AN ABSTRACT OF THE DISSERTATION OF

Ievgen Motorykin for the degree of Doctor of Philosophy in Chemistry presented on April 17 2015.

Title: Mass-spectrometric Methods for Quantitative Proteomics and Post-translational Modification Mapping

Abstract approved: _____

Claudia S. Maier

Signal transduction within and between cells is at the core of biological activity in all living systems. Signaling networks are required for regulating biological functions, including growth, development and survival. Deregulation of signaling cascades has been linked to chronic and acute diseases and disorders This thesis focuses on mass spectrometry as a high resolution and high mass accuracy technique for the detection and characterization of proteins in biological systems.

The thesis presents applications of contemporary mass-spectrometric methods to identify proteins, determine changes in their expression levels, and characterize post-translational modifications, in an effort to study changes in cell signaling in response to stress or disease. Various sample preparation methods to successfully suit needs of different biological questions were developed and applied: (1) extraction of the proteome, (2) chemical tagging and enrichment of the ATPome, a sub proteome comprising nucleotide binding proteins in particularly ATP-binding proteins including kinases, and (3) metal

affinity complexation and enrichment of phosphopeptides. We used the following hybrid mass analyzer configurations: a quadrupole time-of-flight (qToF) instrument with ion mobility, a linear ion trap hyphenated with a FT-ICR mass spectrometry (LTQ-FT) and a hybrid ion trap-orbitrap mass spectrometer (Orbitrap Elite). The bioinformatic analysis of the proteomics data required multiple combinations of software packages to sequence proteins, perform absolute and relative quantification, statistically analyze and visualize data.

Chapter 3 describes the use of zebrafish (*Danio rerio*) as one of the few vertebrate models that similar to humans cannot synthesize vitamin C to investigate the system-wide consequences of deficiencies in two essential micronutrients, vitamins E and C, on the proteome biology. A label-free proteomics workflow was applied to detect changes in protein abundance estimates dependent on vitamin regimes. The study reveals suppression in an energy metabolism cycle, glycolysis, in vitamin C and E deficient zebrafish. It was discovered that alternative energy cycle, glutaminolysis, is activated to fulfill energy requirement.

Chapter 4 focuses on the determination of proteome differences that can be linked to the propensity of metastasis in osteosarcoma (OS), a bone cancer that predominantly targets the adolescent age group. OS has a high propensity to metastasize to the lungs, which is associated with a poor prognosis. The study utilizes canine osteosarcoma cell lines that were originally obtained from orthotropic primary OS and metastatic cells. *Canis familiaris*, the domestic dog, is an established large animal model of OS that recapitulate

many biological and clinical features of the human malignancy. We applied a two-prone comparative proteomics approach that consisted of: (a) determination of protein abundance levels and (b) focus on kinases, a functional sub-proteome, using a chemical affinity tag for enrichment of ATP-binding proteins. Findings of this study indicated that in the highly metastasizing canine osteosarcoma cell line proteins associated with extracellular adhesion were deregulated, which may enhance metastogenesis. Mitogen activated protein kinases MAP2K6, MAP4K3, MAP4K4, MAP4K5, ZAK and v-akt murine thymoma viral oncogene homolog 1, AKT1, are among those expressed in significantly lower abundance in the highly metastasizing canine osteosarcoma cell line indicating changes in cell signaling.

Chapter 5 describes the development and application of a multiple protease protocol (Trypsin, LysC, AspN, Chymotrypsin and GluC) for improving the number of phosphosite identifications in a large-scale phosphoproteomics studies. The method combines immobilized titanium ion affinity chromatography (Ti^{4+}-IMAC) with a data-dependent, decision tree-based data acquisition technique utilizing two complementary fragmentation methods, namely collision induced dissociation (CID) and electron transfer dissociation. The multiple protease protocol was applied to human leukemic T cell lymphoblasts (Jurkat E6.1) and resulted in the detection of >11,000 unique phosphosites, the most comprehensive identification among methods that use similar phosphopeptide enrichment approach.

Mass-spectrometric Methods for Quantitative Proteomics and Post-translational Modification Mapping

by
Ievgen Motorykin

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 17, 2015
Commencement June 2015

Doctor of Philosophy dissertation of Ievgen Motorykin presented on April 17, 2015

APPROVED:

Major Professor, representing Chemistry

Chair of the Department of Chemistry

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Ievgen Motorykin, Author

ACKNOWLEDGEMENTS

I wish to express a sincere gratitude to my major adviser, Dr. Claudia Maier, for her guidance and support over these years. She introduced me to the mass-spectrometry and proteomics, fields I knew little about, and for that I am forever grateful. I am also grateful to my committee members, Dr. Staci Simonich, Dr. Neal Sleszynski, Dr. Robert Tanguay and Dr. John Selker for their timely support and advices. Special thanks to Dr. Albert Heck and Dr. Thin-Thin Aye for their warm reception and exciting project at Utrecht University.

I am thankful to my current and past lab-mates (Sasidhar, New City, Liping, Fereshte, Lin, Jeremiah, Zifeng, Samanthi and Jeff) for their support and suggestions. Special thanks to Dr. Cristobal Miranda for much help with experiments. Thanks to the staff of the Chemistry Department, especially to Dr. Carter, for their efforts to guide me through graduate school.

To my Corvallis friends I say many thanks for fun, unforgettable times together and all the exciting trips to the wild.

I wish to express true gratitude to my brother Oleksii, it was always fun to go through everything together. My many thanks to Valeriya Bychkova, with whom I was experiencing the most exciting time in my life. Thanks to my mother Olga, and friends from Ukraine for their long distance support.

CONTRIBUTION OF AUTHORS

Dr. Claudia S. Maier from Oregon State University provided advice and guidelines in all aspects of this dissertation.

Chapter 3. Dr. Robert L. Tanguay and Dr. Maret G. Traber developed the feeding protocol of the zebrafish study. Zebrafish was grown and maintained in the Sinnhuber Aquatic Research Laboratory (SARL).

Chapter 4. Dr. Milan Milovancev and Dr. Shay Bracha helped with developing the experimental design for the osteosarcoma studies. OS cells were grown by Dr. Shay Bracha.

Chapter 5. Dr. Albert Heck conceived the project. Dr. Thin-Thin Aye grew the cells and assisted in the operation of the LC-MS/MS mass spectrometry system .This project was in part made possible by a Marie Curie Actions - International Research Staff Exchange Scheme FP7-PEOPLE-2010_IRSES-269-256-MSLIFE.

TABLE OF CONTENTS

Page

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 METHODS 4

Mass-spectrometry workflow 6

Quantitative proteomics 10

CHAPTER 3 Proteome-Driven Elucidation of Adaptive Responses to Combined Vitamin E and C Deficiency in Zebrafish 15

Abstract 16

Introduction 16

Materials and Methods 21

Chemiscals 21

Fish Husbandry 22

Feeding Experiment 22

Protein Extraction and Digestion 23

Liquid Chromatography−Mass Spectrometry 24

Peak Detection 25

Protein Database Search 26

Protein Identification, Quantification, and Validation 26

TABLE OF CONTENTS (Continued)

Page

Western Blot 27

Network Construction 28

Results and Discussions 28

Overall Design and Strategy Applied for the Label-Free Quantitative Proteomic Study 28

Effects on Body Weight and Growth Rate 29

Optimization of Sample Preparation and Analytical Workflow for the Comparative "Bottom-up" Proteomics Study 30

Data Analysis Strategy Used for Obtaining Protein Abundance Estimates 31

Proteome Changes Linked to Vitamin C Deficiency 33

Responsive Protein Interaction Network of Vitamin C Deficiency 36

Biological Inferences of Altered Protein Expressions in Vitamin-E- and -C-Deficient Zebrafish 37

Conclusions 40

CHAPTER 4 Application of an ATP-affinity Tag Complements a Full Proteomics Experiment in Dissecting Regulatory Pathways Linked to Metastasis in Osteosarcoma ..42

TABLE OF CONTENTS (Continued)

Page

Abstract 43

Introduction 44

Materials and Methods 49

Chemicals 49

Enrichment of ATP-binding proteins 50

Liquid Chromatography - Mass Spectrometry 54

Peak Detection and Protein Database Search 55

Data Analysis and Representation 56

Western Blots 59

Results and Discussions 60

Experimental Strategy 60

Assessment of the ATPome Enrichment Efficiency 61

Quantitative Proteome Changes in Metastatic vs. Primary OS Cells 70

Pathway Analysis – Adherens Junctions and Cytoskeleton 71

Pathway Analysis – AKT1 and Downstream Signaling 80

TABLE OF CONTENTS (Continued)

Page

Summary 82

CHAPTER 5 Ti^{4+}-IMAC Approach Reveals Complementarity of Proteases to Cover the Phosphoproteome of Jurkat Cell Line 84

Abstract 85

Introduction 86

Materials and Methods 89

Chemicals 89

Experimental Workflow 89

Phosphopeptide Enrichment 91

Liquid Chromatography - Mass Spectrometry 92

Peak Detection and Protein Database Search 93

Results and Discussions 94

Conclusions 101

CONCLUSIONS 103

BIBLIOGRAPHY 105

APPENDICES 128

TABLE OF CONTENTS (Continued)

Page

APPENDIX A. 129

APPENDIX B 139

APPENDIX C. 146

LIST OF FIGURES

Figure Page

Figure 2.1 The generic mass-spectrometric workflow 6

Figure 2.2 The schematics of application of a chemical affinity probe.....7

Figure 2.3 The schematics of application of Ti^{4+}-IMAC phosphopeptide enrichment 8

Figure 2.4 The schematics of label free quantification.....12

Figure 3.1 Data analysis workflow for label-free quantitative proteomics using LC−IMS−MS^E acquisitions. 29

Figure 3.2 Dynamic range of the method based on protein abundance estimates. 32

Figure 3.3 Volcano plot representation of protein abundance changes caused by vitamin C deficiency. 35

Figure 3.4 Protein-protein interaction network visualization of subnetworks that contain proteins that showed changes in expression levels upon vitamin C deficiency in the vitamin-E-deficient state. 37

Figure 3.5 Schematic of the metabolic pathways that were associated with proteins that showed changes in abundance estimates as a consequence of vitamin C deficiency. 38

Figure 4.1 Diagram of the data analysis workflow.....56

Figure 4.2 Experimental design. 60

Figure 4.3 Number of proteins identified after application of the ID Filter. 62

Figure 4.4 Comparison of the number of individual proteins between cell lines within each experiment. 62

Figure 4.5 Comparison of Gene Ontology Molecular Functions terms within each experiment.....64

Figure 4.6 KEGG pathways enriched in both experiments with (A) and without (B) mutually found proteins. 65

LIST OF FIGURES (Continued)

Figure Page

Figure 4.7 Protein kinases detected with application of the ATP affinity tag along with their quantitative changes labeled on the Kinome Map. 67

Figure 4.8 Distribution of the number of protein kinases detected within their families (A) and relative enrichment of protein kinases within their families (B). 68

Figure 4.9 Protein interaction network for proteins found in the ATPome enrichment experiment. 69

Figure 4.10 Cycle of Rho GTPases between GTP bound active and GDP bound inactive state. 73

Figure 4.11 Building blocks and changes in the pathways related to the RHOA mediated weakening of extracellular adhesion. 75

Figure 5.1. Project aims 95

Figure 5.2 Aim 1. Ti^{4+}-IMAC reproducibility. 97

Figure 5.3 Aim 2. Complementarity of protease workflows. 98

Figure 5.4 Aim 3. Benefit from combining different digestion strategies. 99

Figure 5.5. Aim 4. The number of PS identification is increased by including experimental replicates and by adding results from a different protease. 100

LIST OF TABLES

Table Page

Table 2.1 Common MS-enabled experimental strategies for studying signaling pathways.5

Table 2.2 Cleavage specificity of various proteases. 7

Table 2.3 Characteristics of common mass spectrometers used in LC–MS. 9

Table 4.1 Proteins uniquely expressed in either cell line for each workflow (FP and ATPome enrichment). 63

Table 5.1 Number of phosphosites identified 96

Table 5.2 Number of unique phosphosites identified 96

Table 5.3 Complementarity of proteases 96

Table 5.4 Phosphorylation of residues 101

LIST OF APPENDIX FIGURES

Figure Page

Appendix Figure A.1 Design of the feeding experiment. 130

Appendix Figure A.3 Workflow used for preparing the samples for the label-free "bottom-up" quantitative proteomics study. 132

Appendix Figure A.5 Fish characteristics dependent on the vitamin regime fed. 134

Appendix Figure A.7 A boxplot of the distribution of the total protein concentration after extraction and lysis. 135

Appendix Figure A.8 Gradient profiles with respective base peak chromatograms. 135

Appendix Figure A.9 Venn diagrams comparing the number of peptides eluting from a column and identified with LC−IM−MS^E. 136

Appendix Figure A.10 Correlation plot between E−C+ sample 9 and E−C+ sample 8. .136

Appendix Figure A.11 Correlation plot matrix of the E−C+ group. 137

Appendix Figure A.12 The responsive protein network for the vitamin transition C+/C−and both groups deficient in vitamin E status (E−C+ vs E−C+). 138

Appendix Figure B.1 The probe used for enriching ATP-binding proteins 140

Appendix Figure B.2 The schematics of ATPome and FP workflows.....141

Appendix Figure B.3 The most abundant proteins after ATPome enrichment 142

Appendix Figure B.4 Distributions of protein abundance fold changes between cell lines143

Appendix Figure B.5 Western blotting against ACTB (A) and AKT1 (B). 144

Appendix Figure B.6 Observed proteins and their level changes mapped on the MAPK signaling cascade. 145

Appendix Figure C.1 Experimental workflow.....147

LIST OF APPENDIX TABLES

Table Page

Appendix Table A.2 Pipetting scheme used for sample lysis and preparation of digests.131

Appendix Table A.4 Synapt G2 HDMS instrument settings used for Ion-Mobility-MS^E acquisitions. 133

Appendix Table A.6 Fish characteristics dependent on the vitamin regime fed. 134

CHAPTER 1 INTRODUCTION

Signal transduction within and between cells is at the core of biological activity in all living systems. Studying signaling pathways has become increasingly relevant in discovering marker of diseases and is the focus of much of the ongoing biological research. The signal transfer goes mostly through proteins, the workforce of the cells, and is reflected by changes in protein expression levels or via post-translational modifications. Typical signaling cascade starts when an extracellular ligand binds to a protein receptor at the cell's surface. This triggers a chain of phosphorylation events that transfers the signal to the nucleus and affects gene expressions, which regulate cellular activities.

The system of communications called signaling networks is required for almost all biological functions, including growth, development and survival, which is why a disruption of the signaling cascade may cause a disorder or a disease. One of the ways to mediate signal transduction is by reversible protein phosphorylation, hence, studying this post-translational modification along with key regulators, protein kinases and phosphatases, is critical for understanding pathways of diseases.

This thesis presents applications of contemporary mass-spectrometric methods to identify proteins, determine changes in their expression levels, and characterize post-translational modifications, in effort to study changes in cell signaling in response to stress or disease. Separation of analytes is carried out by ultra-high performance liquid chromatography (UHPLC) and the acquired data is processed by multiple combinations of tools for

proteins search, statistical analysis and visualization.

The first project uses zebrafish (*Danio rerio*) as one of the few vertebrate models that similar to humans cannot synthesize vitamin C to investigate the system-wide consequences of deficiencies in the two essential micronutrients, vitamin E and C, on the proteome biology. A label-free proteomics workflow was applied to detect changes in protein abundance estimates dependent on vitamin regimes. For mass spectrometric detection, a high resolution quadrupole time-of-flight (qTOF) instrument equipped with an ion mobility separation device was used. The study reveals suppression in an energy metabolism cycle, glycolysis, in vitamin C and E deficient zebrafish. It was discovered that an alternative energy cycle, glutaminolysis, is activated to fulfill energy requirement.

Use of a chemical affinity tag, desthiobiotin-ATP, for enrichment of the ATPome, a functional sub-proteome, in canine osteosarcoma cell lines is described in the second project. For this study, ultra-high resolution, high mass accuracy quadrupole linear ion trap Fourier Transform ion cyclotron resonance mass spectrometry (LTQ FT-ICR MS) was used for enabling protein identification, determination of protein level changes and PTM mapping. Results of this investigation led to the hypothesis that in highly metastasizing canine osteosarcoma cell line, a complex chain of signaling that starts with the GTPase RHOA leads to weakening of extracellular adhesion, which potentially may promote enhanced propensity of cells to detach from their neighbors. Many protein kinases were expressed in significantly different abundances, including AKT1, mitogen activated protein kinases (MAPKs) and cyclin dependent kinases (CDKs), indicating

disparate cell signaling in the two OS cell lines.

The last project introduces a powerful method for phosphosite enrichment and shows that application of Ti^{4+} immobilized ion affinity chromatography (Ti^{4+}-IMAC) with the use of various proteases greatly improves the number of identifiable phosphosites. The detection of phosphopeptides was carried out with a quadrupole linear ion trap Orbitrap analyzer equipped with an electron-transfer dissociation (ETD) source to preserve protein post translational modifications (PTM). To this end this study identified the highest number of unique phosphorylation sites described for Jurkat E6.1 cells after treatment with PGE2.

CHAPTER 2 METHODS

During recent years many approaches have been utilized to address various biological questions. They present diverse methods of sample preparation, analyte enrichment, detection, quantification and data analysis. The need for high-throughput analytical methods to analyze a large number of analytes in complex matrices has greatly increased. The main objective of an analytical laboratory is to develop a method of fast, quantitative and qualitative analysis. High performance liquid chromatography (HPLC) and its modern variation, ultra high performance liquid chromatography (UHPLC), have become standard techniques for analytical separations.

Mass spectrometry has emerged as the primary technology for the detection, characterization and quantification of proteins and peptides. The combination of liquid chromatography and mass spectrometry (LC-MS) has become the premier technique of choice in analytical proteomics. Mass spectrometry techniques have a sufficient dynamic range for enabling quantitative proteomics studies. MS-based approaches have been developed for determining protein abundance levels and/or relative fold-changes. In addition, tandem mass spectrometry technology provides access to the identification and site location of PTMs. Common strategies of how mass spectrometry can be implemented for studying cell signaling are listed in Table 2.1 (adapted from Li, Wang and Chen[1]). Cells can be subjected to analysis in their basal state, or treated with drugs, inhibitors or other chemicals. Parts of the proteome can be fractionated (enriched) or separated to investigate changes in specific pathways. Mass spectrometry will be used to measure

Table 2.1 Common MS-enabled experimental strategies for studying signaling pathways.

<table>
<tr><th>Step1: Sample preparation</th><th>Step 2: MS application</th><th colspan="2">Step 3: Pre-MS strategies</th><th>Step 4: MS and data analysis</th></tr>
<tr><td>Drug/inhibitor treatment</td><td rowspan="2">Protein abundance</td><td rowspan="3">Protein labels</td><td>Isobaric-tag labeling</td><td rowspan="7">Mass Spectrometry</td></tr>
<tr><td rowspan="2">Ligand treatment</td><td rowspan="2">Post-translational modifications (PTM)</td><td>“shotgun”, no labels</td></tr>
<tr><td>Isotope labeling</td></tr>
<tr><td rowspan="2">Basal state</td><td rowspan="4">Protein-protein interactions</td><td rowspan="4">Affinity purification</td><td>TAP</td></tr>
<tr><td rowspan="2">Fractionation/ Separation of organelles</td><td rowspan="2">One-step purification</td></tr>
<tr><td rowspan="2">Crosslinking of protein complexes</td></tr>
<tr><td>Antibodies</td></tr>
</table>

abundances of proteins, or to follow changes in post-translational modifications (PTM) and protein-protein interactions (PPI).

Various pre-MS techniques are available to enable the identification and quantification of proteins. Labeling can be done by metabolic incorporation of isotopes, or chemical modifications using isotopically coded tags or isobaric tandem mass tags (additional details on these techniques are given later in the text). Alternatively, label-free “shotgun” strategies have recently gained popularity for quantifying protein levels. Affinity purification uses antibodies against endogenous proteins or epitope-tagged bait proteins for one step and tandem affinity purification (TAP). Diverse mass spectrometry instrumentation is then used to detect and sequence ions (usually peptides). Database search engines are used to match the mass spectral data to protein amino acid sequences

curated in diverse databases. As the last step a statistical analysis is performed and data is visualized to reflect a stated hypothesis.

Mass-spectrometry workflow

The generic mass-spectrometric workflow used in this work consists of 6 steps and is presented in Figure 2.1. Protein mixtures are fractionated with gel electrophoresis, isoelectric focusing or liquid chromatography. On the other hand, proteins may be subjected to an enrichment technique, with which a protein is selectively pulled out of a complex matrix. Chemical affinity tags are effective ways for enriching proteins. Affinity labeling strategies use affinity probes that consist of an affinity tag and a substrate that targets the substrate binding site of a protein or protein family. After binding of the affinity probe to the substrate binding site of a protein, the latter reacts with the probe and under loss of the substrate, the protein is chemically tagged. The tag can be used to pull out the protein using a molecule that has high affinity to the tag (Figure 2.2). This technique has high selectivity toward distinct protein-substrate interactions. In Chapter 4 of this dissertation the application of chemical affinity tagging is described for the enrichment of low abundant ATP binding proteins.

Figure 2.1 The generic mass-spectrometric workflow

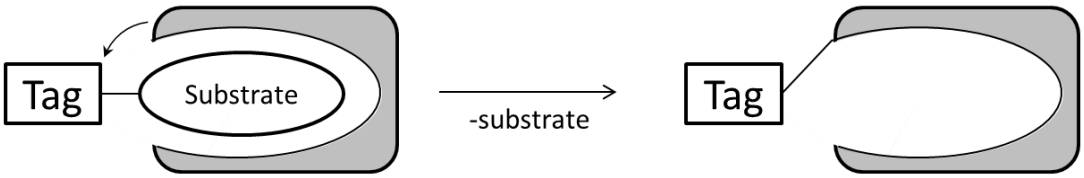


Figure 2.2 The schematics of application of a chemical affinity probe

In step 3 of the workflow (as outlined in Figure 2.1), proteins are proteolytically digested. Proteases commonly employed in "bottom up" shotgun proteomics workflows are listed in (Table 2.2). The endoprotease trypsin is most commonly used due to its specificity for Lys (K) and Arg (R) residues. Other proteases with different cleavage specificities can be used to obtain complementary sets of peptides for subsequent mass spectrometry analysis (Chapter 5 of this work).

Table 2.2 Cleavage specificity of various proteases.

Protease	Cleavage specificity
Trypsin	-K,R-|-X, not -K,R-|-P
LysC	-K-|-X
GluC	-E-|-X, much slower -D-|-X
AspN	-X-|-D and -X-|-cysteic acid, but not -X-|-C
Chymotrypsin	-W,F,Y-|-X, slower –I,M,A,D,E-|-X

After proteolytic digestion, peptides can be separated by various chromatographic techniques, e.g. reversed phase chromatography, strong cation exchange (SCX), strong anion exchange (SAX) or hydrophilic interaction liquid ion chromatography (HILIC). As with proteins, rather than separating peptides, one can selectively pull them out from a solution leaving the matrix behind. For enrichment of phosphorylated peptides, which are relevant in cell signaling pathways, immunoprecipitation, TiO_2 beads or titanium ion immobilized metal affinity chromatography (Ti^{4+}-IMAC) can be used. The schematics of the Ti^{4+}-IMAC phosphopeptide enrichment is presented in the Figure 2.3. After the Ti^{4+} ions are immobilized on the polymer, the peptide solution is applied to the beads. The phosphorylated peptides are attracted to the beads while non- phosphorylated washed out. The IMAC-based phosphopeptide enrichment workflow is described in Chapter 5 of this

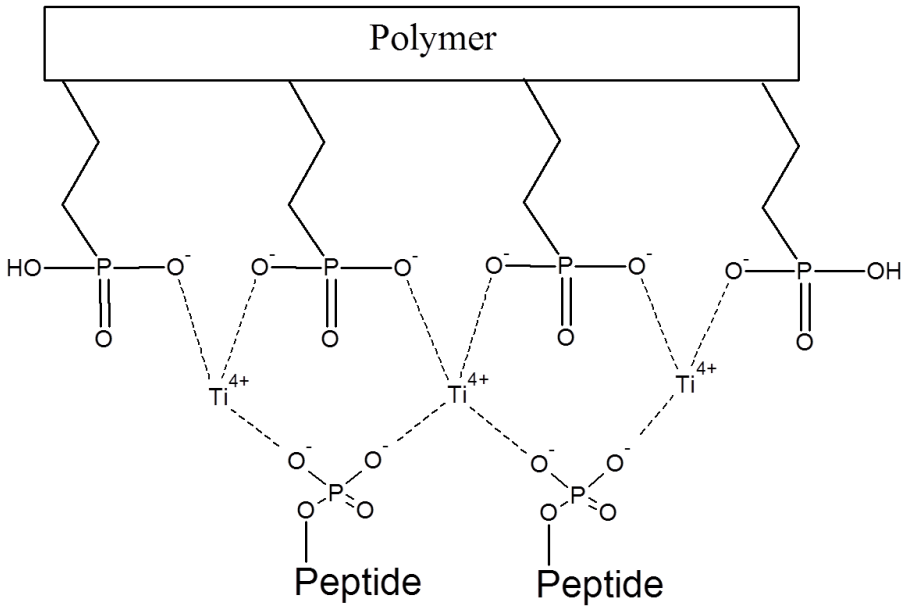


Figure 2.3 The schematics of application of Ti^{4+}-IMAC phosphopeptide enrichment

dissertation and was applied towards the identification of the phosphoproteome of Jurkat E6.1 cells.

Recent developments of new mass spectrometry techniques and instrumentation[2], sample preparation[3] and enrichment methods[4], data analysis[5] and visualization tools[6,7] have made mass spectrometry a method of choice for the analysis of complex biological samples, including proteins[8,9] and metabolites[10,11]. Several thousand proteins can be routinely identified and quantified[12] to measure changes in protein expressions in an organism subjected to external stimuli[13], stress[9] or disease[14,15].

The market of MS is fast and dynamic, driving manufacturers to invest in the development of new technologies and their implementation in state-of-the-art instruments. UHPLC and matrix assisted laser desorption/ionization (MALDI) are the two most popular methods to introduce samples into the MS analyzer. There are many types of mass analyzers in MS to fit the research needs. The main five of them are

Table 2.3 Characteristics of common mass spectrometers used in LC–MS.

Mass analyzer type	Resolving power, ($\times 10^3$)	Mass accuracy (ppm)	*m/z* range (upper limit), ($\times 10^3$)	Acquisition speed (Hz)	Linear dynamic range	Price
Q	3–5	Low	2–3	2–10	10^5–10^6	Lower
IT	4–20	Low	4–6	2–10	10^4–10^5	Moderate
TOF	10–60	1–5	10–20	10–50	10^4–10^5	Moderate
Orbitrap	100–240	1–3	4	1–5	5×10^3	High
ICR	750–2500	0.3–1	4–10	0.5–2	10^4	Very High

presented in Table 2.3 along with their basic characteristics (the table is adapted from Holčapek et al.[16]).

A basic ability of a mass spectrometer to resolve peaks is called resolving power (RP) and defined as the *m/z* value of a peak divided by the peak's full width at half maximum (FWHM): RP = m/z / FWHM. Mass accuracy (MA) is the error in measuring the m/z value. It is defined as $((m/z)_{exp} - (m/z)_{theor}) / (m/z)_{theor} \times 10^6$ and expressed in ppm. To increase mass accuracy one can use external calibration (similar compound run in parallel) or internal calibration (similar compound run within the sample). The rest of specifications are self-explanatory: the upper limit of m/z values that can be detected, the rate of scans, dynamic range of linear response, and relative cost of an instrument.

Quantitative proteomics

The proteome was defined as the total protein complement of a genome present in a cell and/or tissue.[17] However, this definition arguably implies a one-to-one ratio between the genome and the proteome. In the past years much of the experimental data that were collected reveal the highly increased complexity of the proteome over the genome. Complexity is achieved due to the fact one gene may lead to a several proteins through alternative splicing or a wide variety of post-translational modifications. Also, protein expression is cell type dependent and varies over time, further complicating studying of the proteome. This poses an analytical challenge for quantitative proteome analyses, e.g. for assessing differently expressed proteins between healthy cells and cells affected by a

disease or a stress. Therefore, the identity of a protein is insufficient for describing the proteome; it is critical to determine protein expression levels. This has become the emerging and fast growing field of quantitative proteomics.[18]

Quantification of proteins traditionally relied on the use of antibodies in combination with staining techniques that offer a high selectivity and sensitivity of detection. However, this is only a semi-quantitative technique and cannot be applied for a high throughput analysis, when thousands of proteins are detected simultaneously. 2D gel electrophoresis has enabled a more global analysis of protein expression levels.[19–21] In this approach protein are extracted from cells separated in a gel by their charge in the first dimension and the molecular weight in the second. Staining with a dye is used then to mark proteins and the densities of spots are then attributed as the measure of protein abundance. Serious disadvantages of this technique are limited dynamic range and the need of perfectly reproducible gels.

Alternative methods for accurate measurements of proteins expression levels have been developed, some based on the use of peptide (or protein) ion signals in mass spectra as an indicator of protein abundance. Even though quantification analysis in mass spectrometry may seem straightforward due to the fact that peptide ion intensities can be accurately determined, the peak intensity depends on the physicochemical properties of a peptide and its surrounding matrix, which may lead to suppression effects. Several studies have evaluated if there is dependence between protein concentration in a sample and its peak height/area in a mass spectrum or its corresponding ion chromatogram.[22,23] These studies

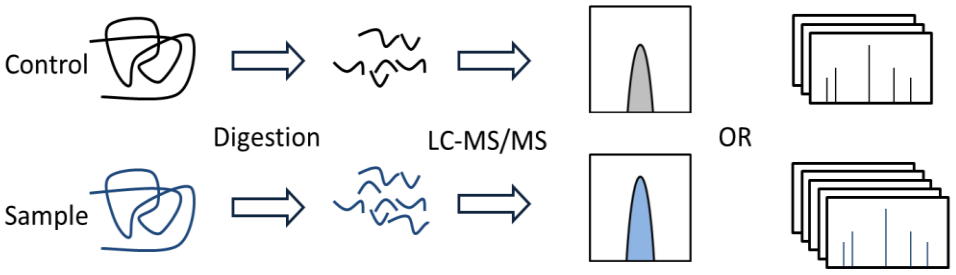


Figure 2.4 The schematics of label free quantification.

discovered that peak areas of peptides (taking at least three peptides per protein) correlate to their concentration in solution. The correlations were sufficiently good for determining the stoichiometry of individual protein constituents in multi-protein complexes.[22,23]

One approach to assess the absolute level of an analyte is the use of an internal standard. Isotope dilution methods have been widely implemented in bioanalysis.[24] The internal standard should be as close as possible to the analyte with respect to the chemo-physical properties. Ideally, the internal standard is the same molecule, but has stable isotope incorporated.[25–29] In a typical experiment (Figure 2.4), the standard, a.k.a. control, of known amount is added to the sample mixture or processed in parallel. After enzymatic digestion, the samples are subjected to the LC-MS/MS analysis and peak intensities or spectra counts are taken as a measure of the abundance of a protein with respect to the amount of the internal standard.

In quantitative proteomics it is often more desirable to investigate differentially expressed proteins between two or more sets of samples rather that to measure absolute levels of proteins. The stable isotope labeling approach incorporates different isotope atoms (or

isotope functional groups) to samples before introduction into mass spectrometer. Then, after sample preparation, the samples are pooled and proteins/peptides are analyzed with conventional proteomics methods.[30,31] The mass spectrometer separates the isomers due to their different m/z values. Assignment of isotopomeric pairs is based on the known mass difference between the isotopomers. The ratios of the intensities or areas of the isotopomeric peptides directly relate to the ratios of the protein expression levels between the two samples.

Some of the most popular methods that use isotopes for relative and absolute protein quantification are listed below:

1. Chemical isotope labeling uses differential isotopes to acetylate N-terminal amines using chemical tags;[32,33]
2. Isotope coded affinity tag (ICAT) uses a molecular mass tag, which specifically derivatizes cysteine, a linker region that is isotopically labeled and a biotin group for specific recovery of labeled peptides;[34]
3. Isobaric tandem mass tags: Newly developed version of the chemical isotope labeling introduces a tag after proteolysis. Different tags are incorporated to multiple sample digests. Samples are pooled and analyzed by MS. Unlike other techniques, the quantification step takes place in the MS/MS mode. This technique allows the use of isotopic labels for relative and absolute quantification (iTRAQ, TMT);[35]
4. Metabolic isotope labeling incorporates isotopes in proteins via the biological

rather than chemical route. Metabolic labeling requires the organism to be grown in predefined growth conditions. Stable isotope labeling by amino acids in cell culture (SILAC)[36] is a special form of metabolic labeling that requires depletion of the cell media from an essential amino acid and subsequent replenishment with a labeled version of the depleted amino acid;

5. For absolute protein quantification (also terms AQUA), an identical peptide with a stable isotope tag is used as internal standard.[37,38] Unlike previous techniques that can only compare protein expression between sets of samples the "AQUA" strategy is a great way to assess the absolute expression level of a protein or proteins, but the strategy requires synthetic peptides for all proteins of interest.

The isotope labeling techniques SILAC and the chemical tagging techniques, ICAT and iTRAQ, are conceptually very similar. The difference resides in the stage of the introduction of the isotope coding: in vivo (metabolic) labeling and chemical labeling of proteins and peptides.

CHAPTER 3 Proteome-Driven Elucidation of Adaptive Responses to Combined Vitamin E and C Deficiency in Zebrafish

Ievgen Motorykin,[1] Maret G. Traber,[2] Robert L. Tanguay,[3] and Claudia S. Maier*,[1]

[1]Department of Chemistry, Oregon State University, 153 Gilbert Hall, Corvallis, Oregon 97331, United States

[2]Linus Pauling Institute and School of Biological and Population Health Sciences, Oregon State University, 307 Linus Pauling Science Center, Corvallis, Oregon 97331, United States

[3]Department of Environmental and Molecular Toxicology, Oregon State University, 1007 Agriculture & Life Sciences Building, Corvallis, Oregon 97331, United States

J Proteome Res. 2014 Mar 7;13(3):1647-56.

doi: 10.1021/pr401108d. Epub 2014 Feb 10.

*Corresponding author, E-mail: claudia.maier@oregonstate.edu

Abstract

The purpose of this study was to determine the system-wide consequences of deficiencies in two essential micronutrients, vitamins E and C, on the proteome using zebrafish (Danio rerio) as one of the few vertebrate models that similar to humans cannot synthesize vitamin C. We describe a label-free proteomics workflow to detect changes in protein abundance estimates dependent on vitamin regimes. We used ion-mobility-enhanced data-independent tandem mass spectrometry to determine differential regulation of proteins in response to low dietary levels of vitamin C with or without vitamin E. The detection limit of the method was as low as 20 amol, and the dynamic range was five orders of magnitude for the protein-level estimates. On the basis of the quantitative changes obtained, we built a network of protein interactions that reflect the whole organism's response to vitamin C deficiency. The proteomics-driven study revealed that in vitamin-E-deficient fish, vitamin C deficiency is associated with induction of stress response, astrogliosis, and a shift from glycolysis to glutaminolysis as an alternative mechanism to satisfy cellular energy requirements.

Introduction

The purpose of this study was to determine the consequences of deficiencies in dietary antioxidants, vitamins C and E, on the zebrafish proteome composition. The combination of deficiencies was used because this dietary manipulation is known to cause severe abnormalities.[39,40] Protein abundance estimates were obtained by label-free accurate

quantification based on data-independent (MS^E) acquisition. Functional interpretation of the systems-wide proteome responses revealed the biomolecular networks affected by vitamin C and E deficiency.

In zebrafish, the lack of vitamin E potentiates vitamin C deficiency and causes degenerative myopathy.[41] The classical roles of vitamin C (ascorbic acid, AA) are in maintaining redox homeostasis[42–44] and facilitating the recycling of vitamin E.[45,46] More recently, vitamin C has been identified as an essential cofactor for multiple enzymes involved in the hydroxylation of procollagen[47,48] and several dioxygenases that are involved in maintaining endothelial function with relevance to atherosclerosis[49] and sepsis.[50,51] Inadequate vitamin C levels, potentially leading to hypovitaminosis, remain widespread health concerns in humans affecting particularly underprivileged populations. Recent research indicates that smoking[52] and Western diets rich in fat and cholesterol[53] exacerbate vitamin C deficiency.

In most animals, vitamin C is synthesized in the liver or kidneys and transferred to tissues via circulation.[54,55] Humans are not capable of synthesizing ascorbic acid due to the lack of the functional gene for L-gulonolactone oxidase.[56,57] For studying the effect of vitamin C deficiency, the most widely used animal model is the guinea pig that similarly to humans cannot synthesize vitamin C.[58] The gulonolactone oxidase knockout ($Gulo^{-/-}$) mouse is an established transgenic animal model for deciphering the physiological roles of vitamin C.[59,60] Similarly to humans, zebrafish also cannot synthesize vitamin C.[61–66] A first report using MS-based metabolomics to study the effect of vitamin C deficiency in

zebrafish revealed that vitamin C deficiency affects the purine metabolism.[11] A classical antioxidant function of vitamin C entails the recycling of the tocopheroxyl radical that is generated when vitamin E (α-tocopherol, α-T) functions as a peroxyl radical, chain-breaking antioxidant.[67,68] To evaluate the interrelating roles of vitamins C and E, we used a 2 × 2 design in the present study resulting in four dietary groups, namely, vitamin-E- and -C-adequate (E+C+), vitamin-E- and -C-inadequate (E−C−), vitamin-C-adequate but inadequate in vitamin E (E−C+), and vitamin-C-inadequate but adequate in vitamin E (E+C−). We subsequently applied a label-free comparative proteomics approach for examining the systems-wide consequences of the different vitamin regimes in adult zebrafish.

In a typical "bottom-up" shotgun experiment, a protein mixture is proteolytically digested with a protease, such as trypsin, to create peptides for further analysis. Liquid chromatography in combination with tandem mass spectrometry (LC−MS/MS) has become the predominantly used technique to both qualitatively and quantitatively analyze a complex mixture of peptides.[69,70] The peptides are chromatographically separated using reversed-phase (RP) chromatography and subsequently subjected to collision-induced dissociation (CID) in a tandem mass spectrometer. The product ion spectra are searched against protein sequence databases to identify the peptides. Unlike comparative isotope-coded labeling methods (e.g., ICAT or iTRAQ), label-free mass spectrometry-based quantitative proteomic approaches are based on the fact that in electrospray ionization (ESI) the peak signal intensity is linearly proportional to the concentration of a peptide

(over an approximate range of 5−5000 fmol).[71,72] Currently, two acquisition techniques have been technically realized and applied for detecting and quantifying peptides, namely, data-dependent acquisitions (DDAs) and data-independent acquisitions (DIAs).

Data-dependent acquisitions are based on tracking of precursor ion intensities on-the-go.[73]The peptide precursor ions that fulfill preselected features are sequentially subjected to MS/MS during the time course of the chromatography. Proteolytic digests in proteomics-type workflow are often complex and even when using ultraperformance liquid chromatography (UPLC) at every moment multiple peptides coelute. In DDA techniques, survey scan and MS/MS scans are used sequentially and compete for duty cycle time. Therefore, both the number of the most intense peaks subjected to MS/MS analysis and the timespan of a survey scan need to be optimized to acquire accurate mass information for the peptide precursor ions and to ensure information-rich fragment ion spectra.[74,75] Albeit somewhat dependent on the mass spectrometry platform used, a possible caveat of DDAs is that a substantial portion of the acquisition cycle is spent acquiring fragment ion spectra; therefore, the intensities of the peptide ions are measured inaccurately and, consequently, are less suited for estimating peptide abundances.[75]

DIA techniques perform parallel fragmentation of precursor ions, meaning that all peptide precursors are fragmented simultaneously regardless of their characteristics. For proteomic-type applications, currently most DIAs are realized by coupling nanoUPLC systems to high-resolution quadrupole time-of-flight (Q-TOF) mass spectrometers. This combination of equipment takes advantage of the high peak and separation capacity of

UPLCs and the measurement of precursor and product ions with accurate mass. In the DIA mode, the Q-TOF mass spectrometer conducts a low-energy MS scan providing access to precursor ion mass and intensity data for quantitation and an elevated energy scan to obtain product ion information. A critical post acquisition step of the DIA technique is the alignment of the fragment ions with the respective precursor ion. DIA-based acquisitions enable quantification based on the peptide ion intensities. The estimation of protein abundance is based on the observation that the average signal intensity of the three most intense tryptic peptides relates to the protein abundance level regardless of protein size.[76] The estimation of protein levels (moles of each identified protein) is based on using an internal protein standard (spiked in at a defined amount).

Recently, a commercially available hybrid high-resolution ion mobility quadrupole time-of-flight (IM-Q-TOF) mass spectrometry system that enables the acquisition of ion mobility enhanced DIA (aka IM-MS^E) data has become available.[77] In this instrument configuration, ions are separated according to their velocity with which they transverse the traveling wave ion mobility device. Parameters that affect the ion's mobility are its charge, size, and shape.[78] In ion mobility-enhanced MS^E acquisitions, ions are separated according to their mobility and then dependent on the energy regime applied in the transfer region directly passed on to the TOF analyzer or collisionally dissociated, and the resulting fragment ions are measured in the TOF analyzer. In the IM-MS^E mode, fragment ions are drift time-aligned with their precursors. An often touted advantage of ion mobility-enhanced MSE acquisitions is the orthogonal separation space gained

additionally to the chromatographic separation of the peptides prior to collision induced dissociation leading to spectral decongestion, improved peptide detection, and dynamic range, all leading cooperatively to more protein identifications.[79]

To investigate the consequences of the combination of vitamin E and C deficiency on protein networks in zebrafish, we profiled the changes in protein abundances. The observed proteome changes were then statistically evaluated and functionally annotated. The protein expression data were used for constructing a protein interaction network. Sub-networks were extracted that comprised proteins related to stress response, glycolysis, and TCA cycles. The observed changes in protein abundances for zebrafish deficient in vitamins E and C suggest a metabolic switch from glycolysis to glutaminolysis as a means of alternative energy production.

Materials and Methods

Chemiscals

Dithiolthreitol (DTT) and iodoacetamide were purchased from Bio-Rad (Hercules, CA). MS-grade water, acetonitrile, and formic acid (FA, 99%) were obtained from EMD Chemicals (Gibbstown, NJ). Sequencing-grade trypsin, resuspension buffer, and protease MAX solution were purchased from Promega (Madison, WI). [Glu^1]-fibrinopeptide B ([Glu^1]-Fib) and *Saccharomyces cerevisiae* enolase digest were obtained from Waters (Milford, MA). Phosphate buffer saline (PBS) solution was purchased from Fisher (Fair Lawn, NJ).

Fish Husbandry

Housing of the wild-type tropical 5D strain zebrafish was carried out in the Sinnhuber Aquatic Research Laboratory at Oregon State University, Corvallis, Oregon. The study was performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC).[80]

Feeding Experiment

The experiment was designed to allow analyzing changes to the proteome upon vitamin C supplementation of fish that were vitamin-E-adequate or -deficient. To do that, we fed fish with predefined diets as previously described.[11] In brief, at 42 days of age, 40 fish were divided in two groups and fed diets low in vitamin C (C−, 50 mg AA/kg diet, added as Stay-C, DSM

Nutritional Products, Parsippany, NJ) with vitamin E at sufficient levels (E+, 178 μmol RRR-α-tocopherol/kg diet) or deficient levels (E−, 22 μmol RRR-α-tocopherol/kg diet). Thus, two vitamin groups were created: E+C− and E−C− with 20 fish in each group. Induction of vitamin C deficiency continued for 56 days, after which half of the fish population in each group was harvested. The diet of remaining fish was altered to have a high amount of vitamin C (C+, 350 mg AA/kg diet, added as Stay-C), thus creating another two vitamin groups: E+C+ and E−C+. After an additional 21 days, the fish were harvested. Thus, four groups with 10 fish in each were created, each having different vitamins levels that varied with supplementation: E+C+, E−C+, E+C−, and E−C− (Appendix Figure A.1). Vitamin E and C concentrations have been published.[11]

Protein Extraction and Digestion

Each sample consisted of one whole flash-frozen fish. The whole flash-frozen fish was individually ground under liquid nitrogen using a mortar and pestle. The pipetting scheme used for preparing the protein extracts and digests for each sample is provided in the Appendix Table A.2. While still dry, the fish powder was transferred to a microcentrifuge tube and weighed (Appendix Table A.2, column 2). A PBS solution with ProteaseMAX (0.04%) was added to the microtubes for the extraction of proteins; assuming the density as 1 mg/mL, its volume was adjusted to six times the fish weight (Appendix Table A.2, column 5). The tubes containing the fish powder suspensions were submitted to three freeze−thaw cycles that included flash-freezing in liquid nitrogen for 2 min, thawing for 5 min, and sonicating for 15 min. After lysis, tubes were centrifuged at 4 °C for 10 min at 15 000 relative centrifugal force (rcf), and the supernatants were transferred to new microcentrifuge tubes (1.3 mL) while solid residue was discarded. The sample preparation workflow is shown in Appendix Figure A.3.

The concentration of proteins in each sample was determined by a photometric (Bradford) assay (Appendix Table A.2, column 6). Proteins in each sample were digested according to the manufacturer's protocol (Promega). The pipetting scheme for preparing the digest of each sample is outlined in Appendix Table A.2, columns 7−9. In brief, a volume of solution containing 50 μg of proteins was taken to 93.5 μL with a freshly prepared 50 mM NH_4HCO_3. Next, 1 μL of 0.5 M DTT (in Millipore water) was added to each vial, and solutions were incubated at 56 °C for 20 min. Then, 2.7 μL of 0.55 M

iodoacetamide (in 50 mM NH_4HCO_3) was added, and solutions were incubated at room temperature in the dark for 15 min. For protein digestion, 1 μL of 1% ProteaseMAX Surfactant in (in 50 mM NH_4HCO_3) and 1.8 μL of trypsin (1 μg/μL in 50 mM acetic acid) were added. Solutions were incubated at 37 °C for 3 h. The final volume of each solution of digested proteins was 100 μL. Solutions were centrifuged at 12 000 rcf for 10 s, and trifluoroacetic acid (TFA) was added at a final concentration of 0.5%. After snap-freezing in liquid nitrogen, samples were stored at −20 °C until analyzed. At the end, we conducted LC−IM−MS^E analyses of 8, 10, 10, and 9 samples of groups E+C+, E−C+, E+C−, and E−C−, respectively (Appendix Table A.2).

Liquid Chromatography−Mass Spectrometry

The analysis of all samples was performed using a Synapt G2 hybrid quadrupole time-of-flight mass spectrometer (Waters, Milford, MA) controlled by MassLynx 4.2 software (Waters). The sample peptide solution (9 μL) was mixed with 1 μL of internal standard (*Saccharomyces cerevisiae* enolase digest, 1 pmol/μL). With 1 μL injection volume the amount of the internal standard per injection on column was 100 fmol. Peptides were separated with a nanoAcquity Ultra Performance LC system (Waters, Milford, MA) equipped with 100 μm × 100 mm BEH130 C18 column with a particle size of 1.7 μm (Waters, Milford, MA). The mobile phase A consisted of 0.1% formic acid in water, and B consisted of 0.1% formic acid in acetonitrile (ACN). Each sample was first retained on a trapping column and then washed using 99.5% A for 3 min at a flow rate 5 μL/min. Peptides were separated using a 120 min gradient (3−40% B for 90 min, 40−90% for 2

min, 90% B for 1 min, 90−3% B for 2 min, and 3% B for 25 min) and then electrosprayed into the mass spectrometer, fitted with a nanoSpray source, at a flow rate of 400 nL/min. External calibration of the TOF analyzer was performed using NaI solution over the range of m/z 50 to 2000. The instrument operated in positive V-mode over the calibration range. Mass spectra were acquired in the MS^E mode alternating between a low energy scan (6 eV) to acquire peptide precursor data and a high-energy scan (ramping from 27 to 50 eV) to acquire fragmentation data. The capillary voltage was 2.5 kV, and the source temperature was 40 °C. Scan time was 1.25 s. The instrument settings are listed in Appendix Table A.4. An auxiliary pump delivered a [Glu^1]-Fib solution as an external calibrant (lock-mass) with a concentration of 500 fmol/μL at a rate 0.2 μL/min. For lock mass acquisition, a low-energy scan was acquired for 0.75 s every 60 s throughout a run.

Peak Detection

Elevated energy mass spectra were extracted, charge-statedeconvoluted, deisotoped, and lock-mass-corrected with [Glu^1]-Fib (MH+ m/z 785.8426). All ion mobility-MSE samples were analyzed using IdentityE by ProteinLynx Global Server (PLGS) version 2.5 (Waters Corporation, Milford, MA). The following processing parameters and their respective settings were used: chromatographic peak width and MS TOF resolution, automatic; low-energy threshold, 100 counts; elevated energy threshold, 10 counts; intensity threshold, 750 counts.

Protein Database Search

Waters' Identity^{E} was configured to search *Danio rerio* protein database (26 812 entries) with the digestion enzyme trypsin. The database was modified by adding the yeast enolase sequence (ENO1_YEAST, accession number P00924, Uniprot) that was spiked in as internal standard to facilitate labelfree quantification. IdentityE searched the protein database with a fragment ion mass tolerance of 0.025 Da and a parent ion tolerance of 0.0100 Da. The iodoacetamide derivative of cysteine and the oxidation of methionine were specified as variable modifications of amino acids in Identity^{E}.

Protein Identification, Quantification, and Validation

Scaffold (version 3.3.1, Proteome Software, Portland, OR) was used to validate MSE-based peptide and protein identifications. Peptide probabilities were assigned by the Peptide Prophet algorithm.[81] Protein probabilities were assigned by the Protein Prophet algorithm.[82] Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. The peptide and protein probabilities were adjusted to allow the best quantification of proteins and keeping the false identification rates at low levels. For each sample, the numbers of peptides assigned with high confidence are compiled in Supplementary Material Table S1. In addition, Supplementary Material Table S2 compiles proteins identified, unique peptides assigned per protein, and protein sequence coverage. This information was directly obtained from Scaffold.

The average of intensity of the three most intense peptides of each protein was

normalized to that of the internal standard with an assigned amount of 100 fmol. Proteins could not be quantified if they ambiguously shared peptides that cannot be uniquely attributed to one protein. Ultimately, the information about group name, sample name, protein ID, and quantitative values for proteins was acquired from Scaffold and exported to spreadsheets. The data were then imported into Perseus (http://www.perseus-framework.org/) and statistically analyzed. Perseus was developed by the Max Planck Institute of Biochemistry, Martinsried, Germany and designed to perform all downstream bioinformatics and statistics on proteomics output tables.

Western Blot

Three samples were randomly selected from each group. Samples were prepared for Western Blot analyses by diluting with sample buffer (5% mercaptoethanol in Laemmli buffer, Bio-Rad, Hercules, CA) to a concentration of 1 mg/mL of protein. For SDS-PAGE analysis, 20 μg (20 μL) of proteins was loaded into a pre-casted 10% gel (Bio-Rad). After electrotransfer to a nitrocellulose (NC) membrane (Bio-Rad), the latter was blocked overnight in 5% nonfat milk in TBS-T (10 mM Tris, pH 8, 150 mM NaCl, 0.05% Tween). The NC paper was blotted for 1 h with antibodies (Thermo, Rockford, IL) against pyruvate kinase M2 (Pkm2), glutamate dehydrogenase (Glud1), and glial fibrillary acidic protein (Gfap). A horseradish-peroxidase-labeled anti-goat and anti-rabbit IgG (Bio-Rad) and an enhanced chemoluminescence kit (SuperSignal West Pico Chemiluminescent Substrate, Thermo) were used to detect signals on a film (Kodak, Rochester, NY). The molecular mass of a protein was estimated using a protein

molecular standard (Bio-Rad).

Network Construction

To construct the vitamin-C-deficiency network, protein IDs were uploaded to STRING, Search Tool for the Retrieval of Interacting Genes (www.string-db.org).[6] The data found on protein interactions and the scores information were extracted. The protein network was constructed based on protein interactions using Cytoscape 2.8.2 (The Cytoscape Consortium). The latter is an open-source platform for complex network analysis and visualization.[7]

Results and Discussions

Overall Design and Strategy Applied for the Label-Free Quantitative Proteomic Study

The goal of this study was to investigate the consequences of vitamin E and C deficiency on the zebrafish proteome. The analysis and validation of quantitative bottom-up shotgun proteomics data sets depend critically on the optimal combination of data processing software tools because each commercially available software package can perform only a few steps of the data analysis pipeline (Figure 3.1). Here we implemented MassLynx and PLGS for peak extraction, alignment, and database search; Scaffold for spectra validation and protein quantitation; Perseus for statistical analysis; and STRING and Cytoscape for protein interactions analysis and network construction, respectively.

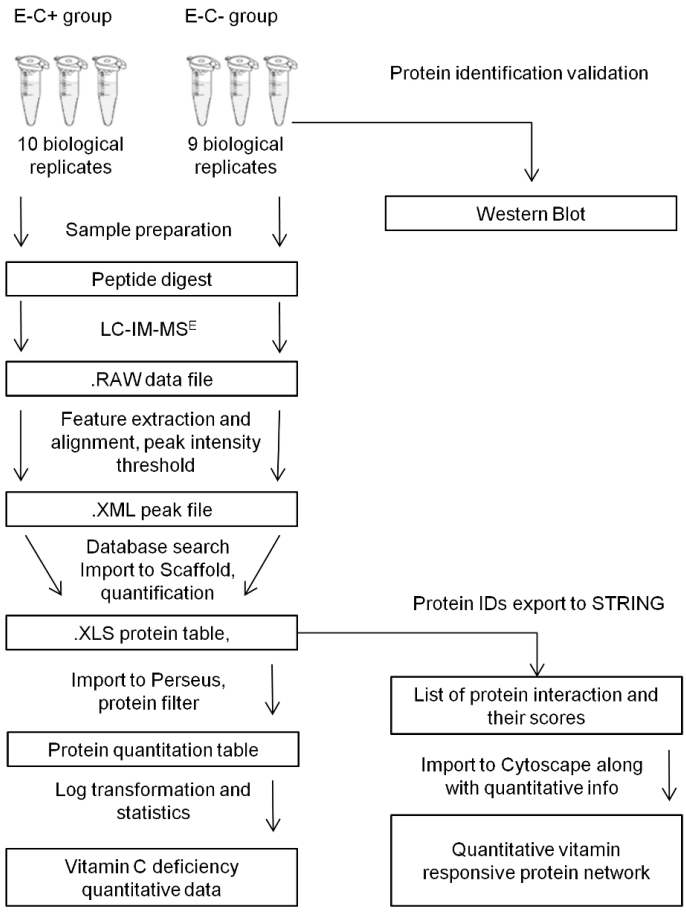


Figure 3.1 Data analysis workflow for label-free quantitative proteomics using LC−IMS−MS^E acquisitions.

Effects on Body Weight and Growth Rate

The experimental design of the feeding study afforded that zebrafish of the E+C− and E−C− groups were harvested after 98 days and the E−C+ and E+C+ groups were sacrificed after 119 days. We therefore expected that the different groups would result in different median fish body weights. Indeed, as shown in the Appendix Figure A.5, the median fish weights were lower for E+C− and E−C−groups compared with E−C+ and

E+C+ groups. The ANOVA test between all four groups did not show significant results, but the median body weights were significantly different between the E+C+ and E−C− groups (*p* value = 0.006, *t* test). Also, not totally unexpected, the growth rate (median body mass/age) of fish adequate in both vitamins (E+C+ group) was approximately three times higher compared with the group deficient in both vitamins (E−C− group). Vitamin C deficiency seemed to have a more drastic effect on the growth rate compared with inadequate vitamin E levels (Appendix Table A.6).

Optimization of Sample Preparation and Analytical Workflow for the Comparative "Bottom-up" Proteomics Study

To account for differences in fish body weights, we normalized the total protein concentration, which resulted in solutions with equal total concentration of proteins independent of fish weights, age, and vitamin supplementation. The total protein concentration in each sample was verified to confirm the validity of the sample preparation technique and for further use. The median total concentration of proteins for all samples was 2.382 mg/mL with first quartile 2.258 mg/mL and third quartile 2.478 mg/mL. As was expected from the sample preparation that accounted for the differences in fish weights, the total concentration of proteins was distributed in a narrow range (Appendix Figure A.7). After lysis, insoluble pellets that remained in the microcentrifuge tubes were discarded. The supernatant was used for protein digestion.

Because of the complexity of the whole fish protein extracts, we optimized the LC conditions to increase the number of peptides detected.[83] For optimization, a protein

extract was run using different LC gradients, and the number of eluted peptides was compared. By changing the mobile phase composition over time, the gradient was optimized after each run to make the elution of peptides even across a chromatogram and to reduce the coelution of peptides. The gradients were named as “Gradient 1”, “Gradient 2”, “Gradient 3”, and “Gradient 4”. The gradient profiles and respective chromatograms are presented in the Appendix Figure A.8. Using Venn diagrams, a comparison of the number of eluted peptides for each gradient is presented in the Appendix Figure A.9. “Gradient 4” resulted in the highest number of peptides detected and was used for all following LC−IM−MS^E analyses.

Data Analysis Strategy Used for Obtaining Protein Abundance Estimates

Label-free approaches are liable to technical variability, such as LC retention time drift, nanospray instability, and sample matrix effects. Therefore, to overcome these caveats, label-free approaches depend on the use of an internal standard and the application of proper mathematical tools for data processing. Log transformation of data is commonly used to make the distribution of protein abundances more Gaussian, with subsequent application of parametric statistical tests, which have higher statistical power after data transformation.[84]

In this study, a digest of enolase from *Saccharomyces cerevisiae* was used as an internal standard. The same volume of the internal standard was added to each sample, making its concentration equal across samples. Data were acquired, deisotoped, peak-aligned, and exported to Scaffold for quantitative analysis. The quantification of proteins was done by

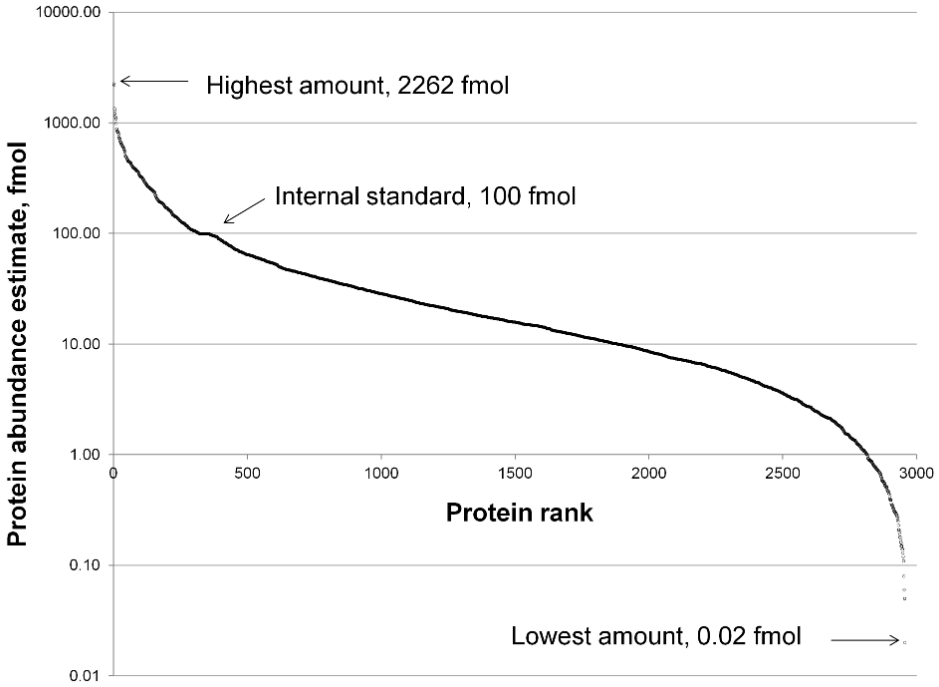


Figure 3.2 Dynamic range of the method based on protein abundance estimates.

normalization of the protein abundances to the internal standard. For the total of 2956 quantitative values obtained, a graph of protein amounts of the identified proteins from all samples against protein ranks was created (Figure 3.2). The value for the detection limit was extracted as the lowest amount of a protein that was calculated by Scaffold. The detection limit of this method was as low as 20 amol (Heat shock protein 8, sample E+C+#5), while the highest detected concentration was 2.3 pmol (Myosin light chain, sample E−C+#8). The dynamic range of estimates of protein amounts was calculated to be about five orders of magnitude.

Next, data were exported to Perseus and $\log_2$-transformed. To estimate the biological

variation between fish samples, we calculated the coefficient of determination, R^2, based on the protein estimates between biological replicates. As an example, Appendix Figure A.10 shows the correlation plot between the biological replicates E−C+ #9 and E−C+ #8 resulting in a R^2 value of 0.654. Additional correlation plots for the E−C+ group are shown in Appendix Figure A.11. As expected, variation between samples was observed. The correlation of protein amounts between different fish (within one feeding group) was found to be stronger in samples containing higher numbers of proteins.

Proteome Changes Linked to Vitamin C Deficiency

To investigate the proteome changes caused by vitamin C deficiency in vitamin-E-adequate fish, we evaluated the fold changes of the protein estimates obtained for the E+C+ versus the E+C− group. For quantitative analysis, peptide identifications were accepted if they could be established at a >95% probability. Protein identifications were accepted if they could be established at >99.0% probability and contained at least two identified peptides. The quantitative values calculated for the comparison of groups C+ versus C− (both sufficient in vitamin E) were based on 8134 spectra with peptide FDR 0.1% and 203 protein with protein FDR 2.0%. Next, internal standard and contaminants were removed and quantitation filter was applied. All proteins that were not detected at least two times across a group were filtered out. This conservative approach resulted in 119 proteins. Among them, 61% were up-regulated in vitamin-C-deficient fish (72 of 119). A change in concentration of twofold or more was observed for 28% of the proteins in this group.

The same quantitative analysis was done to investigate the proteome changes caused by vitamin C deficiency in vitamin-E-deficient fish. For quantitative analysis, peptide identifications were accepted if they could be established at a >50% probability. Protein identifications were accepted if they could be established at >99.9% probability and contained at least two identified peptides. The quantitative values calculated for the comparison of the C+ versus C− group (both deficient in vitamin E) were based on 11 308 spectra with peptide FDR 0.2% and 221 protein with protein FDR 2.3%. Again, internal standard and contaminants were removed and quantitation filter was applied. This resulted in a list of 112 proteins (Supplementary Material Table S2). Among them, 52% were up-regulated in vitamin-C-deficient fish (58 of 112). A change in concentration of two-fold or more was observed for 29% of the proteins.

Figure 3A shows a volcano plot for the distribution of *p* values versus fold-changes calculated for the proteins that were assigned for the comparison of the E+C+ versus E+C− groups (E+C+/E+C− transition). A negative logarithm of the p values and the logarithm of the ratio between amount of proteins from E+C+ and E+C− groups are displayed on the y and x axes, respectively. Dashed lines show a cutoff of two-fold change and a *p*-value threshold of 0.05 to define differential regulation of proteins between groups. Four proteins, namely, type-II cytokeratin, alpha tropomyosin, myosin heavy polypeptide 6, and fatty acid binding protein 6, showed a significant change upon vitamin C deficiency in vitamin-E-adequate fish, and all were upregulated. Regulation of alpha tropomyosin by α-tocopherol aligns well with previously published data.[85]

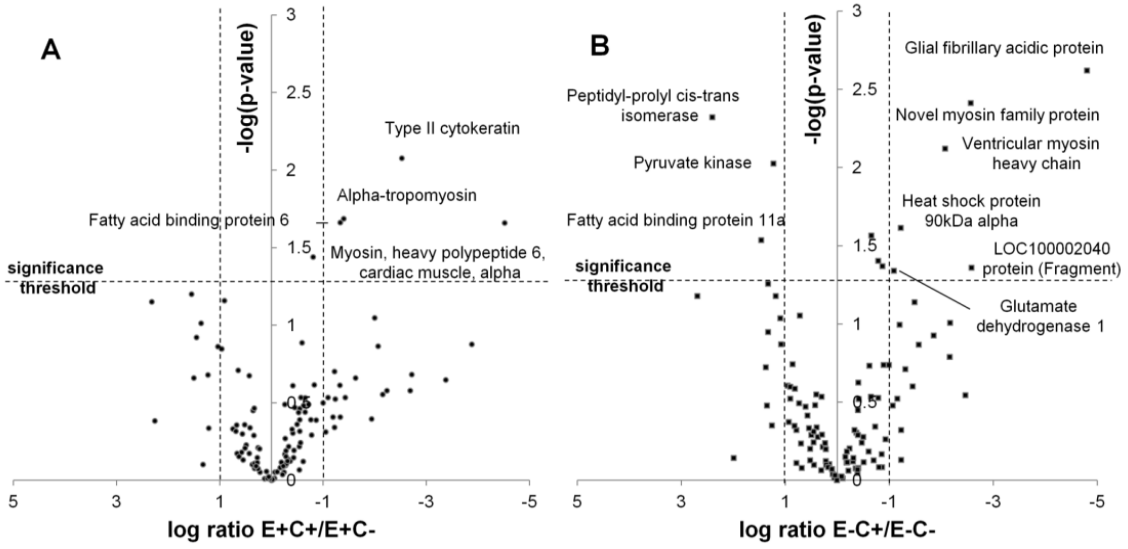


Figure 3.3 Volcano plot representation of protein abundance changes caused by vitamin C deficiency.

In the Figure 3B, the volcano plot is shown for evaluating differential regulation of proteins in group E−C+ versus group E−C− (E−C+/E−C− transition). We found that three proteins were down-regulated (pyruvate kinase M2b, fatty acid binding protein 11, and peptidyl-prolyl cis−trans isomerase) and six up-regulated (glial fibrillary acidic protein, ventricular myosin heavy chain, glutamate dehydrogenase 1, heat shock protein 90 kDa alpha, LOC100002040 protein, and novel myosin family protein CH211-158M24.10-001) as a consequence of vitamin C deficiency in vitamin-E-deficient fish.

A side-by-side comparison of the two volcano plots reveals the permissive role of vitamin E in the effect of vitamin C deficiency on the fish. In the vitamin-E-deficient fish the additional lack of vitamin C causes changes in expression levels of proteins responsible for energy metabolism, namely, pyruvate kinase and glutamate

dehydrogenase, along with other marker proteins. To obtain functional insight into proteins and pathways that are affected by vitamin C deficiency, we functionally annotated the combined proteomic data sets of the E−C− and E−C+ groups using the Gene Ontology (GO) classification (Supplementary Material Table S3).

Responsive Protein Interaction Network of Vitamin C Deficiency

A protein interaction network was constructed to provide a comprehensive visualization of the proteins identified, their abundance estimates, as well as differential regulation caused by vitamin C deficiency in fish low on vitamin E (Appendix Figure A.12). In the network, each node represents a protein found in either group E−C+ or E−C−. The change in protein amounts between vitamin-C-adequate and –deficient fish is color coded: green nodes represent proteins that increased their amounts, red represent those that decreased, and gray nodes show proteins that were not quantified. The size of a node represents an average of the protein abundance estimates in the E−C− group. The width of a line connecting proteins represents the strength of proteins interaction, as extracted from STRING. Figure 3.4 provides a closer look at sub-networks that encompassed proteins that were differentially regulated in response to vitamin E and C deficiency. Upregulated proteins were members of the sub-networks representing glutaminolysis, TCA cycle, and stress response.

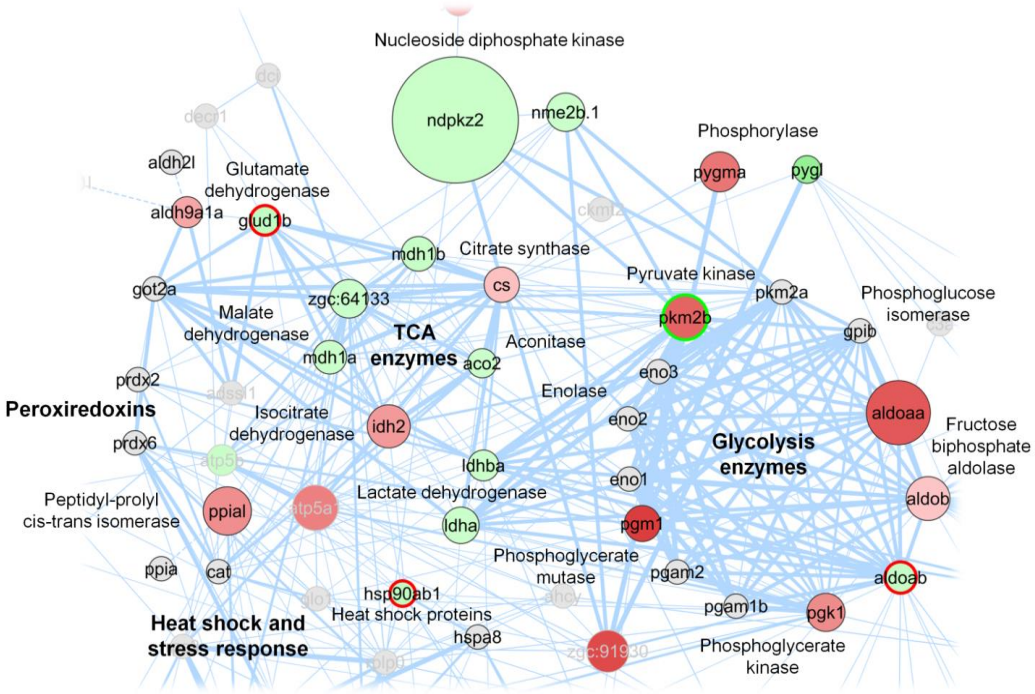


Figure 3.4 Protein-protein interaction network visualization of subnetworks that contain proteins that showed changes in expression levels upon vitamin C deficiency in the vitamin-E-deficient state.

Biological Inferences of Altered Protein Expressions in Vitamin-E- and -C-Deficient Zebrafish

The interaction network highlights the modulation of several proteins involved in two interconnected metabolic pathways, the glycolysis and TCA cycle (Figure 3.5). Proteins are shown as upregulated (green arrow up), downregulated (red arrow down), unquantified (gray arrow), and not detected (no arrow) with their respective numerical fold changes (Supplementary Material Table S2). Proteins that showed a statistically significant fold change are marked with an asterisk.

Upon vitamin C deficiency in groups low on vitamin E, glycolytic enzymes were

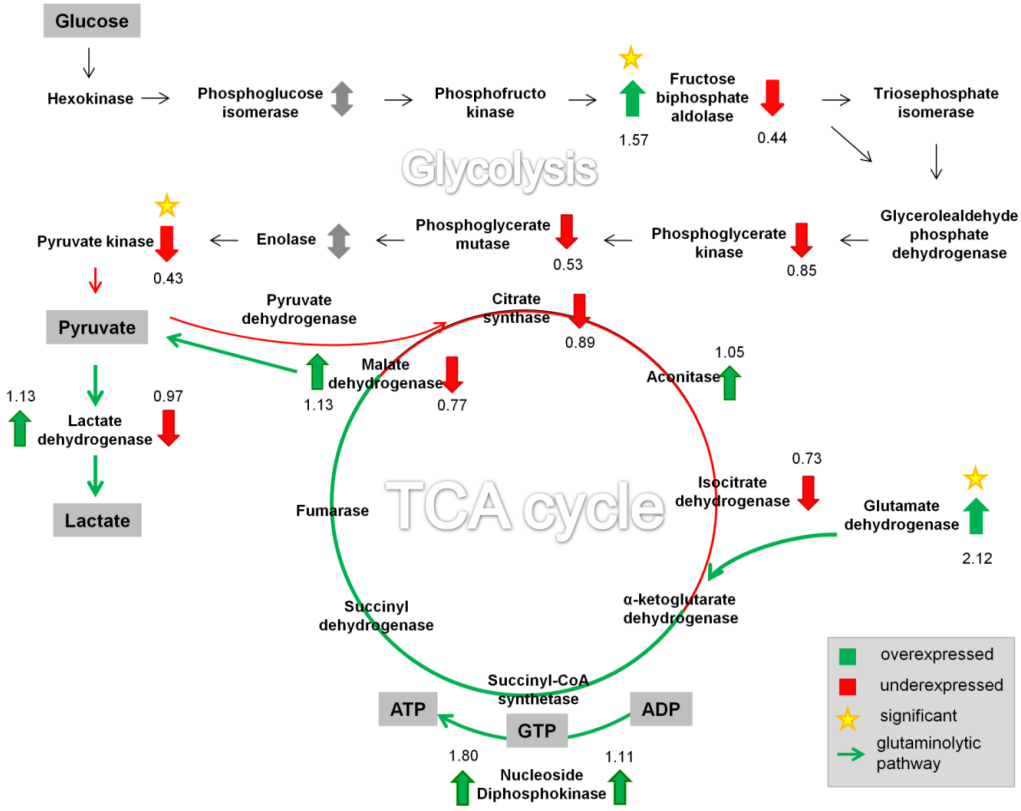


Figure 3.5 Schematic of the metabolic pathways that were associated with proteins that showed changes in abundance estimates as a consequence of vitamin C deficiency.

downregulated. The last step in the glycolysis pathway involves a transfer of the phosphate group from phosphoenolpyruvate to ADP, producing a molecule of pyruvate and a molecule of ATP. This step is catalyzed by pyruvate kinase (PK), which was significantly downregulated in vitamin-E- and -C-deficient fish (p value = 0.009, t test for logtransformed data). In humans and other mammals, four isoforms of pyruvate kinase are expressed: the L and R isoforms are found in liver and red blood cells, the M1 isoform is expressed in most adult tissues, and the M2 isoform, a splice variant of M1, is the predominant form found in proliferating cells and tumor cells.[86,87] In the current

study, pyruvate kinase M2 was detected but not M1 nor L/R. Western Blot analysis confirmed the presence and change of pyruvate kinase M2 (data not shown). Under the conditions of suppressed glycolysis there will be less production of energy (ATP) and reducing power (NADH), the main products of the cycle. We may therefore speculate that in vitamin-E- and -C-deficient fish compensatory mechanisms for the production of energy are activated. Kirkwood et al. studied vitamin-C-deficient zebrafish at the metabolome level and reported activation of the purine nucleotide cycle as possible compensatory mechanism to satisfy intracellular ATP requirements.[11]

After conversion of glutamine to glutamate, glutamate dehydrogenase catalyzes the production of α-ketoglutarate, which subsequently enters the TCA cycle. The conversion of glutamine to lactate is commonly referred as glutaminolysis. [88,89] Our quantitative proteomic study indicates that several proteins that play key roles in glutaminolysis are elevated in vitamin-E- and -C-deficient fish (E−C− group). Our data sets show a significant increase in levels of glutamate dehydrogenase (Glud1b) under vitamin E and C deficiency compared with E deficiency alone (p value = 0.046, t test for log-transformed data). The change in this protein was confirmed with Western Blot (data not shown). Malate dehydrogenase and lactate dehydrogenase, enzymes involved in glutaminolysis, were increased as well. The elevation of several key enzymes involved in glutaminolysis supports the notion that vitamin E and C deficiency promotes a metabolic phenotype in which glutaminolytic pathways are activated for alternative energy production (ATP) and reducing power (NADPH) in cells under conditions of suppressed

glycolysis.

In addition, our quantitative proteomics screens revealed that Glial fibrillary acidic protein (Gfap) was significantly elevated (27 times) in fish deficient in both vitamins (*p* value = 0.002, *t* test for log transformed data), while there was no significant difference in Gfap expression between E+C+ and E−C+ groups consistent with the notion that vitamin E seems to exhibit a permissive role in governing the proteome biology of adult zebrafish. The adverse effect of deficiency of both vitamins has been previously reported.2 The identification of Gfap was based on a conservative probability for protein identification of at least 85% and at least two unique peptides with peptide identification probability of at least 50%. The similarity between zebrafish and human Gfap (GFAP_HUMAN, entry P14136 in Uniprot) is 65%, as calculated by Blast search. Despite the good homology of the human and zebrafish protein, the human antibody used in the Western blot analyses showed nonspecific binding and Western blots were not conclusive. Gfap is a recognized marker of neurologic injury and trauma and is released during astrogliosis in higher vertebrates.[90–92]

Conclusions

We applied a label-free comparative proteomics approach for examining the systems-wide consequences of insufficient levels of two essential micronutrients, vitamins E and C, in adult zebrafish. The experimental design adopted allowed the study of vitamin C deficiency at the whole organism level. Using a label-free quantitative proteomic

workflow, it was possible to assess for the first time changes in the vertebrate proteome upon vitamin C deficiency. The label-free quantitative bottom-up strategy with LC−ion mobility−MS^E is a powerful approach to determine protein abundance estimates in zebrafish. We report sensitivity of the method as low as 20 amol and a dynamic range of five orders of magnitude for protein level estimates. Our results indicate the modulation of expression of proteins involved in stress response and metabolic pathways associated with energy production. Our findings suggest that severe vitamin C deficiency potentiated by vitamin E deficiency causes the suppression of glycolysis and the activation of glutaminolysis as an alternative way to fulfill cellular energy requirements. Glial fibrillary acidic protein (Gfap) showed a significant overexpression in fish low in both vitamins C and E, proving that severe vitamin C deficiency in combination with low vitamin E status causes injury of the central nervous system.

Acknowledgements

We thank Dr. Cristobal Miranda for assistance in sample preparation. This study was partially made possible by grants from the National Institutes of Health, P30ES000210, S10RR025628, and R01HD062109.

CHAPTER 4 Application of an ATP-affinity Tag Complements a Full Proteomics Experiment in Dissecting Regulatory Pathways Linked to Metastasis in Osteosarcoma

Ievgen Motorykin,[1] Milan Milovancev,[2] Shay Bracha[2] and Claudia S. Maier*,[1]

[1]Department of Chemistry, Oregon State University, 153 Gilbert Hall, Corvallis, Oregon 97331, United States

[2]College of Veterinary Medicine, Magruder Hall, Corvallis, Oregon, 97331, United States

This manuscript will be submitted to Journal of Proteome Research

*Corresponding author, E-mail: claudia.maier@oregonstate.edu

Abstract

Osteosarcoma (OS) is the most common primary malignant bone tumor in humans and dogs with a high rate of pulmonary metastasis, which significantly reduces the chance of survival. In this study we describe a multi-tiered proteomics workflow aimed to detect differences in the proteome and the ATPome between two canine OS cells that were originally obtained from an orthotropic primary OS (POS) and metastatic cells (HMPOS).The goal of this study was to establish proteome difference that typify a highly metastatic phenotype. We complemented a MS-driven quantitative proteomics workflow with a chemical proteomics strategy that uses an ATP-affinity tag to target ATP-binding proteins including protein kinases (i.e. the ATPome). This multi-tiered approached allowed us to profile > 1100 proteins including 228 ATP-binding proteins and protein kinases.

Differences in protein levels were observed for proteins associated with cytoskeleton, cell motility and adhesion, most notably for RHOA, AKT1, several cyclin dependent kinases and the check-point kinase, CHK2. We conclude that in highly metastasizing osteosarcoma cells (HMPOS), RHOA is linked to the disruption of the cytoskeleton and the adherens junction complex which in turns weakens intercellular adhesion and increases cell motility. Changes in the cell cycle pathway may indicate an inability of metastatic cells to activate a proper defense response to abnormal behavior. The current findings may contribute to the identification of metastatic biomarkers that can be used for screening and development of novel, therapeutic drugs for OS.

Introduction

Osteosarcoma (OS) is the most common primary malignant bone tumor in humans, most frequently occurring in tall human adolescents.[93] A specific biologic behavior exhibited by OS is a high propensity of metastasis. Twenty percent of humans may develop pulmonary metastases despite treatment[94], significantly reducing overall survival. The 5 year survival rate for patients with localized tumors ranges between 60-80%, in comparison to 15-30% for patients with metastasis.[95]

Metastasis, the biologic process of cancer cells spreading from a primary tumor to a distant healthy organ or tissue, is not fully understood. Tumor metastasis is a highly complex process that starts with loss of adhesion of cancer cells to a tumor[96] and leads to the invasion and colonization of cancerous cells in a new tissue. Development of tools that can effectively, and in time, diagnose and/or prevent metastasis is necessary to increase the survival rate for patients with OS. An example of a such tool is the discovery of a protein biomarker, or a network of proteins, associated with a particular cancer's potential to metastasize.[97] For that, a comparative proteomics study needs to be conducted to elucidate differences in protein expression in cancer cells with variable propensities for metastasis.

As access to the human samples is often limited, researchers frequently use animal models to study OS and the process of metastasis. *Canis Familiaris* (the domestic dog) has established species similarities to humans and possesses unique characteristics that make it particularly relevant for the study of OS and metastasis[98,99]. For example, OS is

also most common primary malignant bone tumor in dogs, especially in large and giant breeds. Metastasis is common in dogs, with 90% of dogs dying from metastases to the lungs.[100] Additional features of dogs that make them uniquely suited for studying the metastatic potential of osteosarcoma include: human and canine OS are indistinguishable on the basis of global gene expression signatures,[101] OS occurs spontaneously in dogs and humans, and both species share the same environment; dogs are less inbred than rodents that are widely used as a disease model, and therefore more completely recapitulate the complex biological environment in which OS develops and spreads; human and canine OS are both rare but it occurs more often and metastasize faster in dogs than in humans, providing a good supply of samples for studying[99]. These similarities are the reasons canines are frequently used to study OS, providing a realistic model that represents the human disease.[98,99]

Multiple comparative proteomics studies compared various cancer cells with different metastatic potential: human gastric cells,[102] human hepatocellular carcinoma cell strains,[103] as well as comparison of primary tumor and metastatic cells of human renal cell carcinoma,[97] human colorectal cancer,[104] canine osteosarcoma[105] and human melanoma.[106,107] Common altered pathways are associated with oxidative phosphorylation, cAMP/Protein Kinase A (PKA) signaling, immune responses, cytoskeletal remodeling, focal adhesion,[105] cell motility and proliferation.[102] Suggested proteins biomarkers of metastasis are KIT,[107] pyruvate kinase M2, ubiquitin carboxy-terminal hydrolase L1, laminin receptor 67 kDa, thioredoxin and cytokeratin 19,[103]

Wilms tumor 1 (WT1), tumor protein p53 (TP53), cyclin-dependent kinase inhibitor 2A (CDKN2A), phosphatase and tensin homolog (PTEN), retinoblastoma 1 (RB1) tumor suppressors[108] and S100 proteins.[103,106]

Adenosine triphosphate (ATP), a nucleotide that is abundantly present in cells, is often referred to as "a unit of currency" in the energy transfer within cells.[109] ATP binding and hydrolysis are the driving processes in all living organisms.[110] This process is used to unfold proteins, transport molecules through membranes, or phosphorylate small molecules or proteins. Perhaps the most important class of ATP-binders is that of protein kinases, which transfer the gamma (γ) phosphate group from an ATP to a substrate. Protein kinases, which are involved in nearly all cell signaling cascades, transfer signals from a cell's membrane to its nucleus and are responsible for a wide range of diverse cellular functions. There are 7 major groups of protein kinases[111]: the tyrosine kinases (TK); the tyrosine kinase-like (TKL) group; the STE kinases, the homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; the caseine kinases 1 (CK1); the AGC group containing protein kinase A, G, and C families (PKA, PKC, PKG); the CAM, calcium/calmodulin-dependent protein kinases; and the CMGC group containing CDK, MAPK, GSK3, CLK families. The human kinome contains 518 protein kinases, of which 478 belong to the before mentioned groups.[111] Other protein kinases include atypical types with no sequence similarity to typical kinase, but with predicted enzymatic activity.

Rather than studying protein kinases individually, global approaches enable studying changes in ATP-dependent catalytic pathways that are altered in metastasizing cancer

cells, thereby allowing the discovery of possible drug-able targets and design of small molecule inhibitors.[110] Traditionally, kinase expression surveys were based on high specificity and sensitivity immunoassays[112], which are limited by the access to expensive high quality antibodies and were useful for profiling a small number of kinases in a low throughput experiment. Even though mass spectrometry based proteomics experiment can identify a large number of proteins in a single run, the detection of protein kinases is still a challenging task, mainly because they are low abundant in cells.[113]

Bartlett et al showed that most kinases have similar ATP binding domains,[114] which can be used to study interactions between protein kinases and ATP by utilizing probes that are modified on the γ-phosphate group of ATP.[4] Since then, many chemical affinity tags coupled to the γ-phosphate group of ATP have been synthesized and used for the enrichment of protein kinases. For instance, desthiobiotin-ATP and other similar probes were used for profiling of protein kinases and other ATP-binding proteins.[110,115–121] It has been previously observed that the desthiobiotin-ATP probe labels preferentially nucleotide binding proteins with binding sites that contain a lysine residue that is capable of directly interacting with the γ-phosphate group of the bound ATP probe.[4] At least one lysine residue is present in close proximity to the ATP binding site in virtually all protein kinases.[122] Because the reactivity of the probe is based on a labile acyl phosphate linkage, any lysine residue in close proximity has the potential to be labeled by the probe.

The proteome-wide, large scale profiling of pathways altered during a disease is facilitated by the use of chemical tags that can selectively enrich ATP binding proteins

and improve their detection in mass spectrometry-based proteome analyses.[116,123] In Hela S3 cells, 551 proteins were identified by Xiao et. al.[120] A total of 539 ATP-binding protein candidates, including 178 novel candidates, were described by Adachi et. al.[124] Also, 122 ATP-binding proteins were found in *Mycobacterium tuberculosis*[110] and 242 labeled sites were found in *Arabidopsis thaliana*.[121] These studies show the effectiveness of using a biotin-conjugated acyl ATP probe in conjunction with a biotin-streptavidin affinity capture strategy for large-scale sensitive profiling of the ATPome. While the desthiobiotin probe shows good selectivity toward a lysine residue, it has been reported that it will not be able to distinguish between lysine residue in the ATP-binding pocket of an enzyme or less specific binding sites outside the ATP binding site and, therefore, affinity tagging occurs relatively unspecifically.[124]

Large scale MS-driven proteomics in conjunction with an ATPome enrichment technique, usually utilize the data dependent acquisition (DDA) mode, in which 10-20 most abundant ions found in a survey scan are selected and subjected to fragmentation to acquire peptide fragment ion spectra.[2] Although, this untargeted "shotgun" approach can detect novel protein biomarker candidates, the sample complexity and variation in automated peak selection lead to compromised reproducibility and sensitivity for protein quantification.[119] This caveat leads to only partial overlap of the sets of proteins that are identified from similar samples,[125]which can be overcome at least in part by incorporation of technical and analytical replicates.

In this study we describe a chemical proteomics method aimed to detect differences in the

proteome and the ATP-binding proteome, the ATPome, between osteosarcoma cells that differ in their propensity to metastasize. The discovered differences in protein expressions and their quantitative values were then analyzed by diverse bioinformatics approaches that incorporate the use of multiple software packages and databases. Application of the chemical affinity tag yielded 85 protein kinases, showing good reproducibility and selectivity. Proteins that displayed significant level changes in the two cell lines indicate alterations in cell signaling and metabolic pathways that have been shown to regulate metastasis, including cytoskeletal regulation, cell adherens and motility, AKT1 signaling, MAPK and cell cycle pathways. The proposed biomarkers for metastatic osteosarcoma will need to be rigorously tested and validated in future studies.

Materials and Methods

Chemicals

Dithiolthreitol (DTT), iodoacetamide (IAA) and Laemmli buffer were acquired from Bio-Rad (Hercules, CA). MS-grade water, acetonitrile and formic acid (99%) were purchased from EMD Chemicals (Gibbstown, NJ). Sequencing-grade trypsin, resuspension buffer and protease MAX solution were from Promega (Madison, WI). PBS solution was purchased from Fisher (Fair Lawn, NJ). Trifluoroacetic acid (TFA) and Pierce Kinase Enrichment Kit were purchased from Thermo Fisher Scientific (Rockford, IL, product number 88310). The kit includes: ActivX Desthiobiotin-ATP Probe (product number 88311), Pierce IP Lysis Buffer (product number 87787), Reaction Buffer, Halt™

Protease/Phosphatase Inhibitor Cocktail, 100X (product number 78440), Zeba™ Spin Desalting Columns, 7K MWCO, 5mL (product number 89891), High Capacity Streptavidin Agarose Resin, 50% slurry (product number 20357), $MgCl_2$ and Urea. $MnCl_2$ powder was obtained from Sigma (St. Louis, MO) and used to make 1 M solution in MQ water. Tris was obtained from Mallinckrodt (Paris, KY) and used to make a 1 M solution of pH 8.0, adjusted with HCl.

Enrichment of ATP-binding proteins

The probe used for enriching ATP-binding proteins, including kinases, consists of the nucleotide, ATP, connected through a labile acyl-phosphate bond to desthiobiotin, a biotin analog. (Appendix Figure B.1).The probe covalently modifies the conserved lysine residue within the nucleotide binding site.[123] Desthiobiotin, the affinity tag, binds to biotin-binding proteins less tightly than biotin, which is beneficial for the release of the tagged peptides by biotin displacement, low pH (this study) or heat denaturation. The details of the protocol are presented below.

POS and HMPOS[126] cell suspensions (each in 15 mL tubes) were obtained from Drs. Bracha and Milovancev. The labeling kit was used according to the manufacture's protocol and applied to both cell lines. Briefly, 3.2×10^7 cells of each cell line were transferred to a new 15 mL tube and recovered from solution by centrifugation at 500×g for 5 min. Cells were resuspended in ice cold PBS and transferred to a 1.5 mL centrifuge tube. The suspensions were centrifuged again at 500×g for 2 minutes and PBS was removed. To each sample, 1 mL of IP Lysis Buffer containing Halt™ protease and

phosphatase inhibitor cocktail (1:100) was added and solutions incubated on ice for 10 minutes with periodic mixing. The cell lysates were cleared by centrifugation at 16,000×g at 4 ºC for 5 minutes and the supernatants (total lysates) were transferred to new microcentrifuge tubes.

Next, sample lysates were run through a Zeba gel-filtration column to remove endogenous ATP and to exchange buffer to the Reaction Buffer. Briefly, desalting columns were placed in 15 mL collection tube and centrifuged at 1,000×g for 2 minutes to remove storage solution. Then, 3 mL of Reaction Buffer was added and each column was centrifuged at 1,000×g for 2 minutes again to remove buffer. The last step was repeated two more times. Columns were placed in new separate collection tubes and 1 mL of total lysate was applied to the center of each resin bed. After centrifugation at 1,000×g for 2 minutes the solutions were collected and the columns were discarded.

Total concentration of proteins in each sample was determined by the Bradford assay. After this point, the sample preparation workflow splits into two parallel branches: (1) labeling with the ATP probe, affinity enrichment followed by protein digestion and (2) no labeling and protein digestion. The schematics of both workflows are shown in Appendix Figure B.2.

ATPome Profiling: ATP Probe Labeling, In-Solution Digestion, and Affinity Enrichment

To profile the ATPome subproteome, one milligram of proteins from each sample was transferred to a microcentrifuge tube and the volume was brought up to 500 µL with the Reaction Buffer. This step was done in triplicate, i.e. the sample preparation for each cell

line was done three times in parallel. To create a control – no desthiobiotin-ATP tag – group, 250 µg of proteins from each sample were pooled together and the volume of the solution was brought up to 500 µL with the Reaction Buffer. To achieve best results for labeling the ATP binding sites, 10 µL of 1 M $MnCl_2$ was added to each sample and solutions were incubated for 1 minute at room temperature. The desthiobiotin-ATP reagent was reconstituted in 10 µL of MS-grade water and this amount was added to each sample except control and solutions were incubated at room temperature for 10 min.

After this, 500 µL of 10M urea in IP Lysis Buffer was added to each sample. Proteins were reduced by adding 20 µL of 500 mM DTT and incubated at 65 ºC for 30 min. Subsequent alkylation of proteins was done by addition of 40 µL of 1 M iodoacetamide and incubation at room temperature for 30 min, protected from light.

Before enzymatic digestion, the sample buffer was changed to Digestion Buffer containing 2 M urea in 20 mM Tris, pH 8. The procedure was carried using a Zeba Spin Desalting Column as described above. For protein digestion, 20 µg of trypsin was reconstituted with 20 µL of trypsin re-suspension buffer and this amount was added to each sample, making final protein:trypsin ratio 50:1. Samples were incubated at 37 ºC with shaking for three hours.

Labeled peptides were captured by adding 50 µL of 50% High Capacity Streptavidin Agarose resin slurry to each digest and incubated at room temperature for one hour under constant mixing on a rotator. For subsequent steps, the mixtures were vortexed briefly after addition of a buffer (the details of each step are described below), centrifuged at

1,000×g for 10 min and supernatant discarded. To remove unbound peptides the resin was washed three times with 500 µL of IP Lysis Buffer, then four times with 500 µL of PBS and then four times with 500 µL of LCMS-grade water. The desthiobiotin-tagged peptides were eluted three times with 75 µL of Elution Buffer, containing 0.1% TFA in 50/50 acetonitrile/H_2O solution, and incubation for 3 minutes each time. For each sample, the eluates were pooled and samples were frozen before lyophilization. Sample were dried in a vacuum rotator and stored at -80 ºC. Lyophilized peptides were resuspended in 10 µL of 0.1% (v/v) formic acid prior to LC-MS/MS analysis.

Full proteome Profiling – In solution digestion without prior labeling and capture/release

Protein preparations from POS and HMPOS cells were digested without any labeling and enrichment to obtain full comprehensive proteome samples. For that, cells were lysed as described above and proteins digested as per manufacturer's protocol (Promega). In brief, a volume of solution containing 50 µg of proteins was brought to 43.5 µL with a freshly prepared 50 mM NH_4HCO_3 and 2 µL of 1% proteaseMAX Surfactant was added for better protein solubility. Proteins were reduced in 10 mM DTT for 20 min at 56 ºC and alkylated in 20 mM iodoacetamide for 15 min at room temperature in the dark. Next, 1 µL 1% ProteaseMAX Surfactant and 1 µL 1 µg/µL of trypsin were added and samples incubated at 37 ºC for 3 hours. Final volume of each reaction was 50 µL and protein:trypsin ratio 50:1. Trifluoroacetic acid (TFA) was added to the final concentration of 0.5% to quench the digestion. Samples were centrifuged at 16,000 rcf for 10 seconds

to remove possible particulates. . The samples were then frozen at -80 ºC and stored until analysis.

Liquid Chromatography - Mass Spectrometry

For peptide separation a Waters nanoAcquity UPLC system equipped with 300 µm x 100 mm Zorbax 300SB-C18 column (Agilent, Santa Clara, CA) with 5 µm particles size equilibrated at 37 ºC was used. A binary solvent system was used: solvent A was 0.1% formic acid in water and solvent B was acetonitrile containing 0.1% formic acid. Analytes were first retained on a Michrom peptide CapTrap column for 3 minutes using 99.5% A at the flow rate 5 µL/min. Peptide elution was achieved using a binary gradient with a flow rate of 4 µL/min : 3-10% B in 3 min, get to 10-30% B in 102 min, get to 30-90% B in 3 min, 90% B for 4 min, 90-3 % B in 1 min, 3% B for 16 min and 50% B in the end of the run. The total gradient time was 130 min, but the analysis run time was specified as 120 min.

LC-MS/MS analyses were conducted using a Thermo LTQ FT-ICR mass spectrometer equipped with an IonMax ion source operated at a spray voltage of 5 kV and a source temperature of 200 ºC. All data were acquired in the positive mode and conducting data dependent acquisitions. The mass-spectrometer acquired data by the FT-MS analyzer over the range 350-2000 m/z with resolution 100,000 (at m/z 400) and the top 5 precursor ions within isolation width 1 Da that exceeded the threshold of 500 counts were subjected to collision induced dissociation (CID) in the linear ion trap with normalized collision energy of 30% and activation time of 30 ms. Exclusion rule was applied (exclusion size

list 100, exclusion duration 45s) and charge state filtering was enabled (precursors with unknown charge state or a charge state of 1 were excluded).

Peak Detection and Protein Database Search

Tandem mass spectra were extracted, deisotoped and charge state deconvoluted by Xcalibur, version 2.0.5 (Thermo Scientific, Rockford, IL). Proteome Discoverer (version 1.4, Thermo Scientific, Rockford, IL) was used to analyze all MS data. Mascot search engine (version 2.3, Matrix Science, London, UK) was set to search *Canis Familiaris* (25,440 entries, released on 7.5.14, Uniprot) assuming digestion enzyme trypsin. The search engine was configured to allow two missed cleavages and have a precursor ion mass tolerance of 20 ppm, a fragment ion mass tolerance of 1.0 Da, relaxed false discovery rate (FDR) 0.05 and strict FDR 0.01. Carbamidomethylation of cysteine was set as static amino acid modification and oxidation of methionine was set as dynamic modification. For labeled peptides only, static desthiobiotin modification of lysine (delta mass +196.1212 Da) was specified. Only peptides with Mascot confidence identified as "high" were accepted. For labeled peptides the additional criteria was lysine residue modified by the desthiobiotin tag. The minimum number of identified peptides was two and one for unlabeled and labeled proteins, respectively. The workflow also quantified peptides by their respective peak areas. The abundance of a protein was then calculated as an average of its three most intense peptides.

Data Analysis and Representation

When analyzing data from proteomics experiments a common challenge is that there is not a single software package that can carry out the analysis of the data along all the steps. Therefore the assembling of software tools that are used for specific steps of the analysis workflow should be carefully made to prevent incompatibilities and to achieve the best performance.[9] For the current study we utilized multiple software platforms to analyze the data in a comprehensive manner. The full diagram of data analysis workflow is shown in Figure 4.1.

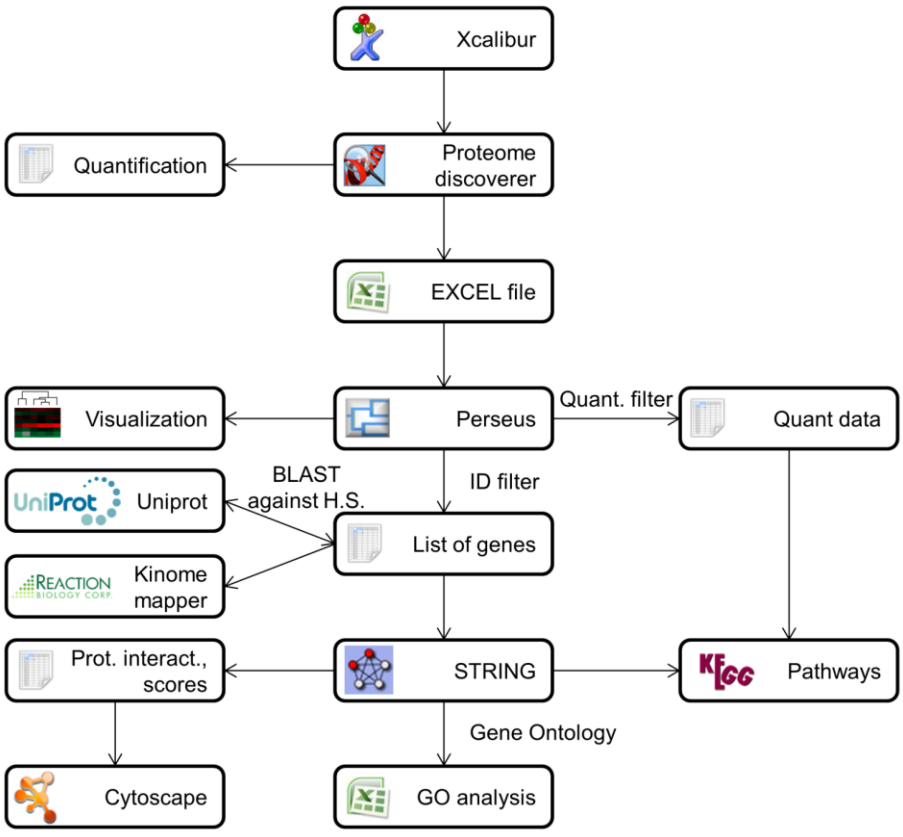


Figure 4.1 Diagram of the data analysis workflow

Data dependent acquisition of LC-MS/MS data was recorded with Xcalibur software, version 2.0.5 (Thermo Scientific). The resulting .MSF files were then imported to Proteome Discoverer v. 1.4 (Thermo Scientific) for peptide filter, and protein assignment and relative quantification. As a quantitative representative of protein abundance the sum of intensities of the protein's three most intense peptide peaks was used. In the next step the protein names and their abundances were exported as an Excel spreadsheet and imported to Perseus for statistical analysis and visualization.

Although the canine genome is fully sequenced, the protein database contains many entries with no assigned protein names, i.e. "uncharacterized proteins". Also, the gene interaction databases lack information for *Canis Familiaris* species. To overcome this problem it was assumed that canine proteins with the same amino acid sequence as those in humans are coded by the same genes and have the same functions in the cells. Proteins from the *Canis familiaris* database that did not have a proper name or gene name were BLAST searched against *Homo sapiens* using Uniprot database (www.uniprot.org) and the top human protein was chosen as a substitute for the missing canine protein.

The study design included three biological replicates for each experimental condition: POS versus HMPOS with enrichment and POS versus HMPOS without enrichment. As a control group was employed the mixture of unlabeled lysates from both cell lines containing equal mass of proteins that was subjected to the identical workflow that was used for labeled proteins including the capture/release step. Control group was created to assess unspecific binding of proteins to streptavidin agarose in the absence of

desthiobiotin label. All samples were injected in the mass-spectrometer three times making nine injections per group. Perseus software (v. 1.4.0.20, www.perseus-framework.org) was used to filter and statistically analyze the data. Protein identification criteria were as follow (a.k.a ID Filter): a protein has to be detected in at least one analytical replicate and in at least two biological replicates. To quantify changes in protein abundances between groups, an additional filter (a.k.a. Quant Filter) was applied: a protein has to be detected in at least one analytical replicate and in at least two biological replicates in each group. The t-test was used to compare the abundances of proteins between samples (all replicates from POS vs. all replicates from HMPOS, with and without enrichment) with p-value cut-off 0.05.

To construct the kinome map the ID list of the ATP enrichment experiment was then cross-matched with the list of kinases in the Reaction Biology Kinome Activity Mapper v. 2.4 (Reaction Biology Corporation, www.reactionbiology.com). Gene IDs from two experiments, FP and ATP, were uploaded to STRING v.9.2, Search Tool for the Retrieval of Interacting Genes (www.string-db.org)[127,128] with a reference organism *Homo sapiens*. The information about protein interactions (extracted from sources including PubMed, www.ncbi.nlm.nih.gov/pubmed), KEGG pathways (extracted from Kyoto Encyclopedia of Genes and Genomes, www.kegg.jp)[39] and gene ontology (GO, extracted from Uniprot, www.uniprot.org) was extracted as separate text files for the two experiments and imported to a spreadsheet. These data were then compared between enrichment and no enrichment experiments and between the two osteosarcoma cell lines. The data found on

gene interactions and their quantitative information were then used to visualize the gene network using Cytoscape 3.2.0 (The Cytoscape Consortium, www.cytoscape.org), an open source platform for network analysis and visualization[7]. In the network, proteins were grouped according to their binding abilities: protein binders, nucleotide binders, poly-RNA binders, all or non-binders. Finally, the KEGG pathways were combined with the observed quantitative changes in proteins to rationalize changes in cell signaling between cell lines.

Western Blots

Cell lysate samples from each group (POS an HMPOS) were compared in triplicate. In brief, protein solutions were diluted with a sample buffer (5% mercaptoethanol in Laemmli buffer, Bio-Rad, Hercules, CA) to the final concentration 1 mg/mL. For the SDS-PAGE separation, 20 μg (20μL) of proteins were loaded onto a two pre-casted 10% gel for testing against various proteins (Bio-Rad). The proteins were then electrotransferred to a nitrocellulose (NC) membrane (Bio-Rad) and the latter was blocked overnight in 5% non-fat milk in TBS-T (10 mM Tris, pH 8, 150 mM NaCl, 0.05% Tween). The first NC paper was blotted for 1 hr with a mixture of goat anti-AKT1 and rabbit anti-RHOA (Santa Cruz Biotechnology, Dallas, TX). The second NC paper was blotted with antibodies against rabbit ACTB and mouse TUBBA (Sigma Aldrich, St. Louis, MO). A mixture of a horseradish peroxidase-labeled anti-goat and anti-rabbit IgG (both 1:10000, Bio-Rad) and a mixture of a horseradish peroxidase-labeled anti-mouse and anti-rabbit IgG (both 1:10000, Bio-Rad) were used to label the primary antibodies on

the first and second NC papers, respectively. Enhanced chemoluminescence kit (SuperSignal West Pico Chemiluminescent Substrate, Thermo) was used to detect signals on a film (Kodak, Rochester, NY). The molecular mass of a protein was estimated using a protein molecular standard (Precision Plus Protein™ Dual Color Standards #161-0374, Bio-Rad).

Results and Discussions

Experimental Strategy

The experiment was designed to determine changes in metabolic pathways and cell signaling cascades between two osteosarcoma cell lines that differ in metastatic potential (Figure 4.2). Cells from both cell lines were subjected to two proteomics workflows: the full proteome assessment (FP) and the ATPome subproteome analysis. The results were

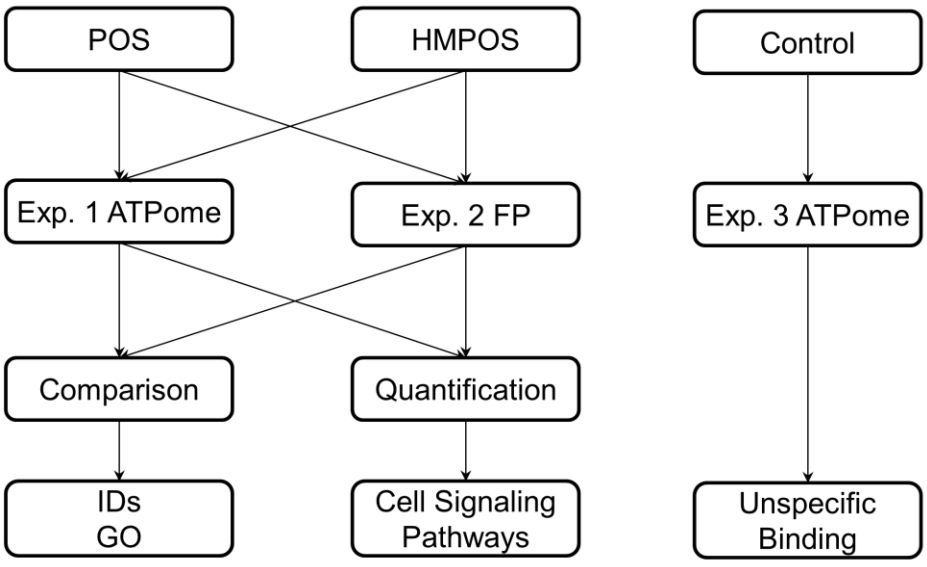


Figure 4.2 Experimental design.

assessed in the context of the high metastatic propensity of HMPOS cells. The efficiency of the ATPome enrichment (ATP) when using the commercial reagent desthiobiotin-ATP was first evaluated. The number, identities of proteins, enrichment of protein kinases and GO terms were compared between the two workflows. Then, after determination of differentially regulated proteins between cell lines the information from the ATPome enrichment experiment and the full proteome was used to determine changes in cell signaling and metabolic pathways. As a control group the equal mixture of both cell lines was used and the experiment was performed as in the ATPome enrichment experiment except that no ATP probe was added.

Cells from the two osteosarcoma cell lines were lysed with addition of protease and phosphatase inhibitors, then samples were cleaned-up with a molecular cut-off cartridge and the protein solution was equilibrated with desthiobiotin-ATP (for the ATPome analysis only). Proteins were reduced, alkylated and then tryptically digested. For the ATPome analysis, the labeled peptides were captured with streptavidin agarose, washed and eluted with acetonitrile. Peptides were subjected to LC-MS/MS analysis (Appendix Figure B.2).

Assessment of the ATPome Enrichment Efficiency

The specificity and advantage in the number of the proteins identified when applying the affinity tag, was evaluated (Appendix Figure B.3). In the control experiment (in which the protocol was performed as per the ATPome enrichment workflow, but no affinity tag was added), there were no proteins identified, which demonstrates that streptavidin

agarose did not capture any biotin-containing proteins from the cell lysate.

Figure 4.3 provides a breakdown of the number of proteins identified after application of the ID Filter. After application of the ID Filter, a total of 847 proteins were identified in the full proteome (FP) analysis and 611 proteins that were modified by the ATP probe. Some of the proteins were uniquely identified in only one experiment, while only 326 (29% of all) proteins were identified with both workflows (Figure 4.3). The number and identities of the individual proteins were also compared between cell lines within each experiment (Figure 4.4). Without

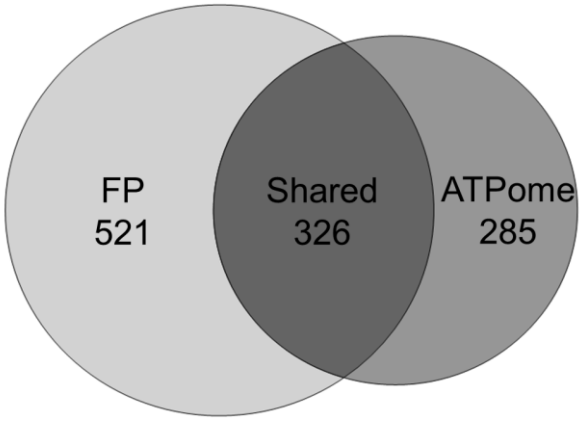


Figure 4.3 Number of proteins identified after application of the ID Filter.

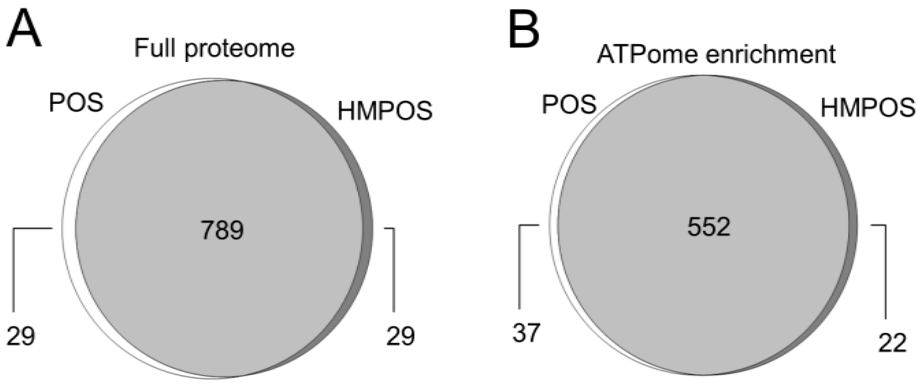


Figure 4.4 Comparison of the number of individual proteins between cell lines within each experiment.

application of the affinity tag, 29 proteins were identified uniquely for the POS and HMPOS cell, respectively (Figure 4.4A). ATPome enrichment led to the detection of 27 and 22 unique proteins for POS and HMPOS cells, respectively (Figure 4.4B).

The identities of the unique proteins is of interest because the programmed behavior of highly metastasizing cells may activate or inhibit the expression of certain proteins and, therefore, influence the biological processes in which these proteins are involved. It should be noted that failure to detect a protein may not indicate the absence of its expression, but it is a valid indicator of a change in the abundance of this protein. The proteins uniquely expressed in either cell lines in either experiments are presented in Table 4.1. Proteins involved in metabolic pathways, kinases, GTPases and receptor

Table 4.1 Proteins uniquely expressed in either cell line for each workflow (FP and ATPome enrichment).

Full proteome				**ATPome enrichment**				
POS		HMPOS		POS			HMPOS	
ACOT7	NAGA	ACO1	NPC2	ACO2	MASTL	RTN4	AARS	ITPA
ACSL1	PAPSS1	ADPGK	NUCB1	ADSS	MRPL12	S100A6	ADPGK	LYN
AHNAK2	PFAS	AK1	PCCB	ANXA8L1	MTPN	SFRP2	AKR1A1	MCM4
BZW2	PFKP	ARPC5L	PLOD3	C1orf58	MYOF	SMAD4	CCAR1	PSAT1
COL6A1	PIR	BAK1	PRPF19	CAPG	NSF	SQLE	CDC42BPB	UBR5
COL6A3	PRPSAP1	DNAJB11	PSMD2	CKB	PDCD5	SSB	CNBP	VIM
CUL4A	PTGIS	EHD4	PTGR1	COL12A1	PFKL	TRIM28	CPNE1	YARS
DCLK1	PYGB	EIF4B	RNPEP	CSNK2A1	PFKP		DDX39	
EHD2	RAB21	FAP	SEC61B	DAB2	PHYH		EPHA5	
ETFB	S100A6	GLOD4	SFXN1	DDX5	PLCD1		FPGS	
GCAT	SQSTM1	H3F3C	TMCO1	EGFR	PNPT1		GMPR	
ISG15	TMEM119	HK1	UGGT1	FASN	PRPF19		HIST1H4L	
ISOC2	TMEM214	HLA-A	UTRN	HARS	RAD23A		HSD17B10	
MGST3	TMSB4X	IGF2R	VAPA	HMOX1	RBM4		IDH2	
MX1		LAMB1		MAP2K3			IQGAP1	

proteins are highlighted. The marked changes in abundance may be associated with the difference in metastatic potential of osteosarcoma cells.

The Gene Ontology Molecular Functions terms show pathways enriched in both experiments, FP and ATPome (Figure 4.5). The diagram shows the percent of all proteins that possess a certain molecular function found in each experiment. For most of the binding groups, ATPome enrichment showed a 10 point increase in percentage, which shows that desthiobiotin labeling in conjunction with enrichment indeed increases the presence of proteins from various binding groups, including nucleotide binders and ATP-binding proteins. The specificity of the ATP affinity tag can also be evaluated here: the probe enriches not only ATP-binding proteins, but also proteins with other nucleotide binding abilities and proteins that were tagged due to the presence of a reactive lysine

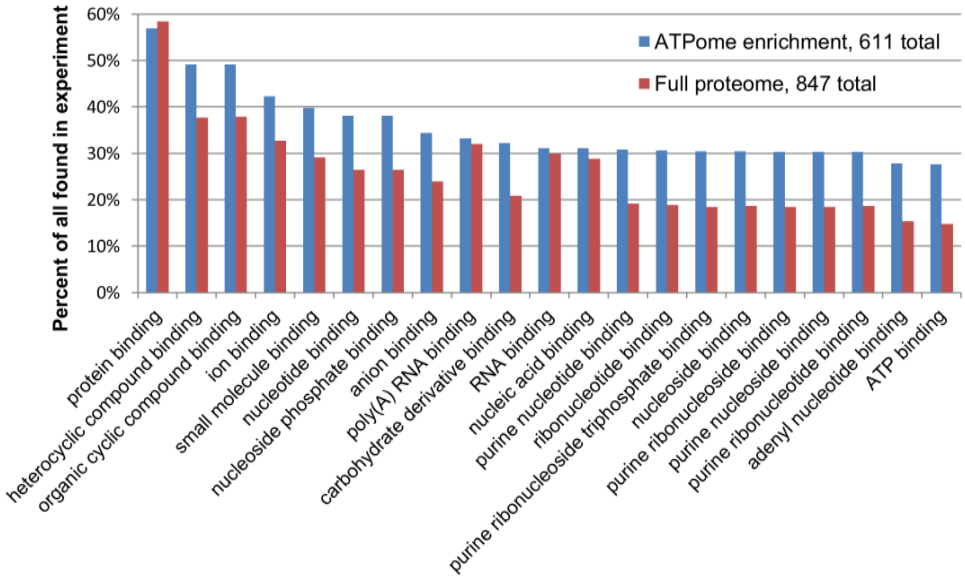


Figure 4.5 Comparison of Gene Ontology Molecular Functions terms within each experiment.

residue in proximity to the probe but which is not necessarily part of a nucleotide-binding site. A few interesting facts may also be noted here: groups such as protein binders, RNA binders and nucleic acid binders were not enriched by the probe. This indicates that the probe is indeed more similar in binding abilities to nucleotides than to nucleic acids.

To evaluate pathways that were differentially enriched in the two experiments, the KEGG pathway terms were compared (Figure 4.6). The diagram indicates that both approaches showed enrichments in pathways, but their extent of significance was dependent on the approach (Figure 4.6A). If proteins that were found in both experiments were removed from the diagram leaving only unique proteins that were detected, it becomes clear that

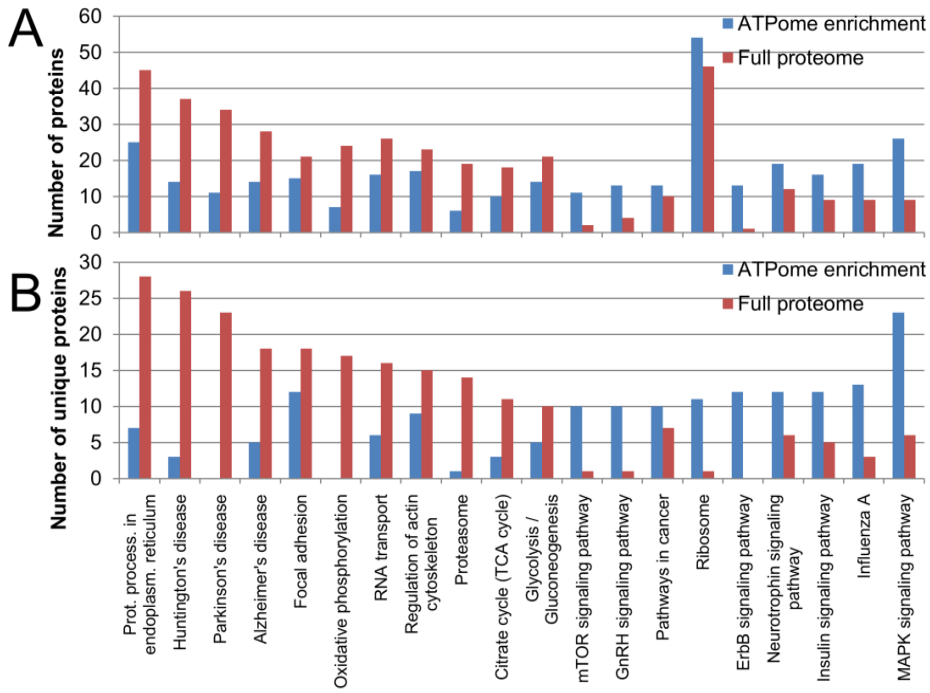


Figure 4.6 KEGG pathways enriched in both experiments with (A) and without (B) mutually found proteins.

the two experiments enrich complementary sets of pathways: the ATPome enrichment finds more proteins of the MAPK, insulin and ErbB signaling pathways, while for the full proteomics experiment, pathways that were found enriched included protein processing in the endoplasmic reticulum, focal adhesion and pathways associated with diseases: Huntington's, Parkinson's and Alzheimer's (Figure 4.6B). As expected, the two workflows provide complementary datasets that allowed a more comprehensive discovery of differentially expressed proteins with potential relation to the disparate cancer biology of the two cell lines studied. Undoubtedly, instead of choosing between the two workflows, a full comprehensive dataset of differences between various conditions can only be obtained with application of both, and, perhaps, even more workflows.

As mentioned previously, protein kinases are low abundant in cells therefore their abundance should be increased upstream of the mass-spectrometric analysis to increase the chance of their detection and identification. The list of all proteins identified with the application of the ATP-affinity probe was sorted in the order of decreasing abundance. The most abundant proteins were investigated using bioinformatics tools and the findings are summarized in Appendix Figure B.3: (A) the dendrogram on top of the heat map shows separation of POS and HMPOS samples; (B) Of the 23 top ranked proteins, 17 had nucleotide binding abilities, 16 were ATP- and ribonucleotide binders and 7 were protein kinases; (C) shows the efficiency of the enrichment procedure – protein kinases that were not detected without the enrichment are now the most abundant proteins in the samples

Again, the ATP-affinity tag shows selectivity not only toward ATP-binders, but also other nucleotide and ribonucleotide binding proteins.

All protein kinases that were detected with application of the ATP affinity tag along with their quantitative changes were labeled on the Kinome Map (Figure 4.7). There were 85 proteins that matched the IDs in the kinome mapper; they present all major families, except the caseine kinase 1 group. In contrast, without kinase enrichment, only 4 protein kinases that matched the kinome map ID list were identified. On the Kinome Map

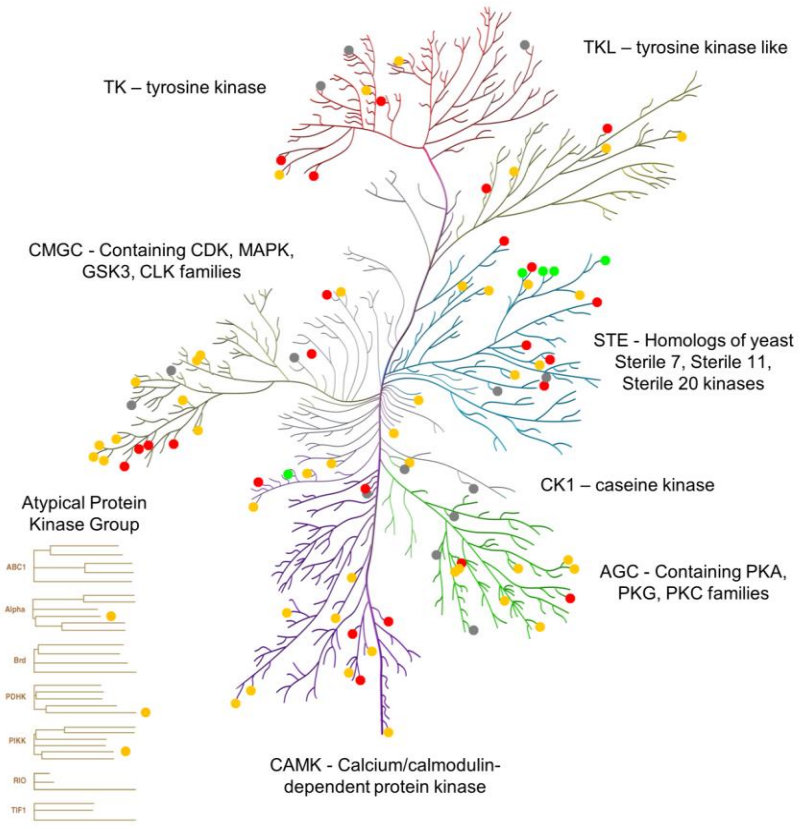


Figure 4.7 Protein kinases detected with application of the ATP affinity tag along with their quantitative changes labeled on the Kinome Map.

(Figure 4.7), the quantitative change in protein abundance for the 85 protein kinases is color coded: green and red dots represent proteins that increased or decreased abundance in HMPOS cells for at least 50%, respectively, grey are those that did not show a large change in abundance and orange dots indicate protein kinases that were detected but were not quantified. Appendix Table B.1 reveals that most of the protein kinases showed decreased abundance levels in the highly metastasizing cell line.

The distribution of the number of protein kinases detected within their families is presented in Figure 4.8A. Figure 4.8B shows relative enrichment of protein kinases within their families. Relative enrichment is calculated by dividing the number of protein kinases within each family by the total number of protein kinases in that family. The histogram shows higher enrichment of protein kinases from STE and CMGC families of

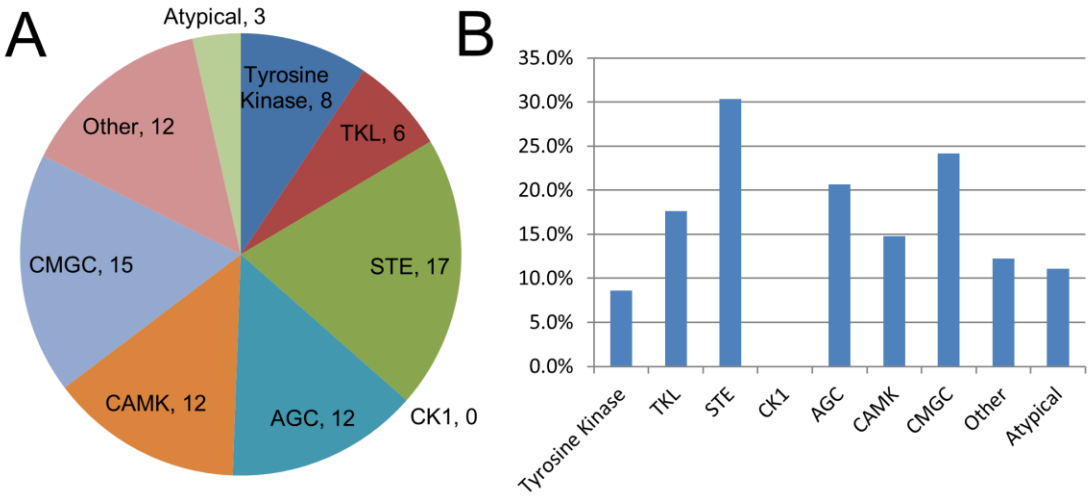


Figure 4.8 Distribution of the number of protein kinases detected within their families (A) and relative enrichment of protein kinases within their families (B).

protein kinases.

Figure 4.9 shows a protein interaction network for proteins found in the ATPome enrichment experiment, as extracted from the STRING database. In the network, each protein is represented by a node while the lines connecting the nodes indicate interactions between proteins. The network consists of 596 proteins that were assigned from a list of 611 proteins minus some isoforms that shared the same gene name. Protein kinases and

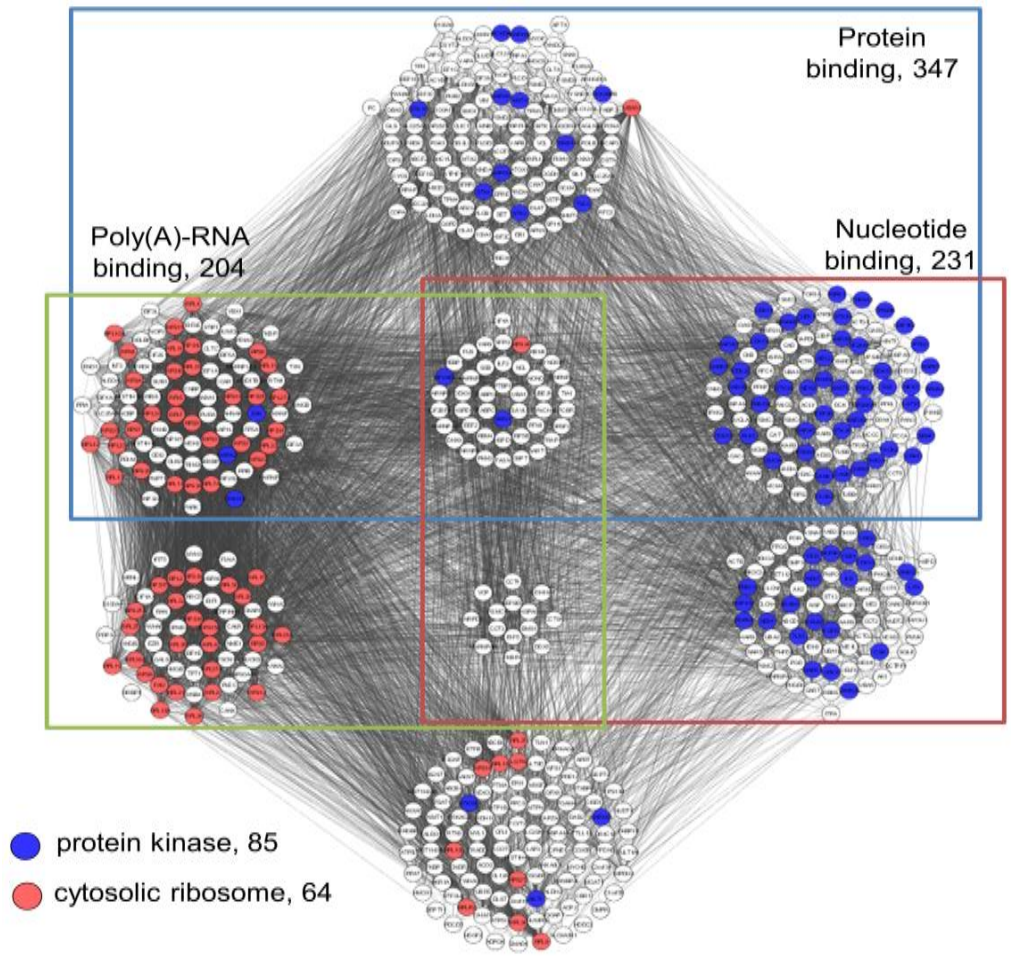


Figure 4.9 Protein interaction network for proteins found in the ATPome enrichment experiment.

ribosomal proteins are labeled with blue and red, respectively. All proteins were grouped according to their binding abilities. The main groups of binders are protein binders with 347 members (58%); poly-RNA binders with 203 proteins (34%) and nucleotide binders with 231 nodes (39%). Some of the proteins possess multiple binding abilities and some were not assigned to any group due to the missing information in the STRING database. The network is dense with interactions showing as many as 7,629 links between proteins. Proteins binders are responsible for 6,351 (83%) of interactions, poly-RNA binders for 6,147 (81%) and nucleotide binders for 3,856 (51%).

In sum, as also observed by other, the ATP-affinity probe enriches proteins that possess nucleotide binding abilities, including ATP-binders, which in turn include protein kinases. Our study also demonstrated that the application of the affinity tag greatly increases the abundance of protein kinases and, hence, the chance of their detection in mass-spectrometry-based proteomics workflow.

Quantitative Proteome Changes in Metastatic vs. Primary OS Cells

After application of the Quant Filter, the list of quantified proteins reduced to 725 and 466 in the FP and ATPome enrichment experiment, respectively. A list of proteins and quantitative changes is presented in the Supporting Information, Table S2 and S3 for the full proteome experiment and the ATPome enrichment, respectively. Appendix Figure B.4 summarizes the observed distributions of protein abundance fold changes between cell lines (HMPOS vs. POS). The histogram of the distribution of the fold changes between cell lines shows that there were more proteins found in the enrichment

experiment that were expressed in lower abundance in the HMPOS cell line (Appendix Figure B.4A). Without enrichment, more proteins were quantified that were expressed in higher abundance (Appendix Figure B.4B). The categorization of the groups is as followed: ">2FC" indicating a fold change greater than 2, "down" and "up" indicate that the change in abundance was more than 50%, "even" are proteins that showed change in abundance less than 50%. The latter group was considered as not showing any change in abundance between cell lines, except if there was change that showed statistical significance.

The quantitative differences are also reflected in boxplots. The distribution of $\log_2$ of fold changes between HMPOS and POS groups in the ATPome enrichment experiment had a median -0.2 (Appendix Figure B.4C), indicating that there were more proteins that decreased their abundance, whereas, in the FP experiment the median of the $\log_2$ fold change is +0.09 (Appendix Figure B.4D). The boxplots also show that some proteins changed their abundance dramatically with the largest fold change value of 13.4.

Pathway Analysis – Adherens Junctions and Cytoskeleton

Cancer metastasis is initiated by weakening of intercellular adhesion, which interferes with tissue integrity. Cell motility may be acquired by reduction, inactivation or loss of intercellular junctions, such as E-cadherin mediated adherens junction (AJ).[131] Strong cellular adhesion critically depends on the association between E-cadherin and the actin cytoskeleton. The cytoplasmic part of E-cadherin binds to β-catenin, which itself is attached to α-catenin, a protein that mediates interactions to the actin cytoskeleton[132]

through many proteins, including α-actinin and vinculin.[133,134]

The abnormal migratory and invasive behavior of tumor cells is a result of a change in the adherens junction formation, maintenance and remodeling, processes that requires precisely regulated mechanism of activation and inactivation of members of the Rho GTPase family of proteins.[131,135–137] Activities of these proteins regulate many cellular processes, such as membrane trafficking, cell cycle and apoptosis,[138–140] cancer development and metastasis.[141]

In mammalian cells, this family of GTPases comprises six groups: the Rho proteins (RHOA, B and C), the Rac members (RAC1, 2, 3 and RHOG), the Cdc42-like proteins (CDC42, TC10, TCL, CHP/WRCH2 and WRCH1), the RhoBTB subgroup (RHOBTB1, RHOBTB2 and RHOBTB3), the Miro proteins (MIRO1 and MIRO2) and the Rnd members (RND1, RND2, RND3/RHOE).[142] The most studied and characterized proteins of this family are RHOA, RAC1 and CDC42.

Rho GTPases cycle between GTP bound active and GDP bound inactive state. Proteins that regulate the exchange of the bound nucleotide are organized into the following three groups: (1) Guanine nucleotide exchange factors (GEFs) activate GTPases by enhancing the release of GDP to enable binding of GTP, (2) GTPase activate proteins (GEFs) and inactivate Rho proteins by catalyzing the hydrolysis of GTP, and (3) guanine dissociation inhibitors (GDIs) bind to Rho GTPases, thus keeping them in the inactive form (Figure 4.10).[143]

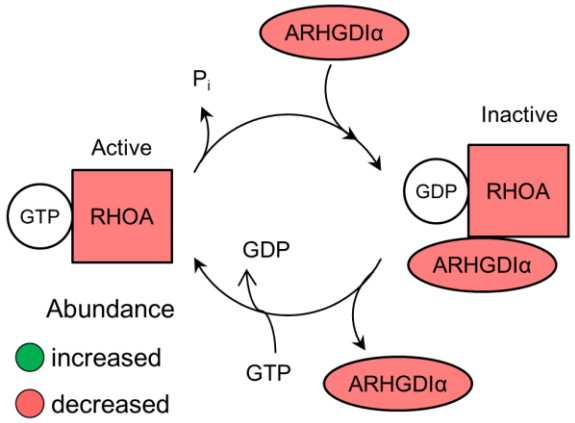


Figure 4.10 Cycle of Rho GTPases between GTP bound active and GDP bound inactive state.

While a comprehensive review of the members of the Rho GTPases, their effects on the AJs and regulation by GEFs, GAPs and GDIs is provided elsewhere,[131,144,145] this pathway analysis will only focus on RHOA and one of its GDI regulators, RHOGDI1.

Earlier studies have shown that inhibition of RHOA activity results in a thinner junction-associated actin filament belt and subsequent disassembly of AJ,[146–150] while the expression of constitutively active mutants of RHOA led to thicker actin filaments and stabilization of AJ.[146,150] RHOA can strengthen or weaken intercellular adhesion and cell spreading depending on cell type.[150–152] In keratinocytes activation of RHOA stabilizes AJs, when its inhibition results in removal of E-cadherin from the junction complex.[147,153] Being a major regulator of the actin-myosin filament system, this GTPase helps to maintain E-cadherin-mediated cell-cell adhesion through action of DIA1[154] and non-muscle myosin II.[155]

At steady state, Rho Guanine nucleotide Dissociation Inhibitor RHOGDI1, which is encoded by *ARHGDIα* gene, is bound to most of the Rho GTPases and keeps them in the inactive state. In mammalian cells, the expression level of RHOGDI1 is roughly equal to

that of RHOA, RAC and CDC42 combined,[156] making the binding competitive between the Rho GTPases. Although upon RHODIA1 depletion by siRNA the level of RHOA was dramatically decreased, the amount of active GTPase was unchanged.[157]

Since this guanidine dissociation inhibitor tightly regulates expression and activity of the main GTPases, which in turn are responsible for proper formation of the adherens junctions, it is hypothesized that dysregulation of RHOGDI1 may cause abnormal behavior of cells. Indeed, ARHGDIα suppresses androgen receptor signaling,[158] while loss of the *ARHGDIα* gene promotes the development and progression of prostate cancer.[159] The latest research reveals that loss of ARHGDIα expression in hepatocellular carcinoma cells contributes to the processes of invasion and metastasis.[131]

The quantitative proteomic results of both experimental workflows, ATPome enrichment and FP, are compiled in a Table 4.2. Cell signaling pathways and protein functional categories are outlined. It is interesting to note that many proteins involved in the regulation of actin cytoskeleton and adherens junction displayed differential expression in the highly metastasizing osteosarcoma cell line. Most notably, GTPase RHOA and its regulator, ARHGDIα, were found as being significantly under-represented in the HMPOS cells. The reduced expression of RHOA inhibitor may indicate that the GTPase is kept in active form, although in lower abundance, a behavior that was described previously.[157] The consequence of lower expression of the active RHOA is traceable through the pathways that are responsible for cellular adhesion.

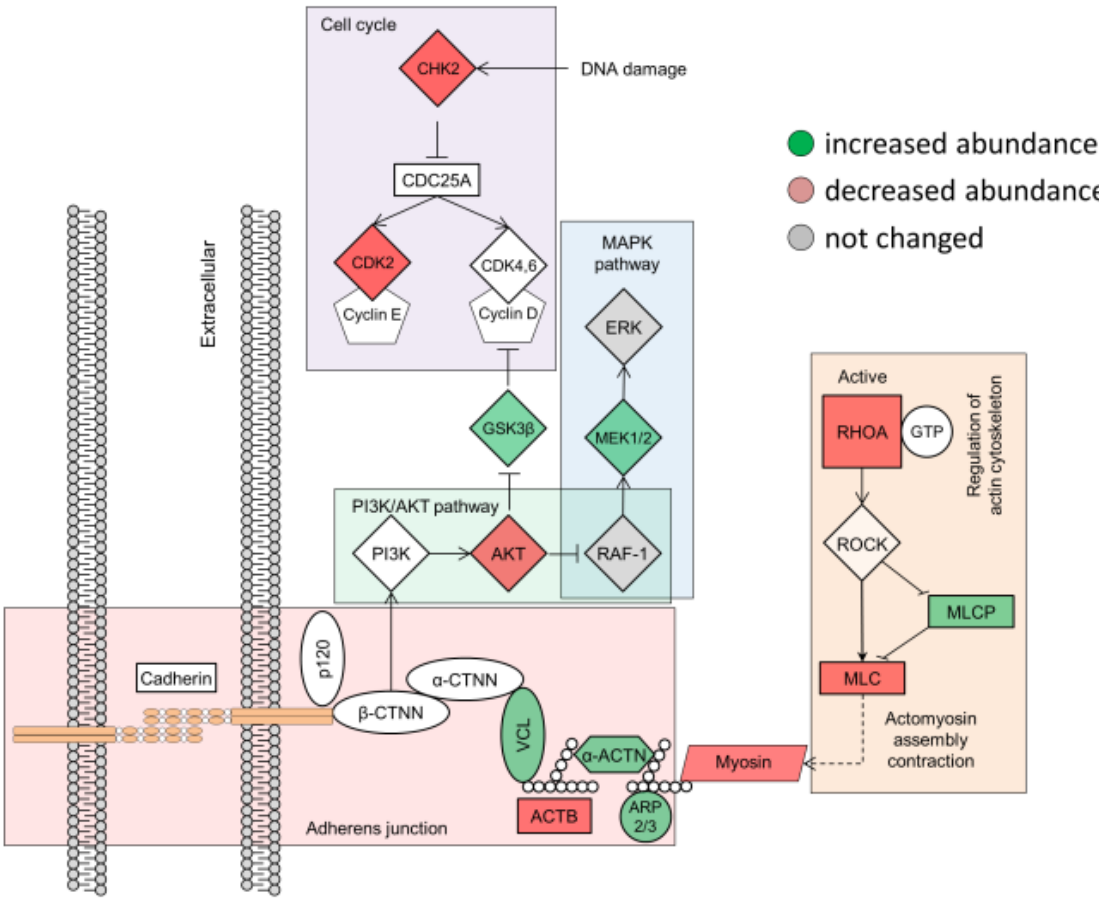


Figure 4.11 Building blocks and changes in the pathways related to the RHOA mediated weakening of extracellular adhesion.

The building blocks and changes in the pathways are visualized in Figure 4.11. The decreased expression of RHOA leads to decreased expression of myosin light chain (MLC), which affects contractility of the myosin fibers. Actin, which connects myosin to the cadherin complex, is under-expressed in HMPOS, indicating weaker structure of the cells' cytoskeleton and, hence, the adherens junction. Actin-β, one of the two actins that compose the cytoskeleton, was expressed in significantly lower abundance (1.23 fold change), which was confirmed by Western blotting (Appendix Figure B.5A). Linker and

binder proteins, vinculin, α-actinins (ACTN1 and ACTN4) and Arp 2/3 complex, showed significantly higher abundance in HMPOS cells (Table 4.2).

The RHOA mediated weakening of extracellular adhesion may have a direct relationship to the metastatic potential of cells because the weaker intercellular connections will promote the migration of cells from the main tumor. Our current findings and the work of others indicate that the main players involved in regulating and forming the adherens junction may be potential targets for preventing osteosarcoma tumor metastasis.

Table 4.2 Cell signaling pathways and protein functional categories from ATPome enrichment and FP.

Accession	Gene	Name	Fold Change	t-test Significant	Experiment
		Regulation of actin cytoskeleton			
O18840	ACTB	actin, beta	-1.23	++	KE
F1PQL8	ACTG1	actin, gamma 1	-1.15		KE
F1P9J3	MYH9	myosin, heavy chain 9, non-muscle	-1.69	+++	KE
F1P9J3	MYH9	myosin, heavy chain 9, non-muscle	-1.69	+++	FP
F1Q3Y0	PFN1	profilin 1	-1.67	+	KE
F1PPJ1	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha	1.50	++	KE
F1PA10	PIP4K2C	phosphatidylinositol-5-phosphate 4-kinase, type II, gamma	-1.81	+	KE
E2RSI6	EZR	ezrin	-1.97	+++	FP
J9P127	TMSB4X	thymosin beta 4, X-linked	-2.74	+++	KE
F1PQN5	CFL1	cofilin 1 (non-muscle)	-1.37	++	KE
J9P2G7	CFL2	cofilin 2 (muscle)	1.58	+++	FP
F1Q0U7	ENAH	enabled homolog (Drosophila)	-1.43	+	FP
		Other actin interacting proteins			
E2QWU0	ARC20	actin related protein 2/3 complex, subunit 4, 20kDa	1.43	+	FP
J9P309	ARPC3	actin related protein 2/3 complex, subunit 3, 21kDa	1.58	+	FP
J9P2J8	CALD1	caldesmon 1	-2.57	+++	FP
F1PMS4	CAPN2	calpain 2, (m/II) large subunit	-2.29	+++	FP
F1P7P3	CAPZA1	capping protein (actin filament) muscle Z-line, alpha 1	1.26	+++	FP
F1PQS3	CAPZB	capping protein (actin filament) muscle Z-line, beta	1.49		FP
F1P679	ACTR3B	ARP3 actin-related protein 3 homolog B (yeast)	1.52	+	FP
J9P1J5	DSTN	destrin (actin depolymerizing factor)	1.60	++	FP
F1PWW0	FLNA	filamin A, alpha	-1.38	++	FP

Table 4.2 Cell signaling pathways and protein functional categories from ATPome enrichment and FP (Cont.).

F6Y3P9	GSN	gelsolin (amyloidosis, Finnish type)	1.52	+++	FP
E2QY85	MYL12A	myosin, light chain 12A, regulatory, non-sarcomeric	-1.34	+++	FP
E2RDU7	PALLD	palladin, cytoskeletal associated protein	-2.81	++	FP
F1PEK5	SEPT2	septin 2	-1.61	+	KE
F1PEK5	SEPT2	septin 2	-1.35	++	FP
F1Q067	SEPT7	septin 7	2.00		FP
J9P4F3	VCL	vinculin	1.57	+	FP
F1PLS4	VIM	vimentin	6.08	+++	FP
		Adherens junction			
E2QY07	ACTN1	actinin, alpha 1	1.73	+++	FP
E2R5T9	ACTN4	actinin, alpha 4	1.38	+++	FP
F1PL93	ARHGDIA	Rho GDP dissociation inhibitor (GDI) alpha	-1.16	+	FP
F1PJ65	IQGAP1	IQ motif containing GTPase activating protein 1	1.36		FP
F1P7Q4	RAC1	ras-related C3 botulinum toxin substrate 1	-1.03		FP
F6XY66	RHOA	ras homolog gene family, member A	-1.47	+	FP
E2RMM5	RHOC	ras homolog gene family, member C	-1.09		FP
P60952	CDC42	cell division cycle 42	1.01		FP
		Cell cycle and apoptosis			
E2R4Y4	GSK3B	glycogen synthase kinase 3 beta	1.31		KE
J9P1S6	CDK1	cyclin-dependent kinase 1	-1.92	+++	KE
E2QW70	CDK2	cyclin-dependent kinase 2	-1.46	+	KE
F1PPV7	CDK20	cyclin-dependent kinase 20	-1.85	++	KE
E2RGN2	CDK5	cyclin-dependent kinase 5	-2.08	+	KE
E2QU88	CHEK2	CHK2 checkpoint homolog (S. pombe)	-2.28	+	KE
E2RJL1	LGALS1	lectin, galactoside-binding, soluble, 1	-1.38	+++	FP

Table 4.2 Cell signaling pathways and protein functional categories from ATPome enrichment and FP (Cont.).

E2RJL1	LGALS1	lectin, galactoside-binding, soluble, 1	-2.10	+++	KE
		Selected kinases			
E2RJX4	AKT1	v-akt murine thymoma viral oncogene homolog 1	-1.86	+++	KE
P05124	CKB	creatine kinase, brain	-13.45	+++	FP
E2RFF7	CSK	c-src tyrosine kinase	-1.41	+	KE
Q2KM13	EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase 2	-3.45	+	KE
E2R5H4	ILK	integrin-linked kinase	-1.43	+	KE
F1P647	LATS1	Serine/threonine-protein kinase LATS1	-1.63	+	KE
E2R7W7	MAP2K6	mitogen-activated protein kinase kinase 6	-1.48	+	KE
J9P806	MAP4K3	mitogen-activated protein kinase kinase kinase kinase 3	-1.44	+++	KE
J9NTP8	MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4	-1.49	++	KE
E2RGP4	MAP4K5	mitogen-activated protein kinase kinase kinase kinase 5	1.84	+	KE
J9P499	SLK	STE20-like kinase (yeast)	-2.05	+++	KE
F6URD2	STK3	Serine/threonine-protein kinase 3	1.24	+	KE
F1PGB1	STK4	serine/threonine kinase 4	1.17	+	KE
J9NW12	TAOK2	TAO kinase 2	-1.46	++	KE
J9P605	ZAK	Mitogen-activated protein kinase kinase kinase MLT	-2.76	+++	KE
F1PPJ1	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha	1.50	++	KE
F1PA10	PIP4K2C	phosphatidylinositol-5-phosphate 4-kinase, type II, gamma	-1.81	+	KE

Negative fold change indicates lower expression in HMPOS cell line, positive – higher. The p-values are coded as follow: $0.01<p\leq0.05$ labeled as "+"; $0.001<p\leq0.01$ labeled as "++"; and $0.0001<p\leq0.001$ labeled as "+++"

Pathway Analysis – AKT1 and Downstream Signaling

AKT1, also known as protein kinase B or v-akt murine thymoma viral oncogene homolog 1, is a serine-threonine protein kinase that plays a central role in cell signaling downstream of growth factors, cytokines and various cellular stimuli. Loss or gain of AKT1 activity correlates with pathogenicity of complex diseases, including cancer. A literature search reveals that AKT1 has over 100 reported substrates and regulate processes of cell survival, growth, proliferation, angiogenesis, migration and invasion.[160]

In the current study we observed AKT1 in the ATPome workflow and it showed a 1.86 fold significantly lower expression in HMPOS cells. The identity and change in abundance of this protein were confirmed by Western blotting (Appendix Figure B.5B). AKT1 regulates many proteins and is involved in pathways with relevance to cell metabolism, cell proliferation and survival, and angiogenesis.

One of the downstream targets of AKT1 is the protein ERK, also known as mitogen activated protein kinase 1 (MAPK1). ERK is involved in the MAPK signaling cascade, which transfers a signal from a receptor on the surface of the cell to the nucleus and is responsible for cell proliferation, differentiation and migration. AKT1 phosphorylates and inhibits ERK signaling via the protein c-Raf (also known as RAFL).[161,162] AKT1 can directly phosphorylate apoptosis signal-regulated kinase (ASK1, also known as MAP3K5) and therefore inhibit the downstream signaling of the stress-activated MAPKs, JNK and p38.[163] Observed proteins and their level changes were mapped on the MAPK signaling cascade presented in Appendix Figure B.6.

Proteins involved in the upper branch of the pathway, also referred to as the classical MAPK pathway, and responsible for cell proliferation, showed no signs of altered protein expression in HMPOS cells. On the other hand, multiple proteins involved in the lower branches, including the JNK and p38 MAPK pathways, showed significantly lower abundance in the highly metastasizing cell line. The lower branch is involved in the regulation of several cellular processes including apoptosis. From the pathway one can see that a balance between AKT1 influence on the proliferative classical MAPK and pro-apoptotic JNK/p38 branches could be established via a crosstalk of the proteins.

In addition to the MAPK pathway, AKT1 targets glucose synthase kinase-3β (GSK-3β). AKT1 phosphorylation and inactivation of this target kinase[164] likely drives cell proliferation through regulation of the activity of proteins involved in the cell cycle entry. The GSK-3β kinase-mediated phosphorylation of proteins D and E cyclins targets them to proteasomal degradation, which plays a central role in the G1/S phase of cell cycle.[165,166] Evidently, the AKT1 inhibition of GSK-3β should increase stability of these cell cycle proteins. In HMPOS cells, lower expression of AKT1 promotes 1.33 fold increase in GSK-3β, a protein that transfers the protein abundance flux to the cell cycle pathway.

Another key player in the cell cycle regulation, check point kinase 2 (CHK2), is a major component of the DNA and other damage response. After genotoxic stress, CHK2 is activated and it phosphorylates more than 20 proteins to promote a proper cellular response - cell cycle checkpoint activation, inducing of apoptosis, DNA repair, or tolerance to the damage - depending on the extent of the damage, cell types and other factors.[167] In the event of DNA damage the progression of cell cycle will be suspended

(arrested) in an attempt to repair the damage. When the damage occurs at the DNA replication time, the S phase, one mechanism of the arrest involves CHK2 phosphorylation of a phosphatase CDC25A with subsequent proteasomal degradation of the latter. The downstream substrates of the phosphatase, cyclin dependent kinases 2,4 and 6 (CDK2, CDK4 and CDK6) are then activated to promote G1/S phase progression.[168] Interestingly, many cyclin dependent kinases, including CDK2, exhibit decreased abundance in the HMPOS cells, implying changes in cell cycle activity that may be linked to the abnormal behavior of metastatic cells. The check point kinase CHK2 showed significant 2.28 fold lower expression in HMPOS (see Table 4.2), which may indicate that in the highly metastatic phenotype the cell's response to DNA damage and associated cell cycle arrest is impeded.

Our current comparative proteomics study in combination with an ATP affinity probe suggests that aberrant protein expression of critical proteins involved in intercellular adhesion, cell cycle regulation and other important pathways may be linked to a higher propensity to metastasis. RHOA and AKT1 were significantly under expressed in the HMPOS cell lines. AKT1 is a currently heavily researched target for modulating cell invasiveness and drug resistance of cancer cells, and the current study underscores that AKT1, should also be investigated as potential therapeutic targets in osteosarcoma.

Summary

In this study the changes in protein expression between osteosarcoma cell lines that differ in metastatic potential were assessed. We applied a chemical affinity tag, desthiobiotin-ATP, for selectively enriching the ATPome of the cells. The enrichment procedure

showed specificity of the tag towards ATP-binding proteins, protein kinases and ribosomal proteins. The ATPome enrichment workflow yielded 611 proteins, including 85 protein kinases. We complemented the ATPome analysis with a full proteome analysis. The full proteome analysis covered different segment of the proteome in particular it allowed an evaluation of proteins of the cytoskeleton and adherens junction complex.

Application of bioinformatics tools revealed substantial changes in protein pathways and networks that were aberrantly regulated in the highly metastatic phenotype, including cytoskeleton, cell adhesion and motility, as well as AKT1 signaling, cell cycle regulation and checkpoint control.

ASSOCIATED CONTENT

This material is available free of charge via the Internet at http://pubs.acs.org.

CHAPTER 5 Ti^{4+}-IMAC Approach Reveals Complementarity of Proteases to Cover the Phosphoproteome of Jurkat Cell Line

Ievgen Motorykin,[1] Thin-Thin Aye,[2] Claudia S. Maier[1] and Albert J. Heck[2]

[1]Department of Chemistry, Oregon State University, 153 Gilbert Hall, Corvallis, Oregon 97331, United States

[2]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

The manuscript is in preparation

*Corresponding author, E-mail: claudia.maier@oregonstate.edu

Abstract

Protein phosphorylation is responsible for many important biological processes, including the transfer of cell signals. Aberrant phosphorylation is linked to many diseases, therefore, the analysis of protein phosphorylation is very important. This remains an analytical challenge because of low abundance of phosphorylated proteins and limitations of mass-spectrometric detection. In this work, we assessed the complementarity of five proteases, trypsin, LysC, GluC, AspN and chymotrypsin, with the goal to determine if their combined use results in an increased number of identifiable phosphorylation sites using mass spectrometry. Jurkat E6.1 cells were used for evaluating the multiplexed protease protocols. Ti^{4+}-IMAC was used for enrichment of phosphopeptides, prior to mass spectrometric analysis using a DDA-dependent acquisition for accurate mass analysis and employing two fragmentation techniques, CID and ETD. A total of 11,338 phosphorylation sites (PS) were identified with probability greater than 75%. Our results reveal that only 0.6% of total phosphosites were detected in experiments with individual proteases, when 55% were unique to only a single protease workflow. The data also indicates that replicates of the trypsin workflow increased the number of detectable PS by 29%. Comparison of the combination of results from the tryspin and LysC workflows and the sequential use of the two protease give rise to a 17% and 7% increase in unique PS identified, respectively. A parallel use of trypsin and another protease increased the PS number by a range from 8.3% for chymotrypsin to 22% for GluC, proving that multi-replicate and multi-protease approach is very valuable in phosphoproteome analyses.

Introduction

Protein phosphorylation is an important mechanism for transferring cell signals and the regulation of many biological processes.[169,170] The correlation of aberrant protein phosphorylation with diseases is apparent from the literature,[171,172] therefore in-depth characterization of protein phosphorylation is a critical step.[173,174] Phosphoproteomics aims to identify at the global level phosphorylation sites of proteins in a biological systems[175,176] at a given point of time. This is not an easy task since protein phosphorylation is of low abundance in a cell proteome and covers a high dynamic range, limiting its detection by mass-spectrometry.[177] Additionally, phosphopeptides have a poor signal response when compared to their non-phosphorylated counterparts and tandem mass spectrometric fragmentation of phosphopeptides often result in fragment ion spectra that are potentially difficult to interpret.[178]

In recent years, a variety of techniques have been developed to improve detection, quantitation of phosphopeptides and site localization of the phosphorylation. These techniques can be separated as those improving mass-spectrometric detection and those improving sample preparation.[179]

The first group consists of a targeted analysis of phosphorylated peptides and includes techniques like neutral-loss scanning using the phosphate fragmentation properties,[180] precursor ion scanning using specific phosphorylated peptide fragments,[181,182] multiple stages of fragmentation,[183,184] multiple reaction monitoring (MRM)[185] and electron transfer dissociation (ETD).[186–188] Efficient mass-spectrometric detection is critical for comprehensive profiling of a phosphoproteome.[177] Various peptide fragmentation

methods possess individual advantages for analysis of phosphopeptides with different physicochemical properties,[189] CID having superior sensitivity and speed,[190] HCD (higher-energy collisional dissociation) providing more comprehensive fragmentation and less neutral loss[191], and ETD being more efficient for sequencing highly charged phosphopeptides.[186,192,193] Hybrid mass-spectrometers, such as the LTQ-Orbitrap Velos or Elite[194], which are high sensitive, fast and can utilize all three fragmentation methods, are superior platforms for in-depth phosphoproteome analyses.[177]

The second group, sample preparation techniques, is based on the utilization of unique features of the phosphate group, namely its intrinsic negative charge and polarity. For instance, strong anion exchange chromatography (SAX),[195,196] strong cation exchange (SCX),[197] hydrophilic interaction liquid chromatography (HILIC)[198] and isoelectric focusing (IEF)[199] are using low pH to separate phosphopeptides. Tyrosine phosphopeptides can also be enriched by specific antibodies that can be applied at the peptide[200,201] or protein level.[202–204] The coordination ability of the phosphate group has also been used for its capturing and led to development of enrichment techniques that are based on immobilized metal affinity chromatography (IMAC)[205–207] with variants based on metal oxides of aluminum,[208] zirconium[209] and titanium.[210–213] Calcium[214] and barium[215] precipitation of phosphopeptides was shown to be an alternative strategy for enrichment.

Studies have shown that the most effective phosphopeptide enrichment workflows combine two techniques, e.g. low pH SCX to collect early fraction containing phosphopeptides and then subsequent enrichment with either IMAC[183,216] or titanium

oxide.[117,217,218] However, it remains a challenge to comprehensively identify phosphorylation sites (PS) due to the fact that all current methods have inherent biases.[210,219,220]

Some of the results of the application of before mentioned separation and enrichment strategies include:

- large scale phosphoproteome analysis of mouse liver resulted in the identification of 5,635 PS;[221]
- large scale phosphoproteome analysis of *Drosophila melanogaster* embryos yielded 13,720 PS;[178]
- IMAC enrichment and HILIC chromatography lead to identification of 8764 unique phosphopeptides in yeast *Saccharomyces cerevisiae*;[222]
- phosphopeptides enriched from human liver were fractionated by an off-line RP-LC at high pH and then the fractions were pooled and submitted to an on-line RP-LC-MS/MS resulting in the identification of about 10,000 phosphorylation sites, the largest data set at the time;[223]
- extensive fractionation methods were applied to the phosphoproteome of the HeLa cells, which resulted in 20,443 phosphosite IDs.[224]

To improve protein identifications and proteome sequence coverage, various proteases with complementary cleavage specificities may be used in parallel,[179,225,226] consecutively[225,226] or simultaneously.[227] This method have been applied for analysis of proteome,[228,229] phosphoproteome[179,225,230] and glycoproteome.[226,231] Using this approach, Bian et al. used double stage, GluC and trypsin, digestion and identified 8,507 unique

phosphorylation sites in HeLa cells.[225] These data, combined with a datsaset from a conventional trypsin workflow resulted in a list of 11,262 phosphosites, of which 6,858 has site localization probability higher than 99%. Gauci et al. used three proteases in parallel experiments and applied Ti^{4+}-IMAC phosphopeptide enrichment. The combined results produced a list of 5,036 non-redundant phosphopeptides.[179]

In this study, we evaluated complementarity of established methods for phosphosite identification when for protein digestion various enzymes and protease combinations were used, namely trypsin, LysC, AspN, GluC, chymotrypsin and a combination of LysC with trypsin. The use of these 5 proteases and a tandem digestion in a single study has not been described before. The total number of phosphosites identified with greater than 75% probability in this study was 11,338, higher than published results from experiments that use similar approaches.

Materials and Methods

Chemicals

Penicillin/streptomycin was supplied by Lonza. Complete EDTA–free protease inhibitor mixture and phosSTOP phosphatase inhibitor mixture was purchased from Roche. Reagents for Bradford assay were supplied by Bio-Rad. AspN, chymotrypsin, GluC, trypsin were purchased from Promega and LysC from Wako. The Sep-Pak C_{18} cartridges were from Waters.

Experimental Workflow

Schematic of the experimental design is outlined in the Appendix Figure C.1.

Jurkat T lymphoma cells (E6.1 clone) were grown in RPMI 1640 medium with 10% fetal bovine serum and penicillin/streptomycin added. Cells were centrifuged at 1,500g for 1 min and growth media were renewed with addition of 10 μM PGE2 to the final concentration of cells 1-2×10^6 cells/mL. After incubation for 10 min cells were washed twice with PBS and harvested.

Lysis was performed by ultrasonication (3 x 1 min at 60 W with 15% duty, 4ºC) the cell suspension in buffer that contains 50 mM ammonium bicarbonate (ABC, pH 8.0), 8 M urea, 100 μM sodium orthovanadate, complete EDTA-free protease inhibitor mixture and phosSTOP phosphatase inhibitor mixture. Cell debris was removed by centrifugation at 20,000g for 15 min at 4 °C and the supernatant was transferred to a new microcentrifuge tube. The total concentration of proteins was measured using a Bradford assay (BioRad, Veenendaal, The Netherlands) and divided onto 1 mg aliquots for enzymatic digestion (6 in total). 10 μL of 2.5% Rapigest SF was added to each aliquot and proteins were reduced by 20 μL of dithiothreitol (1μg/μL in water, 30 min incubation at room temperature) and alkylated by 20 μL of iodoacetamide (5 μg/μL in 50 mM ABC, 20 min incubation at room temperature in the dark). The first aliquot was diluted 10 times with 50 mM ammonium bicarbonate to reach <1 M urea concentration and AspN was added at an enzyme/substrate ratio of 1:150. After incubation for 4 hours at 25 ºC AspN was added again at an enzyme/substrate ratio of 1:150 and mixture was incubated at 25 ºC for 16 hours (overnight). The second aliquot was diluted 10 times with 50 mM ABC to reach <1 M urea concentration and chymotrypsin was added at an enzyme/substrate ratio of 1:50. After incubation for 4 hours at 25 ºC chymotrypsin was added again at an enzyme/substrate ratio of 1:50 and mixture was incubated at 25 ºC for 16 hours

(overnight). The third aliquot was diluted 10 times with 50 mM ABC to reach <1 M urea concentration and Glu-C was added at an enzyme/substrate ratio of 1:50. After incubation for 4 hours at 25 ºC Glu-C was added again at an enzyme/substrate ratio of 1:50 and mixture was incubated at 25 ºC for 16 hours (overnight). Lys-C was added to the fourth aliquot at an enzyme/substrate ratio of 1:50 and mixture was incubated at 25 ºC for 16 hours (overnight). The fifth aliquot was diluted 4 times with 50 mM ABC to reach 2 M urea concentration and trypsin was added at an enzyme/substrate ratio of 1:50 and mixture was incubated at 25 ºC for 16 hours (overnight). Lys-C was added to the sixth aliquot at an enzyme/substrate ratio of 1:50 and mixture was incubated at 25 ºC for 4 hours. After that the solution was diluted 4 times with 50 mM ABC to reach 2 M urea concentration and trypsin was added at an enzyme/substrate ratio of 1:50 and mixture was incubated at 25 ºC for 16 hours (overnight). TFA was added to all aliquots to final concentration of 5% to quench the enzymes. The samples were then desalted and cleaned using Sep-Pak C_{18} columns, dried in vacuum centrifuge and stored at -80 ºC for further analysis.

Phosphopeptide Enrichment

To reduce variation between enrichment, the IMAC columns were simultaneously spinned in a centrifuge. The Ti^{4+}-IMAC beads were prepared as described previously.[232–234] The IMAC tip was built with a Gel-loader tip and a C8 plug. The bottom longer part of the tip was cut off and the tip was placed in the microcentrifuge tube using a centrifuge column adaptor. The suspension of Ti^{4+}-IMAC beads were shaken briefly and 50 µg were loaded in a 200-µL tip. This method allows enriching phosphopeptides from 250 µg of

cell protein digest. The beads were then conditioned with 30 µL of loading buffer (80% ACN/6%TFA) and columns centrifuged at 2500g for 2 min. Protein digests (equivalent to 250 µg of protein) were then loaded onto the beads and columns centrifuged at 500 rpm for 65 min and then at 1000 rpm for 20 min to remove liquid from the tips. After loading, the tips were washed subsequently with 50 µL of washing buffer 1 (50% ACN, 0.5% TFA, 200 mM NaCl) and centrifuged at 1,500 rpm for 30 min, and then with 30 µL of buffer 2 (50% ACN/0.1% TFA), centrifuged at 1,500 rpm for 40 min. Phosphopeptides were eluted in new microtubes (already containing 35 µL of 10% FA) in two steps. First, 25 µL of 10% ammonia were applied and columns centrifuged at 800 rpm for slow elution, and then at 100 rpm for 30 min and 2000 rpm for 10 min to remove remaining liquid from the columns. Second elution was performed with 5 µL of 80% ACN/2% FA and columns centrifuged at 1,000 rpm for 20 min, 1500 rpm for 10 min and lastly at 2000 rpm for 5 min. The collected eluate was acidified with 3 µL of 100% FA and the solutions of phosphopeptides were subjected to LC-MS analysis.

Liquid Chromatography - Mass Spectrometry

The reversed phase nano-LC−MS/MS setup consisted of a Proxeon EASY–nLC 1000 (Thermo Scientific) and Thermo LTQ-Orbitrap Elite equipped with an electron transfer dissociation (ETD) source (ThermoFisher Scientific, Bremen, Germany). The analytes were first trapped on a C_{18} column (Dr Maisch Reprosil C18, 3 µm, 2 cm x 100 µm) at a maximum pressure of 800 psi with 100% solvent A (0.1% FA in water) and then resolved by an analytical column (Agilent Poroshell 120 EC–C18, 2.7 µm, 40 cm x 50 µm). The analytical 150-min gradient was as follow: from 7% to 30% of solvent B (0.1% FA in

ACN) at a flow rate 100 nL/min. Total time of analysis was 180 min per sample. Eluted peptides were introduced into the in-house built nanoESI source (distally coated fused silica emitter with 360 μm outer diameter, 20 μm inner diameter, 10 μm tip inner diameter), and sprayed into the MS at a capillary voltage 1.7 kV. The LTQ-Orbitrap Elite mass-spectrometer was automatically switching between MS and MS/MS for data-dependent mode. Decision tree algorithm[235] was applied when acquiring data. Briefly, a survey scan was acquired in the Orbitrap at resolution 60,000 at m/z 400 after ion accumulation to a target value of 5×10^5 in the linear ion trap in the m/z range 350 – 1500. Charge state filter was applied and precursors with an unknown charge or charge 1 were excluded. Dynamic exclusion rejected up to 500 ions for 40 s. After the survey scan, top 10 most intense precursors were subjected to either CID or ETD fragmentation based on the decision tree for choice of the most appropriate technique for a selected precursor.[236] The algorithm was based on the charge of the ion: doubly charged ions were subjected to CID fragmentation while the rest to ETD.

Peak Detection and Protein Database Search

Proteome Discoverer software (v. 1.3, Thermo Scientific) was used to process all MS/MS data. Peak lists containing CID and ETD fragmentation were generated with a signal-to-noise ratio of 1.5. Also, the ETD non-fragment filter was implemented: precursor peak removal within 4 Da, charge reduced precursor removal within a 2 Da, removal of known neutral losses from charge-reduced precursor within a 2 Da (the maximum neutral loss mass was set to 120 Da) and FT overtones were removed within mass window offset of 0.5 Da. All peak list files were searched against a Uniprot Swiss-Prot database (July

2013, taxonomy Homo Sapience) using Mascot search engine (version 2.3, Matrix science, UK). The following parameters were used for database search: precursor mass tolerance 50 ppm, fragment mass tolerance 0.5 Da. When processing data from various proteases, unless specified otherwise, the following parameters were used: enzyme – trypsin/P, LysC, AspN, GluC, chymotrypsin for respective proteases and trypsin/P for the combination of LysC and trypsin; variable modifications – N-terminal acetylation, methionine oxidation, STY phosphorylation; maximum missed cleavages – 2 when using trypsin/P and 5 for the rest. The instrument type was specified as ESI-TRAP and ETD-TRAP for the corresponding mass spectra. The resulting list of peptides was further reduced by applying following filters: peptide confidence – high; peptide length – minimum 6 residues; rank – 1; peptide search rank – 1; peptide Mascot score – 20, peptide and protein groupings were disabled, protein FDR – 1% at the PSM level utilizing the percolator-based algorithm.[237] Phosphosite localization probability was calculated by Proteome Discoverer incorporated

The resulting file, containing all the information about phosphosites, localization probabilities and other details is supplied in Supplementary material as a text file.

Results and Discussions

This study was designed to compare the number of phosphosites (PS) identified with application of various proteases and combination of proteases. The cell lysate was divided in six parts for 5 proteases and one combination (Figure 5.1). The Ti^{4+}-IMAC phosphopeptide enrichment was performed in triplicate. After the enrichment, each sub-sample (part of the triplicate) was injected twice in the mass-spectrometer. The total

number of samples analyzed and compared was 6 for each protease, and 36 in total.

To compare the number of PS between samples the phosphosite unique identifier was created. For this, the following key was used: [Phosphorylated amino acid][Positions within protein]-[Protein Uniprot accession number]. By using this key every PS has a unique name, for example, S355-A0AVK6, is phosphosite at the serine residue at the position 355 in the protein that has Uniprot accession number A0AVK6.

In this study, four aims were pursued. Aim 1 evaluates the technical reproducibility within each protease. For technical replicates, PS from multiple injections were added and compared to the same from a protease's technical replicate. In other words, the IMAC reproducibility and not injection reproducibility was compared. Aim 2 compares

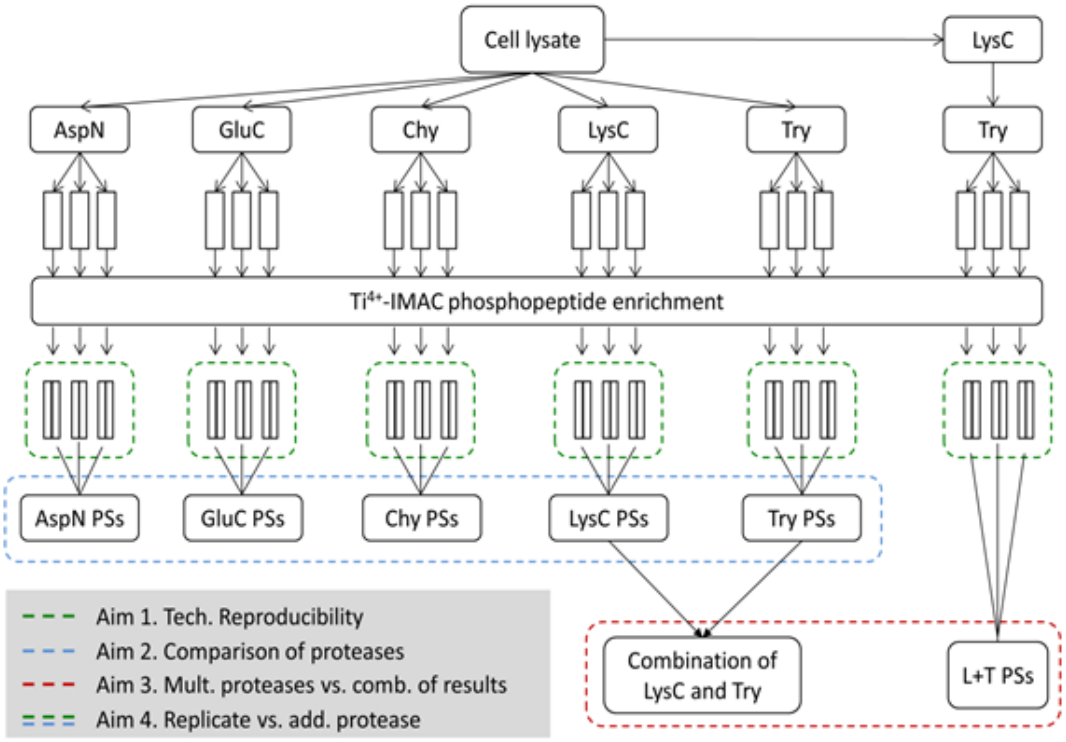


Figure 5.1. Project aims

Table 5.1 Number of phosphosites identified

Protease	Total
ASP_N	2526
CHY	1438
GLU_C	2844
TRY	7205
LYS_C	3938

Table 5.2 Number of unique phosphosites identified

Name	Exclusive	% of all	% of own
ASP_N	876	8.0	34.7
CHY	319	2.9	22.2
GLU_C	1125	10.3	39.6
TRY	3020	27.7	41.9
LYS_C	695	6.4	17.6

the total number of PS identified with a protease. Aim 3 investigates the benefit from adding a LysC protease to the trypsin workflow versus combination of results from these two proteases. And Aim 4 compares the beneficial effect of additional technical replicate and the effect from a parallel workflow with an additional protease. In other word what is more beneficial: to add a technical replicate or to add an experiment that uses another protease?

The total number of PS identified in this study is 11,338. Each protease workflow identified different number of PS (Table 5.1). Interestingly, trypsin workflow identified the most phosphosites, 7,205, while chymotrypsin the least, 1,438. The number of exclusive PS identified with proteases is presented in Table 5.2. Similar pattern is observed: trypsin workflow identified the most unique PS, 3,020 or

Table 5.3 Complementarity of proteases

Times detected	Number of PS	% of all
1	6035	55.3
2	3235	29.7
3	1172	10.7
4	395	3.6
5	70	0.6

27.7% of the total, and chymotrypsin the least – 319 or 2.9%. The complementarity of the workflows was evaluated (Table 5.3). Strikingly, only 0.6% of all identified phosphosites were detected in all five workflows, and 55.3% of the total was identified in only one workflow. These data indicate that protease workflows complement each other in the number of identified phosphosites, the pattern observed before.[179,225]

The Ti^{4+}-IMAC reproducibility, as outlined in the Aim 1, is presented in the Figure 5.2. The phosphosites from each injection replicate were added and then compared to that from another two IMAC replicates. Not surprisingly, with this approach the number of phosphosites that were detected in all three replicates is fairly low, ranging from 36% in lysine experiment to 62% in experiment where trypsin and lysine are subsequently used. On the other hand, by not so strict 2-out-of-3 rule, the reproducibility of the identification

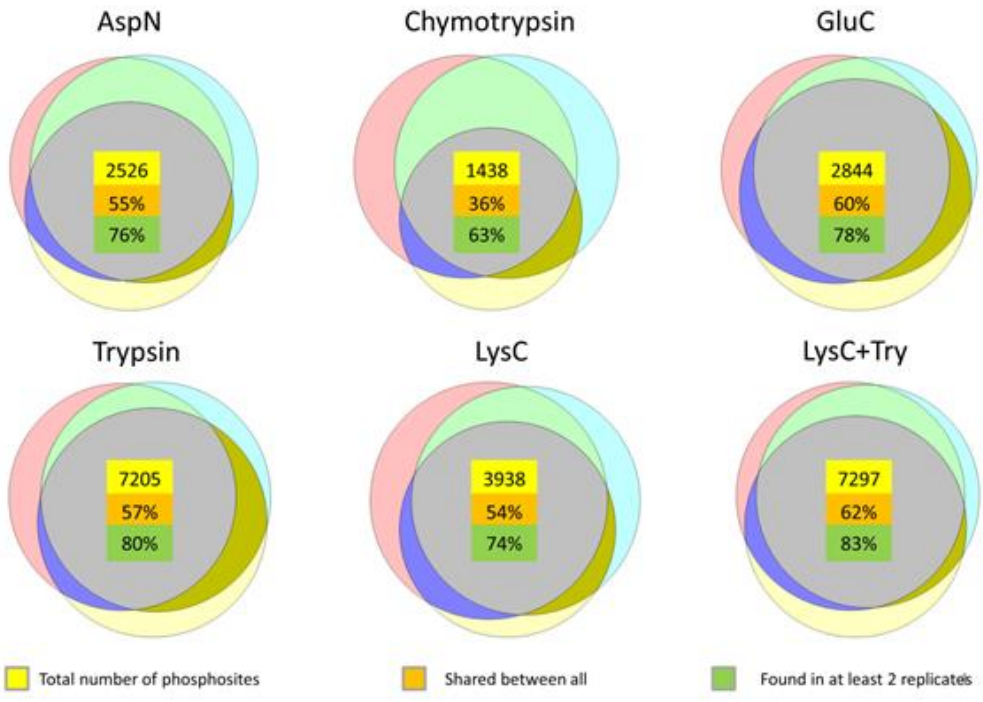


Figure 5.2 Aim 1. Ti^{4+}-IMAC reproducibility.

has risen to 63% and 83% for before mentioned experiments. The number of phosphosites identified and reproducibility of the enrichment was the highest in the experiment where a combination of proteases was used. Lower reproducibility and the number of identified phosphosites for chymotrypsin could be explained by the fact that it produces the shortest peptides compared to the other proteases, therefore the small peptides may not be enriched by the IMAC technique or get lost in the sample preparation steps.

Figure 5.3 represents the results of Aim 2 where number of phosphosites identified in

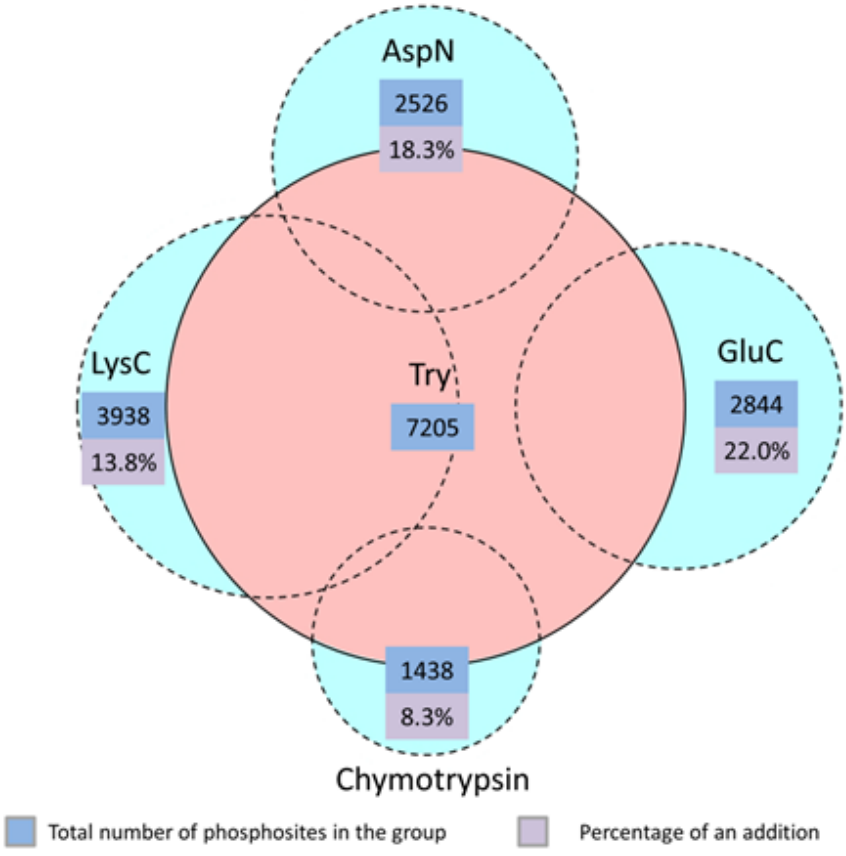


Figure 5.3 Aim 2. Complementarity of protease workflows.

various experiments was compared to the number of phosphosites obtained from the trypsin workflow. The percent of the addition of the number of phosphosites identified from other protease ranges from 8.3% from chymotrypsin, to 22.0 % from GluC workflows. These data supports that from table 3.3 and, once again underscore that the high number of PS identified was only achievable by using workflows employing multiple proteases.

When comparing the PS identified from two workflows with different proteases (namely trypsin and LysC) with a workflow that used the two proteases in a two-step digestion, the Venn diagram does not indicate a strong advantage of one technique over the other (Figure 5.4A). As much as 76% of PS were identified with both techniques. Interestingly, most of the PS in the combined results were identified with trypsin workflow alone (B) indicating that there is no beneficial effect from adding the results of the LysC workflow

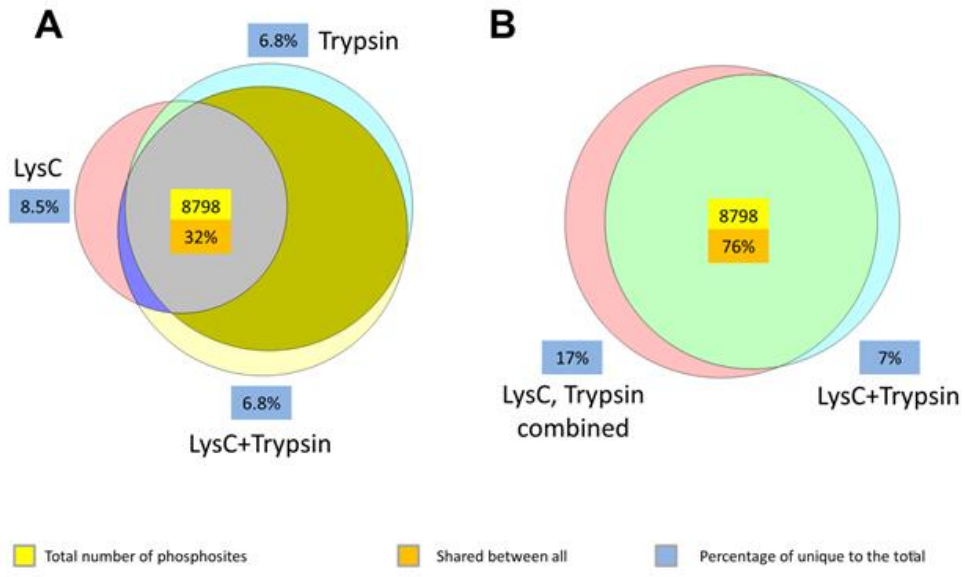


Figure 5.4 Aim 3. Benefit from combining different digestion strategies.

to that of the trypsin workflow.

The results of aim 4 are presented in the Figure 5.5. The left circle represents the number of PS that were detected in the second and third technical replicate in addition to those from the first replicate. The right circle represents the additional number of PS that were detected in the GluC workflow to the total number of PS from trypsin workflow. The data suggests that in order to detect the highest number of identified phosphosites it is beneficial to add both, additional replicates to a workflow and add another workflow using a different protease, in this case GluC showed the highest number of added PS to that of the trypsin workflow.

Next, we evaluated the ratio of phosphorylated amino acids (Table 5.4). Published literature reports that natural representation of the phosphoserine, phosphothreonine and

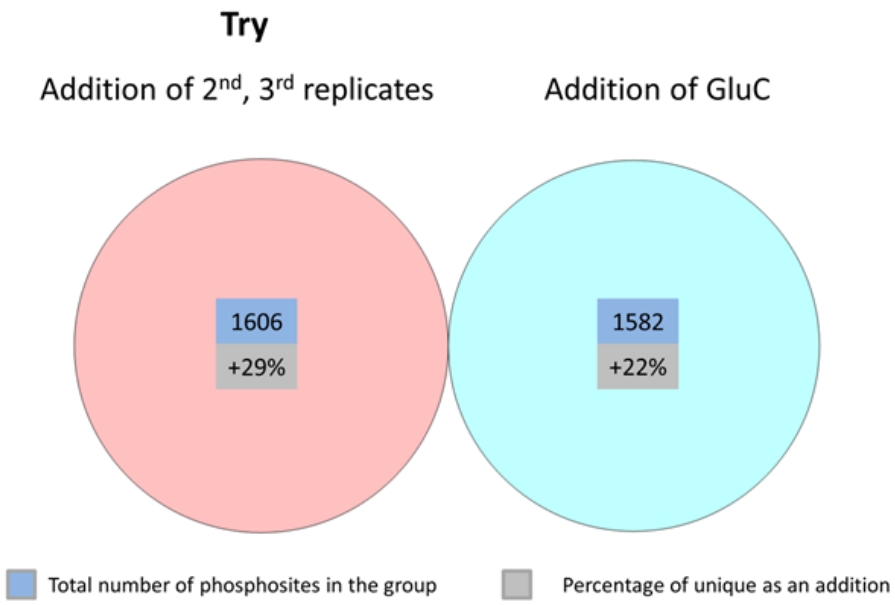


Figure 5.5. Aim 4. The number of PS identification is increased by including experimental replicates and by adding results from a different protease.

Table 5.4 Phosphorylation of residues

	pS:pT:pY	**pS:pT**
AspN	96:17:1	5.6:1
GluC	69:13:1	5.3:1
Chy	43:8:1	5.4:1
LysC	96:15:1	6.2:1
Try	107:16:1	6.7:1
LysC_Try	98:14:1	7:1

phosphotyrosine is 1,800:200:1.[238] From the table 5.4. one can see that the representation of phosphorylation is different from the predicted, which correlates with another study that found ratios 6.75:2:1.[239] This misrepresentation could be described in two ways: the pY is presented in higher abundance than predicted and that pS:pT is underrepresented. Preferential enrichment of pY is unlikely as the immobilized titanium resin enriched all phosphorylated residues equally. Higher abundance of phosphotyrosine could be explained by the cell signaling cascades that are transferred mostly through phosphorylation of tyrosine and over-active in cancer cells. Lower ratio of pS to pT could be attributed to the instability of the two post-translational modifications in the mass-spectrometer.

Conclusions

We described a fast and simple, yet powerful, phosphopeptide enrichment technique, Ti^{4+}-IMAC, with manually packed columns. The method was applied to a protein digests generated from 1 mg of cell lysate by five individual proteases: trypsin, LysC, GluC, AspN, chymotrypsin, and a tandem use of LysC and trypsin. We identified 11,338 phosphosites with high probability, combining all the data generated. The identities of PS reveal that the workflows used are highly complementary, having only 0.6% of PS

identified by all of them. Trypsin digestion alone helped to identify 7,205 PS, the highest number among all proteases. In terms of the added number of PS when applying a parallel protease workflow, GluC has an advantage with 22% increase to the number of PS identified with conventional trypsin workflow. Additional technical replicates added 29% of PS to the first run with trypsin. In sum, the current study underscores that both, multi-replicate and multi-protease workflows are required to improve coverage of the Jurkat T-cell phosphoproteome.

CONCLUSIONS

In this work the application of state-of-art mass spectrometry in combination with ultra-high performance liquid chromatography to study proteomes has been shown. The superiority of this technique is undoubted. We showed that thousands proteins, tens of thousands peptides can be effectively separated and detected in a single run. We used various sample preparation steps to successfully suit needs of different biological questions: from full proteome, to phosphopeptide enrichment and extraction of the ATPome. We present the application of various types of mass spectrometers that specifically satisfy the requirements of a project: qTOF with ion mobility, LTQ FT-ICR and Orbitrap. We presented multiple combinations of software packages to sequence proteins, perform absolute and relative quantification, statistically analyze and visualize data.

Global full-proteome analysis of vitamin C and E deficient zebrafish reveals suppression in an energy metabolism cycle, glycolysis, and activation of an alternative energy cycle, glutaminolysis, to fulfill energy requirement. Application of high resolution qTOF instrument and use of ion mobility with MS^E acquisition resulted in an enhanced separation and identification of peptides. The second project shows the use of a chemical affinity tag for enrichment of ATP binding proteins, aka ATPome, in canine osteosarcoma cell lines. We utilized data dependent acquisition on a quadrupole linear ion trap ion cyclotron resonance mass spectrometer (LTQ-FT) to achieve highly sensitive detection. Results of this investigation led to the hypothesis that in highly metastasizing canine osteosarcoma cell line extracellular adhesion is weakened, causing cells to detach from their neighbors. Many protein kinases were expressed in significantly different

abundance levels indicating changes in cell signaling. The last project introduces a powerful method for phosphosite enrichment and show that application of affinity based enrichment and the use of various proteases during sample preparation in conjunction with an Orbitrap analyzer yields greatly improves the number of identified phosphosites.

In conclusion, mass spectrometry-based proteomics has emerged as a powerful technique to study proteins, post translational modifications, and investigate changes in cell signaling. Mass spectrometry affords high resolution and mass accuracy, selectivity and sensitivity which are only a few advantages of the MS-enabled workflows over others. We show that mass spectrometry-based proteomics excels in high throughput and automation capabilities for protein and peptide sequencing and PTM site mapping.

To further our understanding in biological systems under stress or disease, such as cancer, the mass spectrometry-based proteomics should be applied to *in vivo* models, taking samples of tumor tissues and metastases from canine, and ultimately human, patients. Application of various enrichment techniques will allow us to gain further insight and understanding of the most important sub-proteomes, e.g. protein kinases. Post-translational modifications, including phosphorylation and acetylation, should be studied. Changes in signaling pathways will reveal disturbances in the normal behavior of cells. Ultimately, combination of proteomics, lipidomics and metabolomics in a global systems biology approach will help developing therapeutic targets against cancer and other diseases.

BIBLIOGRAPHY

1. Li X, Wang W, Chen J. From pathways to networks: Connecting dots by establishing protein-protein interaction networks in signaling pathways using affinity purification and mass spectrometry. *Proteomics*. 2015;15(2-3):188-202. doi:10.1002/pmic.201400147.

2. Olsen J V, Schwartz JC, Griep-Raming J, et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics*. 2009;8(12):2759-2769. doi:10.1074/mcp.M900375-MCP200.

3. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;6(5):359-362. doi:10.1038/nmeth.1322.

4. Xiao Y, Wang Y. Global discovery of protein kinases and other nucleotide-binding proteins by mass spectrometry. *Mass Spectrom Rev*. 2014. doi:10.1002/mas.21447.

5. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372. doi:10.1038/nbt.1511.

6. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011;39(Database issue):D561-D568. doi:10.1093/nar/gkq973.

7. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27(3):431-432. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3031041&tool=pmcentrez&rendertype=abstract.

8. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*. 2009;11:49-79. doi:10.1146/annurev-bioeng-061008-124934.

9. Motorykin I, Traber MG, Tanguay RL, Maier CS. Proteome-Driven Elucidation of Adaptive Responses to Combined Vitamin E and C Deficiency in Zebrafish. *J Proteome Res*. 2014;13(3):1647-1656. doi:10.1021/pr401108d.

10. Lorenzo Tejedor M, Mizuno H, Tsuyama N, Harada T, Masujima T. In situ molecular analysis of plant tissues by live single-cell mass spectrometry. *Anal Chem*. 2012;84(12):5221-5228. doi:10.1021/ac202447t.

11. Kirkwood JS, Lebold KM, Miranda CL, et al. Vitamin C deficiency activates the

purine nucleotide cycle in zebrafish. *J Biol Chem*. 2012;287(6):3833-3841. doi:10.1074/jbc.M111.316018.

12. Choudhary C, Kumar C, Gnad F, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009;325(5942):834-840. doi:10.1126/science.1175371.

13. Guo L, Xiao Y, Wang Y. Monomethylarsonous acid inhibited endogenous cholesterol biosynthesis in human skin fibroblasts. *Toxicol Appl Pharmacol*. 2014;277(1):21-29. doi:10.1016/j.taap.2014.02.020.

14. Tao J, Chen S, Lee B. Alteration of Notch signaling in skeletal development and disease. *Ann N Y Acad Sci*. 2010;1192:257-268. doi:10.1111/j.1749-6632.2009.05307.x.

15. Tang Q-L, Xie X-B, Wang J, et al. Glycogen synthase kinase-3β, NF-κB signaling, and tumorigenesis of human osteosarcoma. *J Natl Cancer Inst*. 2012;104(10):749-763. doi:10.1093/jnci/djs210.

16. Holčapek M, Jirásko R, Lísa M. Recent developments in liquid chromatography-mass spectrometry and related techniques. *J Chromatogr A*. 2012;1259:3-15. doi:10.1016/j.chroma.2012.08.072.

17. Wilkins MR, Sanchez JC, Gooley AA, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*. 1996;13:19-50. http://www.ncbi.nlm.nih.gov/pubmed/8948108. Accessed January 14, 2015.

18. Heck AJR, Krijgsveld J. Mass spectrometry-based quantitative proteomics. *Expert Rev Proteomics*. 2004;1(3):317-326. doi:10.1586/14789450.1.3.317.

19. Görg A, Obermaier C, Boguth G, et al. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*. 2000;21(6):1037-1053. doi:10.1002/(SICI)1522-2683(20000401)21:6<1037::AID-ELPS1037>3.0.CO;2-V.

20. Rabilloud T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*. 2002;2(1):3-10. http://www.ncbi.nlm.nih.gov/pubmed/11788986. Accessed February 9, 2015.

21. Hamdan M, Righetti PG. Assessment of protein expression by means of 2-D gel electrophoresis with and without mass spectrometry. *Mass Spectrom Rev*. 2003;22(4):272-284. doi:10.1002/mas.10056.

22. Chelius D, Bondarenko P V. Quantitative profiling of proteins in complex

mixtures using liquid chromatography and mass spectrometry. *J Proteome Res*. 2002;1(4):317-323. http://www.ncbi.nlm.nih.gov/pubmed/12645887. Accessed February 9, 2015.

23. Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, Mann M. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature*. 2003;426(6966):570-574. doi:10.1038/nature02166.

24. Jonckheere JA, Thienpont LM, De Leenheer AP, De Backer P, Debackere M, Belpaire FM. Selected ion monitoring assay for bromhexine in biological fluids. *Biomed Mass Spectrom*. 1980;7(11-12):582-587. http://www.ncbi.nlm.nih.gov/pubmed/7225540. Accessed February 9, 2015.

25. Julka S, Regnier F. Quantification in proteomics through stable isotope coding: a review. *J Proteome Res*. 2004;3(3):350-363. http://www.ncbi.nlm.nih.gov/pubmed/15253416. Accessed February 9, 2015.

26. Lill J. Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom Rev*. 2003;22(3):182-194. doi:10.1002/mas.10048.

27. Ong S-E, Foster LJ, Mann M. Mass spectrometric-based approaches in quantitative proteomics. *Methods*. 2003;29(2):124-130. http://www.ncbi.nlm.nih.gov/pubmed/12606218. Accessed February 9, 2015.

28. Goshe MB, Smith RD. Stable isotope-coded proteomic mass spectrometry. *Curr Opin Biotechnol*. 2003;14(1):101-109. http://www.ncbi.nlm.nih.gov/pubmed/12566009. Accessed February 9, 2015.

29. Sechi S, Oda Y. Quantitative proteomics using mass spectrometry. *Curr Opin Chem Biol*. 2003;7(1):70-77. http://www.ncbi.nlm.nih.gov/pubmed/12547429. Accessed February 9, 2015.

30. Washburn MP, Ulaszek R, Deciu C, Schieltz DM, Yates JR. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem*. 2002;74(7):1650-1657. http://www.ncbi.nlm.nih.gov/pubmed/12043600. Accessed February 10, 2015.

31. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A*. 1999;96(12):6591-6596. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=21959&tool=pmcentrez&rendertype=abstract. Accessed February 10, 2015.

32. Ji J, Chakraborty A, Geng M, et al. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J Chromatogr B Biomed Sci*

Appl. 2000;745(1):197-210. http://www.ncbi.nlm.nih.gov/pubmed/10997715. Accessed February 9, 2015.

33. Geng M, Ji J, Regnier FE. Signature-peptide approach to detecting proteins in complex mixtures. *J Chromatogr A*. 2000;870(1-2):295-313. http://www.ncbi.nlm.nih.gov/pubmed/10722087. Accessed February 9, 2015.

34. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*. 1999;17(10):994-999. doi:10.1038/13690.

35. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004;3(12):1154-1169. doi:10.1074/mcp.M400129-MCP200.

36. Ong S-E, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002;1(5):376-386. http://www.ncbi.nlm.nih.gov/pubmed/12118079. Accessed December 5, 2012.

37. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*. 2003;100(12):6940-6945. doi:10.1073/pnas.0832254100.

38. Stemmann O, Zou H, Gerber SA, Gygi SP, Kirschner MW. Dual inhibition of sister chromatid separation at metaphase. *Cell*. 2001;107(6):715-726. http://www.ncbi.nlm.nih.gov/pubmed/11747808. Accessed February 9, 2015.

39. Burk RF, Christensen JM, Maguire MJ, et al. A combined deficiency of vitamins E and C causes severe central nervous system damage in guinea pigs. *J Nutr*. 2006;136(6):1576-1581.

40. Hill KE, Montine TJ, Motley AK, Li X, May JM, Burk RF. Combined deficiency of vitamins E and C causes paralysis and death in guinea pigs. *Am J Clin Nutr*. 2003;77(6):1484-1488. http://www.ncbi.nlm.nih.gov/pubmed/12791628. Accessed October 25, 2013.

41. Lebold KM, Löhr C V, Barton CL, et al. Chronic vitamin E deficiency promotes vitamin C deficiency in zebrafish leading to degenerative myopathy and impaired swimming behavior. *Comp Biochem Physiol C Toxicol Pharmacol*. 2013;157(4):382-389. doi:10.1016/j.cbpc.2013.03.007.

42. Buettner GR, Schafer FQ. Free radicals, oxidants, and antioxidants. *Teratology*. 2000;62(4):234. doi:10.1002/1096-9926(200010)62:4<234::AID-TERA10>3.0.CO;2-9.

43. Halliwell B. Antioxidant defence mechanisms: from the beginning to the end (of the beginning). *Free Radic Res*. 1999;(31 (4)):261-272.

44. Kalyanaraman B, Darley-Usmar VM, Wood J, Joseph J, Parthasarathy S. Synergistic interaction between the probucol phenoxyl radical and ascorbic acid in inhibiting the oxidation of low density lipoprotein. *J Biol Chem*. 1992;267(10):6789-6795. http://www.ncbi.nlm.nih.gov/pubmed/1313022.

45. Bruno RS, Leonard SW, Atkinson J, et al. Faster plasma vitamin E disappearance in smokers is normalized by vitamin C supplementation. *Free Radic Biol Med*. 2006;40(4):689-697. doi:10.1016/j.freeradbiomed.2005.10.051.

46. Buettner G. The pecking order of free radicals and antioxidants: lipid peroxidation, alpha-tocopherol, and ascorbate. *Arch Biochem Biophys*. 1993;(300 (2)):535-543.

47. Peterkovsky B, Undenfriend S. Enzymatic hydroxylation of proline in microsomal polypeptide leading to formation of collagen. *Proc Natl Acad Sci U S A*. 1965;53:335-342. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=219517&tool=pmcentrez&rendertype=abstract. Accessed August 2, 2013.

48. Myllylä R, Kuutti-Savolainen E-R, Kivirikko KI. The role of ascorbate in the prolyl hydroxylase reaction. *Biochem Biophys Res Commun*. 1978;83(2):441-448. doi:10.1016/0006-291X(78)91010-0.

49. Diaz MN, Frei B, Vita JA, Keaney JF. Antioxidants and atherosclerotic heart disease. *N Engl J Med*. 1997;337(6):408-416. doi:10.1056/NEJM199708073370607.

50. Wilson JX. Mechanism of action of vitamin C in sepsis: ascorbate modulates redox signaling in endothelium. *Biofactors*. 2009;35(1):5-13. doi:10.1002/biof.7.

51. Armour J, Tyml K, Lidington D, Wilson JX. Ascorbate prevents microvascular dysfunction in the skeletal muscle of the septic rat. *J Appl Physiol*. 2001;90(3):795-803. http://jap.physiology.org/content/90/3/795.long. Accessed August 2, 2013.

52. Piyathilake CJ, Macaluso M, Hine RJ, Vinter DW, Richards EW, Krumdieck CL. Cigarette smoking, intracellular vitamin deficiency, and occurrence of micronuclei in epithelial cells of the buccal mucosa. *Cancer Epidemiol Biomarkers Prev*. 1995;4(7):751-758. http://www.ncbi.nlm.nih.gov/pubmed/8672992. Accessed August 14, 2013.

53. Frikke-Schmidt H, Tveden-Nyborg P, Birck MM, Lykkesfeldt J. High dietary fat and cholesterol exacerbates chronic vitamin C deficiency in guinea pigs. *Br J Nutr*.

2011;105(1):54-61. doi:10.1017/S0007114510003077.

54. Bánhegyi G, Braun L, Csala M, Puskás F, Mandl J. Ascorbate metabolism and its regulation in animals. *Free Radic Biol Med*. 1997;23(5):793-803. http://www.sciencedirect.com/science/article/B6T38-3SG8RWD-F/1/a4086f55ae2e66bc2579d5dfb0984a8f.

55. Chatterjee IB, Majumder AK, Nandi BK SN. Synthesis and some major functions of vitamin C in animals. *Ann N Y Acad Sci*. 1975;(258):24-47. doi:doi: 10.1111/j.1749-6632.1975.tb29266.x.

56. Delanghe JR, Langlois MR, De Buyzere ML, et al. Vitamin C deficiency: more than just a nutritional disorder. *Genes Nutr*. 2011;6(4):341-346. doi:10.1007/s12263-011-0237-7.

57. Nishikimi M, Fukuyama R, Minoshima S, Shimizu N, Yagi K. Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J Biol Chem*. 1994;269(18):13685-13688. http://www.ncbi.nlm.nih.gov/pubmed/8175804.

58. Grollman AP, Lehninger AL. Enzymic synthesis of l-ascorbic acid in different animal species. *Arch Biochem Biophys*. 1957;69(null):458-467. doi:10.1016/0003-9861(57)90510-6.

59. Harrison FE, Meredith ME, Dawes SM, Saskowski JL, May JM. Low ascorbic acid and increased oxidative stress in gulo(-/-) mice during development. *Brain Res*. 2010;1349:143-152. doi:10.1016/j.brainres.2010.06.037.

60. Telang S, Clem AL, Eaton JW, Chesney J. Depletion of Ascorbic Acid Restricts Angiogenesis and Retards Tumor Growth in a Mouse Model. *Neoplasia*. 2007;9(1):47-56. doi:10.1593/neo.06664.

61. Dabrowski K. Gulonolactone oxidase is missing in teleost fish. The direct spectrophotometric assay. *Biol Chem Hoppe Seyler*. 1990;371(3):207-214. http://www.ncbi.nlm.nih.gov/pubmed/2340104. Accessed October 25, 2013.

62. Touhata K, Toyohara H, Mitani T, Kinoshita M, Satou M, Sakaguchi M. Distribution of L-Gulono-1, 4-Lactone Oxidase among Fishes. *Fish Sci*. 1995;61:729-730.

63. Toyohara H, Nakata T, Touhata K, et al. Transgenic expression of L-gulono-gamma-lactone oxidase in medaka (Oryzias latipes), a teleost fish that lacks this enzyme necessary for L-ascorbic acid biosynthesis. *Biochem Biophys Res Commun*. 1996;223(3):650-653. doi:10.1006/bbrc.1996.0949.

64. Phromkunthong W, Storch V, Braunbeckw T. Sexual dimorphism in the reaction of zebrafish (Brachydanio rerio) to ascorbic acid deficiency: Induction of steatosis in hepatocytes of male fish. *J Appl Ichthyol*. 1994;10(2-3):146-153. doi:10.1111/j.1439-0426.1994.tb00154.x.

65. Dabrowski K, Ciereszko A. Ascorbic acid protects against male infertility in a teleost fish. *Experientia*. 1996;52(2):97-100. http://www.ncbi.nlm.nih.gov/pubmed/8608823. Accessed October 25, 2013.

66. Drouin G, Godin J-R, Pagé B. The genetics of vitamin C loss in vertebrates. *Curr Genomics*. 2011;12(5):371-378. doi:10.2174/138920211796429736.

67. Brigelius-Flohe R, Traber MG. Vitamin E: function and metabolism. *FASEB J*. 1999;13(10):1145-1155. http://www.fasebj.org/content/13/10/1145.full. Accessed February 27, 2013.

68. Traber MG, Atkinson J. Vitamin E, antioxidant and nothing more. *Free Radic Biol Med*. 2007;43(1):4-15. doi:10.1016/j.freeradbiomed.2007.03.024.

69. McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D YJ 3rd. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem*. 1997;69(4):767-776.

70. Peng J, Elias JE, Thoreen CC, Licklider LJ GS. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res*. 2003;2(1):43-50.

71. Voyksner RD, Lee H. Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Commun Mass Spectrom*. 1999;13(14):1427-1437. doi:10.1002/(SICI)1097-0231(19990730)13:14<1427::AID-RCM662>3.0.CO;2-5.

72. Purves RW, Gabryelski W, Li L. Investigation of the quantitative capabilities of an electrospray ionization ion trap/linear time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom*. 1998;12(11):695-700. doi:10.1002/(SICI)1097-0231(19980615)12:11<695::AID-RCM220>3.0.CO;2-2.

73. Croker CG, Pearcy JO, Stahl DC, Moore RE, Keen DA, Lee TD. LC / MS / MS Analysis of Proteins. *J Biomol Tech*. 2000;11(3):135-141.

74. Plumb RS, Johnson KA, Rainville P, et al. UPLC / MS E ; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun Mass Spectrom*. 2006;(20):1989-1994. doi:10.1002/rcm.

75. Geromanos SJ, Vissers JPC, Silva JC, et al. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics*. 2009;9(6):1683-1695. doi:10.1002/pmic.200800562.

76. Silva JC, Denny R, Dorschel C, et al. Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. *Mol Cell Proteomics*. 2006;5(4):589-607. doi:10.1074/mcp.M500321-MCP200.

77. Ritter S. New and notable at Pittcon. *Chem Eng news*. 2007;85(12):61-65. doi:10.1021/cen-v085n012.p061.

78. Salbo R, Bush MF, Naver H, et al. Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun Mass Spectrom*. 2012;26(10):1181-1193. doi:10.1002/rcm.6211.

79. Hoaglund-Hyzer CS, Li J, Clemmer DE. Mobility labeling for parallel CID of ion mixtures. *Anal Chem*. 2000;72(13):2737-2740. http://www.ncbi.nlm.nih.gov/pubmed/10905301.

80. Miller GW, Labut EM, Lebold KM, Floeter A, Tanguay RL, Traber MG. Zebrafish (Danio rerio) fed vitamin E-deficient diets produce embryos with increased morphologic abnormalities and mortality. *J Nutr Biochem*. 2011. doi:10.1016/j.jnutbio.2011.02.002.

81. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002;74(20):5383-5392. http://www.ncbi.nlm.nih.gov/pubmed/12403597.

82. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003;75(17):4646-4658. http://www.ncbi.nlm.nih.gov/pubmed/14632076.

83. Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in proteome analysis. *Proteomics*. 2003;3(8):1397-1407. doi:10.1002/pmic.200300472.

84. Ramsey FL, Schafer DW. *The Statistical Sleuth: A Course in Methods of Data Analysis*. Cengage Learning; 2012. http://books.google.com/books?id=eSlLjA9TwkUC&pgis=1. Accessed March 11, 2013.

85. Azzi A, Breyer I, Feher M, et al. Nonantioxidant functions of alpha-tocopherol in

smooth muscle cells. *J Nutr*. 2001;131(2):378S - 81S. http://www.ncbi.nlm.nih.gov/pubmed/11160565. Accessed October 25, 2013.

86. Jurica MS, Mesecar A, Heath PJ, Shi W, Nowak T, Stoddard BL. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure*. 1998;6(2):195-210. http://www.ncbi.nlm.nih.gov/pubmed/9519410. Accessed November 14, 2012.

87. Mazurek S, Boschek CB, Hugo F, Eigenbrodt E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin Cancer Biol*. 2005;15(4):300-308. doi:10.1016/j.semcancer.2005.04.009.

88. DeBerardinis RJ, Mancuso A, Daikhin E, et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci U S A*. 2007;104(49):19345-19350. doi:10.1073/pnas.0709747104.

89. Le A, Lane AN, Hamaker M, et al. Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. *Cell Metab*. 2012;15(1):110-121. doi:10.1016/j.cmet.2011.12.009.

90. Vos PE, Lamers KJB, Hendriks JCM, et al. Glial and neuronal proteins in serum predict outcome after severe traumatic brain injury. *Neurology*. 2004;62(8):1303-1310. http://www.ncbi.nlm.nih.gov/pubmed/15111666. Accessed November 14, 2012.

91. Eng LF, Ghirnikar RS, Lee YL. Glial fibrillary acidic protein: GFAP-thirty-one years (1969-2000). *Neurochem Res*. 2000;25(9-10):1439-1451. http://www.ncbi.nlm.nih.gov/pubmed/11059815. Accessed November 11, 2012.

92. Schiff L, Hadker N, Weiser S, Rausch C. A literature review of the feasibility of glial fibrillary acidic protein as a biomarker for stroke and traumatic brain injury. *Mol Diagn Ther*. 2012;16(2):79-92. doi:10.2165/11631580-000000000-00000.

93. Mirabello L, Pfeiffer R, Murphy G, et al. Height at diagnosis and birth-weight as risk factors for osteosarcoma. *Cancer Causes Control*. 2011;22(6):899-908. doi:10.1007/s10552-011-9763-2.

94. Kaste SC, Pratt CB, Cain AM, Jones-Wallace DJ, Rao BN. Metastases detected at the time of diagnosis of primary pediatric extremity osteosarcoma at diagnosis: imaging features. *Cancer*. 1999;86(8):1602-1608. http://www.ncbi.nlm.nih.gov/pubmed/10526292. Accessed September 22, 2014.

95. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results

Program. *Cancer*. 2009;115(7):1531-1543. doi:10.1002/cncr.24121.

96. Cai Z, Chiu J-F, He Q-Y. Application of proteomics in the study of tumor metastasis. *Genomics Proteomics Bioinformatics*. 2004;2(3):152-166. http://www.ncbi.nlm.nih.gov/pubmed/15862116. Accessed September 22, 2014.

97. Masui O, White NM a, DeSouza L V, et al. Quantitative proteomic analysis in metastatic renal cell carcinoma reveals a unique set of proteins with potential prognostic significance. *Mol Cell Proteomics*. 2013;12(1):132-144. doi:10.1074/mcp.M112.020701.

98. Rankin KS, Starkey M, Lunec J, Gerrand CH, Murphy S, Biswas S. Of dogs and men: comparative biology as a tool for the discovery of novel biomarkers and drug development targets in osteosarcoma. *Pediatr Blood Cancer*. 2012;58(3):327-333. doi:10.1002/pbc.23341.

99. Withrow SJ, Powers BE, Straw RC, Wilkins RM. Comparative aspects of osteosarcoma. Dog versus man. *Clin Orthop Relat Res*. 1991;(270):159-168. http://www.ncbi.nlm.nih.gov/pubmed/1884536. Accessed September 22, 2014.

100. Dernell WS. Tumors of the Skeletal System. In: Withrow SJ, Vail DM, eds. *Withrow & MacEwen's Small Animal Clinical Oncology*. Saunders Elsevier: St. Louis, MO; 2007:540-582.

101. Paoloni M, Davis S, Lana S, et al. Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. *BMC Genomics*. 2009;10(1):625. doi:10.1186/1471-2164-10-625.

102. Chen Y-R, Juan H-F, Huang H-C, et al. Quantitative proteomic and genomic profiling reveals metastasis-related protein expression patterns in gastric cancer cells. *J Proteome Res*. 2006;5(10):2727-2742. doi:10.1021/pr060212g.

103. Ding S-J, Li Y, Shao X-X, et al. Proteome analysis of hepatocellular carcinoma cell strains, MHCC97-H and MHCC97-L, with different metastasis potentials. *Proteomics*. 2004;4(4):982-994. doi:10.1002/pmic.200300653.

104. Choi D-S, Choi D-Y, Hong BS, et al. Quantitative proteomics of extracellular vesicles derived from human primary and metastatic colorectal cancer cells. *J Extracell vesicles*. 2012;1. doi:10.3402/jev.v1i0.18704.

105. O'Donoghue L, Ptitsyn A. Expression profiling in canine osteosarcoma: identification of biomarkers and pathways associated with outcome. *BMC Cancer*. 2010;(10). http://www.biomedcentral.com/1471-2407/10/506. Accessed April 22, 2013.

106. Fischer J de S da G, Canedo NHS, Goncalves KM da S, et al. Proteome analysis of formalin-fixed paraffin-embedded tissues from a primary gastric melanoma and its meningeal metastasis: a case report. *Curr Top Med Chem*. 2014;14(3):382-387. http://www.ncbi.nlm.nih.gov/pubmed/24304316. Accessed March 25, 2015.

107. Dai B, Cai X, Kong Y-Y, et al. Analysis of KIT expression and gene mutation in human acral melanoma: with a comparison between primary tumors and corresponding metastases/recurrences. *Hum Pathol*. 2013;44(8):1472-1478. doi:10.1016/j.humpath.2013.01.007.

108. Thomas R, Wang HJ, Tsai P-C, et al. Influence of genetic background on tumor karyotypes: evidence for breed-associated cytogenetic aberrations in canine appendicular osteosarcoma. *Chromosome Res*. 2009;17(3):365-377. doi:10.1007/s10577-009-9028-z.

109. Knowles JR. Enzyme-catalyzed phosphoryl transfer reactions. *Annu Rev Biochem*. 1980;49:877-919. doi:10.1146/annurev.bi.49.070180.004305.

110. Villamor JG, Kaschani F, Colby T, et al. Profiling protein kinases and other ATP binding proteins in Arabidopsis using Acyl-ATP probes. *Mol Cell Proteomics*. 2013;12(9):2481-2496. doi:10.1074/mcp.M112.026278.

111. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002;298(5600):1912-1934. doi:10.1126/science.1075762.

112. Pulford K, Lamant L, Morris SW, et al. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood*. 1997;89(4):1394-1404. http://www.ncbi.nlm.nih.gov/pubmed/9028963. Accessed March 27, 2015.

113. Oppermann FS, Gnad F, Olsen J V, et al. Large-scale proteomics analysis of the human kinome. *Mol Cell Proteomics*. 2009;8(7):1751-1764. doi:10.1074/mcp.M800588-MCP200.

114. Bartlett S, Beddard GS, Jackson RM, et al. Comparison of the ATP binding sites of protein kinases using conformationally diverse bisindolylmaleimides. *J Am Chem Soc*. 2005;127(33):11699-11708. doi:10.1021/ja050576u.

115. Xiao Y, Guo L, Wang Y. Isotope-coded ATP probe for quantitative affinity profiling of ATP-binding proteins. *Anal Chem*. 2013;85(15):7478-7486. doi:10.1021/ac401415z.

116. Patricelli MP, Nomanbhoy TK, Wu J, et al. In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem Biol*. 2011;18(6):699-710.

doi:10.1016/j.chembiol.2011.04.011.

117. Lemeer S, Pinkse MWH, Mohammed S, et al. Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *J Proteome Res*. 2008;7(4):1555-1564. doi:10.1021/pr700667w.

118. Lemeer S, Zörgiebel C, Ruprecht B, Kohl K, Kuster B. Comparing immobilized kinase inhibitors and covalent ATP probes for proteomic profiling of kinase expression and drug selectivity. *J Proteome Res*. 2013;12(4):1723-1731. doi:10.1021/pr301073j.

119. Xiao Y, Guo L, Wang Y. A targeted quantitative proteomics strategy for global kinome profiling of cancer cells and tissues. *Mol Cell Proteomics*. 2014;13(4):1065-1075. doi:10.1074/mcp.M113.036905.

120. Xiao Y, Guo L, Jiang X, Wang Y. Proteome-wide discovery and characterizations of nucleotide-binding proteins with affinity-labeled chemical probes. *Anal Chem*. 2013;85(6):3198-3206. doi:10.1021/ac303383c.

121. Wolfe LM, Veeraraghavan U, Idicula-Thomas S, et al. A chemical proteomics approach to profiling the ATP-binding proteome of Mycobacterium tuberculosis. *Mol Cell Proteomics*. 2013;12(6):1644-1660. doi:10.1074/mcp.M112.025635.

122. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J*. 1995;9(8):576-596. http://www.ncbi.nlm.nih.gov/pubmed/7768349. Accessed March 5, 2015.

123. Patricelli MP, Szardenings AK, Liyanage M, et al. Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry*. 2007;46(2):350-358. doi:10.1021/bi062142x.

124. Adachi J, Kishida M, Watanabe S, Hashimoto Y, Fukamizu K, Tomonaga T. Proteome-wide discovery of unknown ATP-binding proteins and kinase inhibitor target proteins using an ATP probe. *J Proteome Res*. 2014. doi:10.1021/pr500845u.

125. Marx V. Targeted proteomics. *Nat Methods*. 2013;10(1):19-22. http://www.ncbi.nlm.nih.gov/pubmed/23547293. Accessed March 27, 2015.

126. Barroga EF, Kadosawa T, Okumura M, Fujinaga T. Establishment and characterization of the growth and pulmonary metastasis of a highly lung metastasizing cell line from canine osteosarcoma in nude mice. *J Vet Med Sci*. 1999;61(4):361-367. http://www.ncbi.nlm.nih.gov/pubmed/10342286. Accessed February 12, 2015.

127. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res*. 2000;28(18):3442-3444. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=110752&tool=pmcentrez&rendertype=abstract. Accessed February 10, 2015.

128. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41(Database issue):D808-D815. doi:10.1093/nar/gks1094.

129. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102409&tool=pmcentrez&rendertype=abstract. Accessed July 10, 2014.

130. KEGG: Kyoto Encyclopedia of Genes and Genomes. http://www.kegg.jp/.

131. Menke A, Giehl K. Regulation of adherens junctions by Rho GTPases and p120-catenin. *Arch Biochem Biophys*. 2012;524(1):48-55. doi:10.1016/j.abb.2012.04.019.

132. Kemler R. From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends Genet*. 1993;9(9):317-321. http://www.ncbi.nlm.nih.gov/pubmed/8236461. Accessed January 12, 2015.

133. Drees F, Pokutta S, Yamada S, Nelson WJ, Weis WI. Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly. *Cell*. 2005;123(5):903-915. doi:10.1016/j.cell.2005.09.021.

134. Meng W, Takeichi M. Adherens junction: molecular architecture and regulation. *Cold Spring Harb Perspect Biol*. 2009;1(6):a002899. doi:10.1101/cshperspect.a002899.

135. Braga VM, Yap AS. The challenges of abundance: epithelial junctions and small GTPase signalling. *Curr Opin Cell Biol*. 2005;17(5):466-474. doi:10.1016/j.ceb.2005.08.012.

136. Samarin S, Nusrat A. Regulation of epithelial apical junctional complex by Rho family GTPases. *Front Biosci (Landmark Ed*. 2009;14:1129-1142. http://www.ncbi.nlm.nih.gov/pubmed/19273120. Accessed January 12, 2015.

137. Watanabe T, Sato K, Kaibuchi K. Cadherin-mediated intercellular adhesion and signaling cascades involving small GTPases. *Cold Spring Harb Perspect Biol*. 2009;1(3):a003020. doi:10.1101/cshperspect.a003020.

138. Ridley AJ. Rho proteins: linking signaling with membrane trafficking. *Traffic*. 2001;2(5):303-310. http://www.ncbi.nlm.nih.gov/pubmed/11350626. Accessed January 12, 2015.

139. Coleman ML, Marshall CJ, Olson MF. RAS and RHO GTPases in G1-phase cell-cycle regulation. *Nat Rev Mol Cell Biol*. 2004;5(5):355-366. doi:10.1038/nrm1365.

140. Jaffe AB, Hall A. Rho GTPases: biochemistry and biology. *Annu Rev Cell Dev Biol*. 2005;21:247-269. doi:10.1146/annurev.cellbio.21.020604.150721.

141. Sahai E, Marshall CJ. RHO-GTPases and cancer. *Nat Rev Cancer*. 2002;2(2):133-142. doi:10.1038/nrc725.

142. Wennerberg K, Der CJ. Rho-family GTPases: it's not only Rac and Rho (and I like it). *J Cell Sci*. 2004;117(Pt 8):1301-1312. doi:10.1242/jcs.01118.

143. Citi S, Spadaro D, Schneider Y, Stutz J, Pulimeno P. Regulation of small GTPases at epithelial cell-cell junctions. *Mol Membr Biol*. 2011;28(7-8):427-444.

144. Baum B, Georgiou M. Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling. *J Cell Biol*. 2011;192(6):907-917. doi:10.1083/jcb.201009141.

145. Citi S, Spadaro D, Schneider Y, Stutz J, Pulimeno P. Regulation of small GTPases at epithelial cell-cell junctions. *Mol Membr Biol*. 2011;28(7-8):427-444. doi:10.3109/09687688.2011.603101.

146. Nusrat A, Giry M, Turner JR, et al. Rho protein regulates tight junctions and perijunctional actin organization in polarized epithelia. *Proc Natl Acad Sci U S A*. 1995;92(23):10629-10633. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=40665&tool=pmcentrez&rendertype=abstract. Accessed January 12, 2015.

147. Braga VM, Machesky LM, Hall A, Hotchin NA. The small GTPases Rho and Rac are required for the establishment of cadherin-dependent cell-cell contacts. *J Cell Biol*. 1997;137(6):1421-1431. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2132529&tool=pmcentrez&rendertype=abstract. Accessed January 12, 2015.

148. Takaishi K, Sasaki T, Kotani H, Nishioka H, Takai Y. Regulation of cell-cell adhesion by rac and rho small G proteins in MDCK cells. *J Cell Biol*. 1997;139(4):1047-1059. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2139955&tool=pmcentrez&rendertype=abstract. Accessed January 12, 2015.

149. Jou TS, Nelson WJ. Effects of regulated expression of mutant RhoA and Rac1 small GTPases on the development of epithelial (MDCK) cell polarity. *J Cell Biol*. 1998;142(1):85-100. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2133034&tool=pmcentrez&rendertype=abstract. Accessed January 12, 2015.

150. Bruewer M, Hopkins AM, Hobert ME, Nusrat A, Madara JL. RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. *Am J Physiol Cell Physiol*. 2004;287(2):C327-C335. doi:10.1152/ajpcell.00087.2004.

151. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. *Nature*. 2002;420(6916):629-635. doi:10.1038/nature01148.

152. Popoff MR, Geny B. Multifaceted role of Rho, Rac, Cdc42 and Ras in intercellular junctions, lessons from toxins. *Biochim Biophys Acta*. 2009;1788(4):797-812. doi:10.1016/j.bbamem.2009.01.011.

153. Braga VM, Del Maschio A, Machesky L, Dejana E. Regulation of cadherin function by Rho and Rac: modulation by junction maturation and cellular context. *Mol Biol Cell*. 1999;10(1):9-22. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=25150&tool=pmcentrez&rendertype=abstract. Accessed January 12, 2015.

154. Sahai E, Marshall CJ. ROCK and Dia have opposing effects on adherens junctions downstream of Rho. *Nat Cell Biol*. 2002;4(6):408-415. doi:10.1038/ncb796.

155. Shewan AM, Maddugoda M, Kraemer A, et al. Myosin 2 is a key Rho kinase target necessary for the local concentration of E-cadherin at cell-cell contacts. *Mol Biol Cell*. 2005;16(10):4531-4542. doi:10.1091/mbc.E05-04-0330.

156. Michaelson D, Silletti J, Murphy G, D'Eustachio P, Rush M, Philips MR. Differential localization of Rho GTPases in live cells: regulation by hypervariable regions and RhoGDI binding. *J Cell Biol*. 2001;152(1):111-126. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2193662&tool=pmcentrez&rendertype=abstract. Accessed January 13, 2015.

157. Boulter E, Garcia-Mata R, Guilluy C, et al. Regulation of Rho GTPase crosstalk, degradation and activity by RhoGDI1. *Nat Cell Biol*. 2010;12(5):477-483. doi:10.1038/ncb2049.

158. Zhu Y, Liu C, Tummala R, Nadiminty N, Lou W, Gao AC. RhoGDIα downregulates androgen receptor signaling in prostate cancer cells. *Prostate*. 2013;73(15):1614-1622. doi:10.1002/pros.22615.

159. Zhu Y, Tummala R, Liu C, et al. RhoGDIα suppresses growth and survival of prostate cancer cells. *Prostate*. 2012;72(4):392-398. doi:10.1002/pros.21441.

160. Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. *Cell*. 2007;129(7):1261-1274. doi:10.1016/j.cell.2007.06.009.

161. Rommel C, Clarke BA, Zimmermann S, et al. Differentiation stage-specific inhibition of the Raf-MEK-ERK pathway by Akt. *Science*. 1999;286(5445):1738-1741. http://www.ncbi.nlm.nih.gov/pubmed/10576741. Accessed January 28, 2015.

162. Zimmermann S, Moelling K. Phosphorylation and regulation of Raf by Akt (protein kinase B). *Science*. 1999;286(5445):1741-1744. http://www.ncbi.nlm.nih.gov/pubmed/10576742. Accessed January 8, 2015.

163. Kim AH, Khursigara G, Sun X, Franke TF, Chao M V. Akt phosphorylates and negatively regulates apoptosis signal-regulating kinase 1. *Mol Cell Biol*. 2001;21(3):893-901. doi:10.1128/MCB.21.3.893-901.2001.

164. Cross DA, Alessi DR, Cohen P, Andjelkovich M, Hemmings BA. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature*. 1995;378(6559):785-789. doi:10.1038/378785a0.

165. Diehl JA, Cheng M, Roussel MF, Sherr CJ. Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev*. 1998;12(22):3499-3511. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=317244&tool=pmcentrez&rendertype=abstract. Accessed January 27, 2015.

166. Welcker M, Singer J, Loeb KR, et al. Multisite phosphorylation by Cdk2 and GSK3 controls cyclin E degradation. *Mol Cell*. 2003;12(2):381-392. http://www.ncbi.nlm.nih.gov/pubmed/14536078. Accessed January 28, 2015.

167. Zannini L, Delia D, Buscemi G. CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol*. 2014;6(6):442-457. doi:10.1093/jmcb/mju045.

168. Falck J, Mailand N, Syljuåsen RG, Bartek J, Lukas J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature*. 2001;410(6830):842-847. doi:10.1038/35071124.

169. Pawson T, Scott JD. Protein phosphorylation in signaling--50 years and counting. *Trends Biochem Sci*. 2005;30(6):286-290. doi:10.1016/j.tibs.2005.04.013.

170. Hunter T. Tyrosine phosphorylation: thirty years and counting. *Curr Opin Cell Biol*. 2009;21(2):140-146. doi:10.1016/j.ceb.2009.01.028.

171. Brognard J, Hunter T. Protein kinase signaling networks in cancer. *Curr Opin Genet Dev*. 2011;21(1):4-11. doi:10.1016/j.gde.2010.10.012.

172. Harsha HC, Pandey A. Phosphoproteomics in cancer. *Mol Oncol*. 2010;4(6):482-495. doi:10.1016/j.molonc.2010.09.004.

173. Huttlin EL, Jedrychowski MP, Elias JE, et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*. 2010;143(7):1174-1189. doi:10.1016/j.cell.2010.12.001.

174. Olsen J V, Blagoev B, Gnad F, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*. 2006;127(3):635-648. doi:10.1016/j.cell.2006.09.026.

175. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207. doi:10.1038/nature01511.

176. Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. 2011;80:273-299. doi:10.1146/annurev-biochem-061308-093216.

177. Zhou H, Di Palma S, Preisinger C, et al. Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J Proteome Res*. 2013;12(1):260-271. doi:10.1021/pr300630k.

178. Steen H, Jebanathirajah JA, Rush J, Morrice N, Kirschner MW. Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics*. 2006;5(1):172-181. doi:10.1074/mcp.M500135-MCP200.

179. Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJR, Mohammed S. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem*. 2009;81(11):4493-4501. doi:10.1021/ac9004309.

180. Covey T, Shushan B, Bonner R, et al. No Title. *Methods Protein Seq Anal*. 1991:249-256.

181. Carr SA, Huddleston MJ, Annan RS. Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal Biochem*. 1996;239(2):180-192. doi:10.1006/abio.1996.0313.

182. Steen H, Küster B, Mann M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J Mass Spectrom*. 2001;36(7):782-790. doi:10.1002/jms.174.

183. Gruhler A, Olsen J V, Mohammed S, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics*. 2005;4(3):310-327. doi:10.1074/mcp.M400219-MCP200.

184. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem*. 2004;76(13):3590-3598. doi:10.1021/ac0497104.

185. Unwin RD, Griffiths JR, Leverentz MK, Grallert A, Hagan IM, Whetton AD. Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics*. 2005;4(8):1134-1144. doi:10.1074/mcp.M500113-MCP200.

186. Chi A, Huttenhower C, Geer LY, et al. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A*. 2007;104(7):2193-2198. doi:10.1073/pnas.0607084104.

187. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A*. 2007;104(7):2199-2204. doi:10.1073/pnas.0611217104.

188. Taouatas N, Altelaar AFM, Drugan MM, Helbig AO, Mohammed S, Heck AJR. Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. *Mol Cell Proteomics*. 2009;8(1):190-200. doi:10.1074/mcp.M800285-MCP200.

189. Boersema PJ, Mohammed S, Heck AJR. Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom*. 2009;44(6):861-878. doi:10.1002/jms.1599.

190. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics*. 2011;10(12):M111.009910. doi:10.1074/mcp.M111.009910.

191. Nagaraj N, D'Souza RCJ, Cox J, Olsen J V, Mann M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res*. 2010;9(12):6786-6794. doi:10.1021/pr100637q.

192. Zhou H, Low TY, Hennrich ML, et al. Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol Cell Proteomics*. 2011;10(10):M110.006452.

doi:10.1074/mcp.M110.006452.

193. Swaney DL, Wenger CD, Thomson JA, Coon JJ. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A*. 2009;106(4):995-1000. doi:10.1073/pnas.0811964106.

194. Michalski A, Damoc E, Lange O, et al. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics*. 2012;11(3):O111.013698. doi:10.1074/mcp.O111.013698.

195. Nühse TS, Stensballe A, Jensen ON, Peck SC. Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics*. 2003;2(11):1234-1243. doi:10.1074/mcp.T300006-MCP200.

196. Dai J, Jin W-H, Sheng Q-H, Shieh C-H, Wu J-R, Zeng R. Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J Proteome Res*. 2007;6(1):250-262. doi:10.1021/pr0604155.

197. Beausoleil SA, Jedrychowski M, Schwartz D, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A*. 2004;101(33):12130-12135. doi:10.1073/pnas.0404720101.

198. McNulty DE, Annan RS. In 55th ASMS Conference on Mass Spectrometry. 2007.

199. Hung C-W, Kübler D, Lehmann WD. pI-based phosphopeptide enrichment combined with nanoESI-MS. *Electrophoresis*. 2007;28(12):2044-2052. doi:10.1002/elps.200600678.

200. Rikova K, Guo A, Zeng Q, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*. 2007;131(6):1190-1203. doi:10.1016/j.cell.2007.11.025.

201. Luo W, Slebos RJ, Hill S, et al. Global impact of oncogenic Src on a phosphotyrosine proteome. *J Proteome Res*. 2008;7(8):3447-3460. doi:10.1021/pr800187n.

202. Pandey A, Podtelejnikov A V, Blagoev B, Bustelo XR, Mann M, Lodish HF. Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci U S A*. 2000;97(1):179-184. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=26636&tool=pmcentre

z&rendertype=abstract. Accessed July 17, 2014.

203. Grønborg M, Kristiansen TZ, Stensballe A, et al. A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics*. 2002;1(7):517-527. http://www.ncbi.nlm.nih.gov/pubmed/12239280. Accessed July 17, 2014.

204. Zheng H, Hu P, Quinn DF, Wang YK. Phosphotyrosine proteomic study of interferon alpha signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol Cell Proteomics*. 2005;4(6):721-730. doi:10.1074/mcp.M400077-MCP200.

205. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem*. 1999;71(14):2883-2892. http://www.ncbi.nlm.nih.gov/pubmed/10424175. Accessed July 17, 2014.

206. Stensballe A, Andersen S, Jensen ON. Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics*. 2001;1(2):207-222.

207. Ficarro SB, McCleland ML, Stukenberg PT, et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol*. 2002;20(3):301-305. doi:10.1038/nbt0302-301.

208. Chen C-T, Chen W-Y, Tsai P-J, Chien K-Y, Yu J-S, Chen Y-C. Rapid enrichment of phosphopeptides and phosphoproteins from complex samples using magnetic particles coated with alumina as the concentrating probes for MALDI MS analysis. *J Proteome Res*. 2007;6(1):316-325. doi:10.1021/pr0604460.

209. Kweon HK, Håkansson K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem*. 2006;78(6):1743-1749. doi:10.1021/ac0522355.

210. Pinkse MWH, Mohammed S, Gouw JW, van Breukelen B, Vos HR, Heck AJR. Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J Proteome Res*. 2008;7(2):687-697. doi:10.1021/pr700605z.

211. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem*. 2004;76(14):3935-3943. doi:10.1021/ac0498617.

212. Sugiyama N, Masuda T, Shinoda K, Nakamura A, Tomita M, Ishihama Y.

Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol Cell Proteomics*. 2007;6(6):1103-1109. doi:10.1074/mcp.T600060-MCP200.

213. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jørgensen TJD. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics*. 2005;4(7):873-886. doi:10.1074/mcp.T500007-MCP200.

214. Zhang X, Ye J, Jensen ON, Roepstorff P. Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment. *Mol Cell Proteomics*. 2007;6(11):2032-2042. doi:10.1074/mcp.M700278-MCP200.

215. Ruse CI, McClatchy DB, Lu B, et al. Motif-specific sampling of phosphoproteomes. *J Proteome Res*. 2008;7(5):2140-2150. doi:10.1021/pr800147u.

216. Dephoure N, Zhou C, Villén J, et al. A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A*. 2008;105(31):10762-10767. doi:10.1073/pnas.0805139105.

217. Benschop JJ, Mohammed S, O'Flaherty M, Heck AJR, Slijper M, Menke FLH. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol Cell Proteomics*. 2007;6(7):1198-1214. doi:10.1074/mcp.M600429-MCP200.

218. Lemeer S, Jopling C, Gouw J, et al. Comparative phosphoproteomics of zebrafish Fyn/Yes morpholino knockdown embryos. *Mol Cell Proteomics*. 2008;7(11):2176-2187. doi:10.1074/mcp.M800081-MCP200.

219. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods*. 2007;4(3):231-237. doi:10.1038/nmeth1005.

220. Mohammed S, Lorenzen K, Kerkhoven R, et al. Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem*. 2008;80(10):3584-3592. doi:10.1021/ac7024283.

221. Villén J, Beausoleil SA, Gerber SA, Gygi SP. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A*. 2007;104(5):1488-1493. doi:10.1073/pnas.0609836104.

222. Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H. A multidimensional chromatography technology for in-depth phosphoproteome

analysis. *Mol Cell Proteomics*. 2008;7(7):1389-1396. doi:10.1074/mcp.M700468-MCP200.

223. Song C, Ye M, Han G, et al. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Anal Chem*. 2010;82(1):53-56. doi:10.1021/ac9023044.

224. Olsen J V, Vermeulen M, Santamaria A, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal*. 2010;3(104):ra3. doi:10.1126/scisignal.2000475.

225. Bian Y, Ye M, Song C, et al. Improve the coverage for the analysis of phosphoproteome of HeLa cells by a tandem digestion approach. *J Proteome Res*. 2012;11(5):2828-2837. doi:10.1021/pr300242w.

226. Sun Z, Sun D, Wang F, et al. Differential analysis of N-glycoproteome between hepatocellular carcinoma and normal human liver tissues by combination of multiple protease digestion and solid phase based labeling. *Clin Proteomics*. 2014;11(1):26. doi:10.1186/1559-0275-11-26.

227. Nardiello D, Palermo C, Natale A, Quinto M, Centonze D. Strategies in protein sequencing and characterization: multi-enzyme digestion coupled with alternate CID/ETD tandem mass spectrometry. *Anal Chim Acta*. 2015;854:106-117. doi:10.1016/j.aca.2014.10.053.

228. Swaney DL, Wenger CD, Coon JJ. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res*. 2010;9(3):1323-1329. doi:10.1021/pr900863u.

229. Biringer RG, Amato H, Harrington MG, Fonteh AN, Riggins JN, Hühmer AFR. Enhanced sequence coverage of proteins in human cerebrospinal fluid using multiple enzymatic digestion and linear ion trap LC-MS/MS. *Brief Funct Genomic Proteomic*. 2006;5(2):144-153. doi:10.1093/bfgp/ell026.

230. Wang W, Han G, Ye M, Shi H, Zou H, Huo K. Mapping of phosphorylation sites in human MSK1 activated by a novel interaction with MRK-beta. *Electrophoresis*. 2010;31(8):1283-1293. doi:10.1002/elps.200900637.

231. Chen R, Jiang X, Sun D, et al. Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry. *J Proteome Res*. 2009;8(2):651-661. doi:10.1021/pr8008012.

232. Zhou H, Low TY, Hennrich ML, et al. Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol Cell Proteomics*. 2011;10(10):M110.006452.

doi:10.1074/mcp.M110.006452.

233. Zhou H, Ye M, Dong J, et al. Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J Proteome Res*. 2008;7(9):3957-3967. doi:10.1021/pr800223m.

234. Yu Z, Han G, Sun S, et al. Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. *Anal Chim Acta*. 2009;636(1):34-41. doi:10.1016/j.aca.2009.01.033.

235. Raijmakers R, Berkers CR, de Jong A, Ovaa H, Heck AJR, Mohammed S. Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. *Mol Cell Proteomics*. 2008;7(9):1755-1762. doi:10.1074/mcp.M800093-MCP200.

236. Frese CK, Altelaar AFM, Hennrich ML, et al. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res*. 2011;10(5):2377-2388. doi:10.1021/pr1011729.

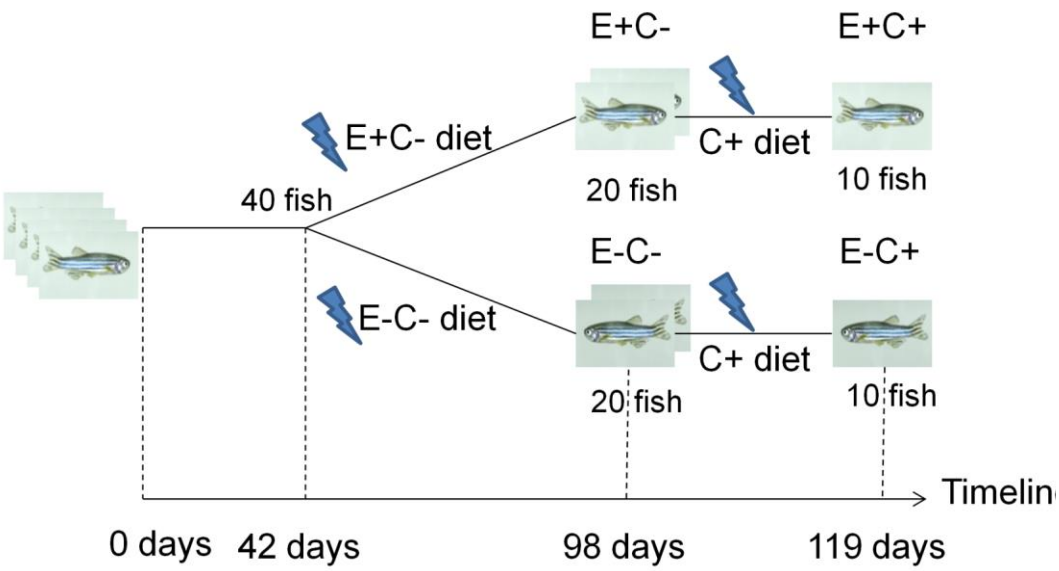
237. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*. 2007;4(11):923-925. doi:10.1038/nmeth1113.

238. Hunter T. The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos Trans R Soc Lond B Biol Sci*. 1998;353(1368):583-605. doi:10.1098/rstb.1998.0228.

239. Macek B, Mijakovic I, Olsen J V, et al. The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis. *Mol Cell Proteomics*. 2007;6(4):697-707. doi:10.1074/mcp.M600464-MCP200.

APPENDICES

APPENDIX A.



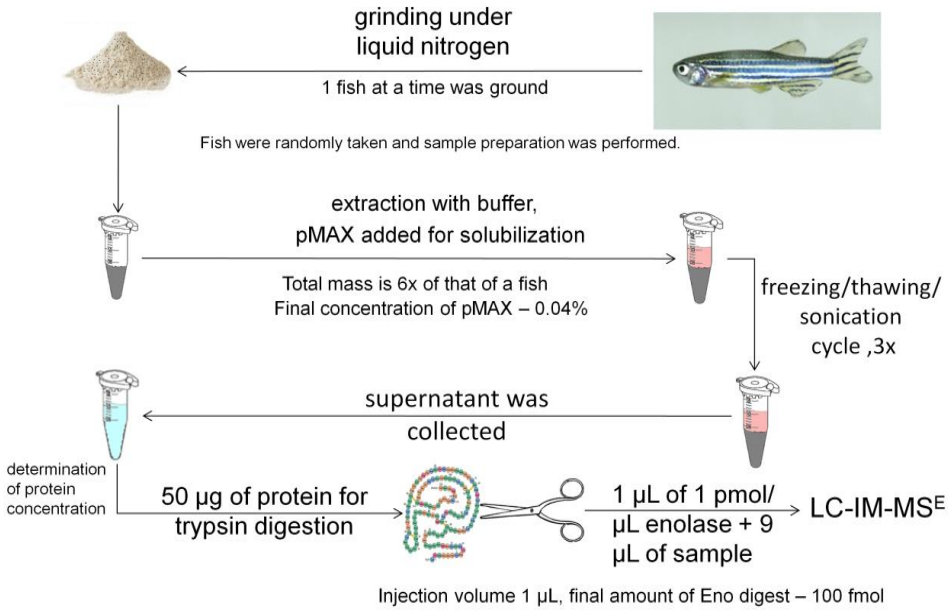
Appendix Figure A.1 Design of the feeding experiment.

Appendix Table A.2 Pipetting scheme used for sample lysis and preparation of digests.

		Solubilization				Digestion		
Sample	Fish weight, mg	Volume of pMAX, uL	Volume of PBS, uL	Total volume, uL	Concentration, ug/mL	Volume for digestion, uL	Mass of proteins for digestion, ug	Volume of NH4HCO3, ul
C+E+ #3	62.0	74	298	372	2080.0	24.0	49.9	69.5
C+E+ #4	115.0	138	552	690	2642.6	18.9	49.9	74.6
C+E+ #5	29.2	35	140	175	1682.1	29.7	50.0	63.8
C+E+ #6	143.4	172	688	860	845.1	59.2	50.0	34.3
C+E+ #7	231.0	277	1109	1386	2423.0	20.6	49.9	72.9
C+E+ #8	207.3	249	995	1244	2217.2	22.6	50.1	70.9
C+E+ #9	242.7	291	1165	1456	2148.6	23.3	50.1	70.2
C+E+ #10	183.5	220	881	1101	1942.8	25.7	49.9	67.8
C+E- #1	127.6	153	612	765	2217.2	22.6	50.1	70.9
C+E- #2	81.2	97	390	487	707.9	70.6	50.0	22.9
C+E- #3	58.9	71	283	354	2477.9	20.2	50.1	73.3
C+E- #4	103.1	124	495	619	2587.7	19.3	49.9	74.2
C+E- #5	124.1	149	596	745	2464.2	20.3	50.0	73.2
C+E- #6	71.8	86	345	431	2697.5	18.5	49.9	75.0
C+E- #7	65.8	79	316	395	2656.3	18.8	49.9	74.7
C+E- #8	207.6	249	996	1245	2258.8	22.1	49.9	71.4
C+E- #9	123.8	149	594	743	2299.5	21.7	49.9	71.8
C+E- #10	215.3	258	1033	1291	2381.9	21.0	50.0	72.5
C-E+ #1	32.5	39	156	195	2450.5	20.4	50.0	73.1
C-E+ #2	201.9	242	969	1211	2313.3	21.6	50.0	71.9
C-E+ #3	168.8	203	810	1013	2258.4	22.1	49.9	71.4
C-E+ #4	41.1	49	197	246	2409.3	20.8	50.1	72.7
C-E+ #5	70.1	84	336	420	2464.2	20.3	50.0	73.2
C-E+ #6	168.2	202	807	1009	2327.0	21.5	50.0	72.0
C-E+ #7	51.1	61	245	306	2519.1	19.8	49.9	73.7
C-E+ #8	42.7	51	205	256	2656.3	18.8	49.9	74.7
C-E+ #9	73.1	88	351	439	2327.0	21.5	50.0	72.0
C-E+ #10	76.9	92	369	461	2368.2	21.1	50.0	72.4
C-E- #1	72.4	87	348	435	1983.9	25.2	50.0	68.3
C-E- #2	14.9	18	72	90	2711.2	18.4	49.9	75.1

Appendix Table A.2 (Continued) Pipetting scheme used for sample lysis and preparation of digests.

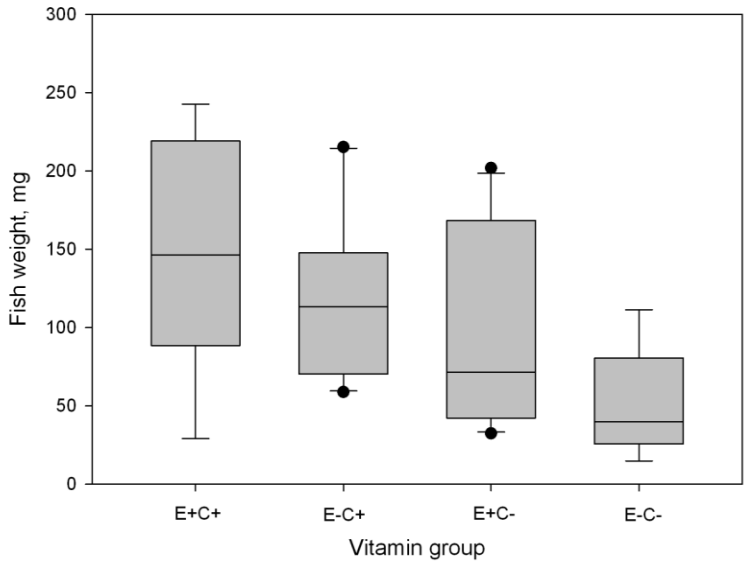
C-E- #3	87.8	105	421	526	2409.3	20.8	50.1	72.7
C-E- #4	34.8	42	167	209	2464.2	20.3	50.0	73.2
C-E- #5	40.0	48	192	240	2423.0	20.6	49.9	72.9
C-E- #6	19.7	24	95	119	2327.0	21.5	50.0	72.0
C-E- #7	73.4	88	352	440	2532.8	19.7	49.9	73.8
C-E- #8	31.7	38	152	190	2834.7	17.6	49.9	75.9
C-E- #9	111.4	134	535	669	2299.5	21.7	49.9	71.8



Appendix Figure A.3 Workflow used for preparing the samples for the label-free "bottom-up" quantitative proteomics study.

Appendix Table A.4 Synapt G2 HDMS instrument settings used for Ion-Mobility-MS^{E} acquisitions.

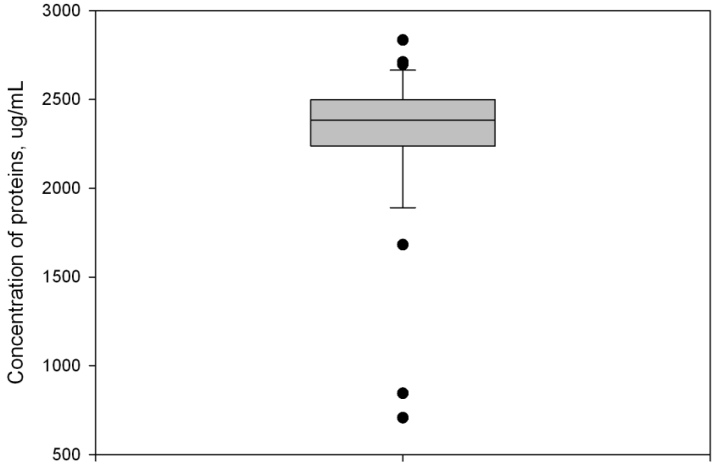
Capillary voltage, kV	2.5
Sampling cone	40
Extraction cone	4.2
Source temperature, ºC	40
Desolvation temperature, ºC	300
Cone gas flow, L/Hr	20
Nanoflow gas pressure, Bar	1
Purge gas flow, mL/Hr	100
Desolvation gas flow, L/Hr	100
IMS gas flow, mL/min	75.30
Reference scan frequency, sec	60
LockSpray capillary, kV	3.0
Use manual trap DC	FALSE
Use manual IMS DC	FALSE
Use manual transfer DC	FALSE
Source manual control	OFF
Trap manual control	OFF
IMS manual control	OFF
Transfer manual control	OFF
Function parameters – Function 1	
Trap collision energy, eV	6.0
Transfer collision energy, eV	0.0
Maintain mobility separation	YES
Function parameters – Function 2	
Using auto trap MS collision energy, eV	4.0
Transfer MS collision energy low, eV	27.0
Transfer MS collision energy high, eV	50.0
Maintain mobility separation	YES
Function parameters – Function 3 (Lock Mass)	
Using auto trap MS collision energy, eV	4.0
Using auto transfer MS collision energy, eV	0.0
Maintain mobility separation	YES



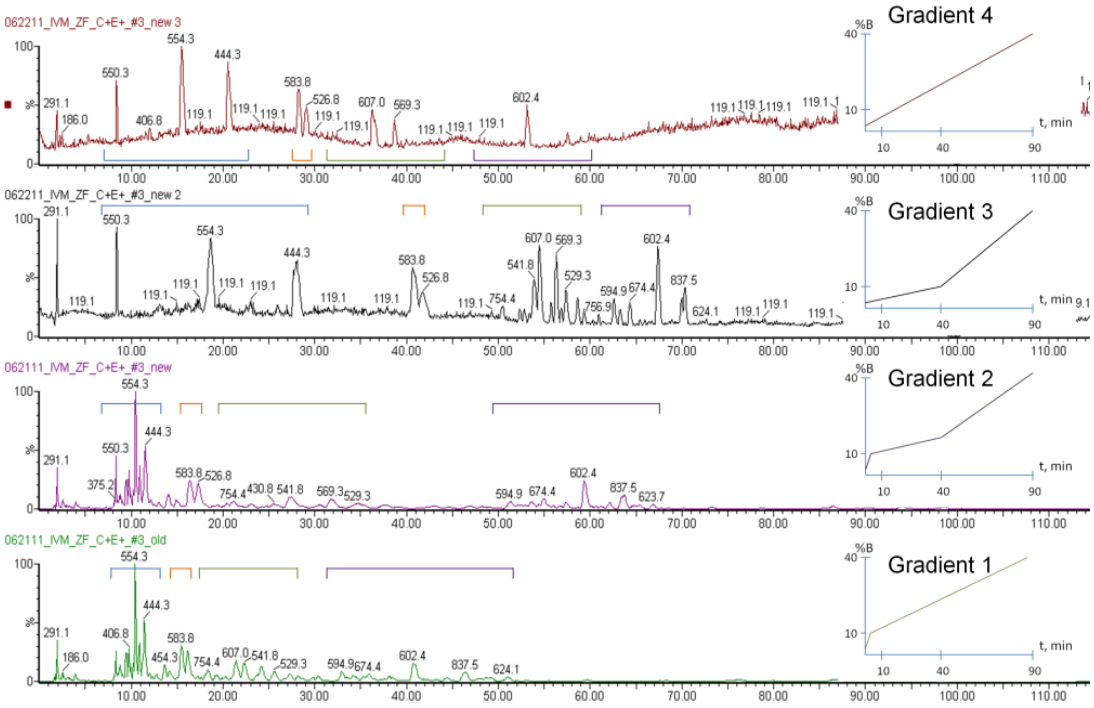
Appendix Figure A.5 Fish characteristics dependent on the vitamin regime fed.

Appendix Table A.6 Fish characteristics dependent on the vitamin regime fed.

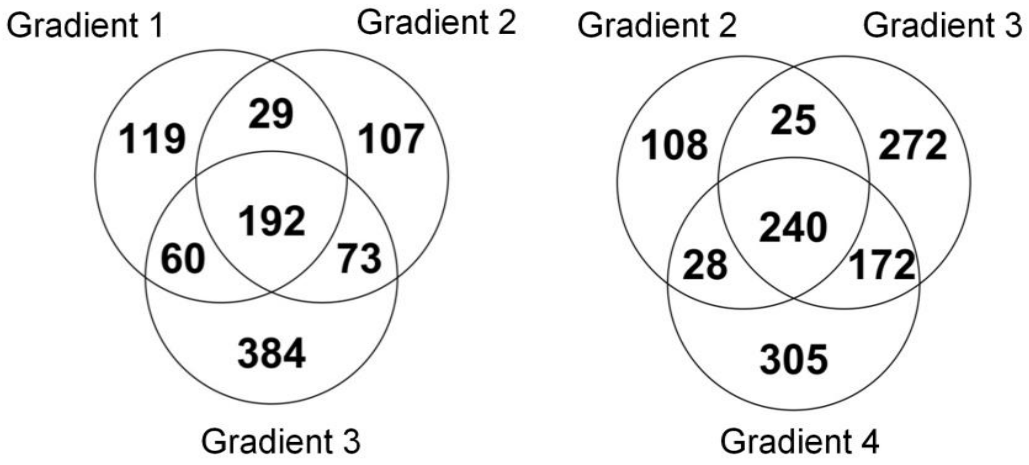
	E+C+	E-C+	E+C-	E-C-
E+C+	1.00			
E-C+	1.29	1.00		
E+C-	1.68	1.30	1.00	
E-C-	3.02	2.34	1.79	1.00



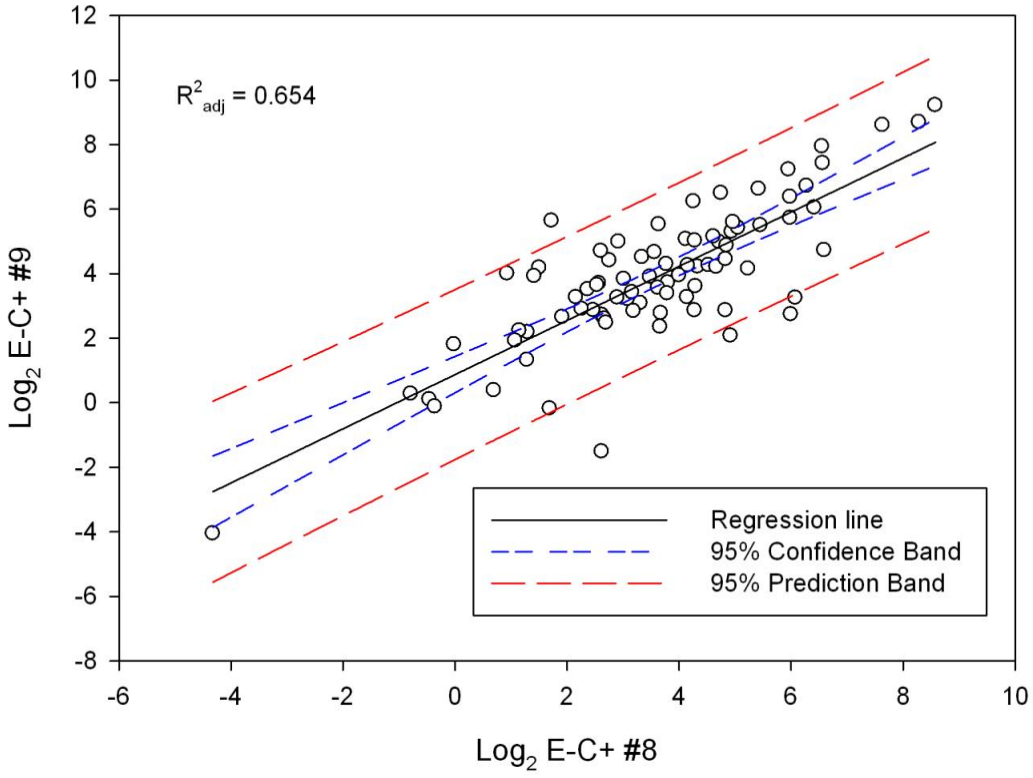
Appendix Figure A.7 A boxplot of the distribution of the total protein concentration after extraction and lysis.



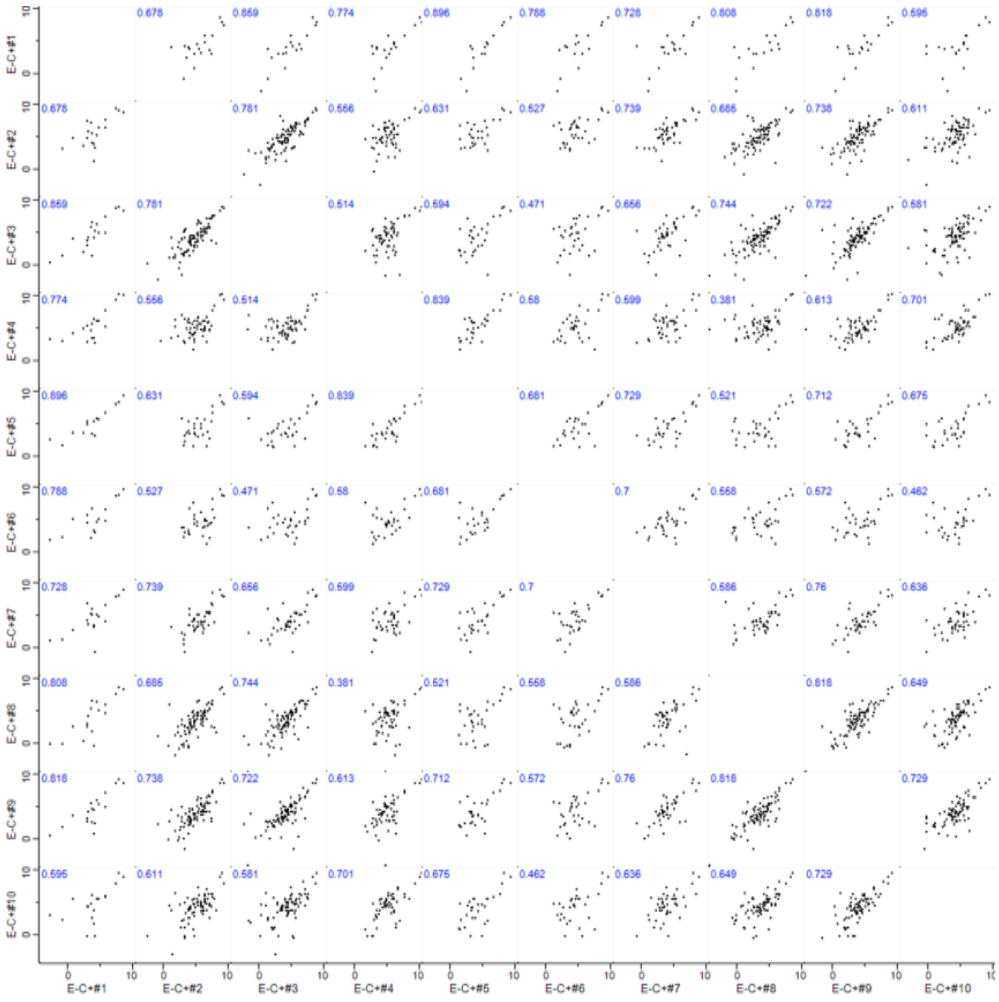
Appendix Figure A.8 Gradient profiles with respective base peak chromatograms.



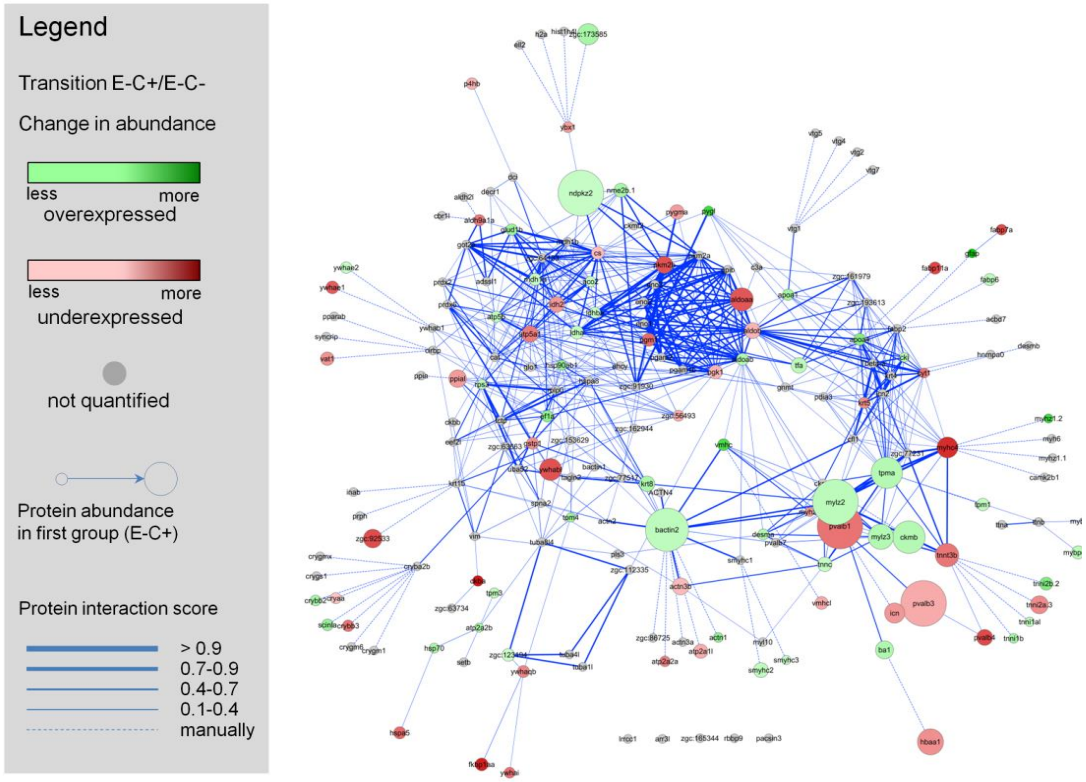
Appendix Figure A.9 Venn diagrams comparing the number of peptides eluting from a column and identified with LC−IM−MS^E.



Appendix Figure A.10 Correlation plot between E−C+ sample 9 and E−C+ sample 8.

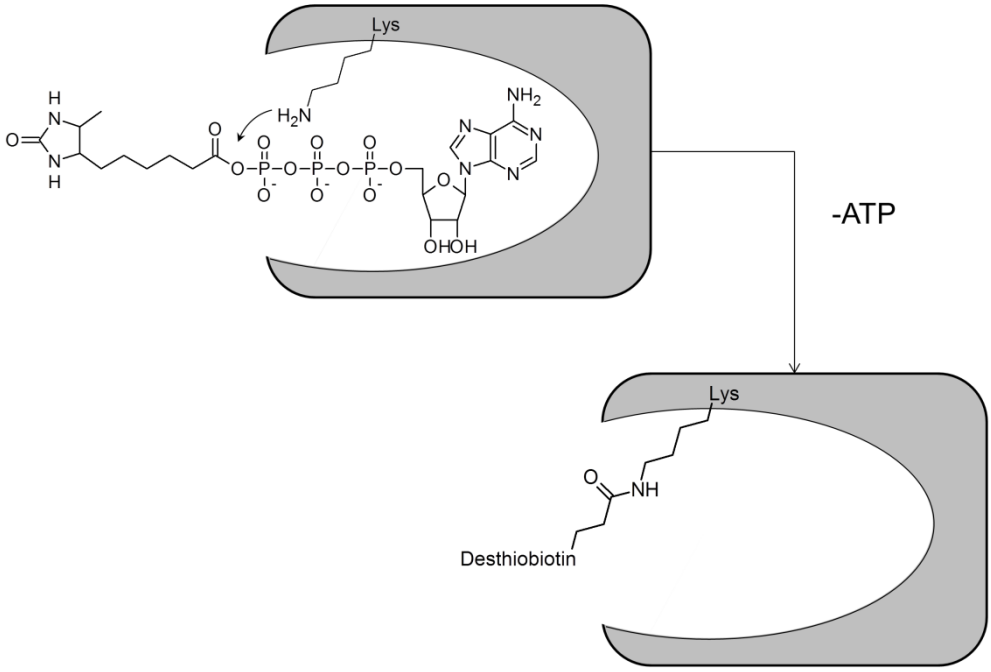


Appendix Figure A.11 Correlation plot matrix of the E−C+ group.

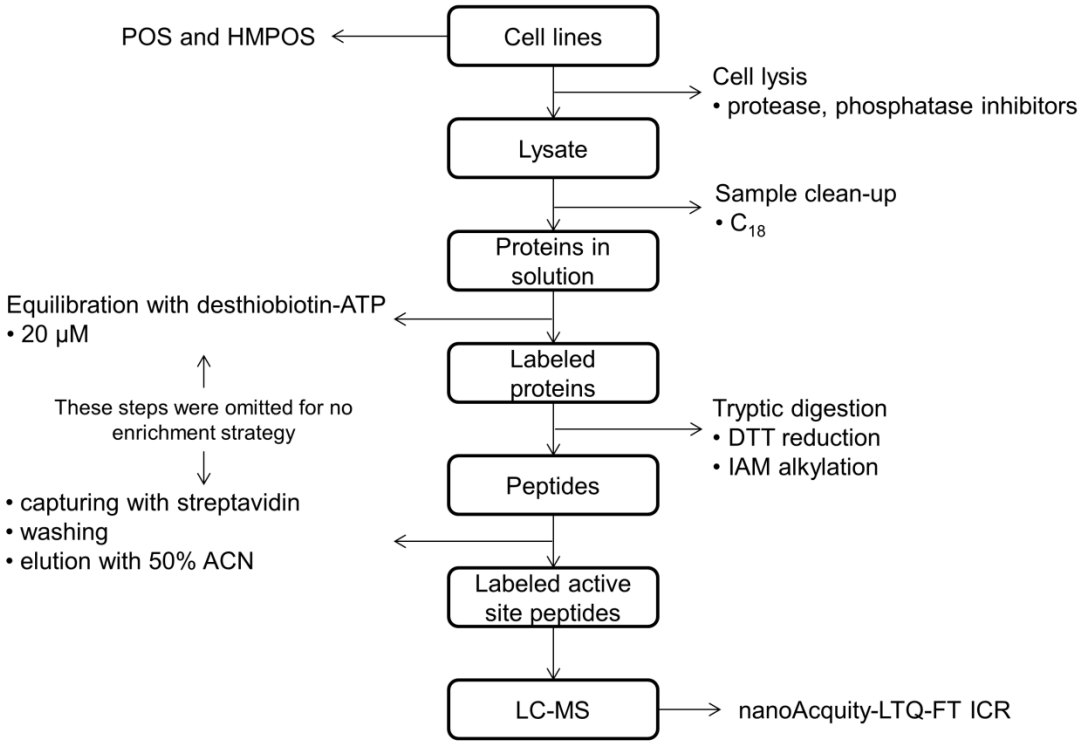


Appendix Figure A.12 The responsive protein network for the vitamin transition C+/C−and both groups deficient in vitamin E status (E−C+ vs E−C+).

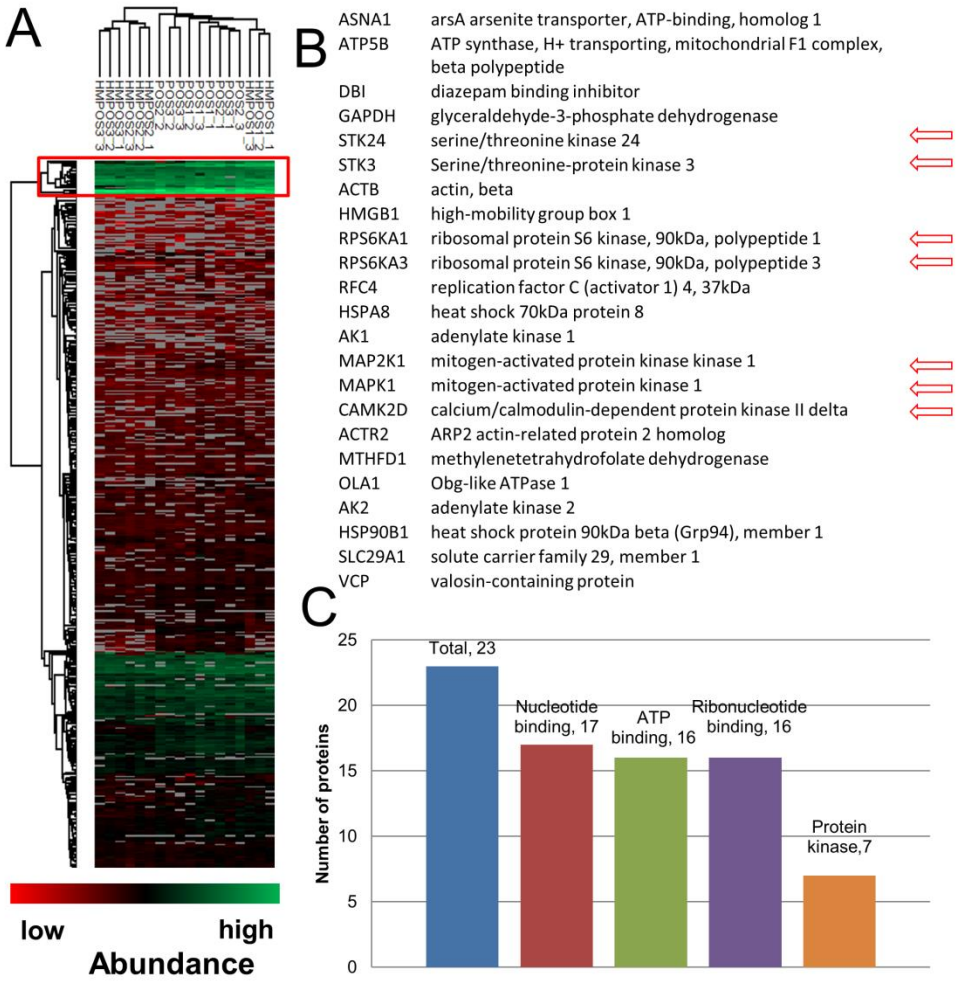
APPENDIX B



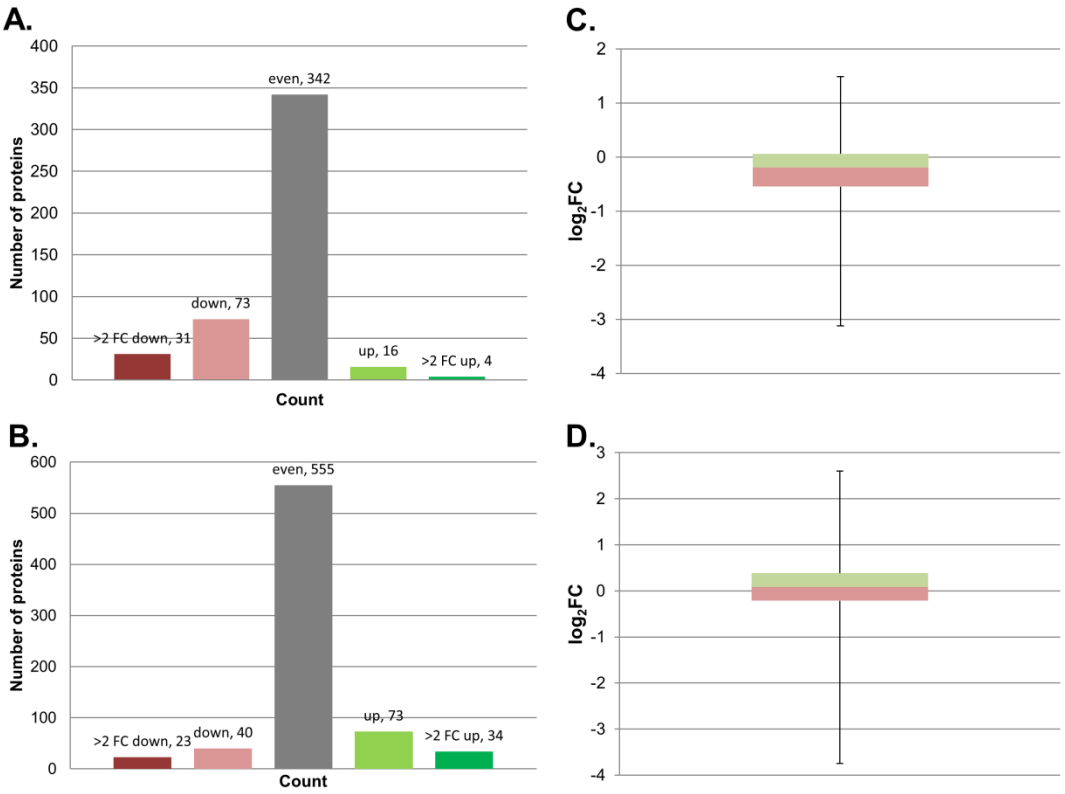
Appendix Figure B.1 The probe used for enriching ATP-binding proteins



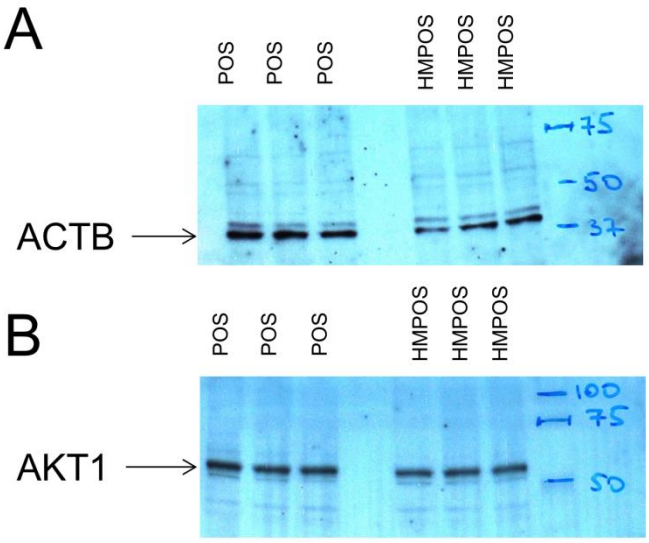
Appendix Figure B.2 The schematics of ATPome and FP workflows



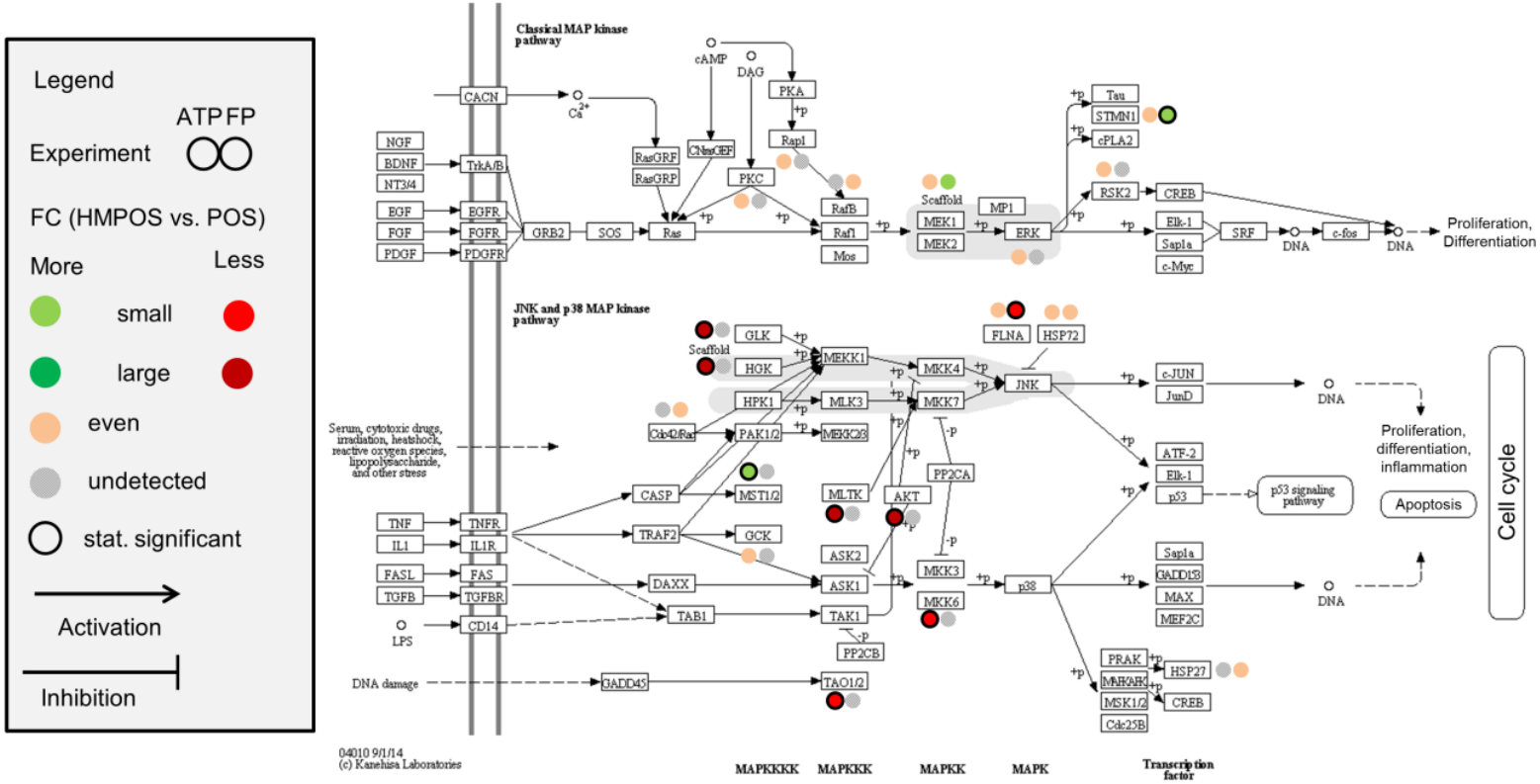
Appendix Figure B.3 The most abundant proteins after ATPome enrichment



Appendix Figure B.4 Distributions of protein abundance fold changes between cell lines

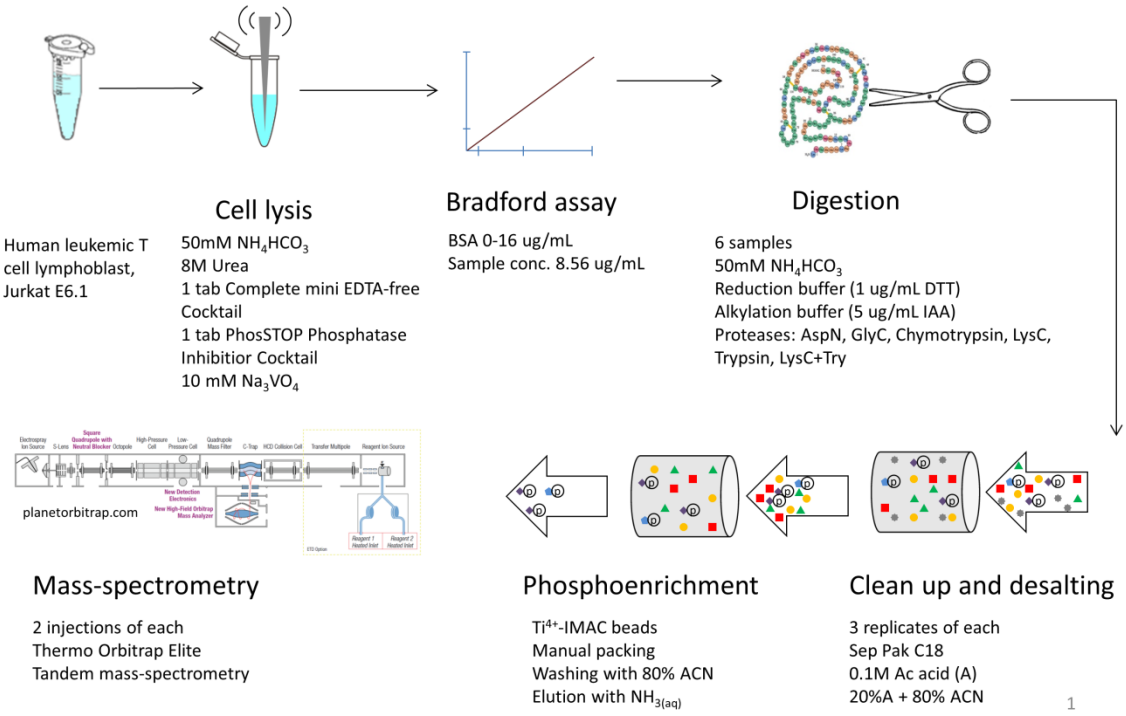


Appendix Figure B.5 Western blotting against ACTB (A) and AKT1 (B).



Appendix Figure B.6 Observed proteins and their level changes mapped on the MAPK signaling cascade.

APPENDIX C.



Appendix Figure C.1 Experimental workflow