-PP in tandem with other energetics-based approaches can provide a means by which to both corroborate protein hits and help eliminate false positives in a proteome-wide ligand binding experiments.

After the successful development and application of the SILAC-based strategies described in the chapters 3 and 4 with yeast lysate and selected ligands, these approaches were utilized to study the thermodynamic changes in disease states modeled by breast cancer cell lines. The chapter that follows in this dissertation discusses the findings of these strategies in identifying proteins with altered thermodynamic stability in two comparative study of breast cancer using cell lines.

5. Global Thermodynamic Profiling of Proteins in Breast Cancer Cell lines

5.1 Introduction

Gene and protein expression level analyses using DNA microarrays and protein mass spectrometry methods have been widely utilized over the past two decades for the characterization of cancer and other disease-related states(26, 138-151). While differential expression profiling studies can provide insights into the cellular pathways and proteins associated with a disease state, the biological significance of proteins with altered expression levels is sometimes dubious because of the indirect link between a gene or protein's expression level and its function. Functionally relevant proteins with the same expression levels in different biological states also go undetected using the current paradigm of expression level profiling to characterize such states. Recently, activity based protein profiling studies have been used to characterize disease states(152). Such studies are attractive because they can probe the functional significance of proteins with a range of expression levels. However, activity-based profiling studies require the development of specialized probes that target specific enzyme activities.

Protein folding and stability measurements can report on a number of different biologically significant phenomena associated with proteins in different disease states. Point mutations, post-translational modifications, and new and/or altered binding

interactions with cellular ligands (e.g., other proteins) can all produce thermodynamic stability changes. Indeed, such thermodynamic stability changes have been measured for specific disease related proteins(62, 153-158), including several proteins linked to cancer(59-61). The thermodynamic analyses of disease-related proteins performed, to date, have largely involved purified protein systems that were already found to be associated with disease states. However, such thermodynamic analyses have not been used for the global analysis of disease states. This is in large part because methods for making protein folding and thermodynamic stability measurements on the proteomic scale have only recently been developed(64-72, 159, 160). Because protein stability changes can arise for such a wide range of functionally relevant reasons, they have the potential to be a general probe of protein function and to be sensitive to a number of different types of functionally relevant and disease-related changes. The characterization of such functionally relevant changes in different disease states has the potential to create new diagnostic signatures of disease states and to produce a better molecular level understanding of the disease, ultimately facilitating the discovery of novel drug therapies.

Described here are the results of a proof-of-principle study involving global measurements of protein folding and stability to differentiate three well-established cell

culture models of breast cancer including the non-tumorigenic MCF-10A breast cell line and the MCF-7 and MDA-MB-231 breast cancer cell lines that have different molecular features and display different degrees of invasiveness. These cell culture models of breast cancer have been well studied using gene and protein expression level studies, and a number of differentially expressed proteins have been identified in these cell lines(12, 97-102). The Stability of Proteins from Rates of Oxidation (SPROX) coupled with Stable Isotope Labeling with Amino acid in Cell Culture (SILAC) technique, termed SILAC-SPROX(65), was used here to assay ~800 proteins in the MCF-10A, MCF-7, and MDA-MB-231 cell lines for protein stability changes. In order to differentiate the true positives from false positives, biological replicates measurements were performed on the same cell lines. Furthermore, the utility of another energetics based approach SILAC-PP was demonstrated to confirm proteins identified with SILAC-SPROX in MCF-10A versus MCF-7 comparative study. Our results demonstrate that protein folding and stability measurements can be used to differentiate the three cell culture models of breast cancer in this study. The differentially stabilized proteins included those with and without significantly altered expression levels. The thermodynamic stability profiles generated here also revealed novel molecular signatures of the breast cancer cell lines in this study. The differential stability of one protein hit (calpain small subunit 1, CAPNS1)

in the MCF-10 and MCF-7 cell lines was correlated with a change in calpain's proteolytic activity. Part of the work described in this chapter has been published in the Journal of Proteome Research³.

5.2 Experimental

5.2.1 Cell Culture and Cell Lysate Preparations

The MCF-10A, MCF-7, and MDA-MB-231 cells were cultured in a humidified 37°C incubator with 5% CO₂. All cell lines were cultured following the ATCC guidelines except for MCF-7 heavy SILAC- labeled cells which were cultured using heavy-labeled lysine and arginine according to established protocols.^(12, 161, 162) Briefly, starting cells for each of the cell lines were acquired from ATCC (provided by the Duke Cell culture facility). MCF10A cell lines were maintained in MEGM media containing MEBM (CC-3151) and all the components of the kit except gentamycin (CC-4136) from Lonza/Clonetics, along with the 100ng/ml of cholera toxin (Sigma). Similarly, MDAMB231 cell lines were grown in MEME (Sigma) media supplemented with 1mM Sodium pyruvate (Gibco), 0.1mM non-essential amino acids, NEAA (Gibco) and 10% fetal bovine serum (FBS) (Hyclone). All the media also contained 1% fungizone (Gibco,

³ **Adhikari, J.**, West, G.M., and Fitzgerald, M.C. Global Analysis of protein folding thermodynamics for Disease State Characterization, Journal of Proteome Research, 2015.

250 µg/ml). For the SILAC heavy labeled cells, initially MCF-7 cell lines were propagated in DMEM (Gibco) media containing 1mM Sodium pyruvate (Gibco), 0.1mM non-essential amino acids, NEAA (Gibco) and 10% fetal bovine serum (Hyclone) supplemented with 10µg/ml of insulin. The MCF-7 cells were passaged twice in this media and then transferred to DMEM media containing all of the above except 10% dialyzed FBS was used instead of the regular serum. The cells were adapted well in this media containing dialyzed FBS before passaging on the heavy SILAC media. The heavy SILAC media comprised of the SILAC DMEM (Thermo Scientific), 10% dialyzed FBS (Sigma), 10µg/ml of insulin with 500µl of ¹³C₆ Arginine(86.2 mg/L) and 500µl of ¹³C₆ ¹⁵N₂ Lysine (181.2 mg/L) (Cambridge Isotope Laboratories) and 500 µl of Proline (200mg/L) (sigma). The MCF-7 cells were passaged at least four times in heavy media before they were harvested. The cells were washed twice with PBS before harvesting with HyQase solution (Hyclone) in a 15 ml tube. These harvested cells were immediately spun down at 1000rpm for 5 min. Washed the pelleted cells with PBS one more time before centrifuging and stored at -80°C until lysis.

Cells were lysed by mechanical disruption using a disruptor genie (Scientific Industries) and 1 mM diameter zirconia/silica beads (Biospec) in 20 mM phosphate buffer (pH 7.4) containing a cocktail of protease inhibitors that included: 1 mM AEBSF,

500 μ M Bestatin, 15 μ M E-64, 20 μ M Leupeptin, and 10 μ M Pepstatin A (Thermo Pierce). The cell lysates were centrifuged at 15000 X g for 15 min at 4°C, and the supernatants were used in the SILAC-SPROX analyses. The total protein concentration of the cell lysates used in each cell line comparison was normalized to the same total protein concentration, which was 3-10 mg/ml depending on the experiment.

The MCF-7 (heavy) and MCF-10A(light) cell lysates used in the SILAC-PP experiment was prepared in similar manner to the SILAC-SPROX experiments except the lysis buffer contained 350-500 μ l of 100mM Tris-HCl buffer (pH 8.0) containing 50mM NaCl, 10 mM CaCl₂, and a cocktail of protease inhibitors. The concentrations of total protein in the supernatants were determined using the Bradford assay. The total protein concentration of the lysate was 6.9 and 13.4 mg/ml for MCF-7 and MCF-10A cell lysates respectively. The protein concentrations in the two cell line lysates were both normalized to 6 mg/ml by diluting with an appropriate amount of lysis buffer before proceeding to the SILAC-PP analysis.

5.2.2 SILAC-SPROX Analyses

Two different SILAC-SPROX comparisons were performed here including one that involved the MCF-7 and MCF-10A cell lines and one that involved the MCF-7 and MDA-MB-231 cell lines. Three biological replicates of each SILAC-SPROX comparisons

were performed. In each biological replicate the test cell lysates were diluted into a series of twelve denaturant-containing buffer stock solutions. The denaturant-containing buffer stock solutions contained 20 mM phosphate buffer (pH 7.4) with urea concentration ranging from 0 to 9 M, and 75 μ L volumes were combined with the 20 μ L sample aliquots of the lysates from respective cell lines. The final denaturant concentrations in each set of denaturant-containing buffers ranged from 0 to 7 M. The samples were equilibrated 16-18 hours in the denaturant-containing buffers. The methionine oxidation reactions were initiated by addition of hydrogen peroxide (0.54 M (Sigma)) to the samples in each set of denaturant-containing buffers. The oxidation reactions were allowed to proceed for 6 minutes before quenching each reaction with the addition of 760 μ l of a 375 mM solution of L-methionine (Sigma). After quenching each reaction, the samples from the two different cell lines in buffers containing the same denaturant concentration in each analysis were combined. The combined samples were precipitated with TCA, and the resulting protein pellets in each of the combined samples were subjected to a quantitative, bottom-up proteomics analysis using SILAC quantitation as described below.

5.2.3 SILAC-Pulse Proteolysis Analysis

The PP reaction conditions were similar to those previously described (91, 94). Briefly, 25 μ l aliquots of the MCF-7 and MCF-10A cell lysates were each combined with 75 μ l of denaturant-containing buffer stock solutions, which were prepared in 100 mM Tris-HCl buffer (pH 8.0) with urea concentrations ranging from 0 to 9 M. The exact concentration of urea in each denaturant-containing buffer stock solution was determined from a refractive index measurement of the buffer as described previously (123). The final urea concentrations in the denaturant-containing buffers were 0, 1.3, 2, 2.6, 3, 3.6, 4, 4.5, 4.9, 5.3, 6, 6.8 M. The samples in the denaturant-containing buffers were equilibrated for ~18 hours at room temperature. The proteolysis reaction was initiated by adding thermolysin protease from *Bacillus thermoproteolyticus rokko* (Sigma) to the protein samples in each denaturant-containing buffer. The protease to protein ratio was approximately 1:10 (w/w). The proteolysis reactions were allowed to proceed for 1 min before each reaction was quenched with the addition of 15 μ l of 500mM EDTA solution (pH 7.9-8.0) (Thermo scientific and Boston Bioproducts, Boston, MA). An aliquot of 100 μ l of the MCF-7 and MCF-10A samples in buffers containing the same denaturant concentration were combined. 200 μ L of the combined sample pairs were mixed with 40 μ L of a 6X stock solution of Laemmli sample buffer, which was comprised of 375 mM

Tris•HCl buffer (pH 6.8) containing 6% SDS, 50% Glycerol and 0.045% Bromophenol (Boston Bioproducts, Boston, MA) and 6-7 % of the 14.3 M β -mercaptoethanol (BME) solution (Aldrich).

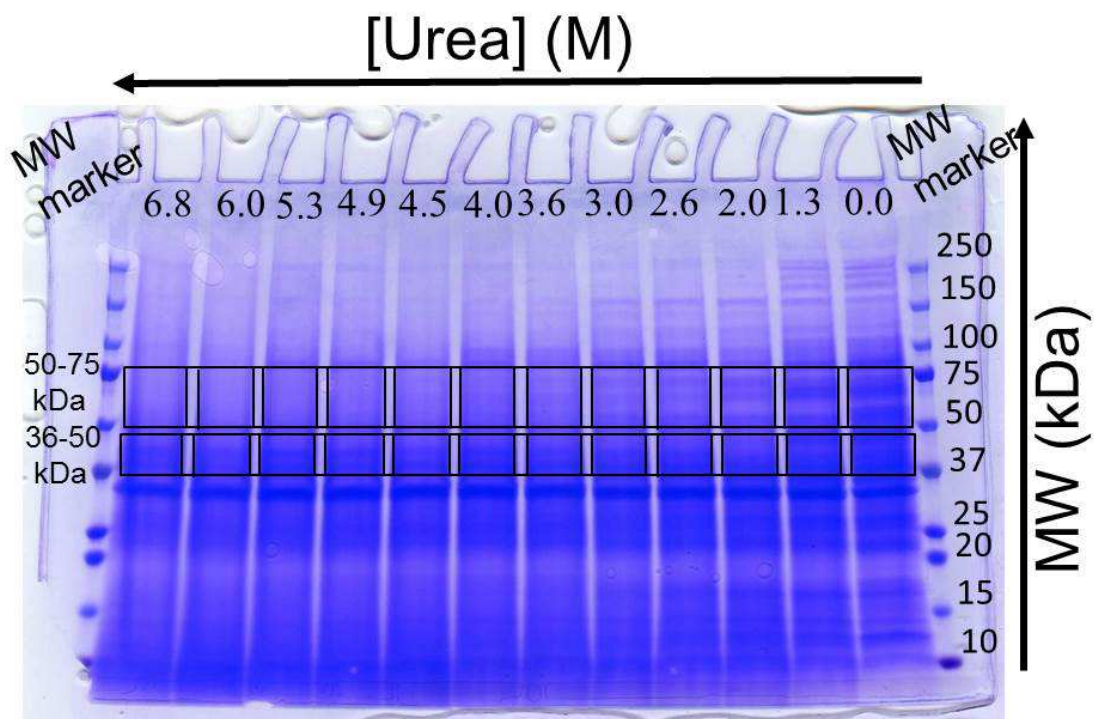


Figure 20: SDS-PAGE gel image for the MCF-7 versus MCF-10A comparative study using SILAC-PP strategy. The rectangular boxes in the gel mark the gel bands excised for the in-gel digestion.

The samples were mixed, centrifuged at $2000 \times g$ for 20 seconds, and heated at 95°C for ~10 mins. Approximately, $45 \mu\text{l}$ of the protein samples in the denaturant-containing buffers, were loaded on to a 4-20% midsize polyacrylamide gel (BioRad Criterion,

Hercules, CA). An aliquot of the Precision Plus standard marker (BioRad, Hercules, CA) was also loaded in the wells at the two ends of the gel.

The SDS-PAGE gels were fixed upon treatment with a solution of 25% Isopropanol, 10% acetic acid and 65% water for approximately 25 minutes. The gels were stained overnight with a staining solution that contained 0.01% R-250 (Bio-Rad) in 10% acetic acid (EMD Chemicals Inc.). Excess dye was removed by treatment of the gels with a 10% acetic acid solution. The protein bands across the twelve lanes in each of the gels were excised in specific molecular weight regions and processed using a standard in-gel digestion protocol as previously described (124). Trypsin was used as the protease. The molecular weight regions of the gel bands excised in this experiment were 36-50 kDa and 50-75 kDa (Figure 20).

5.2.4 Proteomic Sample Preparation

The dried protein pellets from the SILAC-SPROX analyses were dissolved in 60 μ l of 0.5 M triethyl ammonium bicarbonate (TEAB) containing 3 μ l of a 2% stock solution of SDS. The disulfide bonds in each protein sample were reduced upon addition of 5 μ l of 50 mM TCEP and treatment for 1 hour at 60°C. The protein samples were each reacted with 2.5 μ l of 200 mM methyl methane thiosulfonate (MMTS) for 10 minutes at room temperature to block cysteine side chains. Ultimately, 3 μ l of a 1 μ g/ μ l

trypsin solution was added to the protein sample in each tube and the samples were incubated overnight (~15 hours) at 37°C before 5 µl of 10% trifluoroacetic acid (TFA) was added to quench the trypsin digestion. The samples were desalted using C18 resin (The Nest Group, Southborough, MA) according to the manufacturer's protocol.

5.2.5 LC-MS/MS Analyses of Proteomic Samples

In the SILAC-SPROX experiments, the tryptic peptides were desalted using C18 resin (The Nest Group, Southborough, MA) according to the manufacturer's protocol and analyzed by liquid-chromatography-tandem MS (LC-MS/MS) using an EASY-nLC 1000 system coupled to a Q-Exactive linear ion trap mass spectrometer (Thermo Fisher Scientific). Columns were packed in-house with Jupiter 4µ Proteo 90Å reversed phase resin (Phenomenex). Peptides were concentrated and desalted on a trapping column (100µm ID x 20 mm) and eluted on an analytical column (75 µm ID x 150 mm), operating at 300 nl/min and using the following gradient: 5% B for 3 min, 5–35% B in 120 min, 35–80% B in 2 min, and 80% B for 9 min [solvent A: 0.1% formic acid (v/v); solvent B: 0.1% formic acid (v/v), 80% CH₃CN (v/v) (Fisher Scientific)]. The Q-Exactive was operated in a data-dependent MS/MS mode using the top 10 most intense precursors detected in a survey scan from 300 to 1,800 m/z performed at 70K resolution. Tandem

MS was performed by HCD fragmentation with stepped normalized collision energy (NCE) of 20%.

In the SILAC-PP experiment, the peptide samples extracted from the gel-bands generated in the MCF-7 versus MCF-10A comparison experiment were each subjected to quantitative LC/MS/MS using a nanoAcquity UPLC system (Waters Corp) coupled to a Thermo QExactive Plus high resolution accurate mass tandem mass spectrometer (Thermo) via a nanoelectrospray ionization source. First, the sample was trapped on a Symmetry C18 180 mm × 20 mm trapping column for 3 min at 5 uL/min (99.9/0.1 v/v water/acetonitrile 0.1% formic acid), followed by the analytical separation on a 1.7 um Acquity BEH130 C18 75 mm × 250 mm column (Waters Corp). Peptides were separated using a gradient of 5 to 40% acetonitrile with 0.1% formic acid over 60 min at a flow rate of 400 nanoliters/minute (nL/min) with a column temperature of 55C. Data collection on the QExactive Plus mass spectrometer was performed in a data-dependent acquisition (DDA) mode of acquisition with a resolution (r) of 70,000 (for m/z 200) for full MS scan from m/z 375 – 1600 with a target AGC value of $1e6$ ions, followed by 10 MS/MS scans at $r=17,500$ (for m/z 200) at a target AGC value of $5e4$ ions. A 20s dynamic exclusion was employed to increase depth of coverage. The total analysis cycle time for each sample injection was approximately 95 minutes. The LC-MS/MS data generated in the

experiment described here were searched using MaxQuant as in SILAC-SPROX experiments.

5.2.6 Proteomic Data Analysis

In both SILAC-SPROX and SILAC-PP experiments, peak lists were extracted from the raw LC-MS/MS data files and the data were searched against the 20265 human proteins in the 2014-04 release of the UniProt Knowledgebase (downloaded at ftp://ftp.uniprot.org/pub/databases/uniprot/current_releases/release-2014_04/knowledgebase/) using Maxquant 1.3.0.5(163). The following modifications were used: methyl methanethiosulfonate at cysteine as a fixed modification, SILAC labeling of lysine ($^{13}\text{C}_6^{14}\text{N}_2$) and arginine ($^{13}\text{C}_6$), and variable (0-1) oxidation of methionine and deamidation of Asparagine and Glutamine (N and Q), and acetylation of the protein N-terminus. The enzyme was set as Trypsin, and up to 2 missed cleavages were permitted. The false discovery rate for peptide and protein identifications was set to 1%, and all other settings were set at the default parameters. As part of the default settings the mass tolerance for precursor ions was set to 20 ppm for the first search where initial mass recalibration was completed and a 6 ppm precursor mass tolerance was used for the main search. The mass tolerance for fragment ions was 0.5 Da (SILAC-SPROX) and 0.02 Da (SILAC-PP). We also included matches between runs and re-

quantification of the searched peptides. The search results were exported to Excel for further data analysis as described below. In cases where identified peptides could be matched to multiple protein isoforms or multiple members of a protein family, the peptide was assigned to the leading razor protein listed by MaxQuant algorithm.

In the SILAC-SPROX experiments, only the protein and peptide identifications with non-zero positive ratios ($H/L > 0$) were used in subsequent data analysis steps. For the methionine-containing peptides, a single averaged H/L ratio was calculated for each peptide sequence and each charge state at each denaturant concentration. Similarly, a median H/L ratio was determined for each protein in each analysis using the H/L ratios measured for all the non-methionine-containing peptides identified in all the denaturant concentrations for a given protein. These median H/L ratios were used to select hits with $H/L > 2$ in the protein expression level analyses. For hit peptide and protein selection in the thermodynamic analyses, all the H/L ratios for peptides from a given protein were divided by the median H/L ratio calculated for that protein to generate normalized H/L ratios, which were ultimately \log_2 transformed. The normalized and \log_2 transformed H/L ratios generated for the non-methionine-containing peptides in a given analysis (Figure 21) were used to determine the 5th and 95th percentiles. The methionine-containing peptides with \log_2 transformed H/L ratios less than the 5th percentile or

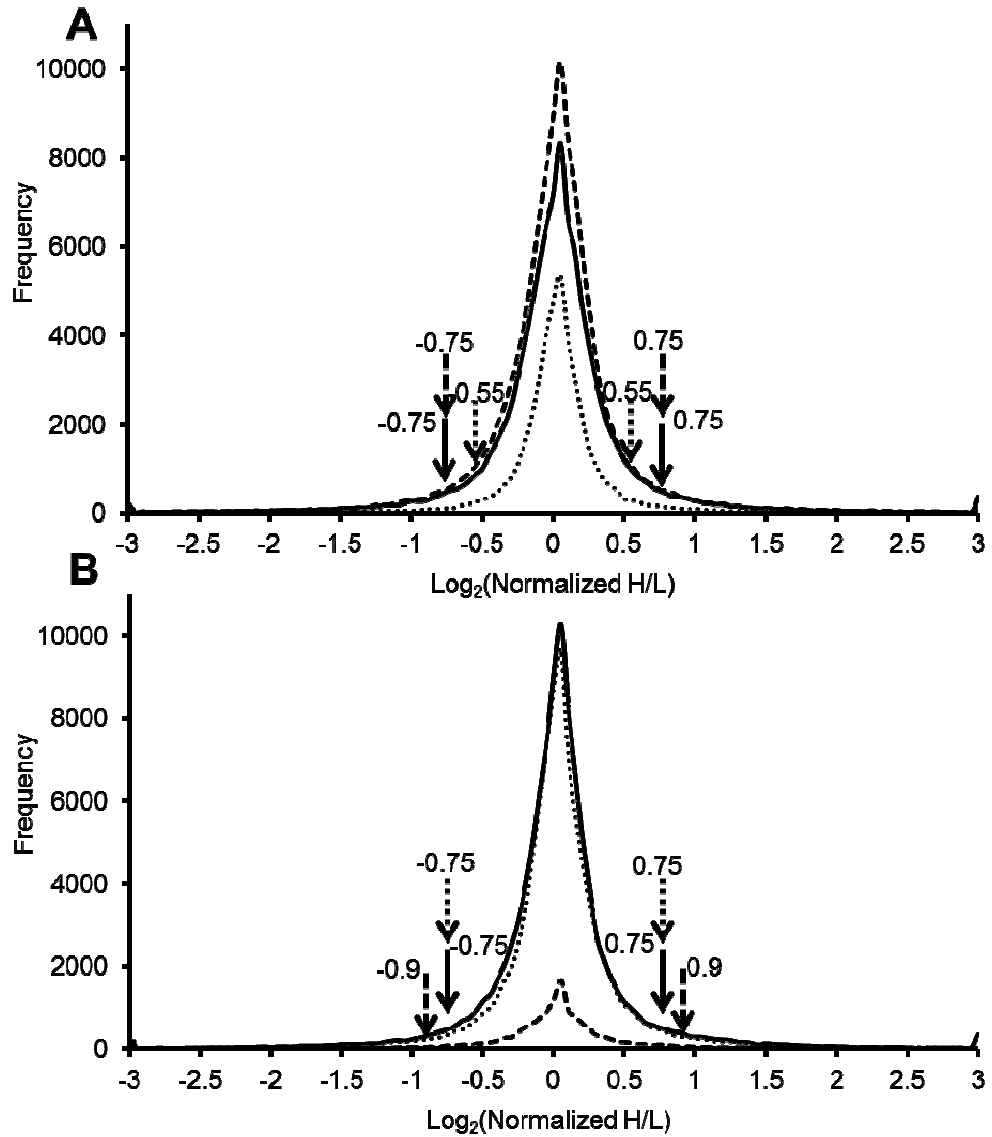


Figure 21: Global distribution of H/L ratios for all the non-methionine containing peptides identified in each experiments using SILAC-SPROX in (A) MCF-7 versus MCF-10A and (B) MCF-7 versus MDA-MB-231. Shown are the data from biological replicates 1 (dotted line), 2 (dashed line) and 3 (solid line) in each comparative studies. The arrows mark 5th and 95th percentile values in each experiment.

greater than the 95th percentile were selected and then visually inspected to determine which peptides had altered H/L ratios at 2 or more consecutive denaturant concentrations to generate an initial list of protein hits.

In order to differentiate the true-positives (i.e., proteins with altered stabilities in the cell line comparisons) from the false-positives, the hit peptides identified in the differential analyses were analyzed for consistency across the three biological replicates. Peptide hits with inconsistencies in the biological replicates (e.g., a peptide was identified as hit in one biological replicate and not in another, different charge states showed different SILAC-SPROX behavior, or the peptide retention times were not consistent in the LC-MS/MS analyses) were deemed false positives and eliminated from the final protein hit lists.

In the SILAC-PP experiment, protein quantitation output file from MaxQuant was exported to excel and used in the subsequent analysis. The median H/L ratios determined for the specific proteins from the non-methionine containing peptides in the SILAC-SPROX experiment were used to determine the baseline for SILAC-PP experiment. The H/L ratios generated in the SILAC-PP experiment for specific proteins in a region of the gel across the twelve different denaturants were inspected for differential thermodynamic stability between the two cell lines, MCF-7 and MCF-10A.

Only those proteins that were identified in at least 6 or more denaturant concentrations were assayed for potential changes in thermodynamic stability. The hit proteins in each experiment were identified as those proteins with >1.7 fold deviations from the median H/L ratio as determined for the same cell line comparison using SILAC-SPROX, at two or more consecutive denaturant concentrations.

5.2.7 Protease Activity Assays

The calpain and cathepsin D activity assay kits (ab65308 and ab65302, respectively) were both purchased from Abcam®. The activity assays were performed according to the manufacturer's protocol using either an Envision® or Victor3V Multilabel Plate Reader (both from Perkin Elmer). The 405 nm excitation and 515 nm emission filters were used in the calpain activity assay, and the 340 nm excitation and 460 nm emission filters were used in the cathepsin D activity assay. In each case the activity assays were performed in duplicate using cell lysates identical to those used in SILAC-SPROX analyses. In the calpain activity assay, fluorescence measurements were recorded on ~30 µg of the total lysate in the absence and presence of 0.5 µl of the calpain inhibitor provided with the kit. In the Cathepsin D activity assay, fluorescence measurements were recorded on ~20 µg of the total lysate in the absence and presence of 10 µl of a 1mM stock solution of pepstastin A. In each case the specific activity of the

enzyme was measured by subtracting the fluorescence intensity of the sample with the inhibitor from the one without the inhibitor and normalizing to the total protein concentration used in each reaction.

5.3 Results and Discussion

5.3.1 Global Analysis of Protein Folding Thermodynamics using SILAC-SPROX Strategy

The experimental workflow that utilized SILAC-SPROX strategy for disease state characterization is outlined in Figure 22. The estrogen positive (ER+) and poorly invasive MCF7 breast cancer cell line was heavy-labeled and used in two different comparative studies. In one study protein stabilities in the heavy-labeled MCF-7 breast cancer cell line were compared to those in a non-tumorigenic MCF-10A cell line. In the second study protein stabilities in the heavy-labeled MCF-7 breast cancer cell line were compared to those in a highly invasive, triple negative (i.e., estrogen, progesterone, and human epidermal growth factor 2 receptor negative) breast cancer cell line, MDA-MB-231.

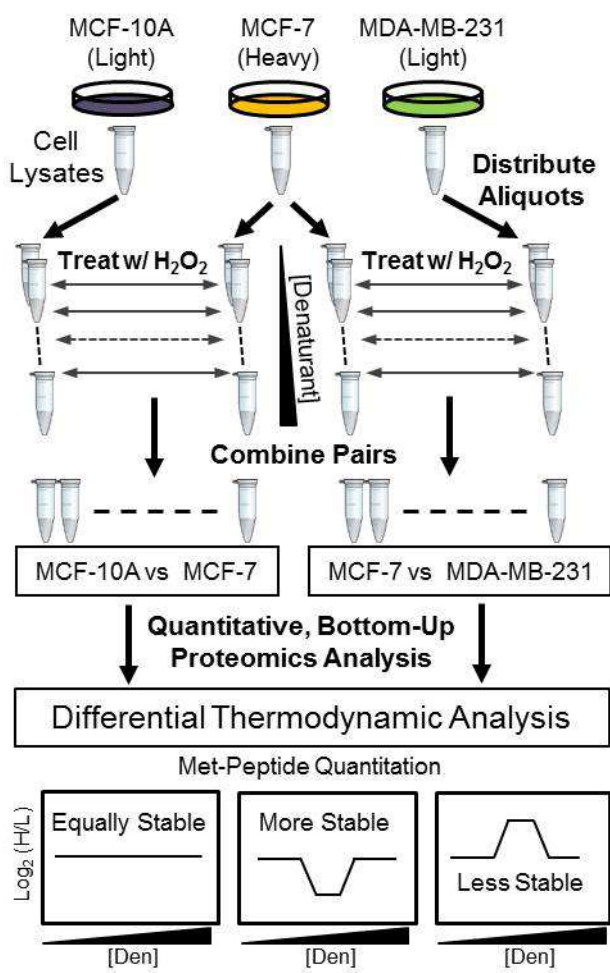


Figure 22: Schematic representation of the SILAC-SPROX experimental workflow used in this study.

Each comparative study utilized the SILAC-SPROX method(65) to compare the thermodynamic stabilities of the proteins in each cell line. Proteins with different chemical denaturation behavior in the two cell lysate samples were identified based on the H/L ratios determined for the methionine-containing peptides identified in a

quantitative bottom-up proteomics analysis of the combined protein samples generated in each comparative analysis (see Figure 22).

Table 12: Proteomic Coverage of peptides and proteins assayed for thermodynamic stability changes in the MCF-7 Vs MCF-10A and MCF-7 vs MDA-MB-231 cell line comparisons.

	<i>MCF-7 Vs MCF-10A</i>		
Biological Replicate	Total Peptides (Proteins) Identified	Total Peptides (Proteins) Assayed	Total Peptides (Proteins) Hits
1	5582 (1394)	987 (432)	120 (80)
2	12578 (1788)	1751 (576)	155 (92)
3	11039 (1863)	1570 (538)	128 (70)
Total	16582 (2431)	2472 (772)	336 (175)
	<i>MCF-7 Vs MDA-MB-231</i>		
Biological Replicate	Total Peptides (Proteins) Identified	Total Peptides (Proteins) Assayed	Total Peptides (Proteins) Hits
1	12504 (2249)	1457 (609)	118 (82)
2	3051 (1210)	548 (359)	35 (34)
3	11299 (1852)	1595 (531)	132 (98)
Total	18147 (2872)	2602 (881)	274 (169)

Shown in Table 12 is a summary of the proteomic data obtained in the MCF-7 versus MCF-10A and MCF-7 versus MDA-MB-231 cell line comparisons. In each of the comparative analyses approximately 800 protein stabilities were surveyed using ~2500 methionine-containing peptide probes (see Table 12). Approximately 10-12% of the methionine-containing peptides assayed in each of the comparative analyses were identified as hits (i.e., yielded different chemical denaturation curves in the comparative analyses). This hit rate is significantly above the false positive rate of 3.5% that we have

previously reported for SILAC-SPROX(65), indicating that a number of the detected hits are true positives.

In order to differentiate the true-positives from the false-positives, the hit peptides were analyzed for consistency across the three biological replicates. Approximately 50% of the peptide hits in Table 1 had inconsistent data in the biological replicates. The final hit list contained 138 peptide hits from 84 unique proteins and 128 peptide hits from 102 unique proteins in the MCF-7 versus MCF-10A and MCF-7 versus MDA-MB-231 cell line comparisons, respectively (see Appendix B.1 and B.2). In these final hit lists 38 and 23% of the protein hits were identified as hits in two or more replicates and/or with multiple peptide probes in the MCF-7 versus MCF-10A and MCF-7 versus MDA-MB-231 comparisons (respectively). The remaining protein hits included those that were only identified as hits in only one experimental replicate with a single peptide probe.

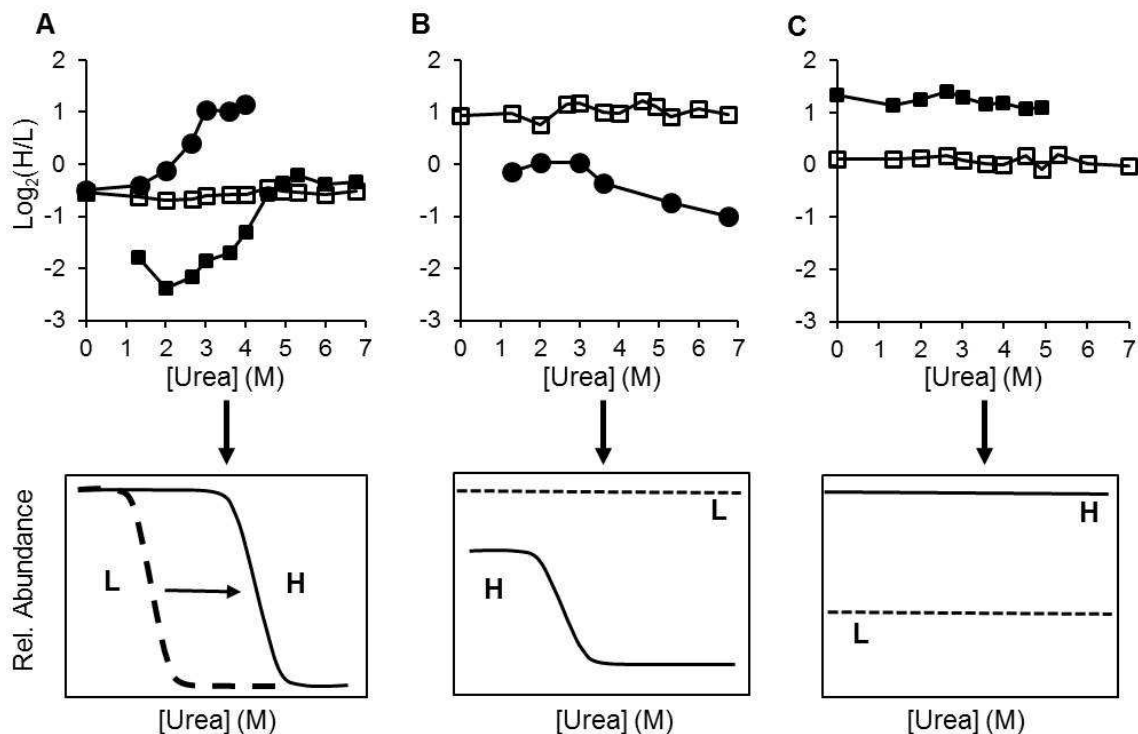


Figure 23: Representative SILAC-SPROX data obtained in the two cell line comparisons in this work. (A) Data obtained on peptides from calpain small subunit 1 in the MCF-7 versus MCF-10A comparison including the wild-type (filled circles) and doubly oxidized (filled squares) forms of the methionine-containing peptide SMVAVMDSDTTGK (detected in biological replicates 3 and 2, respectively), (B) Data obtained on peptides from splicing factor 3B subunit 3 in biological replicate 2 of the MCF-7 versus MCF-10A comparison including the wild-type form of the methionine-containing peptide AVMISAIEK (filled circles), (C) Data obtained on peptides from peptidyl-prolyl cis-trans isomerase protein in biological replicate 3 of the MCF7 versus MDA-MB-231 comparison including the oxidized form of the methionine-containing peptide (ac)MVNPTVFFDIAVDGEPLGR (closed squares). In each case the median data for all the non-methionine-containing peptides from respective proteins are also shown (open squares). Shown on the bottom panels are schematic representations of the expected unfolding curves that produced the observed SILAC-SPROX behavior in each example.

The peptide hits displayed one of three different SILAC-SPROX behaviors (see Figure 23). About half of the final hits detected in each of the two comparisons showed the behavior expected for methionine-containing peptide probes from globally protected regions in proteins (or protein folding domains) that were differentially stabilized in the disease states (Figure 23A). The other hits in each comparison had methionine-containing peptide probes with H/L ratios that were significantly altered across all the denaturant concentrations used in each experiment (Figure 23B and C). Some of these also showed a denaturant dependence to their H/L ratios at least in one cell line (Figure 23B), while others did not (Figure 23C). The SILAC-SPROX behaviors observed in Figures 22B and C would be expected of methionine-containing peptide probes derived from regions of protein structure that are either solvent exposed or only locally protected in one cell line and globally or sub-globally protected in the other cell line (Figure 22B) or one exposed and the other locally protected (Figure 23C) in the disease states.

5.3.2 Hit Proteins Map to a Wide Range of Protein Expression Levels

The SILAC-SPROX approach is especially attractive for disease-state analyses because the quantitative proteomic data collected on non-methionine containing

peptides can be used to evaluate protein expression level changes. This allowed for a concurrent analysis of the expression changes associated with the differentially stabilized protein hits identified in this work. The protein hits with significantly altered thermodynamic stabilities showed a wide range of expression level changes (Figure 24 and Appendix B.1-B.2). The protein expression levels measured in the MCF-7 and MCF-10A comparison in this work were generally in good agreement with those reported in a previous protein expression level study on the same cell lines (Figure 24 and Appendix B.1)(12). Approximately, 45% of the protein hits identified with significantly altered thermodynamic stabilities in the two comparisons had expression levels that changed less than two-fold in the comparative analyses (see Appendix B.1-B.2). These proteins with similar expression levels and different thermodynamic stabilities are likely to have altered functions in the cancer cell lines studied here. These functional differences would go undetected in a typical expression level-based proteomic study.

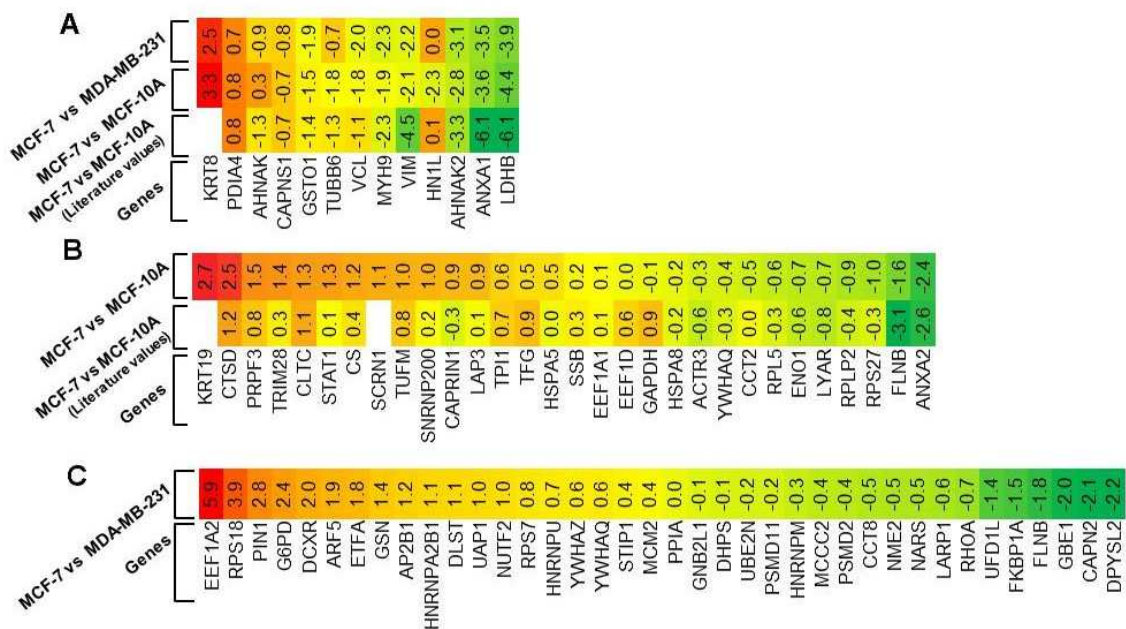


Figure 24: Subset of protein hits detected in the comparative analyses performed here. Shown in (A), (B), and (C) are the gene names of proteins that were detected as hits in both cell line comparisons, only the MCF-7 versus MCF-10A comparison, and only the MCF-7 versus MDA-MB-231 comparison, respectively. The numbers in the boxes represent the protein expression level data (i.e., Log₂(H/L) values) either measured in this work or previously reported in the literature for each protein hits. The literature values are from reference (12). The protein hits highlighted here are those that were assayed with the same peptide probe in both cell line comparisons.

5.3.3 Protein Signatures for Disease State Differentiation

The MCF-10A, MCF-7, and MDA-MB-231 cell lines used here are well-studied cell line models of breast cancer, and a number of the protein biomarkers identified in

this work have been previously associated with cancer. All the protein hits in our comparative studies that were identified in multiple biological replicates have been previously associated with tumorigenesis and/or the progression of cancer in expression level and/or other studies (Appendix B.1 and B.2 and references therein). All but two (protein disulfide-isomerase A4 and stress-induced-phosphoprotein 1) have been previously associated with breast cancer in expression level and/or other studies (see Appendix B.1 and B.2 and references therein).

Several proteins involved in cell building and metabolic processes also displayed differential thermodynamic properties in our study. These included cytoskeletal proteins such as vimentin, keratin, myosin, and filamin that are known to have differentially modulated expression level in cancer(164-167). A significant number of glycolytic enzymes also showed differential thermodynamic stabilities in the normal to cancer transition. These included proteins such as GAPDH, LADHB, TPI1 and ENO1 that are reported to be dysregulated in many cancers(101, 168-171). These glycolytic enzymes have been implicated in tumorigenesis and proliferation of cancer cells including in breast cancer and have been widely studied for diagnostic and therapeutic purposes(168, 169, 171). This helps validate the biological significance of the protein targets discovered here using the described approach.

Table 13: Protein hits detected in both of the cell line comparisons, the MCF-7 versus MCF-10A, and the MCF-7 versus MDA-MB-231.

<i>Peptides</i>	<i>Gene Names</i>	<i>Protein Names</i>
HEAM(ox)ITDLEER***	MYH9	Myosin-9
IMGIPEEQMGLLR	MYH9	Myosin-9
SMVAVMDSDTTGK	CAPNS1	Calpain small subunit 1
FAMEPEEFSDTLR	PDIA4	Protein disulfide-isomerase A4
M(ox)LGQM(ox)TDQVADLR	VCL	Vinculin
M(ox)ILELFSK	GSTO1	Glutathione S-transferase omega-1
EM(ox)GVDVHFPK	AHNAK	Neuroblast differentiation-associated protein AHNAK
LESGMQNM(ox)SIHTK	KRT8	Keratin, type II cytoskeletal 8
SEIDMNDIK	ANXA1	Annexin A1
IHPVSTM(ox)VK	LDHB	L-lactate dehydrogenase B chain
MVVESAYEVIK	LDHB	L-lactate dehydrogenase B chain
MVVESAYEVIK	LDHB	L-lactate dehydrogenase B chain
VEADVSLPSM(ox)QGDLK	AHNAK2	Protein AHNAK2
M(ox)ASTFIGN(de)STAIQELFK	TUBB6	Tubulin beta-6 chain
M(ox)ASTFIGNSTAIQELFK	TUBB6	Tubulin beta-6 chain
AM(ox)KPPGGESSNLFG SPEEATPSSRPNR	HN1L	Hematological and neurological expressed 1-like protein
LGDLYEEEMR	VIM	Vimentin

*** showed different behavior than the rest of the peptides

While many of the protein biomarkers discovered here have been previously associated with cancer, not all associations were discovered using gene and protein expression level changes. For example, the increased catalytic activity of calpain has been associated with ER+ breast cancer even though the protein's expression level is unchanged(172). Similarly, the post-translational modification of myosin-9 has been

shown to regulate the invasiveness of cancer cells in studies on MDA-MB-231 cell lines(173).

The protein hits with altered thermodynamic stabilities in the two comparative studies described here were classified into three groups based on whether or not they were detected as hits in either one or both of the cell line comparisons described here. A total of 80 of the 162 protein hits with altered thermodynamic stabilities in the two cell line comparisons, were assayed with the same methionine-containing peptide probe in both cell line comparisons. Thus, only these 80 proteins hits could be unambiguously classified into one of the three groups described below. These 80 protein hits included 13 proteins and 16 methionine-containing peptide probes with altered thermodynamic stabilities in both the MCF-7 versus MCF-10A and MCF-7 versus MDA-MB-231 cell line comparisons (Figure 24A and Table 13). A total of 15 of these 16 probes had similar SILAC-SPROX behavior in each comparison (Table 13), indicating the thermodynamic stability changes in the proteins or protein domains probed by these 15 peptide probes were specific to the MCF-7 cell line. Representative plots of the SILAC-SPROX behaviors of four of these proteins are shown in Figure 25. Thus, these proteins have the potential to serve as biomarkers for MCF-7 specific disease-related changes.

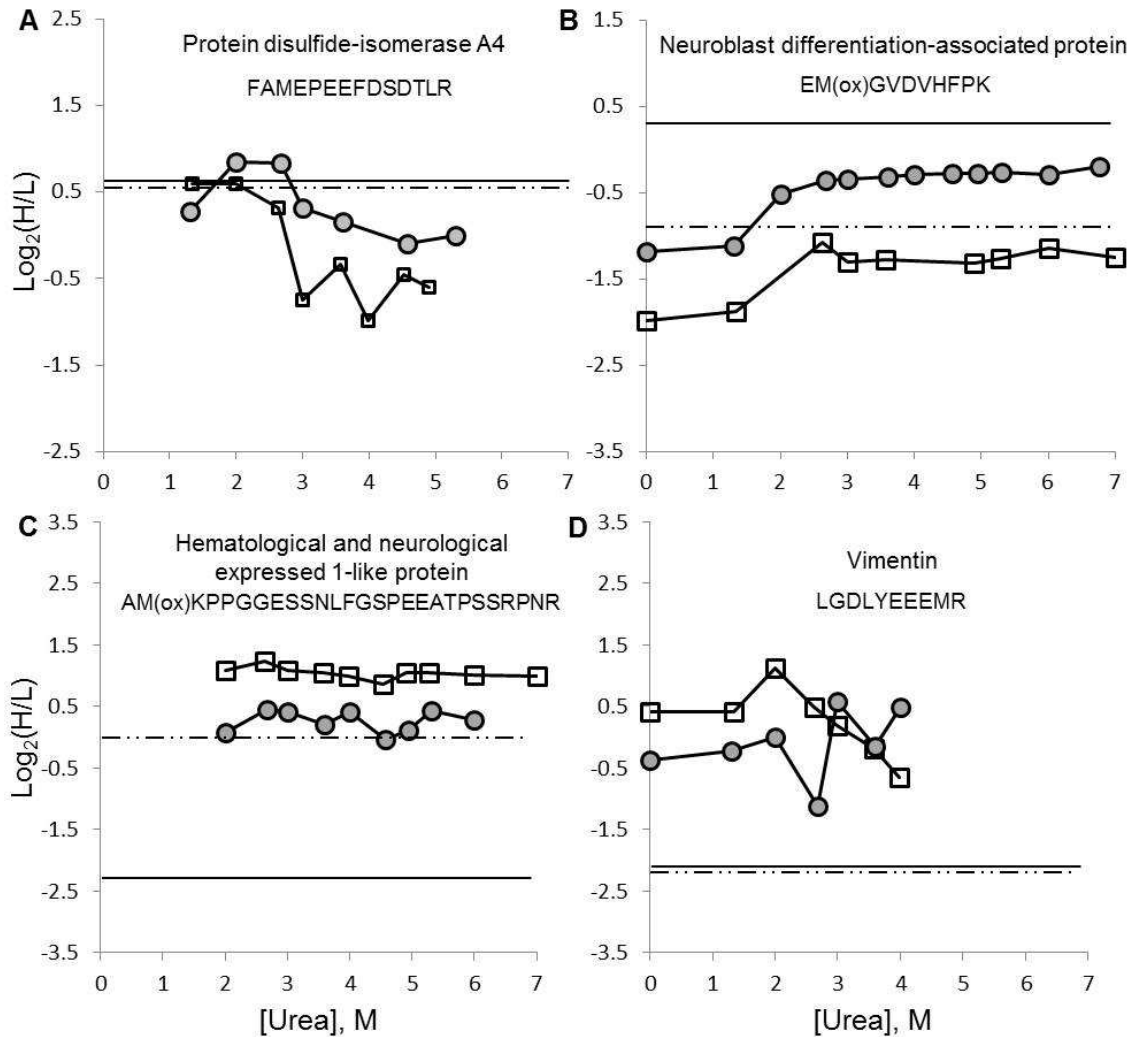


Figure 25: Representative SILAC-SPROX behaviors of four MCF-7 specific proteins. Shown are the data in the MCF-7 versus MDA-MB-231 cell line comparison (empty squares) and the MCF-7 versus MCF-10A cell line comparison (filled circles) from biological replicate 3 and biological replicate 2 respectively. The solid and dotted lines in each plot indicate the H/L ratio determined for each protein in the MCF-7 versus MDA-MB-231 and MCF-7 vs MCF-10A cell line comparisons, respectively.

The different SILAC-SPROX behavior of the HEAM(ox)ITDLEER peptide probe from myosin-9 (Figure 26) in the two comparisons, indicates that this protein has different protein folding behavior in all three cell lines. The SILAC-SPROX behavior of the HEAM(ox)ITDLEER peptide probe in the two comparisons suggests that myosin-9 is destabilized in both of the cancer cell lines, but more so in the MBA-MD-231 cell line, suggesting that the thermodynamic stability of myosin-9 can be used to differentiate breast cancer subtypes. Our observation that myosin-9 was highly destabilized in the MDA-MB-231 cell lysate could be because of the altered phosphorylation of this protein in MDA-MB-231 cells, which has been previously reported(173). Such a destabilization could result from altered protein-protein interaction and/or conformational changes induced by the altered phosphorylation. Significantly, the potential effects (both direct and indirect) of such mutations and PTMs on a protein's folding stability can be detected in the SILAC-SPROX experiment without actually detecting a methionine-containing peptide probe that includes the mutation or PTM site. This is because the chemical denaturation of a single methionine-containing peptide probe can report on the folding behavior of the protein or protein domain to which it maps. However, additional studies are required to understand the biological significance of the specific stability changes observed for each protein hit.

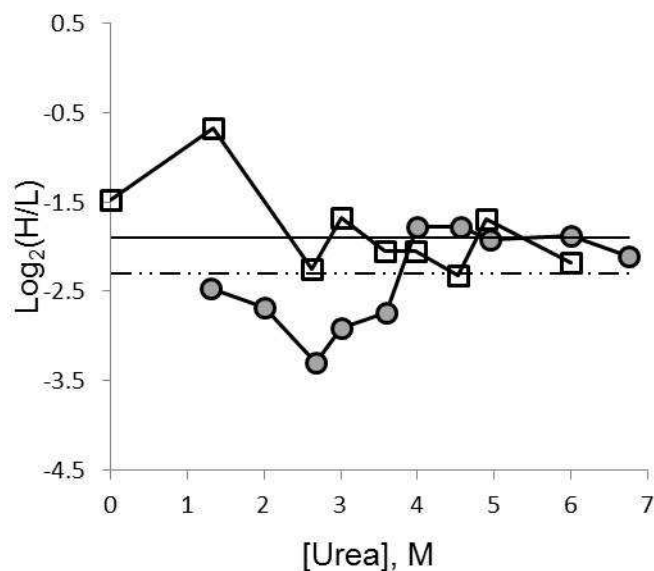


Figure 26: The SILAC-SPROX plots of the HEAM(ox)ITDLEER peptide from myosin-9 in the MCF-7 versus MDA-MB-231 (empty squares) and the MCF-7 versus MCF-10A (filled circles) cell line comparisons from biological replicate 3 and biological replicate 2 respectively. The solid and dotted lines in each plot indicate the H/L ratio determined for each protein in the MCF-7 versus MDA-MB-231 and MCF-7 vs MCF-10A cell line comparisons, respectively.

A total of 30 proteins (Figure 24 B) were assayed with the same 58 methionine-containing peptide probes (Appendix B.1) in both comparative studies and only identified as hits in the MCF-7 versus MCF-10A comparison. The thermodynamic stability changes probed by these 58 methionine-containing peptides were similar in the two cancer cell lines but different with respect to the non-tumorigenic control cell line.

These 30 proteins represent general breast cancer biomarkers for the normal to cancer transition. Similarly, 37 proteins (Figure 24C) were assayed with the same 39 methionine-containing peptide probes in both comparative studies and only identified as hits in the MCF-7 versus MDA-MB-231 comparison (Appendix B.2). These 37 proteins have MDA-MB-231 specific protein folding properties and may be useful biomarkers for the diagnosis and further characterization of triple negative (i.e., ER-, PgR- and HER2-) tumors. Together, these results show that thermodynamic stability profiles can be used to differentiate the cell line models of breast cancer studied here, and they suggest that such thermodynamic stability profiles can be useful for disease state characterizations.

5.3.4 Classification of Proteins with Altered Thermodynamic Stability

A bioinformatics analysis was performed using PANTHER (174, 175) to characterize the distribution of the 84 and 102 protein hits identified in the two cell line comparisons. In general, the biological processes and molecular functions represented in the protein hits identified using the thermodynamic data in this work were similar to the biological processes and molecular functions represented in the protein hits identified using the protein expression level data generated in this work (Appendix B.3). However, one exception was the higher percentage of hits with structural functions observed in the thermodynamic analyses (Appendix B.3 Table B.3.1). Analysis of the protein classes

represented in the hits also revealed that cytoskeletal proteins and chaperone proteins were enriched in the thermodynamic analyses (Figure 27).

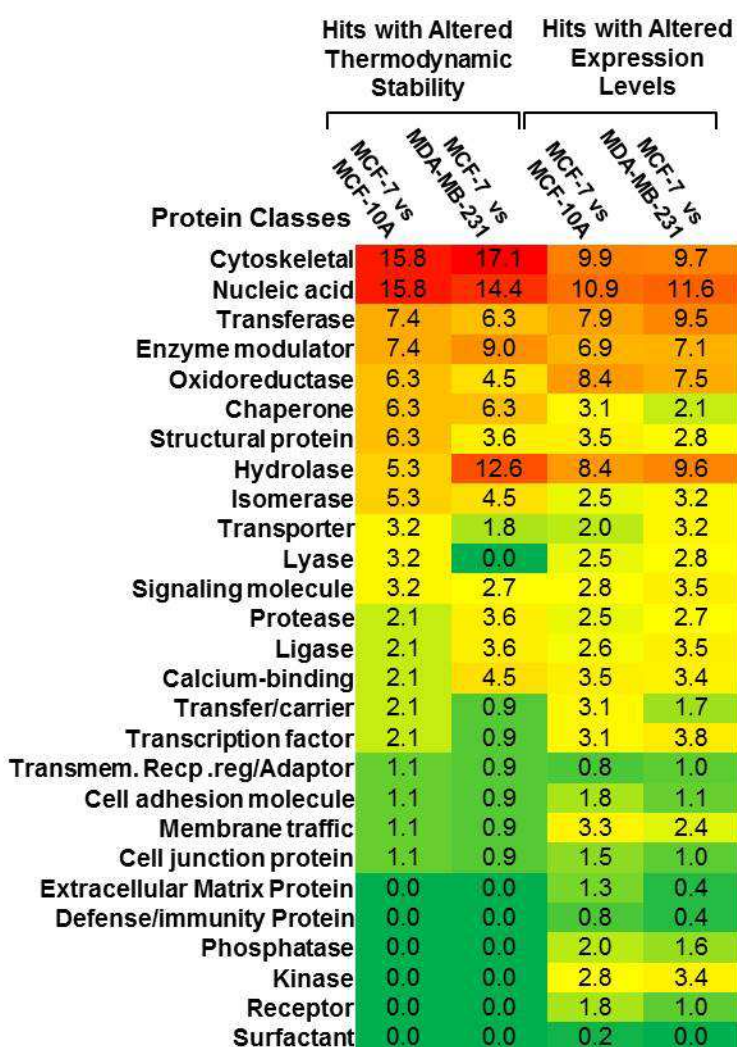


Figure 27: Heat map showing the distribution of protein classes observed in the protein hits identified from the thermodynamic and protein expression level analyses performed in this work. The numbers in the boxes represent the percentage of protein hits from each class in each experiment.

The additional molecular space probed using the thermodynamic stability measurements in this study enabled distinctions between the breast-cancer-related disease states in this study that were not possible using more conventional measurements of protein expression levels. The bioinformatics analysis of the protein hits identified in the two cell line comparisons revealed that the fraction of hydrolases hits in the MCF-7 versus MDA-MB-231 cell line comparison was more than double that found in the MCF-7 versus MCF-10A comparison (Figure 27). This is consistent with activity-based profiling studies, in which serine hydrolase activity profiles were found to differentiate MCF-7 and MDA-MB-231 cell lines (152). Our results add to the growing evidence that hydrolases are important for cancer cell growth and invasiveness. Significantly, the expression level data does not reveal such a signature, as the percentage of hydrolase hits in the expression level analysis was similar in the two subtypes of the breast cancer cell line models (Figure 27). This is especially noteworthy as the fraction of hydrolases in the total assayed proteins in the thermodynamic and expression level analyses was similar (Appendix B.3 Table B.3.5).

5.3.5 Cross-Validation and Discovery using SILAC-PP Protocol

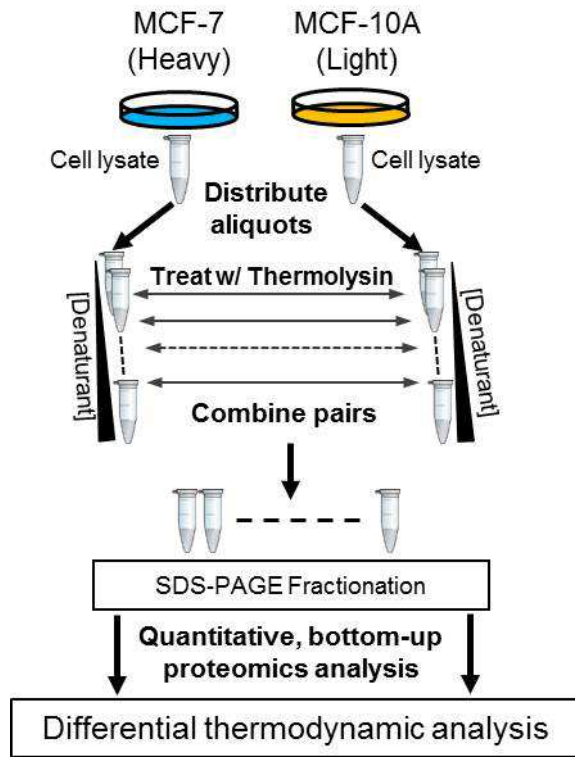


Figure 28: SILAC-PP workflow for MCF-7 versus MCF-10A comparison.

The SILAC-PP strategy used to probe disease related changes in proteome-wide thermodynamic measurement is outlined in Figure 28. The thermolysin cleavage and subsequent gel-based separation of the intact proteins enables every peptide that is successfully identified and quantified in the gel-based proteomics readout (not just the methionine-containing peptides) to report on the thermodynamic properties of the protein from which it was derived. This makes it a suitable technique for cross-

validation of the protein targets with SILAC-SPROX. Thus, specific molecular weight regions of the SDS-PAGE gel were excised and analyzed. The goal of this study was to demonstrate the protocol's ability to discover and cross validate the protein targets in a disease state analysis. In this experiment, two molecular weight regions of the gel, 36-50 kDa and 50-75 kDa, were analyzed (Figure 20).

Table 14: Proteomic Coverage obtained in the MCF-7 versus MCF-10A comparative study using SILAC-PP protocol.

MCF-7 vs MCF-10A Gel Bands (MW)	Total Proteins		
	Identified	Assayed ¹	Hits
36-50kDa	684	183	34
50-75kDa	685	194	40
Total	1043	288	67

¹Includes the proteins that were successfully identified and quantified in samples from at least half of the denaturant concentrations (6 or more).

The proteomic coverage of the SILAC-PP experiment with the MCF-7 and MCF-10A cell lysates are shown in Table 14. The described protocol successfully identified a total of 67 unique protein hits to have altered thermodynamic stability in the two gel regions that were analyzed (see Table 14 and Appendix B.4). One goal of the SILAC-PP experiment was to cross-validate the protein behaviors observed in SILAC-SPROX experiments involving the same cell lines comparison.

Table 15: Proteins with the altered thermodynamics in MCF-7 versus MCF-10A comparison identified in both SILAC-PP and SILAC-SPROX experiments.

Protein Hits ¹	Protein Hits ²
Protein names Glyceraldehyde-3-phosphate dehydrogenase 4-aminobutyrate aminotransferase, mitochondrial Actin-related protein 3 Secernin-1 Elongation factor 1-alpha 1 Keratin, type II cytoskeletal 5 Protein disulfide-isomerase A4 Nuclear autoantigenic sperm protein Glyceraldehyde-3-phosphate dehydrogenase Serpin B5 Neuroblast differentiation-associated protein AHNAK Heat shock cognate 71 kDa protein Protein disulfide-isomerase A4 Moesin Heat shock 70 kDa protein 1A/1B Vimentin 78 kDa glucose-regulated protein Annexin A2 ATP-citrate synthase Keratin, type II cytoskeletal 5	Protein disulfide-isomerase A3 Phosphoglycerate kinase 1 60 kDa heat shock protein, mitochondrial 26S protease regulatory subunit 7 Elongation factor 1-alpha 1 LIM and SH3 domain protein 1 Stress-70 protein, mitochondrial Tubulin beta-4B chain T-complex protein 1 subunit eta Heterogeneous nuclear ribonucleoprotein K Glucose-6-phosphate 1- dehydrogenase
	¹ Consistent in biological replicates in SILAC-SPROX. ² Includes proteins that were deemed false positives due to inconsistent hit behavior in biological replicates in SILAC- SPROX.

A comparison of the final hit lists from SILAC-PP and SILAC-SPROX experiments in MCF-7 versus MCF-10A cell lines successfully confirmed 17 of the 67 protein hits identified here (**Table 15**). One of the 17 protein, Glyceraldehyde-3-

phosphate dehydrogenase, that showed hit behavior in both SILAC-PP and SILAC-SPROX is shown in Figure 29.

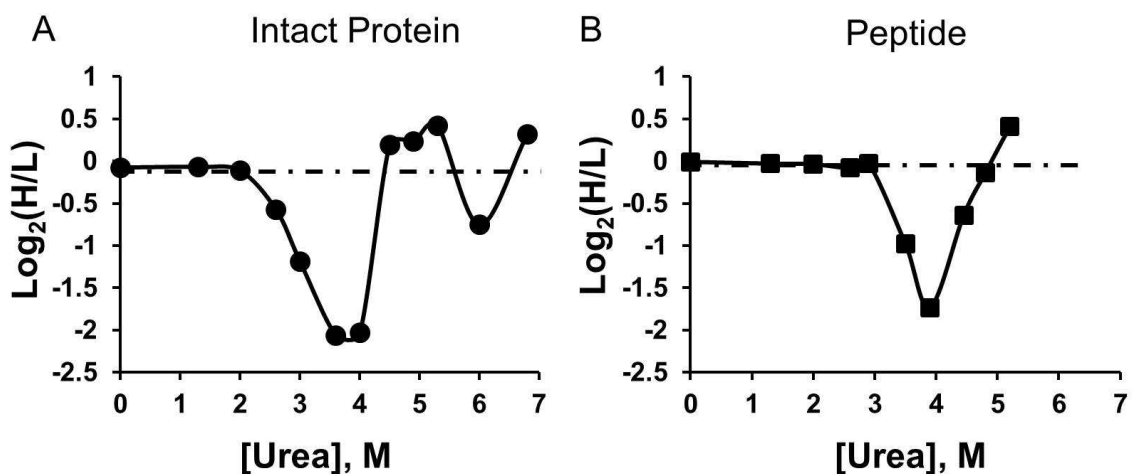


Figure 29: Data confirming the altered thermodynamic stability for the protein, Glyceraldehyde-3-phosphate dehydrogenase, using (A) SILAC-PP and (B) SILAC-SPROX. The data in (A) is for the intact protein and (B) is from the un-oxidized peptide VIISAPSADAPMFVMGVNHEK.

An additional 11 of the 67 protein hits showed hit behavior in at least one of the three biological replicates in SILAC-SPROX but were excluded from the final list because of discrepancies in between the replicates or in retention time and other parameters (Table 15). The SILAC-PP and SPROX experiments utilize two different reaction probes (thermolysin versus oxidation) and mass spectrometry readouts (e.g., quantitation at the peptide versus protein level). Thus, this cross-validation provides more confidence in

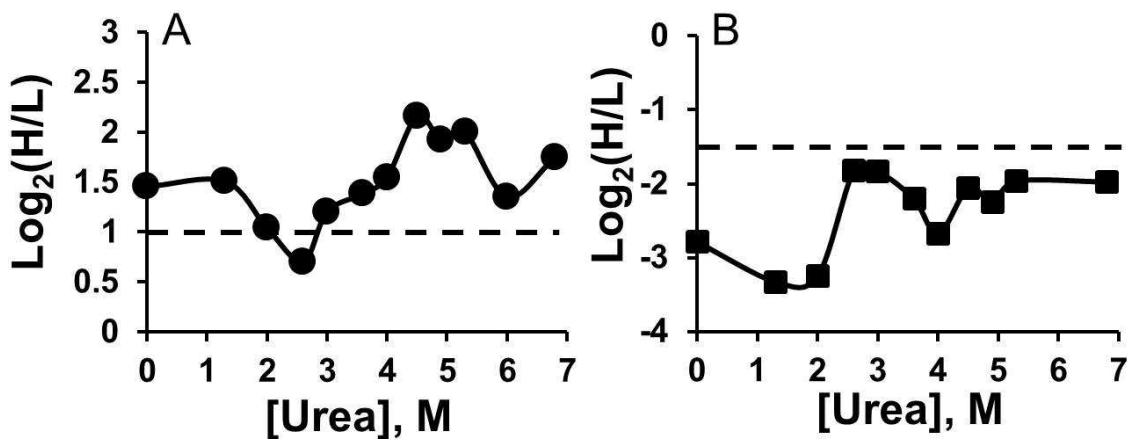


Figure 30: Data showing the altered thermodynamic stability behavior of the catalytic subunit from (A) calpain-1 and (B) calpain-2 obtained from 50-75kDa gel band region of MCF-7 versus MCF-10A SILAC-PP experiment.

these 28 thermodynamically altered protein targets identified in disease state analysis here. A total of 39 of the 67 protein hits identified here by SILAC-PP were not identified as hits in the SILAC-SPROX experiment. Cross-examination of the data revealed that 12 of these 39 proteins were not assayed in the SILAC-SPROX experiment. However, the remaining 27 protein targets were assayed but did not appear as hit in the SILAC-SPROX experiment. Interestingly, some of these 27 proteins includes proteins such as calpain-1 and calpain-2 catalytic subunits (Figure 30) that are known to have catalytic activity differences between the normal and ER+ breast cancer tumors(172).

5.3.6 Correlating Changes in Thermodynamic Stability with Changes in Function

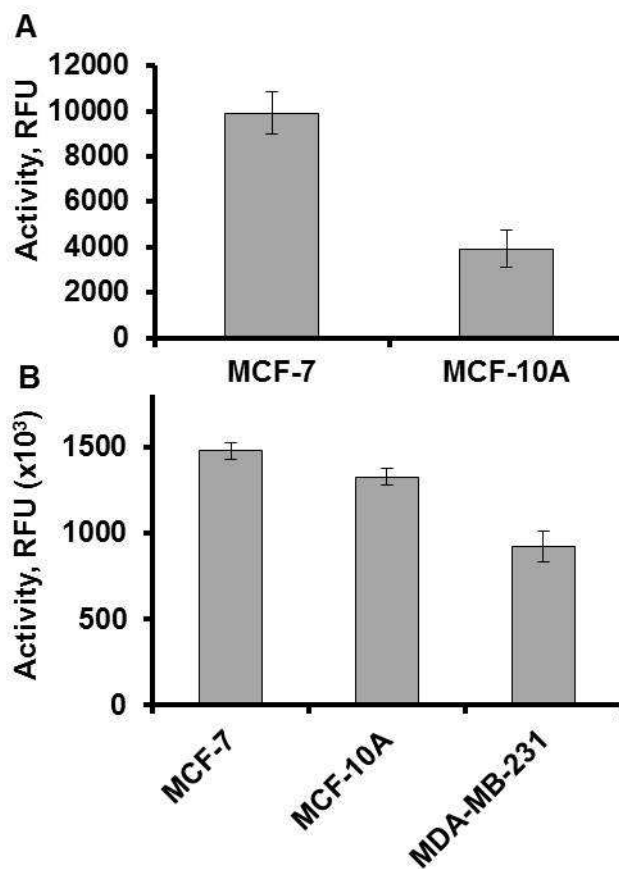


Figure 31: Catalytic activity data for calpain protease in MCF-7 lysate and MCF-10A lysate (A) and for cathepsin D protease in MCF-7, MCF-10A, and MDA-MB-231 cell lysates (B). Error bars represent +/- one standard deviation from duplicate measurements.

The thermodynamic stability changes observed for the protein hits identified in this work could arise for a wide variety of functionally relevant reasons (e.g., point mutations, post-translational modifications, and new and/or altered binding interactions

with cellular ligands). Additional studies are required to understand the biological significance of the specific stability changes observed for each protein hit. Biological activity assays performed with hit proteins that have known catalytic functions provides one way in which to correlate a protein hit's altered stability with a change in function. As part of this work, the catalytic activities of Calpain and Cathepsin D, were assayed to determine if the altered thermodynamic stabilities of these proteins could be correlated with changes in their known proteolytic activity. The proteolytic activity of calpain that showed increased stability (for calpain I with probes from both calpain small subunit-1, CAPNS1 (SILAC-SPROX), and calpain-1 catalytic subunit, CAPN1 (SILAC-PP)) in the MCF-7 cell line, was evaluated using a known peptidic substrate of calpain that was spiked into MCF-10A and MCF-7 cell lysates. The catalytic activity of calpain was approximately 2.5-fold greater in the MCF-7, ER+ cell line compared to the activity measured in the normal cell line model (Figure 31A). These catalytic activity measurements are consistent with earlier studies of calpain activity in cell lysates from normal and ER+ breast cancer tissues, in which the proteolytic activity of calpain was also found to be greater in ER+ breast cancer tissues than in normal breast tissues(172).

The catalytic activity of Cathepsin D was assayed in the MCF-10A, MCF-7, and MDA-MB-231 cell lines using a known peptidic substrate that was spiked into cell

lysates derived from each cell lines. The thermodynamic stability of Cathepsin D was decreased in the MCF-7 and MDA-MB-231 cell lysates compared to that in the MCF-10A cell lysate. The catalytic activity of Cathepsin D was slightly enhanced in the MCF-7 cell line and slightly reduced in the MDA-MB-231 cell line relative to the activity measured in the MCF-10A cell line (Figure 31B). Thus, in the case of Cathepsin D there was not a correlation between the protein's proteolytic activity and its decreased stability in the MCF-7 and MDA-MB-231 cell lysates.

5.4 Conclusion

The differential stability measurements described here create a discovery platform for disease state biomarkers. The results of this work demonstrate that global measurement of thermodynamic stability can be used to detect disease-related changes in mammalian cell lines. The discovery of hit proteins in this comparative study with a wide range of altered expression levels suggests that the described methodology can be used to complement protein expression level studies. The thermodynamic stability measurements strategy presented here shows promise for both validating protein targets from other strategies utilizing proteomic platforms such as protein expression levels and also for uncovering novel molecular targets in future disease-related characterization studies.

6. Conclusions and Future directions

The work presented in this thesis is the first demonstration of the utility of the thermodynamic protein folding measurements in disease state characterization studies. The major contribution of this study involve i) the development of two SILAC-based approaches for protein folding and thermodynamic measurements, and ii) subsequent demonstration of the differential measurement of protein folding thermodynamics in breast cancer cell lines. Initially, the SILAC-based energetics approaches, SILAC-SPROX and SILAC-PP, were developed and their applicability was demonstrated in proof-of-principle studies using proteins in a yeast cell lysate and two test ligands. Both of the approaches utilized the shotgun quantitative proteomic strategy using mass-spectrometry based platform to assay thermodynamic stability changes and offer many advantages to ligand binding experiments. A major advantage of energetics-based methods, like the SILAC-SPROX and SILAC-PP approaches described here, is that they do not require the synthesis of specially designed ligands. This makes the methodology easy to adapt for the study of other ligands. The SILAC-based quantitation coupled with SPROX allowed isotopic labeling of proteins (metabolic labeling) and thus samples could be combined immediately after the SPROX analysis. The quantitation in SILAC is performed with precursor ions (MS1). This eliminated the several drawbacks in iTRAQ-

SPROX protocol arising from peptide level labeling of the sample and product ion (MS2) quantitation. These improvements slightly lowered the false positive rates in SILAC-SPROX protocol to 0.4-3.4% from 1-5% previously reported for iTRAQ-SPROX experiments (64, 90). The incorporation of SILAC into PP enabled the PP technique to be used for the unbiased detection and quantitation of protein-ligand binding interactions in complex biological mixtures (e.g., cell lysates) without the need for pre-fractionation.

The two SILAC-based approaches developed as part of this thesis complement each other and help cross-validate the protein targets identified with these strategies. The SILAC-PP and the previously reported Stability of Protein from Rates of Oxidation (SPROX) technique both report on the same thermodynamic properties of proteins and protein-ligand complex. However, they employ different probes and mass spectrometry based-readouts. This created the opportunity to cross-validate SPROX results with SILAC-PP results, and vice-versa. As part of this work, the SPROX results were cross-validated with SILAC-PP results obtained on the same model systems to help differentiate true positives from false positives in the two experiments.

Using the SILAC-SPROX technique, ~800 proteins were assayed for changes in their protein folding behavior in three different cell line models of breast cancer including the MCF-10A, MCF-7, and MDA-MB-231 cell lines. A total of 84 and 102

proteins in the MCF-7 versus MCF-10A and MCF-7 versus MDA-MB-231 cell line comparative analyses, respectively exhibited differential stability in cell lysates prepared from the different cell lines. Thermodynamic profiling differences of 28 proteins identified with SILAC-SPROX strategy in MCF-10A versus MCF-7 cell line comparison were also confirmed with SILAC-pulse proteolysis (SILAC-PP) technique. The differentially stabilized protein hits in these studies encompassed those with a wide range of functions and protein expression levels, and they included a significant fraction (~45%) with similar expression levels in the cell line comparisons.

In the case of protein hits for which cell-based activity assays do not exist, or for which a specific known activity is not differentially modulated (such as was the case for cathepsin D), other studies are required to connect a protein hit's altered stability with a change in function. For example, immunoprecipitation strategies can be used to purify the "hit" proteins from the different cell lysates so that additional SPROX experiments can be performed to determine if the originally detected stability differences in the unpurified samples are also observed in the purified protein samples. Detection of the same stability difference would suggest that a differential post-translational modification or another covalent modification to the "hit" protein (e.g., a point mutation) might be responsible for the detected stability change. In this case, the

isolated protein samples could be subjected to bottom-up and/or top-down MS-based strategies to more completely characterize the covalent structure of the protein. On the other hand, if the isolated protein samples do not show similar stability differences in the purified and unpurified samples, it may be that the originally observed stability difference in the unpurified sample was the result of an altered protein interaction network. In such cases, a differential proteomic analysis of the protein complexes pulled down with the “hit” protein in the immunoprecipitation experiments could be used to help dissect the protein interaction networks that may be altered in the biological states being compared.

The use of well-established cell lines was especially convenient in this initial work as the cell lines provide a convenient source of highly homogeneous cells representing initial tumors. It has been shown that some cell lines, including the breast cancer cell line in this study, can faithfully represent many features of cancer cells *in vivo*(176). However, there are certain limitations when using cell lines such as the loss of heterogeneity of the original tumors and potential for phenotypic changes in *in vitro* culture flasks(177). Thus, it is certainly more ideal to directly study clinical samples. In theory the described approach can also be applied to the analysis of tumor tissues cell lysates derived from clinical samples provided sufficient amounts of protein (~2-3 mg of

total protein) can be obtained from such samples. It is also possible that protein hits identified with stability changes in cell line analyses could be validated in clinical samples using the SPROX methodology to directly measure the stability of potential hits in tumor tissue cell lysates. Such validation studies could utilize targeted proteomics methods, which require significantly less sample than the shot-gun proteomics methods described here. Alternatively, the physical cause (e.g., point mutation) and/or the functional effect (e.g., catalytic activity change) of the altered stability differences could be exploited to validate protein hits in clinical samples.

The differential stability measurements described here create a novel discovery platform for disease state biomarkers. The results of this work demonstrate that global measurement of thermodynamic stability can be used to detect disease-related changes. The discovery of hit proteins in this comparative study with a wide range of altered expression levels suggests that the described methodology can be used to complement protein expression level studies. The thermodynamic stability measurements strategy presented here shows promise for both validating protein targets from other strategies utilizing proteomic platforms such as protein expression levels and also for uncovering novel molecular targets in future disease-related characterization studies.

Appendix A: SILAC-Pulse Proteolysis Protocol.

A.1 CsA-Binding Experimental Parameters

<i>CsA-binding Experiment(s)</i>	1	2	3	4
<i>Final CsA concentration in each reaction tubes</i>	250 μ M	250 μ M	250 μ M	250 μ M
<i>Reaction time, min</i>	1	1	1	1
<i>Total Protein/reaction tube ,μg</i>	160	88	159	143
<i>Thermolysin used/reaction tube, μg</i>	16	10	12	16
<i>Amount of EDTA Quenching solution</i>	35 μ l of 50mM	15 μ l of 0.5M	15 μ l of 0.5M	15 μ l of 0.5M
<i>Denaturant Used,(M)</i>	[Urea]	[Urea]	[Urea]	[Urea]
1	0	0	0	0
2	1	0.95	0.85	0.9
3	1.35	1.34	1.28	1.29
4	2	2	2	2
5	2.6	2.6	2.6	2.58
6	3	3	2.94	2.94
7	3.4	3.36	3.36	3.28
8	3.6	3.6	3.5	3.5
9	4	3.94	3.9	3.85
10	5	5.3	5.22	5.13
11	6	6	6	5.87
12	6.8	6.9	6.66	6.8

A.2 ATP-Binding Experimental Parameters

<i>ATP-binding Experiment(s)</i>	1	2
<i>Final ATP Concentration</i>	20.6 mM	20.6 mM
<i>Final MgCl₂ concentration</i>	25 mM	25 mM
<i>Reaction time, min</i>	1	1
<i>Total Protein/sample, µg</i>	452.5	130
<i>Thermolysin used/sample, µg</i>	13	13.9
<i>Amount of EDTA Quenching solution</i>	2.08µl of 0.5M EDTA	35µl of 50mM EDTA
<i>Denaturants Used</i>	[Urea] (M)	[Urea] (M)
1	0	0
2	0.98	0.9
3	1.5*	1.3
4	2	2
5	2.48	2.6
6	2.78	3
7	3.3	3.3
8	3.98	3.6
9	4.9	3.9
10	6	4.8
11	-	5.9
12	-	6.7

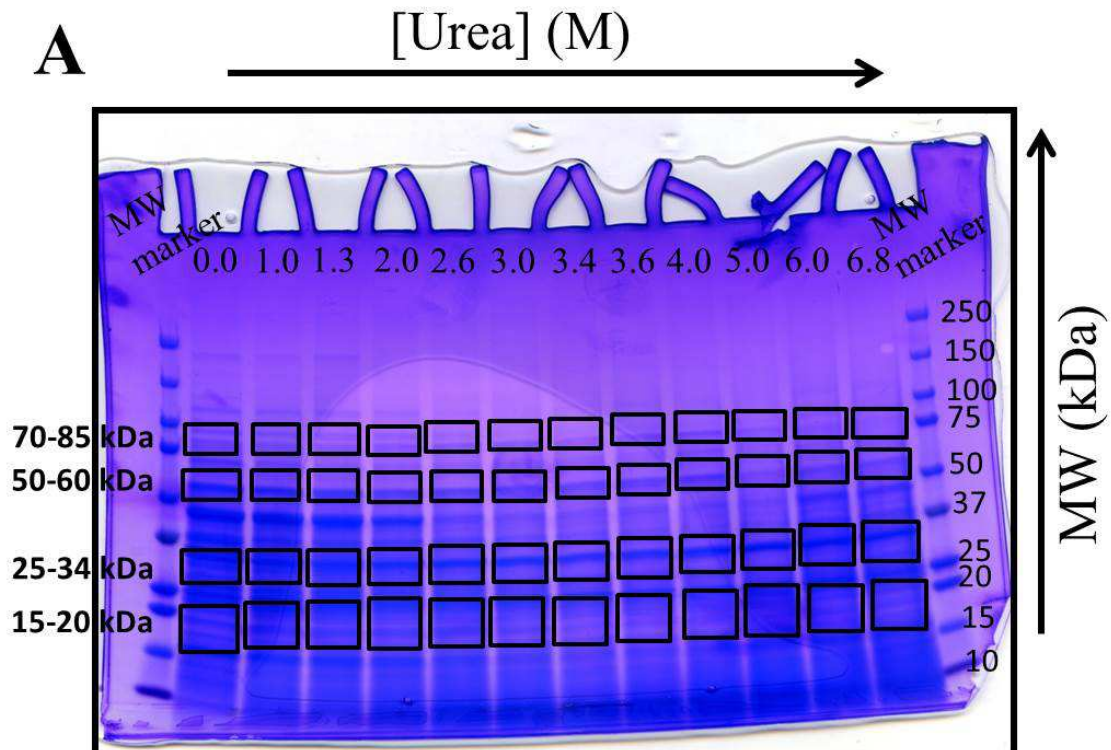
* No data from this denaturant concentration was included in the analysis of the 80 kDa band. This is because the LC-MS/MS analysis of this sample was unsuccessful (i.e., no peptide and protein Id's were obtained).

A.3 Control Experiment Experimental Parameters

<i>Control Experiment(s)</i>	1	2
<i>Ligand Mimic Used</i>	<i>Lysis Buffer</i>	<i>Lysis Buffer</i>
<i>Reaction time, min</i>	1	1
<i>Total Protein/sample, μg</i>	160	200
<i>Thermolysin used/sample, μg</i>	15.66	18
<i>Amount of EDTA Quenching solution</i>	35 μl of 50mM	15 μl of 0.5M
<i>Denaturants Used</i>	<i>[Urea] (M)</i>	<i>[Urea] (M)</i>
1	0	0
2	1	0.92
3	1.35	1.33
4	2	2.06
5	2.6	2.66
6	3	3.03
7	3.4	3.41
8	3.6	3.6
9	4	3.98
10	5	4.94
11	6	6.05
12	6.8	6.76

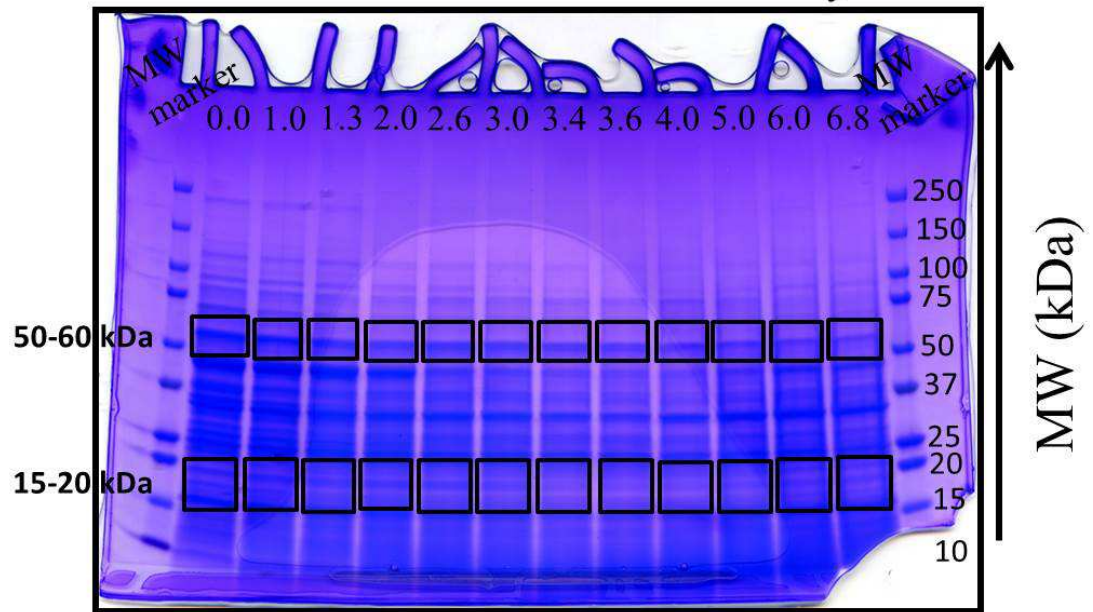
A.4 SDS-PAGE Gels Generated in the SILAC-PP Experiments (A-E).

A.4.A-E The gels are those generated in (A) Control Experiment 1, (B) CsA-binding Experiment 1, (C) CsA-binding Experiment 3, (D) CsA-binding Experiment 4, and (E) ATP Experiment 1. The gel regions excised in each of the experiments are indicated with rectangular boxes.

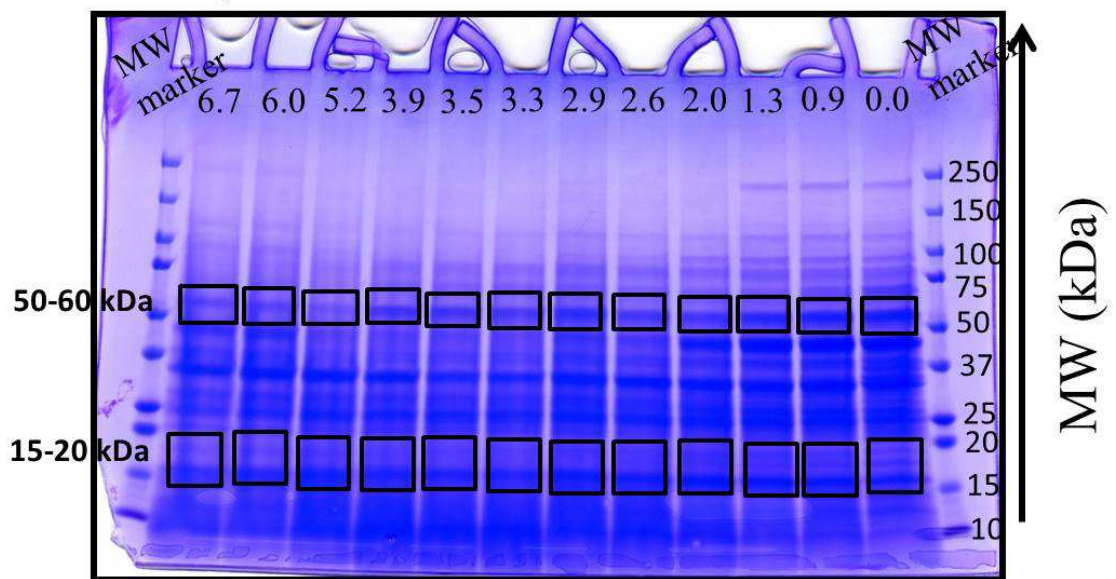


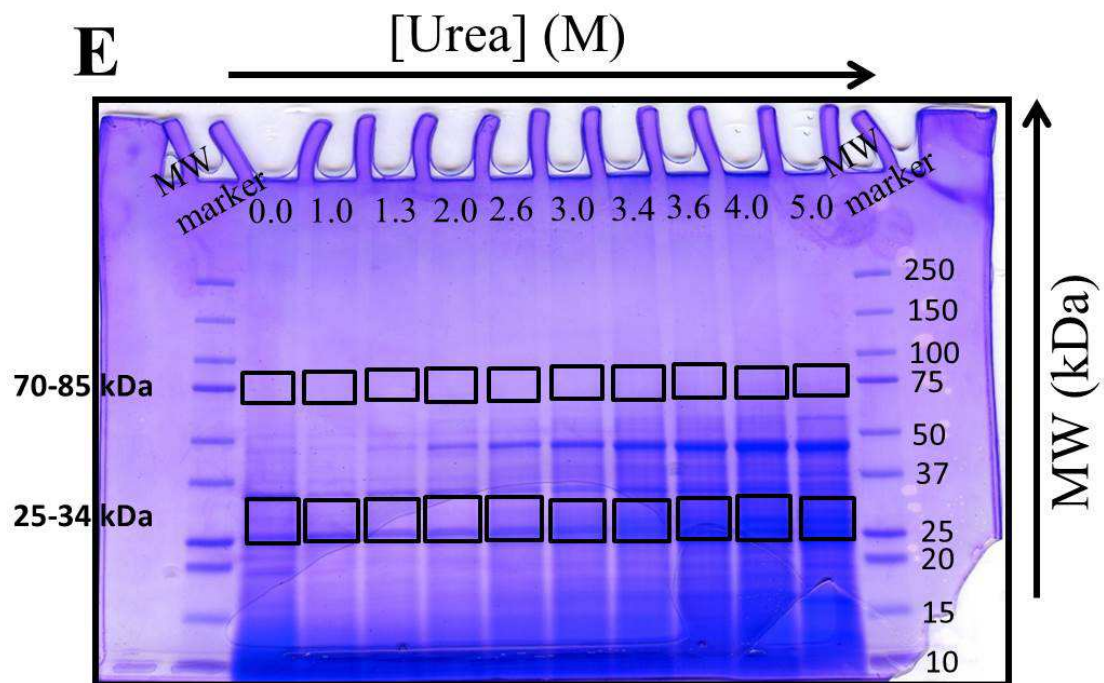
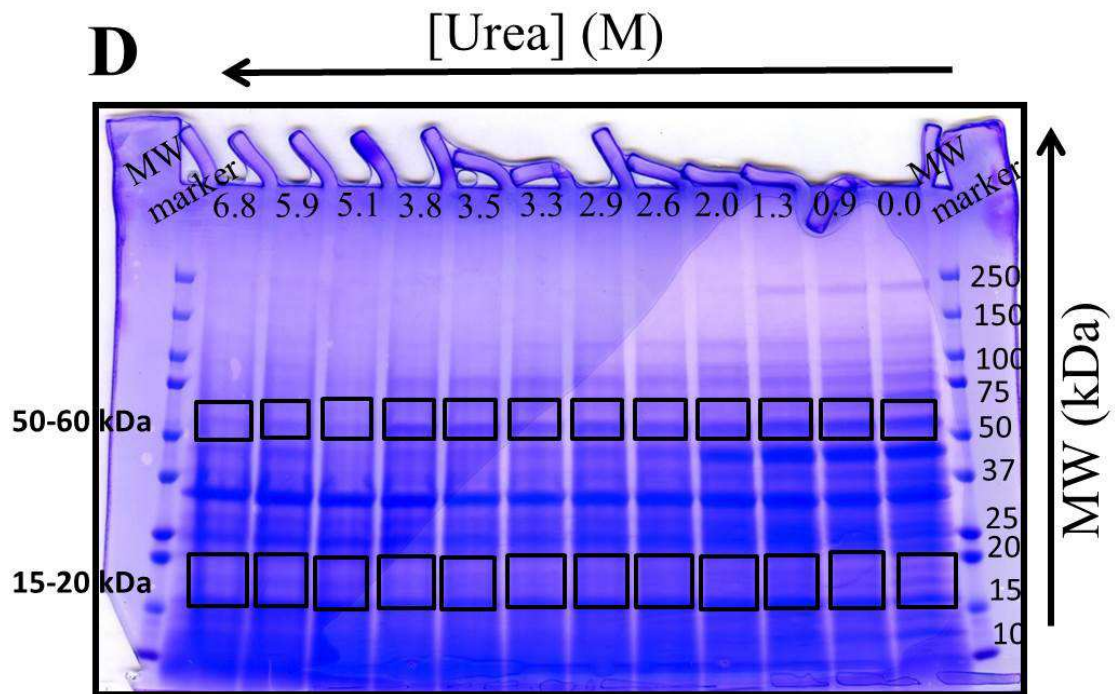
B

[Urea] (M)

**C**

[Urea] (M)





Appendix B: Disease State Analysis with Breast Cancer Cell Lines

B.1 Peptides and Proteins with Altered Thermodynamics in MCF-7 vs MCF-10A

Peptide(s)	Protein(s)	Bio. Replicate(s)			Median H/L **	Literature H/L ***	Refs
		1*	2*	3*			
<u>_ATPM(ox)PSRPSTTPFIDK_</u>	E7EVA0	X	NA	NA	0.46	0.21	1,2
<u>_ATPM(ox)PSRPSTTPFIDKKPTSAPKPSSTTPR_</u>	E7EVA0	NA	X	NA	0.46	0.21	
<u>_VIISAPSADAPM(ox)FVM(ox)GVNHEK_¶</u>	P04406	X	X	X	0.94	1.87	3,4
<u>_VIISAPSADAPM(ox)FVMGVNHEK_¶</u>	P04406	X	X	X	0.94	1.87	
<u>_VVDLM(ox)AHM(ox)ASK_¶</u>	P04406	X	X	TR	0.94	1.87	
<u>_VVDLM(ox)AHM(ox)ASKE_¶</u>	P04406	X	X	X	0.94	1.87	
<u>_LTGMAFR_¶</u>	P04406	NA	X	X	0.94	1.87	
<u>_RVIISAPSADAPM(ox)FVM(ox)GVNHEK_¶</u>	P04406	NA	X	X	0.94	1.87	
<u>_VIISAPSADAPM(ox)FVM(ox)GVN(de)HEK_¶</u>	P04406	TR	X	X	0.94	1.87	
<u>_VIPELN(de)GKLTGMAFR_¶</u>	P04406	NA	X	NA	0.94	1.87	
<u>_VVDLMAHMASK_¶</u>	P04406	NA	X	X	0.94	1.87	
<u>_VIISAPSADAPMFVMGVN(de)HEK_¶</u>	P04406	NA	TR	X	0.94	1.87	
<u>_VIISAPSADAPMFVMGVNHEK_¶</u>	P04406	NA	TR	X	0.94	1.87	
<u>_VVDLMAHMASKE_¶</u>	P04406	NA	TR	X	0.94	1.87	
<u>_SM(ox)VAVM(ox)DSDTTGK_</u>	P04632	X	X	TR	0.62	0.61	5,6
<u>_SMVAVMDSDTTGK_</u>	P04632	NA	X	X	0.62	0.61	

<u>_ETM(ox)QSLNDR_</u>	P05783	PD	NA	X	11.04	9.34	7,8,9
<u>_IDKLM(ox)IEM(ox)DGTENK_</u>	P06733	TR	X	X	0.61	0.65	10,11
<u>_IDKLMIEMDGTENK_</u>	P06733	NA	X	NA	0.61	0.65	
<u>_LAM(ox)QEFM(ox)ILPVGAAN(de)FR_</u>	P06733	NA	X	X	0.61	0.65	
<u>_LAM(ox)QEFM(ox)ILPVGAANFR_</u>	P06733	NA	X	X	0.61	0.65	
<u>_LAQAN(de)GWGVM(ox)VSHR_</u>	P06733	NA	X	X	0.61	0.65	
<u>_LMIEMDGTENK_</u>	P06733	NA	NA	X	0.61	0.65	
<u>_IHPVSTMVK_</u>	P07195	NA	X	X	0.05	0.01	12,13
<u>_GEM(ox)M(ox)DLQHGSFLQTPK_</u>	P07195	NA	AH	X	0.05	0.01	
<u>_EALPAPSDDATALMTDPK_</u>	P07203	TR	X	X	0.29	0.08	14,15
<u>_DPDAQPGGELMLGGTDSK_</u>	P07339	NA	X	TR	5.47	2.34	16,17
<u>_TMSEVGGSVEDLIAK_</u>	P07339	NA	X	X	5.47	2.34	
<u>_ISVNNVLPVFDNLMQOK_</u>	P07339	NA	NA	X	5.47	2.34	
<u>_SYSPYDM(ox)LESIR_</u>	P07355-2	X	TR	X	0.18	0.16	18
<u>_SYSPYDM(ox)LESIRK_</u>	P07355-2	NA	X	NA	0.18	0.16	
<u>_DAGTIAGLNVM(ox)R_</u>	P11021	X	TR	TR	1.43	0.97	19,20
<u>_DAGTIAGLNVMR_</u>	P11021	X	T/N	TR	1.43	0.97	
<u>_TFAPEEISAM(ox)VLTk_</u>	P11021	X	X	X	1.43	0.97	
<u>_M(ox)KETAEAYLGKK_</u>	P11021	NA	X	NA	1.43	0.97	
<u>_MKETAEAYLGK_</u>	P11021	NA	TR	X	1.43	0.97	
<u>_TFAPEEISAMVLTk_</u>	P11021	NA	TR	X	1.43	0.97	
<u>_N(de)SLESYAFNMK_</u>	P11142	X	TR	TR	0.90	0.88	21
<u>_NSLESYAFNMK_</u>	P11142	X	TR	TR	0.90	0.88	

<u>_SFYPEEVSSMVLTK_</u> ¶	P11142	X	TR	X	0.90	0.88	
<u>_NTDQASMPDNTAAQK_</u>	P35579	X	NA	NA	0.27	0.20	
<u>_HEAM(ox)ITDLEER_</u>	P35579	NA	X	NA	0.27	0.20	22,23
<u>_M(ox)QQNIQELEEQLLEESAR_</u>	P35579	NA	X	TR	0.27	0.20	
<u>_QIATLHAQVADMK_</u>	P35579	NA	X	NA	0.27	0.20	
<u>_SM(ox)AVAAR_</u>	P35579	NA	X	NA	0.27	0.20	
<u>_LTEMETLQSQLM(ox)AEK_</u>	P35579	NA	NA	X	0.27	0.20	
<u>_VTN(de)GAFTGEISPGM(ox)IK_</u> ¶	P60174	X	X	X	1.57	1.64	24
<u>_VTN(de)GAFTGEISPGMIK_</u> ¶	P60174	X	X	TR	1.57	1.64	
<u>_NALESYAFNMK_</u>	P08107	NA	TR	X	1.62	2.48	
<u>_FYEQM(ox)N(de)GPVAGASR_</u> ¶	E9PRY8	X	NA	NA	0.98	1.56	
<u>_VAAAESM(ox)PLLECAR_</u>	O00410-3	NA	NA	X	0.21	0.46	
<u>_N(de)GM(ox)LNVSPIGR_</u>	O15305	X	NA	NA	1.42	1.73	
<u>_GEMM(ox)DLQHGSFLR_</u>	P00338-3	NA	X	NA	0.20	0.56	
<u>_M(ox)AALEAK_</u> ¶	P05455	NA	NA	X	1.14	1.27	
<u>_VEVERDNLAEDIM(ox)R_</u>	P08670	NA	X	NA	0.24	0.05	
<u>_FAMEPEEFSDTLR_</u>	P13667	NA	X	NA	1.74	1.72	
<u>_AGEVINQPM(ox)MMAAR_</u>	P18206	X	NA	NA	0.29	0.46	
<u>_M(ox)LGQM(ox)TDQVADLR_</u>	P18206	X	NA	NA	0.29	0.46	
<u>_YDDMATCMK_</u> ¶	P27348	X	NA	NA	0.78	0.82	
<u>_ASANMDLM(ox)R_</u> ¶	P28838	X	NA	NA	1.92	1.04	
<u>_VMAAENIPENPLK_</u> ¶	P42224	X	NA	NA	2.43	1.07	
<u>_LVQSPNSYFMDVK_</u> ¶	P42677	NA	NA	X	0.51	0.81	

<u>_HIM(ox)GQNVADYM(ox)R_</u>	P46777	NA	X	NA	0.66	0.82	
<u>_GLVM(ox)VKPGSIKPHQK_</u>	P49411	NA	X	NA	2.07	1.70	
<u>_SIDTGM(ox)GLER_</u>	P49588	X	NA	NA	3.32	3.94	
<u>_EAATLEVERPLPMEVEK_</u>	P51858	NA	NA	X	1.08	1.34	
<u>_AFDSGIIPMEFVNK_</u>	P53396	NA	NA	X	0.31	0.96	
<u>_M(ox)DELQLFR_</u>	P55072	NA	X	NA	1.15	0.82	
<u>_AAHSEGNTTAGLDMR_</u>	P78371	NA	X	NA	0.72	1.00	
<u>_M(ox)ILELFSK_</u>	P78417	NA	X	NA	0.35	0.37	
<u>_SGEGEVSGLMR_</u>	Q13263	X	NA	NA	2.64	1.26	
<u>_SFM(ox)ALSQDIQK_</u>	Q14444	X	NA	NA	1.92	0.81	
<u>_(ac)M(ox)N(de)GQLDLGK_</u>	Q92734	X	NA	NA	1.43	1.86	
<u>_M(ox)KLPEHPEGGEPEDDEAPAK_</u>	Q9NX58	X	NA	NA	0.61	0.59	
<u>_EM(ox)GVDVHFPK_</u>	Q09666	NA	X	NA	1.20	0.42	
<u>_YGIQMPAFSK_</u>	P46940	NA	X	TR	0.83	0.72	
<u>_LESGMQNM(ox)SIHTK_</u>	P05787-2	NA	X	NA	10.14	NaN	
<u>_LTM(ox)QNLNDR_</u>	CON_P08727	X	NA	NA	6.57	NA	
<u>_SQYEVM(ox)AEQNR_</u>	CON_P08727	X	TR	X	6.57	NA	
<u>_GLVEPVNVVDN(de)GDGTHVTYTPSQEGPYM(ox)VSVK_</u>	O75369	NA	X	X	0.34	0.12	
<u>_SEIDMNDIK_</u>	P04083	X	X	NA	0.08	0.01	
<u>_KEESEESDDDM(ox)GFGFLD_</u>	P05387	NA	X	X	0.54	0.75	
<u>_KDVDEAYMNK_</u>	P05787-2	X	NA	NA	10.14	NaN	
<u>_LESGM(ox)Q(de)NM(ox)SIHTK_</u>	P05787-2	X	NA	NA	10.14	NaN	
<u>_LESGM(ox)QNM(ox)SIHTK_</u>	P05787-2	X	X	HN	10.14	NaN	

<u>_SLDM(ox)DSIIAEVK_</u>	P05787-2	X	X	HN	10.14	NaN	
<u>_TEM(ox)ENEFVLIKK_</u>	P05787-2	NA	X	NA	10.14	NaN	
<u>_LESGM(ox)QNMSIHTK_</u>	P05787-2	NA	TR	X	10.14	NaN	
<u>_IHPVSTM(ox)VK_</u>	P07195	NA	X	NA	0.05	0.01	
<u>_MVVESAYEVIK_</u>	P07195	NA	X	X	0.05	0.01	
<u>_GEM(ox)M(ox)DLQHGSFLQTPK_</u>	P07195	NA	X	TR	0.05	0.01	
<u>_IMGIPEEQMGLLR_</u>	P35579	NA	X	TR	0.27	0.20	
<u>_LTEM(ox)ETLQSQLMAEK_</u>	P35579	NA	NA	X	0.27	0.20	
<u>_HIFSEDTSDFSGM(ox)SETK_</u>	P36952	NA	X	X	0.12	0.02	
<u>_VCVETVESGAM(ox)TK_</u>	P48735	X	TR	HN	7.20	9.18	
<u>_LIDDMVAQVLK_</u>	P48735	NA	HN	X	7.20	9.18	
<u>_DAM(ox)QYASESK_</u>	Q00610	X	NA	NA	2.45	2.19	
<u>_VVGAMQLYSVDR_</u>	Q00610	NA	X	NA	2.45	2.19	
<u>_VEADVSLPSM(ox)QGDLK_</u>	Q8IVF2	X	X	X	0.15	0.10	
<u>_LEGDSLADKDM(ox)TAK_</u>	Q8IVF2	NA	X	X	0.15	0.10	
<u>_M(ox)ASTFIGN(de)STAIQELFK_</u>	Q9BUF5	NA	X	X	0.29	0.41	
<u>_M(ox)ASTFIGNSTAIQELFK_</u>	Q9BUF5	NA	X	X	0.29	0.41	
<u>_(ac)AAEPVEDN(de)CINRVAM(ox)K_</u>	Q14116	NA	X	NA	0.24	0.02	
<u>_(ac)M(ox)ESGFTSK_</u>	P40261	X	NA	NA	0.03	0.08	
<u>_AGM(ox)SAEQAQGLLEK_</u>	Q9Y2Q3-2	X	NA	NA	2.76	1.22	
<u>_AM(ox)SQDGASQFQEVIR_</u>	E7EUC7	X	NA	NA	0.58	0.07	
<u>_DLM(ox)ACAQTGSGK_</u>	O00571	X	NA	NA	0.32	0.79	
<u>_ELQEM(ox)DKDDESLIK_</u>	P52566	X	NA	NA	0.06	0.05	

<u>_IM(ox)N(de)TFSVM(ox)PSPK_</u>	Q13885	X	NA	NA	0.62	0.81	
<u>_LGLM(ox)PPPEPK_¶</u>	O43395	X	NA	NA	2.74	1.75	
<u>_M(ox)DTDLETM(ox)DLDQGGEALAPR_¶</u>	O75643	X	NA	NA	2.03	1.11	
<u>_VGN(de)GEETPM(ox)IGDK_</u>	Q13451	X	NA	NA	0.12	0.18	
<u>_AKQDMACLR_</u>	CON_Q14533	NA	X	NA	0.14	NA	
<u>_AM(ox)KPPGGESSNLFSGPEEATPSSRPNR_</u>	B4DLH4	NA	X	NA	0.21	1.10	
<u>_AVMISAIEK_</u>	Q15393	NA	X	NA	2.07	1.39	
<u>_EHGLIFM(ox)ETSAK_</u>	P61019	NA	X	NA	4.95	1.08	
<u>_EPAAEIEALLGM(ox)DLVR_¶</u>	Q12765-2	NA	X	NA	2.14	NA	
<u>_IM(ox)QDFESDTFFPEIDLEK_</u>	P00374	NA	X	NA	0.59	1.23	
<u>_LGDLYEEEMR_</u>	P08670	NA	X	NA	0.24	0.05	
<u>_M(ox)INLSVPDTIDER_</u>	P13797	NA	X	NA	0.22	0.60	
<u>_M(ox)SVIEEGDCKR_</u>	P11216	NA	X	NA	0.23	0.36	
<u>_MINLSVPDTIDER_</u>	P13797	NA	X	NA	0.22	0.60	
<u>_NIVLSGGSTM(ox)FR_¶</u>	P61158	NA	X	NA	0.79	0.66	
<u>_SLAKPETDKEQDSEM(ox)EK_</u>	P49321-3	NA	X	NA	1.06	0.78	
<u>_SM(ox)STEGLM(ox)K_¶</u>	O75390	NA	X	NA	2.22	1.36	
<u>_SVEM(ox)HHEALSEALPGDNVGFNVK_¶</u>	P68104	NA	X	NA	1.08	1.10	
<u>_GAPGVSVDM(ox)NITYM(ox)SPAK_</u>	Q9NPJ3	NA	NA	X	5.76	0.95	
<u>_GILAADESVGSM(ox)AK_</u>	P09972	NA	NA	X	1.59	2.25	
<u>_GVGISVLEM(ox)SHR_</u>	Q9Y617	NA	NA	X	0.23	1.14	
<u>_IIVVM(ox)DAYGSELVER_</u>	P33121	NA	NA	X	0.29	0.74	
<u>_LEGM(ox)LSQSVSSQYNM(ox)AGVR_</u>	P54105	NA	NA	X	2.20	0.87	

<u>_LERENAEVDGDDDAEEM(ox)EAKAED_</u>	P05198	NA	NA	X	0.93	1.10	
<u>_MADCGGLPQISQPAK_</u>	Q92597	NA	NA	X	0.20	0.55	
<u>_MRPGVACSVSQAQK_</u>	P32969	NA	NA	X	0.43	0.80	
<u>_TM(ox)SEVGGSVEDLIAK_¶</u>	P07339	NA	NA	X	5.47	2.34	
<u>_TQEQLALEM(ox)AELTAR_</u>	P26038	NA	NA	X	0.38	0.49	
<u>_VDALM(ox)DEINFMK_</u>	CON_P13647	NA	NA	X	0.17	NA	
<u>_VDVEFDYDGPLM(ox)K_</u>	H3BNQ7	NA	NA	X	13.75	NA	
<u>_WEQGQADYM(ox)GADSFENIK_</u>	P46976	NA	NA	X	4.94	2.60	

* X: Hit, NA=Not Assayed, TR= shows similar behavior but has missing data points or only one denaturants point with H/L higher than cutoff value, ** Averaged median H/L from triplicate measurements.

¶ only hits in MCF-10A versus MCF-7 cell line comparison.

References:

1. Tala, et al. (2014) Microtubule-associated protein Mdp3 promotes breast cancer growth and metastasis. *Theranostics* 4(10):1052-1061.
2. Bash-Babula J, et al. (2002) A phase I/pilot study of sequential doxorubicin/vinorelbine: Effects on p53 and microtubule-associated protein 4. *Clin Cancer Res* 8(5):1057-1064.
3. Nicholls C, et al. (2012) Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) induces cancer cell senescence by interacting with telomerase RNA component. *P Natl Acad Sci USA* 109(33):13308-13313.
4. Krasnov GS, Dmitriev AA, Snezhkina AV, & Kudryavtseva AV (2013) Dereglulation of glycolysis in cancer: glyceraldehyde-3-phosphate dehydrogenase as a therapeutic target. *Expert Opin Ther Tar* 17(6):681-693.
5. Shiba E, et al. (1996) Ca²⁺-Dependent Neutral Protease (Calpain) Activity in Breast Cancer Tissue and Estrogen Receptor Status. *Breast Cancer* 3(1):13-17.
6. Wang XD, Rosales JL, Magliocco A, Gnanakumar R, & Lee KY (2003) Cyclin E in breast tumors is cleaved into its low molecular weight forms by calpain. *Oncogene* 22(5):769-774.

-
7. Woelfle U, Sauter G, Santjer S, Brakenhoff R, & Pantel K (2004) Down-regulated expression of cytokeratin 18 promotes progression of human breast cancer. *Clin Cancer Res* 10(8):2670-2674.
 8. Olofsson MH, et al. (2007) Cytokeratin-18 is a useful serum biomarker for early determination of response of breast carcinomas to chemotherapy. *Clin Cancer Res* 13(11):3198-3206.
 9. Cimpean AM, et al. (2008) Relevance of the immunohistochemical expression of cytokeratin 8/18 for the diagnosis and classification of breast cancer. *Rom J Morphol Embryo* 49(4):479-483.
 10. Tu SH, et al. (2010) Increased expression of enolase alpha in human breast cancer confers tamoxifen resistance in human breast cancer cells. *Breast Cancer Res Tr* 121(3):539-553.
 11. Capello M, Ferri-Borgogno S, Cappello P, & Novelli F (2011) alpha-Enolase: a promising therapeutic and diagnostic tumor target. *The FEBS journal* 278(7):1064-1074.
 12. Brown NJ, et al. (2013) Lactate Dehydrogenase-B Is Silenced by Promoter Methylation in a High Frequency of Human Breast Cancers. *Plos One* 8(2).
 13. Dennison JB, et al. (2013) Lactate Dehydrogenase B: A Metabolic Marker of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Clin Cancer Res* 19(13):3703-3713.
 14. Jardim BV, et al. (2013) Glutathione and glutathione peroxidase expression in breast cancer: an immunohistochemical and molecular study. *Oncology reports* 30(3):1119-1128.
 15. Hu YJ & Diamond AM (2003) Role of glutathione peroxidase 1 in breast cancer: loss of heterozygosity and allelic differences in the response to selenium. *Cancer research* 63(12):3347-3351.
 16. Johnson MD, Torri JA, Lippman ME, & Dickson RB (1993) The role of cathepsin D in the invasiveness of human breast cancer cells. *Cancer research* 53(4):873-877.
 17. Glondu M, et al. (2001) A mutated cathepsin-D devoid of its catalytic activity stimulates the growth of cancer cells. *Oncogene* 20(47):6920-6929.
 18. Deng S, Jing B, Xing T, Hou L, & Yang Z (2012) Overexpression of annexin A2 is associated with abnormal ubiquitination in breast cancer. *Genomics, proteomics & bioinformatics* 10(3):153-157.

-
19. Cook KL, et al. (2012) *Glucose-regulated protein 78 controls cross-talk between apoptosis and autophagy to determine antiestrogen responsiveness. Cancer Res* 72(13):3337-3349.
 20. Fernandez PM, et al. (2000) *Overexpression of the glucose-regulated stress gene GRP78 in malignant but not benign human breast lesions. Breast Cancer Res Tr* 59(1):15-26.
 21. Nirde P, et al. (2010) *Heat shock cognate 70 protein secretion as a new growth arrest signal for cancer cells. Oncogene* 29(1):117-127.
 22. Derycke L, et al. (2011) *The role of non-muscle myosin IIA in aggregation and invasion of human MCF-7 breast cancer cells. Int J Dev Biol* 55(7-9):835-840.
 23. Dulyaninova NG, House RP, Betapudi V, & Bresnick AR (2007) *Myosin-IIA heavy-chain phosphorylation regulates the motility of MDA-MB-231 carcinoma cells. Mol Biol Cell* 18(8):3144-3155.
 24. Nagaraja GM, et al. (2006) *Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. Oncogene* 25(16):2328-2338.

B.2 Peptides and Proteins with Altered Thermodynamics in MCF-7 vs MDA-MB-231

Peptide(s)	Protein(s)	Bio. Replicates			Median H/L**	Refs
		1*	2*	3*		
<u>DM(ox)TLPPETNVILTK</u>	E7EVA0	X	NA	NA	0.51	1,2
<u>KPM(ox)SLASGLVPAAPPK</u>	E7EVA0	X	NA	X	0.51	
<u>SMVAVMDSDTTGK</u>	P04632	X	NA	X	0.56	3,4
<u>YGIVDYM(ox)IEQSGPPSK</u>	P13667	X	NA	X	1.64	
<u>FAMEPEEFDSDTLR</u>	P13667	NA	NA	X	1.64	
<u>(ac)MEQVNELK</u>	P31948	X	NA	NA	1.36	
<u>GDYPQAM(ox)K</u>	P31948	NA	X	NA	1.36	
<u>(ac)MDKNELVQK</u>	P63104	X	NA	NA	1.53	5,6
<u>YDDMAACM(ox)K</u>	P63104	X	NA	NA	1.53	
<u>DSTLIMQLLR</u>	P63104	NA	NA	X	1.53	
<u>EM(ox)GVDVHFPK</u>	Q09666	X	NA	TR	0.52	7
<u>ISM(ox)QDVDSLGSPPK</u>	Q09666	X	NA	TR	0.52	
<u>M(ox)EGDLKGPEVDIK</u>	Q09666	NA	NA	X	0.52	
<u>M(ox)ADCGLPQISQPAK</u>	Q92597	TR	NA	X	0.15	
<u>HEAM(ox)ITDLEER</u>	P35579	NA	NA	X	0.20	
<u>NMDPLNDNIATLLHQSSDK</u>	P35579	NA	NA	X	0.20	
<u>(ac)TEQMTLR</u>	P63244	X	NA	NA	0.94	
<u>ATCIGNNSAAAVSM(ox)LK</u>	P25789	X	NA	NA	1.36	

<u>_DNAM(ox)LEYLK_</u>	P15311	X	NA	NA	0.49	
<u>_EAFNM(ox)IDQNR_</u>	J3QRS3	X	NA	NA	0.26	
<u>_LAEKEETGM(ox)AM(ox)R_</u>	P21980	X	NA	NA	0.09	
<u>_LESGMQNM(ox)SIHTK_</u>	P05787-2	X	NA	NA	5.70	
<u>_M(ox)DENQFVAVTSTN(de)AAK_¶</u>	Q16555	X	NA	NA	0.22	
<u>_M(ox)DENQFVAVTSTNAAK_¶</u>	Q16555	X	NA	NA	0.22	
<u>_M(ox)GLALSSLNK_</u>	O43765	X	NA	NA	0.68	
<u>_VAKPVVEM(ox)DGDEM(ox)TR_</u>	P48735	X	NA	NA	9.14	
<u>_YPMVAVGLNK_</u>	Q9Y3U8	X	NA	NA	0.95	
<u>_LIFQM(ox)PQNK_</u>	P46940	X	NA	NA	0.66	
<u>_M(ox)AEGEAEGSVK_</u>	Q96CV9	NA	X	NA	0.17	
<u>_TAM(ox)AAAK_</u>	P83731	NA	X	NA	0.90	
<u>(ac)AAGVEAAAEEVAATEIKM(ox)EEESGAPGVPSGN(de)GAP GPK_¶</u>	P52272	NA	NA	X	0.82	
<u>_(ac)AAPM(ox)TPAARPEDYEALNAALADVPELAR_¶</u>	Q04446	NA	NA	X	0.24	
<u>_(ac)ADDLDFETGDAGASATFPMQCSALR_</u>	I3L504	NA	NA	X	0.89	
<u>_DPNNLFMVR_¶</u>	Q13200	NA	NA	X	0.76	
<u>_IMLNTPEDVQALVSGK_¶</u>	O00231-2	NA	NA	X	0.89	
<u>_LAGPQLVQMFIGDGAK_</u>	P17980	NA	NA	X	0.95	
<u>_M(ox)RPGVACSVSQAQK_</u>	P32969	NA	NA	X	0.86	
<u>_TILMMGR_</u>	P13639	NA	NA	X	1.11	
<u>_VADMALHYANK_¶</u>	P50990	NA	NA	X	0.72	
<u>_VALSNMNVDR_</u>	P55786	NA	NA	X	1.13	

VKNPEDLSAETMAK	Q9Y570-4	NA	NA	X	0.27	
VWDAVSGDELM(ox)TLAHK	B4DNJ6	NA	NA	X	0.72	
ADEDPIM(ox)GFHQM(ox)FLLK¶	P61970	NA	NA	X	1.95	
KPGMFFNPEESELDLTYGNR	P11413-2	NA	NA	X	5.29	
VLGMTLIQK	Q04760	NA	NA	X	0.81	
M(ox)ILELFSK	P78417	NA	NA	X	0.26	
ESM(ox)ATGSIPITVR¶	P49736	X	NA	NA	1.28	
M(ox)LGQM(ox)TDQVADLR	P18206	X	NA	TR	0.24	
VM(ox)YTPM(ox)APGNYLISVK¶	O75369-8	X	NA	X	0.30	
DGTAVTYVPLTAGMYTLTM(ox)K	O75369-8	NA	NA	X	0.30	
LITSVADVNNNDPM(ox)VGSK	P06737	X	NA	NA	0.45	
DIIN(de)MLFYHDRFK	P06737	NA	NA	X	0.45	
TAM(ox)STPHVAEPAENEQDEQDEN(de)GAEASADLR	P26038	X	NA	NA	0.24	
AM(ox)LENEK	P26038	NA	X	NA	0.24	
AQM(ox)VQEDLEK	P26038	NA	X	NA	0.24	
LM(ox)SNLDSNR	P26447	X	NA	NA	0.04	
ALDVM(ox)VSTFHK	P26447	NA	NA	X	0.04	
TDM(ox)DQIITSK	Q96TA1	NA	X	NA	0.35	
VQQVQPAMQAVIR	Q96TA1	NA	NA	X	0.35	
IHPVSTM(ox)VK	P07195	X	NA	NA	0.07	
MIVESAYEVIK	P07195	X	NA	NA	0.05	
SEIDMNDIK	P04083	TR	NA	NA	0.09	
(ac)M(ox)DEM(ox)ATTQISKDELDELK	P13797	X	NA	NA	0.21	

<u>_NYPNTFWM(ox)NPQYLIK_</u>	P17655	NA	NA	X	0.23	
<u>_VAKPVVEMDGDEM(ox)TR_</u>	P48735	NA	NA	X	9.14	
<u>_IM(ox)NTFSVMPSPK_</u>	Q13885	NA	NA	X	0.41	
<u>_FVPQEM(ox)GPHTVAVK_</u>	Q14315	NA	NA	X	0.16	
<u>_(ac)AAVTM(ox)SVPGR_</u>	G8JLL9	X	NA	NA	1.21	
<u>_(ac)M(ox)ELFLAGR_</u>	Q7Z4W1	X	NA	NA	4.07	
<u>_(ac)MEPSSLELPADTVQR_</u>	Q16719	X	NA	NA	8.31	
<u>_ALAAQNIVEDM(ox)EQR_</u>	Q16222	X	NA	NA	2.04	
<u>_FM(ox)ATN(de)DLM(ox)TELQK_</u>	Q86VP6	X	NA	NA	1.90	
<u>_GYM(ox)KIN(de)M(ox)LVIELK_</u>	Q14204	X	NA	NA	0.41	
<u>_IGAFGYM(ox)ECSAK_</u>	P61586	X	NA	NA	0.62	
<u>_ILM(ox)VGLDAAGK_</u>	P84085	X	NA	NA	3.70	
<u>_ISVM(ox)GGEQAANVLATITK_</u>	Q9HCC0	X	NA	NA	0.76	
<u>_M(ox)M(ox)GLEVLGEK_</u>	Q08J23	X	NA	NA	1.41	
<u>_M(ox)SAEINEIIR_</u>	P18206	X	NA	NA	0.24	
<u>_NVEGQDM(ox)LYQSLK_</u>	Q10567	X	NA	NA	2.35	
<u>_VEADVSLPSM(ox)QGDLK_</u>	Q8IVF2	X	NA	NA	0.12	
<u>_VSM(ox)APDGN(de)GGLYR_</u>	Q16222	X	NA	NA	2.04	
<u>_VVPEMTEILK_</u>	P13804	X	NA	NA	3.38	
<u>_VVSGM(ox)VNCNDDQGVLLGR_</u>	P21980	X	NA	NA	0.09	
<u>_(ac)M(ox)EQLSSANTR_</u>	P30740	NA	X	NA	0.53	
<u>_(ac)M(ox)FSSSAK_</u>	P62081	NA	X	NA	1.71	
<u>_(ac)M(ox)TM(ox)DKSELVQK_</u>	P31946	NA	X	NA	1.15	

<u>_AQTM(ox)SNSGIK_</u>	Q12904-2	NA	X	NA	5.01	
<u>_GM(ox)GYM(ox)PK_</u>	P31146	NA	X	NA	2.26	
<u>_M(ox)FNAEN(de)GK_</u>	P37837	NA	X	NA	1.06	
<u>_M(ox)GAN(de)SLER_</u>	P52272	NA	X	NA	0.82	
<u>_M(ox)M(ox)VAGFK_¶</u>	Q00839	NA	X	NA	1.61	
<u>_N(de)DLM(ox)EYAK_</u>	P00966	NA	X	NA	7.35	
<u>_NGLM(ox)FVK_</u>	Q5TBB1	NA	X	NA	0.45	
<u>_NM(ox)QDM(ox)VEDYR_</u>	P04264	NA	X	NA	0.14	
<u>_QEM(ox)QEVQSSR_¶</u>	P22626	NA	X	NA	2.18	
<u>_TDM(ox)AGPK_</u>	P16152	NA	X	NA	0.95	
<u>_VITIM(ox)QNPR_¶</u>	P62269	NA	X	NA	15.10	
<u>_(ac)AAPMTPAARPEDYEAALNAALADVPELAR_</u>	Q04446	NA	NA	X	0.24	
<u>_(ac)M(ox)VNPTVFFDIAVDGEPLGR_¶</u>	P62937	NA	NA	X	1.03	
<u>_(ac)MDVLVSECSAR_</u>	P54136	NA	NA	X	1.02	
<u>_AM(ox)KPPGGESSNLFGSPEEATPSSRPNR_</u>	B4DLH4	NA	NA	X	1.02	
<u>_AVAEAVETGEEDVIM(ox)EALR_</u>	Q9NPQ8-3	NA	NA	X	0.87	
<u>_AVQQVNAM(ox)IEK_¶</u>	P49366	NA	NA	X	0.91	
<u>_EMVQNLM(ox)VLR_¶</u>	P11413-2	NA	NA	X	5.29	
<u>_ETASELLM(ox)R_</u>	Q01082	NA	NA	X	0.19	
<u>_FGDPVVQSDMK_</u>	P08107	NA	NA	X	3.07	
<u>_FSLTSLPQQQLPSQQLM(ox)SK_¶</u>	Q6PKG0	NA	NA	X	0.67	
<u>_GQM(ox)QKPFEDASFALR_¶</u>	Q13526	NA	NA	X	7.21	
<u>_GWEEGVAQMSVGQR_¶</u>	P62942	NA	NA	X	0.35	

<u>_IIMPPSALDQLSR_‡</u>	Q92890-1	NA	NA	X	0.38	
<u>_IMGIPEEEQMGLLR_</u>	P35579	NA	NA	X	0.20	
<u>_KAEIGIAM(ox)GSGTAVAK_</u>	P16615	NA	NA	X	0.36	
<u>_LELFLPEEYPMAAPK_‡</u>	P61088	NA	NA	X	0.90	
<u>_LESGMQNMSIHTK_</u>	P05787-2	NA	NA	X	5.70	
<u>_LGDLYEEEMR_</u>	P08670	NA	NA	X	0.22	
<u>_LM(ox)TDTINEPILLCR_‡</u>	O43776	NA	NA	X	0.69	
<u>_LMVPLLK_</u>	O00410-3	NA	NA	X	0.32	
<u>_LQGQKEPGDQGPAHPPGADM(ox)SHSL_</u>	Q9Y399	NA	NA	X	12.13	
<u>_M(ox)ASTFIGN(de)STAIQELFK_</u>	Q9BUF5	NA	NA	X	0.61	
<u>_M(ox)ASTFIGNSTAIQELFK_</u>	Q9BUF5	NA	NA	X	0.61	
<u>_M(ox)ILIQDGSQNTNVDKPLR_</u>	Q92945	NA	NA	X	0.75	
<u>_M(ox)VVEHPEFLK_‡</u>	P06396	NA	NA	X	2.60	
<u>_M(ox)VWEGLNVVK_‡</u>	F6XY72	NA	NA	X	0.69	
<u>_MSSTFIGN(de)STAIQELFK_</u>	Q13509	NA	NA	X	4.34	
<u>_NVEAM(ox)NFADIER_‡</u>	P36957	NA	NA	X	2.10	
<u>_SGDAAIVEM(ox)VPGKPM(ox)CVESFSQYPPLGR_‡</u>	Q05639	NA	NA	X	60.08	
<u>_VIQGDGVDINTLQEIVEGM(ox)K_</u>	P43490	NA	NA	X	0.29	
<u>_VKYETELAMR_</u>	P05783	NA	NA	X	5.57	

* X: Hit, NA=Not Assayed, TR= shows similar behavior but has missing data points or only one denaturants point with H/L higher than cutoff value, ** Averaged median H/L from triplicate measurements.

‡ only hits in MCF-7 versus MDA-MB-231 cell line comparison.

References:

-
1. Tala, et al. (2014) *Microtubule-associated protein Mdp3 promotes breast cancer growth and metastasis. Theranostics* 4(10):1052-1061.
 2. Bash-Babula J, et al. (2002) *A phase I/pilot study of sequential doxorubicin/vinorelbine: Effects on p53 and microtubule-associated protein 4. Clin Cancer Res* 8(5):1057-1064.
 3. Shiba E, et al. (1996) *Ca²⁺-Dependent Neutral Protease (Calpain) Activity in Breast Cancer Tissue and Estrogen Receptor Status. Breast Cancer* 3(1):13-17.
 4. Wang XD, Rosales JL, Magliocco A, Gnanakumar R, & Lee KY (2003) *Cyclin E in breast tumors is cleaved into its low molecular weight forms by calpain. Oncogene* 22(5):769-774.
 5. Neal CL, et al. (2009) *14-3-3 zeta Overexpression Defines High Risk for Breast Cancer Recurrence and Promotes Cancer Cell Survival. Cancer Res* 69(8):3425-3432.
 6. Hodgkinson VC, et al. (2012) *Proteomic identification of predictive biomarkers of resistance to neoadjuvant chemotherapy in luminal breast cancer: A possible role for 14-3-3 theta/tau and tBID? J Proteomics* 75(4):1276-1283.
 7. Kabir MH, Suh EJ, & Lee C (2012) *Comparative phosphoproteome analysis reveals more ERK activation in MDA-MB-231 than in MCF-7. Int J Mass Spectrom* 309:1-12.

B.3 Bioinformatics Analyses of Peptides and Proteins with Altered Thermodynamics

Table B.3.1: Summary of molecular functions represented in the protein hits identified in the thermodynamic and protein expression level analyses performed here. Data in parentheses is from expression level analyses.

Molecular Functions	MCF-7 Vs MCF-10A Comparison	MCF-7 Vs MDA-MB-231 Comparison
Catalytic activity	34.0 (40.4)%	37.8 (40.6)%
Binding	26.6 (29.0)%	30.6 (29.2)%
Structural molecule activity	20.2 (13.1)%	21.6 (11.5)%
Enzyme regulator	5.3 (6.0)%	4.5 (5.4)%
Translational regulator	5.3 (2.1)%	2.7 (2.5)%
Transporter activity	4.3 (2.4)%	1.8 (1.6)%
Nucleic acid binding transcription factor	2.1 (3.6)%	0.9 (4.3)%
Antioxidant activity	1.1 (0.6)%	0.0 (0.0)%
Protein binding transcription factor	1.1 (0.7)%	0.0 (0.8)%
Receptor activity	0.0 (2.1)%	0.0 (1.1)%

Table B.3.2: Summary of biological processes represented in the protein hits identified in the thermodynamic and protein expression level analyses performed here. Data in parentheses is from expression level analyses.

	MCF-7 Vs MCF-10A Comparison	MCF-7 Vs MDA-MB-231 Comparison
Biological Processes		
Metabolic process	32.2 (33.8)%	31.1 (33.7)%
Cellular process	16.8 (19.2)%	21.1 (19.9)%
Cellular component organization or biogenesis	12.1 (6.8)%	11.6 (7.6)%
Developmental process	11.4 (9.5)%	11.1 (8.7)%
Biological regulation	8.1 (7.0)%	5.3 (7.2)%
Localization	6.0 (7.8)%	7.9 (8.3)%
Response to stimulus	4.7 (3.9)%	2.6 (3.9)%
Immune system	4.7 (4.0)%	3.7 (4.5)%
Multicellular organismal process	2.0 (3.5)%	3.2 (3.2)%
Apoptotic process	1.3 (2.0)%	2.1 (1.6)%
Biological adhesion	0.7 (1.8)%	0.5 (0.9)%
Reproduction	0.0 (0.8)%	0.0 (0.6)%

Table B.3.3: Summary of molecular functions represented by the assayed proteins used in the thermodynamic and protein expression level analyses performed here. Data in parentheses is from expression level analyses.

	MCF-7 Vs MCF-10A	MCF-7 Vs MDA-MB-231
Molecular Functions	Comparison	Comparison
Catalytic activity	38.6 (39.1)%	39.0 (39.1)%
Binding Activity	31.3 (31.3)%	31.6 (31.8)%
Structural molecule activity	13.4 (12.0)%	13.6 (11.5)%
Enzyme regulator	5.3 (5.8)%	4.9 (5.9)%
Translational regulator	4.1 (3.1)%	3.6 (3.0)%
Nucleic acid binding transcription factor	2.7 (3.3)%	2.9 (3.8)%
Transporter activity	2.6 (2.7)%	2.5 (2.7)%
Receptor activity	0.8 (1.6)%	1.2 (1.0)%
Protein binding transcription factor	0.8 (0.7)%	0.6 (0.8)%
Antioxidant activity	0.4 (0.5)%	0.2 (0.4)%

Table B.3.4: Summary of biological processes represented by the proteins assayed used in the thermodynamic and protein expression level analyses performed here. Data in parentheses is from expression level analyses.

	MCF-7 Vs MCF-10A Comparison	MCF-7 Vs MDA-MB-231 Comparison
Biological Processes		
Metabolic process	39.5 (38.4)%	38.9 (37.7)%
Cellular process	16.2 (17.3)%	17.3 (17.5)%
Biological regulation	8.7 (8.4)%	7.9 (8.6)%
Localization	8.6 (8.8)%	8.7 (8.9)%
Cellular component organization or biogenesis	7.4 (6.9)%	7.7 (6.7)%
Developmental process	7.3 (7.2)%	7.6 (7.6)%
Response to stimulus	3.6 (3.4)%	3.0 (3.2)%
Immune system	3.0 (3.0)%	2.8 (2.9)%
Apoptotic process	2.2 (1.9)%	2.1 (2.0)%
Multicellular organismal process	2.2 (3.0)%	2.5 (3.0)%
Reproduction	1.0 (0.8)%	1.2 (1.0)%
Biological adhesion	0.3 (0.9)%	0.3 (1.0)%

Table B.3.5: Summary of protein classes represented by the assayed proteins used in the thermodynamic and protein expression level analyses performed here. Data in parentheses is from expression level analyses.

Protein Classes	MCF-7 Vs MCF-10A Comparison	MCF-7 Vs MDA-MB-231 Comparison
Nucleic acid binding	19.4 (18.6)%	18.6 (18.1)%
Hydrolase	10.4 (10.7)%	11.1 (10.6)%
Cytoskeletal protein	8.1 (7.5)%	8.8 (7.8)%
Transferase	8.4 (7.9)%	8.1 (8.2)%
Enzyme modulator	8.3 (8.2)%	8.1 (8.5)%
Ligase	4.9 (4.1)%	5.0 (4.3)%
Chaperone	4.7 (3.8)%	5.0 (3.5)%
Oxidoreductase	6.1 (6.0)%	4.9 (5.9)%
Calcium-binding protein	2.7 (2.2)%	3.3 (2.5)%
Protease	3.8 (2.8)%	3.1 (2.9)%
Isomerase	3.2 (2.5)%	2.9 (2.6)%
Transcription factor	2.9 (3.3)%	2.9 (3.9)%
Transporter	2.4 (2.5)%	2.4 (2.6)%
Membrane traffic protein	2.2 (3.0)%	2.4 (2.9)%
Kinase	2.1 (2.5)%	2.4 (2.8)%
Transfer/carrier protein	2.2 (2.1)%	2.1 (1.9)%
Phosphatase	1.6 (1.8)%	1.9 (1.8)%
Signaling molecule	2.2 (2.1)%	1.8 (2.3)%
<i>Lyase</i>	1.6 (1.6)%	1.3 (1.6)%
Receptor	0.8 (1.5)%	1.2 (1.0)%
Structural protein	1.0 (1.3)%	1.1 (1.2)%
Cell adhesion molecule	0.5 (0.7)%	0.4 (0.6)%
Defense/immunity protein	0.2 (0.5)%	0.4 (0.5)%
Cell junction protein	0.1 (0.7)%	0.4 (0.8)%
Transmembrane receptor	0.2 (0.6)%	0.2 (0.7)%
regulatory/adaptor protein		
Storage protein	0.1 (0.1)%	0.0 (0.0)%
Surfactant	0.0 (0.1)%	0.0 (0.1)%
Extracellular matrix protein	0.0 (0.6)%	0.0 (0.3)%

B.4 Peptides and Proteins with Altered Thermodynamics in MCF-7 versus MCF-10A using SILAC-PP.

Protein IDs	Protein names	Gel Bands
P61158	Actin-related protein 3	36-50kDa
P48643	T-complex protein 1 subunit epsilon	36-50kDa
CON__P13647	Keratin, type II cytoskeletal 5	36-50kDa 50-75kDa
P07237	Protein disulfide-isomerase	36-50kDa
P30101	Protein disulfide-isomerase A3	36-50kDa 50-75kDa
P04075-2	Fructose-bisphosphate aldolase A	36-50kDa
P00558	Phosphoglycerate kinase 1	36-50kDa
P04406	Glyceraldehyde-3-phosphate dehydrogenase	36-50kDa 50-75kDa
P10809	60 kDa heat shock protein, mitochondrial	36-50kDa
P13489	Ribonuclease inhibitor	36-50kDa
P13667	Protein disulfide-isomerase A4	36-50kDa 50-75kDa
P15311	Ezrin	36-50kDa 50-75kDa
P23246	Splicing factor, proline- and glutamine-rich	36-50kDa 50-75kDa
P35241-5	Radixin	36-50kDa
P35998	26S protease regulatory subunit 7	36-50kDa
P36952	Serpin B5	36-50kDa
P39023	60S ribosomal protein L3	36-50kDa
P43686	26S protease regulatory subunit 6B	36-50kDa
P49189	4-trimethylaminobutyraldehyde dehydrogenase	36-50kDa
P49321-3	Nuclear autoantigenic sperm protein	36-50kDa
P53396	ATP-citrate synthase	36-50kDa
P61978-2	Heterogeneous nuclear ribonucleoprotein K	36-50kDa

		50-75kDa
P68104	Elongation factor 1-alpha 1	36-50kDa
Q12765-2	Secernin-1	36-50kDa
Q14847	LIM and SH3 domain protein 1	36-50kDa
Q14974	Importin subunit beta-1	36-50kDa
P20073	Annexin A7	36-50kDa
P27361	Mitogen-activated protein kinase 3	36-50kDa
H3BNQ7	4-aminobutyrate aminotransferase, mitochondrial	36-50kDa
P07355-2	Annexin A2	36-50kDa
P14866	Heterogeneous nuclear ribonucleoprotein L	36-50kDa
P30740	Leukocyte elastase inhibitor	36-50kDa
P58107	Epiplakin	36-50kDa
P08559-4	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	36-50kDa
Q16881	Thioredoxin reductase 1, cytoplasmic	50-75kDa
CON_P60712	Actin, cytoplasmic 1	50-75kDa
Q13200	26S proteasome non-ATPase regulatory subunit 2	50-75kDa
Q16658	Fascin	50-75kDa
O75083	WD repeat-containing protein 1	50-75kDa
O95831	Apoptosis-inducing factor 1, mitochondrial	50-75kDa
P07384	Calpain-1 catalytic subunit	50-75kDa
P07437	Tubulin beta chain	50-75kDa
P08107	Heat shock 70 kDa protein 1A/1B	50-75kDa
P11021	78 kDa glucose-regulated protein	50-75kDa
P11142	Heat shock cognate 71 kDa protein	50-75kDa
P11413-2	Glucose-6-phosphate 1-dehydrogenase	50-75kDa
P12956	X-ray repair cross-complementing protein 6	50-75kDa
P13010	X-ray repair cross-complementing protein 5	50-75kDa
P14625	Endoplasmin	50-75kDa
P17655	Calpain-2 catalytic subunit	50-75kDa
P26038	Moesin	50-75kDa
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2- oxoglutarate dehydrogenase complex,	50-75kDa

	mitochondrial	
P38646	Stress-70 protein, mitochondrial	50-75kDa
P50990	T-complex protein 1 subunit theta	50-75kDa
P54727	UV excision repair protein RAD23 homolog	50-75kDa
	B	
P68371	Tubulin beta-4B chain	50-75kDa
Q01082	Spectrin beta chain, brain 1	50-75kDa
Q01518	Adenylyl cyclase-associated protein 1	50-75kDa
Q09666	Neuroblast differentiation-associated protein AHNAK	50-75kDa
Q14166	Tubulin--tyrosine ligase-like protein 12	50-75kDa
Q99832	T-complex protein 1 subunit eta	50-75kDa
Q9UHD8	Septin-9	50-75kDa
P08670	Vimentin	50-75kDa
P09622	Dihydrolipoyl dehydrogenase, mitochondrial	50-75kDa
Q15084-2	Protein disulfide-isomerase A6	50-75kDa
Q14152	Eukaryotic translation initiation factor 3 subunit A	50-75kDa
P04264	Keratin, type II cytoskeletal 1	50-75kDa

References

1. International agency for Research on Cancer and Cancer research UK. World Cancer Factsheet. *Cancer Research UK, London*. 2014.
2. American Cancer Society. Cancer Facts & Figures 2014. American Cancer Society, Atlanta. 2014.
3. Miranda, J. J., Kinra, S., Casas, J. P., Davey Smith, G., and Ebrahim, S. (2008) Non-communicable diseases in low- and middle-income countries: context, determinants and health policy, *Tropical medicine & international health : TM & IH* 13, 1225-1234.
4. Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., Radich, J., Anderson, G., and Hartwell, L. (2003) The case for early detection, *Nature reviews. Cancer* 3, 243-252.
5. Reiman, E. M., Quiroz, Y. T., Fleisher, A. S., Chen, K., Velez-Pardo, C., Jimenez-Del-Rio, M., Fagan, A. M., Shah, A. R., Alvarez, S., Arbelaez, A., Giraldo, M., Acosta-Baena, N., Sperling, R. A., Dickerson, B., Stern, C. E., Tirado, V., Munoz, C., Reiman, R. A., Huentelman, M. J., Alexander, G. E., Langbaum, J. B., Kosik, K. S., Tariot, P. N., and Lopera, F. (2012) Brain imaging and fluid biomarker analysis in young adults at genetic risk for autosomal dominant Alzheimer's disease in the presenilin 1 E280A kindred: a case-control study, *The Lancet. Neurology* 11, 1048-1056.
6. American Cancer Society. Breast Cancer Facts & Figures 2013-2014. *American Cancer Society*, Atlanta. 2013.
7. American Cancer Society. Cancer Treatment and Survivorship Facts & Figures 2014-2015. American Cancer Society, Atlanta. 2014.
8. Srinivas, P. R., Kramer, B. S., and Srivastava, S. (2001) Trends in biomarker research for cancer detection, *The lancet oncology* 2, 698-704.
9. Papassotiropoulos, A., Fountoulakis, M., Dunckley, T., Stephan, D. A., and Reiman, E. M. (2006) Genetics, transcriptomics, and proteomics of Alzheimer's disease, *The Journal of clinical psychiatry* 67, 652-670.

10. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y. D., Su, Y. A., and Trent, J. M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer, *Nat Genet* 14, 457-460.
11. Paweletz, C. P., Charboneau, L., Bichsel, V. E., Simone, N. L., Chen, T., Gillespie, J. W., Emmert-Buck, M. R., Roth, M. J., Petricoin, E. F., and Liotta, L. A. (2001) Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front, *Oncogene* 20, 1981-1989.
12. Geiger, T., Madden, S. F., Gallagher, W. M., Cox, J., and Mann, M. (2012) Proteomic portrait of human breast cancer progression identifies novel prognostic markers, *Cancer research* 72, 2428-2439.
13. Gronborg, M., Kristiansen, T. Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O. N., Hruban, R. H., Goggins, M. G., Maitra, A., and Pandey, A. (2006) Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach, *Molecular & cellular proteomics : MCP* 5, 157-171.
14. Poste, G. (2011) Bring on the biomarkers, *Nature* 469, 156-157.
15. Ankerst, D. P., and Thompson, I. M. (2006) Sensitivity and specificity of prostate-specific antigen for prostate cancer detection with high rates of biopsy verification, *Archivio italiano di urologia, andrologia : organo ufficiale [di] Societa italiana di ecografia urologica e nefrologica / Associazione ricerche in urologia* 78, 125-129.
16. Pegram, M. D., Pauletti, G., and Slamon, D. J. (1998) HER-2/neu as a predictive marker of response to breast cancer therapy, *Breast cancer research and treatment* 52, 65-77.
17. Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M., Pinter, T., Valero, V., Liu, M. C., Sauter, G., von Minckwitz, G., Visco, F., Bee, V., Buyse, M., Bendahmane, B., Tabah-Fisch, I., Lindsay, M. A., Riva, A., and Crown, J. (2011) Adjuvant trastuzumab in HER2-positive breast cancer, *The New England journal of medicine* 365, 1273-1283.

18. Ntzani, E. E., and Ioannidis, J. P. (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment, *Lancet* 362, 1439-1444.
19. Iida, K., and Nishimura, I. (2002) Gene expression profiling by DNA microarray technology, *Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists* 13, 35-50.
20. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995) Serial analysis of gene expression, *Science* 270, 484-487.
21. Weindruch, R., Kayo, T., Lee, C. K., and Prolla, T. A. (2001) Microarray profiling of gene expression in aging and its alteration by caloric restriction in mice, *The Journal of nutrition* 131, 918S-923S.
22. Ricciarelli, R., d'Abramo, C., Massone, S., Marinari, U., Pronzato, M., and Tabaton, M. (2004) Microarray analysis in Alzheimer's disease and normal aging, *IUBMB life* 56, 349-354.
23. Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. (2001) Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer, *Cancer research* 61, 5974-5978.
24. Buckhaults, P., Rago, C., St. Croix, B., Romans, K. E., Saha, S., Zhang, L., Vogelstein, B., and Kinzler, K. W. (2001) Secreted and Cell Surface Genes Expressed in Benign and Malignant Colorectal Tumors, *Cancer research* 61, 6996-7001.
25. Bertucci, F., Salas, S., Eysteris, S., Nasser, V., Finetti, P., Ginestier, C., Charafe-Jauffret, E., Lloriod, B., Bachelart, L., Montfort, J., Victorero, G., Viret, F., Ollendorff, V., Fert, V., Giovaninni, M., Delpero, J. R., Nguyen, C., Viens, P., Monges, G., Birnbaum, D., and Houlgatte, R. (2004) Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters, *Oncogene* 23, 1377-1391.
26. van't Veer, L. J., Dai, H. Y., van de Vijver, M. J., He, Y. D. D., Hart, A. A. M., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H.

- (2002) Gene expression profiling predicts clinical outcome of breast cancer, *Nature* 415, 530-536.
27. Booij, B. B., Lindahl, T., Wetterberg, P., Skaane, N. V., Saebo, S., Feten, G., Rye, P. D., Kristiansen, L. I., Hagen, N., Jensen, M., Bardsen, K., Winblad, B., Sharma, P., and Lonneborg, A. (2011) A gene expression pattern in blood for the early detection of Alzheimer's disease, *Journal of Alzheimer's disease : JAD* 23, 109-119.
 28. Michiels, S., Koscielny, S., and Hill, C. (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy, *Lancet* 365, 488-492.
 29. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast, *Molecular and cellular biology* 19, 1720-1730.
 30. Zhu, H., and Snyder, M. (2001) Protein arrays and microarrays, *Current opinion in chemical biology* 5, 40-45.
 31. Zhu, H., and Snyder, M. (2003) Protein chip technology, *Current opinion in chemical biology* 7, 55-63.
 32. Spurrier, B., Ramalingam, S., and Nishizuka, S. (2008) Reverse-phase protein lysate microarrays for cell signaling analysis, *Nature protocols* 3, 1796-1808.
 33. Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G. B., and Kornblau, S. M. (2006) Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells, *Molecular cancer therapeutics* 5, 2512-2521.
 34. Nishizuka, S., Chen, S. T., Gwadry, F. G., Alexander, J., Major, S. M., Scherf, U., Reinhold, W. C., Waltham, M., Charboneau, L., Young, L., Bussey, K. J., Kim, S., Lababidi, S., Lee, J. K., Pittaluga, S., Scudiero, D. A., Sausville, E. A., Munson, P. J., Petricoin, E. F., 3rd, Liotta, L. A., Hewitt, S. M., Raffeld, M., and Weinstein, J. N. (2003) Diagnostic markers that distinguish colon and ovarian adenocarcinomas: identification by genomic, proteomic, and tissue array profiling, *Cancer research* 63, 5243-5250.

35. MacBeath, G. (2002) Protein microarrays and proteomics, *Nat Genet* 32 Suppl, 526-532.
36. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, *Nature* 422, 198-207.
37. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proceedings of the National Academy of Sciences of the United States of America* 97, 9390-9395.
38. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Molecular & cellular proteomics : MCP* 1, 376-386.
39. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, *Analytical chemistry* 75, 1895-1904.
40. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nature biotechnology* 17, 994-999.
41. Zhou, C., Zhong, Q., Rhodes, L. V., Townley, I., Bratton, M. R., Zhang, Q., Martin, E. C., Elliott, S., Collins-Burow, B. M., Burow, M. E., and Wang, G. (2012) Proteomic analysis of acquired tamoxifen resistance in MCF-7 cells reveals expression signatures associated with enhanced migration, *Breast cancer research : BCR* 14, R45.
42. Gutman, S., and Kessler, L. G. (2006) The US Food and Drug Administration perspective on cancer biomarker development, *Nature reviews. Cancer* 6, 565-571.
43. Nicholson, J. K. (2006) Global systems biology, personalized medicine and molecular epidemiology, *Molecular systems biology* 2, 52.

44. Weckwerth, W., and Morgenthal, K. (2005) Metabolomics: from pattern recognition to biological interpretation, *Drug discovery today* 10, 1551-1558.
45. Armitage, E. G., and Barbas, C. (2014) Metabolomics in cancer biomarker discovery: current trends and future perspectives, *Journal of pharmaceutical and biomedical analysis* 87, 1-11.
46. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature* 340, 245-246.
47. Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M., and Fields, S. (2003) Protein analysis on a proteomic scale, *Nature* 422, 208-215.
48. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edlmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature* 415, 141-147.
49. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*, *Nature* 440, 637-643.
50. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proceedings of the National Academy of Sciences of the United States of America* 98, 4569-4574.

51. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*, *Nature* 403, 623-627.
52. Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y. V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J. P., Duewel, H. S., Stewart, II, Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S. L., Moran, M. F., Morin, G. B., Topaloglou, T., and Figeys, D. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry, *Molecular systems biology* 3, 89.
53. Walhout, A. J., and Vidal, M. (1999) A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens, *Genome research* 9, 1128-1134.
54. Brandts, J. F., and Lin, L. N. (1990) Study of strong to ultratight protein interactions using differential scanning calorimetry, *Biochemistry* 29, 6927-6940.
55. Hill, J. J., and Royer, C. A. (1997) Fluorescence approaches to study of protein-nucleic acid complexation, *Methods in enzymology* 278, 390-416.
56. Jessani, N., Liu, Y., Humphrey, M., and Cravatt, B. F. (2002) Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness, *Proceedings of the National Academy of Sciences of the United States of America* 99, 10335-10340.
57. Kobe, B., and Kemp, B. E. (1999) Active site-directed protein regulation, *Nature* 402, 373-376.
58. Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry, *Annual review of biochemistry* 77, 383-414.

59. Bullock, A. N., Henckel, J., DeDecker, B. S., Johnson, C. M., Nikolova, P. V., Proctor, M. R., Lane, D. P., and Fersht, A. R. (1997) Thermodynamic stability of wild-type and mutant p53 core domain, *Proceedings of the National Academy of Sciences of the United States of America* 94, 14338-14342.
60. Mayer, S., Rudiger, S., Ang, H. C., Joerger, A. C., and Fersht, A. R. (2007) Correlation of levels of folded recombinant p53 in escherichia coli with thermodynamic stability in vitro, *Journal of molecular biology* 372, 268-276.
61. Gyorgy, B., Toth, E., Tarcsa, E., Falus, A., and Buzas, E. I. (2006) Citrullination: a posttranslational modification in health and disease, *The international journal of biochemistry & cell biology* 38, 1662-1677.
62. Proctor, E. A., Ding, F., and Dokholyan, N. V. (2011) Structural and thermodynamic effects of post-translational modifications in mutant and wild type Cu, Zn superoxide dismutase, *Journal of molecular biology* 408, 555-567.
63. Strickland, E. C., Geer, M. A., Tran, D. T., Adhikari, J., West, G. M., DeArmond, P. D., Xu, Y., and Fitzgerald, M. C. (2013) Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation, *Nature protocols* 8, 148-161.
64. Dearmond, P. D., Xu, Y., Strickland, E. C., Daniels, K. G., and Fitzgerald, M. C. (2011) Thermodynamic analysis of protein-ligand interactions in complex biological mixtures using a shotgun proteomics approach, *Journal of proteome research* 10, 4948-4958.
65. Tran, D. T., Adhikari, J., and Fitzgerald, M. C. (2014) SILAC-Based Strategy for Proteome-Wide Thermodynamic Analysis of Protein-Ligand Binding Interactions, *Molecular & Cellular Proteomics*.
66. Adhikari, J., and Fitzgerald, M. C. (2014) SILAC-Pulse Proteolysis: A Mass Spectrometry-Based Method for Discovery and Cross-Validation in Proteome-Wide Studies of Ligand Binding, *Journal of the American Society for Mass Spectrometry* 25, 2073-2083.
67. Feng, Y. H., De Franceschi, G., Kahraman, A., Soste, M., Melnik, A., Boersema, P. J., de Laureto, P. P., Nikolaev, Y., Oliveira, A. P., and Picotti, P. (2014) Global

analysis of protein structural changes in complex proteomes, *Nature biotechnology* 32, 1036-+.

68. Savitski, M. M., Reinhard, F. B. M., Franken, H., Werner, T., Savitski, M. F., Eberhard, D., Molina, D. M., Jafari, R., Dovega, R. B., Klaeger, S., Kuster, B., Nordlund, P., Bantscheff, M., and Drewes, G. (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome, *Science* 346, 55-+.
69. Liu, P. F., Kihara, D., and Park, C. (2011) Energetics-based discovery of protein-ligand interactions on a proteomic scale, *Journal of molecular biology* 408, 147-162.
70. Chang, Y., Schleich, J. P., VerHeul, R. A., and Park, C. (2012) Simplified proteomics approach to discover protein-ligand interactions, *Protein Sci* 21, 1280-1287.
71. Lomenick, B., Hao, R., Jonai, N., Chin, R. M., Aghajan, M., Warburton, S., Wang, J. N., Wu, R. P., Gomez, F., Loo, J. A., Wohlschlegel, J. A., Vondriska, T. M., Pelletier, J., Herschman, H. R., Clardy, J., Clarke, C. F., and Huang, J. (2009) Target identification using drug affinity responsive target stability (DARTS), *Proceedings of the National Academy of Sciences of the United States of America* 106, 21984-21989.
72. Jafari, R., Almqvist, H., Axelsson, H., Ignatushchenko, M., Lundback, T., Nordlund, P., and Molina, D. M. (2014) The cellular thermal shift assay for evaluating drug target interactions in cells, *Nature protocols* 9, 2100-2122.
73. West, G. M., Tucker, C. L., Xu, T., Park, S. K., Han, X., Yates, J. R., 3rd, and Fitzgerald, M. C. (2010) Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements, *Proceedings of the National Academy of Sciences of the United States of America* 107, 9078-9082.
74. Straume, M., and Freire, E. (1992) Two-dimensional differential scanning calorimetry: simultaneous resolution of intrinsic protein structural energetics and ligand binding interactions by global linkage analysis, *Analytical Biochemistry* 203, 259-268.

75. Sigurskjold, B. W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry, *Analytical Biochemistry* 277, 260-266.
76. Konermann, L., Stocks, B. B., Pan, Y., and Tong, X. (2010) Mass spectrometry combined with oxidative labeling for exploring protein structure and folding, *Mass spectrometry reviews* 29, 651-667.
77. Englander, J. J., Del Mar, C., Li, W., Englander, S. W., Kim, J. S., Stranz, D. D., Hamuro, Y., and Woods, V. L., Jr. (2003) Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry, *Proceedings of the National Academy of Sciences of the United States of America* 100, 7057-7062.
78. Xu, G., and Chance, M. R. (2005) Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting, *Analytical chemistry* 77, 4549-4555.
79. Powell, K. D., and Fitzgerald, M. C. (2001) Measurements of protein stability by H/D exchange and matrix-assisted laser desorption/ionization mass spectrometry using picomoles of material, *Analytical chemistry* 73, 3300-3304.
80. Tang, L., Hopper, E. D., Tong, Y., Sadowsky, J. D., Peterson, K. J., Gellman, S. H., and Fitzgerald, M. C. (2007) H/D exchange- and mass spectrometry-based strategy for the thermodynamic analysis of protein-ligand binding, *Analytical chemistry* 79, 5869-5877.
81. Shcherbakova, I., Mitra, S., Beer, R. H., and Brenowitz, M. (2006) Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins, *Nucleic acids research* 34, e48.
82. Roulhac, P. L., Powell, K. D., Dhungana, S., Weaver, K. D., Mietzner, T. A., Crumbliss, A. L., and Fitzgerald, M. C. (2004) SUPREX (Stability of Unpurified Proteins from Rates of H/D Exchange) analysis of the thermodynamics of synergistic anion binding by ferric-binding protein (FbpA), a bacterial transferrin, *Biochemistry* 43, 15767-15774.

83. McClintock, C., Kertesz, V., and Hettich, R. L. (2008) Development of an electrochemical oxidation method for probing higher order protein structure with mass spectrometry, *Analytical chemistry* 80, 3304-3317.
84. Hamuro, Y., Coales, S. J., Southern, M. R., Nemeth-Cawley, J. F., Stranz, D. D., and Griffin, P. R. (2003) Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry, *J Biomol Tech* 14, 171-182.
85. Molina, D. M., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., Sreekumar, L., Cao, Y., and Nordlund, P. (2013) Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay, *Science* 341, 84-87.
86. Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis, *Acta biochimica Polonica* 51, 299-321.
87. Reid, G. E., Roberts, K. D., Simpson, R. J., and O'Hair, R. A. (2005) Selective identification and quantitative analysis of methionine containing peptides by charge derivatization and tandem mass spectrometry, *Journal of the American Society for Mass Spectrometry* 16, 1131-1150.
88. Gevaert, K., Van Damme, J., Goethals, M., Thomas, G. R., Hoorelbeke, B., Demol, H., Martens, L., Puype, M., Staes, A., and Vandekerckhove, J. (2002) Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins, *Molecular & cellular proteomics : MCP* 1, 896-903.
89. Brosnan, J. T., and Brosnan, M. E. (2006) The sulfur-containing amino acids: an overview, *The Journal of nutrition* 136, 1636S-1640S.
90. Strickland, E. C., Geer, M. A., Hong, J. Y., and Fitzgerald, M. C. (2014) False-Positive Rate Determination of Protein Target Discovery using a Covalent Modification- and Mass Spectrometry-Based Proteomics Platform, *Journal of the American Society for Mass Spectrometry* 25, 132-140.

91. Park, C., and Marqusee, S. (2005) Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding, *Nature methods* 2, 207-212.
92. Pace, C. N., and Barrett, A. J. (1984) Kinetics of tryptic hydrolysis of the arginine-valine bond in folded and unfolded ribonuclease T1, *The Biochemical journal* 219, 411-417.
93. Imoto, T., Yamada, H., and Ueda, T. (1986) Unfolding rates of globular proteins determined by kinetics of proteolysis, *Journal of molecular biology* 190, 647-649.
94. Chang, Y., Schleich, J. P., Verheul, R. A., and Park, C. (2012) Simplified proteomics approach to discover protein-ligand interactions, *Protein science : a publication of the Protein Society* 21, 1280-1287.
95. Everley, P. A., Krijgsveld, J., Zetter, B. R., and Gygi, S. P. (2004) Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research, *Molecular & cellular proteomics : MCP* 3, 729-735.
96. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling, *Nature biotechnology* 21, 315-318.
97. Kulasingam, V., and Diamandis, E. P. (2007) Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets, *Molecular & Cellular Proteomics* 6, 1997-2011.
98. Lai, T. C., Chou, H. C., Chen, Y. W., Lee, T. R., Chan, H. T., Shen, H. H., Lee, W. T., Lin, S. T., Lu, Y. C., Wu, C. L., and Chan, H. L. (2010) Secretomic and Proteomic Analysis of Potential Breast Cancer Markers by Two-Dimensional Differential Gel Electrophoresis, *Journal of proteome research* 9, 1302-1322.
99. Shi, Y., Xiang, R., Horvath, C., and Wilkins, J. A. (2005) Quantitative analysis of membrane proteins from breast cancer cell lines BT474 and MCF7 using multistep solid phase mass tagging and 2D LC/MS, *Journal of proteome research* 4, 1427-1433.

100. Xiang, R., Shi, Y., Dillon, D. A., Negin, B., Horvath, C., and Wilkins, J. A. (2004) 2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474, *Journal of proteome research* 3, 1278-1283.
101. Nagaraja, G. M., Othman, M., Fox, B. P., Alsaber, R., Pellegrino, C. M., Zeng, Y., Khanna, R., Tamburini, P., Swaroop, A., and Kandpal, R. P. (2006) Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics, *Oncogene* 25, 2328-2338.
102. Bateman, N. W., Sun, M., Hood, B. L., Flint, M. S., and Conrads, T. P. (2010) Defining central themes in breast cancer biology by differential proteomics: conserved regulation of cell spreading and focal adhesion kinase, *Journal of proteome research* 9, 5311-5324.
103. Aka, J. A., and Lin, S. X. (2012) Comparison of functional proteomic analyses of human breast cancer cell lines T47D and MCF7, *PloS one* 7, e31532.
104. Edwards, M. S., Chadda, S. D., Zhao, Z., Barber, B. L., and Sykes, D. P. (2012) A systematic review of treatment guidelines for metastatic colorectal cancer, *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland* 14, e31-47.
105. Harding, J., and Burtness, B. (2005) Cetuximab: an epidermal growth factor receptor chimeric human-murine monoclonal antibody, *Drugs Today (Barc)* 41, 107-127.
106. Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., Ferrara, N., Fyfe, G., Rogers, B., Ross, R., and Kabbinavar, F. (2004) Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer, *The New England journal of medicine* 350, 2335-2342.
107. Alberts, S. R., Sargent, D. J., Nair, S., Mahoney, M. R., Mooney, M., Thibodeau, S. N., Smyrk, T. C., Sinicrope, F. A., Chan, E., Gill, S., Kahlenberg, M. S., Shields, A. F., Quesenberry, J. T., Webb, T. A., Farr, G. H., Jr., Pockaj, B. A., Grothey, A., and Goldberg, R. M. (2012) Effect of oxaliplatin, fluorouracil, and leucovorin with or

without cetuximab on survival among patients with resected stage III colon cancer: a randomized trial, *Jama* 307, 1383-1393.

108. Hah, S. S., Sumbad, R. A., de Vere White, R. W., Turteltaub, K. W., and Henderson, P. T. (2007) Characterization of oxaliplatin-DNA adduct formation in DNA and differentiation of cancer cell drug sensitivity at microdose concentrations, *Chemical research in toxicology* 20, 1745-1751.
109. Lehky, T. J., Leonard, G. D., Wilson, R. H., Grem, J. L., and Floeter, M. K. (2004) Oxaliplatin-induced neurotoxicity: acute hyperexcitability and chronic neuropathy, *Muscle & nerve* 29, 387-392.
110. Martinez-Balibrea, E., Plasencia, C., Gines, A., Martinez-Cardus, A., Musulen, E., Aguilera, R., Manzano, J. L., Neamati, N., and Abad, A. (2009) A proteomic approach links decreased pyruvate kinase M2 expression to oxaliplatin resistance in patients with colorectal cancer and in human cell lines, *Molecular cancer therapeutics* 8, 771-778.
111. Lee, H., Flaherty, P., and Ji, H. P. (2013) Systematic genomic identification of colorectal cancer genes delineating advanced from early clinical stage and metastasis, *BMC medical genomics* 6, 54.
112. Tsunoda, T., Nakamura, T., Ishimoto, K., Yamaue, H., Tanimura, H., Saijo, N., and Nishio, K. (2001) Upregulated expression of angiogenesis genes and down regulation of cell cycle genes in human colorectal cancer tissue determined by cDNA macroarray, *Anticancer research* 21, 137-143.
113. Peng, Y., Li, X., Wu, M., Yang, J., Liu, M., Zhang, W., Xiang, B., Wang, X., Li, G., and Shen, S. (2012) New prognosis biomarkers identified by dynamic proteomic analysis of colorectal cancer, *Molecular bioSystems* 8, 3077-3088.
114. Fan, N.-J., Gao, J.-L., Liu, Y., Song, W., Zhang, Z.-Y., and Gao, C.-F. (2015) Label-Free Quantitative Mass Spectrometry Reveals a Panel of Differentially Expressed Proteins in Colorectal Cancer, *BioMed Research International* 2015, 13.
115. Zhang, Y., Ye, Y., Shen, D., Jiang, K., Zhang, H., Sun, W., Zhang, J., Xu, F., Cui, Z., and Wang, S. (2010) Identification of transgelin-2 as a biomarker of colorectal

cancer by laser capture microdissection and quantitative proteome analysis, *Cancer science* 101, 523-529.

116. Tonus, C., Sellinger, M., Koss, K., and Neupert, G. (2012) Faecal pyruvate kinase isoenzyme type M2 for colorectal cancer screening: a meta-analysis, *World journal of gastroenterology : WJG* 18, 4004-4011.
117. Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2007) Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ), *Journal of proteome research* 6, 821-827.
118. Wu, W. W., Wang, G., Baek, S. J., and Shen, R. F. (2006) Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF, *Journal of proteome research* 5, 651-658.
119. Walther, T. C., Olsen, J. V., and Mann, M. (2010) Yeast expression proteomics by high-resolution mass spectrometry, *Methods in enzymology* 470, 259-280.
120. Nozaki, Y. (1972) The preparation of guanidine hydrochloride, *Methods in enzymology* 26, 43-50.
121. Lomenick, B., Olsen, R. W., and Huang, J. (2011) Identification of Direct Protein Targets of Small Molecules, *Acs Chem Biol* 6, 34-46.
122. Villamor, J. G., Kaschani, F., Colby, T., Oeljeklaus, J., Zhao, D., Kaiser, M., Patricelli, M. P., and van der Hoorn, R. A. L. (2013) Profiling Protein Kinases and Other ATP Binding Proteins in Arabidopsis Using Acyl-ATP Probes, *Molecular & Cellular Proteomics* 12, 2481-2496.
123. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods in enzymology* 131, 266-280.
124. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes, *Nature protocols* 1, 2856-2860.

125. Wang, M. Z., Shetty, J. T., Howard, B. A., Campa, M. J., Patz, E. F., Jr., and Fitzgerald, M. C. (2004) Thermodynamic analysis of cyclosporin a binding to cyclophilin a in a lung tumor tissue lysate, *Analytical chemistry* 76, 4343-4348.
126. West, G. M., Tang, L., and Fitzgerald, M. C. (2008) Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy, *Analytical chemistry* 80, 4175-4185.
127. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding, *Protein Sci* 4, 2138-2148.
128. Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone, *Cell* 90, 65-75.
129. Harding, M. W., and Handschumacher, R. E. (1988) Cyclophilin, a primary molecular target for cyclosporine. Structural and functional implications, *Transplantation* 46, 29S-35S.
130. Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J., and Speicher, D. W. (1984) Cyclophilin: a specific cytosolic binding protein for cyclosporin A, *Science* 226, 544-547.
131. Liu, J., Albers, M. W., Chen, C. M., Schreiber, S. L., and Walsh, C. T. (1990) Cloning, expression, and purification of human cyclophilin in *Escherichia coli* and assessment of the catalytic role of cysteines by site-directed mutagenesis, *Proceedings of the National Academy of Sciences of the United States of America* 87, 2304-2308.
132. Jorgensen, P. L., Hakansson, K. O., and Karlsh, S. J. D. (2003) Structure and mechanism of Na,K-ATPase: Functional sites and their interactions, *Annu Rev Physiol* 65, 817-849.
133. Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D. L., and O'Halloran, T. V. (2002) Metallochaperones and metal-transporting ATPases: A comparative analysis of sequences and structures, *Genome research* 12, 255-271.

134. Pause, A., Methot, N., and Sonenberg, N. (1993) The HriGrxxr Region of the Dead Box Rna Helicase Eukaryotic Translation Initiation Factor-4a Is Required for Rna-Binding and Atp Hydrolysis, *Molecular and cellular biology* 13, 6789-6798.
135. Weng, Y. M., Czaplinski, K., and Peltz, S. W. (1998) ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities, *Rna* 4, 205-214.
136. Apweiler, R., Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alam-Faruque, Y., Alpi, E., Antunes, R., Arganiska, J., Casanova, E. B., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chan, W. M., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Fazzini, F., Gane, P., Castro, L. G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Legge, D., Liu, W. D., Luo, J., MacDougall, A., Mutowo, P., Nightingale, A., Orchard, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G. Y., Rosanoff, S., Sawford, T., Shypitsyna, A., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Corbett, M., Donnelly, M., Van Rensburg, P., Goujon, M., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Binz, P. A., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., de Castro, E., Cerutti, L., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., James, J., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lew, J., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C. M., Chen, Y. X., Garavelli, J. S., Huang, H. Z., Laiho, K., McGarvey, P., Natale, D. A., Suzek, B. E., Vinayaka, C. R., Wang, Q. H., Wang, Y. Q., Yeh, L. S., Yerramalla, M. S., Zhang, J., and Consortium, U. (2014) Activities at the Universal Protein Resource (UniProt), *Nucleic acids research* 42, D191-D198.
137. DeArmond, P. D., West, G. M., Huang, H. T., and Fitzgerald, M. C. (2011) Stable isotope labeling strategy for protein-ligand binding analysis in multi-component protein mixtures, *Journal of the American Society for Mass Spectrometry* 22, 418-430.

138. Walther, T. C., and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology, *J Cell Biol* 190, 491-500.
139. Oppermann, F. S., Gnad, F., Olsen, J. V., Hornberger, R., Greff, Z., Keri, G., Mann, M., and Daub, H. (2009) Large-scale Proteomics Analysis of the Human Kinome, *Molecular & Cellular Proteomics* 8, 1751-1764.
140. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. G., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L. M., Marti, G. E., Moore, T., Hudson, J., Lu, L. S., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature* 403, 503-511.
141. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science* 286, 531-537.
142. Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebet, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., and Golub, T. R. (2005) MicroRNA expression profiles classify human cancers, *Nature* 435, 834-838.
143. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Aksten, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000) Molecular portraits of human breast tumours, *Nature* 406, 747-752.
144. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lonning, P. E., and Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, *Proceedings of the National Academy of Sciences of the United States of America* 98, 10869-10874.

145. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L., and Botstein, D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets, *Proceedings of the National Academy of Sciences of the United States of America* 100, 8418-8423.
146. Volinia, S., Calin, G. A., Liu, C. G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., Prueitt, R. L., Yanaihara, N., Lanza, G., Scarpa, A., Vecchione, A., Negrini, M., Harris, C. C., and Croce, C. M. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets, *Proceedings of the National Academy of Sciences of the United States of America* 103, 2257-2261.
147. Adam, B. L., Qu, Y. S., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z. D., and Wright, G. L. (2002) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men, *Cancer research* 62, 3609-3614.
148. Hanash, S. (2003) Disease proteomics, *Nature* 422, 226-232.
149. Li, J. N., Zhang, Z., Rosenzweig, J., Wang, Y. Y., and Chan, D. W. (2002) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer, *Clinical Chemistry* 48, 1296-1304.
150. Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z. P., Stokes, M., Sullivan, L., Mitchell, J., Wetzell, R., MacNeill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T. L., Polakiewicz, R. D., Rush, J., and Comb, M. J. (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer, *Cell* 131, 1190-1203.
151. Yanagisawa, K., Shyr, Y., Xu, B. G. J., Massion, P. P., Larsen, P. H., White, B. C., Roberts, J. R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J. H., Caprioli, R. M., and Carbone, D. P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer, *Lancet* 362, 433-439.

152. Jessani, N., Liu, Y. S., Humphrey, M., and Cravatt, B. F. (2002) Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness, *Proceedings of the National Academy of Sciences of the United States of America* 99, 10335-10340.
153. Apetri, A. C., Surewicz, K., and Surewicz, W. K. (2004) The effect of disease-associated mutations on the folding pathway of human prion protein, *Journal of Biological Chemistry* 279, 18008-18014.
154. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G., and Dobson, C. M. (2000) Mutational analysis of the propensity for amyloid formation by a globular protein, *Embo Journal* 19, 1441-1449.
155. Liemann, S., and Glockshuber, R. (1999) Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein, *Biochemistry* 38, 3258-3267.
156. Ma, J. Y., Wollmann, R., and Lindquist, S. (2002) Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol, *Science* 298, 1781-1785.
157. Qu, B. H., and Thomas, P. J. (1996) Alteration of the cystic fibrosis transmembrane conductance regulator folding pathway - Effects of the Delta F508 mutation on the thermodynamic stability and folding yield of NBD1, *Journal of Biological Chemistry* 271, 7261-7264.
158. Varani, L., Hasegawa, M., Spillantini, M. G., Smith, M. J., Murrell, J. R., Ghetti, B., Klug, A., Goedert, M., and Varani, G. (1999) Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17, *Proceedings of the National Academy of Sciences of the United States of America* 96, 8229-8234.
159. West, G. M., Tucker, C. L., Xu, T., Park, S. K., Han, X. M., Yates, J. R., and Fitzgerald, M. C. (2010) Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements, *Proceedings of the National Academy of Sciences of the United States of America* 107, 9078-9082.

160. Strickland, E. C., Geer, M. A., Tran, D. T., Adhikari, J., West, G. M., DeArmond, P. D., Xu, Y., and Fitzgerald, M. C. (2013) Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation, *Nat Protoc* 8, 148-161.
161. Ong, S. E., and Mann, M. (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC), *Nature protocols* 1, 2650-2660.
162. Bendall, S. C., Hughes, C., Stewart, M. H., Doble, B., Bhatia, M., and Lajoie, G. A. (2008) Prevention of amino acid conversion in SILAC experiments with embryonic stem cells, *Molecular & Cellular Proteomics* 7, 1587-1597.
163. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nature biotechnology* 26, 1367-1372.
164. Vora, H. H., Patel, N. A., Rajvik, K. N., Mehta, S. V., Brahmbhatt, B. V., Shah, M. J., Shukla, S. N., and Shah, P. M. (2009) Cytokeratin and vimentin expression in breast cancer, *Int J Biol Marker* 24, 38-46.
165. Cimpean, A. M., Suci, C., Ceausu, R., Tatucu, D., Muresan, A. M., and Raica, M. (2008) Relevance of the immunohistochemical expression of cytokeratin 8/18 for the diagnosis and classification of breast cancer, *Rom J Morphol Embryo* 49, 479-483.
166. Derycke, L., Stove, C., Vercoutter-Edouart, A. S., De Wever, O., Dolle, L., Colpaert, N., Depypere, H., Michalski, J. C., and Bracke, M. (2011) The role of non-muscle myosin IIA in aggregation and invasion of human MCF-7 breast cancer cells, *Int J Dev Biol* 55, 835-840.
167. Tian, H. M., Liu, X. H., Han, W., Zhao, L. L., Yuan, B., and Yuan, C. J. (2013) Differential expression of filamin A and its clinical significance in breast cancer, *Oncol Lett* 6, 681-686.
168. Krasnov, G. S., Dmitriev, A. A., Snezhkina, A. V., and Kudryavtseva, A. V. (2013) Deregulation of glycolysis in cancer: glyceraldehyde-3-phosphate dehydrogenase as a therapeutic target, *Expert Opin Ther Tar* 17, 681-693.

169. Dennison, J. B., Molina, J. R., Mitra, S., Gonzalez-Angulo, A. M., Balko, J. M., Kuba, M. G., Sanders, M. E., Pinto, J. A., Gomez, H. L., Arteaga, C. L., Brown, R. E., and Mills, G. B. (2013) Lactate Dehydrogenase B: A Metabolic Marker of Response to Neoadjuvant Chemotherapy in Breast Cancer, *Clin Cancer Res* 19, 3703-3713.
170. Brown, N. J., Higham, S. E., Perunovic, B., Arafa, M., Balasubramanian, S., and Rehman, I. (2013) Lactate Dehydrogenase-B Is Silenced by Promoter Methylation in a High Frequency of Human Breast Cancers, *PLoS one* 8.
171. Capello, M., Ferri-Borgogno, S., Cappello, P., and Novelli, F. (2011) alpha-Enolase: a promising therapeutic and diagnostic tumor target, *The FEBS journal* 278, 1064-1074.
172. Shiba, E., Kambayashi, J. I., Sakon, M., Kawasaki, T., Kobayashi, T., Koyama, H., Yayoi, E., Takatsuka, Y., and Takai, S. I. (1996) Ca²⁺-Dependent Neutral Protease (Calpain) Activity in Breast Cancer Tissue and Estrogen Receptor Status, *Breast Cancer* 3, 13-17.
173. Dulyaninova, N. G., House, R. P., Betapudi, V., and Bresnick, A. R. (2007) Myosin-IIA heavy-chain phosphorylation regulates the motility of MDA-MB-231 carcinoma cells, *Mol Biol Cell* 18, 3144-3155.
174. Mi, H. Y., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013) Large-scale gene function analysis with the PANTHER classification system, *Nature protocols* 8, 1551-1566.
175. Mi, H. Y., Muruganujan, A., and Thomas, P. D. (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees, *Nucleic acids research* 41, D377-D386.
176. Lacroix, M., and Leclercq, G. (2004) Relevance of breast cancer cell lines as models for breast tumours: an update, *Breast cancer research and treatment* 83, 249-289.
177. van Staveren, W. C., Solis, D. Y., Hebrant, A., Detours, V., Dumont, J. E., and Maenhaut, C. (2009) Human cancer cell lines: Experimental models for cancer cells in situ? For cancer stem cells?, *Biochimica et biophysica acta* 1795, 92-103.

Biography

Jagat Adhikari was born in Belahani of Nawalparasi district in Nepal on September 30, 1984. He graduated high school from Budhanikantha School in Kathmandu, Nepal in May, 2003 and was awarded a presidential scholarship to attend Lee University in Cleveland, Tennessee in the fall of 2005. Jagat graduated Magna cum laude with a B.S. in biochemistry in May, 2009. He entered graduate school in Biochemistry department at Duke University the following fall. Jagat joined the laboratory of Professor Michael C. Fitzgerald with a doctoral fellowship from the Duke University graduate school. He was married to Ritu Poudel on June 12, 2013. He was also the recipient of several travel grants, including a United States Human Proteome Organization Travel award and an American Society for Mass Spectrometry Travel Stipend. He is a member of the American Society for Mass Spectrometry since 2012 and United States Human Proteome Organization since 2013. Jagat received his PhD in Biochemistry in May of 2015.

Publications

1. **Adhikari, J.**, West, G.M., and Fitzgerald, M.C. Global Analysis of protein folding thermodynamics for Disease State Characterization, Journal of Proteome Research, 2015 (In press).

2. **Adhikari, J.** & Fitzgerald, M.C. SILAC-Pulse Proteolysis: A Mass Spectrometry-Based Method for Discovery and Cross-Validation in Proteome-Wide Studies of Ligand Binding. *J Am Soc Mass Spectrom* 25, 2073-83 (2014).
3. Tran, D.T., **Adhikari, J.** & Fitzgerald, M.C. SILAC-Based Strategy for Proteome-Wide Thermodynamic Analysis of Protein-Ligand Binding Interactions. *Molecular & Cellular Proteomics* (2014).
4. Strickland, E. C.; Geer, M. A.; Tran, D. T.; **Adhikari, J.**; West, G. M.; DeArmond, P. D.; Xu, Y.; Fitzgerald, M. C., Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation. *Nature protocols* (2013).